Spatial Dynamic Wind power forecasting using lightGBM and multi-variate LSTM with hierarchical coherence constraints(Team name: Dynamo)

Ai Hongfeng*
OPPO Research Center

Wu Wenqi*
OPPO Research Center

Zhang Chaodong*
OPPO Research Center

ABSTRACT

Renewable energy has drawn growing attention and wind power forecasting is becoming increasingly important to optimize the operations of wind farms. Numerous methods have been used to forecast wind turbine powers, including numerical weather prediction, simulation, statistical time series forecasting, machine learning, deep learning and hybrid methods. Spatial information of wind turbines has been utilized to enhance forecast accuracy. In this work, we developed a hybrid machine learning model composed of a lightGBM model for middle-term forecasts(24-48 hours) and a multi-variate LSTM model for short-term forecasts(0-24 hours). Spatial information of turbines was implicitly incorporated in the LSTM model by enforcing hierarchical coherence constraints and applying a self-attention output layer trying to catch the interactions among turbines. The multi-variate LSTM model was implemented in PaddlePaddle. Code has been uploaded to https://github.com/cdzhang/wind_power_forecast.

KEYWORDS

wind power forecasting, lightGBM, LSTM, support vector, hierarchical forecasting

ACM Reference Format:
Ai Hongfeng, Wu Wenqi, and Zhang Chaodong. 2022. Spatial Dynamic Wind power forecasting using lightGBM and multi-variate LSTM with hierarchical coherence constraints(Team name: Dynamo). In . ACM, New York, NY, USA, 7 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 INTRODUCTION

The consumption of renewable energy has become increasingly fast in recent years, including wind, solar and hydrogen energy. Among them, wind power generation is relatively simple to realize because of its wide availability and uncomplicated mechanical energy transformation. However, due to the instability of weather and wind speed, the power generated by wind fluctuates greatly, which in turn affects the operation of regional power grids [1, 3]. Therefore, an effective wind power prediction method is essential to find the most economical solution for grid operation. This can help power dispatch departments organize power generation optimization plans, and thereby improve reliability.

Wind power forecasting is usually regarded as a complex task due to the variability and stochastic characteristics of wind speed and internal turbine controls. The mainstream models for wind power forecasting include numerical weather prediction, statistical time series model like ARIMA[6] and Kalman Filtering[11]. Multiple machine learning methods have been used for wind power forecasting, like SVM [19], wavelet analysis[10] and fuzzy logic[7] methods. Deep learning methods have also been used for wind power forecasting recently[16]. Hybrid models are becoming increasingly popular as they can exploit the advantages of each model based on their own characteristics[17, 20], and thus achieve superior performance[15].

According to forecast horizons, wind power forecasting problems can be classified into ultra-short-term forecasting(minutes), short-term forecasting(hours), middle-term forecasting (days) and long-term forecasting(months). Different methods have advantages for specific forecasting horizons [13].

Spatial information of turbines has been exploited for wind power forecasting. KNN methods was used to find nearest neighbors of turbines and features of these neighbors were then feed into the encoder part of a GRU-based sequence-to-sequence model[9]. Attention-based graph neural network structure was proposed to learn the interactions among turbines and their complex physical properties[2].

Before applying model training and prediction, anomaly detection and data pre-processing are significant as they affect the performance of the whole forecasting task. Various anomaly detection methods have been proposed to enhance forecast performance or for wind turbine failure detection[14, 18].

This article proposes a hybrid methods based on lightGBM and multi-variate LSTM for wind power forecasting, with novel anomaly detection methods and implicit incorporation of spatial information. The main contributions are summarized as follows:

- Combination of lightGBM model and multi-variate LSTM model for middle-term and short-term forecasts respectively can take advantage of suitable model characteristics at specific forecasting horizons.
- Applied a method to find the wind speed - max power curve of turbines solely from historical data by solving an optimization problem. The curve is decided by a few support vectors and can effectively detect anomalies in the data.
- Spatial information of wind turbines was implicitly incorporated in our method through hierarchical coherence constraints and self-attention. To the best of our knowledge, this is the first time hierarchical coherence constraints are used to utilize spatial structures of turbines.

*All authors contributed equally. Sorted by alphabetical order

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM…$15.00
https://doi.org/10.1145/nnnnnnn.nnnnnnn

The dataset used is from KDD CUP 2022 [21].

