Short-term Offshore Wind Power Forecasting - A Hybrid Model based on Discrete Wavelet Transform (DWT), Seasonal Autoregressive Integrated Moving Average (SARIMA), and Deep-learning-based Long Short-Term Memory (LSTM)

Wanqing Zhang[a,b], Zi Lin[c], Xiaolei Liu[a 1]

[a]*James Watt School of Engineering, University of Glasgow, Glasgow, G12 8QQ, United Kingdom*
[b]Glasgow College, University of Electronic Science and Technology of China, Chengdu, 611731, China
[c]*Department of Mechanical & Construction Engineering, Northumbria University, Newcastle, NE1 8ST, United Kingdom*

Abstract

Short-term time series wind power predictions are extremely essential for accurate and efficient offshore wind energy evaluation and, in turn, benefit large wind farm operation and maintenance (O&M). However, it is still a challenging task due to the intermittent nature of offshore wind, which significantly increases difficulties in wind power forecasting. In this paper, a novel hybrid model, using unique strengths of Discrete Wavelet Transform (DWT), Seasonal Autoregressive Integrated Moving Average (SARIMA), and Deep-learning-based Long Short-Term Memory (LSTM), was proposed to handle different components in the power time series of an offshore wind turbine in Scotland, where neither the approximation nor the detail was considered as purely nonlinear or linear. Besides, an integrated pre-processing method, incorporating Isolation Forest (IF), resampling, and interpolation was applied for the raw Supervisory Control and Data Acquisition (SCADA) datasets. The proposed DWT-SARIMA-LSTM model provided the highest accuracy among all the observed tests, indicating it could efficiently capture complex times series patterns from offshore wind power.

Keywords: Short-term wind power forecasting; Offshore wind turbine; Wavelet transform; Seasonal auto-regression integrated moving average (SARIMA); Deep learning.

[1] Corresponding author, E-mail: xiaolei.liu@glasgow.ac.uk

NOMENCLATURE

Latin symbols

$(1-B^s)^D$	Seasonal difference operator
$(1-B)^d$	Regular difference operator
$\hat{L}_t^{app}$	Prediction of linear part of reconstructed approximation
$\hat{L}_t^{det}$	Prediction of linear part of reconstructed detail
$\hat{N}_t^{app}$	Prediction of nonlinear part of reconstructed approximation
$\hat{N}_t^{det}$	Prediction of nonlinear part of reconstructed detail
$\hat{y}_t$	Prediction of original time series power data
$\hat{y}_t^{app}$	Prediction of reconstructed approximation
$\hat{y}_t^{det}$	Prediction of reconstructed detail
h_t	Overall output at time step t
h_{t-1}	Cell state vector at time step t-1
$\max(x)$	Maximum value of signal
H_i	Net input of neuron j
L_t^{app}	Linear part of reconstructed approximation
L_t^{det}	Linear part of reconstructed detail
N_t^{app}	Nonlinear part of reconstructed approximation
N_t^{det}	Nonlinear part of reconstructed detail
X_{scaled}	Normalized value of signal
w_{ij}	Weight linking neuron i and neuron j
x_t	Input neuron at time step t
y_t	Original time series power data

51	y_t^{app}	Reconstructed approximation
52	y_t^{det}	Reconstructed detail
53	a_n	Low frequency component at n decomposition level
54	B	Backward shift operator
55	c(n)	Average path length of unsuccessful search in a Binary Search Tree
56	d	Difference order
57	D	Seasonal difference order
58	d_n	High frequency component at n decomposition level
59	E(x)	Average value of x
60	h	Output of neuron j
61	h(x)	Path length of data x
62	m	Scaling parameter
63	$\min{(x)}$	Minimum value of signal
64	n	Number of external nodes
65	n	Translation parameter
66	p	Autoregressive order
67	P	Seasonal autoregressive order
68	q	Moving average order
69	Q	Seasonal moving average order
70	s	Anomaly score
71	s	Number of time steps for a single seasonal period
72	t	Discrete time parameter
73	T	Length of signal
74	tanh	Hyperbolic tangent function

x	An observation
x(t)	Wind power signal
Z_t	Time series
L	Number of decomposition level
N	Length of signal
W	Corresponding weight connecting the input signal
b	Bias along with corresponding activation function

Greek symbols

Θ_Q	Seasonal moving average polynomial
ε_t	Estimated residual at time t
θ_q	Regular moving average polynomial
φ_P	Seasonal autoregressive polynomial
ϕ_p	Regular autoregressive polynomial
$\odot$	Element level multiplication
σ	Activation function

ABBREVIATION

ACF	Autocorrelation function
AdaGrad	Adaptive gradient algorithm
Adam	Adaptive Moment Estimation
AIC	Akaike's information criterion
ANN	Artificial Neural Network
AR	Autoregressive

ARIMA	Autoregressive integrated moving average
BIC	Bayesian information criterion
cA	Component of approximation at level 1
cA2	Component of approximation at level 2
cD/cD1	Component of detail at level 1
cD2	Component of detail at level 2
CWT	Continuous wavelet Transform
DWT	Discrete wavelet Transform
I	Integrated
IDWT	Inverse discrete wavelet Transform
IEC	International Electrotechnical Commission
IF	Isolation Forest
LSTM	Long Short-Term Memory
MA	Moving average
MSE	Mean square error
MAPE	Mean absolute percentage error
NaN	Not a number
NMAE	Normalised mean absolute error
NRMSE	Normalised root mean square error
NWP	Numerical weather prediction
ORE	Offshore Renewable Energy
PACF	Partial autocorrelation function
PMG	Permanent Magnet Generator
R^2	R-square

ReLU	Rectified Linear Unit
RMSE	Root mean square error
RMSProp	Root Mean Square Propagation
RNN	Recurrent Neural Network
SARIMA	Seasonal Autoregressive Integrated Moving Average
SCADA	Supervisory Control and Data Acquisition
SVM	Support Vector Machine
WT	Wavelet Transform

1. Introduction

In recent years, renewables have been considered as an effective alternative that can replace conventional power sources. Among them, wind energy has become one of the most attractive supplies, which is expected to provide 20% of electricity for the global demand by 2030 [1]. It can be seen that wind turbine installations are growing sharply [2], especially offshore wind turbines [3], which are expected to own over 234 GW capacity worldwide in recent decades [4]. As one of the most suitable locations for wind energy developments, the United Kingdom has committed to greatly extending offshore wind capacity [5]. However, as the demand for wind energy continues to upgrade, the uncertainty of wind power integration also increases due to the intermittent, uncertainty and volatility of the wind power, and thus trigger difficulties in grid operation. Therefore, accurate wind power prediction is highly desired to effectively dispatch these issues on a reasonable schedule.

1.1 Motivation and incitement

The operation security of the power network relies on the stability of power generations, where the balance between electricity generation and consumption needs to be maintained, otherwise disturbances in power quality/supply may occur and thus leads to significant financial loss. An accurate wind power prediction can optimize the integration of wind energy into the electricity grid. It showed that an increase of 10% in prediction accuracy can achieve about a 30% improvement of wind power generation [6]. Therefore, it is of great practical significance to develop a wind power prediction model of high accuracy.

1.2 Literature review

Over the years, various wind power prediction models have been developed, which can be coarsely categorized as a physical model, statistical model, intelligent model and hybrid model. Physical models mainly refer to numerical weather prediction (NWP) models. One advantage of physical models is the capability to make power predictions directly from real-time data. But using NWP parameters requires a large amount of historical data with high precision, increasing the difficulty in data collection as well as economic cost. On the other hand, statistical models treat weather changes as a random process, in which prediction errors can be reduced if the input signal is under normal conditions [7]. This type of model can efficiently exploit historical data and explain linear signals well [8], while it cannot effectively capture nonlinear signals. These difficulties can be addressed by using intelligent models, which are mainly based on Artificial Neural Networks (ANN). These models used non-linear methods to predict targets based on historical variables. Later, deep learning with a deeper neural network has been proposed as a powerful tool to dig out useful information in complex signals, especially for those time series data with extreme variations.

Among these models, Autoregressive Integrated Moving Average (ARIMA) [9] is one of the most commonly used methods for univariate time series predictions. For example, Yatiyana *et al*. [10] used ARIMA to predict wind speed and direction for wind power generation, where collected signals were processed to get an hourly average data. This single ARIMA model presented Mean Absolute Percentage

Error (MAPE) of 4.9% for wind speed and 15.6% for wind direction. Additionally, Seasonal ARIMA (SARIMA) has been proposed later as an extension of ARIMA, which can support signals with an additional seasonal component. This model removes characteristics of seasonal variations using seasonal differencing, improving the prediction accuracy of wind power.

Recently, Long Short-Term Memory (LSTM) has also become a widely used deep learning method, which addressed the problem of gradient explosion in traditional neural networks. LSTM has the capability of learning and remembering both short and long-term information, which is suitable to be used for time series predictions. For instance, Zhang *et al*. [11] used LSTM models to predict wind power generation, where the first 24 historical data were used to predict the data at the next hour. It presented lower Normalised Mean Absolute Error (NMAE) and Normalised Root Mean Square Error (NRMSE) of 0.059 and 0.06 than that of using Support Vector Machine (SVM) (0.087 and 0.11), respectively.

Although these single methods have made a breakthrough in terms of prediction performance, they are still not sufficient for accurate wind power prediction. Wind power generation is caused by various natural factors, such as wind speed/direction, air pressure and wind turbine friction, which makes the output power of wind turbines non-stationary and volatile. When comes to times series power data mixed with both linear and nonlinear information, neither statistical models nor intelligent models can solely make an accurate prediction. That is, although ARIMA/SARIMA and LSTM models can be used to predict times series data, each of them is only suitable for either linear or nonlinear problems. In specific, ARIMA/SARIMA can effectively explain linear information, such as trends in time series power, while failing to capture nonlinear ones. On the other hand, LSTM with a deep learning neural network can address this problem while cannot process purely linear information or signal with the characteristic of seasonality. Based on this fact, hybrid models were proposed in this paper, aiming to utilize the unique strength of each model to achieve more accuracy and robust predictions than those using a single model.

Hybrid models can be further combined with decomposition strategies. Among these decomposition methods, wavelet transform (WT) has attracted the most interest nowadays [12]. WT decomposes a signal into a high-frequency component (detail) and low-frequency component (approximation), which make them more stationary and easier for further analysis. When combining WT with hybrid models, the decomposed components can be fitted into models individually. This type of hybrid model utilizes the strength of different prediction models as well as the ability of WT. Recently, Khandelwal *et al*. [13] have proved that using WT can enhance prediction accuracy for time series forecasting. According to the authors, time series could be decomposed into high and low-frequency components and then be reconstructed using inverse transform. The reconstructed approximation and detail are fit into ARIMA and ANN, respectively. The prediction accuracy was improved compared with using either single ARIMA or single ANN, which presented MAPEs of 1.97%, 4.11% and 3.71%, respectively. Instead of using ARIMA in a hybrid model, SARIMA could also be combined with WT and ANN [14]. Unlike the methods mentioned above, in the current proposed hybrid model, the approximation is fitted into SARIMA and detail is fitted into ANN, where a higher prediction accuracy was achieved. Besides, the proposed hybrid model has been designed without linear or nonlinear assumptions on the approximation and the detail [15]. Time series data is first decomposed by discrete WT (DWT) to obtain the approximation and the detail. Then the two decomposed components were separately analyzed by both ARIMA and ANN.

