Information systems modeling

UDC 004.85: 004.94: 551.553 **doi:** https://doi.org/10.20998/2522-9052.2024.4.02

Nataliia Ausheva[1], Svitlana Shapovalova[1], Kateryna Petrenko[2], Oleksandr Kardashov[1], Anton Sofiienko[1]

[1] National Technical University of Ukraine «Igor Sikorsky Kyiv Polytechnic Institute», Kyiv, Ukraine
[2] The Institute of Renewable Energy of the National Academy of Sciences of Ukraine, Kyiv, Ukraine

CONSTRUCTION OF A SPATIAL DISTRIBUTION MODEL OF WIND ENERGY CHARACTERISTICS

Abstract. The aim of this article is to develop a model for the spatial distribution of wind energy characteristics across the territory of Ukraine. **The subject of the study** includes datasets of wind speed values, as well as methods of data correlation, validation, and interpolation. **Research results.** Based on NASA reanalysis datasets and measurement results from 70 meteorological stations in Ukraine, a dataset of paired wind speed data corresponding to the same location but obtained through different methods was created. Through a comparative analysis of regression task results, evaluated using machine learning models trained on the dataset, the Random Forest model was selected as the most accurate (based on RMSE, R^2, and Pearson correlation coefficient) for predicting wind speedc deviations in NASA reanalysis data to bring them closer to actual values. The Pearson correlation coefficient improved by 0.07 in the worst case and by 0.66 in the best case. Using the Random Forest model's predictions, corrections were made to all wind speed values in the NASA reanalysis data. The accuracy of the corrected data was also confirmed by the study of trend dynamics over the course of a year using three different data sources: meteorological station measurements, NASA reanalysis data, and corrected NASA reanalysis data. Using the universal kriging method, the corrected wind speed values were interpolated at grid nodes across the entire territory of Ukraine. The accuracy of the interpolation results was validated using the cross-validation method. **Conclusion.** Based on these results, a GIS-based tool was created, enabling the determination of reliable wind energy characteristics at any given point across Ukraine. The proposed GIS can primarily be used for the design of wind power plants and for selecting optimal locations for their deployment.

Keywords: machine learning; global atmospheric models; MERRA-2; meteorological data; reanalysis data; wind speed interpolation; inverse distance weighting; universal kriging.

Introduction

Renewable energy is currently a priority area for energy development in most developed countries. As of 2023, wind power plants (WPP) account for 26.3% (1,017.4 GW) of the total installed capacity in the global renewable energy market [1]. Between 2000 and 2022, global wind energy production increased from 31.2 TWh to 2,098.5 TWh, representing a 67-fold increase [2]. Despite a significant reduction in the cost of WPP installation, their construction still requires substantial investments [3], which highlights the importance of accurate forecasting of the potential annual energy production of a wind farm. Therefore, a reliable assessment of wind resources at the future WPP site is a crucial stage in the design process.

The most reliable source of information on wind speed and direction at a potential WPP site comes from measurements taken directly at the location. However, since wind farms have lifespans of several decades, the World Meteorological Organization recommends the use of wind speed time series of at least 10 years to account for long-term trends. Such measurements are costly and time-consuming.

A source of long-term wind speed and direction data is the network of meteorological stations (MS), where wind parameters are measured every 3 hours at a height of 10 meters. However, the distribution of MS across Ukraine is highly uneven, and their data do not provide insights into wind speeds at significant altitudes, whereas the hub height of modern wind turbines exceeds 100 meters. In global practice, databases from reanalysis models [4, 5], which are the result of complex modeling of satellite data and remote sensing data [6], are used to provide information on wind at heights of 10, 50, 100, and 300 meters for wind energy calculations.

However, reanalysis data do not always align with real observation data. The verification and adaptation of reanalysis data for wind energy applications in specific regions are conducted by researchers worldwide. Developing models that can improve energy production forecasts based on reanalysis data and bring them closer to the more accurate predictions derived from ground meteorological station data would significantly enhance the precision of energy production forecasts for planned WPPs. Therefore, spatial modeling of wind speed in a specific area is a relevant and practically significant task.

Analysis of recent research and publications

The first international validation of wind speed data for the MERRA and MERRA-2 reanalysis datasets with observational results was conducted for 23 European countries [7]. Both reanalysis datasets were determined to have a significant spatial error, significantly overestimating wind speed in some regions and underestimating it in others. National correction factors were proposed, making obtaining data with acceptable accuracy possible.

In the study [8] NasaPower reanalysis products were evaluated for daily mean temperature, solar radiation, relative humidity, and wind speed parameters compared to data from 14 weather stations in southern Portugal with a Mediterranean climate. The research results have shown that NASA POWER data can be useful for creating weather datasets when data from ground-based weather stations are missing or unavailable. The study showed good agreement for all parameters except wind speed. Even after correcting the data for bias, the data for this indicator show poor correlation, are inconsistent with most of the observations, and still need improvement.

