BULLETIN OF THE POLISH ACADEMY OF SCIENCES
TECHNICAL SCIENCES, Vol. 73(5), 2025, Article number: e155038
DOI: 10.24425/bpasts.2025.155038

Wind power prediction in Poland using temporal fusion transformers and numerical weather prediction

Weronika JACHUŁA* **and Michał WYDRA**

Lublin University of Technology, Faculty of Mathematics and Applied Computer Science,
ul. Nadbystrzycka 38, 20-618 Lublin, Poland

Abstract. Predicting wind power generation is essential to ensure the stability and efficiency of power systems. Accurate predictions enable better planning and management of energy reserves, minimizing operational costs and helping grid operators mitigate the adverse effects of wind generation fluctuations. The primary objective of the presented study is to develop an accurate wind power prediction method and apply it to Poland's conditions. Among many emerging methods, the temporal fusion transformers (TFT) method is particularly well-suited for wind power generation forecasting, as it models complex, nonlinear dependencies in time series data. The TFT method combines self-attention mechanisms and recurrent networks, capturing long-term dependencies and short-term changes in input data. Additionally, TFT enables the effective use of contextual information, improving forecast accuracy. The numerical weather data was collected, and the feature extraction was performed. The features, such as time series data, have been used to train and test the different TFT networks. After the training and testing stage, an error analysis was performed. The final results showed similar or improved accuracy in wind generation forecasts compared to other methods in increased variability of weather conditions.

Keywords: wind power; prediction; temporal fusion transformers; machine learning.

1. INTRODUCTION

As wind energy plays an increasingly crucial role in the transition to renewable energy, its successful integration with the interconnected power grid is essential, given its variable and unpredictable nature. Lately, wind power forecasting has drawn significant interest from researchers, as it supports efficient grid management and stabilization by allowing for the optimized scheduling of controllable thermal power units [1, 2]. Conventional power plants generally have slower response times, making precise wind power forecasts essential. Accurate forecasts that extend 24 hours enable better thermal plant scheduling and overall grid optimization. Wind power forecasting often uses numerical weather prediction (NWP) models and location-specific and technical details about wind farms [3–5]. Recent advancements in computing have made it possible to enhance traditional wind power forecasting models with various machine learning (ML) techniques, improving prediction accuracy and reliability [6, 7]. Recent advancements in ML methods, especially attention-based transformers [8], improve the accuracy of long-term predictions [9–11]. Based on this, the authors present results on wind power predictions using transformer methods to enhance the accuracy of wind generation in Poland. Data were acquired from the Polish Transmission System Operator (TSO) and NWP. The novelty of the proposed method lies in the fact that the most significant predictors, such as maximal and minimal generation, were calculated and assumed to be known 60 hours ahead of the time horizon. Utilizing the NWP forecast of wind direction, atmospheric pressure, and temperature in all turbine coordinates, also known as predictors in forecasting horizon, significantly improved prediction accuracy.

2. METHOD

The publication of the so-called attention mechanism presented in [8] significantly added to the acceleration of the artificial intelligence technologies. The presented method effectively and efficiently captures dependencies in sequences without requiring recurrence or convolutions in neural network architectures. It revolutionized the development of neural network models, leading to the invention of temporal fusion transformer (TFT) models. For the presented work, the TFT models were developed and adapted from [12].

In the presented method, the authors constructed a database of wind turbine types and production curves in Poland, utilizing data obtained from manufacturers and investors. Based on data from the Polish Wind Energy Association, the locations (coordinates) of individual turbines were determined and ultimately verified using publicly available satellite images from Google Maps.

Using Numerical Weather Prediction (NWP) data sequentially obtained from the ICM (Interdisciplinary Center for Mathematical and Computational Modeling). The NWP is updated every 6 hours, with a forecast horizon of 60 hours and 10-minute granularity for every turbine location stored in the database, as shown in Fig. 2. Such an organized NWP enabled the calcula-

*e-mail: w.jachula@pollub.pl

Manuscript submitted 2024-02-10, revised 2025-05-21, initially accepted for publication 2025-07-04, published in August 2025.

tion of the theoretical power generation for all turbines, based on their production curves (Fig. 3). The resulting data were then aggregated in multiple contexts, creating additional input features for the TFT model, as shown in Table 1. The suffixes ('min', 'max', 'std', 'mean') refer to statistical values, specifically the minimum, maximum, standard deviation, and mean value, respectively. The prefix 'average' in the variable name (Table 1) refers to the average value calculated from all turbine locations at the same timestamp. Subsequently, the TFT model was trained using the actual values of total wind generation in Poland obtained from the National Power System (NPS).

Table 1
Input variables to the TFT model

Description	Variable name
Values of the average atmospheric pressure from a single hour	averagepressure_min averagepressure_max averagepressure_std averagepressure_mean
Values of the average temperature at a height of 90 m for a single hour	averagetemperature90m_min averagetemperature90m_max averagetemperature90m_std averagetemperature90m_mean
Values of the average speed of wind at a height of 90 m for a single hour	averagewind90m_min averagewind90m_max averagewind90m_std averagewind90m_mean
Values of the average wind direction at a height of 90 m for a single hour	averagewinddirection90m_min averagewinddirection90m_max averagewinddirection90m_std averagewinddirection90m_mean
Values of the wind power generation depending on wind gusts from a single hour	gust_min gust_max gust_std gust_mean
Values of the wind power generation depending on wind speed at a height of 90 m for a single hour	wind90m_min wind90m_max wind90m_std wind90m_mean
Values of the wind direction change at a height of 90 m for a single hour	winddirectionchange90m_min winddirectionchange90m_max winddirectionchange90m_std winddirectionchange90m_mean

Based on previous research in the field of wind power forecasting conducted for Poland [7, 13], the data used for this research were acquired from NWP and calculated according to Fig. 2 to determine the total wind power in Poland. The process begins by gathering the coordinates and technical specifications of the wind turbines. This data is then integrated with a numerical weather prediction model, which generates wind-related forecasts spanning 60 hours every 6 hours as shown in Fig. 1.