1.3 Objective and methodology

The major objective of this study is to utilize the unique strength of both linear and non-linear techniques to construct a hybrid model to predict wind power generation from historical turbine data collected from a target offshore wind turbine. The proposed hybrid model is based on SARIMA and deep-learning-based LSTM without assumptions of linear and nonlinear components. Meanwhile, WT was applied to further improve the prediction accuracy, where the effect of decomposition level is critically investigated. Additionally, to improve the quality of used datasets, several techniques are used in data pre-processing, including Isolation Forest (IF), re-sampling, and interpolation. The methodology of this study is summarized in **Fig. 1**.

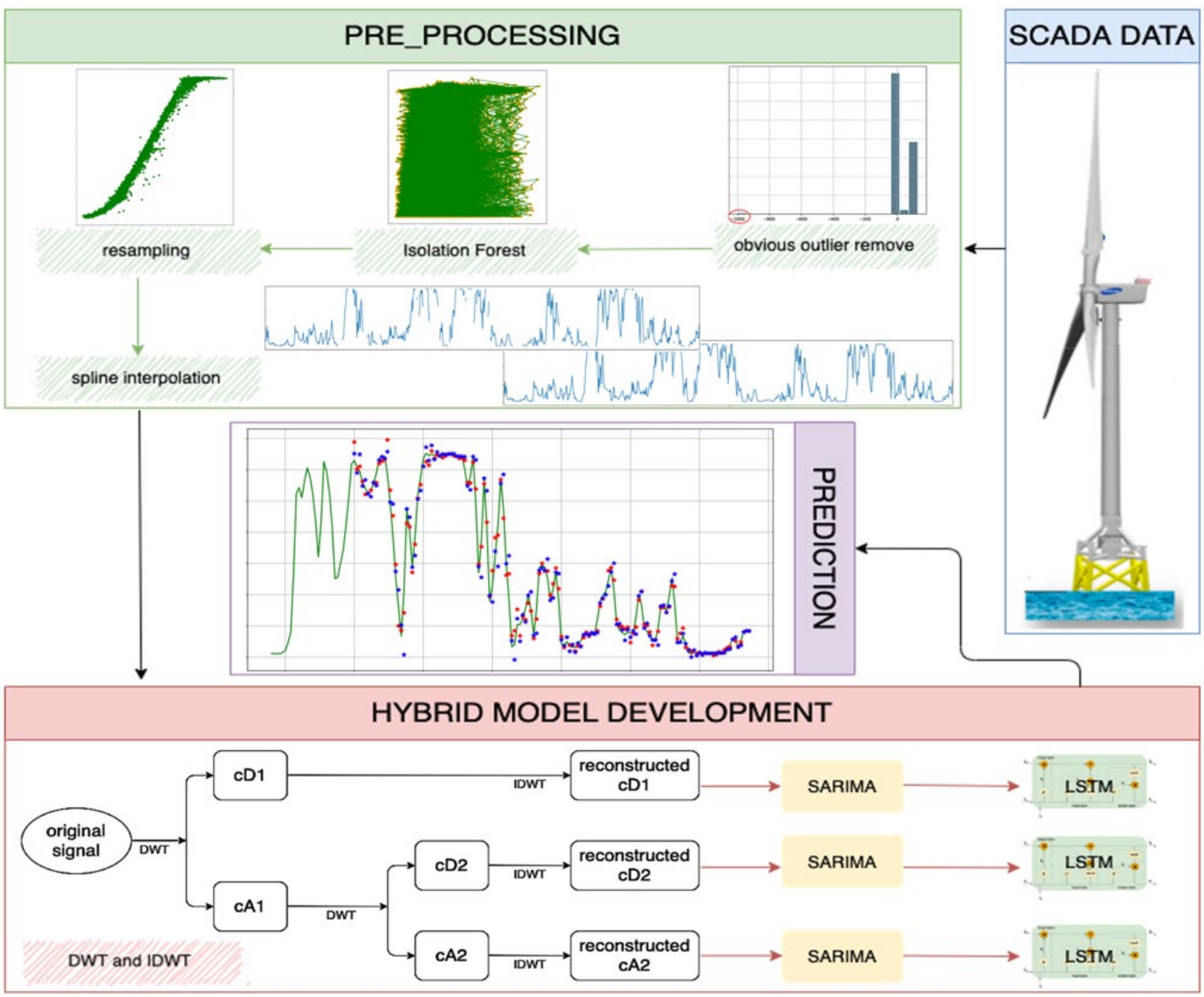

Fig. 1. Diagram of the applied methodology.

1.4 Contribution and paper organization

The key contributions of this paper to the current knowledge gaps can be summarised as follows:

- Existing studies on wind power prediction using hybrid models have been mainly based on the assumption of using linear and nonlinear models, to process approximation and detail components of wind power data, respectively. However, time series after DWT cannot be divided into linear and nonlinear data. This study has proposed to process approximation and detail components with both linear and nonlinear models, such as ARIMA and LSTM. Besides, to date, no study has considered the seasonality effect on time series on wind power. In this paper, a novel hybrid model using the unique strength of SARIMA and LSTM is proposed, predicting both approximation and detail components for an offshore wind turbine in Scotland.
- Many studies developing linear models for wind energy forecasting have not considered a thoroughly pre-process step. However, unsatisfied datasets may cause inaccurate prediction performance. For example, SARIMA models, which can be applied for time series with seasonality, require a dataset with continuous time stamps. In this study, interpolation was used to mitigate the effect of missing data, which improved the reliability and accuracy of the SARIMA model.
- Besides, IF is used in pre-processing to detect and remove outliers in the used dataset after obvious outlier removal. This outlier detection algorithm has recently been proved to be suitable for wind power forecasting [16]. It can effectively and efficiently eliminate error data far away from normal points, reducing computation time and costs.

The remainder of this paper is organized as follows. Section 2 provided a description of the target wind turbine and the used Supervisory Control and Data Acquisition (SCADA) database. Section 3 presented the used pre-processing strategies, including outlier detection/removal, resampling and missing data treatment. Section 4 introduced the theories and background of model development, including WT, SARIMA and LSTM. Section 5 presented results and discussion of the proposed hybrid model, where the prediction accuracy of various models was analyzed. Section 6 concluded this study by summarizing the key findings and contributions of the current paper, and also limitations and future perspectives.

2. SCADA data description

The target offshore wind turbine is owned by Offshore Renewable Energy (ORE) Catapult, located at Levenmouth, Fife, Scotland, UK (see **Fig. 2**). It is a 7MW offshore wind turbine with a total height of 196 m. As for operating regions, it has a designed cut-in speed of 3.5 m/s, a rated speed of 10.9 m/s and a cut-out speed of 25 m/s, respectively. The target turbine was controlled and monitored by a SCADA system, which can deliver power outputs by default without extra costs [17]. In this study, SCADA datasets were extracted for wind power forecasting. The investigated SCADA datasets were recorded with a sampling rate of 1-s. A one-month time series database (January 2019) was selected as the used dataset for model developments. The train-test split percentage of 0.8-0.2 is selected in this study. The used dataset (744 points) is split into two parts: a training set (600 points) and a testing set (144 points).

Properties	**Values**
Wind Class	IEC Class IA/ SB
Rotor diameter	171.2 m
Capacity	7 MW at grid side
Hub height	110.6 m
Blade length	83.5 m
Total height	196 m blade tip to sea level
Generator	Medium voltage PMG (3.3 kV)
Converter	Full Power Conversion
Drive train	Medium speed (400 rpm)
Rated frequency	50 Hz
Rotor speed	5.9 ~ 10.6 rpm
Wind speed	3.5 ~ 25 m/s
Temp. range	Survival: -20°C to +50 °C Operating: -10°C to +25 °C
Design Life	25 years

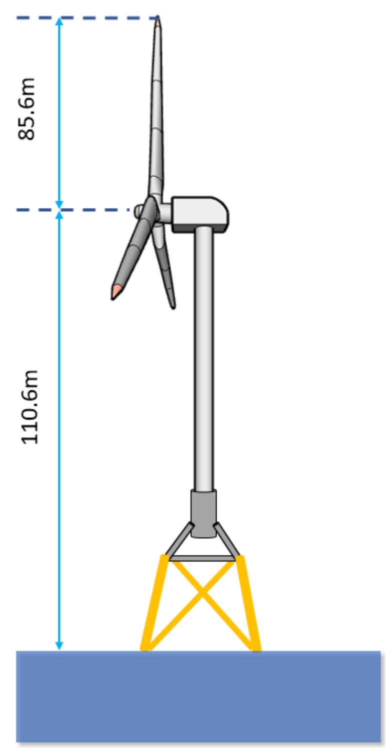

Fig. 2. Schematic and major characteristics of Levenmouth offshore wind turbine, after [18].

3. SCADA data pre-processing

Although SCADA data can be used for wind turbine power prediction, it is still challenging to achieve an optimum strategy due to possible erroneous data points within the datasets. These invalid data points mainly originate from maintenance, sensor malfunction/degradation or system processing errors during wind turbine operations, which are detrimental to the prediction model. Therefore, it is expected to pre-process SCADA data before using them to build a model [17].

3.1 Obvious outlier removal

The histograms of wind speed, active power and blade pitch angle in the raw SCADA datasets are shown in **Fig. 3**, where negative values are representing obvious outliers. For example, an extremely negative value of around -1000° (in red circle) is located in the case of the blade pitch angle histogram. These negative values are physically possible but have no practical meaning in terms of wind power generation. Therefore, these obvious outliers would be removed along with the corresponding variables under the same time stamps.

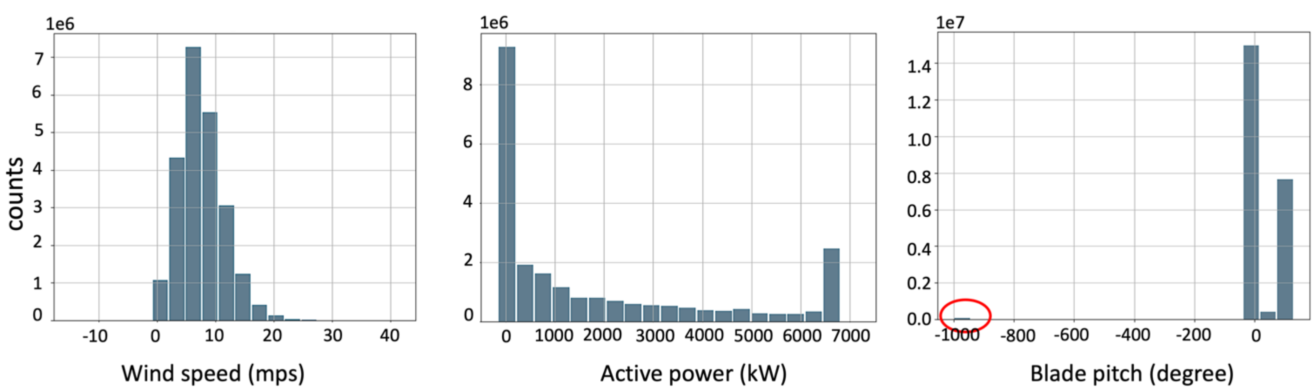

Fig. 3. Histograms of wind speed (left), active power (middle) and blade pitch (right) in the used SCADA database.

