Periodicity Enhanced Long Short-term Memory Network for Wind Power Forecasting

Zhechun Liang[1] , **Kexuan Shi**[2] , **Wu Li**[3] and **Weiwei Wang**[4] *
School of Mathematics and Statistics,Xidian University,Xian,China
c1601631746@outlook.com, {20079100050,20079100081}@stu.xidian.edu.cn,
vivianwang919@126.com

Abstract

Global warming and energy crisis have brought sever challenges to the world. Wind power provides a significant renewable energy. Nevertheless, volatility of wind power causes structural risk to power grid and thus hiders its application. Short-term power forecasting(STPF) provides immediate information so that the power load can be promptly adjusted. In this paper, we propose a new neural network for STPF, wherein the periodic information in wind power series is exploited to improve the prediction accuracy. Specifically, to capture the intrinsic seasonal trend and periodic patterns of wind power series, we propose to preprocess the wind power series by using the variational mode decomposition(VMD). We also define a seasonal contribution factor (SCF) to determine the optimal number of decomposition. For STPF, we introduce the periodicity enhancing gate into the long short-term memory (LSTM) structure, so that the new network can fully exploit the periodic pattern in the subseries. Experiments on two datasets from different districts show the superiority of our proposed method in prediction accuracy.

1 Introduction

The renewable energy is an important measure to cope with the increasingly urgent energy and environmental crisis. Abundant and readily available, wind power has become an important source of renewable energy. Nevertheless, wind power has strong uncertainty, volatility and randomness, which increases the structural risk of the power grid, and even worsely, grid overload or power cut. Wind power forcasting provides an indispensable means for overcoming this.

According to time horizon,wind power forecasting can be divided into short-term power forecasting(minutes ahead)[Shi *et al.*, 2011; Li *et al.*, 2020] and long-term power forecasting(hours or days ahead)[Ahmadi *et al.*, 2020].Compared to the long-term forecasting, the short-term power forecasting(STPF) provides more immediate information so that the power load can be promptly adjusted. However, due to the unstable essence of the wind power time series, accurate STPF is still a challenging task.

Existing models can be roughly classified into three categories: the traditional models [Niu *et al.*, 2022; Hodge *et al.*, 2011; Eldali *et al.*, 2016] ,machine learning methods[Zendehboudi *et al.*, 2018; Zeng and Qiao, 2011; Kavousi-Fard *et al.*, 2014],and recently developed deep learning methods[Lin and 0044, 2021; Wang *et al.*, 2022b; Wang *et al.*, 2022a; Yu *et al.*, 2022]. The numerical weather prediction (NWP) model [Niu *et al.*, 2022] , and the auto regressive integrated moving average model (ARIMA) [Hodge *et al.*, 2011; Eldali *et al.*, 2016] are two typical representatives of the first category.In contrast, the machine learning methods, especially the support vector machine(SVM) [Zendehboudi *et al.*, 2018; Zeng and Qiao, 2011; Kavousi-Fard *et al.*, 2014] have shown better nonlinear curve fitting capability, and better performance in STPF. Recently, deep learning methods, such as the transformer-like neural networks[Lin and 0044, 2021; Wang *et al.*, 2022b]and variants of recurrent neural networks (RNN), arouse more interests in STPF research[Wang *et al.*, 2022a; Yu *et al.*, 2022].

Among them, the long short term memory (LSTM) structure has shown outstanding performance in processing series data and exploiting long-term data. For example, [Duan *et al.*, 2021] develops correntropy long short-term memory(CLSTM) neural network, which employs correntropy to replace traditional mean square error(MSE) as loss function..[Han *et al.*, 2019a] utilizes an operator vector acted on input to suppress LSTM's long-term memory of random components. [Fu *et al.*, 2018] compares gated recurrent unit with LSTM and claimed that LSTM is more suitable for multi-step forecasting.

To facilitate forecasting, data processing methods are usually used to reduce the instability and mine the seasonal trend of wind power data. Decomposition methods[Wang *et al.*, 2020; Niu *et al.*, 2022; Qiu *et al.*, 2017] and cluster methods[Wang *et al.*, 2018] are two major means utilized in STPF, among which variational mode decomposition(VMD) is the most widely used one.[Sun and Zhao, 2020; Han *et al.*, 2019b]employs VMD before forecasting, and the results prove that VMD improves the accuracy greatly.