2 SOLUTION OVERVIEW

Our method is based on the ensemble of lightGBM model and multivariate LSTM model. The pipeline is as follows:

(1) Anomaly detection and data preprocessing: A lightGBM model was trained to fill in missing and abnormal wind power values, and noise was added to alleviate overfitting. Temperature values were fixed as there are some unreal values which might be caused by broken thermometers. Curves of max turbine power to wind speed was trained for each turbine and power higher than this curve was marked as abnormal and adjusted to the respective max value on the curve. Wind direction respect to some zero point was adjusted based on some simple assumptions.
(2) Feature engineering: Features based on minutes of the day, wind direction, wind speed, temperature and historical wind power were calculated by window sliding statistics and extracted by `tsfresh`.
(3) Forecasting: multiple models of lightGBM and LSTM were trained using different hyper-parameters, and during the forecasting step, multiple forecasts were generated by both types of models. A lightGBM model has a forecast horizon of 288, and generates forecasts of 144-288 steps ahead by rolling forecast. A LSTM model forecasts h steps ahead simultaneously but different models have different horizons.
(4) Ensemble: The median forecasts of LSTM models with the same forecast horizon were firstly calculated and then the median values of different forecast horizons were concatenated to form the final forecast of 1-144 steps ahead. Forecasts of 145-288 steps ahead were averaged from LSTM and lightGBM models.
(5) Rule-based adjustment: Based on the assumption that wind power will not change rapidly in the very near future if wind turbines are working properly, forecasts of the ultra-short future were adjusted by rules to satisfy this constraints.

Spatial information of turbines was implicitly incorporated by enforcing hierarchical coherency constraints and adding a self-attention layer along the turbine axis in the LSTM model.

3 DETAILED METHOD

3.1 Exploratory Data Analysis

In order to better design solutions, we conducted detailed exploration and analysis of data. Some conclusions are as follows:

- According to the official anomaly standard, the proportion of anomalies in this dataset is nearly 30%, so the processing of anomalies is particularly important.
- Power generation is mostly affected by the wind speed at the current moment. The relationship of max power and wind speed is a very regular curve. It is suggested that when wind speed is fixed, there is a corresponding max power generation rate, however, there may be other factors, like the wind turbine is not along the wind direction and the pitch angles are not appropriate, that cause the power to be below its maximum value.
- Some turbines report extremely abnormal temperature values which might be caused by broken thermometers. The temperature has a robust daily seasonality component but in the training data, the mean temperature per day slowly decreases which might indicate the data has been collected from summer to winter. Under fixed wind speed, temperature has a negative correlation with power generation, possibly caused by higher air density at lower temperature. However, power generation has a positive correlation with temperature because temperature is positively correlated with wind speed.
- When wind speed is greater than $2.5m/s$, the angle between wind direction and turbine `Wdir` is around 0°. But the wind direction respect to north `Wdir+Ndir` is nearly randomly distributed and we suppose this is because the absolute zero angle is different for different turbines.
- The absolute value of reactive power is positively correlated with active power.

Our general conclusion of the task from exploratory data analysis was that only short-term wind power can be effectively foretasted, and for the middle-term forecast, appropriate mean values can be foretasted at best, as it is difficult to forecast wind speed hours later.

3.2 Anomaly Detection and Data Preprocessing

Based on official instructions [21], wind power may have zero values, missing values, unknown values, and outliers for different reasons, which can be treated as data anomalies in general. The given dataset has nearly 30% anomalies and these anomalies may bring negative influences. For example, extreme values will make time series constrained to a narrow region after standardization, and these anomalies will cause difficulty for model learning temporal dependencies .

Table 1 shows our anomaly analysis and preprocessing methods. Specifically, Wdir and Ndir has occasional measurement errors. For example, in day 49 and 50, Wdir values of many turbines are not within the range $[-180°180°]$. On the other hand, only turbine 2 had an abnormal Ndir from 3:40 to 12:00 in the day of 185. For the anomalies of Wdir and Ndir, linear interpolation was used. In addition, there are ten turbines recorded extreme Etmp, possibly caused by broken thermometers. Since the external temperature should be similar for turbines at the same time, turbines reported Etmp out of 3 times standard deviation were treated as anomalies and adjusted to the median values at the current time. The anomalies of Pab1/2/3 were difficult to repair, so we only used the normal ratio of this value as a feature.

More anomaly detection and data processing methods will be described in detail in the remaining part of this section.

3.2.1 Filling in Unknown Wind Powers. As there are nearly 30% unknown values for wind powers, a lightGBM model was trained from data with normal power values, and was then used to predict those unknown values. We also added noise by sampling from the distribution of normal data at the same wind speed to alleviate overfitting.