The NWP model is then supplemented with real-world wind power measurements provided by the National System Operator, PSE S.A.

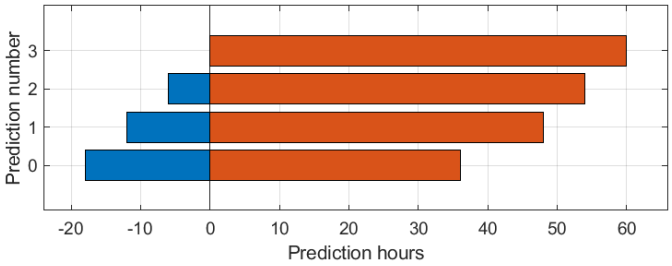


Fig. 1. Diagram illustrating how the prediction is made

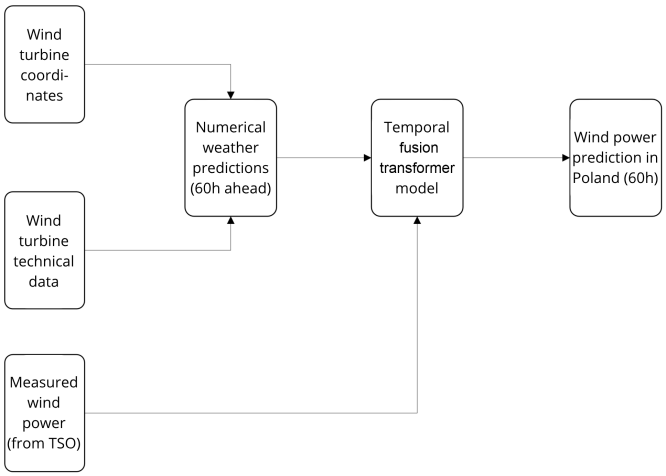


Fig. 2. Schema used in research for TFT model training and testing

Accurate forecasting of wind generation requires models that combine numerical weather prediction data with the geographical and technical details of wind farms. The proposed framework utilizes NWP data, including wind speed and gusts at 90 meters, air temperature, atmospheric pressure, and wind direction, in conjunction with coordinates of wind farms and turbine specifications, to estimate the upper and lower limits of wind generation. However, inaccuracies arise due to natural variability, limiting predictions to potential generation ranges within areas such as Poland.

The uncertainty arises from the reliance on turbine power curves and predicted wind conditions, as described in (1) and (2).

$$P_{\text{wind}} = \sum_{i=1}^{N} V_w \cdot f_i\big(V_w(j)\big), \tag{1}$$

$$P_{\text{gust}} = \sum_{i=1}^{N} V_g \cdot f_i\big(V_g(j)\big), \tag{2}$$

where N – number of wind turbines, V_w – wind speed at 90 meters altitude above ground level (m/s), V_g – wind speed during gusts (m/s), f_i – i-th wind turbine generation curve as a function of wind speed V_w or V_g at j-th time (Fig. 3), j – NWP wind forecast sample number.

Equations (1) and (2) are shown in general form because it is assumed that the f_i functions contain the whole production curve of wind turbines, containing the cut-in and cut-out wind

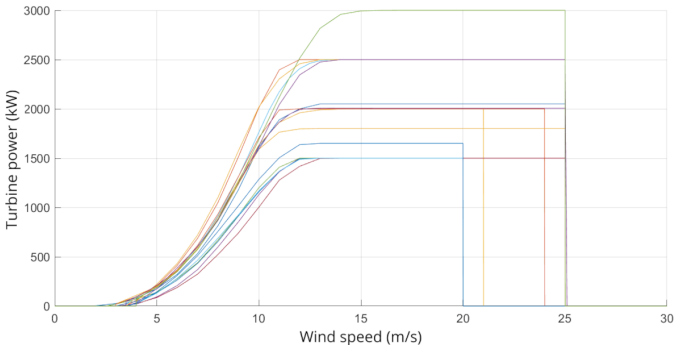


Fig. 3. Characteristics of chosen turbines used in research

speed points. Example characteristics of turbines used in the tests and simulations are shown in Fig. 3.

Grid-connected wind farms must align their production schedules with NWP forecasts, considering grid constraints set by the Transmission System Operator (TSO). Machine learning (ML) techniques improve forecast accuracy by integrating NWP data, turbine locations, and power curves with results compared to TSO-measured values. Analysis shows that predicted generation generally stays within calculated bounds.

To address the challenges of nonadaptive models, the authors propose an ML model with an adaptive-ensemble architecture, enhancing long-term forecast precision by accounting for nonlinear dynamics, which nonadaptive models struggle to capture [14].

The inputs enable the development of a temporal fusion transformer model. After thorough training and testing presented in this paper, the advanced model can forecast a 60-hour wind power generation with a high degree of accuracy compared to that generated by the standalone NWP model. As stated in the Introduction, the following features/predictors were prepared, calculated as averages across all wind turbine locations.

The input features for a temporal fusion transformer (TFT) provide the data necessary to identify temporal patterns and contextual relationships, enabling accurate forecasting. They capture historical trends, external influences, and any static factors that impact the target variable. By leveraging these inputs, the model learns complex dependencies and interactions. The set of predictors shown in Table 1 enabled a multidimensional approach, thereby improving the accuracy of the predictions.

2.1. Data used in the research

Figures from Fig. 4a to Fig. 4d show the characteristics of total wind power in Poland, measured by TSO, as a function of the above averages.

Data used in this research were gathered hourly from the end of October to the end of June. This period provides a valuable opportunity to observe how individual predictors behave under varying conditions, including seasonal and weather-related changes. Such variability significantly affects the accuracy and reliability of the prediction results, allowing for an evaluation of how well the model adapts to different influencing factors over an extended period.