3.2 Anomalies observation

The power curve of a wind turbine could show the relationship between the amount of generated wind power and the corresponding wind speed, which is an important metric of wind turbine performance [19]. Theoretically, the power curve should be in the shape of the sigmoid function ('S' shape) [20]. As shown in **Fig. 4**, compared to a normal 'S' shape, the power curve of the target offshore wind turbine still shows some outliers that deviate from normal observations after obvious outlier removal. These outliers are caused during operation and can be mainly categorized into three types of anomalies [2]:

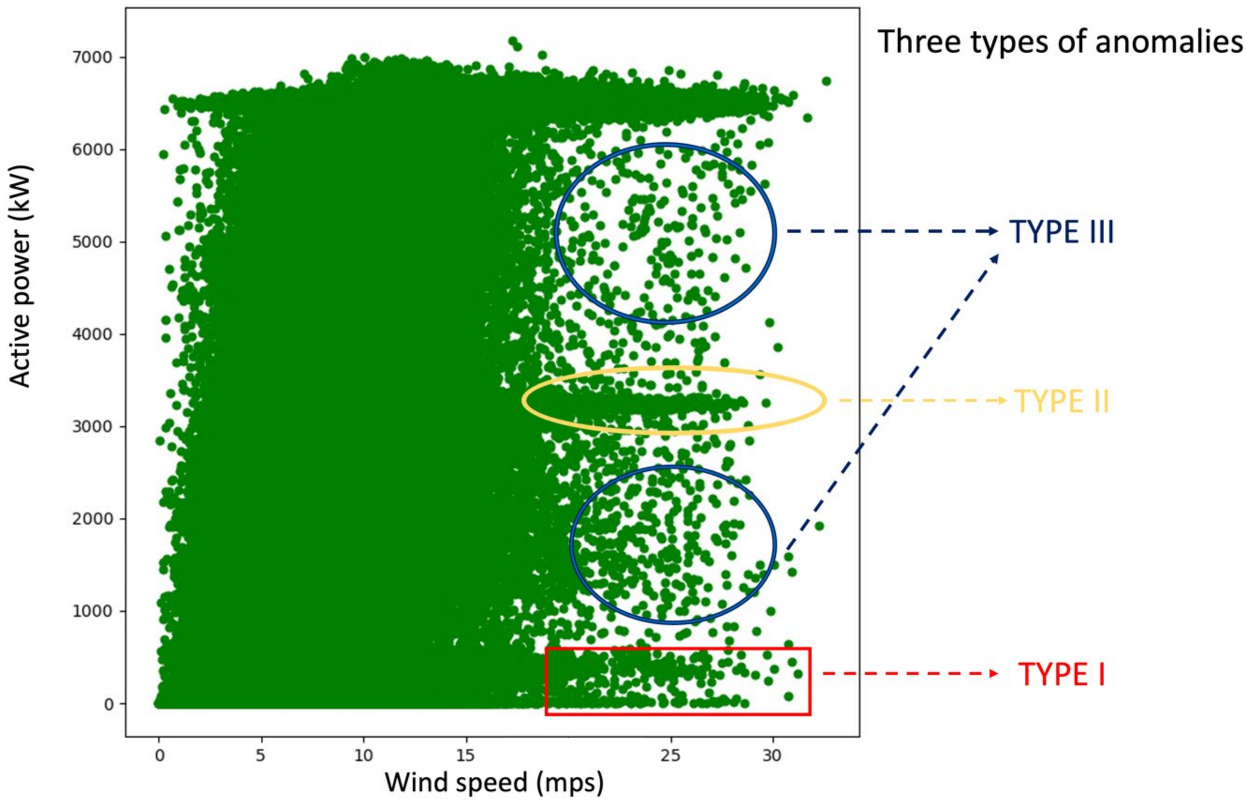

Fig. 4. Wind power curve after obvious outlier removal.

- Type I: this type of anomaly are mainly caused by turbine downtime [21], where the wind speed is larger than its cut-in speed (3.5 m/s) while the wind power is about zero.
- Type II: this type of anomaly is mainly caused by wind curtailment, where the output power is artificially limited by its operator due to different factors i.e., challenges in large capacity power storing or grid supply limitations.
- Type III: this type of anomaly is mainly caused by sensor malfunction/degradation [22].

3.3 Anomalies detection and treatment

The issue of power curve outlier rejections discussed above leads to degradation of forecasting model performance, which should be considered in the pre-processing stage. A novel outlier detection method of IF is used in this study. IF is an outlier detection method based on a binary tree structure, which has been proposed as an effective algorithm in wind power prediction [23]. Besides, IF can be more effective to process datasets of large size [24], where SCADA datasets usually have multiple input features and a large data size due to their high sampling rate. The principle of IF is isolating anomalies explicitly because occurrence frequencies and values of normal/abnormal data are usually significantly different so that outliers are usually far away from these normal data points. The anomaly score 's' of an observation x can be defined as **Eq. 1**:

$$s(x,n) = 2^{-\frac{E(h(x))}{c(n)}} \tag{1}$$

where n is the number of external nodes, $h(x)$ is the path length of data x, $E(h(x))$ means the average of $h(x)$ from a collection of isolation tress and $c(n)$ is the average path length of unsuccessful search in a Binary Search Tree.

After removing obvious outliers from the used dataset, IF is applied to detect and remove anomalies. The anomaly score 's' is set to 1 for normal points and -1 for anomalies. The range of contamination ratio from 1% ~ 20% was investigated. Subsequently, the contamination ratio of 14% was identified as the optimal parameter for the current dataset. As shown in **Fig. 5**, detected outliers are represented by red dotted points and normal points are linked via blue lines. Most detected anomalies are located at the boundaries of the pattern, which are then be removed from the used SCADA datasets.

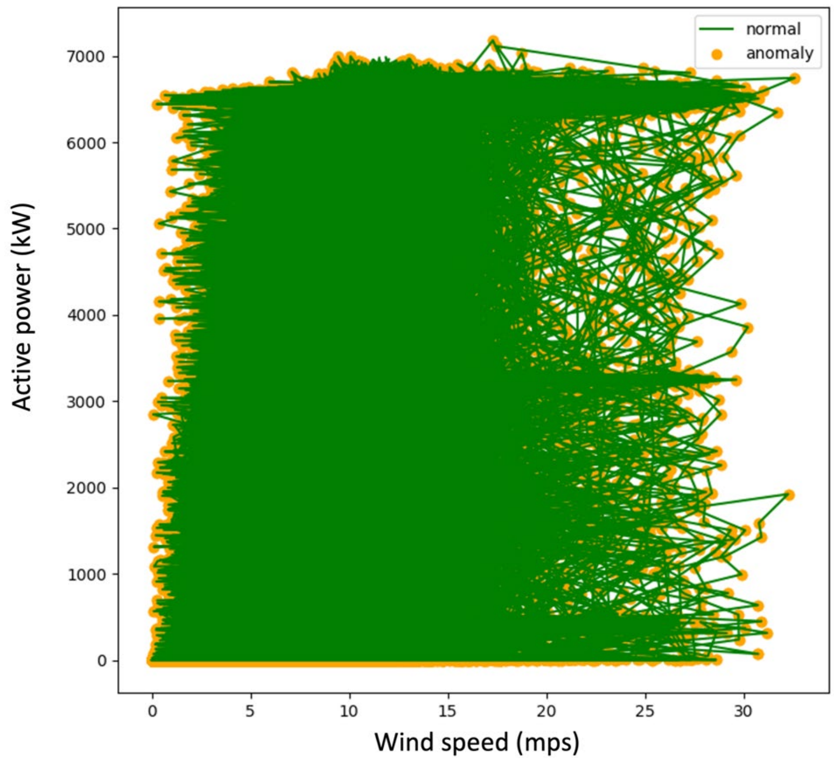

Fig. 5. Anomaly detection and removal by using IF, where the contamination ratio is set as 0.14.

3.4 Re-sampling

One challenge of using high-frequency SCADA datasets is the turbulence caused by the strong volatility of wind. A relatively small time interval leads to high computation costs and makes models sensitive. This effect can be addressed by averaging the sampled data over an appropriate period [25]. Usually, the sampling rate for short time prediction is 10 minutes, 15 minutes or 1 hour. In the current study, power data were resampled over 1-hour averaging period with mean values. After resampling, the power curve of hourly data is plotted in **Fig. 6**. Compared to curves of other contamination ratios and the power curve of the raw dataset (without IF process), the selected power data (14%) showed a smoother power curve, in which most outliers are cleaned successfully.

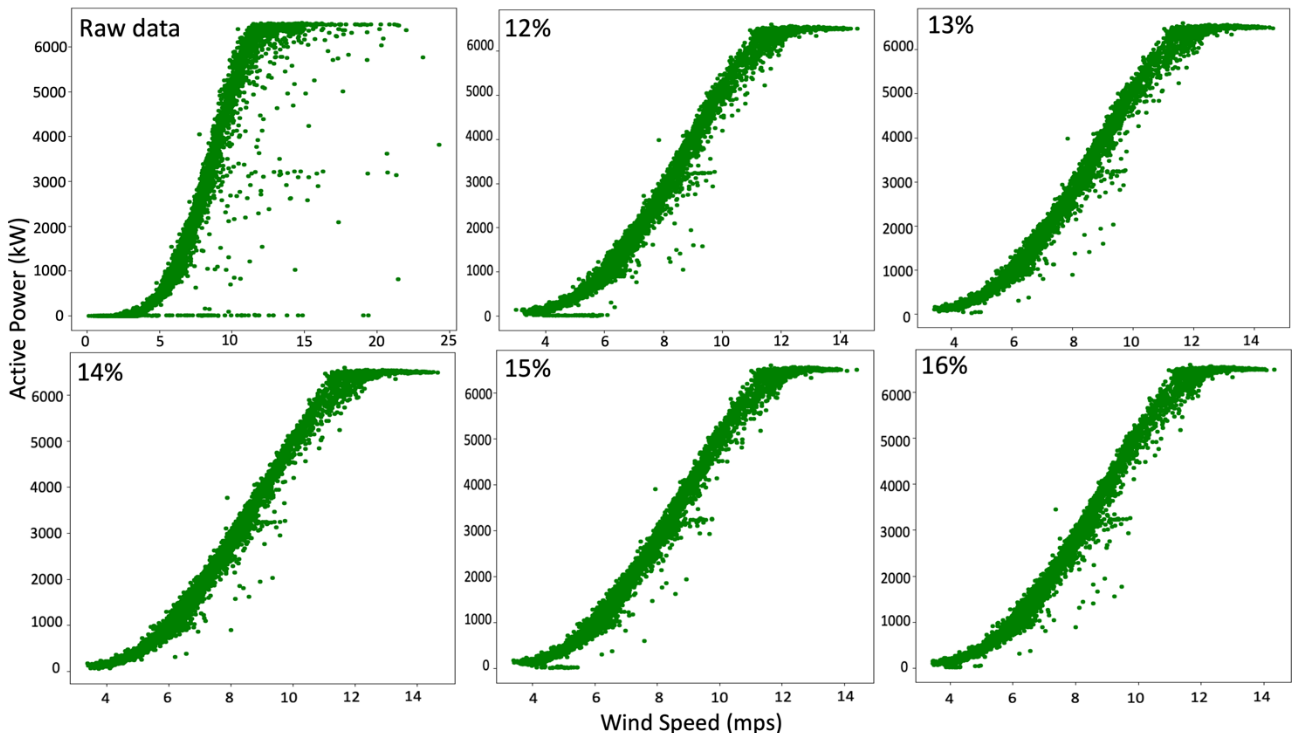

Fig. 6. Comparisons of power curves after using IF with different contamination ratios.

