**Oleksandr Kryvoshein**
*ORCID: 0000-0001-5029-4228*
*krivoshein@uhmi.org.ua*

**Oleksii Kryvobok**
*ORCID: 0000-0002-1730-1809*
*kryvobok@uhmi.org.ua*

**Olena Zabolotna**
*ORCID: 0009-0009-6338-7672*
*zabolotnaolena@ukr.net*

*Ukrainian Hydrometeorological Institute of the State Emergency of Ukraine and the National Academy of Sciences of Ukraine, Kyiv*

УДК 551.509.313:004.942(477)

DOI: https://doi.org.ua/10.15407/Meteorology2025.08.092

# VERIFICATION OF THE ICON NUMERICAL WEATHER PREDICTION MODEL IN UKRAINE

*This study presents a comprehensive verification of the ICON numerical weather prediction model over Ukraine for the year 2024. The evaluation covers key meteorological parameters — air temperature, wind speed, relative humidity, precipitation, and cloud cover — at 24-, 48-, and 72-hour forecast lead times. Both continuous metrics (correlation, mean absolute error, root mean square error, bias) and categorical metrics (POD, FAR, CSI) were applied, along with seasonal and spatial analyses. The model demonstrated high accuracy in forecasting mean temperature, with a correlation coefficient of $r = 0.95$ at 24 hours, low RMSE (≈2.6°C), and near-zero bias. Cloud cover forecasts also showed excellent performance, with POD > 0.94 and CSI up to 0.86 at a 10% threshold, maintaining stability across regions and seasons. By contrast, wind speed forecasts were less reliable, with lower correlations ($r = 0.40$ at 24 h), RMSE ~1.75 m/s, and consistent overestimation. Forecasts of relative humidity were moderately accurate ($r = 0.88$), although a persistent negative bias (≈–4.2%) was observed. Precipitation forecasts exhibited the lowest skill, especially at longer lead times and higher thresholds. At a 0.1 mm threshold and 24-hour forecast, POD reached 0.61, but FAR remained high (>0.50), particularly in southern regions with frequent convective activity. Seasonal analysis indicated the best model performance in autumn and winter, with reduced accuracy in summer, especially for humidity and precipitation. Spatial verification at 24-hour lead time revealed regional differences: the lowest RMSE for mean temperature was found in Kherson (2.29°C), while the highest wind speed error occurred in Donetsk (4.88 m/s). Overall, the ICON model provides robust forecasts for temperature and cloud cover, acceptable performance for humidity, and highlights the need for further refinement in wind and precipitation prediction. These findings offer valuable guidance for improving regional forecast applications and adjusting physical parameterizations under Ukrainian climate and topography conditions.*

**Keywords:** *ICON model, verification, weather prediction, forecast accuracy.*

## INTRODUCTION

Numerical Weather Prediction (NWP) models are fundamental tools in modern meteorology, providing critical forecasts that support public safety, economic activities, and environmental management worldwide (Kalnay, 2003; Bauer et al., 2015). The ICON (ICOsahedral Nonhydrostatic) model, developed by the German Weather Service (Deutscher Wetterdienst, DWD), represents a state-of-the-art global and regional forecasting system that leverages advanced grid structures and physics parameterizations to improve forecast accuracy across scales (Zängl et al., 2015; Dipankar et al., 2015).

In recent years, the ICON model has been used for operational regional forecasting in Ukraine, a country characterized by complex orography, significant climatic variability, and transitional weather regimes between continental and maritime influences (Shyian et al., 2018; Holoborodko et al., 2020). Previous studies have examined the application and verification of regional NWP models in Ukraine, highlighting methods for forecast evaluation and interpolation (Shpyg & Budak, 2015; Doroshenko et al., 2020). These factors pose unique challenges for NWP, requiring rigorous model verification to ensure reliable weather predictions, especially under extreme and high-impact weather events such as severe storms, heatwaves, and winter weather (Ebert et al., 2016).

Moreover, ongoing geopolitical and logistical issues have impacted the density and quality of observational data in certain Ukrainian regions, complicating the data assimilation process and verification efforts (Ukrainian Hydrometeorological Center, 2023). Against this backdrop, an in-depth assessment of the ICON model's forecast skill over Ukraine for 2024 is essential. This study aims to systematically evaluate the ICON model's performance across multiple meteorological parameters, temporal scales, and spatial domains, providing insights into its strengths, limitations, and opportunities for further development.

By comparing ICON forecasts with extensive observational datasets from 133 meteorological stations, this work contributes to enhancing NWP reliability in Ukraine and serves as a reference for similar mid-latitude regions with diverse climatic and topographic settings. The results also provide a foundation for targeted model improvements, bias correction strategies, and informed decision-making in operational meteorology.

## METHODOLOGY

**Data and Parameters.** Forecast verification was based on daily observations from 133 surface meteorological stations across Ukraine (*Fig. 1*). The ICON-EU configuration was applied with a horizontal resolution of 6.5 × 6.5 km. The computational domain covered the territory of Ukraine, approximately between 44°–53° N and 22°–41° E. The model data were obtained from the DWD open data portal (https://opendata.dwd.de/). The parameters verified included: minimum, maximum, and mean 2 m air temperature (Tmin, Tmax, Tmean); 10 m wind speed; relative humidity at 2 m; total cloud cover; daily precipitation totals.

**Verification Framework.** Forecasts at 24, 48, and 72-hour lead times were evaluated using continuous and categorical metrics depending on parameters (in this research, metrics applied to continuous variables are referred to as *continuous metrics*, while those applied to categorical variables are referred to as *categorical metrics*).

To evaluate the accuracy of temperature, wind speed, and humidity forecast, the following metrics were employed.