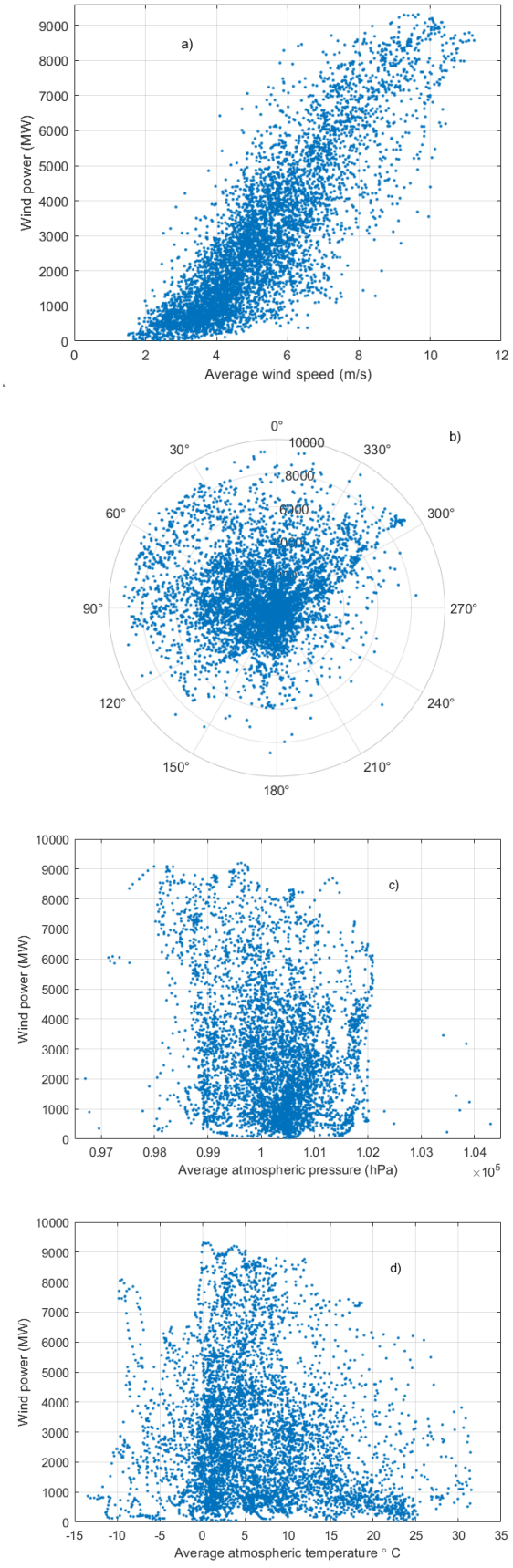


Fig. 4. Characteristics of total wind power in Poland: (a) average wind speed (m/s); (b) average wind direction (°); (c) average atmospheric pressure (Pa); (d) average atmospheric temperature (°C)

As illustrated in Figs. 4a and 5 and in Table 2, the maximum power generation occurs when the average wind speeds in Poland range between 4.5 and 9.5 m/s.

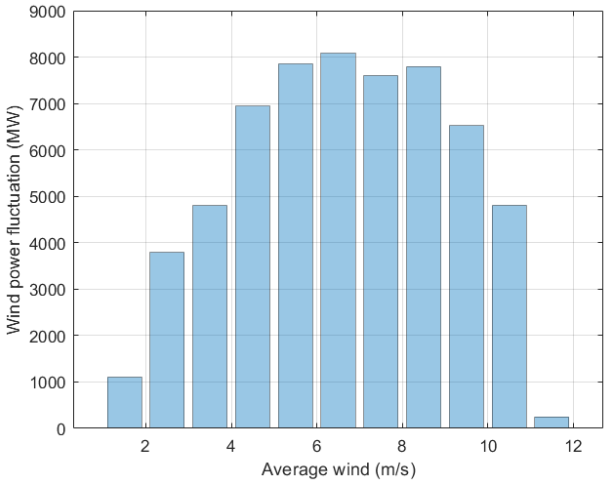


Fig. 5. Fluctuations of wind power as a function of average wind speed in Poland

Table 2
Values of wind power fluctuations

Average wind speed in Poland (m/s)	Wind power variability (MW)
1.5	1105.4
2.5	3793.1
3.5	4801.8
4.5	6945.6
5.5	7853.2
6.5	8083.7
7.5	7602.2
8.5	7795.0
9.5	6524.3
10.5	4806.5
11.5	251.2

Figure 4b shows that the highest power is produced when the wind comes from the northern direction. Meanwhile, Fig. 4c shows that the optimal air pressure for generating wind power is between 0.98 and 1.02 Pa. Finally, as depicted in Fig. 4d, the highest levels of wind energy generation are observed at average temperatures in Poland ranging from 0°C to 10°C.

These findings highlight the critical environmental factors that significantly influence the efficiency of wind power generation.

2.2. Model

The temporal fusion transformer (TFT) is a deep learning model for time series forecasting. It effectively combines the strengths of attention mechanisms and recurrent architectures to capture short-term patterns and long-term dependencies in sequential data. A key feature of the TFT is its ability to handle static, temporal, and time-varying covariates, allowing it to integrate various contextual information into the forecasting process [15]. The model employs attention mechanisms to identify the most relevant features and time points, thereby enhancing interpretability by highlighting which inputs contribute the most to the predictions. It also incorporates gating layers to enhance efficiency by selecting only the necessary information for each step, avoiding overfitting. With its ability to process complex and dynamic data, the TFT model excels in producing accurate forecasts for various applications [16].

Forecasting wind power generation using a temporal fusion transformer (TFT) model represents one potential approach to address the challenges of imbalances in electrical supply and demand. This is particularly important in power grids that rely heavily on wind power as a generation source, where variability and unpredictability can create significant risks [17]. By providing more accurate predictions of wind power output, the TFT model enables grid operators to better plan and balance energy production and consumption, thereby reducing the likelihood of disruptions or inefficiencies.