The paper [9] presents a study of the influence of distance and altitude of the regions above sea level on the deviation of the average daily values of temperature, relative humidity, and wind speed of the NASA POWER reanalysis database from the observation data of 3 meteorological stations in the Mediterranean and Continental regions of Turkey. The research results also showed a high correlation between both datasets for all parameters except wind speed.

In Ukraine, studies have been conducted on the correlation of wind speed values based on the results of measurements at 70 MS and the MERRA-2 data set [10]. It was found that for the territory of Ukraine, the average correlation coefficient for average daily pairs of values is 0.8 (minimum 0.45 for a mountainous MS, maximum 0.95 for an MS located on the open territory of the airport).

Thus, the reanalysis data needs to be corrected based on real-world observation datasets. For this purpose, it is necessary to determine the wind speed correction function from the NasaPower reanalysis database in relation to the corresponding observational data. This regression problem can be solved by comparing the wind speed measurements at the MS with the MERRA-2 reanalysis data, which are given for grid nodes reasonably close to the MS coordinates. The regression problem is a classic machine-learning task. Usually, applied problems, reduced to classical ones, are solved by choosing the best-trained machine learning model in terms of accuracy [11–14].

The next data preparation stage is to adjust the wind speed for all MERRA-2 grid nodes in Ukraine based on the forecasts of the selected trained artificial intelligence model.

The next task, after creating the relevant dataset, is to interpolate wind speed values for the entire territory of Ukraine using the available sample of values in the coordinates of deterministic points.

The paper [15] presents models of the spatial distribution of the mean wind speed and wind energy density developed on the basis of the spatial interpolation method based on the results of historical measurements at 22 observation stations in Latvia. The results are presented in the form of color contour maps. To make it possible to determine the desired characteristic at any point, it is advisable to present models of the spatial distribution of wind energy characteristics in the form of GIS [16].

The study [17] demonstrates the impracticality of using deterministic interpolation methods, such as the Inverse Distance Weighting (IDW) method, in cases where observation points are unevenly distributed. In the research [18], a comparative analysis is conducted between the classical IDW method, the Modified Inverse Distance Weighting (MIDW) method, and the Gradient Inverse Distance Weighting (GIDW) method. Based on the analysis results, it is concluded that the classical IDW method cannot be applied without modifying the calculation of the weighting coefficients for the points.

Another approach to interpolation is the geostatistical one, which is based on the spatial autocorrelation of data [19, 20]. One of the most common geostatistical interpolation methods is kriging, which is divided into ordinary kriging, typically used in most cases, and universal kriging, which assumes the presence of dominant trends in the data. The study [21] compares the results of wind speed forecasting using the methods of ordinary and universal kriging, assessing the appropriateness of these methods for short-term wind speed forecasting. Universal kriging is used in cases where the data are known to contain scientifically established trends. Wind speed distribution properties tend to exhibit dominant trends, such as prevailing winds and topographical features like mountain ranges and coastal areas [22]. The study [23] examines global trends in wind flow distribution, particularly wind speed, which significantly affects data forecasting. Thus, given the specific nature of the subject area, it is appropriate to apply the universal kriging method to solve the problem of wind speed interpolation across the territory of Ukraine.

Problem statement. To design wind power plants and select optimal locations, it is necessary to create a software resource that will allow determining reliable values of wind energy characteristics at any given point in Ukraine. The purpose of the article is to create a model of the spatial distribution of wind energy characteristics in Ukraine. To achieve this goal, the following tasks need to be accomplished:

1. To form a dataset of pairs of relevant data that refer to the same location and belong to different existing data sets.
2. To create a machine learning model to predict the shift in the wind speed value in the NasaPower reanalysis data to approximate the real value.
3. To correct the wind speed values in the *NASAPower* reanalysis data.
4. To interpolate the obtained values for the entire territory of Ukraine.

Formation of a dataset of correlation of measured and calculated wind speed

Two sufficiently large and representative arrays of data were obtained for the same territory:

- *NASAPower* - by the reanalysis method.
- *WeatherStation* - by ground meteorological observation.

The *NASAPower* dataset is derived from NASA's open reanalysis database [24], which provides daily data on surface air temperature, relative humidity, precipitation, solar radiation, wind speed, and wind direction in a coordinate grid (0.5◦ latitude by 0.5◦ longitude resolution). Wind speed data are provided for heights of 10 m and 50 m above the ground. The *NASAPower* dataset uses 4939200 wind speed records at 10 m height at 533 grid nodes with specified geographic coordinates for the period from January 1, 2001 to January 1, 2024 at 12 noon.

The *WeatherStation* dataset contains daily meteorological data from 70 ground-based weather stations in Ukraine obtained from the Institute of Renewable Energy of the National Academy of Sciences of Ukraine [25] and open sources. The dataset consists of 264543 wind speed measurements at 12 noon for the period from January 1, 2011 to December 31, 2020. The data of the *WeatherStation* dataset, unlike the *NASAPower* dataset, contains gaps in measurements for different time periods at different MSs. The geographical coordinates of the corresponding weather station were determined and added to each record.

