

Wind Power Prediction with ETSformer

Jill Baumann and Oliver Kramer

Computational Intelligence Lab
Department of Computer Science
Carl-von-Ossietzky University of Oldenburg
26111 Oldenburg, Germany
jill.baumann@uni-oldenburg.de

Abstract. With growing environmental awareness, power generation from wind and other renewable sources is becoming increasingly important. Accurate short-term predictions of wind turbine power are needed to keep the grid stable and secure. This paper investigates the use of ETSformer, a time series approach based on the transformer architecture, for wind power prediction. ETSformer incorporates exponential smoothing principles and introduces mechanisms such as exponential smoothing attention and frequency attention to improve accuracy, efficiency and interpretability. This study compares ETSformer and LSTM on a sample dataset of a wind farm and its surrounding sites within a three kilometer radius from the Wind Integration National Dataset Toolkit with five minute interval measurements. The investigation shows promising results and improvements of ETSformer in ultra-short and short-term wind power prediction.

1 Introduction

Wind, solar and other renewable energy sources are seen as important in the fight against climate change. According to the Global Wind Report 2022, installed wind power capacity has grown rapidly in recent years to 837 GW worldwide [1]. Due to the intermittency and variability of wind power, short-term wind power forecasting is necessary to keep the grid safe and stable when integrating wind power. Wind power prediction based on time series data plays an important role in developing a more accurate alternative to short-term weather simulations. Previous work has focused on machine learning (ML) methods such as support vector regression (SVR) or long short-term memory (LSTM). Transformer architectures, which belong to the deep learning (DL) methods, have attracted attention due to their promising performance in Natural Language Processing (NLP) and are already being used in other fields outside NLP.

This paper focuses on transformers, in particular the ETSformer by Woo et al. [2], which extends the standard transformer architecture to improve its performance for time series and long term forecasts. The aim of this paper is to evaluate the use of the ETSformer in the field of wind power forecasting. The paper is structured as follows Section 2 gives a brief overview of related work. The wind data set based on the experimental analysis is presented in Section 3. Section 4 presents the ETSformer architecture, which is applied to the wind prediction problem in Section 5. The paper ends with a conclusion in Section 6.

2 Related Work

Woon et al. [3] demonstrate LSTM's effectiveness in short-term wind energy forecasts using spatio-temporal information. Their approach shows significant improvements over state-of-the-art algorithms. Chen et al. [4] proposed a combined LSTM-CNN network to predict wind power for multiple turbines. Their method has separate LSTM modules for each turbine, with the output fed into a matrix. A convolutional neural network (CNN) then captures spatial features and generates final predictions. Peng et al. [5] proposed a three-step multi-integration algorithm combining DL and ensemble learning. They use variational mode decomposition (VMD) and wavelet transform (WT) to decompose original data and construct sub-models with stacked denoising autoencoder (SDAE), LSTM, and bidirectional long short-term memory (BLSTM). Support vector machine (SVM) and other networks are applied to sub-models for final predictions. Qu et al. [6] evaluated a standard transformer model for wind power prediction based on NLP success. Grigsby et al. [7] developed the Spacetimeformer, considering spatial and temporal information using a transformer architecture.

3 Wind Power Dataset

The Wind Integration National Dataset (WIND) Toolkit contains the calculated turbine output for more than 126,000 sites in the continental US from 2007 to 2013. The power dataset was created using hub wind data at a height of 100 meters and site-matched turbine power curves. The data was measured every five minutes, resulting in 631,086 available wind power measurements for each site [8].

Using the haversine distance, the neighboring wind turbines within a three kilometer radius of the target turbine with ID 17423 are determined. The eight neighboring wind turbines define together with the target turbine the Cheyenne wind farm, which were previously used in [3].

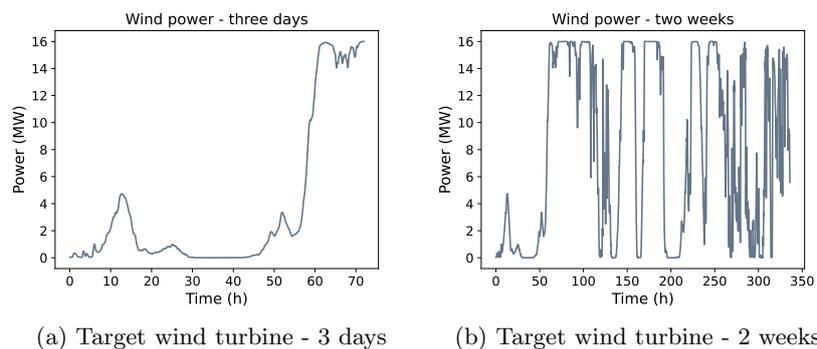


Fig. 1: Extracts of the time series of the target wind turbine

Figure 1 shows extracts of the target wind turbine time series for a period of three days and two weeks. It shows that the wind power time series contains noise as well as periods of more and less wind power, which occur without any apparent periodicity in this short extract.