When creating the model, the previously mentioned predictor values were entered, as shown in Table 1.

An additional parameter, the time index, was introduced as the data was collected at hourly intervals. This index was designed to organize the data and facilitate easier work with it in the subsequent stages of analysis.

Models were developed with variations in how predictors were combined across two training parameters:

- Continuous variables that change over time and are known in the future.
- Continuous variables that change over time and are not known in the future.

Moreover, the models also include variables such as:

- The maximum amount of data that is passed to the model.
- The maximum prediction length.
- The maximum encoder length.

After training and testing various combinations of predictors and variable values, one set was selected for its superior performance, resulting in the lowest loss (Table 3).

In the next stage, four scripts were prepared that, based on the TFT model, perform predictions for a given time in specific time frames. Additionally, mean absolute percentage error (MAPE), mean absolute error (MAE), and absolute error (AE) are calculated.

$$\mathrm{MAPE} = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{P_i - A_i}{A_i}\right|, \tag{3}$$

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}|P_i - A_i|, \tag{4}$$

$$\mathrm{AE} = P_i - A_i, \tag{5}$$

where n – number of observations, A_i – actual value (MW), P_i – predicted value (MW).

Table 3
Set of training parameters

Parameter	Value
Time-varying known reals – a list of continuous variables that change over time and are known in the future	Time index, wind speed at 90 meters altitude, wind gust at 90 meters altitude, wind direction, atmospheric pressure, atmospheric temperature
Time-varying unknown reals – a list of continuous variables that change over time and are not known in the future	None
Time max – the maximum amount of data that is passed to the model	5760 hours
Max prediction length – the maximum prediction length	60 hours
Max encoder length – the maximum encoder length	300 hours

The four scripts represent configurations of a transformer model inspired by the architecture introduced in [8].

The transformer model utilizes multi-head self-attention and feedforward layers to process input data and capture long-range dependencies efficiently. In the presented work, the models vary based on two critical parameters: the hidden size (also referred to as the hidden continuous size) and the attention head size.

The hidden size and the hidden continuous size determine the dimensionality of the model embeddings and intermediate representations. These parameters are integral to both the self-attention mechanism and the feedforward layers of the transformer. The attention head size refers to the number of parallel attention mechanisms in the multi-head self-attention layer. Each head focuses on a different input subspace, allowing the model to attend to various aspects of the sequence. Together, these parameters influence the model capacity to learn and generalize (Table 4).

Table 4
Specification of created scripts: (a) – four attention head size + 160 neurons, (b) – 16 attention head size + 160 neurons, (c) – four attention head size + 320 neurons, (d) – 16 attention head size + 320 neurons

Model	Attention head size	Hidden size/Hidden continuous size
(a)	4	160
(b)	16	160
(c)	4	320
(d)	16	320

Scripts were prepared based on a modified PyTorch Forecasting model in Python 3.12. The training was conducted on a PC with a 64-bit operating system, 32 GB RAM, an Intel Core i7 processor, and an NVIDIA RTX A2000 12 GB GPU card.

The model was trained on a workstation with the aforementioned specifications, and the training process took six hours. For the trained model, generating a 60-hour ahead prediction required less than one minute. Reducing the values of max encoder length, hidden size, and attention head size did not yield the expected trade-off between training speed and the quality of the results.

3. RESULTS AND DISCUSSION

Once the training and prediction processes were completed, collective graphs and histograms were generated to visualize and communicate the results better. Figure 6 provides an overview of six sequential predictions, each initiated six hours apart, alongside the actual wind energy generation data.

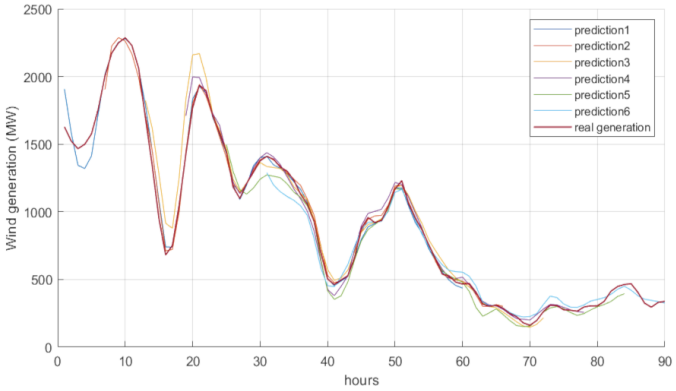


Fig. 6. Predictions plot against actual generation – forecasting using TFT model starts every six hours: shown predictions from 1 to 6, where every is of 60 hours length

Each forecast encompasses a temporal scope of 60 hours, providing a thorough comparative analysis of the forecasted and actual energy outputs over this duration. This enables an exhaustive assessment of the model alignment with empirical data. Concurrently, Fig. 7 meticulously delineates the errors associated with each prediction, presenting them separately to facilitate a more precise and detailed analysis in conjunction with the data depicted in Fig. 6. Figure 10 presents histograms that illustrate the distribution of the mean absolute error (MAE) and standard deviation for the various models under analysis.