*Pearson Correlation Coefficient* (*r*)

$$r = \frac{\sum_{i=1}^{N}\left(P_i - \bar{P}\right)\left(O_i - \bar{O}\right)}{\sqrt{\sum_{i=1}^{N}\left(P_i - \bar{P}\right)^2 \sum_{i=1}^{N}\left(O_i - \bar{O}\right)^2}},$$

where $N$ is the number of observations; $P_i$ — the predicted value; $O_i$ — the observed value; $\bar{P}$ — the mean of predictions; $\bar{O}$ — the mean of observations.

*Mean Absolute Error* (*MAE*)

$$MAE = \frac{1}{N}\sum_{i=1}^{N}\left|P_i - O_i\right|,$$

where $N$ is the number of observations; $O_i$ — the observed value; $P_i$ — the predicted value.

*Root Mean Square Error* (*RMSE*)

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(P_i - O_i\right)^2},$$

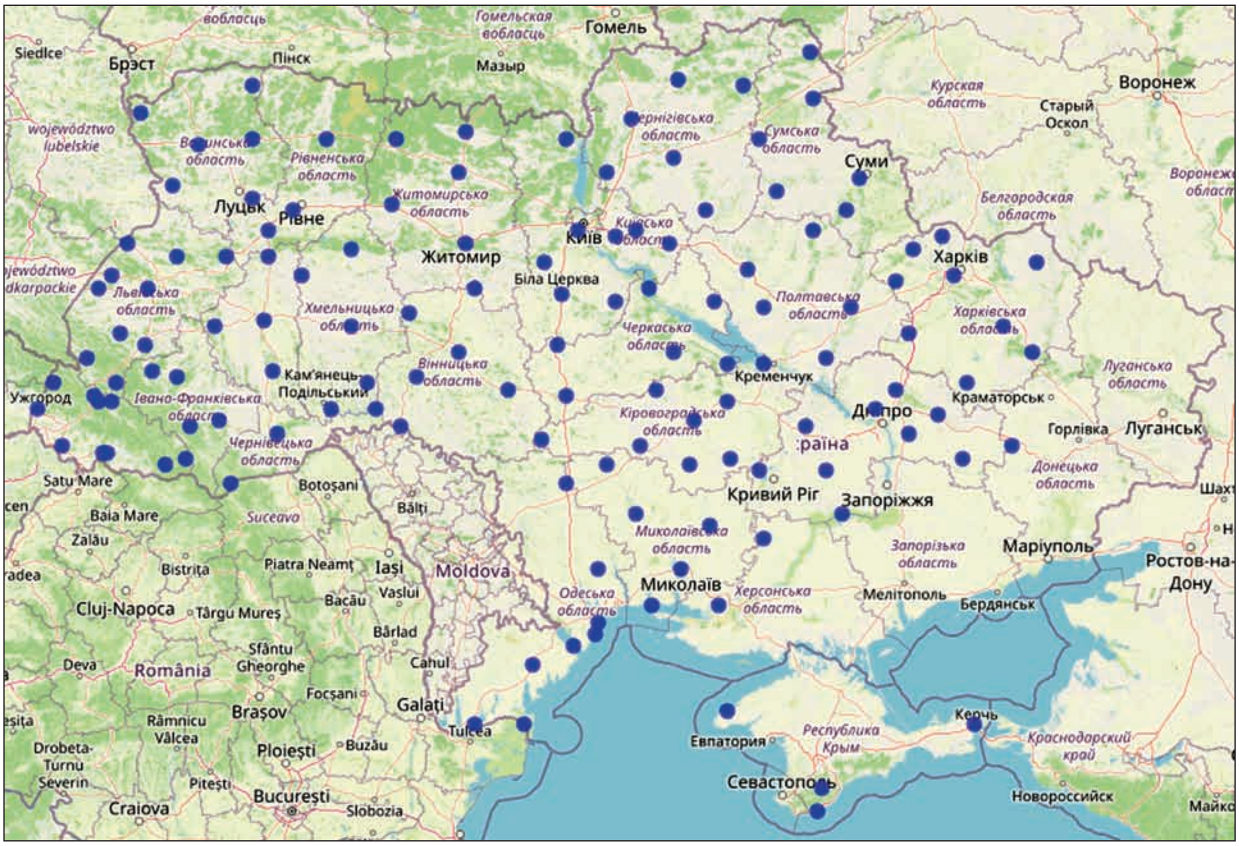

**Fig. 1.** Locations of meteorological network stations used for the verification of the ICON model data

where $N$ is the number of observations; $O_i$ — the observed value; $P_i$ — the predicted value.

*Bias*

$$Bias = \frac{1}{N}\sum_{i=1}^{N}\left(P_i - O_i\right),$$

where $N$ is the number of observations; $P_i$ — the predicted value; $O_i$ — the observed value.

For assessing the quality of forecasts of binary event variables, such as the presence or absence of precipitation or cloud cover, a set of metrics based on the confusion matrix were applied. These metrics include:

- *Probability of Detection* (POD): Measures the likelihood of correctly forecasting the occurrence of an event.
- *False Alarm Ratio* (FAR): Represents the proportion of false alarms among all forecasted events.
- *Critical Success Index* (CSI): An integrated metric accounting for both correct and incorrect forecasts, providing a comprehensive measure of forecast performance.

To clarify, the confusion matrix elements for binary classification are:

- *True Positives* (TP): Number of cases where the event occurred and was correctly forecasted.
- *False Positives* (FP): Number of cases where the event did not occur but was incorrectly forecasted.
- *True Negatives* (TN): Number of cases where the event did not occur and was correctly forecasted as such.
- *False Negatives* (FN): Number of cases where the event occurred but was not forecasted.

