

DEVELOPMENT OF A MACHINE-LEARNING-BASED UP-SAMPLE ALGORITHM FOR WIND TURBINE TIME SERIES

*Nergiz Küçükğacal, *Christian Fröhlich

FGH e.V., Roermonder Str. 199, 52067 Aachen, +49 151 4201 86 49,
christian.froehlich@fgh-ma.de, <https://www.fgh-ma.de/de/>

Abstract: In the context of the global energy transition, wind energy has emerged as a key renewable energy source. However, due to the insufficient availability of high-resolution time-series datasets, producing reliable results in grid optimization simulations becomes increasingly difficult. Most publicly available datasets are limited to low temporal resolutions, such as hourly measurements, while acquiring high-resolution data requires costly sensors and extensive storage infrastructure.

This study addresses this gap by proposing a novel machine learning-based up-sampling approach that converts low-resolution wind turbine data into minute-level time series. The proposed Frequency-Based LSTM (LSTM-FB) model employs the Fast Fourier Transform (FFT) to learn complex signal dynamics in the frequency domain. By decomposing the wind power signal into a deterministic trend and stochastic fluctuations, the model estimates spectral coefficients to reconstruct short-term variability.

Model performance is evaluated using standard error metrics as well as the Volatility Similarity Score (VSS), which specifically measures the accuracy of short-term fluctuation reconstruction. Experimental results indicate that the model is promising in capturing short-term fluctuations in real-world data; in this context, a VSS of 0.0162 and an R^2 of 0.88 were obtained.

Keywords: Fast Fourier Transform, Long Short-Term Memory, Up-Sampling, Wind Turbine, Renewable Energy Resources

1 Introduction

The global demand for renewable and clean energy sources is steadily increasing in line with worldwide efforts to reduce greenhouse gas emissions, decrease the dependence on fossil fuels, and achieve carbon neutrality. Among various energy sources, wind energy stands out as an environmentally sustainable alternative. Although wind turbines provide a significant advantage by converting wind energy into electrical energy, they also introduce several challenges. Wind speed is highly uncertain and exhibits variable fluctuations. This directly affects generation capacity and complicates the efficient integration of wind power into the grid, by introducing power and voltage fluctuations. As a result, new difficulties arise in the design and reliable operation of modern energy systems, and one of the most important challenges is the need for high resolution temporal data [1].

High resolution generation data, for example data recorded at minute intervals, is essential for simulations focusing on power system stability, frequency dynamics, and power quality analysis [2], [3]. Such detailed information allows accurate modeling of system dynamics that occur within short time scales. However, current data sources often fail to meet this

requirement. Widely used meteorological datasets (e.g., ERA5) and even some SCADA systems usually provide low temporal resolution, often in the form of hourly measurements [4], [5].

This limitation leads to significant inaccuracies in energy system modeling. Simulations that rely on low resolution data may underestimate critical events such as zero power situations or overestimate cumulative generation [6], [7]. Consequently, the accuracy and reliability of grid simulations become severely restricted, critical for high-fidelity energy simulations, such as grid stability and power quality assessments. On the other hand, obtaining high resolution data is often impractical for most research and operational environments because it requires costly sensors and advanced data infrastructure.

While literature extensively covers machine learning for wind speed forecasting, there is a distinct lack of research on artificially increasing the temporal resolution of wind turbine time series. Traditional techniques like interpolation fail to capture the high-frequency fluctuations inherent in these signals.

This study addresses this gap by presenting a frequency-based LSTM algorithm that up-samples hourly values to minute-level resolution. By combining the Long Short-Term Memory (LSTM) network with the Fast Fourier Transform (FFT), the model learns complex fluctuation components in the frequency domain, demonstrating superior performance over traditional time-domain approaches.

2 Related Works

Capturing high-frequency dynamics in non-stationary wind energy time series is quite difficult, and traditional interpolation methods are insufficient for capturing these dynamics [8]. Consequently, machine learning models, specifically Long Short-Term Memory (LSTM) networks and Generative Adversarial Networks (GANs), have emerged as robust solutions for up-sampling. LSTMs are particularly effective due to their ability to model long-term temporal dependencies and reconstruct missing information with high precision [9], [10]. For instance, Quentin et al. demonstrated that LSTMs outperform interpolation in increasing the resolution of ERA5 wind speed data [8], while Liu et al. integrated LSTMs into a super-resolution framework for enhanced wind power forecasting [11].