While observing the curve of selected data (see **Fig. 7**), its operation characteristics, such as cut-in speed (~3.5. m/s) and rated speed (~11.1 m/s), are consistent with the references (3.5 m/s and 10.9 m/s respectively). It further verified that using a contamination ratio of 14% is reasonable for the used SCADA datasets. Therefore, the hourly time series power data using IF at 14% contamination ratio is selected because it represents the ideal shape of the wind power curve, considering the proper cut-in, rated, and cut-off speeds.

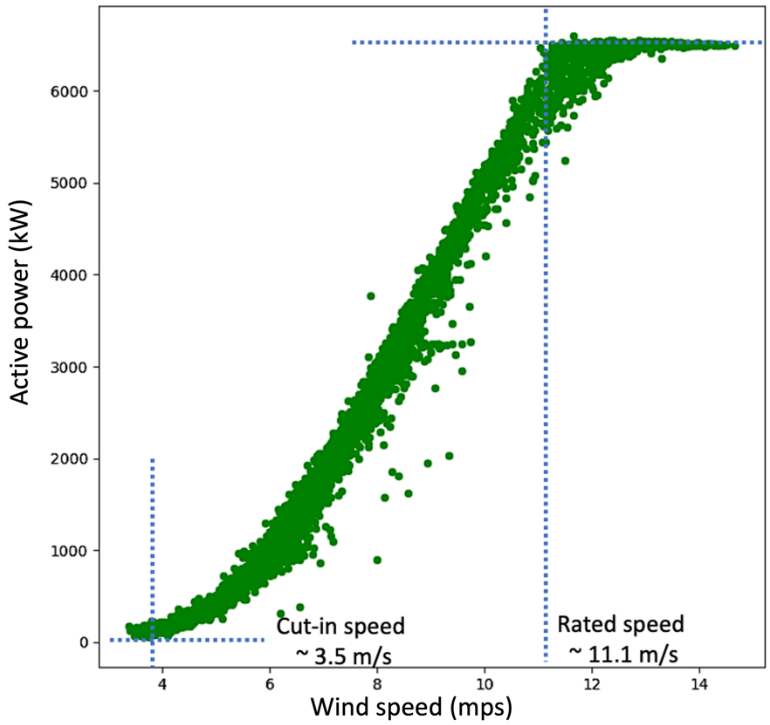

Fig.7. Wind power curve after resampling (1-hour sampling rate).

3.5 Interpolation

Missing values in a dataset cannot meet the requirement of prediction modelling. The problem of data discontinuity should be fixed before fitting time-series data into any models. In this study, missing points in the resampled time series data were first replaced with flags named 'not a number (NaN)'. Then the spline interpolation method in the 'interpolate' library was used to fill these NaN positions. In **Fig. 8**, one-month data (744 points) in January 2019 is used, considering the number of missing values in this month is smallest compared with the others. As shown in **Fig. 8**, the shape of the data after interpolating (**Fig. 8b**) is similar to that before interpolating (**Fig.8a**). It further verified the spline interpolation method can effectively complete the missing values in the used datasets.

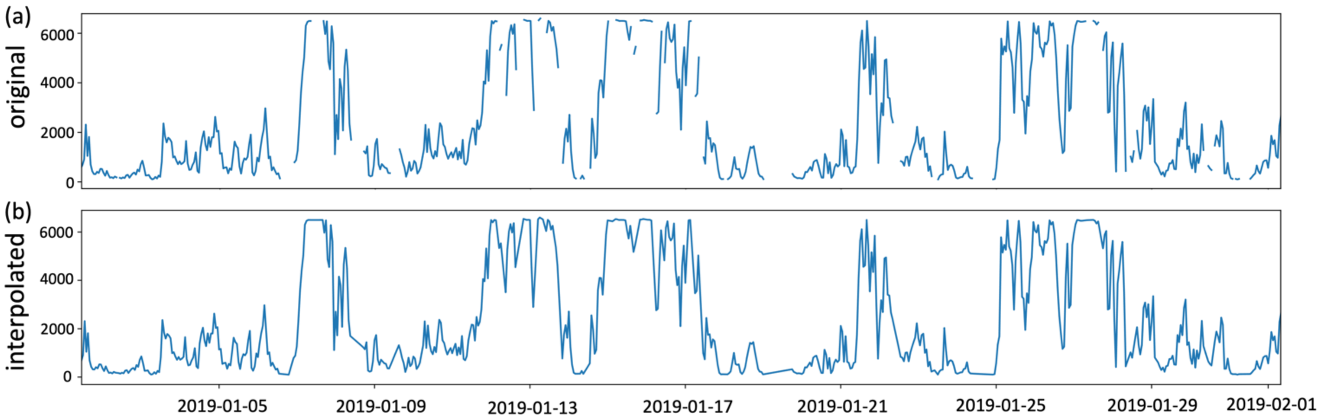

Fig. 8. Time series power (a) before and (b) after using spline interpolation.

4. Methodology

4.1 WT

Compared with the Fourier transform which is not suitable to analyse non-stationary signals [26], WT has the advantage of temporal resolution, which can analyse both time and frequency of signal simultaneously. Besides, WT has the flexibility in choosing mother wavelet types based on time series [27] while enhancing prediction accuracy.

WT can be categorized into two different types, including continuous WT (CWT) and DWT. CWT can capture all information in a given time series signal, but it is of high computational complexity and implementation difficulty [26]. DWT is more suitable to time series signals in practical applications as it samples wavelets discretely. Besides, DWT can reduce the computational complexity and bypass information redundancy caused by CWT. Therefore, DWT was selected to be used in this paper, which can be represented as **Eq. 2**:

$$W(m,n) = 2^{-\left(\frac{m}{2}\right)} \sum_{t=0}^{T} \psi\left(\frac{t - n \cdot 2^m}{2^m}\right) \cdot x(t) \tag{2}$$

where 't' is a discrete-time parameter, 'T' is the length of signal $x(t)$, variable m is the scaling parameter and variable n is the translation parameter.

Decomposed components are produced by downsampling and their length is reduced as the number of decomposition increases. Commonly, a reconstruction via inverse DWT (IDWT) [28] is applied before combining them to reproduce the original signal. The relationship of the original signal and n-level decomposed components contains approximation and details, which can be expressed in **Eq. 3**:

$$x(t) = a_n + d_n + d_{n-1} + \cdots + d_1 \quad (3)$$

4.2 SARIMA

SARIMA is an extension of ARIMA. Compared with ARIMA that cannot support seasonal data, SARIMA is sensitive to time series with seasonal components, considering seasonal features in data. Thus, it can be used for non-stationary datasets i.e., wind power, with improved prediction accuracy. The model can be represented as SARIMA (p, d, q) (P, D, Q)$_s$. 'AR' stands for autoregressive, where its order 'p' indicates the number of time series lags. 'I' stands for integrated. It is differencing time series instead of taking them directly, which makes the target variable more stationary and thus allows the model to support time series with a trend. Its order can be presented as 'd', which is the times to difference time series. 'MA' stands for moving average. It uses lagged prediction errors as inputs, push the model toward actual values and thus improve prediction accuracy, where its order can be represented as 'q'. 'P', 'D' and 'Q' have the same associations as 'p', 'd' and 'q' while they correspond with the seasonal components. 's' represents the seasonality length of data. For example, the time series $\{Z_t|1,2,\ldots,k\}$ can be presented by the SARIMA in **Eq. 4** [9]:

$$\phi_p(B)\varphi_P(B^s)(1-B)^d(1-B^s)^D Z_t = \theta_q(B)\Theta_Q(B^s)\varepsilon_t \quad (4)$$

where p, d, q, P, D, Q are order numbers, s is season length, B is the backward shift operator, $\phi_p(B)$ and $\varphi_P(B^s)$ are the regular and seasonal AR polynomials, $(1-B)^d$ and $(1-B^s)^D$ are the regular and seasonal I operators, $\theta_q(B)$ and $\Theta_Q(B^s)$ are the regular and seasonal MA polynomials, respectively, and ε_t is the estimated residual at time t.

In this study, both ARIMA and SARIMA models were developed by Python3 using the 'Statsmodels' library, where one-step ahead univariate prediction with 50 iterations was implemented on each model.

4.3 Deep-learning-based LSTM

LSTM is a type of ANN. As a variant of Recurrent Neural Network (RNN), LSTM addresses the issue of gradient disappearance/explosion in traditional neural networks [29]. Compared with conventional models that lack the memory function of historical information, LSTM has a unique structure based on memory cells. The capability of learning and remembering both short and long-term dependent information allows it to forecast time series. As shown in **Fig. 9**, an LSTM unit is composed of a forget gate, an output gate and an input gate.

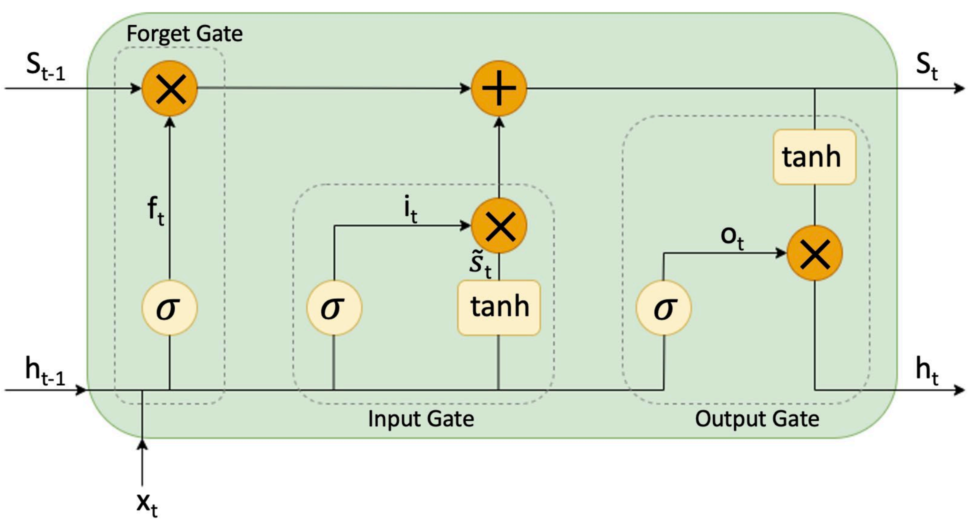

Fig. 9. Long short-term memory unit structure.

In LSTM, a recursive hidden layer includes various memory modules, where each of them has one or more self-connected memory units with three gates. The three gates (input gate i_t,, forget gate f_t and output gate o_t) can control information flow into/out of cells. The cell state (s_t) obtaining from previous state cell state (s_{t-1}) can remember previous values over arbitrary time intervals while $\tilde{s}_t$ is the newly assessed value of s_t.The formulations related to LSTM structure can be defined as follows (**Eq. 5** ~ **Eq. 10**) [30]:

$$f_t = \sigma\left(W_f[h_{t-1}, x_t] + b_f\right) \tag{5}$$

$$i_t = \sigma(W_i[h_{t-1}, x_t] + b_i) \tag{6}$$

$$o_t = \sigma(W_o[h_{t-1}, x_t] + b_o) \tag{7}$$

$$\tilde{s}_t = tanh(W_s[h_{t-1}, x_t] + b_s) \tag{8}$$

$$s_t = s_{t-1} \odot f_t + g_t \odot i_t \tag{9}$$

$$h_t = \tanh(s_t) \odot o_t \tag{10}$$

where $[h_{t-1}, x_t]$ is the input signal consisting of the input of the neuron x_t at time step t and the cell state vector h_{t-1} at time step t-1; h_t is the overall output at time step t; W_f, W_i, $W_o$and W_s are the corresponding weights connecting the input signal; b_f, b_i, b_o and b_s are bias along with corresponding activation function σ; tanh represents the hyperbolic tangent function and $\odot$ represents the element level multiplication.