Although the methods mentioned above have made achievements, we observe the following issues:

- **Lacking a criterion to evaluate the decomposition's**

effect. The VMD[Dragomiretskiy and Zosso, 2013] involves some hyper-parameters, which seriously affect the decomposition results and the forecasting accuracy. Existing methods[Wang *et al.*, 2020; Niu *et al.*, 2022; Qiu *et al.*, 2017] set the parameters manually. As a result, the empirical setting may bias the succeeding forecasting results. In addition, manually selection takes time.

- **Ignoring the seasonal trend and periodic pattern of preprocessed data.** Most Existing deep learning methods[Lin and 0044, 2021; Wang *et al.*, 2022b; Wang *et al.*, 2022a; Yu *et al.*, 2022] focus on designing novel and sophisticated network modules and structures, but ignore the intrinsic seasonal trend and periodic pattern of the wind power data.

In this work, we propose a two-phase STPF method, wherein we provide solutions to the problems mentioned above. In the first phase, we apply VMD to wind power series, hoping to exploit latent seasonal and periodic components. In order to evaluate the results of decomposition before making prediction, we define the Seasonal Contribution Factor (SCF). It reflects the strength of seasonal trend of each component and the proportion of each sub-series in the original sequence. Decomposition with high SCF contains more seasonal and periodic components, which is believed to be helpful for forecasting. In the second phase, the decomposition results of the first phase is fed into our proposed network, named Periodicity Enhanced Long Short term Memory network(PELSTM) to make prediction. To fully exploit the seasonal trend and periodic pattern captured by the VMD, we introduce a new gate module, named Periodicity Enhanced Gate, which encourages the periodic component to be preserved in the hidden state.

2 Related Work

2.1 Wind Power Series And STPF.

Let real number $\{X_t\}$ represent the numerical value of power generated from wind-driven generator at time t.

Wind power series $\{X_t\}$ is a typical time series. or, a discrete sequence taken at successive equally spaced time instants. Therefore, STPF can be considered as a time series forecasting problem. It aims to predict future values based on previously observed values.

In order to utilize the previously observed data dynamically and avoid accumulation of error, 'Sliding Window' is a common strategy for time series forecasting.The detailed process of 'Sliding Window' is shown in Figure 2. Let $X_{t-T}, X_{t-T+1}...X_t$ be previously observed data, $\hat{X}_{t+1}$ be the predicted value at instant t,f be a regression function, STPF can be formulated as follows:

$$X_{t-T}, X_{t-T+1}...X_t \xrightarrow{f} \hat{X}_{t+1} \tag{1}$$

Besides wind power series$\{X_t\}$, we employ $\{X_t^i\}$ to represent the descomposed sub-series.

2.2 Variational Mode Decomposition.

VMD [Dragomiretskiy and Zosso, 2013].is an adaptive and entirely non-recursive variational mode decomposition model, where the modes are extracted concurrently. It looks for an ensemble of modes denoted $\{X_t^k\}$ and their respective center frequencies, such that the modes collectively reproduce the 1D input signal $\{X_t\}$ while each being smooth after demodulation into baseband. VMD has found wide applications in signal decomposition in audio engineering, climate analysis, various flux and neuromuscular signal analysis in medicine and biology, etc.

For a given wind power series $\{X_t\}$, let $\{X_t^k\}^{k=1,...K}$ represent the decomposed sub-series, K be the number of intrinsic mode function(IMF), or the number of sub-series. VMD achieves the decomposition by solving the following constrained optimization problem:

$$\begin{aligned} &min \sum_k^K ||\partial_t[(\delta(t) + \frac{j}{\pi t}) * X_t^k)]e^{-jw_k t}||_2^2 \\ &s.t. \sum_k^K X_t^k = X_t \end{aligned} \tag{2}$$

where $\delta(t)$ denotes the Dirac distribution; $*$ denotes convolution;w_k denotes the central frequency of sub-series$\{X_t^k\}$ and j is the imaginary unit.

2.3 Auto-correlation Function.

As a function of delay time lag, the auto-correlation function (ACF) is the correlation of a series with a delayed copy of itself, measuring the dependency between present data and past data. ACF is an important analytical tool employed in time series analysis, helping discover periodic patterns and seasonal trend obscured by noise.For a given wind power series $\{X_t\}$, ACF, denoted by γ_k, is defined as

$$\gamma_k = \frac{1}{N-k} \sum_{t=k+1}^{N} (X_t - \bar{X})(X_{t-k} - \bar{X}) \tag{3}$$