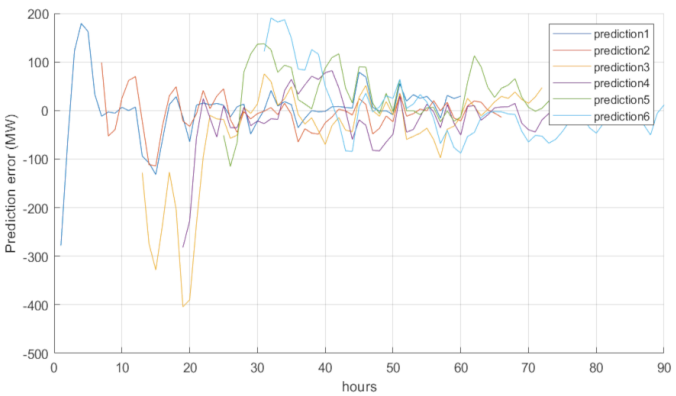


Fig. 7. The prediction errors plot of the TFT model started every six hours: predictions errors from 1 to 6 are shown, and colors correspond to predictions from Fig. 3

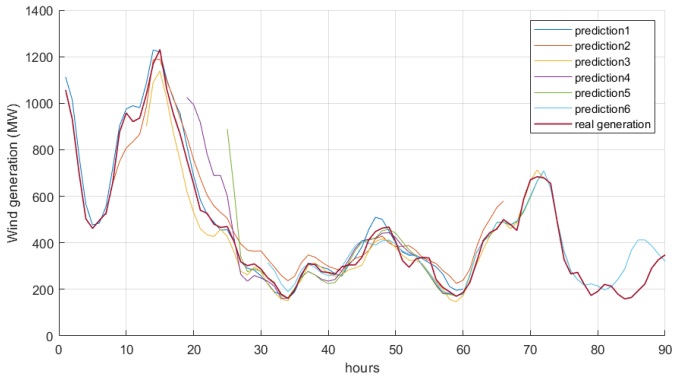


Fig. 8. Predictions plot against actual generation – forecasting using TFT model starts every six hours: shown predictions from 7 to 12, where every is of 60 hours length

These visual representations elucidate the variations in error magnitudes across the different forecasting methodologies. Specifically, models (a) and (b) display a propensity to overestimate wind power generation, whereas models (c) and (d) exhibit the contrary tendency, often underestimating the actual output. To augment the graphical insights provided in Fig. 9, Table 5 was compiled to present comprehensive numerical values pertaining to the performance of each evaluated model. By furnishing precise statistical data, this table facilitates a more comprehensive and quantitative assessment of model accuracy. It enables direct comparisons between the models, thus providing enhanced clarity regarding their respective strengths and weaknesses in predicting wind power generation.

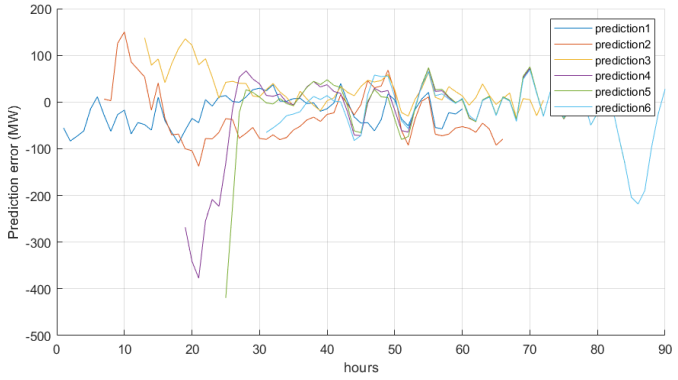


Fig. 9. The prediction errors plot of the TFT model started every six hours: predictions errors from 7 to 12 are shown, and colors correspond to predictions from Fig. 5

Table 5
Prediction absolute error (AE) statistics for a 60-hour-ahead prediction horizon

TFT model	Architecture	ΔP_{avg} MW	σ MW
(a)	4 attention head size + 160 neurons	−16.14	81.09
(b)	16 attention head size + 160 neurons	−26.06	102.43
(c)	4 attention head size + 320 neurons	2.94	105.57
(d)	16 attention head size + 320 neurons	9.65	100.26

In Table 6, there are the percentage error analysis histograms for the four investigated models.

Table 6
Prediction percentage error (PE) statistics for a 60-hour prediction horizon

TFT model	Architecture	MAPE	σ
(a)	4 attention head size + 160 neurons	–7.36%	15.2%
(b)	16 attention head size + 160 neurons	–13.12%	19.4%
(c)	4 attention head size + 320 neurons	–5.45%	18.3%
(d)	16 attention head size + 320 neurons	–2.47%	17.2%

Below is the feature importance analysis of the decoder variables (Fig. 11) and encoder variables (Fig. 12). As we can see in both cases, the values of wind speed at 90 meters altitude (mean, average, and standard deviation) contribute the most to the predictions. Additionally, the standard deviation of wind gusts at 90 meters altitude played a crucial role in creating predictions.

Wind generation prediction calculations were also performed for several new prediction techniques, including NHITS, DeepVAR, and DeepAR; the results are shown in Table 7. On this basis, it is possible to compare the results and clearly indicate that the TFT model proved to be the most effective.

Table 7
Comparison of prediction analysis for other new prediction techniques: nhits, deepvar, deepar

Model	ΔP_{avg}	σ	MAPE	σ
nhits	–499.01 MW	577.96 MW	–125.33%	180.95%
deepvar	51.40 MW	283.46 MW	30.48%	–0.98%
deepar	97.52 MW	251.17 MW	16.24%	25.47%

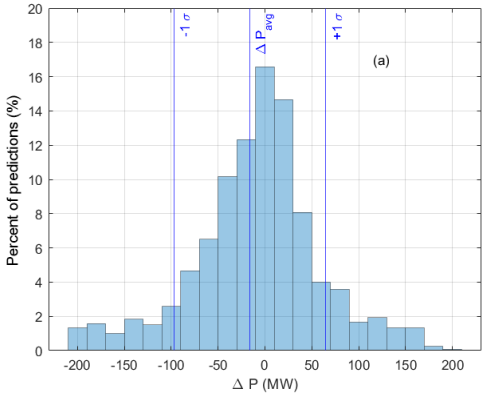
An ANOVA analysis was conducted, which confirmed the initial assumptions.