Identifying data for correcting the wind speed value from the *NASAPower* dataset for the corresponding *WeatherStation* observations first requires extracting data from both datasets, which are defined by the same location.

Fig. 1 shows a map of Ukraine with a grid of dot nodes containing reanalysis data and triangles indicating the locations of weather stations.

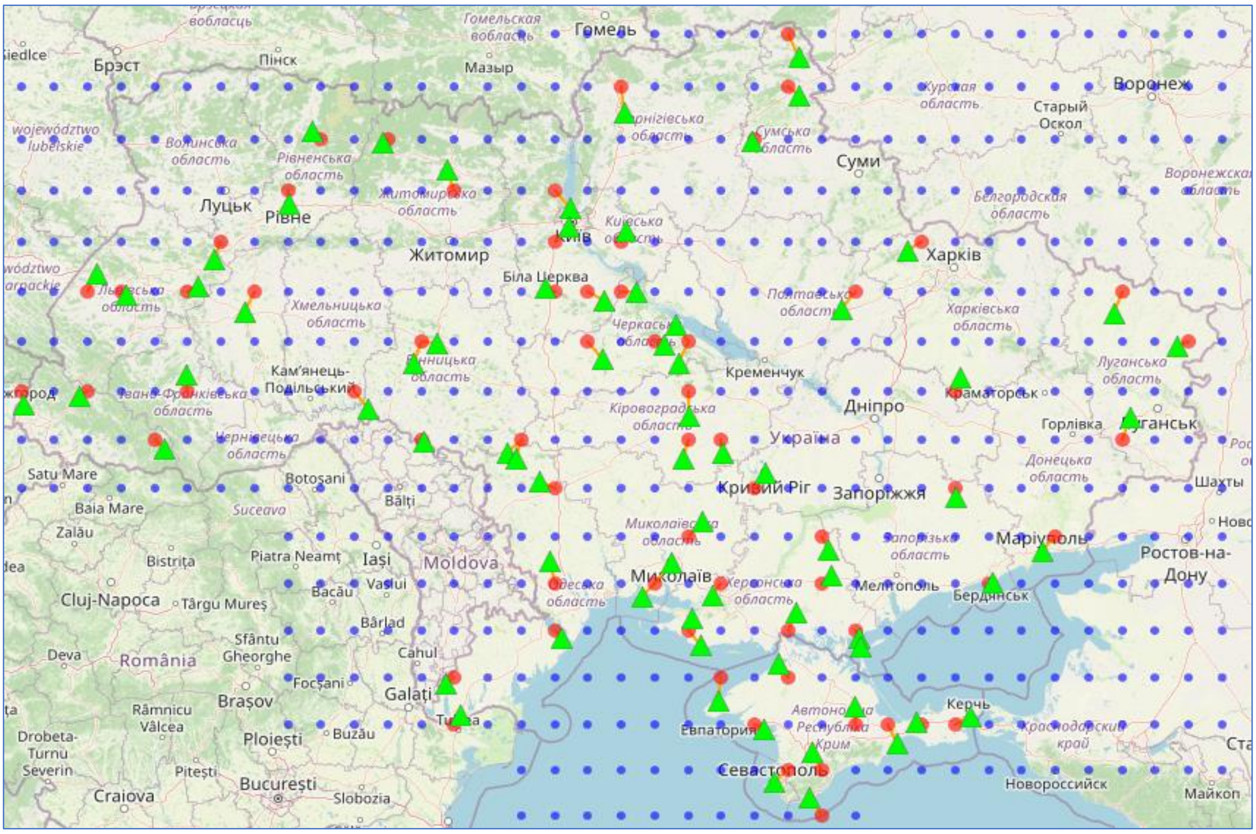


Fig. 1. Locations of available wind speed data

Based on the pairs of relevant data calculated/measured at the same location, the *MEASURvsMODEL* dataset of 264543 records was created. A record represents:

- a pair of time series of wind speeds measured at weather stations and calculated by a reanalysis model at the nearest NASA grid node,
- the height of the location point above sea level,
- the serial number of the day of the year on which the measurement was performed.

The last two features are added to the initial datasets because altitude and season significantly affect wind speed. The height above sea level was determined by the location coordinates from the Shuttle Radar Topography Mission [26].

A machine learning model for wind speed shifting prediction

The regression problem was solved based on the obtained dataset *MEASURvsMODEL*. The adjusted value of wind speed was determined for each pair of time series of wind speeds.

The following models were selected from the sklearn library [27]: Linear Regression, Ridge Regression, Lasso Regression, Decision Tree, Random Forest, KNeighbors Regressor, Perceptron, Gradient Boosting Regressor, and Historical Gradient Boosting Regressor.

The root mean square error (MSE) was chosen as the optimization metric.

To evaluate the performance of the models, five-fold cross-validation was applied, which ensures the distribution of data into five subsets, where each model is trained on four folds and tested on the fifth.

The results for each model were averaged over all folds, which made it possible to obtain a generalized assessment of the quality of the models. Table 1 shows the average values of the root mean square error (RMSE), the coefficient of determination R2, and the increase in the Pearson correlation coefficient of each of the models. These metrics allow us to quantify how well the model corrects the data.

