BeFORECAST

Deliverable: Impact of most recent weather and WTG data on short-term forecasting

Deliverable No.: D3.3

DOCUMENT CONTROL SHEET

This project has received funding from the Energy Transition Fund of the Belgian Federal Public Service for Economy, SMEs, and Energy (FOD Economie, KMO, Middenstand en Energie)

DOCUMENT INFORMATION	
Document title	D3.3 Impact of most recent weather and WTG data on ST forecasting
Author(s), (organization)	Karthik Badarinath (3E), Hasan Yazicioglu (3E), Patrick Hoebeke (3E), Daniele D'Ambrosio (3E)
Deliverable No.	D3.3
Work Package No.	WP3
Lead beneficiary	3E
Dissemination level	Public
Date of issue	22/11/2024

DOCUMENT HISTORY		
Version	Status	Date
V1	Final	22/11/2024

DOCUMENT APPROVAL	
Date	Approver
22/11/2024	David Schillebeeckx
Date	WP leader
27/11/2024	KUL
Date	Coordinator
27/11/2024	VKI

DOCUMENT SUMMARY

This document summarizes the developments that are made in Task 3.6. Integration of most recent weather and wind turbine data for very short term forecasting (< 6h). The main outcome of this task is to develop machine learning and engineering algorithms to accurately perform short term forecasting and to detect wind ramps compared to solely relying on numerical weather forecasts (NWP) that do not assimilate operational WTG data and using solely operational WTG data and combination of both.

Table of Contents

1 Introduction

The most recent Numerical Weather Prediction (NWP) data cannot be directly applied to continuous wind power forecasting due to the inherent latency of 5-6 hours required for data delivery from the provider. Short-term wind power forecasting, which typically spans from a few minutes to several hours ahead, is crucial for addressing the challenges of wind energy integration into the power grid. This study aims to develop multiple AI-based models that incorporate real-time operational data from wind turbine generators (WTGs) to enhance power forecasting accuracy at the plant level. Additionally, each model's ability to estimate wind ramp events, characterized by their magnitude and duration, will be assessed.

1.1 Related work

Wind power forecasting models usually include physical models, statistical models, AI/DL models, and hybrid models. This literature survey explores these methods, their evolution, and recent advancements in wind power forecasting and ramp detection.

1.1.1 Physical Models

Physical models are based on numerical weather prediction (NWP) techniques, utilizing the physics of the atmosphere to predict wind speed and power. These models rely on inputs such as wind speed, temperature, air pressure, and terrain topography to predict wind patterns and, subsequently, the output of wind turbines. Physical models, like those implemented in the Weather Research and Forecasting (WRF) model, are widely used for medium- to long-term forecasting. However, they are computationally intensive and often suffer from low accuracy in short-term forecasting due to the nonlinear nature of atmospheric conditions (Giebel et al., 2011).

1.1.2 Statistical Methods

Statistical methods focus on finding relationships between historical wind power output and wind speed data. Autoregressive (AR), autoregressive integrated moving average (ARIMA), and persistence models have been traditionally used to model wind power. ARIMA models, for instance, are employed to capture temporal dependencies in wind power data. These models perform reasonably well in short-term forecasting, but their accuracy decreases with longer time horizons (Costa et al., 2008). Furthermore, these models assume linearity, which limits their ability to capture the complex and nonlinear nature of wind power fluctuations (Wang et al., 2014).

1.1.3 Machine Learning Approaches

Machine learning methods have gained traction due to their ability to model complex nonlinear relationships in wind power data. Techniques such as artificial neural networks (ANNs), support vector machines (SVMs), and random forests have been applied to wind power forecasting with promising results. ANNs, particularly, have shown high accuracy in short-term forecasts as they can learn from large historical datasets and adjust to new conditions quickly (Mohandes et al., 2004). In recent years, more sophisticated deep learning models, such as long short-term memory (LSTM) networks and convolutional neural networks (CNNs), have further improved forecasting performance by capturing both temporal and spatial dependencies in wind data (Zhang et al., 2016).

Recent advancements have introduced transformer-based architectures, which have demonstrated superior performance in wind power forecasting due to their ability to model long-range dependencies in time-series data more effectively than traditional recurrent models like LSTMs. Transformers, leveraging self-attention mechanisms, have been shown to outperform other methods by focusing on the most relevant parts of the input sequences, thus providing better forecasting accuracy (Chen et al., 2021). Attention mechanisms have been particularly useful in multi-step forecasting scenarios, where they help to capture the dynamic relationships between different time steps more efficiently (Zhou et al., 2021). These approaches have revolutionized forecasting by reducing computation time and improving scalability over traditional methods like CNN-LSTM hybrids.

1.1.4 Hybrid Models

Hybrid models combine the strengths of different forecasting approaches to enhance prediction accuracy. For example, hybrid models often integrate physical and machine learning methods, where NWP forecasts are used as inputs for machine learning models to fine-tune predictions. These models have demonstrated improved performance compared to standalone physical or statistical models (Bessa et al., 2017). Additionally, some hybrid approaches incorporate ensemble learning techniques, where multiple models are combined to reduce uncertainty and increase forecast robustness (Li et al., 2010).

CNN-LSTM hybrid models, which combine convolutional neural networks' ability to extract spatial features with LSTMs' temporal modelling capability, have shown superior accuracy in short-term wind power forecasting. By capturing both spatial correlations in wind farm data and temporal dependencies, CNN-LSTM models achieve higher accuracy and robustness, especially in volatile weather conditions (Liu et al., 2020). Further advancements in attention-based models have also led to the development of CNN-LSTM-Attention architectures, where attention mechanisms are used to focus on important time steps, further enhancing forecasting performance (Yan et al., 2020).

1.1.5 Recent Advancements and Trends

Recent advancements in wind power forecasting have focused on enhancing the accuracy of short-term predictions, which are critical for grid operation and market participation. Techniques like transfer learning and domain adaptation have been proposed to improve model generalization across varying geographical regions (Liu et al., 2020). The growth of big data and high-performance computing has enabled the use of large datasets to train more complex models, leading to significant improvements in forecast performance.

A notable development is the application of transformer-based architecture. These models leverage self-attention mechanisms to effectively model global dependencies in wind data. Transformer-based models have been increasingly applied to time-series forecasting, including wind power, due to their ability to capture long-range dependencies more efficiently than traditional recurrent neural networks (RNNs) like LSTMs, which often struggle with handling long-term dependencies and require recursive computation (Chen et al., 2021). For instance, the Hierarchical Spatial-Temporal Transformer Network (HSTTN), proposed by Zhou et al. (2023), applies transformer architectures specifically for long-term wind power forecasting, demonstrating significant improvements in capturing both spatial and temporal dependencies at a global scale.