3.2.2 Wind Speed - Max Power Curve. The relation of turbine power and wind speed was deduced from physical laws as shown

Table 1: Anomalies Overview and Preprocessing Methods

data	Anomalies	Abormal Percentage	Possible Causes	Preprocessing Methods
Wspd	<0	0	-	-
Wdir	out of [-180,180]	0.002%	Occasional Measurement Error	Interpolation
Ndir	out of [-170 170]	0.3%	Occasional Measurement Error	Interpolation
Wind direction(Wdir + Ndir)	nearly random distribution	-	zero angle difference	wind angle adjustment
Etmp	out of $[\mu_t - 3\sigma_t : \mu_t + 3\sigma_t]$	6%	Broken Thermometer	Median Filling
Pab1/2/3	>89	20.9%	Power Scheduling	Drop
Patv	<0 Wspd>2.5 and Patv<0 Pab1/2/3 >89 Wdir is out of [-180 180] is out of [-170 170]	29%	Power Scheduling Turbine Renovation Measurement Error	treat <0 as 0 LightGBM wind speed - max power curve adjustment

in equation 1[15],

$$P_e = \frac{1}{2}\rho S v^3 C_p \tag{1}$$

where ρ is air density, S is turbine sweap area, v is wind speed and C_p is efficiency of wind turbine. However, actual power is dependent on operating conditions, such as the turbine Patv, moisture, temperature, etc. Especially, C_p is dependent on the wind speed v. For the dataset of KDD2022[21], we formulated a hypothesis: for a fixed wind speed v, there is a maximum power $P_{max} = f(v;\theta)$, and under different operational conditions, the actual power $P \le P_{max}$. Data points with $P > f(v;\theta)$ were regarded as abnormal values and the corresponding P were adjusted to $f(v;\theta)$.

Inspired by the theory of sparse kernel machines, we inferred appropriate envelopes $P_{max} = f(v;\theta)$ for each turbines from data. Here the problem was formulated a bit differently: The data was split into G histograms according to wind speed. The center of each group is $v_g, g = 1, 2, \ldots, G$. For this dataset, $f(v;\theta)$ is a nonlinear function described by equation 2.

$$f(v,\theta) = f(v;\alpha,\beta,s) = \begin{cases} \beta v^\alpha, v < s \\ \beta s^\alpha, v \ge s \end{cases} \tag{2}$$

$\theta = (\alpha, \beta, s)$ was calculated by minimizing the function in equation 3.

$$\theta^* = \underset{\theta}{\operatorname{argmin}} \left[C \sum_{n, n\in g} (P_n - f(v_g;\theta))^+ + \sum_g (f(v_g;\theta) - \max_{n\in g, P_n \le f(v_g;\theta)} P_n) \right] \tag{3}$$

where C is a hyper-parameter describing the penalty of data points above the curve.

Under this formulation, the max power to wind speed envelope only depends on a very small subset of the data points(the support vectors): those above the curve and those just below it, and invariant to other data samples. The minimization problem was solved by `scipy` package in python. The result showed that the α values of nearly all turbines were around 2, instead of 3 in equation 1. Sample curves are shown in figure 1.

3.2.3 Wind angle adjustment. According to the official report and discussion replies, wind direction respect to north is the sum of

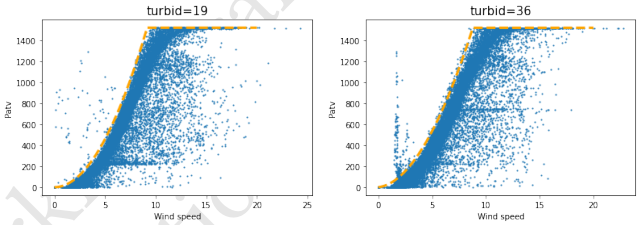

Figure 1: Relationship of max turbine power and wind speed. Yellow lines represent max power to wind envelope. Data points above the envelope were treated as anomalies adjusted to respective values on the envelope.

nacelle position and wind direction respect to the nacelle, as shown in figure 2. However, we found that at a time snapshot, the wind directions were nearly randomly distributed (figure 3(a)). We suppose this was because the zero angles of each turbines were different. So the wind directions at each turbine were adjusted based on a very simple assumption: the wind directions were the same for all turbines on the farm if the wind speed is greater than $10m/s$. During

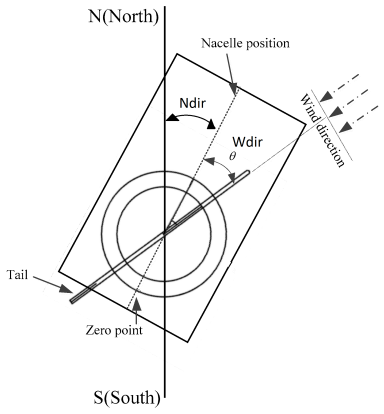

Figure 2: Wind direction respect to north is the sum of nacelle position and wind direction respect to the nacelle.

the adjustment, the mean of a group of angles was calculated by transforming each angles to their respective unit vectors on a 2D surface and the mean angle is the angle of the vector sum of all the unit vectors.