The best performance was obtained by the Random Forest ensemble method, which uses the prediction from a large number of decision trees. Since each tree is trained on a random subset of data and features, the model becomes robust to noise and anomalies, which is important for meteorological data where there may be random fluctuations in wind speed.

Table 1 – **Results of testing machine learning models on the regression task**

Method	**RMSE**	**R2**	**Pearson Coefficient Delta**
Linear Regression	1.8185	0.3913	0.0027
Ridge Regression	1.8185	0.3913	0.0027
Lasso Regression	1.8193	0.3907	0.0027
Decision Tree	2.1980	0.1107	0.2855
Random Forest	**1.6770**	**0.4823**	**0.2944**
K-Neighbors Regressor	1.7895	0.4106	0.1331
Perceptron	6.3224	-6.381	-0.5047
Gradient Boosting	1.7033	0.4660	0.0627

The average value of the R2 metric improved by 0.73 points. The smallest increase was observed in Nyzhni Sirohozy, where the indicator increased from 0.613 to 0.8.

The largest increase was recorded in Zolochiv, where R2 increased from -2.06 to 0.64. The RMSE metric improved by 1 on average. The smallest increase was observed in Nyzhni Sirohozy, where the indicator decreased from 1.418943 to 1.019769.

The largest increase in model quality by this metric was observed in Zvenyhorodka, where the RMSE decreased from 2.132813 to 0.329809. The value of the Pearson correlation coefficient improved by 0.294 on average. The smallest increase was observed in Strelkove, where the coefficient increased from 0.782145 to 0.853491.

The largest increase was in Velykyi Burluk, where the coefficient increased from 0.17 to 0.826, indicating a significant increase in correlation accuracy.

Correction of wind speed in *NASApower* real-time data

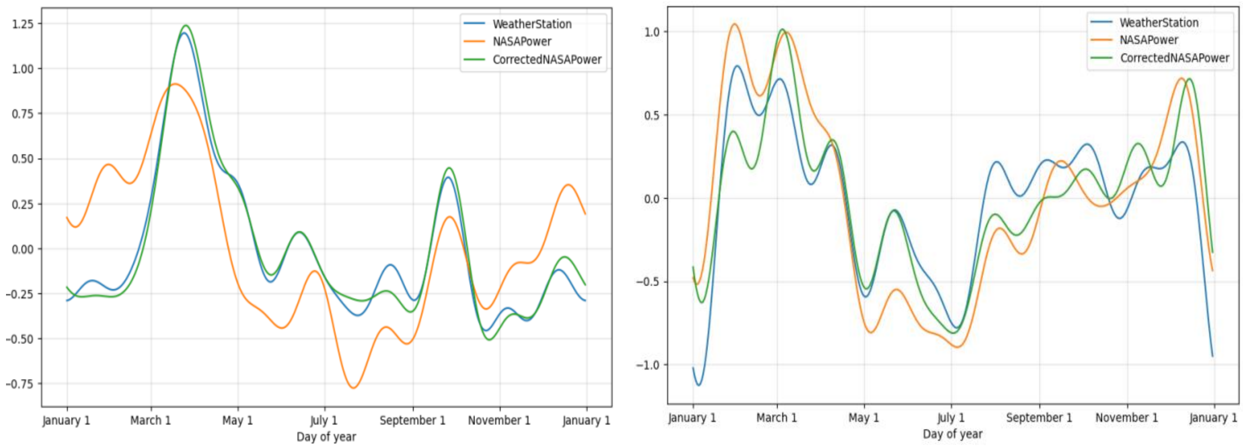
Based on the predictions of the selected Random Forest regression model, wind speed values were corrected for all grid nodes with *NASAPower* reanalysis data in Ukraine.

To verify the authenticity of the corrected *NASAPower* data in relation to the corresponding wind speed measurements of meteorological stations, the annual dynamics of changes in trend values were investigated using the Prophet library [28]. The graphs below show annual wind speed trends for three different data sources: *WeatherStation* (weather station data), *NASAPower* (reanalysis data) and *CorrectedNASAPower* (corrected data). All three data series have been normalized to reflect the annual cycle of wind speed, which allows to reveal the main patterns of fluctuations throughout the year. The graphs (Fig. 2) show seasonal fluctuations.

The *NASAPower* data show similar seasonal changes but with larger deviations in amplitude, especially in the summer and fall periods.

The adjusted data provide a smoother trend that is less prone to the large seasonal variations characteristic of *NASAPower*. Example 2, a shows a correct correlation and corresponds to most of the corrected data in the nodes. Example 2, b is an example of a not-so-good correlation, which is explained by the location on the sea coast.

Saving all 3 trends confirms the authenticity of the *NASAPower* data correction



a b

Fig. 2. Locations of available wind speed data: a – Cherkasy, b – Strilkove

Wind speed distribution construction on the territory of Ukraine

To determine the wind speed value at any point of the territory, it is necessary to interpolate the wind speed specified in the coordinate grid nodes. The input to the interpolation is an array of point coordinates, each of which contains a wind speed value generated on the basis of the *CorrectedNASAPower* dataset.