Researchers are also exploring probabilistic forecasting, where uncertainty is explicitly modelled to provide grid operators with a range of possible outcomes, thereby enhancing risk management (Gönül & Sen, 2019). Probabilistic models are increasingly integrated with machine learning frameworks to improve the reliability and robustness of wind power forecasts.

Another emerging trend is the hybridization of models. Recent studies have combined different deep learning architectures to improve both accuracy and computational efficiency. For example, a model that integrates Conditional Generative Adversarial Networks (CGAN) with CNN-LSTM architectures has shown significant improvements in ultra-short-term forecasting accuracy. In this approach, CGAN handles missing data, while the CNN-LSTM extracts spatial and temporal features, followed by an attention mechanism to enhance convergence speed. This hybrid model has proven highly effective in real-world applications across multiple regions, outperforming standalone CNN-LSTM models (Zhao et al., 2023).

1.1.6 Wind power ramp detection

Wind power ramp detection has become an increasingly important area of research in the field of renewable energy. A wind power ramp is defined as a sudden variation in wind power output over a short period of time. The accurate detection and characterization of these ramps are crucial for grid stability, power balance, and efficient wind farm operation.

Early approaches to ramp detection often relied on binary threshold-crossing methods. These methods typically defined a ramp based on a specific percentage change in power output over a given time frame (Bossavy et al., 2012). However, these approaches were found to be limited in their ability to capture the diverse range of ramp events that occur in real-world scenarios.

To overcome the limitations of traditional methods, researchers have developed more sophisticated approaches:

1. Wavelet Transform: Continuous Wavelet Transform (CWT) has emerged as a powerful tool for ramp detection (Cheneka et al., Hannesdóttir et al.,). This technique allows for the decomposition of wind power time series into time-frequency space, enabling the identification of ramps across various timescales and magnitudes.
2. Optimized Algorithms: Zhang et al. developed an optimized swinging-door algorithm that allows for flexible adaptation of ramp definition parameters.

2 Data

Our dataset contains SCADA data from 1 Belgian offshore wind farm with 2 years of 10-minute SCADA data. We use Numerical Weather Prediction (NWP) forecasts for the corresponding periods for the identical locations. The NWP forecasts are derived from ECMWF IFS model. This data comprises computed forecasts of the wind speed, wind direction, temperature, specific pressure and air density at 1-h frequency.

2.1 Data cleaning

Historical SCADA data often contains measurements that are contaminated by faulty sensors, turbine downtime, or communication issues.

Filtering for flatlines and extreme values–

The dataset is refined by excluding turbine measurements that exhibit specific anomalies. Measurements are removed if the turbine of interest reports a negative wind speed or negative power production. Additionally, data points where the wind speed exceeds the turbine's cut-out speed or where power production is above 1% of the turbine's rated capacity are also excluded. This data cleaning step eliminates physically implausible or erroneous values, ensuring subsequent analysis is based on valid and realistic measurements only. Sensor value is stuck for 30 minutes or 6 consecutive 10-mintue timestamps (depending on the sensor), or more are also removed.

Filtering power curve outliers –

Data points associated with curtailment periods were then eliminated from the dataset. An envelope filtering technique was applied to remove additional outlier datapoints, utilizing bounds defined based on the contractual power curve, as illustrated in Figure 2. The right bound is defined by the curve x_{rb} =1.01 x_{nom} + 0.05 and y_{rb} =0.99 y_{nom} – 2.0 and the left bound is defined by the curve x_{lb} =0.99 x_{nom} - 0.05 and y_{lb} = 1.01 y_{nom} + 2.0 where x_{nom} and y_{nom} represent the nominal values from the contractual power curve [2]. Following the automated filtering, the power curve was again manually inspected to ensure no outliers were overlooked.

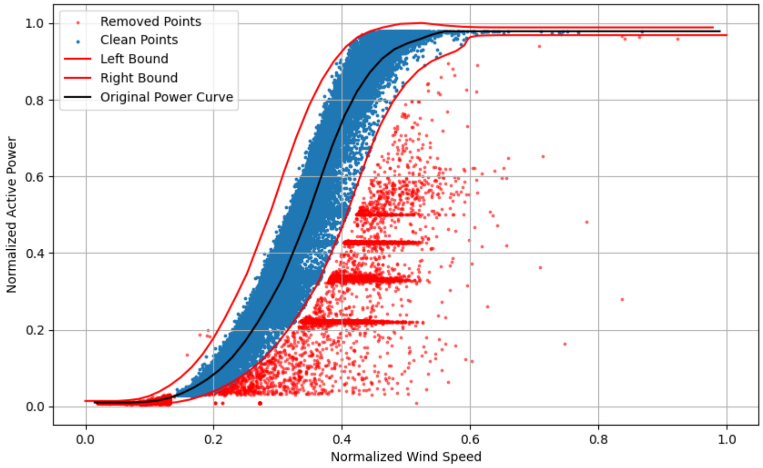

Figure 1: Power curve of a wind turbine, where blue points indicate clean datapoints retained for analysis, while red points represent measurements excluded from the dataset.

2.2 Missing data imputation

Missing data poses significant challenges to effective data analysis. Most data mining algorithms and techniques lack built-in mechanisms to directly handle missing values. This limitation is particularly problematic during the training phase, where the presence of missing data can substantially impact model performance and reliability. Consequently, appropriate handling of missing values is a crucial preprocessing step to ensure robust and accurate results in data mining applications.

We design an algorithm to address missing data in wind turbine datasets, building upon the work of Morshedizadeh et al (2018). The method has been enhanced through the incorporation of iterative imputation techniques and random forest regression.

1. Feature Enhancement: For each feature requiring imputation, the dataset is augmented with contextual information:

    - Site-wide daily statistics: Calculated mean and standard deviation of the feature across all turbines for each day.
    - Turbine-specific statistics: Computed mean and standard deviation of the feature for individual turbines. These additional features provide crucial context for the imputation process, capturing both overall site conditions and individual turbine characteristics.

2. Model Development: A separate Random Forest Regressor model is developed for each feature with missing values:
    - Initial imputation: IterativeImputer with RandomForestRegressor is applied to handle any remaining NaN values in the training data.
    - Data normalization: StandardScaler is employed to normalize feature ranges.
    - Model training: A RandomForestRegressor is trained on the prepared data to predict the target feature.
3. Imputation Implementation: The algorithm processes each row in the incomplete dataset sequentially:

    - For each missing value, input data is prepared using the same methodology as in the training process (feature enhancement, initial imputation, and scaling).
    - The corresponding trained model is then utilized to predict the missing value.
    - If more than 3 hours of data is continuously missing, no data imputation is done.