where X_t denotes original series and X_{t-k} denotes the delayed series with delayed time lag of k. While $\bar{X}$ represents the expected value of the overlapping parts of $\{X_t\}$ and $\{X_{t-k}\}$.The normalized ACF, also called auto-correlation coefficient(ACC) is defined as

$$\rho(k) = \frac{\gamma_k}{\gamma_0} \tag{4}$$

2.4 Variance Contribution Rate.

The variance measures the variability in the data, and can be used to measure the amount of information carried by a sequence. In principal component analysis(PCA), variance contribution rate(VCR) is the rate of sub-series variance in the original sequence variance, and it measures the amount of information that a sub-series carry. It can be computed as follows:

$$\begin{aligned} r_i &= \frac{\lambda_i}{\sum_{j=1}^K \lambda_j} \\ \lambda_i &= \frac{1}{N} \sum_{t=1}^{N} (X_t^i - \bar{X}^i)^2 \end{aligned} \tag{5}$$

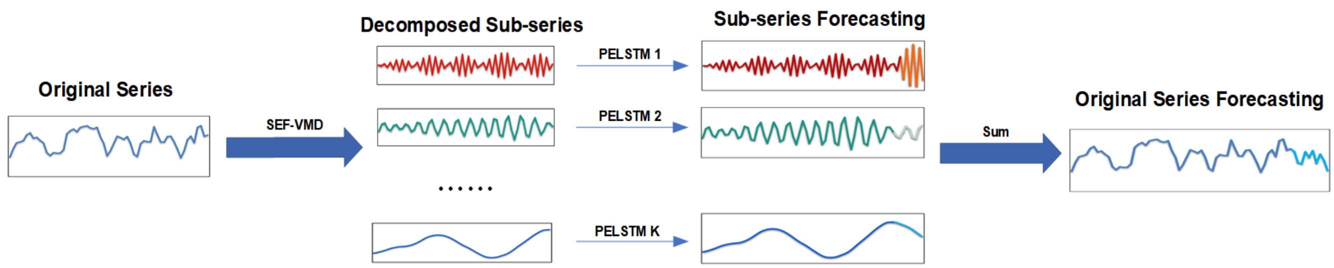

Figure 1: Illustration of STPF Framework

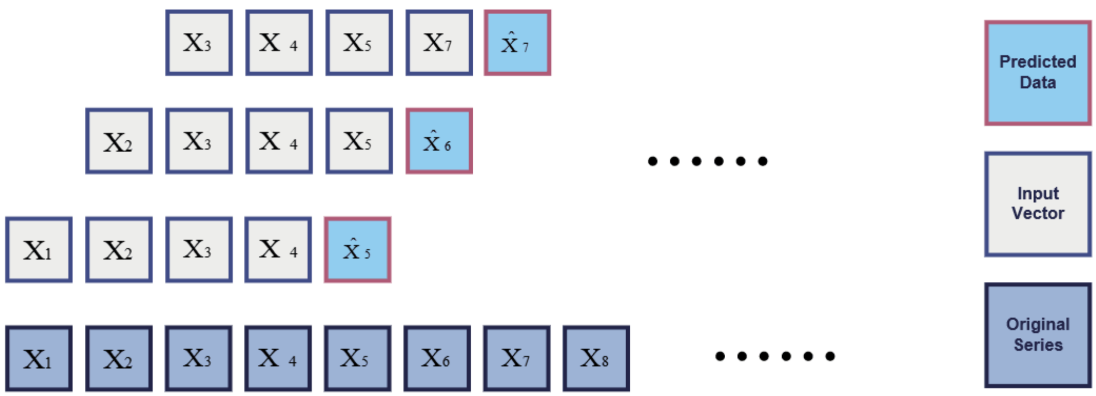

Figure 2: Illustration of Sliding Window

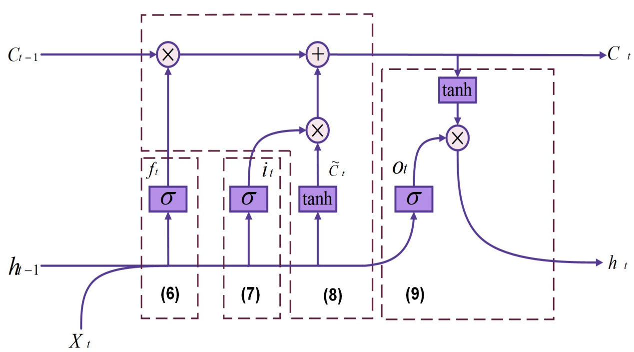

Figure 3: Illustration of LSTM Unit

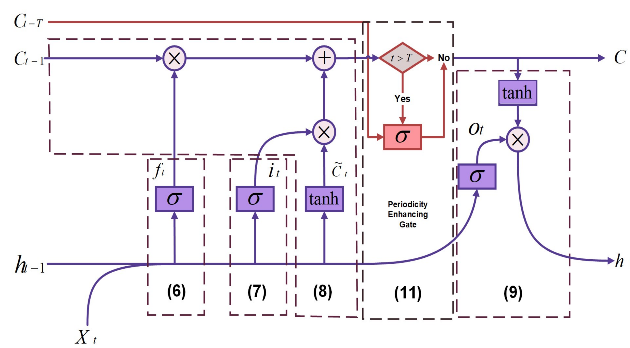

Figure 4: Illustration of PELSTM Unit

where r_i resembles VCR of sub-series $\{X_t^i\}$, λ_i denotes the variance of sub-series i, and $\bar{X}^i$ denotes the expected value of $\{X_t^i\}$.

2.5 Long Short-term Memory.

LSTM [Hochreiter and Schmidhuber, 1997] can make full use of long-term data and avoid vanishing or exploding gradients, and performs well in processing series data.

As illustrated in Figure 3, a LSTM network consists of an input layer, an output layer, and several cascaded recursive hidden layers. A recursive hidden layer is composed of several LSTM units. Each unit contains a memory cell and three gate functions that control the flow of information: forgetting gate, input gate and output gate. The gate functions are defined as follow:

The forgetting gate:

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \tag{6}$$

The input gate:

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \tag{7}$$

The memory cell state updates:

$$\begin{aligned} \widetilde{C_t} &= tanh(W_c \cdot [h_{t-1}, x_t] + b_c) \\ C_t &= f_t \circ C_{t-1} + i_t \circ \widetilde{C_t} \end{aligned} \tag{8}$$

The output gate:

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o)$$
$$h_t = o_t \circ tanh(C_t) \quad (9)$$

where σ is the sigmoid activator and h_{t-1} is the hidden state or rather, the output of time $t - 1$. W_f, W_i, W_c, W_o are the weight matrices and b_f, b_i, b_c, b_o bias variables of three units and memory cell. What's more, $\circ$ denotes the hadamard product. The forward propagation of LSTM unit is to execute the computation of gate functions in order, which is illustrated in Figure 2.