The result is an output raster - a GIS layer, each pixel of which corresponds to the predicted wind speed value using the appropriate interpolation method. The accuracy of the method is evaluated by comparing the predicted value with the wind speed value at the coordinates of each point of the input sample.

Geostatistical interpolation using the universal kriging method was performed using the Geostatistical Analyst module in ArcGIS Pro application [29]. The result of the kriging interpolation is an output raster, where the predicted value of wind speed is determined for each pixel (Fig. 3). GIS technology enables the determination of wind speed at any specific location. For instance, as illustrated in Fig. 4, the forecast near the city of Novoselytsia indicates a wind speed of 4 m/s.

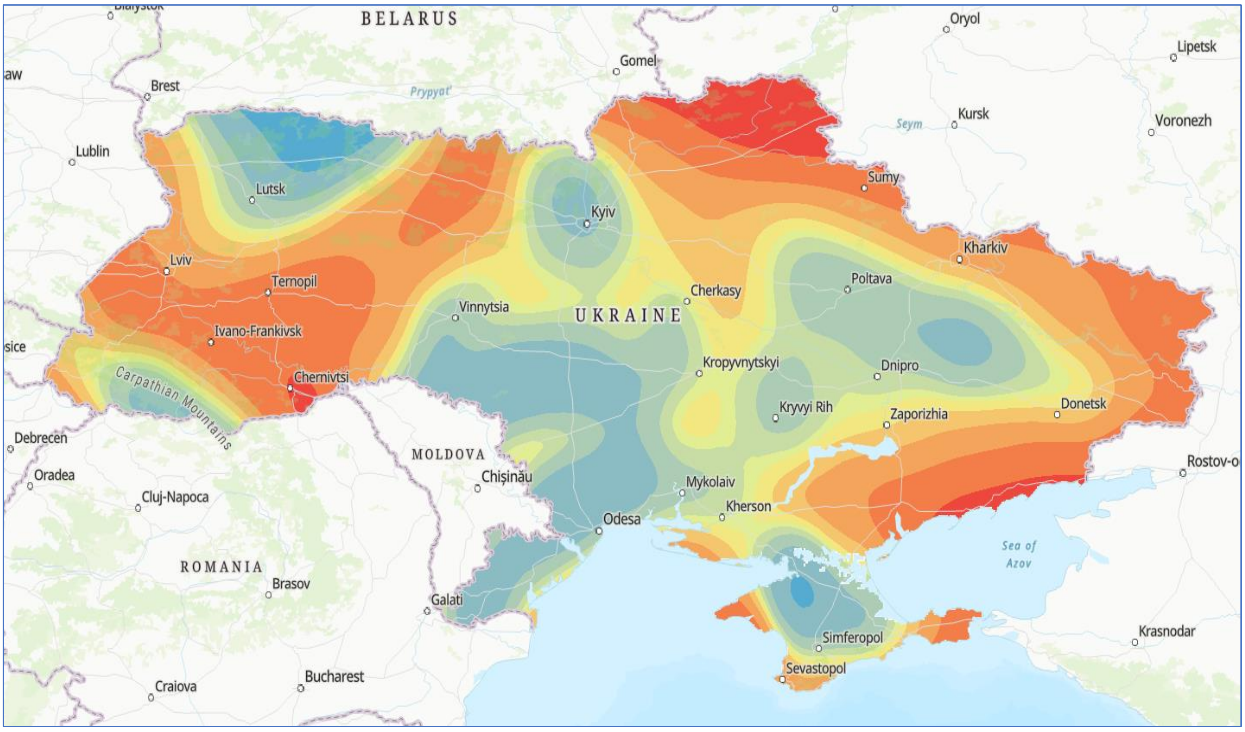


Fig. 3. Distribution of wind speed on the territory of Ukraine as of 20.08.2020, 12:00, at an altitude of 10m