Generative models have also shown promise in synthetic data generation. Zhang et al. applied a GAN-based approach for one-minute up-sampling, achieving an Output Variance Consistency (OVC) score exceeding 90% compared to target distributions [8]. Beyond pure machine learning, hybrid approaches like the one proposed by Olauson et al. utilize scale separation. By using cubic spline interpolation for low-frequency trends and stochastic modeling for high-frequency fluctuations, they ensured physical consistency and excellent agreement in Power Spectral Density (PSD) characteristics [12].

Despite these advancements, a gap exists in integrating the sequential modeling strength of LSTMs with the scale separation principle using FFT. Current LSTM applications are either prediction-oriented or operate exclusively in the time domain, often struggling with high-frequency volatility. This study addresses this gap by integrating the frequency-based separation principle with the sequential learning capability of LSTMs, ensuring accurate

reconstruction of both the underlying trend and the essential high-frequency dynamics required for power system simulations.

3 Methodology

This section outlines the framework of the proposed up-sampling approach, which aims to reconstruct high-resolution wind power signals from low-resolution data. First, the core components of the methodology and the theoretical foundations of the employed models are established. Subsequently, the performance metrics used to evaluate the model's accuracy in capturing both overall trends and short-term volatility are introduced. Finally, the step-by-step workflow of the frequency-based LSTM algorithm is described in detail, encompassing signal decomposition, feature learning, and signal reconstruction stages.

3.1 Fundamental Components of the Approach

Wind turbine data is collected through sensors and SCADA systems, and by its nature it exhibits a highly stochastic and non-stationary structure [13]. This data is generally stored with low temporal resolution, often in the form of hourly averages [14]. Low resolution data conceals short term system dynamics, such as sudden turbulence in wind and rapid fluctuations in power output. This situation creates the need for up-sampling [15]. Up-sampling refers to the process of increasing the sampling rate of a time series by inserting new data points between the existing points [16]. Although traditional interpolation techniques can be used for this task, these methods fail to capture the instantaneous and random high frequency fluctuation components of the signal [8], [17].

To address these limitations, LSTM networks are prominent due to their capacity to model long-term temporal dependencies in complex time series [9], [10]. The LSTM architecture, characterized by its gating mechanisms that regulate information flow and a cell state that preserves information over extended periods, provides an ideal framework for modeling the intricate and delayed relationships inherent in wind data [18], [19]. This capability enables the model to reconstruct not only the general trend but also the characteristic fluctuations of the wind power profile with high fidelity.

3.2 Evaluation Metrics

The performance of the developed up-sampling algorithm was evaluated using metrics that measure both: the ability of the model to capture the overall data trend and its success in reproducing high frequency fluctuations, also referred to as volatility.

3.2.1 Coefficient of Determination (R^2)

The Coefficient of Determination (R^2) is employed to measure the proportion of variance explained by the model [20]. In this context, R^2 assesses how accurately the generated series follows the overall trend and is formulated as:

$$R^2 = \frac{\sum(y_a - y_g)^2}{\sum(y_a - y_m)^2} \quad (1)$$

where y_a represents the actual high-resolution data points, y_g denotes the generated data, and y_m is the mean of the actual data. Values approaching “one” indicate that the model successfully reconstructs the general trend of the wind power signal [20].

3.2.2 Volatility Similarity Score (VSS)

Traditional metrics such as R-squared primarily focus on capturing the overall trend of the signal. However, for energy system simulations, it is critically important to accurately evaluate the instantaneous power fluctuations that represent the volatility of the system. For this reason, the most important performance indicator of the model is the Volatility Similarity Score, referred to as VSS.

The VSS measures the similarity between the volatility patterns of the predicted time series and the actual time series. A high VSS indicates that the model successfully reconstructs not only the trend but also the high frequency components required for grid dynamics. The computation of the VSS combines two components to evaluate volatility in terms of both: direction and magnitude.