The metrics are calculated as follows:

$$POD = \frac{TP}{TP + FN}.$$

Values range from 0 to 1, where 1 indicates a perfect forecast.

$$FAR = \frac{FP}{FP + TP}.$$

Values range from 0 to 1, where 0 indicates a perfect forecast (no false alarms).

$$CSI = \frac{TP}{TP + FP + FN}.$$

Values range from 0 to 1, with 1 indicating a perfect forecast. CSI is a more stringent measure than POD as it accounts for all three types of errors related to event forecasting.

Verification was performed:

- For all meteorological stations collectively (overall)
- By seasons (spring, summer, autumn, winter)
- By regions of Ukraine (spatial verification)

## RESULTS AND DISCUSSION

**Overall verification results.**

*Continuous variables*. For each forecast horizon (24 hours, 48 hours, 72 hours), the average values of the accuracy metrics were calculated. *Table 1* presents the verification results for temperature, wind speed, and relative humidity.

The model demonstrated generally high accuracy in forecasting temperature indicators. For mean temperature, the correlation between forecasted and actual values was $r = 0.95$ at the 24-hour horizon and decreased to $r = 0.90$ at 72 hours. The mean absolute error (MAE) remained within 2–3°C, and the bias was close to zero (ranging from –0.1 to –0.01°C), indicating minimal systematic error.

Similar results were obtained for maximum temperature, where the correlation also declined from $r = 0.95$ (24 hours) to $r = 0.90$ (72 hours), although the bias tended toward underestimation (up to –0.21°C).

For minimum temperature, the correlation was slightly lower ($r = 0.85$ at 72 hours), and the bias was positive, meaning the model tended to slightly overestimate minimum values.

In the case of wind speed, forecast quality was significantly lower: the correlation coefficient dropped from $r = 0.40$ at 24 hours to $r = 0.18$ at 72 hours, indicating a weak relationship between forecast and observation. Meanwhile, the MAE was approximately 1.3–1.5 m/s, and the bias remained consistently positive (around +0.68 m/s), pointing to a systematic overestimation of wind speed.

The forecast for relative humidity showed fairly stable performance: the correlation decreased from $r = 0.88$ to $r = 0.78$ over the 72-hour horizon, and the MAE remained steady at about 6.3% across all time ranges. However, a persistent negative bias (approximately –4.2%) was observed, indicating a tendency to underpredict humidity levels.

*Discrete (event-based) variables*. Since for event-based variables such as precipitation and cloud cover, threshold values must be defined to distinguish the occurrence of an event (“yes” / “no”) before evaluating forecast accuracy, a sensitivity analysis was conducted to assess the model’s response to different threshold selections. Specifically, for precipitation, three threshold values were tested: 0.1 mm, 0.5 mm, and 1.0 mm. For cloud cover, the tested thresholds were: 10%, 30%, and 50%. These thresholds were chosen to evaluate the model’s ability to detect the presence of cloudiness across different levels of intensity. Each threshold represents a separate binary classification task: 10% serves to test whether the model can capture any cloud formation, 30% corresponds to moderate cloudiness, and 50% targets the detection of dense

Table 1. **Overall accuracy rating for temperature, wind, and humidity**

| Variable | Metric | 24 h | 48 h | 72 h |
|---|---|---|---|---|
| Temperature (min), °C | *r* | 0.92 | 0.88 | 0.85 |
| | *MAE* | 2.34 | 2.94 | 3.25 |
| | *RMSE* | 3.11 | 3.84 | 4.23 |
| | *Bias* | 0.32 | 0.26 | 0.23 |
| Temperature (max), °C | *r* | 0.95 | 0.92 | 0.9 |
| | *MAE* | 2.35 | 3.2 | 3.59 |
| | *RMSE* | 3.18 | 4.16 | 4.62 |
| | *Bias* | -0.12 | -0.17 | -0.21 |
| Temperature (avg), °C | *r* | 0.95 | 0.92 | 0.9 |
| | *MAE* | 1.97 | 2.68 | 3.03 |
| | *RMSE* | 2.61 | 3.49 | 3.92 |
| | *Bias* | 0.1 | 0.04 | 0.01 |
| Wind (м/с) | *r* | 0.4 | 0.25 | 0.18 |
| | *MAE* | 1.33 | 1.45 | 1.49 |
| | *RMSE* | 1.75 | 1.91 | 1.97 |
| | *Bias* | 0.69 | 0.68 | 0.67 |
| Humidity (%) | *r* | 0.88 | 0.82 | 0.78 |
| | *MAE* | 6.29 | 6.31 | 6.3 |
| | *RMSE* | 7.88 | 7.91 | 7.92 |
| | *Bias* | –4.22 | –4.22 | –4.21 |

or overcast conditions. This approach is more robust than multiclass classification, where even small differences between observed and forecasted values (e.g., 40% vs. 50%) may be penalized as errors. Instead, the threshold-based binary evaluation better reflects practical needs in real-world applications, where users are typically interested in general sky conditions rather than exact numerical cloud cover percentages.