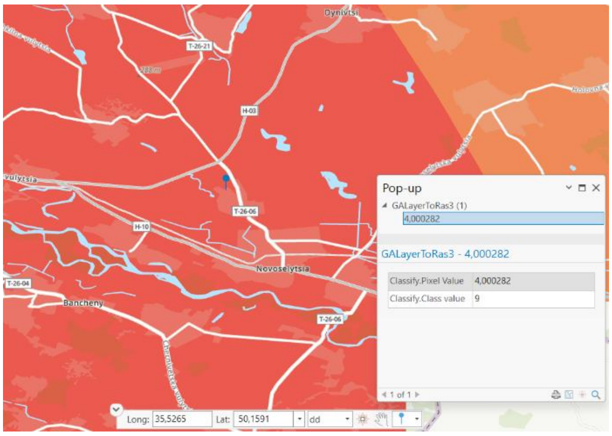


Fig. 4. Input data for wind speed forecasting

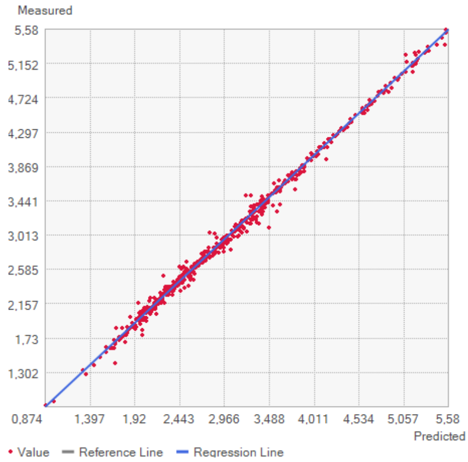


Fig. 5. Graph of predicted value

The accuracy of the obtained results was evaluated using the cross-validation method. The resulting graph of predicted values (Figure 5) demonstrates that these values are sufficiently close to both the reference line and the regression line, which almost coincides with the reference line.

This confirms the correctness of the interpolation.

Conclusions

1. Based on NASA reanalysis datasets and measurement results from meteorological stations in Ukraine, a dataset of wind speed data pairs was created, corresponding to the same location but obtained through different methods.

2. Through a comparative analysis of regression task results, evaluated using machine learning models trained on the dataset, the Random Forest model was selected as the most accurate (based on RMSE, R^2, and Pearson correlation coefficient) for predicting wind speed deviations in NASA reanalysis data, bringing them closer to the actual values.

3. Using the forecasts of the Random Forest model, corrections were made to wind speed values in the NASA reanalysis data. The accuracy of the corrected

data was confirmed by the results of the study on the dynamics of trend changes throughout the year from three different data sources: meteorological station measurements, NASA reanalysis data, and corrected NASA reanalysis data. The Pearson correlation coefficient improved by 0.07 in the worst case and by 0.66 in the best case.

4. Using the universal kriging method, the corrected wind speed values were interpolated at grid nodes across the entire territory of Ukraine. The accuracy of the interpolation results was verified using the cross-validation method. Based on the obtained results, a spatial distribution model of wind energy characteristics across Ukraine was created in GIS format.