- Correlation Component (C): This component measures the pearson correlation coefficient between the successive differences of the actual series (ΔA), and the successive differences of the predicted series (ΔP), representing the directional consistency of fluctuations. To isolate these dynamics, a successive difference is defined as the discrete change between two consecutive time steps [21]. In this study, the successive difference of the actual series at time t , ΔA_t , is formulated as:

$$\Delta A_t = A_t - A_{t-1} \quad (2)$$

where A_t and A_{t-1} represent the actual wind power values at the current and previous minutes, respectively. Similarly, the successive difference of the predicted series, ΔP_t is defined as:

$$\Delta P_t = P_t - P_{t-1} \quad (3)$$

where P_t and P_{t-1} represent the predicted wind power values at the current and previous minute. By calculating these differences for all t in the evaluation period, we obtain the vectors $\Delta A = [\Delta A_1, \Delta A_2, \dots, \Delta A_n]$ and $\Delta P = [\Delta P_1, \Delta P_2, \dots, \Delta P_n]$. These vectors specifically target the volatility of the wind power signal and the correlation component is computed as:

$$C = \rho(\Delta A - \Delta P) \quad (4)$$

where ρ denotes the pearson correlation coefficient.

- Variance Ratio Component (VR): This component measures the ratio between the variance of the differences in the actual series and the variance of the differences in the predicted series. By taking the minimum of the two possible ratios, the metric ensures that the predicted variance is neither excessively higher nor excessively lower than the actual variance [21].

$$VR = \min \left(\frac{\sigma^2(\Delta A)}{\sigma^2(\Delta P)}, \frac{\sigma^2(\Delta P)}{\sigma^2(\Delta A)} \right) \quad (5)$$

The final VSS value is computed as the product of these two components and is evaluated within the range from 0 to 1.

$$VSS = C \times VR \quad (6)$$

A high VSS serves as a fundamental indicator that the Frequency Based LSTM model successfully reconstructs the instantaneous power variations that remain hidden within low resolution data. This confirms that the model provides reliable outputs for power system simulations.

3.3 Frequency Based Long-Short Term Memory Algorithm

A frequency-based LSTM framework is proposed to reconstruct high-resolution minute-level wind power signals from low resolution hourly data. Unlike conventional interpolation methods, this approach captures the stochastic nature of wind dynamics by analyzing the signal in the frequency domain. As illustrated in Figure 1 and Figure 2, the overall workflow consists of three main stages: signal decomposition, learning of frequency characteristics, and signal reconstruction.