3 Methodology

In most existing work[Wang *et al.*, 2020; Qiu *et al.*, 2017], the hyperparameter K of VMD is set manually[Niu *et al.*, 2022]utilizes the envelope entropy as the criterion to measure the effect of VMD, but didn't discuss the improvements to forecasting accuracy. Meanwhile it ignored the seasonal trend and periodic pattern captured by the VMD. To solve this problem, we proposed SCF as a novel criterion to measure the effect of VMD, then select suitable hyperparameter K. SCF is aimed to maximize exploitable periodic components of decomposition, which is helpful for forecasting. We conducted experiments to prove that SCF improve the accuracy by 57.92% and 26.37% at most on two datasets compared to other hyperparameters.

Inspired by the seasonal trend and periodic pattern of data captured by VMD, we proposed PELSTM to exploit these information, which introduced 'Periodicity Enhanced Gate' structure to traditional LSTM. Experiments results show that PELSTM outperform LSTM on both datasets, especially on volatile series, which is close to application. Meanwhile, PELSTM avoids the abnormal fluctuations of LSTM forecasting outcome in flat series, which significantly improves the accuracy.

As shown in figure 1, our proposed framework consists of two major components: the wind power series decomposition module based on VMD and SCF, the forecasting module based on PELSTM. We employs SCF to evaluate the effect of decomposition, thus choosing the best number of IMF. Then, we deploy PELSTM on each IMF series, utilizing sliding windows to forecast the results.

3.1 Seasonal Contribution Factor.

We define the SCF to measure the effect of decomposition. Usually, if a decomposition contains more seasonal trend and periodic pattern, it is easier for forecasting. The first maximum value of ACC of a sub-series measures the seasonal trend of the sub-seriesis, thus we define it as seasonal factor(SF). The VCR measures the amount of information a sub-series contains. We define the SCF of a sub-series as the product of SF and VCR, and the sum of all sub-series' SCF as the SCF of a decomposition. The mathematical formulation of SCF is as follows:

$$SCF^i = SF^i \times VCR^i$$
$$SCF = \sum_{i}^{K} SCF^i \quad (10)$$

where SF^idenotes the seasonal factor of sub-series$\{X_t^i\}$ and SCF^idenotes the seasonal contribution factor of sub-series$\{X_t^i\}$.While SCF denotes the seasonal contribution factor of the decomposition.