The cleaned farm power curve used for training of AI models is shown in Figure 2.

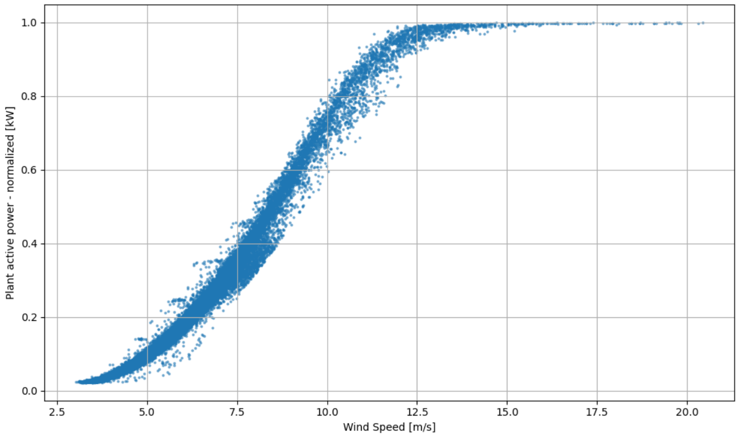

Figure 2: Farm power curve used to train AI models

3 Development of forecast models and ramp detection method

Various approaches are tested for predicting wind power at the park level, ranging from simple baseline models to sophisticated deep learning architectures. These approaches can be categorized into three main groups:

1. Single-source models:

    - Baseline models using only NWP forecast data with power curve conversion
    - ARIMA models utilizing only SCADA historical data
    - Classical machine learning models trained on NWP forecast data

2. Deep learning architectures:

    - Advanced neural networks designed to naturally combine both NWP and SCADA data
    - Architectures including LSTM, CNN-LSTM, CNN-LSTM with Attention, and Transformers

The flowchart in Figure 3 illustrates a comprehensive process for developing and evaluating wind power prediction models using deep learning techniques and ramp detection. The process begins by combining SCADA (operational) data with numerical weather prediction (NWP) forecasts. This merged dataset undergoes cleaning and is then split into training/validation and test sets, both of which are normalized. The training set is used to develop and validate various deep learning models, including stacked LSTM, CNN-LSTM, CNN LSTM with Attention, and Transformers. Meanwhile, the test set is fed into the developed wind power prediction models. The models' performance is then evaluated using several metrics: Mean Absolute Error (MAE), Root Mean Square Error (RMSE), Mean Squared Error (MSE), and R-squared (R2). The model predictions are then used to detect wind power ramps.

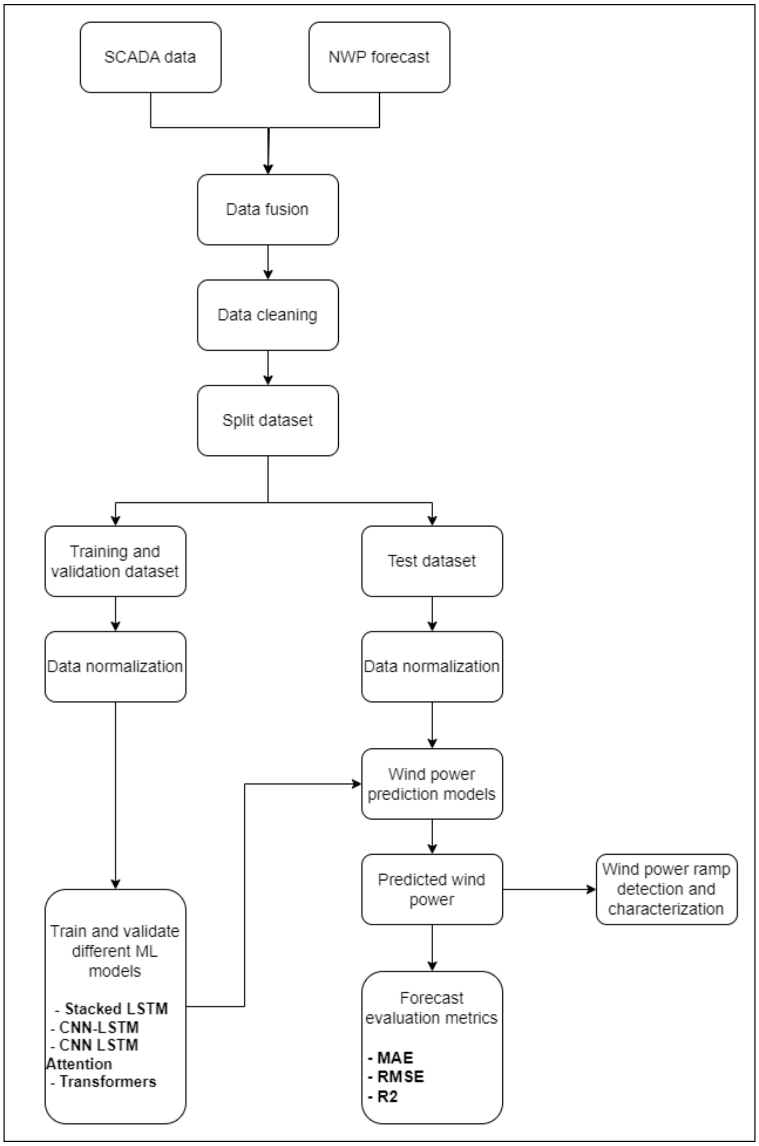

Figure 3: Workflow for wind power forecasting and ramp detection combining NWP forecast and SCADA data.

3.1 Combining NWP and SCADA data

The following variables are used to predict next 12-hour wind farm power production –

- SCADA – wind speed, active power
- NWP (IFS forecast) – air density, temperature (2m), wind speed (100m), wind direction (100m)

To optimize data fusion, we adjusted the time resolutions of our two data sources: the ECMWF IFS forecast (originally 1-hour intervals) and SCADA measurements (originally 10-minute intervals). We standardized both to 30-minute intervals, downsampling the higher-resolution SCADA data and upsampling the IFS forecast data. This approach maximizes the number of usable SCADA measurements while maintaining consistent time intervals for analysis.