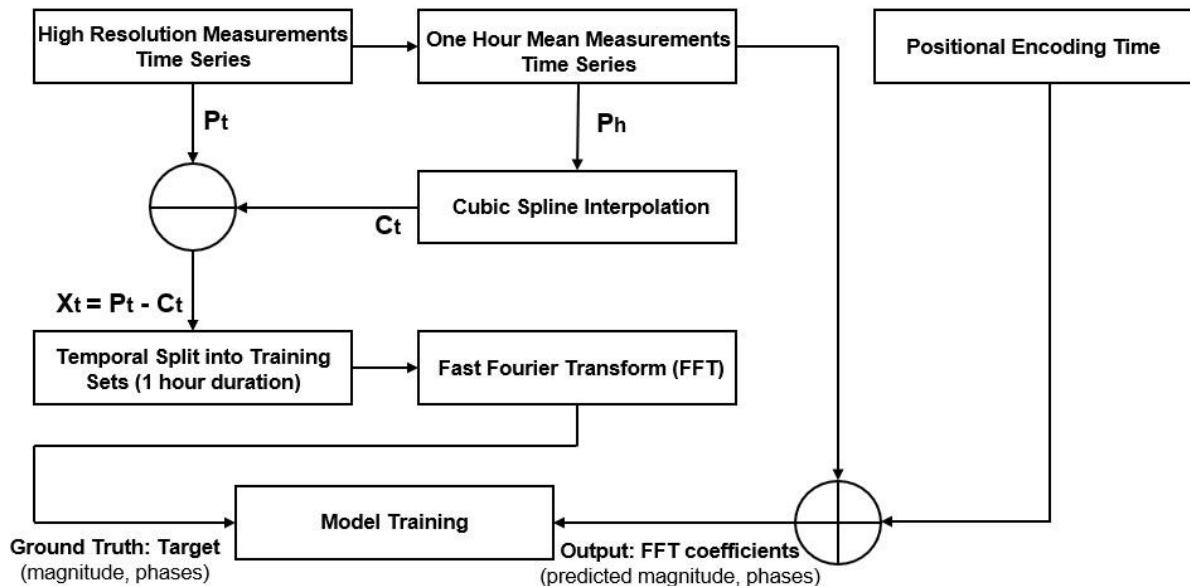


Figure 1 Training Step for LSTM-Frequency Based Model

In the training stage shown in Figure 1, the high-resolution active power signal in the dataset (P_t), is modeled as a combination of a deterministic trend and stochastic fluctuations. Hourly average power values (P_h) are obtained by computing 60-minute averages from the minute level data, for example between 01:00 and 01:59. Cubic spline interpolation is then applied to these hourly averages to construct a smooth baseline representing the general trend (C_t) of wind power. The fluctuation component (X_t) is derived by subtracting the trend component (C_t), from the original signal (P_t). This component contains high frequency oscillations that are critical for realistic up-sampling.

To enable the model to learn periodic patterns, the fluctuation signal (X_t) in the time domain is transformed into the frequency domain using the FFT. The resulting complex Fourier coefficients are decomposed into magnitude and phase components. These components

serve as the ground truth targets for the machine learning model. Focusing on frequency coefficients rather than raw time series data reduces the prediction burden of the model.

At the core of Figure 1 lies a LSTM network designed to capture temporal dependencies in wind dynamics. The model takes hourly features, such as active power, wind speed, and temperature, as inputs, together with positional encodings based on sine and cosine functions that provide cyclical time information.

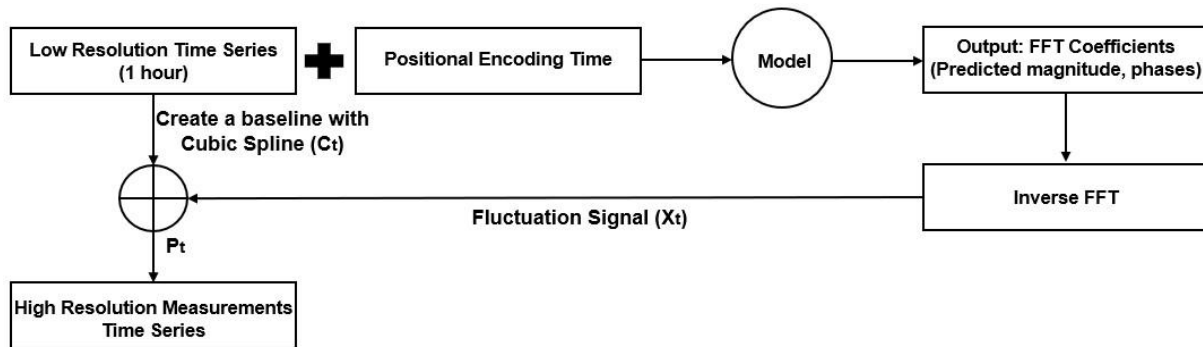


Figure 2 Application for LSTM-Frequency Based Model

Figure 2 describes the testing stage, in which the high-resolution signal is reconstructed through the following steps. The trained LSTM model predicts the magnitude and phase coefficients corresponding to the given hourly inputs. These predicted coefficients are converted back into complex numbers, and the Inverse Fast Fourier Transform (Inverse FFT) is applied to obtain the reconstructed fluctuation signal in the time domain, denoted as (X_t). The final high resolution power series is obtained by adding the reconstructed fluctuations to the cubic spline trend ($P_t = (C_t) + (X_t)$).

The sequential structure of the LSTM enables the model to leverage trends from previous hours when predicting the future characteristics of the fluctuation signal. For instance, if an increasing trend in wind speed is observed over the past few hours, the model can infer the frequency characteristics of the fluctuation signal for the subsequent hour. This provides a significant advantage over methods that process each hour independently.