We choose the best number of sub-series as the one corresponding to the largest SCF. In practice, we search time lag k from 0 to a finite number for SF.The finite number is set according to the computation time requirements.We definite SF as 0 if there does not exist a maximum value. What's more, the maximum point we can be viewed as a period of the sub-series, which we will discuss in the next section.

3.2 Periodicity Enhanced Long Short-term Memory.

To fully exploit the seasonal trend in data, we introduce a gate function named 'Periodicity Enhancing Gate' just following the update of memory cell state. Each time it propagates forward, except for the first period, it enhances the current cell state with the cell state one period ago, The structure of PELSTM is illustrated in Figure 3 and the operation of the periodicity enhancing gate is

$$C_t = \begin{cases} \sigma(W_{ps} * C_{t-T} + (1 - W_{ps} * C_t)) & t > T \\ C_t & t \leq T \end{cases} \quad (11)$$

where T is the period mentioned in section 3.1, W_{ps} is the weight matrix that can be updated in the training process.

4 Experiments

In this section, we conduct experiments to validate the effectiveness of SCF and PELSTM, respectively. We first introduce the datasets and data preprocessing.Then we conduct an experiment to verify the effectiveness of SCF. Finally, we conduct an experiment to compare PELSTM to other machine learning methods and LSTM variants.

4.1 Datasets

We conduct experiments on two datasets collected from one wind farm in China(Datasets 1) and one from Turkey(Datasets 2).The data are extracted from sensors in wind turbine, which generates real-time wind power data. In Datasets 1, samples were taken at 10 minute intervals, and in Datasets 2, samples were taken at 15 minute intervals.

Both datasets contain some missing values. In order to maintain the seasonal trend of data, we only exploit the the longest continuous segment of data. In addition, the data are normalized to [0,1].

After the data cleaning , Datasets 1 has 25234 data and Datasets 2 has 22712 data.

4.2 Setting

We split these two datasets into training set, validation set and test set with ratio 7 : 1 : 2.Let T^i denote the period of a sub-series. We set the length of 'Sliding Window' as T^i. If T^i is larger than 20 or T^i doesn't exist, we set it as 20. We set the batch size of data as 5.

The hidden size of LSTM and PELSTM is 48. The models are trained by the optimizer 'ADAM',We set the initial learning rate as 0.1 and choose exponential learning rate decay.

Datasets1		8 IMF	9 IMF	**10 IMF**	11 IMF	12 IMF
LSTM	RMSE	36.07	34.55	**29.64**	37.76	39.17
	MAE	23.24	23.05	**18.77**	26.20	24.30
	MAPE(%)	12.46	12.80	**10.19**	14.06	13.71
PSLSTM	RMSE	30.98	30.72	**23.88**	28.43	34.77
	MAE	19.84	19.14	**14.77**	18.87	21.22
	MAPE(%)	10.44	10.73	**8.54**	10.92	12.46

Table 1: Prediction Errors of Datasets 1

Datasets2		11 IMF	12 IMF	**13 IMF**	14 IMF	15 IMF
LSTM	RMSE	10.80	9.58	**8.52**	11.37	13.91
	MAE	8.75	7.85	**7.18**	8.71	10.81
	MAPE(%)	4.92	4.05	**3.97**	5.39	6.96
PSLSTM	RMSE	7.77	9.05	**7.25**	8.84	8.33
	MAE	6.46	7.60	**5.97**	6.79	6.54
	MAPE(%)	3.47	3.99	**3.43**	3.92	3.50

Table 2: Prediction Errors of Datasets 2

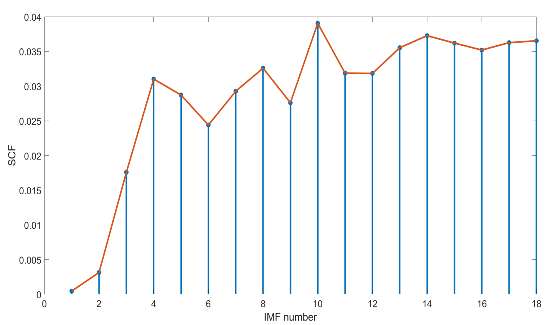

Figure 5: SCF of Datasets 1

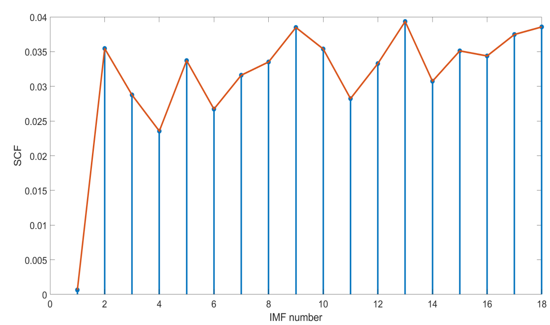

Figure 6: SCF of Datasets 2

4.3 Effectiveness of PELSTM and SCF

In this subsection we empirically validate the effectiveness of our PELSTM and SCF.