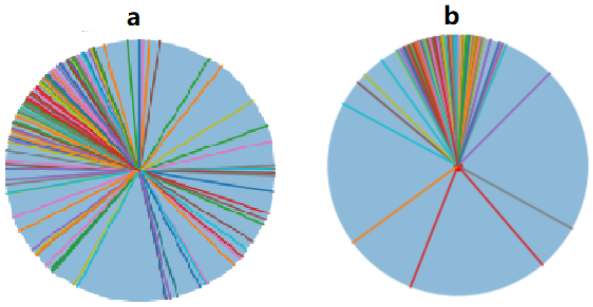

Figure 3: Wind direction at day 1 06:20 of the training data. a: original distribution of wind directions at each turbines. b: wind directions after adjustment.

After angle adjustment, most of the wind angles at the same time were much similar, expect for only a few outliers, as shown in figure 3(b).

3.3 Feature engineering

Features used for our methods are summarized as follows:

- `Wind Direction`: The wind direction reflects the projection angle of the wind speed in a specific direction. Since it is a circular feature, its sin and cos values are used as features for models.
- `Minutes of the day`: This is also a circular feature, and $sin\frac{2\pi m_t}{1440}, cos\frac{2\pi m_t}{1440}$ are used as features for models, where m_t is minutes of the day.
- `Max power features`: The max power based on the current wind speed, and the difference of max power and actual power was used as features.
- `Sliding window Features`: For values wind speed, external and internal temperature, active and reactive powers, max power values and wind directions, we calculated the mean, maximum, minimum, variance, and median values within the history window.
- `tsfresh Features`: Since the manual features mining may not be comprehensive enough, the `tsfresh` library [5] was used to automatically generate time series features. However, automatic feature generation is memory expensive. So feature selection was performed according to the correlation with target values. The final top 59 features generated include fourier entropy and coefficient, auto-correlation coefficient, etc.

3.4 Hierarchical coherency constraints

Turbine locations were implicitly incorporated in multi-variate LSTM model by two mechanisms: hierarchical coherency constraints and a final self-attention layer.

Nearby turbines were clustered into groups. The power of groups and turbines should satisfy hierarchical coherence that the sum powers of turbines within a group should be equal to the total power of the group. For example, as shown in figure 4, $x_{c1} = \sum_{i=113}^{134} x_i$. The constraints can be expressed by equation 4, where x is the power of all turbines and groups and A is a matrix solely decided by the hierarchical structure of the turbines and groups.

$$Ax = 0 \tag{4}$$

The power of all turbines and groups were predicted simultaneously with multi-variate LSTM, and then pass through a coherence layer, after which, constraints equation 4 will be satisfied.

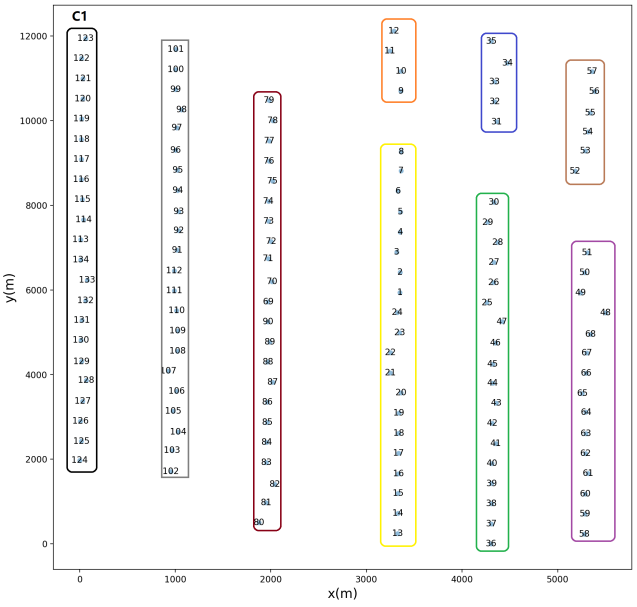

Figure 4: Wind turbines were separated into clusters based on their locations.