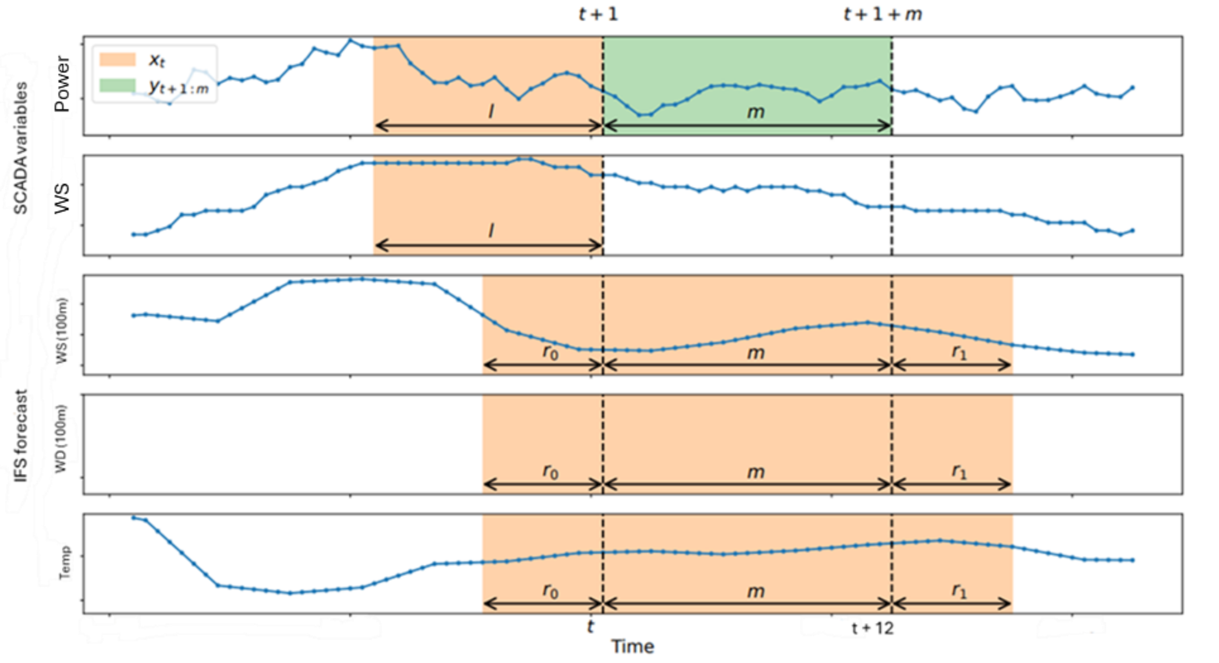

Figure 4: Summary of time windows used for wind power prediction. Adapted from Bouche, D et al (2023)

Dataset building – The formulation of the dataset for input into the ML models are adapted from the work of Bouche D et al., 2023 as show in Figure 4.

Consider a forecast that predicts m future values of wind speed or wind power, where m is 11 hours ahead in our case. While we naturally include the corresponding ECMWF forecast variables for this period, Bouche D et al., 2023 found that incorporating additional data points improved performance. Specifically:

- r0: represents how many past ECMWF predictions we include
- r1: represents how many additional future ECMWF predictions we include beyond the m prediction points
- l: represents how many past SCADA observations we include

The following values are used in this case: m=22, l=8, r0=2, r1=2. With data being sampled at 30-minute intervals, 11 hours of potential future wind power is predicted.

As shown in Figure 4, we combine all relevant observations (x_t) from these time windows (marked in orange). The model then uses this input to predict $y_{t+1:m}$, which represents m future values (shown in the green zone).

3.2 Data driven forecast models

The following models are developed to assess their performance in predicting wind power using NWP forecasts and SCADA data.

Baseline model (Wind speed – power curve model) – The baseline model uses a simple power curve, provided by the turbine manufacturer, to convert raw weather forecast data into predicted wind power output. For any given time, it takes the forecasted wind speed and calculates how much power the turbine should generate, without applying any additional corrections. For the baseline model, SCADA data is not used.

ARIMA (AutoRegressive Integrated Moving Average) is a statistical model for analyzing and forecasting time series data that combines three components: AutoRegression (AR) which uses past values to predict future ones, Integration (I) which differences the data to make it stationary and Moving Average (MA) which incorporates past forecast errors into the model. ARIMA is particularly effective for linear, univariate time series forecasting when the data can be made stationary through differencing, though it has limitations with non-linear patterns and multiple variables. In our case, the only input to ARIMA is total power generated by the wind farm for the given time t.

Classical ML models – Multiple classical ML models such as Random Forest regression, kNN regressor, gradient boosting, and so on is experimented with to predict wind power using just IFS forecast as input. This is done using PyCaret python library. No SCADA data is used in this model as classical ML models do not handle multi horizon forecasting naturally. Traditional ML models are often designed to make single-step predictions rather than forecasting multiple future steps simultaneously. This makes them less suitable for true multi-horizon forecasting.

We also explore some modern deep learning architectures which are designed for long term multi-horizon forecasting which can also help combine both NWP forecasts and SCADA more naturally. The following architectures are explored -

Long Short-Term Memory (LSTM) networks are a type of recurrent neural network specifically designed to handle long-term dependencies in sequential data. They use a gating mechanism consisting of input, forget, and output gates that control information flow through the cell state. This architecture allows the network to selectively remember or forget information over long sequences, making them particularly effective for tasks like time series prediction.

CNN-LSTM combines the spatial feature extraction capabilities of Convolutional Neural Networks (CNN) with the temporal modeling of LSTMs. In this architecture, the CNN first processes input data to extract relevant spatial features, which are then fed into an LSTM layer for temporal processing. This combination is particularly powerful for tasks that require both spatial and temporal understanding.

CNN-LSTM with Attention adds an attention mechanism on top of the CNN-LSTM architecture to help the model focus on the most relevant parts of the input sequence when making predictions. The attention layer computes importance weights for different parts of the input sequence, allowing the model to dynamically focus on relevant features at each time step. This addition is particularly useful for complex sequence modeling tasks where different parts of the input sequence have varying levels of importance for the final prediction, such as in time series forecasting with multiple variables.

Temporal Fusion Transformers (TFT) (Lim, B et al. 2021) represent a more modern architecture specifically designed for multi-horizon forecasting with multiple variables. TFTs incorporate several innovative components: variable selection networks to determine relevant inputs, gating mechanisms to suppress irrelevant components, and a temporal attention layer that can learn complex temporal relationships across different time steps. Unlike simpler architectures, TFTs can handle static covariates, known future inputs, and unknown future inputs separately, making them particularly effective for real-world forecasting problems where different types of information are available at different times. TFTs are also interpretable.

3.3 Forecast evaluation metrics

The following forecast evaluation metrics are used to assess the model performance – Mean Absolute Error (MAE), Root Mean Squared Error (RMSE) and R-squared (R^2).

Mean Absolute Error (MAE) - MAE measures the average magnitude of errors in a forecast, without considering their direction. It is calculated as the average of the absolute differences between the predicted values and the actual values.

Root Mean Square Error (RMSE) - RMSE is the square root of the average of squared differences between predicted and actual values.