- **Advantages of PELSTM over LSTM.** To show advantages of PELSTM over LSTM, we make predictions by using PELSTM and LSTM, respectively, and the root mean squared errors(RMSE), mean absolute deviations(MAE), and mean absolute percentage errors(MAPE) of the predictions are given in Table 1 and Table 2, respectively for Dataset 1 and Dataset 2. Before prediction, we first apply VMD to obtain different numbers of IMFs. Specifically, for Dataset 1,$K \in \{8, 9, 10, 11, 12\}$ and for Dataset 2, $K \in \{11, 12, 13, 14, 15\}$. We can note that, on Dataset 1, K=10 leads to the best results for both PELSTM and LSTM, while on Dataset 2, K=13 leads to the best results for both PELSTM and LSTM. On the other hand, we can note that, on both datasets, our PELSTM outperforms LSTM consistently for different values of K.

  We also illustrate some predicted values by PELSTM and LSTM for Dataset 1 in Figure 7 and Figure 8. We can note that, the predicted values obtained by LSTM shows abnormal periodic fluctuation in flat part of some sub-series while the predicted values obtained by PELSTM are more consistent with the observed values. This is because PELSTM can timely capture the periodic pattern.

  Besides, our PELSTM shows better accuracy in predicting peaks. Figure 9 and Figure 10 shows the prediction results by LSTM and PELSTM for one subseries of the data from Dataset 1. We can see that PELSTM can better track local pattern in the data.

		ARIMA	SVM	BP	Transformer	LSTM	BiLSTM	GRU	ConvLSTM	CLSTM	**PELSTM**
Datasets1	RMSE	87.04	35.31	131.37	33.28	28.72	26.72	44.28	31.27	33.58	**23.17**
	MAE	50.95	21.12	100.16	19.45	18.40	18.35	38.75	23.52	19.43	**14.45**
	MAPE(%)	31.35	12.69	50.37	11.28	10.12	9.56	18.41	14.71	13.01	**8.51**
Datasets2	RMSE	15.82	129.56	12.12	20.34	8.52	8.02	23.68	10.22	12.52	**7.25**
	MAE	10.19	92.59	10.31	14.56	7.18	6.89	14.25	8.98	9.36	**5.97**
	MAPE(%)	6.04	49.67	5.73	9.59	3.97	3.52	9.59	5.62	6.01	**3.43**