3.5 Models

3.5.1 LightGBM Model. LightGBM model [8] was used to forecast 145-288 steps ahead, a kind of long sequence time-series forecasting (LSTF). The model predicts the 288th value ahead, and continuous forecasts were achieved by rolling prediction. LightGBM uses the histogram algorithm to find the optimal split point. Also it uses the Gradient-based One-Side Sampling (GOSS) which retains all samples with large gradients, and then random sample among samples with small gradient. In addition, for the sparsity of the feature space, Exclusive Feature Bundling (EFB) is designed to reduce the feature dimension. Although compared with XGBoost [4], LightGBM sacrifices a bit of accuracy but it reduces the risk of overfitting and increases training and inference speed. For LSTF tasks, especially when the data distribution is unknown, LightGBM with better generalization performance is more preferable. In addition, the parallel optimization of LightGBM improves the efficiency of iterative experiments.

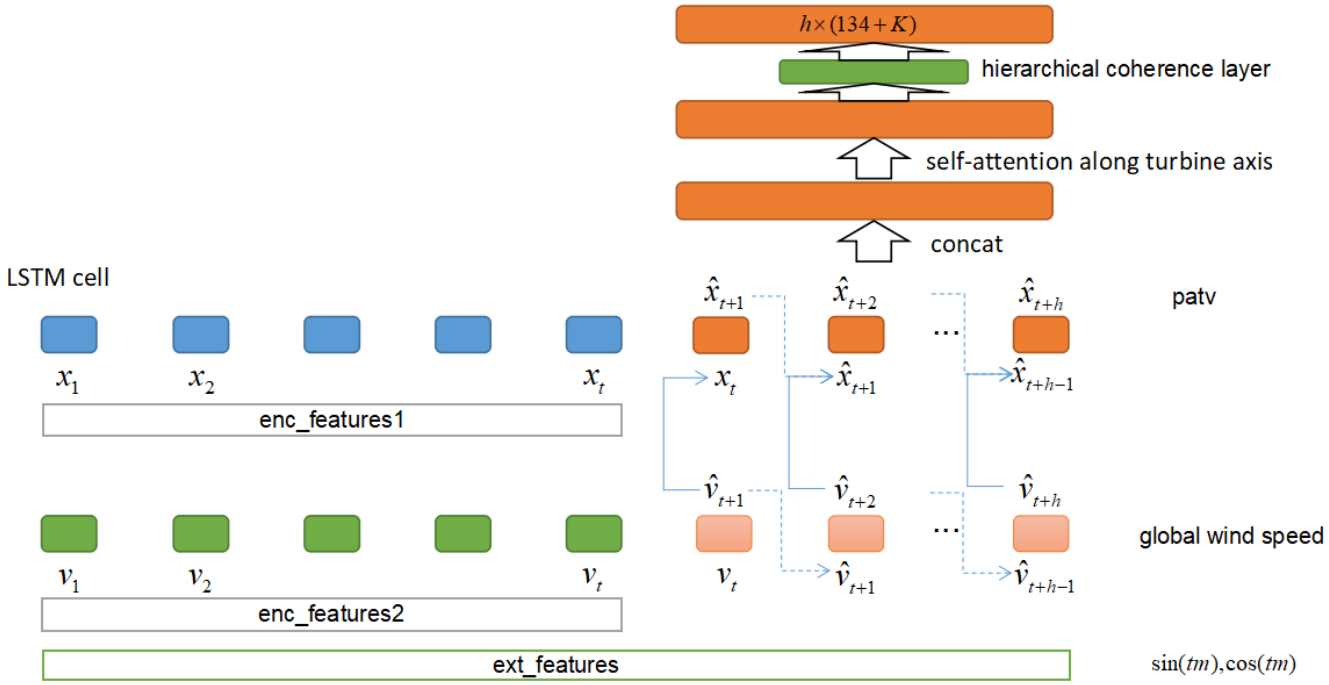

Figure 5: Multi-variate LSTM model with hierarchical coherence constraints.

3.5.2 Multi-variate LSTM with Hierarchical Coherence Constraints. Our second model structure is a multi-variate sequence-to-sequence model based on LSTM structure. Two LSTM layers were included in our model. The first layer forecasts the multi-variate turbine powers for each turbines and turbine clusters. The second layer forecasts the global wind speed, the median wind speed of all turbines.