References

1. International Renewable Energy Agency (2024), *Renewable energy statistics 2024*, 298 p., available at: https://www.irena.org/-/media/Files/IRENA/Agency/Publication/2024/Jul/IRENA_Renewable_Energy_Statistics_2024.pdf
2. Amber and Energy Institute (2023), *Wind power generation*, available at: https://ourworldindata.org/grapher/wind-generation
3. International Renewable Energy Agency (2022), *Renewable power generation costs in 2022*, 208 p., available at: www.irena.org/-/media/Files/IRENA/Agency/Publication/2023/Aug/IRENA_Renewable_power_generation_costs_in_2022.pdf
4. Samal, R.K. (2021), "Assessment of wind energy potential using reanalysis data: A comparison with mast measurements", *Journal of Cleaner Production*, vol. 313, doi: https://doi.org/10.1016/j.jclepro.2021.127933
5. Gonzalez-Arceo, A., de Musitu, M.Z.-M., Ulazia, F., del Rio, M. and Garcia, O. (2020), "Calibration of Reanalysis Data against Wind Measurements for Energy Production Estimation of Building Integrated Savonius-Type Wind Turbine", *Applied Sciences*, vol. 10 (24), doi: https://doi.org/10.3390/app10249017
6. Gelaro, R., McCarty, W., Suárez, Max J. and Zhao, B. (2017), "The Modern-Era Retrospective Analysis for Research and Applications, Version 2 (MERRA-2)", *Journal of Climate*, vol. 30 (14), doi: https://doi.org/10.1175/JCLI-D-16-0758.1
7. Staffell, I. and Pfenninger, S. (2016), "Using bias-corrected reanalysis to simulate current and future wind power output", *Energy*, vol. 114, pp. 1224–1239, doi: https://doi.org/10.1016/j.energy.2016.08.068
8. Rodrigues, G.C. and Braga, R.P. (2021), "Evaluation of NASA POWER Reanalysis Products to Estimate Daily Weather Variables in a Hot Summer Mediterranean Climate", *Agronomy*, vol. 11(6), no. 1207. doi: https://doi.org/10.3390/agronomy11061207
9. Halİmİ, A. H., Karaca, C. and Büyüktaş, D. (2023), "Evaluation of NASA POWER Climatic Data against Ground-Based Observations in The Mediterranean and Continental Regions of Turkey", *JOTAF*, vol. 20, is. 1, pp. 101–114, doi: https://doi.org/10.33462/jotaf.1073903
10. Petrenko, K., Kuznietsov, M., Ivanchenko, I., Karmazin, O. and Borsuk, A. (2023), "ssessment of the possibility of using wind speed reanalysis data for wind energy calculations", *Vidnovluvana Energetika*, is. 3, pp. 75–85, doi: https://doi.org/10.36296/1819-8058.2023.3(74).75-85
11. Shapovalova, S., Matiakh, S. and Titov, V. (2024), "Determination of network traffic anomalies in a distributed computer system with energy facilities", *Vidnovluvana Energetika*, vol. 2(77), pp. 52–57. doi: https://doi.org/10.36296/1819-8058.2024.2(77).52-57
12. Sobchuk, V., Pykhnivskyi, R., Barabash, O., Korotin, S. and Omarov, S. (2024), "Sequential intrusion detection system for zero-trust cyber defense of IOT/IIOT networks", *Advanced Information Systems*, vol. 8, no. 3, pp. 92–99. doi: https://doi.org/10.20998/2522-9052.2024.3.11
13. Aniskevich, S., Bezrukovs, V., Zandovskis, U. and Bezrukovs, D. (2017), "Modelling the spatial distribution of wind energy resources in Latvia", *Latvian Journal of Physics and Technical Sciences*, vol. 6. pp. 10–20, doi: https://doi.org/10.1515/lpts-2017-0037
14. Abdulhussein, A.-M. Aqeel, Smirnova, T., Buravchenko, K. and Smirnov, O. (2023), "The method of assessing and improving the user experience of subscribers in software-configured networks based on the use of machine learning", *Advanced Information Systems*, vol. 7, no. 2, pp. 49–56, doi: https://doi.org/10.20998/2522-9052.2023.2.07
15. Poliarush, O., Krepych, S. and Spivak, I. (2023), "Hybrid approach for data filtering and machine learning inside content management system", *Advanced Information Systems*, vol. 7, no. 4, pp. 70–74, 2023, doi: https://doi.org/10.20998/2522-9052.2023.4.09
16. Andrieiev, S. and Zhilin, V. (2020) "Methods of construction of hydrological cartographic models according to remote sensing of the Earth data", *Advanced Information Systems*, vol. 4, no. 3, pp. 22–40, doi: https://doi.org/10.20998/2522-9052.2020.3.04
17. Changyeon, L. (2022), "Long-term wind speed interpolation using anisotropic regression kriging with regional heterogeneous terrain and solar insolation in the United States", *Energy Reports*, vol. 8, pp. 12–23, doi: https://doi.org/10.1016/j.egyr.2021.11.285
18. Zhao, W., Zhong, Y., Li, Q., Li, M., Liu, J. and Tang, L. (2022), "Comparison and correction of IDW based wind speed interpolation methods in urbanized Shenzhen", *Front. Earth Sci*., is. 16, pp. 798–808, doi: https://doi.org/10.1007/s11707-021-0948-z
19. Wackernagel, H. (2003), *Multivariate Geostatistics: An Introduction with Applications*, Springer, Berlin, 387 p., doi: https://doi.org/10.1007/978-3-662-05294-5
20. Kuchuk, N., Mozhaiev, O., Semenov, S., Haichenko, A., Kuchuk, H., Tiulieniev, S., Mozhaiev, M., Davydov, V., Brusakova, O. and Gnusov, Y. (2023), "Devising a method for balancing the load on a territorially distributed foggy environment", *Eastern-European Journal of Enterprise Technologies*, vol. 1(4 (121), pp. 48–55, doi: https://doi.org/10.15587/1729-4061.2023.274177
21. Wang, Yu, Wang, D., Zhao, J. and Zhu, C. (2020), "Wind speed spatial estimation using geostatistical kriging", *IOP Conf. Ser.: Earth Environ. Sci*., vol. 619, no. 012049, doi: https://doi.org/10.1088/1755-1315/619/1/012049
22. Brower, M. C., Barton, M. S., Lledo, L. and Markus, M. J. (2012), *Wind Resource Assessment: A Practical Guide to Developing a Wind Project*, Wiley, 280 p., doi: https://doi.org/10.1002/9781118249864
23. Jung, C. and Schindler, D. (2024), "Global trends of wind direction-dependent wind resource", *Energy*, vol. 304, no. 132235, doi: https://doi.org/10.1016/j.energy.2024.132235
24. (2024), "NASA Prediction of Worldwide Energy Resources", *The POWER Project*, available at: https://power.larc.nasa.gov/
25. Kudrya, S. (2023), *Wind energy*, The Institute of Renewable Energy, Kyiv, 135 p., available at: https://www.ive.org.ua/wp-content/uploads/%D0%92%D1%96%D1%82%D1%80%D0%BE%D0%B5%D0%BD%D0%B5%D1%80%D0%B3%D0%B5%D1%82%D0%B8%D0%BA%D0%B0_%D0%BC%D0%BE%D0%BD%D0%BE%D0%B3%D1%80%D0%B0%D1%84%D1%96%D1%8F-2023_%D0%BD%D0%B0_%D1%81%D0%B0%D0%B9%D1%82.pdf

26. (2000), "Earthdata", *Shuttle Radar Topography Mission (SRTM)*, available at: https://www.earthdata.nasa.gov/sensors/srtm
27. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Brucher, M., Perrot, M. and Duchesnay, E. (2011), "Scikit-learn: Machine Learning in Python", *The Journal of Machine Learning Research*,, vol. 12, pp. 2825–2830, available at: https://dl.acm.org/doi/10.5555/1953048.2078195
28. (2023), *GitHub Prophet*, available at: https://github.com/facebook/prophet
29. (2024), "Advancing the power of geography", *ArcGIS Geostatistical Analyst*, available at: https://www.esri.com/en-us/arcgis/products/geostatistical-analyst/overview