Table 3: Prediction Errors by Baselines and Our Proposed Method

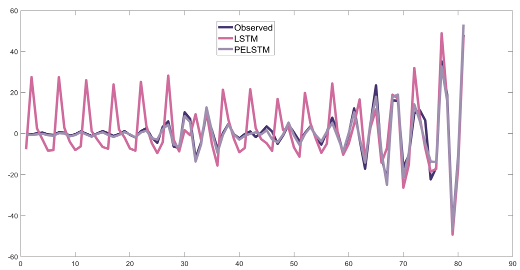

Figure 7: Abnormal Fluctuation Part of Sub-series from Dataset 1

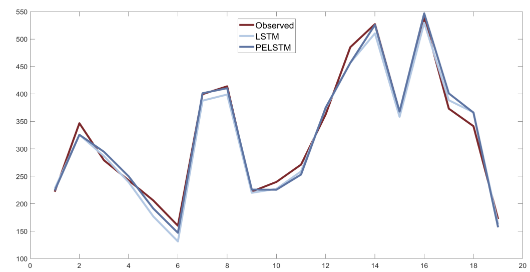

Figure 9: Peak Values of Original Series from Dataset 1

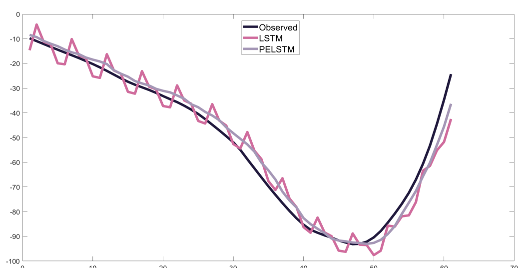

Figure 8: Abnormal Fluctuation Part of Sub-series from Dataset 1

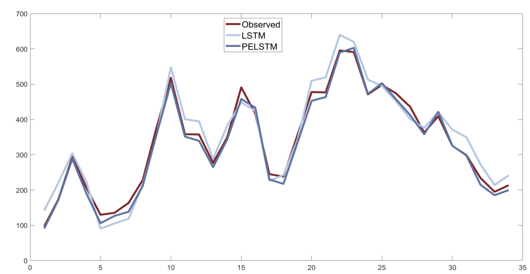

Figure 10: Peak Values of Original Series from Dataset 1

- **Effectiveness of SCF.** We apply VMD to Dataset 1 and Dataset 2, with $K \in \{1, 2...18\}$ and the corresponding SCFs are presented Figure 5 and Figure 6. For dataset1, the maximum SCF corresponds to the empirical best value of K=10(Table 1); and on for Dataset 2, the maximum SCF corresponds to the empirical best value of K=13(Table 2). Thus, we can conclude that our SCF is effective in inferring the best value of K.

4.4 Comparison of related methods with ours

In this subsection, we conduct experiments to compare our proposed method with related methods: ARIMA, SVM, BP, LSTM and Transformer. Besides, some important variants of LSTM applied in wind power forecasting like BiLSTM[Niu *et al.*, 2022], GRU[Fu *et al.*, 2018], ConvLSTM[Sun and Zhao, 2020] and CLSTM[Duan *et al.*, 2021] are also compared. All prediction methods work on the same decomposition results obtained by VMD, with the best K determined by the maximum SCF. The prediction errors of all methods are given in Table 3. We can note that, our PELSTM significantly outperforms the baselines for both datasets. For Datasets 1 in particular, which is rather volatile (with a large standard deviation), the prediction errors obtained by our method are much less than that obtained by others. In this sense, we believe that our PELSTM is more suitable for dealing with real data.

5 Conclusion

In order to solve STPF, this paper proposes a novel factor named 'Seasonal Contribution Factor' to select the number of IMF in the decomposition of power series.Then we propose a novel neural network architecture called 'Periodicity Enhanced Long Short-term Memory Network' to forecast the decomposed sub-series. Extensive experiments were conducted in two datasets from different districts, which proves the effectiveness of SCF and the better performance of PELSTM than baselines.

In the future, we plan to extend the model to enable more complex deep learning architectures such as attention mechanism. Furthermore, we will try to explore the accuracy of our proposed model in multi-step forecasting problem.

Acknowledgments

This research is supported by "the Fundamental Research Funds for the Central Universities"(5007-20108235343).

References

[Ahmadi *et al.*, 2020] Amirhossein Ahmadi, Mojtaba Nabipour, Behnam Mohammadi-Ivatloo, Ali Moradi Amani, Seungmin Rho, and Md. Jalil Piran. Long-term wind power forecasting using tree-based learning algorithms. *IEEE Access*, 8:151511–151522, 2020.

[Dragomiretskiy and Zosso, 2013] Konstantin Dragomiretskiy and Dominique Zosso. Variational mode decomposition. *IEEE transactions on signal processing*, 62(3):531–544, 2013.

[Duan *et al.*, 2021] Jiandong Duan, Peng Wang, Wentao Ma, Xuan Tian, Shuai Fang, Yulin Cheng, Ying Chang, and Haofan Liu. Short-term wind power forecasting using the hybrid model of improved variational mode decomposition and correntropy long short-term memory neural network. *Energy*, 214:118980, 2021.

[Eldali *et al.*, 2016] Fathalla A Eldali, Timothy M Hansen, Siddharth Suryanarayanan, and Edwin KP Chong. Employing arima models to improve wind power forecasts: A case study in ercot. In *2016 North American Power Symposium (NAPS)*, pages 1–6. IEEE, 2016.

[Fu *et al.*, 2018] Yiwei Fu, Wei Hu, Maolin Tang, Rui Yu, and Baisi Liu. Multi-step ahead wind power forecasting based on recurrent neural networks. In *2018 IEEE PES Asia-Pacific Power and Energy Engineering Conference (APPEEC)*, pages 217–222. IEEE, 2018.

[Han *et al.*, 2019a] Li Han, Huitian Jing, Rongchang Zhang, and Zhiyu Gao. Wind power forecast based on improved long short term memory network. *Energy*, 189:116300, 2019.