The geological structure was incorporated implicitly in the LSTM model through two mechanisms: hierarchical coherence constraints and a self attention layer. $\hat{x}_t$ is the forecasts of multi-variate time series of all turbines and clusters t steps ahead by the first LSTM layer. But the coherence constraints described by equation 4 are now not satisfied. $\hat{x}_t$ is then passed into a hierarchical coherence layer. The output of this layer $\hat{x}'_t$ satisfies the coherence constraints with minimal change of $\hat{x}_t$, as shown in equation 5, where subscript t is dropped for convenience.

$$\begin{aligned} &\hat{x}' = \underset{\hat{x}'}{\operatorname{argmin}} \frac{1}{2}||\hat{x}' - \hat{x}|| \\ &\text{s.t.} \\ &A\hat{x}' = 0 \end{aligned} \tag{5}$$

The minimization problem 5 can be solved analytically using Lagrange multiplier by minimizing $L(\hat{x}', \lambda) = \frac{1}{2}||\hat{x}' - \hat{x}|| + \lambda A\hat{x}'$, and the solution is shown in equation 6.

$$\hat{x}' = (I - A^T(AA^T)^{-1}A)x \tag{6}$$

where I is the identity matrix .

Finally, forecasts at different horizons are concatenated and passed into a self-attention layer, trying to capture the relations among different turbines and groups. The multi-variate LSTM model was implemented in PaddlePaddle[12], an open-source deep learning platform from industrial practice.

3.6 Rule-based Adjustment

Results of both lightGBM and multi-variate LSTM models showed that, as models minimize the loss of all horizons of their forecasts during training, the forecasts fluctuate very little and just try to learn the mean of wind power, especially for lightGBM and LSTM with long forecasting horizons. Based on the assumption that wind power will not change rapidly in the very near future if wind turbines are working properly, forecasts of the ultra-short future were adjusted by rules to satisfy this constraints, as shown in equation 7.

$$\hat{y}_\tau = \hat{x}_\tau + w_\tau(\overline{y}_b - \hat{x}_\tau) \tag{7}$$

$$w_\tau = \frac{(\tau - \Gamma)^2}{\Gamma^2}, \tau \in [0, \Gamma]$$

where $\hat{x}_\tau$ is the forecast τ steps ahead; $\hat{y}_\tau$ is the adjusted forecast; w_τ is the weight of adjustment; $\overline{y}_b$ is the mean historical power start from b steps back; Γ is the maximum length of modification. Before the adjustment, the abnormal ratio in historical power will be calculated and if this value is higher than a threshold, the forecasts will not be adjusted. Figure 6 shows the rule-based adjustment of lightGBM forecast on the test dataset.

4 EXPERIMENTS

In this section, we will not as usual show our method comparing with other methods on multiple datasets, but compare and discuss the effect of different model configurations based on results on the same datasets.

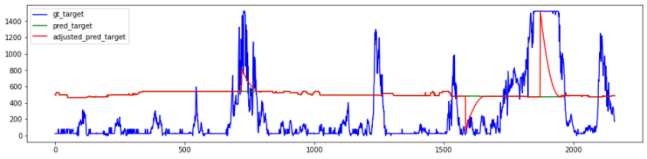

Figure 6: Rule-based Adjustment on the prediction of the test dataset by lightGBM

4.1 lightGBM model

For lightGBM models, 214 days of data was used for training, 16 days for validation and the final 15 days for test. The baseline lightGBM model used historical wind power and wind speed features of sliding window statistics. As shown in table 2, filling unknown wind power by another lightGBM model decreased offline performance but increased online performance. Adding max power - wind speed adjustment and features somehow reduced offline error. And `tsfresh` features compromised offline but slightly increased online performance.

4.2 Multi-variate LSTM model

For multi-variate LSTM model, we show the results where first 109 days of data was used for training and then 16 days for validation. And 200 random samples from the last 15 days were used for test. The large time span between training and test data might show the effect of data drift in this forecasting task.

As shown in table 3, attention among turbines increased offline performance. Then adding hierarchical coherence constraint slightly increased forecast error but the difference was not significant. Actually, for the online dataset of Phase I, hierarchical coherence constraint decreased the error from 41.79 to 41.37.

Multi-variate LSTM model trained by different horizons showed different forecasts for the same input data. An example is shown in figure 7. Generally, model trained with shorter forecasting horizons

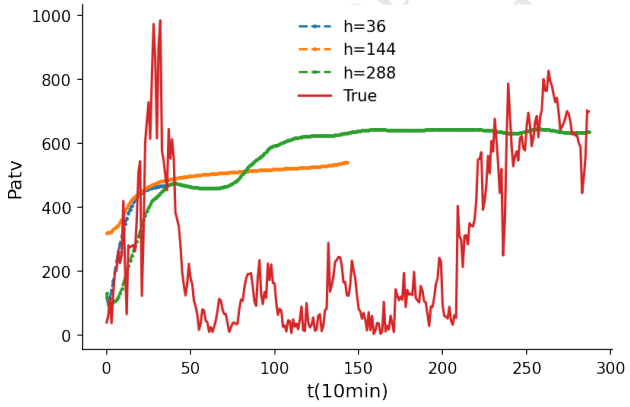

Figure 7: An instance of forecasts by models trained with different horizons.

will have better performance for short-term forecasts, as shown by figure 8. Thus concatenating forecasts of models trained by different horizons might be a better solution than a single model.