In this study, TensorFlow was used as the platform for deep-learning-based LSTM development. The prediction is one-step univariate time series forecasting using walk-forward model validation with four-step input.

4.4 Integrated DWT-SARIMA-LSTM model

In this study, a novel hybrid model named DWT-SARIMA-LSTM is presented. The core idea of the proposed model is summarized as follows:

At the first step, DWT was applied to decompose wind power time series into approximation and detail. Then IDWT is used to reconstruct each component before developing prediction models, which can be represented as **Eq. 11**. Instead of using the whole time series directly, fitting approximation and detail into independent models can make signal analysis more effective, which is expected to improve model performance.

$$y_t = y_t^{app} + y_t^{det} \quad (11)$$

where y_t is the original time series power data; y_t^{app} is the reconstructed approximation; y_t^{det} is the reconstructed detail.

At the second step, unlike previous studies that assumed approximation is purely linear and detail is purely nonlinear [31], this study considers that each decomposed time series contains both linear and nonlinear components, which was represented in **Eq. 12** and **Eq. 13**.

$$y_t^{app} = L_t^{app} + N_t^{app} \quad (12)$$

$$y_t^{det} = L_t^{det} + N_t^{det} \quad (13)$$

In the third step, considering wind power generation highly relies on natural factors and has potential seasonality component, SARIMA models combined with LSTM models are developed. Firstly, SARIMA models are used to estimate and analyze both approximation and detail components. The linear components in both approximation ($\hat{L}_t^{app}$) and detail ($\hat{L}_t^{det}$) are assumed as prediction results from SARIMA models. Secondly, LSTM models are used to estimate and analyze the corresponding residuals after SARIMA models. The nonlinear components in both approximation ($\hat{N}_t^{app}$) and detail ($\hat{N}_t^{det}$) are assumed as prediction results from LSTM models. Then the predicted linear and nonlinear signal from approximation is combined to obtain the final prediction of approximation and the predicted linear and nonlinear signals from detail are combined to obtain the final prediction of detail. This step can be summarized as **Eq. 14** and **Eq. 15**:

$$\hat{y}_t^{app} = \hat{L}_t^{app} + \hat{N}_t^{app} \tag{14}$$

$$\hat{y}_t^{det} = \hat{L}_t^{det} + \hat{N}_t^{det} \tag{15}$$

Finally, the prediction is obtained by an additive combination of predicted approximation and predicted detail, which can be represented as **Eq. 16**:

$$\hat{y}_t = \hat{y}_t^{app} + \hat{y}_t^{det} \tag{16}$$

5. Results and discussions

5.1 WT parameter selection

For mother wavelet selection, considering the applied mother wavelet coefficients should have an easily physical interpretation and a fast computation [32], the most commonly used wavelet-Daubechies wavelet (db3) [14] was chosen in this study. For decomposition level selection, a formulation, which described the relationship between the signal length and the level number, was taken as a reference to determine the proper number of decomposition levels. The corresponding formulation is shown as **Eq .17** [33],

$$L = int(\log(N)) \tag{17}$$

where N is the length of the signal and L is the number of levels.

In the current study, the time series power data has a length of 744 points so that the optimal number of decomposition levels would be L=2. Besides, data at the decomposition level of L=1 and L=3 were also studied for investigation purposes. After DWT processing, each decomposed component would be reconstructed using IDWT individually. At first, the reconstruction accuracy is verified by comparing additive combinations of reconstructed and original signals. As presented in **Fig. 10**, the signals can be accurately reconstructed at level 1 (L=1) and level 2 (L=2) decomposition, while when the further decomposition (L=3) was carried out, distortion can be observed between the reconstructed signal and the original one. The forecast horizon in the plot is the sequence of data points (hours). It indicated that the reconstruction accuracy is limited at level 3, where this limitation may be related to the inherent properties of used SCADA datasets [34]. In summary, times series with decomposition level 1 and level 2 were investigated in this paper. **Fig. 11** illustrates the relationship between the decomposed components and the original signal, where the approximation of the previous level is the input of a higher decomposition level in terms of multiple decomposition levels (L=2).

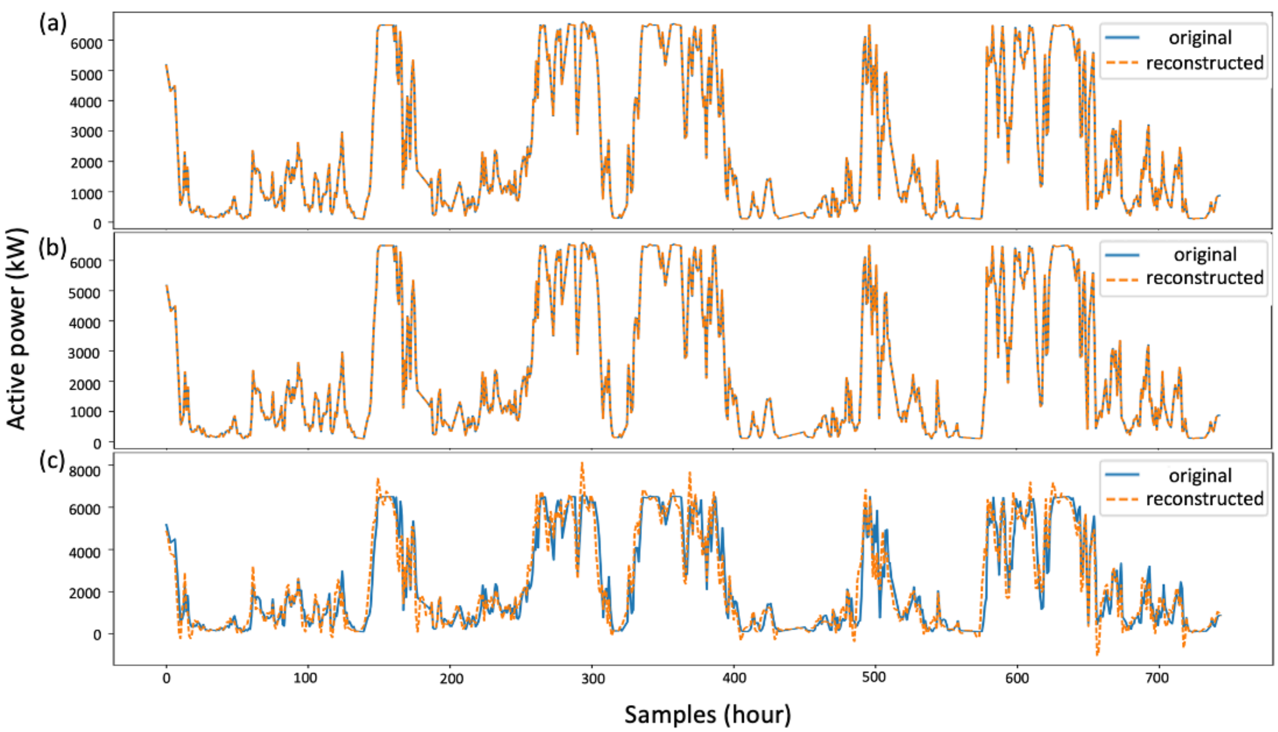

Fig. 10. Comparison among reconstructed times series power at different decomposition levels, including (a) level 1, (b) level 2 and (c) level 3.

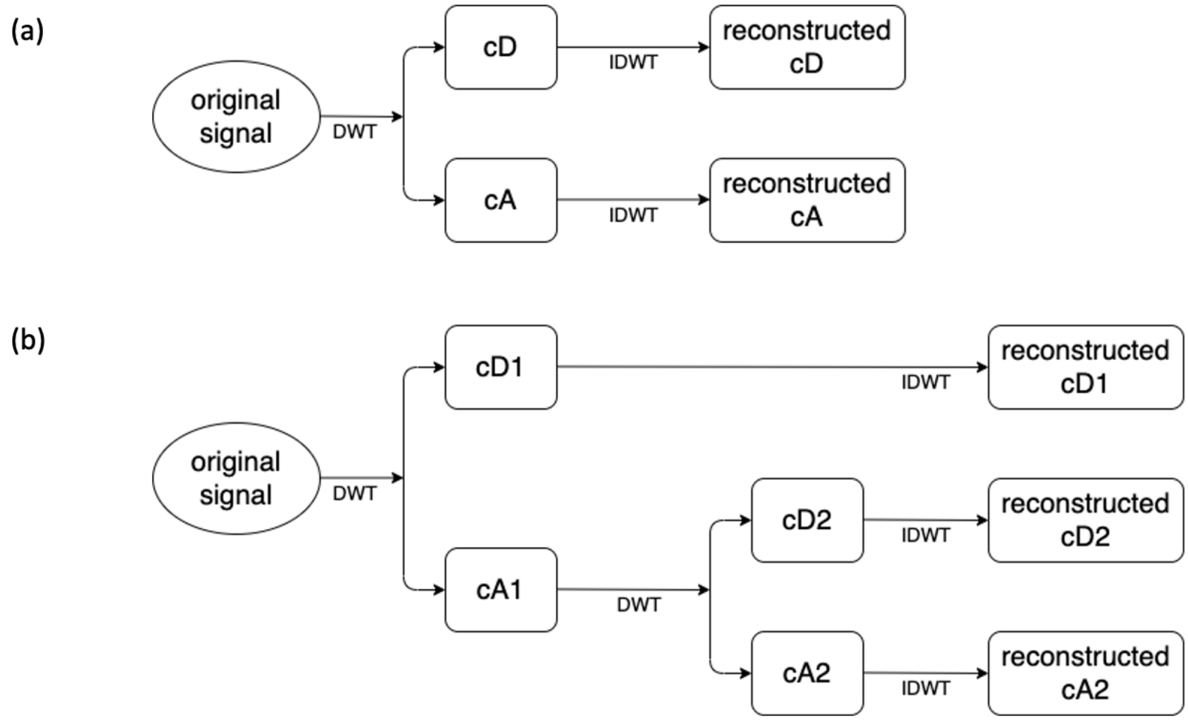

Fig. 11. Structure of decomposed wind power time series: (a) original signal is decomposed at level 1 to cD and cA components, and then be reconstructed separately by using IDWT; (b) original signal is decomposed at level 2 to cD1, cA2 and cD2 components, and then be reconstructed separately by using IDWT.

The time series of cD and cA at level 1 is presented in **Fig. 12a** while cD1, cD2 and cA2 at level 2 is shown in **Fig. 12b**. As cD1 represented the same time series as cD, their prediction models should be the same. Therefore, four components (cA, cD, cA2 and cD2) were taken as target signals in the following sections.