Before feeding the dataset into the ETSformer, we conducted a thorough data analysis. There are no missing values in the dataset. For illustration the decomposition of the time series of the target wind turbine for May 2007 was performed. It shows no trend as there is no decreasing or increasing trend in the level of the time series. Due to recurring fluctuations it can be concluded that there is a seasonal value. The residuals are distributed around the expected value of 0. According to the Augmented Dickey-Fuller (ADF) test, the time series is stationary. The result of the ADF test seems surprising, since the data show some seasonality according to the time series decomposition. However, the critical values indicate that the null hypothesis can be rejected with a significance level of less than one percent, i.e. there is a low probability that the result is a statistical coincidence.

4 ETSformer

Woo et al. [2] proposed ETSformer, a transformer variant for time series prediction that takes advantage of time series decomposition and exponential smoothing to improve transformer models for time-series forecasting. For this purpose, two attention mechanisms exponential smoothing attention (ESA) and frequency attention (FA) are developed to replace the original self-attention mechanism. The result is a DL model that generates forecasts with level, growth and seasonal components that can be analyzed.

Figure 2 shows the overall encoder-decoder architecture of the ETSformer.

In each encoder layer the seasonal component is determined by FA and the growth component by multi-head ESA for more granular extrapolation. The seasonal component shows periodic fluctuations, while the growth and level component show the current level of the time series. The level is calculated in the level layer. In contrast to other approaches, residual learning is used within the ETSformer encoder, so each further encoder layer gets the residual of the previous encoder layer as input. Each decoder layer includes a growth and season (G+S) stack. Growth damping (GD) and FA in every G+S stack are processed to predict the season and growth component using its extracted components. The level component is predicted in the level stack. Finally, the extrapolated seasonal, growth and level components are combined in the decoder to produce a final forecast.

5 Experiments

For the input data, the wind power measurements of the target turbine and the eight neighboring turbines from 2007 to 2008 are concatenated. In this way, a multivariate data set of wind power is created by presenting the wind power

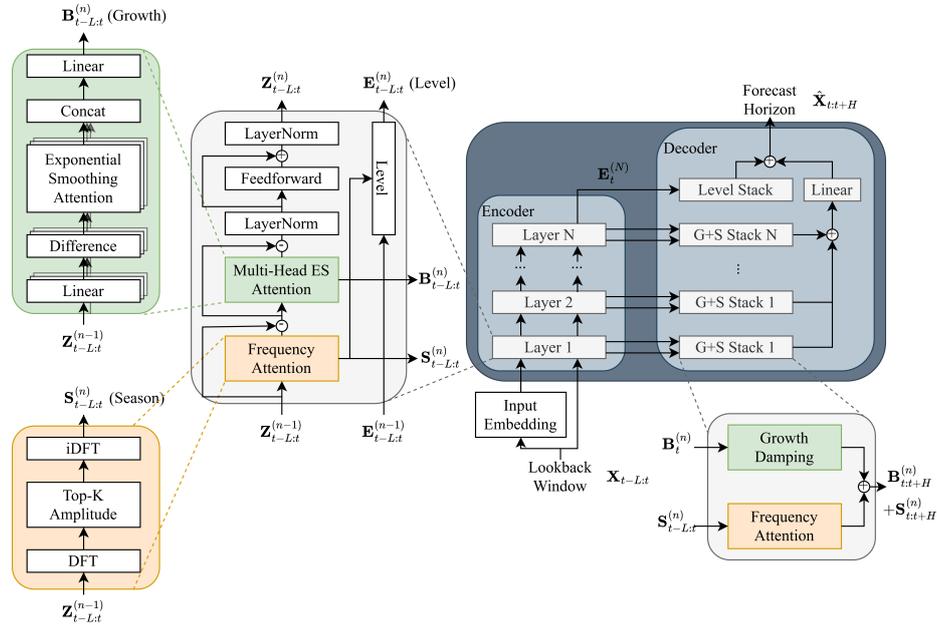


Fig. 2: ETSformer model architecture (oriented to [2])

data of a single wind turbine in chronological order in each column. The data is divided into training, test and validation data sets with a split of 70:20:10 and is pre-processed in the ETSformer by standardizing the features and dividing it into different batches. The time dependency is taken into account when dividing the data, i.e. the first 70 percent of the data set is used for training and the remaining part of the data set is used for validating and testing. The mean squared error (MSE) is used as the loss function and Adam as the optimization algorithm. Grid search was used to find optimal hyperparameters for the transformer architecture.

For the model, inspired by the original ETSformer paper, a small selection of hyperparameters were tested: sequence length $sl \in \{24, 36, 48, 60, 96, 192, 336, 720\}$; prediction length $pl \in \{24, 36, 48, 60, 96, 192, 336, 720\}$; number of frequencies of the fourier bases within the FA $K \in \{0, 1, 2, 3\}$ and learning rate $lr \in \{0.001, 0.00001\}$. The learning rate was chosen on the basis of a few experiments. Further parameter choices are number of encoders and decoder layer $N = 2$ and number of attention heads $h = 8$, while other adjustable parameters were left at their original values.