[Han *et al.*, 2019b] Li Han, Rongchang Zhang, Xuesong Wang, Achun Bao, and Huitian Jing. Multi-step wind power forecast based on vmd-lstm. *IET Renewable Power Generation*, 13(10):1690–1700, 2019.

[Hochreiter and Schmidhuber, 1997] Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural computation*, 9(8):1735–1780, 1997.

[Hodge *et al.*, 2011] Bri-Mathias Hodge, Austin Zeiler, Duncan Brooks, Gary Blau, Joseph Pekny, and Gintaras Reklatis. Improved wind power forecasting with arima models. In *Computer Aided Chemical Engineering*, volume 29, pages 1789–1793. Elsevier, 2011.

[Kavousi-Fard *et al.*, 2014] Abdollah Kavousi-Fard, Haidar Samet, and Fatemeh Marzbani. A new hybrid modified firefly algorithm and support vector regression model for accurate short term load forecasting. *Expert systems with applications*, 41(13):6047–6056, 2014.

[Li *et al.*, 2020] Ling-Ling Li, Xue Zhao, Ming-Lang Tseng, and Raymond R Tan. Short-term wind power forecasting based on support vector machine with improved dragonfly algorithm. *Journal of Cleaner Production*, 242:118447, 2020.

[Lin and 0044, 2021] Weixuan Lin and Di Wu 0044. Residential electric load forecasting via attentive transfer of graph neural networks. In *IJCAI*, pages 2716–2722, 2021.

[Niu *et al.*, 2022] Dongxiao Niu, Lijie Sun, Min Yu, and Keke Wang. Point and interval forecasting of ultra-short-term wind power based on a data-driven method and hybrid deep learning model. *Energy*, page 124384, 2022.

[Qiu *et al.*, 2017] Xueheng Qiu, Ye Ren, Ponnuthurai Nagaratnam Suganthan, and Gehan AJ Amaratunga. Empirical mode decomposition based ensemble deep learning for load demand time series forecasting. *Applied Soft Computing*, 54:246–255, 2017.

[Shi *et al.*, 2011] Jing Shi, Xiuli Qu, and Songtao Zeng. Short-term wind power generation forecasting: Direct versus indirect arima-based approaches. *International Journal of Green Energy*, 8(1):100–112, 2011.

[Sun and Zhao, 2020] Zexian Sun and Mingyu Zhao. Short-term wind power forecasting based on vmd decomposition, convlstm networks and error analysis. *IEEE Access*, 8:134422–134434, 2020.

[Wang *et al.*, 2018] Kejun Wang, Xiaoxia Qi, Hongda Liu, and Jiakang Song. Deep belief network based k-means cluster approach for short-term wind power forecasting. *Energy*, 165:840–852, 2018.

[Wang *et al.*, 2020] Yun Wang, Yifen Li, Runmin Zou, Aoife M Foley, Dlzar Al Kez, Dongran Song, Qinghua Hu, and Dipti Srinivasan. Sparse heteroscedastic multiple spline regression models for wind turbine power curve modeling. *IEEE Transactions on Sustainable Energy*, 12(1):191–201, 2020.

[Wang *et al.*, 2022a] Jianing Wang, Hongqiu Zhu, Yingjie Zhang, Fei Cheng, and Can Zhou. A novel prediction model for wind power based on improved long short-term memory neural network. *Energy*, page 126283, 2022.

[Wang *et al.*, 2022b] Lei Wang, Yigang He, Xiaoyan Liu, Lie Li, and Kaixuan Shao. M2tnet: Multi-modal multi-task transformer network for ultra-short-term wind power multi-step forecasting. *Energy Reports*, 8:7628–7642, 2022.

[Yu *et al.*, 2022] Guang Zheng Yu, Liu Lu, Bo Tang, Si Yuan Wang, Ru Si Chen, and CY Chung. Ultra-short-term wind power subsection forecasting method based on extreme weather. *IEEE Transactions on Power Systems*, 2022.

[Zendehboudi *et al.*, 2018] Alireza Zendehboudi, M Abdul Baseer, and R Saidur. Application of support vector machine models for forecasting solar and wind energy resources: A review. *Journal of cleaner production*, 199:272–285, 2018.

[Zeng and Qiao, 2011] Jianwu Zeng and Wei Qiao. Support vector machine-based short-term wind power forecasting. In *2011 IEEE/PES power systems conference and exposition*, pages 1–8. IEEE, 2011.