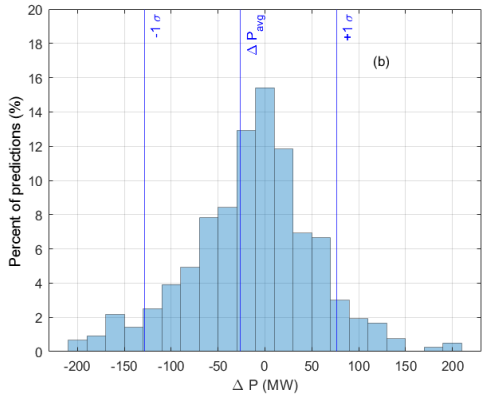
Firstly, the significant features are associated with variability in wind direction, average temperature, and average and minimum wind speed. Secondly, the model analysis demonstrated that the prediction accuracy is inversely proportional to variations in the above parameters.

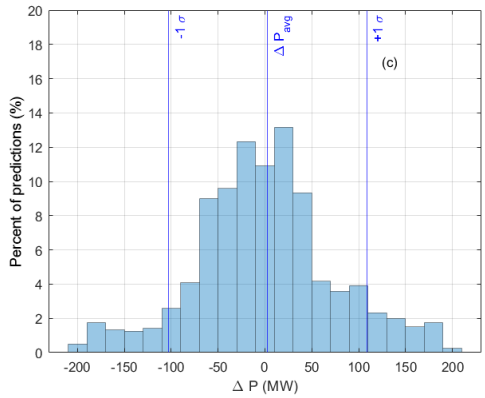
Specifically, minor changes in averaged values are weakly detected by the model, which is a consequence of the use of mean statistics that inherently possess a compensatory property.

Consequently, minor changes in the average may stem from large fluctuations in extreme values. Below are the results of the analysis for the model test predictions. The p-value indicates the probability that the observed result could have occurred by chance if there were no real effects. It is used to assess statistical significance (Table 8).

The F-statistics represent the ratio of between-group variance to within-group variance listed in Table 1, Fig. 12, and Fig. 13.







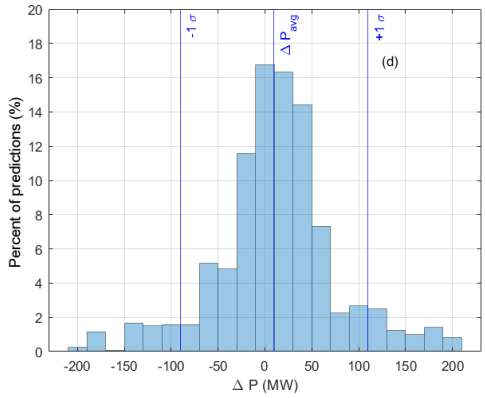
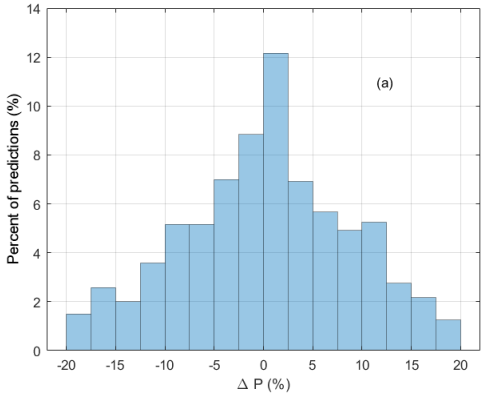
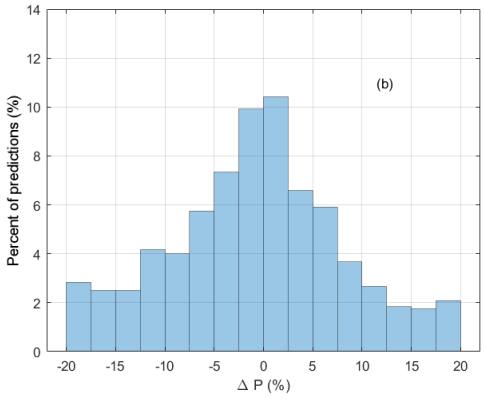
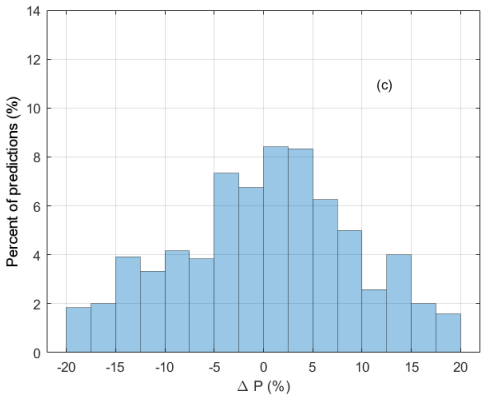


Fig. 10. Histograms of wind power prediction absolute errors (AE) in 60 hours ahead horizon for four different model configurations, respectively, (a), (b), (c), (d) to Table 5, where ΔP_{avg} – MAE: mean average error (MW), σ – STD: standard deviation of prediction error (MW)







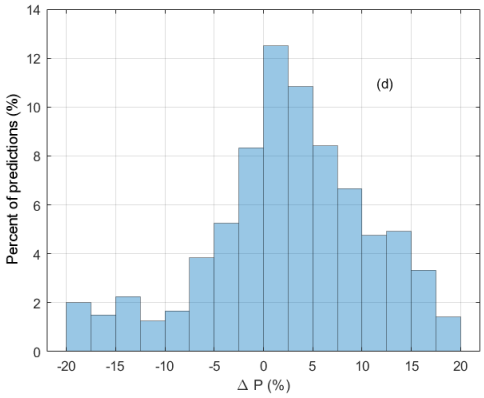


Fig. 11. Histograms of wind power prediction percentage errors (PE) in 60 hours ahead horizon for four different model configurations, (a) four attention head size + 160 neurons, (b) 16 attention head size + 160 neurons, (c) four attention head size + 320 neurons, (d) 16 attention head size + 320 neurons