4 Experimental Setup

4.1 Data Description

The up-sampling algorithm was trained and validated using the high resolution publicly available Smarteole dataset. The data were collected from a wind farm consisting of seven Senvion MM82 wind turbines each with a nominal capacity of 2.05 MW and a rotor diameter of 82 m. The dataset includes essential SCADA measurements such as active power wind speed blade pitch angle and generator speed. To enhance the predictive capability of the model, the raw features were enriched with rolling statistics and lagged variables [22].

4.2 Data Preprocessing

The experimental workflow was implemented in Python 3.12 using PyTorch for deep learning and Pandas, NumPy for numerical operations. The preprocessing pipeline consisted of following steps:

- Down-sampling: High resolution (1 minute) data were averaged into 1-hour intervals to create low-resolution input.
- Normalization: All input features were scaled to a range of [0,1] to ensure numerical stability.
- Data Splitting: To preserve temporal dependencies, the dataset was split chronologically into 70% for training and 30% for testing.

4.3 Model Training and Hyperparameters

The model was trained using the Adam optimizer with the mean absolute error (L1 Loss) function to mitigate the impact of outliers inherent in wind data. To prevent overfitting, an early stopping strategy was implemented with a patience of 50 epochs, ensuring the retention of weights from the best performing validation epoch.

To achieve the highest reconstruction accuracy, the model was trained using a comprehensive set of input features derived from the SCADA system. The primary variables include active power, wind speed, and temperature. To capture the temporal dynamics and local trends within the data, rolling mean statistics were incorporated as engineered features, calculated over 5, 10 and 30-minute windows.

Table 1 Hyperparameters of LSTM-Frequency Based

Hyperparameters	Values
Input sequence length	4
Number of layers	2
Dropout rate	0.2
Epochs	1000
Learning rate	0.0005
Batch size	64
Early stopping patience	50
Hidden dimension	256

A systematic optimization process was conducted to identify the most significant feature combinations and to fine tune the model's architecture. The results presented in this study correspond to the best performing configuration identified through this optimization. The optimal hyperparameters for the LSTM Frequency-Based model, yielded to the most superior performance, are detailed in Table 1.

5 Results

The quantitative performance results of the proposed LSTM frequency-based model are summarized in Table 2. The model's convergence behavior and a visual comparison of the high-resolution power signal are presented in Figure 3 and Figure 4, respectively.

Analysis of the loss curve in Figure 3 reveals that the training loss exhibits a steady decline, while the validation loss reaches a plateau after approximately the 100th epoch. To mitigate overfitting and preserve the model's generalization capability, an early stopping mechanism was employed to terminate training at the point of minimum validation loss. This strategy ensured that the model weights corresponding to the optimal validation performance were retained for final evaluation.