Based on *Table 2*, it is evident that for precipitation forecasts at all thresholds, there is a decline in the Probability of Detection (POD) and the Critical Success Index (CSI) with increasing forecast lead time, indicating

Table 2. **Overall accuracy rating for precipitation and cloudiness**

| Variable | Threshold | Metric | 24 h | 48 h | 72 h |
|---|---|---|---|---|---|
| Precipitation | 0.1 mm | POD | 0.61 | 0.53 | 0.49 |
| | | FAR | 0.52 | 0.58 | 0.60 |
| | | CSI | 0.36 | 0.30 | 0.28 |
| | 0.5 mm | POD | 0.50 | 0.41 | 0.38 |
| | | FAR | 0.56 | 0.64 | 0.66 |
| | | CSI | 0.30 | 0.23 | 0.21 |
| | 1.0 mm | POD | 0.43 | 0.33 | 0.31 |
| | | FAR | 0.64 | 0.71 | 0.73 |
| | | CSI | 0.24 | 0.18 | 0.16 |
| Cloudiness | 10% | POD | 0.94 | 0.93 | 0.92 |
| | | FAR | 0.09 | 0.10 | 0.11 |
| | | CSI | 0.86 | 0.83 | 0.82 |
| | 30% | POD | 0.88 | 0.84 | 0.84 |
| | | FAR | 0.19 | 0.22 | 0.22 |
| | | CSI | 0.73 | 0.68 | 0.67 |
| | 50% | POD | 0.8 | 0.74 | 0.72 |
| | | FAR | 0.31 | 0.35 | 0.37 |
| | | CSI | 0.58 | 0.52 | 0.50 |

a reduction in accuracy over time. The highest POD (0.61) and CSI (0.36) values were observed at the 0.1 mm threshold for the 24-hour forecast. However, even at this lowest threshold, the False Alarm Ratio (FAR) remained relatively high (ranging from 0.52 to 0.60), suggesting low quality of precipitation occurrence estimation. As the threshold increased to 0.5 mm and 1.0 mm, POD and CSI gradually decreased, while FAR increased, indicating a reduced ability of the model to detect precipitation, when the value of threshold increased.

For cloud cover forecasts, three threshold values were used — 10%, 30%, and 50% — to define the event "cloudy" which was considered to occur when the corresponding threshold was exceeded (*Table 2*). At all thresholds, the model demonstrated high POD values, indicating a good ability to detect cloudy conditions, with the highest value (0.94) recorded at the 10% threshold for the 24-hour forecast. The FAR remained low — between 0.09 and 0.11 for the lowest threshold — gradually increasing to 0.37 at the 50% threshold for the 72-hour forecast. The CSI was also highest at the 10% threshold, reaching up to 0.86 for the 24-hour forecast, with only a slight decline over time. Overall, the model performs best in forecasting even minimal cloud cover, while accuracy decreases somewhat for higher cloud density. The model predicts cloudiness more accurately at lower thresholds—for example, when detecting as little as 10% cloud cover. However, its predictive accuracy decreases as the threshold increases, with noticeably reduced performance at higher levels such as 50% cloud cover.

Compared to precipitation forecasts, the model demonstrates significantly higher accuracy for cloud cover prediction. Notably, the POD values for cloud cover are considerably higher (up to 0.94 at a 24-hour lead time and a 10% threshold), and FAR values are lower (ranging from 0.09 to 0.37), indicating a strong ability to detect cloud presence with a relatively low rate of false alarms. CSI values for cloud cover remain high even at longer forecast horizons — for instance, up to 0.82 at 72 hours with a 10% threshold — whereas for precipitation, CSI did not exceed 0.36 even at the most favorable threshold (0.1 mm) and the shortest forecast horizon (24 hours). This discrepancy can be explained by more complex physical processes responsible for precipitation formation in clouds than cloud formation itself. Generally, the higher uncertainty in precipitation forecasts compared to cloud forecasts stems from a combination of factors: the extreme sensitivity of precipitation formation to tiny-scale microphysical processes, the inherent difficulty of representing sub-grid convection and the highly localized nature of precipitation events. The Icon model for cloud forecast uses well-known and widely-used diagnostic cloud cover scheme (Sundqvist et al., 1989), its fundamental purpose is to determine the fractional cloud cover within a grid cell based on the specifically relative humidity. Our results show that relative humidity is forecasted quite well at the surface level (Tab. 1), and we therefore assume a similar accuracy at higher levels; however, this interpretation should be regarded as a reasonable assumption rather than a verified relationship. Forecasting precipitation with the ICON model is a more complex process that relies on a combination of the model's high-resolution dynamical core and sophisticated physical parameterizations, including cloud condensation and ice nucleation, evaporation, melting and sedimentation (Giorgetta et al., 2018). Thus, cloud cover forecasts are more reliable and consistent, whereas precipitation forecasts remain more challenging due to the complex and localized nature of rainfall events.

**Seasonal Verification.** Seasonal verification of the ICON model forecasts revealed that forecast accuracy strongly depends on both the type of meteorological parameter and the time of year. The highest forecast quality was observed at shorter lead times (24 hours), while errors increased and correlations with observed values decreased as the lead time extended.

Temperature-related parameters ($T_{\min}$, $T_{\text{mean}}$, $T_{\max}$) exhibited the greatest stability and accuracy, particularly in autumn. However, the model showed a tendency to overestimate temperatures during the cold season and to underestimate them in autumn.

Wind speed forecasts demonstrated the lowest predictive skill, with correlation coefficients significantly lower than for other parameters. The model systematically overestimated wind speed across all seasons.

Relative humidity was reproduced with acceptable accuracy during spring, autumn, and winter, but showed significant underestimation in summer (Bias up to +12.5%), indicating a need for additional model correction in this season.

The evaluation of event-based parameters revealed substantial seasonal differences. For precipitation, the model performed best in winter (POD up to 0.71, CSI up to 0.48) and worst in autumn. Warm seasons were characterized by a high rate of false alarms (FAR > 0.6), reducing the reliability of precipitation forecasts. A single threshold of 0.1 mm was used in this analysis, as it yielded the best results during overall verification.

In contrast, cloud cover forecasts showed the highest consistency and accuracy among all parameters. POD values exceeded 0.89 across all seasons, FAR remained low (especially in winter, down to 0.03), and CSI reached 0.94. These results indicate a very high capability of the ICON model to reliably predict cloud cover, with no significant degradation at longer lead times.