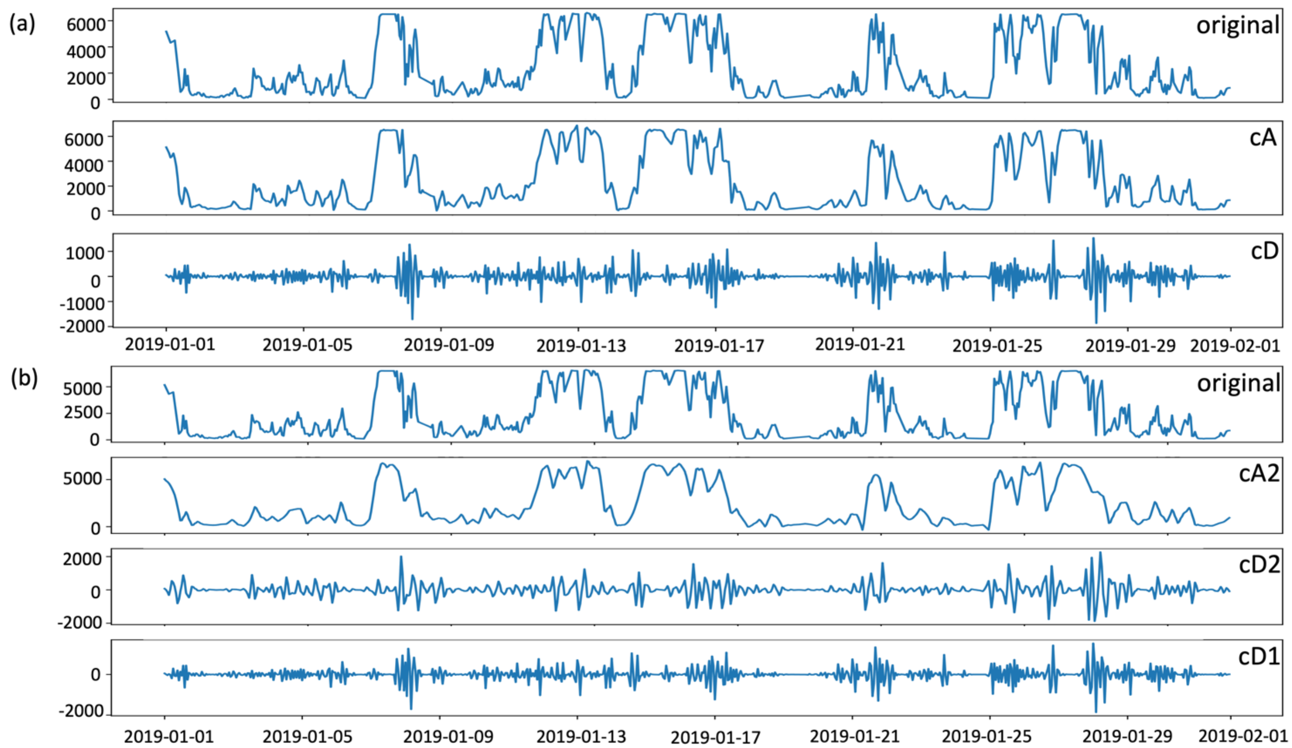

Fig. 12. The original time series power signal and its decomposed components; (a) Signal under level 1 decomposition is divided into approximation (cA) and detail (cD); (b) Signal under level 2 decomposition is divided into approximation (cA2), detail at level 2 (cD2) and detail at level 1 (cD1).

5.2 SARIMA

5.2.1 SARIMA model selection

In this study, Dickey-Fuller Test was used to analyse the stationarity of the time series of wind power at first, determining the order of differencing. Then, autocorrelation function (ACF) and partial autocorrelation function (PACF) are applied to make the first screening for AR and MA parameter selection for ARIMA/SARIMA models. As SARIMA models potentially have a large number of parameters as well as a combination of these terms, a range of models was investigated. The best-fitting model is selected based on the lowest value of Akaike's information criterion (AIC) and Bayesian information criterion (BIC) as well as suitable ACF and PACF of residuals.

5.2.1.1 AIC and BIC

AIC [35] and BIC [36] statistical criteria were employed in model selections. AIC is an estimator of the relative quality of statistical models and its value presents how well a model fits the given data considering the complexity of a model, which can be defined as **Eq. 18**:

$$AIC = -2\ln(L) + 2k \tag{18}$$

BIC is related to the sum of squared errors (SSE) from the estimated model, which can be defined as **Eq. 19**:

$$BIC = n\ln\left(\frac{SSE}{n}\right) + kln(n) \tag{19}$$

where n is the length of data, L is the maximized value of the maximum likelihood function and k is the number of parameters used in the model.

In this study, AIC and BIC are used to estimate these potential models, where the model with the lowest AIC and BIC value is preferred.

5.2.1.2 Dickey-Fuller Test

The Dickey-Fuller test [37] is a method to measure stationarity in the given time series. It is a statistical test, which determines how strongly the time series is defined by a trend. The null hypothesis is that time series with a unit root is non-stationary. If the p-value is smaller than 0.05 and the test statistic is much smaller than the critical value of 1%, we can reject the null hypothesis and assume the time series dataset is stationary. Lower p-values and more negative statistic values mean a higher degree of stationarity.

In this study, the stationarity of wind power data was conducted using the Dickey-Fuller test and the results are summarized in **Table 1**. The p-value of cA, cD cA2 and cD2 (0.006, 0, 0.013 and 0, respectively) are below the threshold of 0.05. The test statistic values of cD (-16.241) and cD2 (-12.552) are significantly less than the value of -3.439 at 1% while the test statistic value of cA (-3.573) is slightly lower than that at 1% and that of cA2 (-3.336) is only less than the value of -2.866 at 5%. Therefore, we assume the time series of cD and cD2 are stationary and the time series of cA and cA2 are non-stationary. The following experiment will set differencing orders of cD and cD2 as zero and consider both differencing orders of 0 and 1 for cA and cA2 to investigate model performance.

Table 1. Stationary check for decomposed components using Dickey-Fuller Test.

	cA	cD	cA2	cD2
P-value	0.006	0.000	0.013	0.000
Test statistic	-3.573	-16.241	-3.336	-12.552
Critical value 1%	-3.439	-3.439	-3.439	-3.439
Critical value 5%	-2.866	-2.866	-2.866	-2.866
Critical value 10%	-2.569	-2.569	-2.569	-2.569

5.2.1.3 ACF and PACF

In this study, both ACF and PACF of each decomposed component were analyzed to select possible SARIMA models. The seasonal parameter ‘s’ is selected based on knowledge of the problem, setting 24 as the initial parameter because there are 24 hours in one day and adjusting the order according to the previously possible model based on ACF and PACF plots. Potential values of p and q were estimated by looking at the correlations of recent time steps. Potential values of P and Q are estimated using a similar way as above while considering seasonality by looking at the correlations at seasonal lag time steps. Generally, increase AR order if the first several lags in both ACF and PACF are positive while increasing MA order if the first several lags in both plots are negative. After trials and errors, the possible combination of models for cA, cD, cA2 and cD2 with corresponding AIC and BIC values are summarized in **Table 2**. Based on the error criteria of AIC and BIC values, SARIMA(2,1,1)$(1,0,0)_3$, SARIMA(2,0,2)$(1,0,2)_{24}$, SARIMA(1,0,1)(1,1,2)$_{24}$ and SARIMA(1,0,5)(1,0,0)$_{12}$ are selected for cA, cD, cA2 and cD2 component, respectively (optimal models are bolded in **Table 2**).

Table 2. Characteristics for possible ARIMA/SARIMA models with AIC and BIC values.

Signal	Model parameters	AIC	BIC
cA	ARIMA (2,0,1)	10905.115	10923.552
	ARIMA (2,1,1)	10829.585	10848.017
	SARIMA (2,1,1) (1,0,0)3	**10748.843**	**10771.863**
	SARIMA (1,1,0) (0,0,1)24	11013.362	11027.091
cD	ARIMA (1,0,2)	9495.352	9513.784
	SARIMA (1,0,2) (1,0,1)24	9330.056	9357.507
	SARIMA (2,0,2) (1,0,1)24	9092.683	9124.709
	SARIMA (2,0,2) (1,0,2)24	**8802.442**	**8838.770**
cA2	ARIMA (1,0,0)	10882.245	10891.467
	SARIMA (1,0,1) (1,1,0)24	9894.654	9912.830
	SARIMA (1,0,1) (1,1,2)24	**9289.332**	**9316.375**
	SARIMA (1,0,1) (2,0,2)24	9528.168	9559.965
cD2	ARIMA (0,0,2)	10319.402	10333.226
	SARIMA (1,0,4) (1,0,0)24	9368.341	9400.386
	SARIMA (1,0,4) (1,0,0)12	9381.281	9413.442
	SARIMA (1,0,5) (1,0,0)12	**9190.597**	**9231.946**

5.2.2 SARIMA model diagnostic

The goodness-of-fit test was conducted on residuals from the selected models, as shown in **Fig. 13**. This step considered standardized residual, correlogram, histogram with an estimated density of standardized residual (KDE curve) and a reference curve of normal (0,1) density and normal Q-Q plot, where the blue dots are residuals of ordered distribution and a reference line of normal (0,1) distribution.

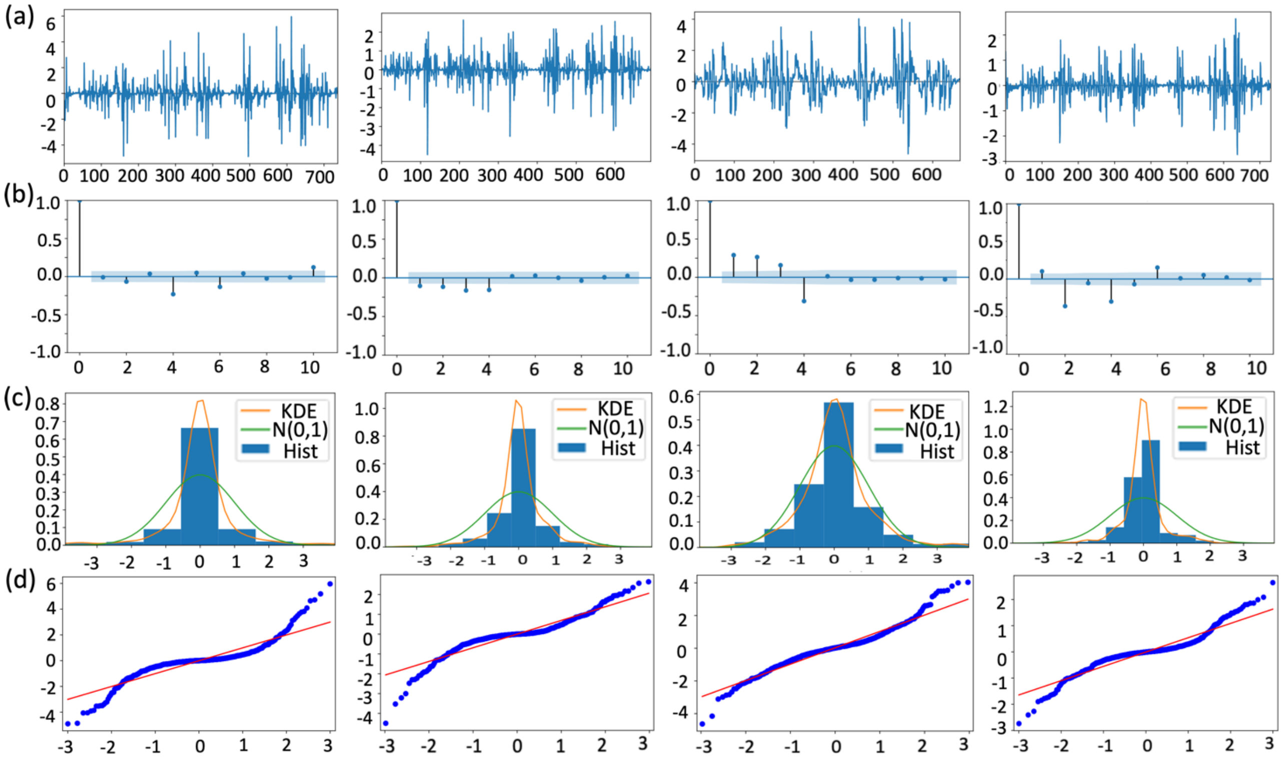

Fig. 13. Goodness-of-fit test including (a) standardized residual, (b) correlogram, (c) histogram with an estimated density of standardized residual and a reference curve of normal (0,1) density and (d) normal Q-Q plot of selected models of cA component, cD component, cA2 component and cD2 component (from left to right respectively).