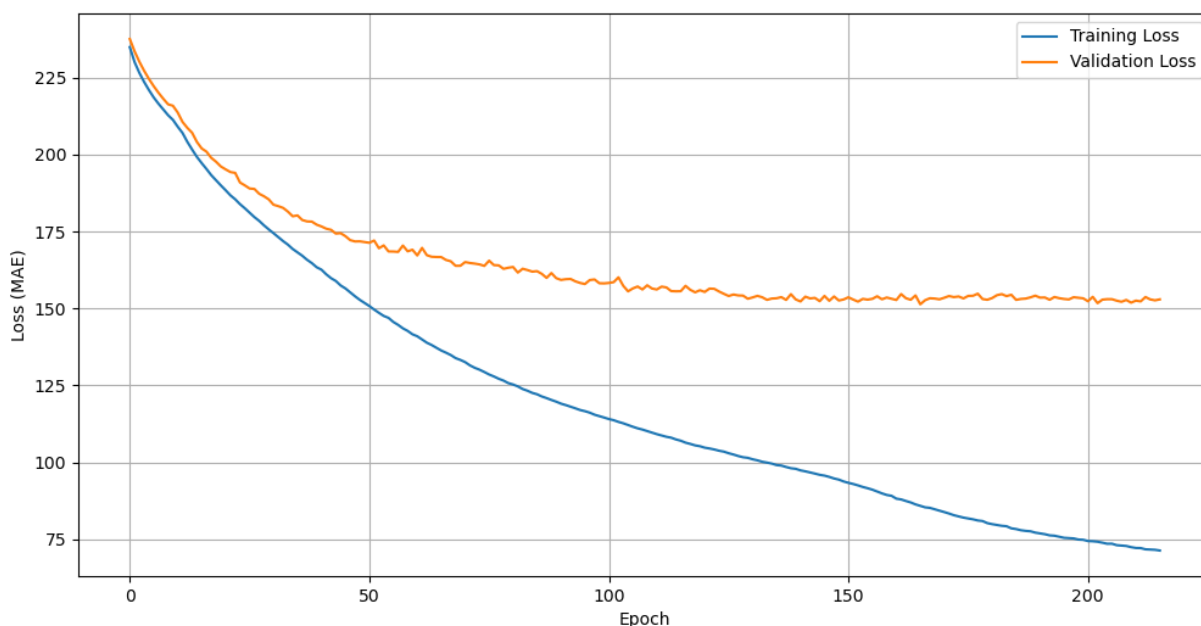


Figure 3 Loss Curve of LSTM-Frequency Based

The quantitative findings in Table 2 demonstrate that the frequency-based LSTM model establishes an effective balance between trend accuracy and volatility reconstruction. The model achieved a high Coefficient of Determination ($R^2=0.883$), proving its proficiency in following the overall generation trend. However, in wind power analysis, a high R^2 value alone can be misleading if the model fails to capture abrupt power fluctuations.

Table 2 Evaluation Metrics Results of LSTM-Frequency Based

Evaluation Metrics	Results
R^2	0.883
VSS	0.0162

To address this limitation, the Volatility Similarity Score (VSS) was adopted as the primary performance indicator, yielding a value of 0.0162. The visual assessment in Figure 4 supports these quantitative results: the model not only follows the hourly mean trend, indicated by the blue dashed lines, but also reproduces high frequency instantaneous fluctuations, shown by

the green signal, with high fidelity. Nevertheless, since VSS is ideally expected to be close to 1, the obtained value remains considerably low, and the sharp spikes appearing at the beginning of each hour in Figure 4 indicate a notable deficiency in the predicted series. Therefore, while the frequency-based LSTM approach is promising, it remains open to further improvements, particularly with respect to mitigating these periodic artifacts and increasing volatility similarity.