Overall, the ICON model demonstrated high accuracy in forecasting temperature and cloud cover, satisfactory performance for humidity (except in summer), limited skill for wind, and seasonally dependent performance for precipitation. These findings provide a foundation for further spatial analysis of forecast errors, enabling a more detailed assessment of the model's regional behavior.

**Spatial Verification (24-hour Forecasts).** To identify the spatial characteristics of forecast accuracy in the ICON model, a regional evaluation of verification metrics was performed by grouping observational stations at the oblast (regional) level. This approach makes it possible to pinpoint geographic areas with the highest and lowest forecast quality. For ease of visual analysis, the results are presented as thematic maps (fig.2), where each oblast is characterized by the average values of RMSE metric computed from all stations within its territory.

This spatial approach is particularly important given the heterogeneous climatic and topographic conditions that can significantly influence forecast accuracy. Furthermore, based on findings from previous verification stages — which showed a consistent decline in forecast quality with increasing lead time (notably, accuracy was lower for the 48-hour forecast compared to 24 hours, and even lower for 72 hours) — the spatial analysis was limited to the 24-hour lead time. This decision allows the focus to remain on the most reliable forecast interval, where the model exhibits the highest predictive skill.

It should also be noted that due to the absence of ground-based meteorological observations in the Zaporizhzhia and Luhansk regions as a result of ongoing hostilities, these areas were excluded from the spatial verification.

***Temperature.*** The spatial verification of minimum temperature forecasts produced by the ICON model revealed certain regional differences in prediction accuracy. Overall, the model demonstrates a high level of consistency in forecast quality across most of Ukraine, with a few regional exceptions. The root mean square error (RMSE) values for the majority of regions range between 2.8°C and 3.3°C, indicating reasonably good forecast accuracy. The lowest RMSE values were recorded in Poltava (2.65°C), Kherson (2.78°C), and Khmelnytskyi (2.83°C) oblasts, suggesting stable conditions for minimum temperatures and a relatively good adaptation of the model to local climatic features. The highest RMSE values were observed in Donetsk (5.69°C) and Crimea (4.28°C). These elevated errors may be attributed not only to more complex synoptic conditions (particularly over Crimea) but also to the limited availability of observational data (only one station in the Donetsk region) resulting from the ongoing military situation in these areas.

For the maximum temperature the results indicate stable and relatively high forecast accuracy by the ICON model across most regions of Ukraine. RMSE values generally range from 2.8°C to 3.4°C, which can be considered a favorable outcome for short-term forecasting. The lowest RMSE values were found in Mykolaiv (2.76°C), Kherson (2.80°C), and Cherkasy (2.88°C) oblasts, reflecting a high degree of consistency between model forecasts and actual observations under predominantly flat terrain and lower diurnal temperature variability. In contrast, higher RMSE values were recorded in Donetsk (3.54°C), Ivano-Frankivsk (3.71°C), and Crimea (3.69°C). For Donetsk, as previously noted, the limited number of observations (with only one meteorological station operational in the region) complicates verification. In the cases of Crimea and Ivano-Frankivsk, the increased errors may result from complex topographic features, where the model may not fully account for localized thermal processes.

The verification of mean temperature forecasts shows the highest stability among all temperature parameters. RMSE values predominantly range from 2.4°C to 2.7°C, indicating a strong agreement between model outputs and observed data. The lowest RMSE values were observed in Kherson (2.29°C), Mykolaiv (2.37°C), and Kirovohrad (2.42°C) oblasts, reflecting high forecast accuracy under conditions of lower thermal variability. Similar values were also recorded in Odesa, Poltava, Cherkasy, and Chernivtsi oblasts. Higher RMSE values were found in Crimea (3.70°C) and Donetsk (3.52°C). In Crimea, this may be attributed to complex marine circulation, local breezes, and increased temperature variability in coastal zones.

Since mean temperature is typically the most useful and representative parameter, only this type of temperature is illustrated on the accompanying map (*Fig. 2a*). It also demonstrates the best overall model performance. To avoid overloading the article with graphics, maps for minimum and maximum temperatures have been omitted.

***Wind.*** The results of spatial verification indicate that wind speed forecasting is one of the least accurate components of the ICON model. In most regions, the root mean square error (RMSE) exceeds 1.5 m/s (*Fig. 2b*), and the mean bias is consistently positive, suggesting a systematic overestimation of actual wind speeds by the model. This indicates a tendency of the model to overestimate wind speed, which is a common issue among many global and regional models that do not always accurately represent the effects of surface fric-