As for standardized residual (**Fig.13a**), the mean of cA, cD, cA2 and cD2 are about zero while there are some obvious patterns. This can be reflected on correlogram plots (**Fig.13b**) in which there are some correlations for lags that are outside the confidential levels. Their KDE curves (**Fig.13c**) are similar to the normal distribution, indicating the residuals are normally distributed. But some data points deviated away from the straight line in the normal Q-Q plot (**Fig.13d**), especially for cA and cD2 components. Therefore, we could conduct that these residuals are not purely white noise. There is still some useful information left in residuals that cannot be extracted from their corresponding SARIMA models. This assessment of models is reasonable because the time series signals used in this study are collected from the real world, where nonlinear information exists in both approximation and details.

5.2.3 SARIMA model evaluation

As shown in **Fig. 14**, the green line represents the original time series and the red line represents the predicted values from the SARIMA model. For prediction results at level 1 decomposition, the forecasting accuracies are 96.72% for cA and 88.16% for cD, where the accuracy for cD is lower about 9% than that for cA. This indicated that there is more nonlinear information in cD than in cA because cD is the high-frequency component. As for the prediction result of level 2 decomposition, the accuracy of cA2 and cD2 is 98.77% and 94.26% respectively. Accordingly, the accuracy of high-frequency component cD2 is lower than that of cA2, which is similar to what occurred in level 1, while the accuracy difference is not larger than that in level 1. We suppose that this is because cA2 and cD2 are derived from the same component of cA. The difference between the portion of nonlinear information in cA2 and cD2 is smaller than that in cA and cD. Thus, the difference in prediction accuracy at level 2 is smaller. The prediction performance is summarized in **Table 3**.

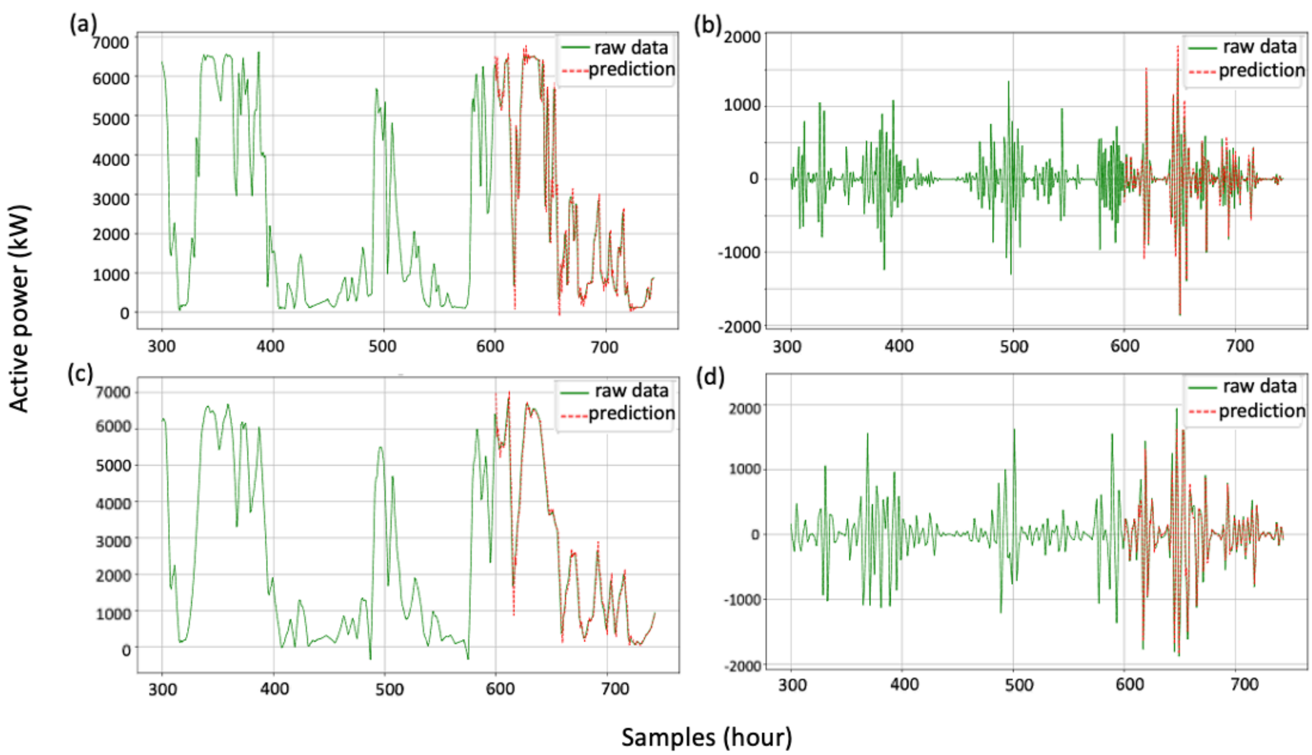

Fig. 14. Prediction results for (a) cA component, (b) cD component, (c) cA2 component and (d) cD2 component.

Further investigations on prediction performance at different decomposition levels were also conducted, considering that the combination of cA2 and cD2 equals cA. The zoom-in graph **Fig. 15** compares prediction results between cA and combined cA. After summing cA2 and cD2, the accuracy of combined cA achieved 98.74% with an increase of about 2% compared with cA (96.72%), indicating the further decomposition of cA can make cA2 and cD2 more stationary. The final prediction accuracy at level 1 decomposition reached 96.17% and was increased to 98.51% with level 2 decomposition, which verified the advantage of using DWT before fitting data into the model.

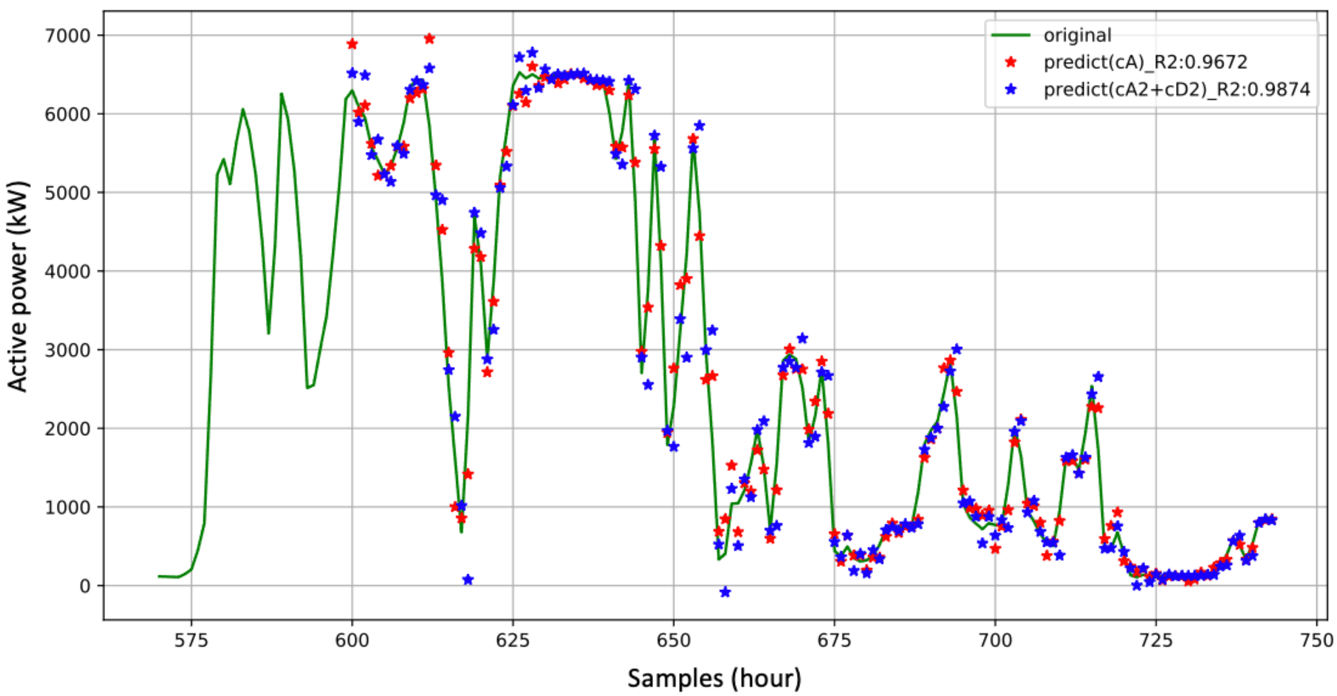

Fig. 15. Comparisons of prediction accuracy between cA and combined cA using cA2 and cD2.

Table 3. Prediction accuracy of decomposed components.

Components	cA	cD	cA2	cD2	cA2 + cD2
R2	0.967	0.881	0.987	0.942	0.987

5.3 LSTM

This section aims to use a deep-learning-based LSTM to dig out the remained useful information that cannot be extracted by SARIMA models. Because it has been proved in the previous section that prediction accuracy at level 2 decomposition is higher than that at level 1, the following experiment is focused on analysing data at level 2 decomposition. The three residuals of cD, cA2 and cD2 from their corresponding SARIMA are used in the following session.

5.3.1 LSTM model configuration

5.3.1.1 Normalization

Because LSTM models are sensitive to the scale of input data, normalization was implemented before fitting data into models. The normalized predicted values are then denormalized by using inverse transformation to obtain forecasting results. In this paper, time series were rescaled to the range of 0~1. The corresponding formulation can be represented as **Eq. 20:**

$$X_{scaled} = \frac{x_i - \min(x)}{\max(x) - \min(x)} \tag{20}$$

where x_i is the original value, X_{scaled} is the normalized value, $\max(x)$ and $\min\,(x)$ are the maximum and minimum values, respectively.

5.3.1.2 Batch size and number of epochs

Batch size and number of epochs are two hyperparameters that have a significant effect on overall computation cost and performance for forecasting models. Batch size is the number of samples that are processed before the weights are updated while the number of epochs is the iteration times that are completed through the training dataset. At each epoch, the model randomly samples series from the set that is defined by the batch size. Usually, the number of epochs is about hundreds or thousands. A sufficient number of epochs can minimize model errors. In this study, the number of epochs was initially set as 1000 for each model.

5.3.1.3 Activation function

Activation functions could manipulate and propagate the summed weights through gradient processing in neural networks, which are important for training and optimizing. Nonlinear activation functions, such as sigmoid and hyperbolic tangent (tanh), allow neural networks to learn data with complex structures. But they are not suitable to be used in deep learning neural networks that have multiple layers because of the vanishing gradient problem. This problem can be addressed by using rectified linear activation functions based on stochastic gradient descent with backpropagation of errors. Among them, Rectified Linear Unit (ReLU) is one of the most commonly used activation functions. It is a piecewise linear function but allows the model to account for non-linearities. It outputs zero if receiving negative input while returns any positive input back, where formulations used for the fully connected layer can be represented as **Eq. 21** and **Eq. 22**:

$$H_i = \sum_{j=1}^{m} x_i w_{ij} + b_j \tag{21}$$

$$h = ReLU(H_i) \tag{22}$$

where H_i is the net input of neuron j in the deeper hidden layer; h is the output of neuron j; x_i and b_j is the input and a bias for neuron j, respectively; w_{ij} is a weight that linked neuron i and neuron j.