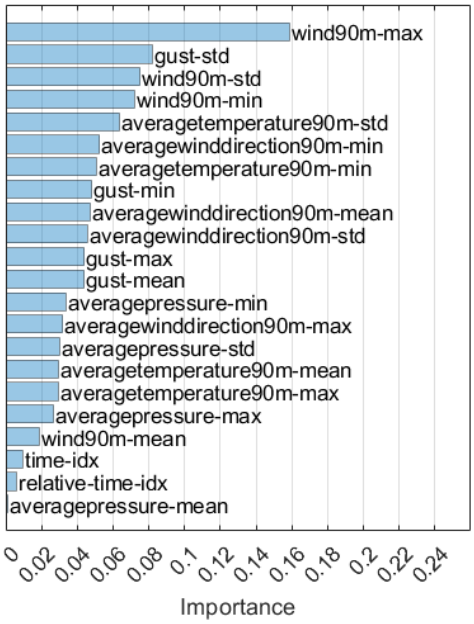


Fig. 12. The importance of decoder variables

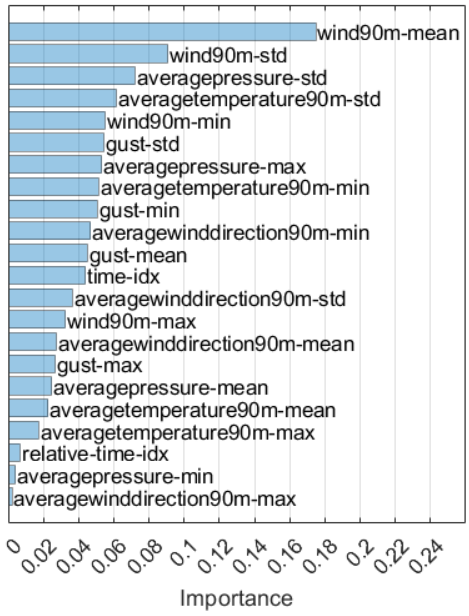


Fig. 13. The importance of encoder variables

Table 8
Results of ANOVA analysis

Feature	F-statistic	p value
winddirectionchange90m_min	14.227	$2.95 \cdot 10^{-6}$
winddirectionchange90m_mean	9.994	$9.84 \cdot 10^{-5}$
winddirectionchange90m_max	7.543	0.00083
averagetemperature90m_min	5.891	0.00365
averagewind90m_min	5.029	0.00803
averagetemperature90m_mean	5.009	0.00818
averagewind90m_mean	4.071	0.01953

Figure 14 presents the prediction trajectory compared to the actual values of the generated wind power (bottom), along with the corresponding variability in atmospheric conditions (top), indicating relatively higher variability of average weather conditions.

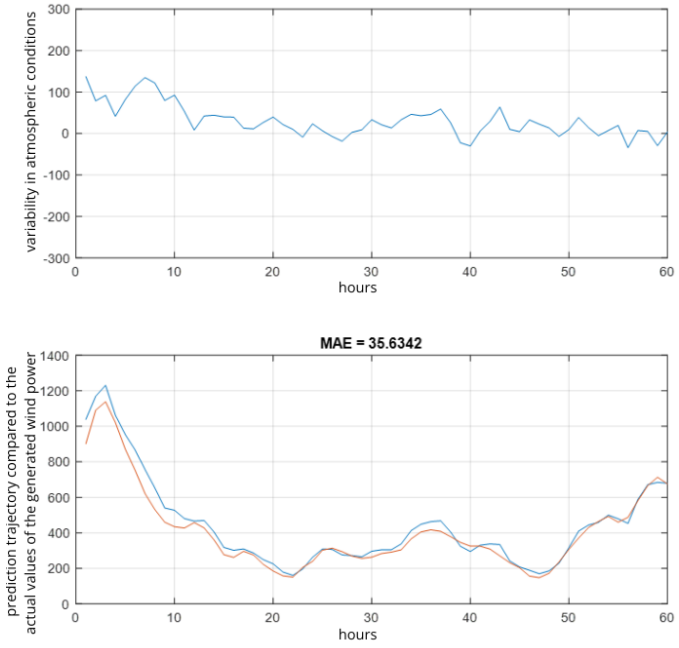


Fig. 14. Wind power prediction using TFT in higher variability weather conditions

In contrast, Fig. 15 illustrates the opposite case: the model exhibited greater difficulty in fitting the data, and the average atmospheric conditions were more stable.

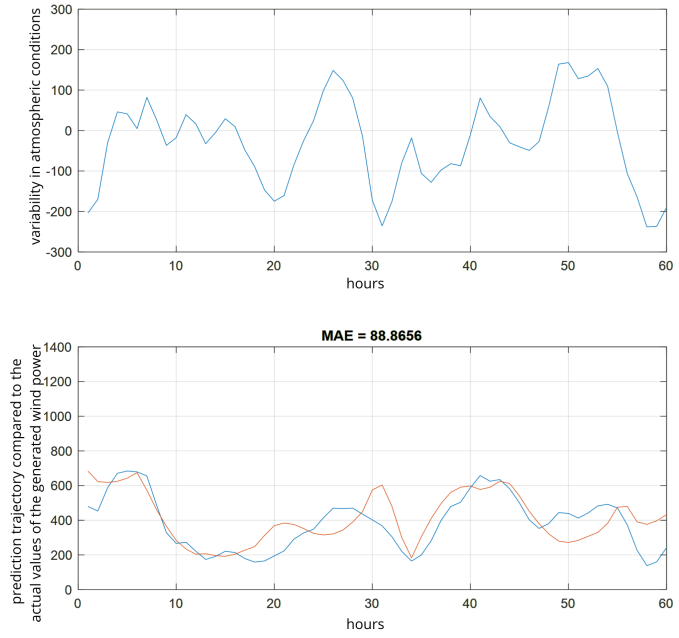


Fig. 15. Wind power prediction using TFT in lower variability of weather conditions

4. CONCLUSIONS

In the article, the authors presented a method for significant wind power forecasting using 60-hour numerical weather predictions acquired for wind farm locations in Poland. The authors proposed specific predictors derived from the preprocessing of NWP, including average wind speed, wind direction, air temperature at a 90-meter height, wind gust at a 90-meter height, and atmospheric pressure. Developed predictors have been implemented and combined with the temporal fusion transformer neural network. Four models have been investigated according to their architecture. The prediction results are presented in Figs. 8 and 9, along with the corresponding values in Tables 5 and 6. Errors within the range of ±100 MW are practically negligible in managing the Polish National Power System (NPS), representing approximately 0.54% in summer and 0.35% in winter, insignificant compared to the NPS load of 28 500 MW in winter and 18 500 MW in summer.