Надійшла (received) 22.06.2024
Прийнята до друку (accepted for publication) 06.09.2024

Відомості про авторів / About the authors

Аушева Наталія Миколаївна – доктор технічних наук, професор, завідувачка кафедри цифрових технологій в енергетиці, Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського», Київ, Україна;
Nataliia Ausheva – Doctor of Technical Sciences, Professor, Head of Department Digital Technologies in Energy, National Technical University of Ukraine «Igor Sikorsky Kyiv Polytechnic Institute», Kyiv, Ukraine;
e-mail: auscheva@gmail.com; ORCID Author ID: https://orcid.org/0000-0003-0816-2971;
Scopus ID: https://www.scopus.com/authid/detail.uri?authorId=57210707106.

Шаповалова Світлана Ігорівна – кандидат технічних наук, доцент, доцент кафедри цифрових технологій в енергетиці, Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського», Київ, Україна;
Svitlana Shapovalova – PhD, Associate Professor, Associate Professor of Department Digital Technologies in Energy, National Technical University of Ukraine «Igor Sikorsky Kyiv Polytechnic Institute», Kyiv, Ukraine;
e-mail: LanaShape@gmail.com; ORCID Author ID: https://orcid.org/0000-0002-3431-5639;
Scopus ID: https://www.scopus.com/authid/detail.uri?authorId=35367847700.

Петренко Катерина Володимирівна - молодший науковий співробітник, Інститут відновлюваної енергетики Національної академії наук України, Київ, Україна;
Kateryna Petrenko - junior researcher, The Institute of Renewable energy of the National Academy of Sciences of Ukraine, Kyiv, Ukraine;
e-mail: Katerynkapetrenko@gmail.com; ORCID Author ID: https://orcid.org/0009-0004-4658-9946;
Scopus ID: https://www.scopus.com/authid/detail.uri?authorId=57219715314.

Кардашов Олександр Вадимович – аспірант, Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського», Київ, Україна;
Oleksandr Kardashov – Post-graduate student, National Technical University of Ukraine «Igor Sikorsky Kyiv Polytechnic Institute», Kyiv, Ukraine;
e-mail: alexanderkardashov3@gmail.com; ORCID Author ID: https://orcid.org/0000-0003-1767-7846.

Софієнко Антон Юрійович – аспірант, Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського», Київ, Україна;
Anton Sofiienko – Post-graduate student, National Technical University of Ukraine «Igor Sikorsky Kyiv Polytechnic Institute», Kyiv, Ukraine;
e-mail: anton.sofiienko@gmail.com; ORCID Author ID: https://orcid.org/0009-0002-7337-2203.

Побудова моделі просторового розподілу характеристик вітрової енергії

Н. М. Аушева, С. І. Шаповалова, К. В. Петренко, О. В. Кардашов, А. Ю. Софієнко

Анотація. Метою статті є створення моделі просторового розподілу характеристик вітрової енергії на території України. **Предмет дослідження** - набори даних значень швидкості вітру, методи кореляції, валідації та інтерполяції даних. **Результати дослідження.** На основі наборів даних реаналізу NASA та результатів вимірювань на 70 метеорологічних станціях України створено датасет пар відповідних даних швидкості вітру, які відносяться до однієї локації, але отримані різними способами. На основі порівняльного аналізу результатів розв'язання задачі регресії навченими на отриманому датасеті моделями машинного навчання обрано оптимальну за точністю (метрики RMSE, R2, коефіцієнта кореляції Пірсона) модель Random Forest для прогнозування зсуву значення швидкості вітру в даних реаналізу NASA для їх наближення до реального значення. Значення коефіцієнта кореляції Пірсона покращилась на 0,07 в гіршому випадку та 0,66 - в кращому. За прогнозами моделі Random Forest проведено корекцію всіх значень швидкості вітру в даних реаналізу NASA. Коректність скоригованих даних також доведено результатами дослідження динаміки зміни трендових значень впродовж року від трьох різних джерел даних: вимірів метеостанцій, даних реаналізу NASA та скоригованих даних реаналізу NASA. Методом універсального кригінгу здійснено інтерполяцію отриманих скоригованих значень у вузлах сітки на всю територію України. Коректність результатів інтерполяції перевірено методом кросс-валідації. **Висновок.** За отриманими результатами створено GIS - програмний ресурс, який дозволить визначати достовірні значення характеристик вітрової енергії в будь-якій заданій точці на території України. Запропонована GIS може використовуватись, насамперед, для проєктування вітрових електростанцій та вибору оптимальних місць їх розташування.

Ключові слова: машинне навчання; глобальні атмосферні моделі; MERRA-2; метеорологічні дані; дані повторного аналізу; інтерполяція швидкості вітру; метод зворотно зважених відстаней; універсальний кригінг.