R-squared (R^2) - R^2 measures the proportion of variance in the dependent variable that is predictable from the independent variable(s)

Multi-horizon metrics (RMSE, MAE, R^2) are calculated by comparing predicted and actual values at each forecast horizon across all timestamps in the dataset. For a 6-hour forecast, all 1-hour-ahead predictions are compared with their corresponding actuals in horizon 1, all 2-hour-ahead predictions are compared in horizon 2, and so on up to horizon 6. Each horizon's metrics are calculated independently - so horizon 1 RMSE represents the root mean square error of all 1-hour-ahead predictions, horizon 2 RMSE for all 2-hour-ahead predictions, etc. The overall metrics (like overall RMSE) are calculated by pooling all errors across all horizons together, effectively treating each horizon's prediction-actual pair as an independent observation. This approach allows to evaluate how forecast accuracy degrades (or maintains) across different prediction horizons while still getting a comprehensive view of the model's overall performance through the pooled metrics.

3.4 Power ramp detection algorithm

The methodology to detect power ramps consists of three main steps, and is a slightly modified version of Hannesdóttir et al., 2019.

First, the power data is normalized by its mean and standard deviation to ensure consistent analysis across different scales. Then, a continuous wavelet transforms (CWT) using a Gaussian wavelet is applied to the normalized data, with the wavelet coefficients normalized by the square root of their respective scales. The averaged power of the wavelet coefficients is analysed to identify significant peaks that exceed a prominence threshold of 0.1, with a minimum separation of 2-time steps between peaks. Prominence measures how much a peak stands out relative to its surrounding values.

For each identified peak, the method fits an idealized ramp function based on the error function, erf to a window of data cantered on the peak. The window size is proportional to the dominant wavelet scale at that peak. The ramp function characterizes the transition between two power levels with parameters for timing and transition speed. For multiple qualifying ramps, the one with the largest percentage change is selected as the best ramp event, with its magnitude, duration, rate of change, and timing recorded.

4 Evaluation of the models

The performance of forecast models and their capability to detect ramp events is detailed in this section.

4.1 Forecast performances of different models

The analysis of multiple time series prediction models for power forecasting reveals significant differences in performance across various prediction horizons. Through examination of multiple performance metrics (Table 1) including Root Mean Square Error (RMSE), Mean Absolute Error (MAE), R² score, and time-horizon specific performance (Figure 5), and visual time series data for select few days (Figure 6), we can draw several conclusions about model effectiveness and reliability.

Table 1: Values of evaluation metrics for 6 months of test data for 12 hour forecast horizon.

	RMSE [kW]	MAE [kW]	R2 score
LSTM	678.1	429.17	0.77
CNN-LSTM	596.97	380.82	0.82
CNN-LSTM-Attention	591.44	374.05	0.83
TFT	564.72	354.88	0.85

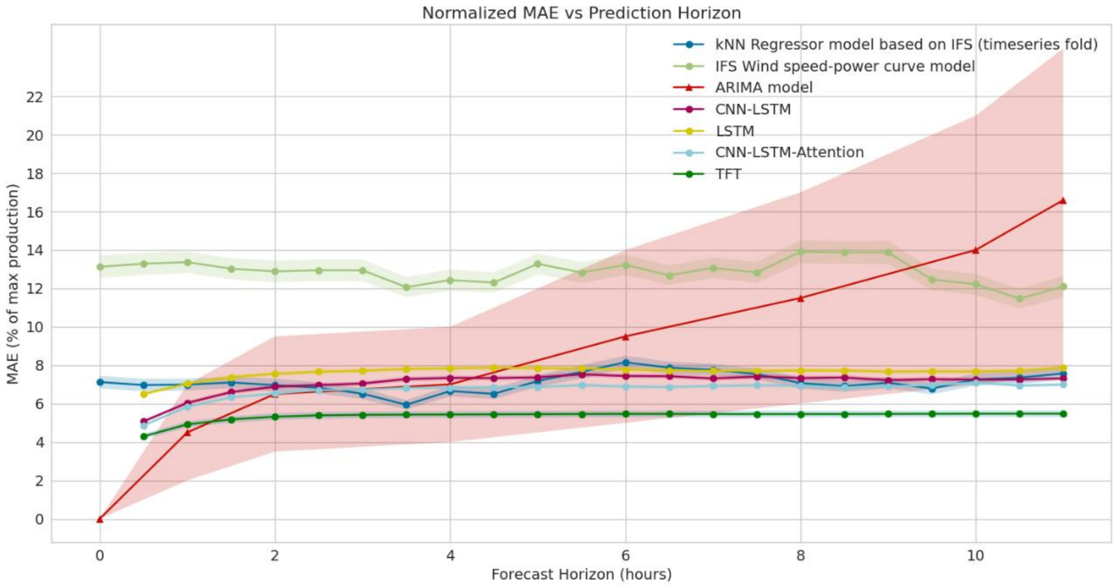

Figure 5: Normalized MAE vs forecast horizon. TFT shows the lowest and most consistent MAE over 10+ hour horizon.

The Temporal Fusion Transformer (TFT) emerges as the best model across all performance metrics. With the lowest RMSE of 564.72 kW, lowest MAE of 354.88 kW, and highest R² score of 0.85, TFT demonstrates superior prediction accuracy. With consistently low Mean Absolute Error (MAE) of about 5-6% across all prediction horizons, TFT exhibits impressive resilience even as other models begin to decline.

The CNN-LSTM and CNN-LSTM-Attention models demonstrate similar performance, achieving nMAE values around 7-8% and remaining relatively stable across different prediction horizons, though they lack the same level of consistency as TFT. The standard LSTM model falls slightly behind its CNN-enhanced variants, displaying greater prediction variance and a gradual performance drop over longer horizons. CNNs, known for detecting local patterns and sudden changes in data (Lin et al., 2017), excel due to sliding window operations (convolutions) that effectively capture abrupt shifts or features in the input, making them more responsive to immediate changes in the time series compared to LSTMs.

Traditional statistical and ML approaches, including the KNN Regressor, IFS Wind speed-power curve model, and ARIMA model, demonstrate the limitations of conventional methods in this complex prediction task. While the KNN Regressor maintains moderate but consistent performance, the ARIMA model has error increases over longer prediction horizons. This performance gap highlights the advantages of modern deep learning approaches in handling the complicated prediction tasks.

The impact of prediction horizon length on model performance is particularly revealing. While all models show some degree of error increase as the prediction horizon extends, the rate and severity of this degradation vary significantly. Traditional statistical models, especially ARIMA, exhibit performance deterioration over longer horizons, with their confidence intervals widening substantially. In contrast, deep learning models, particularly the TFT, maintain stable performance even at extended horizons of 10+ hours.