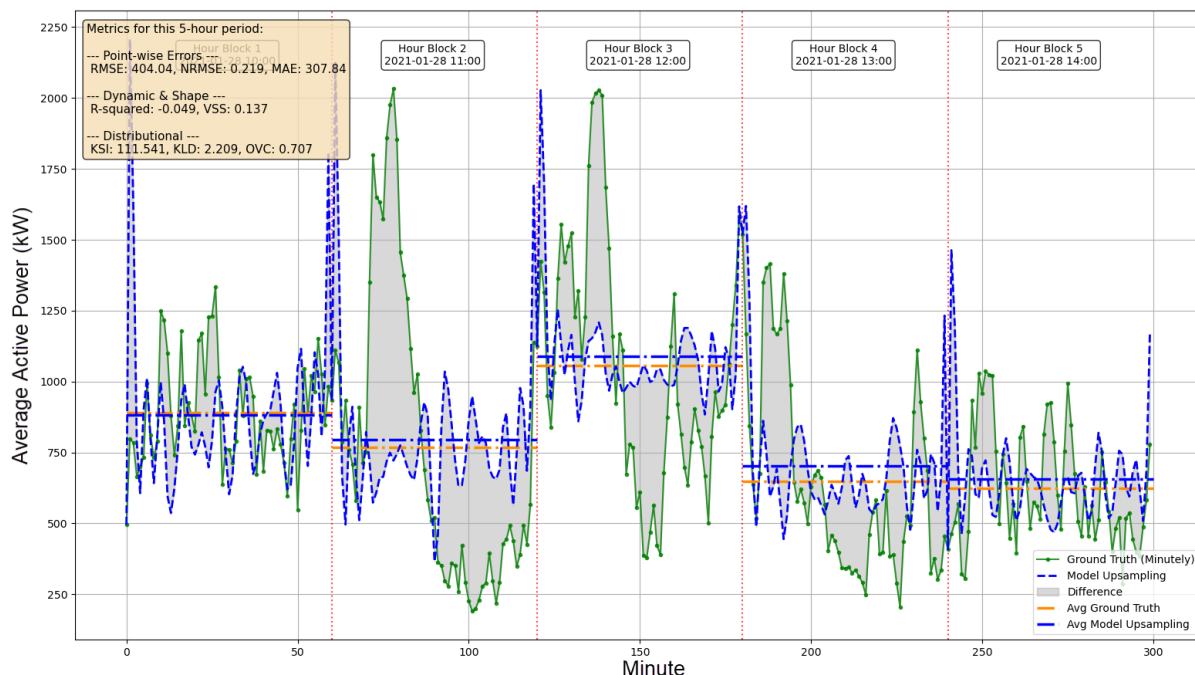


Figure 4 Actual and Predicted Active Power (LSTM-Frequency Based)

6 Discussion and Conclusion

This study addresses the critical scarcity of high-resolution wind power data by developing an innovative up sampling framework that integrates frequency domain signal processing with deep learning. The proposed model aims to reconstruct high frequency fluctuations by leveraging spectral components, and the results indicate that, when compared with real measurements, the model output can meaningfully follow both the hourly mean trend and instantaneous variations. Although the achieved VSS of 0.0162 and R^2 of 0.883 suggest that incorporating spectral information has the potential to represent the stochastic characteristics of wind dynamics more realistically, the current VSS value indicates that volatility similarity remains limited, given that VSS is ideally expected to be close to 1. Moreover, the sharp spikes observed at the beginning of each hour in the predictions emerge as a key aspect requiring improvement, particularly with respect to modeling abrupt and extreme variations. While the study offers a cost-effective alternative to expensive high frequency sensors, the limited availability of high-resolution data for model training constitutes an important factor affecting performance. Future work will focus on increasing volatility similarity while preserving the strengths of this frequency-based approach, with the objective of enhancing the synthesis of instantaneous power variations through the integration of advanced generative architectures

such as GANs or diffusion models and thereby improving the reliability of grid stability simulations.

7 Acknowledgments

Supported by:



on the basis of a decision
by the German Bundestag

The presented work was financially supported by the German Federal Ministry for Economic Affairs and Climate Action (BMWK) under the project number QUIRINUS-Control (FKZ 03EI4048A).

The dataset used in this research was provided by ENGIE Green within the scope of the French national project SMARTEOLE (grant no. ANR-14-CE05-0034). I am grateful to them for making this data publicly available [22].

8 References

- [1] Li, X., Li, K., Shen, S., & Tian, Y. (2023). Exploring time series models for wind speed forecasting: A comparative analysis. *Energies*, 16(23), 7785. <https://www.mdpi.com/1996-1073/16/23/7785>
- [2] Cong, Z., Yu, Y., Li, L., & Yan, J. (2022). Wind power time series simulation model based on typical daily output processes and Markov algorithm. *Global Energy Interconnection*, 5(1), 44–54. <https://doi.org/10.1016/j.gloi.2022.04.004>
- [3] Koivisto, M. (2021, October). Report on suggested improvements for time series simulation tools when working with DW (IEA Task 41 Deliverable 2.3; DTU Wind Energy Report No. TBA). Technical University of Denmark, DTU Wind Energy. <https://iea-wind.org/wp-content/uploads/2021/12/IEA-Task-41-Report-on-suggested-improvements-for-time-series-simulation-tools-when-working-with-DW.pdf>
- [4] Copernicus Climate Change Service (C3S). (n.d.). ERA5 hourly data on single levels from 1940 to present [Data set]. Climate Data Store (CDS). Retrieved January 20, 2026, from <https://cds.climate.copernicus.eu/datasets/reanalysis-era5-single-levels?tab=overview>
- [5] Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Abellan, X., Balsamo, G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita, M., ... Thépaut, J.-N. (2020). The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, 146(730), 1999–2049. <https://doi.org/10.1002/qj.3803>
- [6] Ward, K. R., Bamisile, O., Ejiyi, C. J., & Staffell, I. (2023). Time-averaged wind power data hides variability critical to renewables integration. *Energy Strategy Reviews*, 50, 101235. <https://doi.org/10.1016/j.esr.2023.101235>
- [7] Lehneis, R., Manske, D., & Thrän, D. (2023). Temporally and spatially resolved simulation of the wind power generation in Germany. *Energies*, 16(7), 3239. <https://doi.org/10.3390/en16073239>