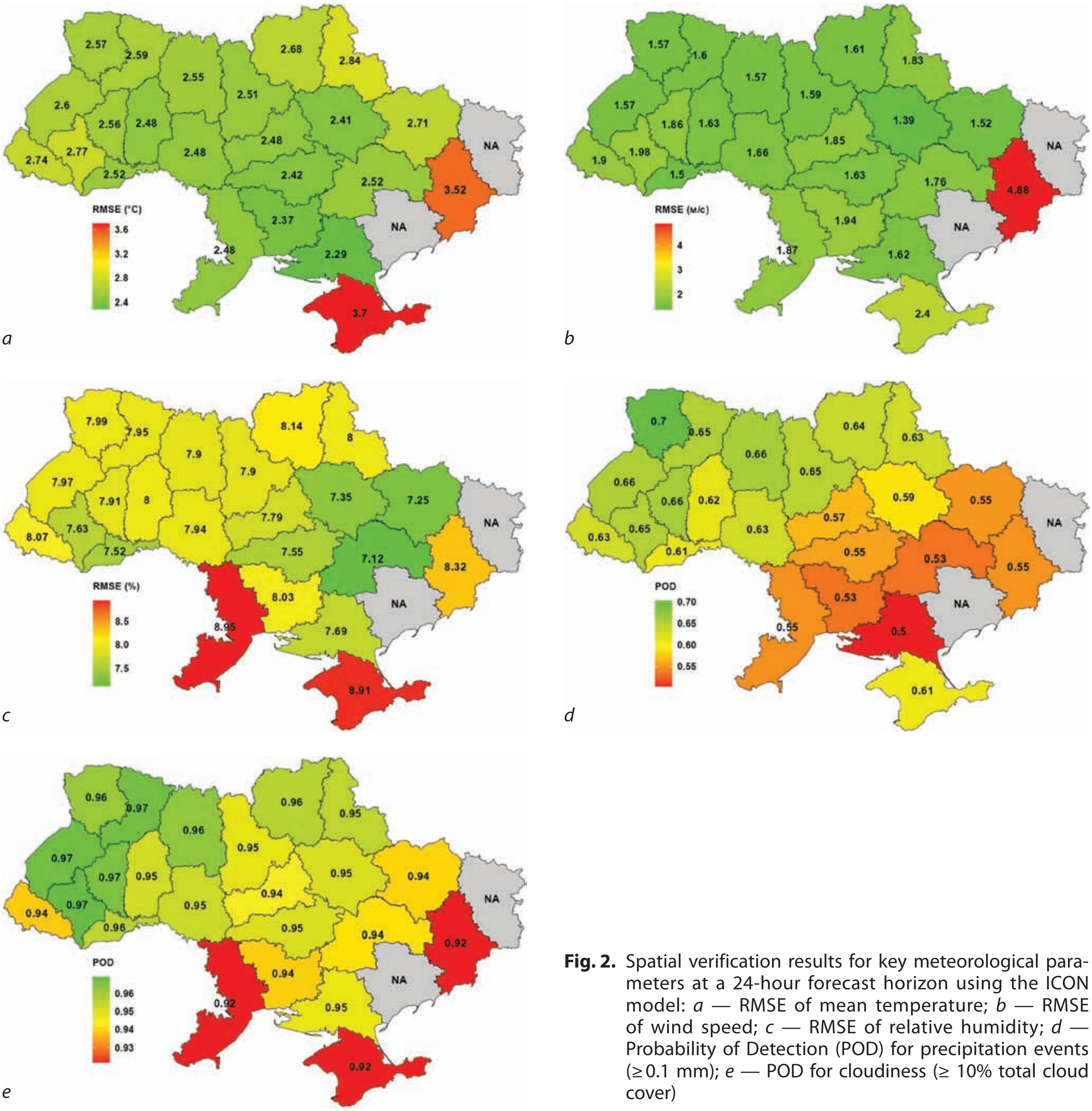

**Fig. 2.** Spatial verification results for key meteorological parameters at a 24-hour forecast horizon using the ICON model: *a* — RMSE of mean temperature; *b* — RMSE of wind speed; *c* — RMSE of relative humidity; *d* — Probability of Detection (POD) for precipitation events (≥ 0.1 mm); *e* — POD for cloudiness (≥ 10% total cloud cover)

tion, complex terrain, or local inversions (Hoffmann, Wilms, Blank, & Ludwig, 2022). The highest RMSE values were recorded in Donetsk (4.88 m/s) and Crimea (2.40 m/s), representing significant deviations compared to other regions. As with previous parameters, the limited availability of observational data remains a key factor for Donetsk. In the case of Crimea, the elevated errors are likely due to complex marine and coastal circulations, which may not be adequately captured by the model. The lowest RMSE values were found in Poltava (1.39 m/s), Chernivtsi (1.50 m/s), and Kharkiv (1.52 m/s) oblasts. These results reflect relatively high local stability of wind fields, which facilitates more accurate modeling in those regions.

***Relative humidity.*** The spatial verification of relative humidity forecasts from the ICON model at a 24-hour lead time demonstrates a high degree of consistency across regions. In most oblasts, the RMSE falls within the range of 7.1% to 8.9% (*Fig. 2c*), which represents an acceptable level of accuracy for synoptic-scale forecasting. The lowest RMSE was recorded in Dnipropetrovsk Oblast (7.12%), while the highest values were observed in Odesa (8.95%) and Crimea (8.91%). These elevated errors may be attributed to coastal circulation features, marine moisture transport, or insufficient observational coverage in shoreline areas. Overall, the results of spatial verification confirm that the ICON model provides relatively stable performance in forecasting relative humidity, though with a consistent negative bias. This effect has also been noted in other studies, which report overestimation of near-surface humidity by atmospheric models, particularly in the lowest atmospheric layers (Bastin et al., 2019; Liu et al., 2021).

***Precipitation.*** The Probability of Detection (POD) values for precipitation forecasts by the ICON model indicate a moderate level of accuracy in identifying precipitation events at a 24-hour lead time. Across most oblasts, POD values range from 0.53 to 0.66 (*Fig. 2d*), reflecting the model's reasonable ability to capture precipitation events, albeit with certain regional variations. For this spatial analysis, an event was defined as the occurrence of precipitation ≥ 0.1 mm, which, according to the results of the overall verification, provided the optimal balance between detection and false alarms. The highest POD values were recorded in Volyn (0.70), Ternopil (0.66), Lviv (0.66), Kyiv (0.65), and Rivne (0.65) oblasts—primarily located in the western and northwestern regions of Ukraine. This may be attributed to higher precipitation frequency in these areas and better model alignment with frontal activity in this sector. In contrast, the lowest values were observed in Kherson (0.50), Dnipropetrovsk (0.53), and Odesa (0.55) oblasts, indicating reduced model efficiency in detecting precipitation in southern and southeastern regions. This could be due to both lower overall precipitation frequency and greater variability in convective processes, which the model may not always adequately capture. In Crimea, the POD value was 0.61, which can be considered an average level of detection performance. Overall, the results of spatial verification suggest that the ICON model exhibits stable, though not optimal, skill in detecting precipitation events, with noticeable regional variation. Forecast accuracy tends to be higher in areas dominated by large-scale (frontal) precipitation systems, while performance declines in regions with a greater prevalence of localized or convective events.