For practical implementation, these results strongly suggest adopting the TFT as the primary prediction model, with CNN-LSTM variants potentially serving as backup systems.

The uncertainty analysis reveals that TFT maintains the narrowest confidence band among all models, indicating highly consistent predictions. This contrasts sharply with the ARIMA model's wide confidence intervals, which suggest high prediction uncertainty. The neural network-based models generally show moderate confidence bands, positioning them between these extremes.

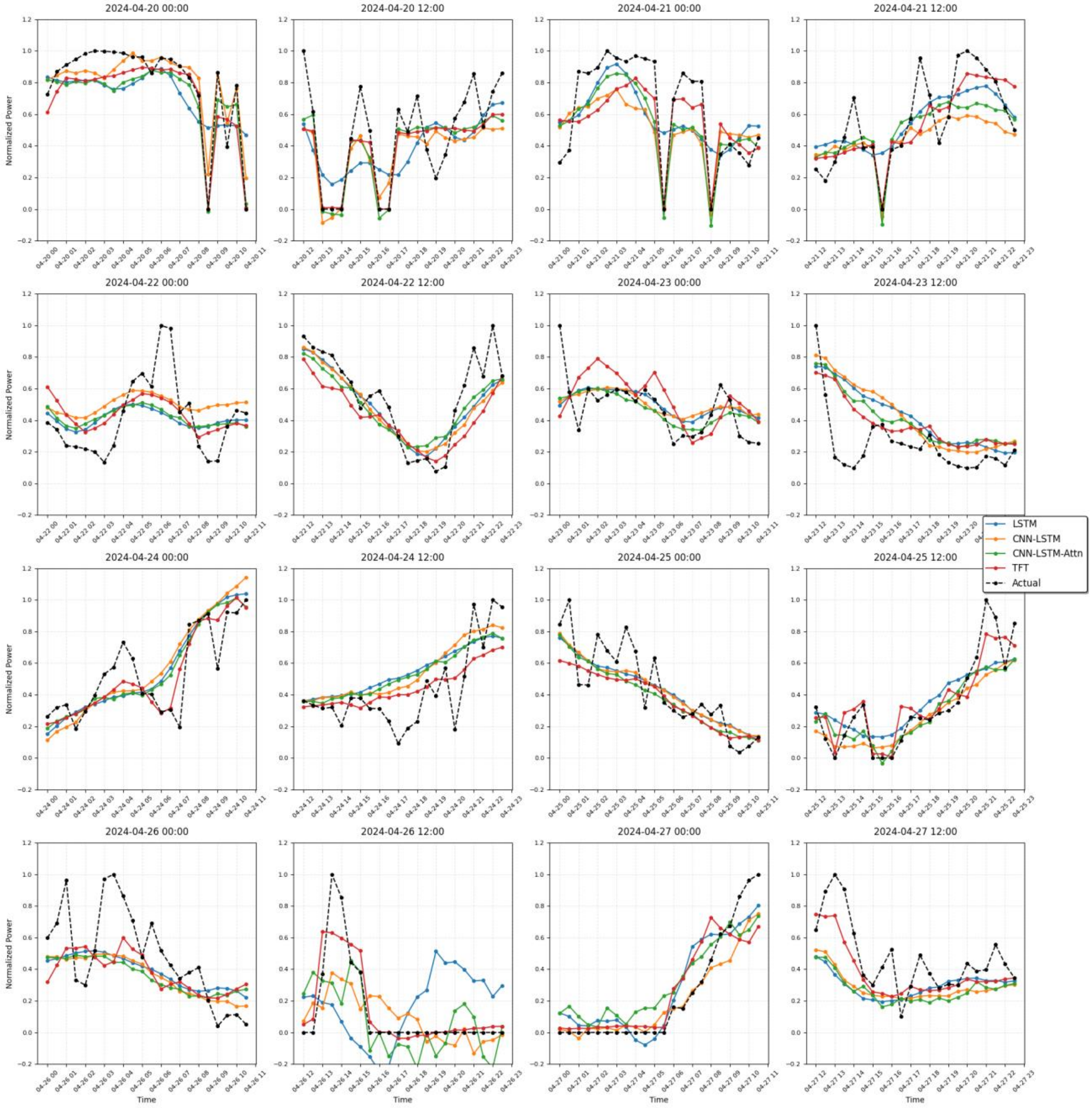

Figure 6: Normalized power prediction of different models for selected run datetimes.

4.2 Best model (TFT) interpretation

Our analysis confirmed that the Temporal Fusion Transformer (TFT) outperformed other models in prediction accuracy. To better understand its effectiveness, the model's interpretability is now examined, with the input features that most strongly influence its predictions being specifically analyzed.

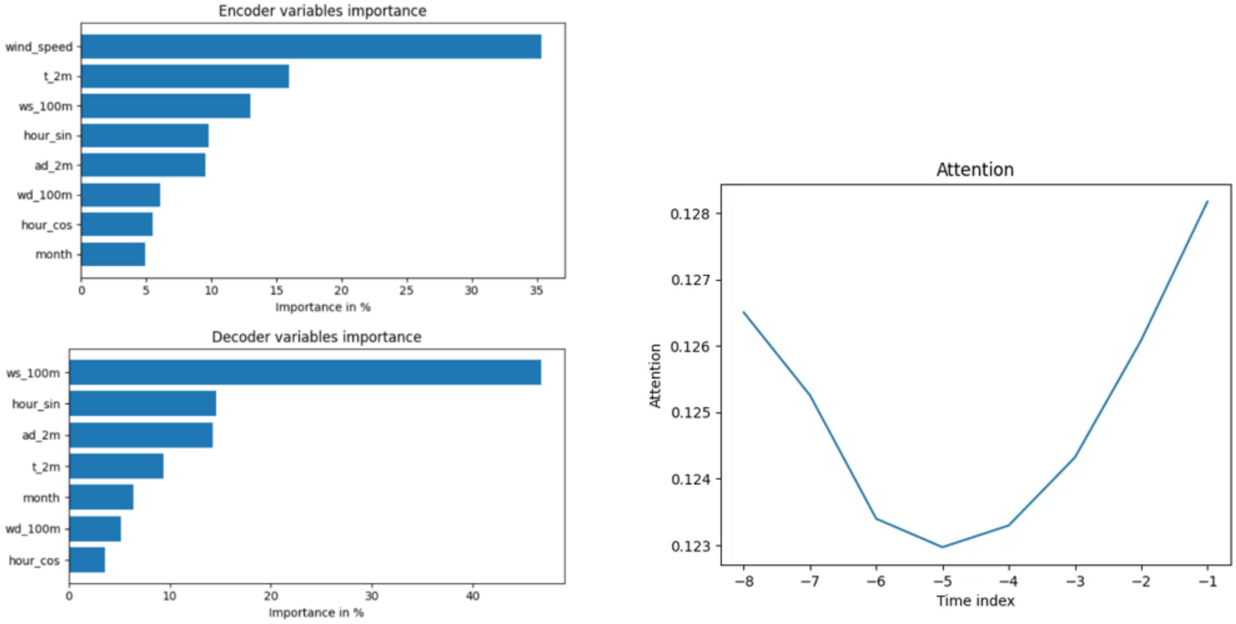

Figure 7: Feature importance (left) and temporal attention (right) of different input variables for TFT model.