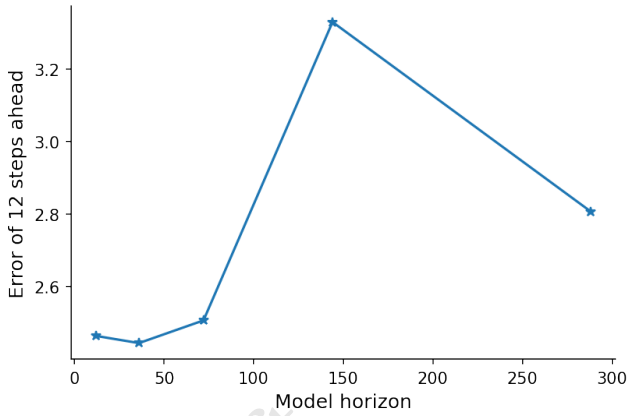

Figure 8: Forecast errors of 12 steps ahead of model trained with different forecasting horizons.

4.3 Final Model Configurations

The model configurations were updated sequentially, not based on searching of optimal combinations. Due to the versatile nature of the data, many results were not consistent on different datasets and conclusions were hard to draw. Thus the final solution were selected as follows, without elimination of subjective bias: multiple lightGBM and LSTM models were trained with different hyperparameters. The median forecasts of LSTM models with the same horizons were firstly calculated and then concatenated to form the LSTM ensemble forecast. The mean forecasts of lightGBM models formed the lightGBM ensemble forecast. The 1-144 steps of LSTM ensemble forecast and the mean value of LSTM and lightGBM ensemble forecasts 145-288 steps ahead were concatenated to form the final forecast.

50 samples were randomly selected from last 31 days of data. The results of lightGBM, multi-variate LSTM model and their combination on both offline and online (phase 3) datasets are shown in table 4.

5 CONCLUSION

In this work, we developed a hybrid method combined of lightGBM and multi-variate LSTM for wind power forecasting of individual turbines in a wind farm. The lightGBM forecasts middle-term steps ahead (24-48hours) and accurately forecast the mean value of future wind power. The LSTM model forecasts short-term steps ahead (within 24 hours). Spatial information of turbines were implicitly incorporated into our methods by enforcing hierarchical coherence constraints and adding a self-attention layer along the turbine axis. We developed novel anomaly detection methods like max power to wind speed curve based on support vectors to enhance forecasting accuracy. For future work, we will explicitly include spatial information by graph attention, and will develop appropriate recurrent model structures for multi-variate time series composed of similar single time series, like the powers of wind turbines.

Table 2: LightGBM Experiments

Model Configuration(successive)	Offline MAE	Offline RMSE	Offline Score	Online Score	offline p-value
Baseline	39.37	42.83	41.10	41.92(Phase I)	-
Filling Unknown	41.85	44.92	43.39	41.02(Phase I)	1.3e-64
Max Power-Wind Speed	41.31	44.42	42.87	46.57(Phrase III)	1.6e-5
`tsfresh` Features	43.04	45.99	44.51	46.54(Phrase III)	4.2e-36

Table 3: Multi-variate LSTM Experiments

Model Configuration(successive)	Offline MAE	Offline RMSE	Offline Score	offline p-value
Baseline	46.18	47.82	47.00	-
Attention	44.47	45.88	45.17	8e-12
Hierarchical Coherence	44.79	47.49	45.64	0.07

Table 4: Test Results of Final Method

dataset	lightGBM	multi-variate LSTM	ensemble
offline	40.88	42.95	41.20
online	46.73	45.87	45.64

6 REFERENCES

REFERENCES

[1] Hanieh Borhan Azad, Saad Mekhilef, and Vellapa Gounder Ganapathy. 2014. Long-term wind speed forecasting and general pattern recognition using neural networks. *IEEE Transactions on Sustainable Energy* 5, 2 (2014), 546–553.

[2] Lars Ødegaard Bentsen, Narada Dilp Warakagoda, Roy Stenbro, and Paal Engelstad. 2022. Wind Park Power Prediction: Attention-Based Graph Networks and Deep Learning to Capture Wake Losses. In *Journal of Physics: Conference Series*, Vol. 2265. IOP Publishing, 022035.