***Cloudiness.*** The results of spatial verification demonstrate a high capability of the ICON model to detect cloud cover across nearly all regions of Ukraine, with Probability of Detection (POD) values ranging from 0.92 to 0.97 (*Fig. 2e*). This reflects the model's consistent performance in forecasting this parameter, regardless of geographic location or climatic conditions. For the purpose of defining a "cloudiness event", a threshold value of 10% total cloud cover was used. The highest POD values were observed in Ivano-Frankivsk (0.97), Lviv (0.97), Rivne (0.97), Ternopil (0.97), Volyn (0.96), and Zhytomyr (0.96) oblasts. These regions are characterized by high annual cloud cover and frequent cyclonic activity, which likely contributes to improved detection of extensive cloud systems by the model. Slightly lower (but still high) POD values were recorded in Odesa (0.92), Donetsk (0.92), Crimea (0.92), and Zakarpattia (0.94). In the southern regions, this may be related to the more frequent occurrence of localized, thin, or fragmented cloud structures, which are more difficult to model accurately within a global forecasting system—particularly under conditions of high turbulence or thermal inversions. Overall, the ICON model demonstrates strong performance in short-term cloudiness forecasting, delivering a high probability of correct cloud detection across all regions of Ukraine. This is a valuable feature for the model's further application in various operational domains, including aviation, agrometeorology, and energy planning.

## CONCLUSIONS

This study presents a detailed verification of the ICON numerical weather prediction model over Ukraine for the year 2024, covering forecasts of temperature, wind speed, relative humidity, precipitation, and cloud cover at 24-, 48-, and 72-hour lead times. The results confirm the model's overall strong performance in forecasting temperature and cloudiness, with correlation coefficients for mean temperature reaching r = 0.95 at 24 hours and Probability of Detection (POD) for cloud cover exceeding 0.94 at a 10% threshold. Temperature forecasts showed low root mean square errors (e.g., 2.61°C for mean temperature at 24 h) and minimal bias, indicating high reliability.

The ICON model also showed moderate accuracy for relative humidity forecasts ($r = 0.88$, MAE≈6.3%), although a consistent underestimation of about –4.2% was observed. Wind forecasts were the least accurate, with weak correlations ($r = 0.40$ at 24 h, dropping to 0.18 at 72 h) and a systematic overestimation of wind speed by ≈0.68 m/s.

For precipitation, the model demonstrated limited predictive skill, with POD values around 0.61 and Critical Success Index (CSI) of 0.36 at a 0.1 mm threshold for 24-hour forecasts, declining steadily with increased lead time and intensity threshold. Precipitation detection was more effective in western and northwestern oblasts (e.g., POD = 0.70 in Volyn), but performance dropped in southern regions (e.g., POD = 0.50 in Kherson), reflecting challenges in modeling convective activity.

Cloud cover forecasts were consistently strong across all regions and seasons, with CSI values up to

0.86 and FAR as low as 0.09, making this parameter one of the model's most reliable outputs.

Spatial verification further highlighted regional differences: ICON achieved the lowest RMSE for mean temperature in Kherson (2.29°C) and highest wind RMSE in Donetsk (4.88 m/s). Seasonal trends also revealed best overall performance in autumn and winter, while summer posed greater challenges, particularly for humidity and precipitation.

In summary, the ICON model shows excellent capability in forecasting temperature and cloudiness, reasonable accuracy for humidity, and room for improvement in wind and precipitation forecasts. These insights support ongoing refinement of model physics and post-processing techniques, especially for high-impact weather events in topographically and climatologically complex regions of Ukraine.

**Acknowledgments.** The authors acknowledge the Ukrainian Hydrometeorological Center and national observation network staff for maintaining the observational infrastructure under challenging conditions. We also thank DWD for the availability to download the data of the ICON model.

## REFERENCES

Bastin, S., Bock, O., Chazette, P., et al. (2019). Modeling and observations of lower tropospheric humidity profiles in the Mediterranean. *Atmospheric Chemistry and Physics*, **19**(5), 3155–3178. https://doi.org/10.5194/acp-19-3155-2019

Bauer, P., Thorpe, A., & Brunet, G. (2015). The quiet revolution of numerical weather prediction. *Nature*, **525**(7567), 47–55. https://doi.org/10.1038/nature14956

Dipankar, A., Schlemmer, L., Gorges, K., et al. (2015). ICON — a novel approach for global modeling of atmospheric dynamics: first results and future perspectives. *Geoscientific Model Development*, **8**(7), 2489–2511. https://doi.org/10.5194/gmd-8-2489-2015

Doroshenko, A., Shpyg, V., Budak I., Huda K. (2020). Numerical atmospheric models and their application in different areas of economics [In: Kvasniy L. and Tatomyr I. (eds) Ukraine in the context of global and national modern servisation processes and digital economy]: monograph. Praha: Oktan Print, 155–171. doi: 10.46489/uitcog0909

Ebert, E.E., Janowiak, J.E. & Kidd, C. (2016). Comparison of near-real-time precipitation estimates from satellite observations and numerical models over global land areas. *Journal of Hydrometeorology*, **17**(8), 2453–2470. https://doi.org/10.1175/JHM-D-16-0010.1