In Figure 7, the encoder section (past data), SCADA wind speed emerges as the dominant predictor, contributing about 35% to the model's decisions. Temperature at 2m height and wind speed at 100m form the second tier of importance, each accounting for roughly 15% of the model's attention. Active power data is not shown in the feature importance as TFT model works in such a way that previous values of the predictor variable is always considered.

The decoder variables (future data) show a different priority pattern. Here, IFS forecast wind speed at 100m height has approximately 45% importance. Wind direction forecast seems to be less important to predict wind power in case of TFTs. This should further be validated with more datasets.

Temporal attention pattern shows stabilization of the attention weights across all available timestamps with the variation between 0.123 to 0.128. This U-shaped distribution suggests the model places slightly more emphasis on both the most recent and oldest timestamps in its prediction window, though the difference is minimal.

The overall stability in attention weights indicates that the model considers all historical timesteps to be similarly valuable for prediction, rather than heavily favouring specific time points. This balanced approach to temporal information processing likely contributes to the model's robust performance.

4.3 Ramp detection

There is no accepted definition or classification of wind power ramps except that they are manifested in terms of a significant change in production over a relatively short time. The quantification of the duration and magnitude of wind power ramps are arbitrary. Hence, we explore many different conditions for power ramps based on change in percentage of magnitude with respect to the rated power of the wind farm and with different durations.

The algorithm matches actual and predicted ramps by comparing three key characteristics: timing, magnitude, and direction. For each actual ramp, it searches through predicted ramps and considers them a match if they meet all three criteria: (1) the time difference between the centers of the ramps (calculated as the midpoint between start and end times) is within a specified time threshold (1-hour here), (2) the difference in percentage change between the ramps falls within a defined percentage threshold (15% here), and (3) both ramps share the same direction (increasing or decreasing). Once a match is found, that predicted ramp is marked as processed to prevent double-matching, and the pair is recorded as a true positive. Any actual ramps without matches are counted as false negatives, while unmatched predicted ramps are counted as false positives.

The contingency table for different ramp magnitudes and durations is shown in Figure 8. This is adapted from beforecast-verification package developed in T5.1.

The evaluation metrics of the contingency table -

Hits: Correctly predicted ramp events. These are cases where the model successfully detects a ramp-up event within the threshold and time constraints.

Misses: Ramp events that were missed by the model, indicating that the model failed to identify an actual event.

False Alarms: Instances where the model predicted a ramp-up event, but no such event occurred. This reflects the model's tendency to incorrectly signal a ramp.

Correct Negatives: Cases where the model correctly identifies the absence of a ramp, representing the accuracy in predicting no event.

The contingency plot compares four models - LSTM, CNN-LSTM, CNN-LSTM-Attn, and TFT - across different thresholds of magnitude (25%, 50%, 75%, 80%) and time windows (≤1 hour and ≤3 hours). The plot represents hits, misses, false alarms, and correct negatives. The TFT (Temporal Fusion Transformer) model shows slightly better consistency across different thresholds, particularly in reducing false alarms while maintaining good hit rates. The CNN-LSTM with attention mechanism demonstrates comparable performance to the base CNN-LSTM, suggesting that the addition of attention may not significantly improve ramp detection in this case. The basic LSTM model, while competent, shows slightly more variability in its performance metrics.

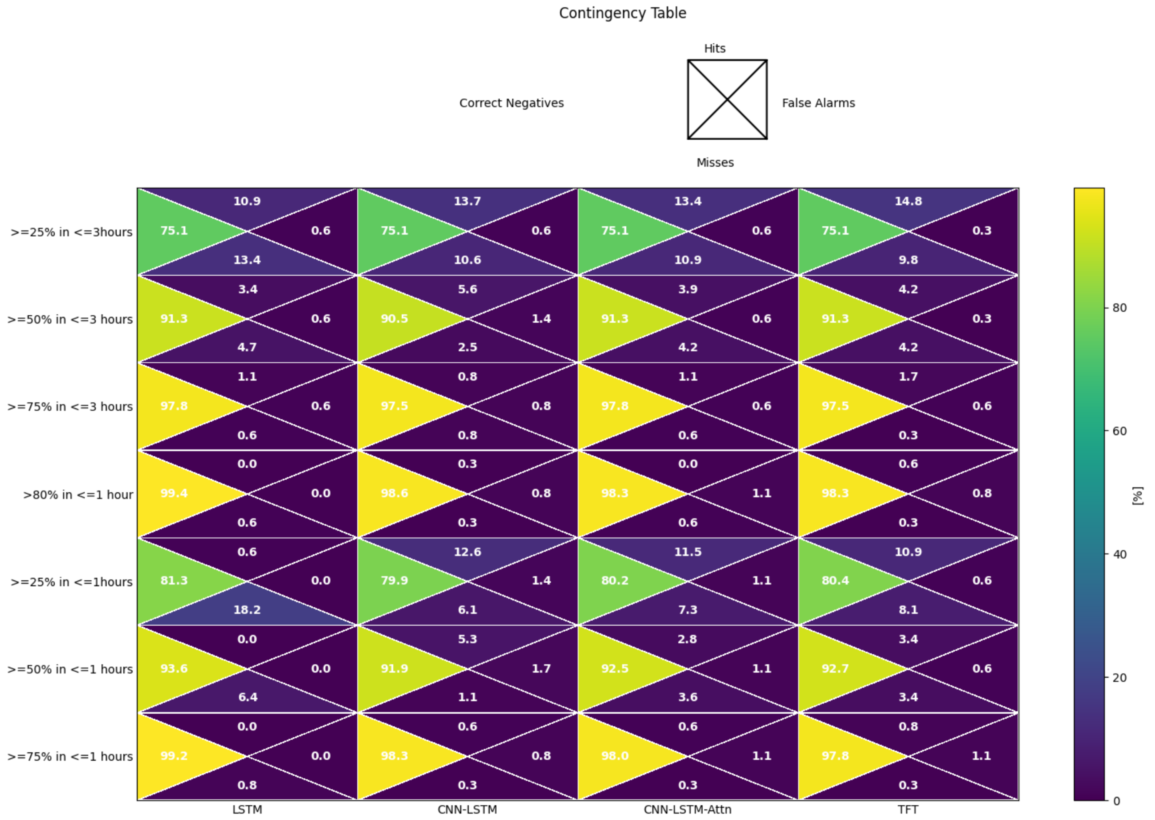

Figure 8: Contingency table for different ramp condition and power prediction models.