[3] Niya Chen, Zheng Qian, Ian T Nabney, and Xiaofeng Meng. 2013. Wind power forecasts using Gaussian processes and numerical weather prediction. *IEEE Transactions on Power Systems* 29, 2 (2013), 656–665.

[4] Tianqi Chen and Carlos Guestrin. 2016. Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*. 785–794.

[5] Maximilian Christ, Nils Braun, Julius Neuffer, and Andreas W Kempa-Liehr. 2018. Time series feature extraction on basis of scalable hypothesis tests (tsfresh–a python package). *Neurocomputing* 307 (2018), 72–77.

[6] Ergin Erdem and Jing Shi. 2011. ARMA based approaches for forecasting the tuple of wind speed and direction. *Applied Energy* 88, 4 (2011), 1405–1414.

[7] G Kariniotakis, E Nogaret, and G Stavrakakis. 1996. A fuzzy logic and a neural network based wind power forecasting model. (1996).

[8] Guolin Ke, Qi Meng, Thomas Finley, Taifeng Wang, Wei Chen, Weidong Ma, Qiwei Ye, and Tie-Yan Liu. 2017. Lightgbm: A highly efficient gradient boosting decision tree. *Advances in neural information processing systems* 30 (2017).

[9] Jiangyuan Li and Mohammadreza Armandpour. 2022. Deep Spatio-Temporal Wind Power Forecasting. In *ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 4138–4142.

[10] Yao Liu, Lin Guan, Chen Hou, Hua Han, Zhangjie Liu, Yao Sun, and Minghui Zheng. 2019. Wind power short-term prediction based on LSTM and discrete wavelet transform. *Applied Sciences* 9, 6 (2019), 1108.

[11] Petroula Louka, Georges Galanis, Nils Siebert, Georges Kariniotakis, Petros Katsafados, Ioannis Pytharoulis, and George Kallos. 2008. Improvements in wind speed forecasts for wind power prediction purposes using Kalman filtering. *Journal of Wind Engineering and Industrial Aerodynamics* 96, 12 (2008), 2348–2362.

[12] Yanjun Ma, Dianhai Yu, Tian Wu, and Haifeng Wang. 2019. PaddlePaddle: An open-source deep learning platform from industrial practice. *Frontiers of Data and Domputing* 1, 1 (2019), 105–115.

[13] Yang Mao and Wang Shaoshuai. 2016. A review of wind power forecasting & prediction. In *2016 International Conference on probabilistic methods applied to power systems (PMAPS)*. IEEE, 1–7.

[14] Rory Morrison, Xiaolei Liu, and Zi Lin. 2022. Anomaly detection in wind turbine SCADA data for power curve cleaning. *Renewable Energy* 184 (2022), 473–486.

[15] Madasthu Santhosh, Chintham Venkaiah, and DM Vinod Kumar. 2020. Current advances and approaches in wind speed and wind power forecasting for improved renewable energy integration: A review. *Engineering Reports* 2, 6 (2020), e12178.

[16] Osamah Basheer Shukur and Muhammad Hisyam Lee. 2015. Daily wind speed forecasting through hybrid KF-ANN model based on ARIMA. *Renewable Energy* 76 (2015), 637–647.

[17] Jianzhou Wang, Fanyong Zhang, Feng Liu, and Jianjun Ma. 2016. Hybrid forecasting model-based data mining and genetic algorithm-adaptive particle swarm optimisation: a case study of wind speed time series. *IET Renewable Power Generation* 10, 3 (2016), 287–298.

[18] Shuangyuan Wang, Yixiang Huang, Lin Li, and Chengliang Liu. 2016. Wind turbines abnormality detection through analysis of wind farm power curves. *Measurement* 93 (2016), 178–188.

[19] Yagang Zhang, Penghui Wang, Tao Ni, Penglai Cheng, and Shuang Lei. 2017. Wind power prediction based on LS-SVM model with error correction. *Advances in Electrical and Computer Engineering* 17, 1 (2017), 3–8.

[20] Dehua Zheng, Yordanos Kassa Semero, Jianhua Zhang, and Dan Wei. 2018. Short-term wind power prediction in microgrids using a hybrid approach integrating genetic algorithm, particle swarm optimization, and adaptive neuro-fuzzy inference systems. *IEEJ Transactions on Electrical and Electronic Engineering* 13, 11 (2018), 1561–1567.

[21] Jingbo Zhou, Xinjiang Lu, Yixiong Xiao, Jiantao Su, Junfu Lyu, Yanjun Ma, and Dejing Dou. 2022. SDWPF: A Dataset for Spatial Dynamic Wind Power Forecasting Challenge at KDD Cup 2022. *Techincal Report* (2022).