Giorgetta, M. A., Brokopf, R., Crueger, T., Esch, M., Fiedler, S., Helmert, J., et al. (2018). ICON-A, the atmosphere component of the ICON Earth system model: I. Model description. Journal of Advances in Modeling Earth Systems, 10. https://doi.org/10.1029/2017MS001242

Hoffmann, P., Wilms, H., Blank, L., & Ludwig, P. (2022). Benchmarking high-resolution regional climate models over complex terrain: A case study for wind simulations in the Alps. *Frontiers in Earth Science*, **10**, 789332. https://doi.org/10.3389/feart.2022.789332

Holoborodko, P., Lykhovyd, O., & Fedirko, V. (2020). Regional climatology and atmospheric dynamics of Ukraine: challenges for numerical weather prediction. *Ukrainian Hydrometeorological Journal*, **12**(1), 45–59.

Kalnay, E. (2003). Atmospheric modeling, data assimilation and predictability. Cambridge University Press. https://doi.org/10.1017/CBO9780511802270

Liu, L., Huang, X., Chen, F., et al. (2021). Evaluation of relative humidity in weather and climate models over complex terrain. *Journal of Geophysical Research: Atmospheres*, **126**(12), e2020JD034123. https://doi.org/10.1029/2020JD034123

Shpyg, V., Budak, I. (2015). WRF reflectivity simulation and verification of thunderstorm forecast by radar and surface observation. 16th International Radar Symposium. Symposium Materials (24–26 June 2015, Dresden, Germany). 610–615. doi: 10.1109/irs.2015.7226388

Shyian, S., Dudnik, I. & Makarenko, A. (2018). Climatic conditions of Ukraine and their variability: implications for weather prediction. *Geographical Review of Ukraine*, **11**(2), 33–42.

Sundqvist, H., Berge, E. & Kristjánsson, J.E. (1989). Condensation and cloud parame- terization studies with a mesoscale numerical weather prediction model. *Monthly Weather Review*, **117**(8), 1641–1657.

Ukrainian Hydrometeorological Center. (2023). Annual report on meteorological observations and challenges. Kyiv: State Hydrometeorological Service of Ukraine.

Zängl, G., Reinert, D., Rípodas, P. & Baldauf, M. (2015). The ICON (ICOsahedral Nonhydrostatic) modelling framework of DWD and MPI-M: description of the nonhydrostatic dynamical core. *Quarterly Journal of the Royal Meteorological Society*, **141**(687), 563–579. https://doi.org/10.1002/qj.2378

---

**Кривошеїн О.О.***
*ORCID: 0000-0001-5029-4228*
*krivoshein@uhmi.org.ua*

**Кривобок О.А.**
*ORCID: 0000-0002-1730-1809*
*kryvobok@uhmi.org.ua*

**Заболотна О.С.**
*ORCID: 0009-0009-6338-7672*
*zabolotnaolena@ukr.net*

## ВЕРИФІКАЦІЯ ЧИСЕЛЬНОЇ МОДЕЛІ ПРОГНОЗУ ПОГОДИ ICON В УКРАЇНІ

*Український гідрометеорологічний інститут*
*Державної служби України з надзвичайних ситуацій*
*та Національної академії наук України, Київ*

*У статті представлено комплексну верифікацію глобальної моделі чисельного прогнозування погоди ICON над територією України за 2024 рік. Оцінювання охоплювало ключові метеорологічні параметри: температуру по-*

*вітря, швидкість вітру, відносну вологість, кількість опадів та хмарність — на горизонтах прогнозу 24, 48 та 72 години. Для аналізу точності були застосовані як неперервні (кореляція, середня абсолютна похибка, середньоквадратична похибка, зміщення), так і категоріальні метрики (POD, FAR, CSI), з урахуванням просторового та сезонного розподілу. Результати показали високу точність прогнозу температури: коефіцієнт кореляції для середньої температури досягав r = 0.95 на горизонті 24 годин, при низьких значеннях RMSE (≈2.6 °C) та мінімальному зміщенні. Прогноз хмарності також продемонстрував високу якість: POD > 0.94 та CSI до 0.86 на низькому порозі (10% хмарності), з незначним зниженням точності при зростанні горизонту прогнозу. Водночас модель продемонструвала обмежену точність у прогнозуванні швидкості вітру (r = 0.40 при 24 год, RMSE ≈ 1.75 м/с), що супроводжувалось систематичним завищенням. Прогноз відносної вологості був задовільним (r = 0.88), хоча спостерігалося стабільне негативне зміщення (близько –4.2%). Найнижча точність зафіксована у прогнозі опадів: при порозі 0.1 мм POD становив 0.61, але FAR залишався високим (>0.50), особливо в південних регіонах, де переважає конвективний характер опадів. Сезонна верифікація показала найвищу точність у холодний період року, зокрема восени та взимку, а також виявила погіршення прогнозу вологості та опадів улітку. Просторовий аналіз на горизонті 24 години дозволив виявити регіональні особливості прогнозної точності. Наприклад, найменше RMSE для середньої температури зафіксовано в Херсонській області (2.29 °C), тоді як найбільше — для вітру в Донецькій області (4.88 м/с). Загалом модель ICON продемонструвала високу надійність у прогнозуванні температури та хмарності, задовільні результати для вологості та значний потенціал для покращення точності прогнозів опадів і вітру. Отримані результати можуть бути використані для підвищення ефективності регіонального прогнозування та адаптації фізичних параметризацій моделі до кліматичних і топографічних особливостей України.*

**Ключові слова:** *модель ICON, верифікація, прогноз погоди, точність прогнозування.*