5 Conclusion

A detailed evaluation of various models for wind power forecasting was conducted, combining Numerical Weather Prediction (NWP) data and SCADA measurements. Among the tested models, which included traditional methods like ARIMA and advanced deep learning architectures, the Temporal Fusion Transformer (TFT) stood out, achieving the lowest RMSE of 564.72 kW and the highest R^2 score of 0.85. The TFT excelled due to its ability to process multiple input variables effectively and deliver consistent performance across different prediction horizons, maintaining a stable normalized MAE of 5–6%, even for predictions beyond 10 hours.

An analysis of the feature importance of the TFT model revealed that SCADA wind speed data accounted for 35% of the importance in historical inputs, while NWP wind speed forecasts at 100m height contributed 45% to future predictions. The model also demonstrated strong ramp detection capabilities across various magnitude thresholds and time windows, outperforming other architectures in reducing false alarms while maintaining high hit rates.

These findings highlight the significant advantages of modern deep learning approaches, especially transformer-based architectures, over traditional statistical methods for accurate wind power forecasting and reliable ramp detection.

In the next steps, more validation on different farms needs to be carried out with different NWP model data.

6 References

Bessa, R.J., Miranda, V., Botterud, A., Wang, J., & Konstantelos, I. (2017). Time adaptive conditional kernel density estimation for wind power forecasting. IEEE Transactions on Sustainable Energy, 8(3), 1220-1229. doi:10.1109/TSTE.2016.2646142

Bossavy, A., Girard, R., & Kariniotakis, G. (2012). Forecasting ramps of wind power production with numerical weather prediction ensembles. Wind Energy, 16(1), 51-63. doi:10.1002/we.526

Bouche, D., et al. (2023). Wind power predictions from nowcasts to 4-hour forecasts: A learning approach with variable selection. Renewable Energy, 211, 938-947.

Chen, J., Wan, C., & Xu, Z. (2021). Wind power forecasting using transformer neural networks: Model comparison and case study. Renewable Energy, 174, 734-749. doi:10.1016/j.renene.2021.04.098

Cheneka, B.R., Watson, S.J., & Basu, S. (2020). A simple methodology to detect and quantify wind power ramps. Wind Energy Science, 5(4), 1731-1741.

Costa, A., Crespo, A., Navarro, J., Lizcano, G., Madsen, H., & Feitosa, E. (2008). A review on the young history of the wind power short-term prediction. Renewable and Sustainable Energy Reviews, 12(6), 1725-1744. doi:10.1016/j.rser.2007.01.015

Doekemeijer, B., Simley, E., & Fleming, P. (2022). Comparison of the Gaussian Wind Farm Model with Historical Data of Three Offshore Wind Farms. Energies, 15, 1964. doi:10.3390/en15061964

Giebel, G., Brownsword, R., Kariniotakis, G., Denhard, M., & Draxl, C. (2011). The state-of-the-art in short-term prediction of wind power: A literature overview. Wind Energy, 14(8), 829-846. doi:10.1002/we.458

Hannesdóttir, Á., & Kelly, M. (2019). Detection and characterization of extreme wind speed ramps. Wind Energy Science, 4(3), 385-396.

Li, G., Shi, J., & Qu, X. (2010). Probabilistic forecasting of wind power generation using extreme learning machine. Energy Conversion and Management, 50(4), 1090-1099. doi:10.1016/j.enconman.2010.02.023

Lim, B., Arık, S.Ö., Loeff, N., & Pfister, T. (2021). Temporal Fusion Transformers for interpretable multi-horizon time series forecasting. International Journal of Forecasting, 37(4), 1748-1764. doi:10.1016/j.ijforecast.2021.03.012

Lin, L., Yang, C., & Zhang, Y. (2022). Deep learning-based wind power forecasting: A transformer approach. Frontiers in Energy Research.

Lin, T., Guo, T., & Aberer, K. (2017). Hybrid Neural Networks for Learning the Trend in Time Series. Proceedings of IJCAI, 2273-2279.

Liu, H., Tian, H.Q., Pan, D., & Li, Y.F. (2015). Forecasting models for wind speed using wavelet, wavelet packet, time series and artificial neural networks. Applied Energy, 156, 203-219. doi:10.1016/j.apenergy.2015.07.023

Liu, Y., Wang, D., Zhang, X., & Ding, H. (2020). A hybrid model of convolutional neural networks and long short-term memory network for wind power forecasting. Energy, 201, 117598. doi:10.1016/j.energy.2020.117598

Messner, J.W., Pinson, P., Browell, J., Bjerregård, M.B., & Schicker, I. (2020). Evaluation of wind power forecasts—An up-to-date view. Wind Energy, 23, 1461-1481. doi:10.1002/we.2497

Mohandes, M., Rehman, S., & Halawani, T.O. (2004). A neural networks approach for wind speed prediction. Renewable Energy, 13(3), 345-354. doi:10.1016/j.renene.2004.01.011

Morshedizadeh, M., Kordestani, M., Carriveau, R., Ting, D.S.K., & Saif, M. (2018). Power production prediction of wind turbines using a fusion of MLP and ANFIS networks. IET Renewable Power Generation, 12(9), 1025-1033.

Wang, J., Zhu, J., Zhang, B., & Li, Z. (2014). A new hybrid wind speed forecasting method based on self-organizing maps, wavelet analysis, and ARIMA model. Energy Conversion and Management, 84, 109-119. doi:10.1016/j.enconman.2014.04.061

Zhang, Y., Wang, J., Wang, X., & Li, J. (2016). Short-term wind power prediction based on LSTM and CNN. Neurocomputing, 277, 317-325. doi:10.1016/j.neucom.2017.02.042

Zhao, Z., Yan, J., & Cheng, P. (2023). Ultra-Short-Term Wind Power Forecasting Based on CGAN-CNN-LSTM Model Supported by Lidar. Sensors, 23(9), 4369. doi:10.3390/s23094369

Zhou, Y., Wang, J., Zhang, G., & Ma, J. (2021). Short-term wind power prediction based on attention mechanism and encoder-decoder structure. Energy, 231, 120946. doi:10.1016/j.energy.2021.120946

Zhou, Y., Wang, J., Zhang, G., & Ma, J. (2023). Hierarchical Spatial-Temporal Transformer Network for long-term wind power forecasting. arXiv Preprint, 2305.18724.