The percentage histogram analysis of error distribution directly correlates to Poland's actual wind generation capacity. Limit errors in the range of ±20% arise from comparing small-scale errors to relatively low generation levels – for instance, a 20 MW error represents 20% of a maximum generation of 100 MW. However, the analysis highlights symmetry around zero, with errors within ±2.5% being the most frequent, accounting for approximately 21% of all predictions across models (a), (b), and (d). This evaluation supports the selection of two complementary models: models (a) and (d). Overall, the presented TFT model satisfactorily predicts the power from wind farms in Poland in the increased variability of weather conditions.

REFERENCES

[1] R. Doherty and M. O'Malley, "A new approach to quantify reserve demand in systems with significant installed wind capacity," *IEEE Trans. Power Syst.*, vol. 20, no. 2, pp. 587–595, 2005, doi: 10.1109/TPWRS.2005.846206.

[2] B.C. Ummels, M. Gibescu, E. Pelgrum, W.L. Kling, and A.J. Brand, "Impacts of wind power on thermal generation unit commitment and dispatch," *IEEE Trans. Energy Conv.*, vol. 22, no. 1, pp. 44–51, 2007, doi: 10.1109/TEC.2006.889616.

[3] Y. Chang *et al.*, "A Hybrid Model for Long-Term Wind Power Forecasting Utilizing NWP Subsequence Correction and Multi-Scale Deep Learning Regression Methods," *IEEE Trans. Sustain. Energy*, vol. 15, pp. 263–275, 2024, doi: 10.1109/TSTE.2023.3283242.

[4] C. Ge *et al.*, "Middle-Term Wind Power Forecasting Method based on Long-span NWP and Microscale Terrain Fusion Correction," *Renew. Energy*, vol. 240, p. 122123, 2024, doi: 10.1016/j.renene.2024.122123.

[5] H. Zhang, Y. Liu, J. Yan, S. Han, L. Li, and Q. Long, "Improved Deep Mixture Density Network for Regional Wind Power Probabilistic Forecasting," *IEEE Trans. Power Syst.*, vol. 35, pp. 2549–2560, 2020, doi: 10.1109/TPWRS.2020.2971607.

[6] S. Hanifi, X. Liu, Z. Lin, and S. Lotfian, "A Critical Review of Wind Power Forecasting," *Energies (Basel)*, vol. 13, no. 15, p. 3764, 2020, doi: 10.3390/en13153764.

[7] M. Wydra, M. Kozlowski, D. Czerwinski, and B. Ksiezopolski, "Automated Adaptive-Ensemble Framework for Large Wind Power Prediction in Poland Using Deep Learning Models," *Adv. Sci. Technol. Res. J.*, vol. 16, no. 6, 2022, doi: 10.12913/22998624/156316.

[8] A. Vaswani *et al.*, "Attention is all you need," in *Advances in Neural Information Processing Systems*, 2017, pp. 5999–6009.

[9] M. Aslam, J.S. Kim, and J. Jung, "Multi-step ahead wind power forecasting based on dual-attention mechanism," *Energy Rep.*, vol. 9, pp. 239–251, 2023, doi: 10.1016/j.egyr.2022.11.167.

[10] Y. Hu, H. Liu, S. Wu, Y. Zhao, Z. Wang, and X. Liu, "Temporal collaborative attention for wind power forecasting," *Appl. Energy*, vol. 357, p. 122502, 2024, doi: 10.1016/j.apenergy.2023.122502.

[11] S. Mo, H. Wang, B. Li, Z. Xue, S. Fan, and X. Liu, "Powerformer: A temporal-based transformer model for wind power forecasting," *Energy Rep.*, vol. 11, pp. 736–744, 2024, doi: 10.1016/j.egyr.2023.12.030.

[12] "PyTorch Forecasting Documentation#." [Online]. Available at: pytorch-forecasting.readthedocs.io/en/stable/

[13] P. Kacejko and M. Wydra, "Wind Energy in Poland – Real Estimation of Generation Possibilities," *Rynek Energii*, vol. 91, no. 6, 2010.

[14] P. Kacejko and M. Wydra, "Wind Energy in Poland – Analysis of Potential Power System Balance Limitations and Influence on Conventional Power Units Operational Conditions.," *Rynek Energii*, vol. 93, no. 2, 2011.

[15] B. Wu, L. Wang, and Y.R. Zeng, "Interpretable wind speed prediction with multivariate time series and temporal fusion transformers," *Energy*, vol. 252, p. 123990, 2022, doi: 10.1016/j.energy.2022.123990.

[16] B. Lim, S. Arık, N. Loeff, and T. Pfister, "Temporal Fusion Transformers for interpretable multi-horizon time series forecasting," *Int. J. Forecast.*, vol. 37, no. 4, pp. 1748–1764, 2021, doi: 10.1016/j.ijforecast.2021.03.012.

[17] L. van Heerden, C. van Staden and H.J. Vermeulen, "Temporal Fusion Transformer for Day-Ahead Wind Power Forecasting in the South African Context," *2023 IEEE International Conference on Environment and Electrical Engineering and 2023 IEEE Industrial and Commercial Power Systems Europe (EEEIC/I&CPS Europe)*, Madrid, Spain, 2023, pp. 1–5, doi: 10.1109/EEEIC/ICPSEurope57605.2023.10194737.