Table 1 shows the comparison between ETSformer and LSTM with different sequence and prediction lengths in terms of mean absolute error (MAE) and MSE. For the ETSformer the parameters that led to the best results were listed. The best MSE results are shown in *italics*, the best MAE results in **bold**.

ETSformer								
sl	192	192	720	720	336	720	720	720
pl	24	36	48	60	96	192	336	720
K	0	0	1	0	1	0	0	2
lr	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001
MAE	0.253	0.314	0.406	0.413	0.511	0.660	0.725	0.768
MSE	0.166	0.237	0.321	0.358	0.480	<i>0.694</i>	<i>0.772</i>	<i>0.830</i>
LSTM								
MAE	0.264	0.314	0.386	0.432	0.525	0.698	0.784	0.940
MSE	<i>0.162</i>	<i>0.226</i>	<i>0.304</i>	<i>0.354</i>	<i>0.475</i>	0.715	0.829	1.098

Table 1: Comparison the between ETSformer and LSTM for different sequence and prediction lengths on the test wind dataset.

In the very short term (0-4 hours) ETSformer shows comparable performance to LSTM. However, for short-term predictions (from four hours to several days), ETSformer outperforms LSTM. Interestingly, we observed that the accuracy of the predictions increases as the prediction length decreases. It is worth noting that the most effective models are characterized by a sequence length that is significantly larger than their prediction length.

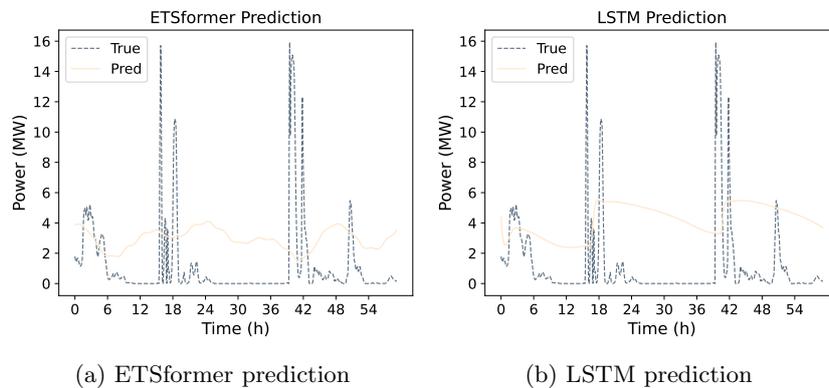


Fig. 3: Plot of actual wind data and (a) ETSformer versus (b) LSTM prediction for an exemplary interval

Figure 3 shows a plot of actual wind data and (a) ETSformer prediction versus (b) LSTM prediction. ETSformer assumes the curves of the original time series, but shows difficulties in predicting the extremes. LSTM uses less of the original time series curves and remains approximately on a par.

6 Conclusions

The results presented show significant improvements over previously used machine learning methods. The ETSformer has demonstrated its ability to predict ultra-short and short-term wind power data with high accuracy. These results are particularly noteworthy as they show that ETSformer not only matches the performance of LSTM for ultra-short-term predictions, but also outperforms the LSTM for short-term predictions.

Future work can build on these promising results by further refining and extending ETSformer and its hyperparameters to optimize its performance for wind power data. We plan to further improve the accuracy and reliability of wind power prediction models, ultimately contributing to the efficient integration of renewable energy sources into the power grid.

References

- [1] Joyce Lee and Feng Zhao. Global Wind Report 2022. Technical report, Global Wind Energy Council, 2022.
- [2] Gerald Woo, Chenghao Liu, Doyen Sahoo, Akshat Kumar, and Steven Hoi. ETSformer: Exponential smoothing transformers for time-series forecasting. Technical report, 2022.
- [3] Wei L. Woon, Stefan Oehmcke, and Oliver Kramer. Spatio-temporal wind power prediction using recurrent neural networks. In *Neural Information Processing*, Lecture Notes in Computer Science, pages 556–563. Springer International Publishing, 2017.
- [4] Xiaojiao Chen, Xiuqing Zhang, Mi Dong, Liansheng Huang, Yan Guo, and Shiyong He. Deep learning-based prediction of wind power for multi-turbines in a wind farm. *Frontiers in Energy Research*, 9: 723775, 2021.
- [5] Xiaosheng Peng, Cong Li, Shiyuan Jia, Liangsong Zhou, Bo Wang, and Jianfeng Che. A short-term wind power prediction method based on deep learning and multistage ensemble algorithm. *Wind Energy*, 25(9):1610–1625, 2022.
- [6] Kai Qu, Gangquan Si, Zihan Shan, XiangGuang Kong, and Xin Yang. Short-term forecasting for multiple wind farms based on transformer model. *Energy Reports*, 8:483–490, 2022.
- [7] Jake Grigsby, Zhe Wang, Nam Nguyen, and Yanjun Qi. Long-range transformers for dynamic spatiotemporal forecasting. *CoRR*, 2021.
- [8] National Renewable Energy Laboratory. Wind Integration National Dataset Toolkit, 2023. Accessed on 27.02.2023.