- [8] Omoyele, O., Hoffmann, M., Koivisto, M., Larraneta, M., Weinand, J. M., Linen, J., & Stolten, D. (2024). Increasing the resolution of solar and wind time series for energy system modeling: A review. *Renewable and Sustainable Energy Reviews*, 189, 113792. <https://doi.org/10.1016/j.rser.2023.113792>
- [9] Song, W., Gao, C., Zhao, Y., & Zhao, Y. (2020). A time series data filling method based on LSTM—Taking the stem moisture as an example. *Sensors*, 20(18), 5045. <https://doi.org/10.3390/s20185045>
- [10] Tian, Y., Song, X., Hammond, G. E., Zachara, J. M., & Chen, X. (2022). Technical note: Using long short-term memory models to fill data gaps in hydrological monitoring networks. *Hydrology and Earth System Sciences*, 26(7), 1727–1743. <https://doi.org/10.5194/hess-26-1727-2022>
- [11] Liu, G., Zhang, S., Zhao, H., Liu, J., Liang, G., Zhao, J., & Sun, G. (2022). Super-resolution perception for wind power forecasting by enhancing historical data. *Frontiers in Energy Research*, 10, 959333. <https://doi.org/10.3389/fenrg.2022.959333>
- [12] Olauson, J., & Rydén, J. (2017). Simulating intra-hourly wind power fluctuations on a power system level. *Wind Energy*, 20(6), 973–985. <https://doi.org/10.1002/we.2074>
- [13] Zou, M., & Djokic, S. (2020). A review of approaches for the detection and treatment of outliers in processing wind turbine and wind farm measurements. *Energies*, 13(16), 4228. <https://doi.org/10.3390/en13164228>
- [14] Oliveira-Filho, A., Zemouri, R., Cambron, P., & Tahan, A. (2024). Wind turbine SCADA data imbalance: A review of its impact on health condition analyses and mitigation strategies. *Energies*, 18(1), 59. <https://doi.org/10.3390/en18010059>
- [15] Yin, Y. (2023). Prediction and analysis of time series data based on granular computing. *Frontiers in Computational Neuroscience*, 17, 1192876. <https://doi.org/10.3389/fncom.2023.1192876>
- [16] Stringfest Analytics. (2024, June 17). Python in Excel: How to upsample and interpolate time series data. *Python Bloggers*. <https://python-bloggers.com/2024/06/python-in-excel-how-to-upsample-and-interpolate-time-series-data/>
- [17] Wen, W., Liu, Y., Sun, R., Zhang, H., & Wang, J. (2022). Research on anomaly detection of wind farm SCADA wind speed data. *Energies*, 15(16), 5869. <https://doi.org/10.3390/en15165869>
- [18] Wikipedia contributors. (n.d.). Long short-term memory. In Wikipedia. Retrieved January 15, 2026, from https://en.wikipedia.org/wiki/Long_short-term_memory
- [19] Olah, C. (2015, August 27). Understanding LSTM networks. colah's blog. <https://colah.github.io/posts/2015-08-Understanding-LSTMs/>

[20] IrishStat. (2014). What is the problem with using R-squared in time series models? Cross Validated (Stack Exchange). Retrived January 11,2026, from <https://stats.stackexchange.com/questions/101546/what-is-the-problem-with-using-rsquared-in-time-series-models>

[21] Kudrat,D., Xie,Z., Sun, Y., Jia, T., Hu, Q. (2025). Patch-wise structural loss for time series forecasting (arXiv:2503.00877). arXiv. <https://doi.org/10.48550/arXiv.2503.00877>

[22] Simley, E., et al. (2021). The Smarteole wind farm control dataset. Wind Energy Science, 6(5), 1427-1440. <https://doi.org/10.5194/wes-6-1427-2021>