As the ReLU activation function is stress-free to train and can learn complex relationships in data, it was selected in this study.

5.3.1.4 Optimizer

Optimizers iteratively update weight parameters in neural networks and can minimize the loss function. Using proper optimization algorithms can lower the expense of the training process in deep learning. In this study, the commonly used Adaptive Moment Estimation (Adam) [38] was selected, which is an extension to the stochastic gradient descent algorithm. It realized the advantages of both adaptive gradient algorithm (AdaGrad) [39] and Root Mean Square Propagation (RMSProp) [40].

5.3.2 LSTM model selection

In this study, hyperparameters of LSTM were selected based on the lowest Mean Squared Error (MSE) value (see **Eq. 23**). It computes the average of the squared differences between actual values and predicted values. To improve the accuracy of the model, the loss value is expected to be reduced as small as possible. Compared with RMSE, the squaring can punish the model for making big mistakes.

$$MSE = \frac{\sum_{i=1}^{n}[\hat{y}_i - y_i]^2}{n} \tag{23}$$

where $\hat{y}_i$ is the predicted value and y_i is the real value in the given dataset.

Considering that random initial conditions for LSTM neural network can bring different results at each time, each experimental scenario for hyperparameter selection was run 10 times. In this paper, various deep learning structures were tested, and all LSTM neural networks are hyperparameter tuned through manual search. After trial and error, a five-layer deep learning LSTM (neuron number of 20, 50, 50, 20 and 1 in each layer) was selected for cD; a five-layer deep learning LSTM (neuron numbers of 20, 50, 50, 20, 1 in each layer) was selected for cD2; a four-layer deep learning LSTM (neuron numbers of 10, 20, 5, 1 in each layer) was selected for cA2. The activation function for each layer was set as ReLU. Optimizer is set as Adam for all three models, where the learning rate for each model is 0.01. Details on each model configuration are summarized in **Table 4**.

Table 4. LSTM model configuration for each decomposed component.

Signal	Network structure	Epochs	Batch size	Activation function	Optimizer
cD	(20,50,50,20,1)	1000	2	ReLU	Adam
cD2	(20,50,50,20,1)	800	2	ReLU	Adam
cA2	(10,20,5,1)	800	1	ReLU	Adam

5.3.3 LSTM model evaluation

The prediction results are shown in **Fig. 16**. The prediction accuracy of cD residual, cA2 residual and cD2 residual is 94.61%, 66.90% and 63.77% respectively, which reflects the model capability of extracting remaining information in residuals from SARIMA models. The highest accuracy of about 95% was achieved at cD residual while the accuracy of cA2 (~67%) and cD2 (~64%) are relatively lower, this can look back to their corresponding SARIMA models. Because higher accuracy of SARIMA model means less useful information left in residuals, where the prediction accuracy of cD is about 88% while higher accuracy achieves for cA2 (~99%) and cD2 (~94%). The prediction performance is summarized in **Table 5**.

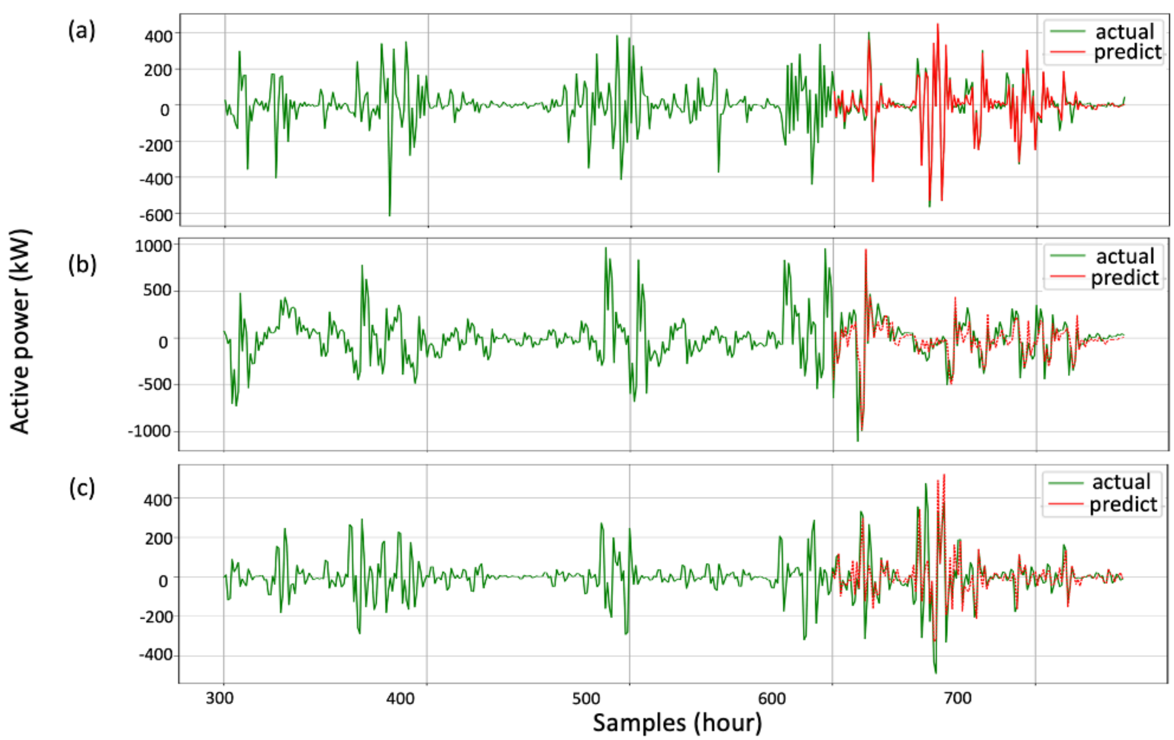

Fig. 16. Comparisons of LSTM model prediction accuracy of residuals, including (a) cD residuals, (b) cA2 residuals and (c) cD2 residuals.

Table 5. Prediction accuracy of LSTM models.

Components	cD residual	cA2 residual	cD2 residual
R2	0.946	0.669	0.638

5.4 Hybrid model prediction evaluation

This section investigated the performance of the proposed hybrid prediction model. For each decomposed component, the prediction was obtained by an additive combination of forecasting from SARIMA and the corresponding residual forecasting from LSTM. The prediction accuracy for cD, cA2 and cD2 was achieved at 99.35%, 99.59% and 97.92%, respectively. It indicated that with the assistant of LSTM modelling, prediction accuracy is enhanced with an increase of 11.19% for cD, 0.82% for cA2 and 3.66% for cD2. It can be seen that major improvements were achieved in detail (both cD and cD2), which should contain more high frequency/nonlinear information. The prediction performance is summarized in **Table 6**.

Table 6. Prediction accuracy of decomposed components based on the proposed hybrid model.

Components	cD	cA2	cD2
R2	0.993	0.996	0.979

The final prediction is obtained by the additive combination of approximation and detail. As shown in **Fig. 17**, the blue start marker represents the prediction power at level 1 decomposition after SARIMA, the orange plus marker represents the prediction power at level 2 decomposition after SARIMA and the red x marker represents the completed hybrid model. The performance of SARIMA is enhanced by increasing the decomposition level from level 1 to level 2, where the prediction accuracy is 96.17% and 98.51%, respectively. After using LSTM models to dig out information in residuals, the prediction accuracy is up to 99.46%. It shows a further increase of 0.94% compared with that at the same decomposition level without LSTM modelling.

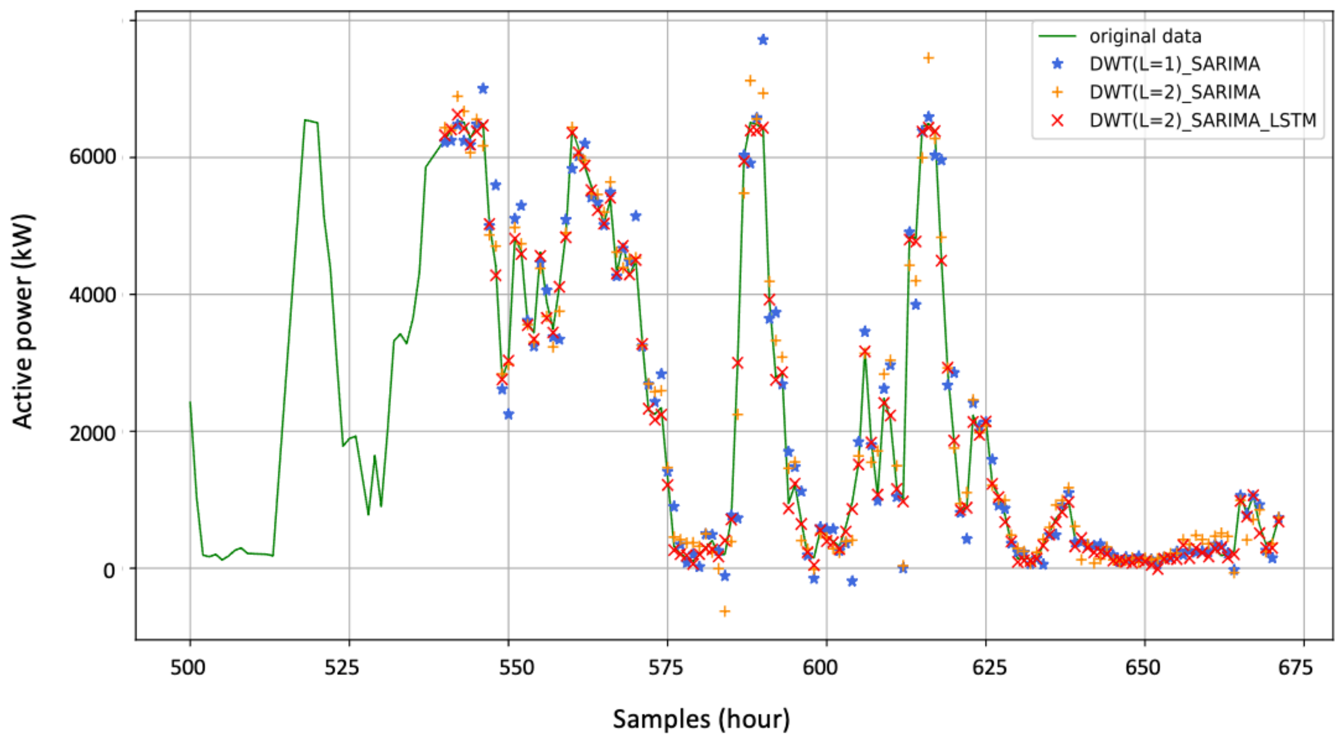

Fig. 17. Comparison of prediction accuracy at different stages during model developments.

5.5 Mode benchmarking

It is essential to build a baseline to time series prediction problem because it can provide a point of comparison. Generally, the baseline prediction should be simple, fast, and repeatable, therefore the naïve model-persistence algorithm is applied for benchmark testing. The dataset used in the benchmark model is pre-processed one considering that there are some missing points in the original time series. The accuracy using the naïve model achieves 84.4%, which has a lower of 15.1% than that using the proposed hybrid model (99.5%). The prediction performance is summarized in **Table 7**.

Table 7. Prediction accuracy of SARIMA model at level1/2 and that of the proposed hybrid model and the naïve model.

	L=1	L=2	Naïve model	Hybrid model
R2	0.962	0.985	0.844	0.995

5.6 Hybrid model evaluation under different weather conditions

To prove the sufficient integrity of the proposed hybrid model, a dataset of different weather conditions is considered. Because the dataset of January 2019 can be considered as in winter, another dataset (April 2019) is chosen as in Spring. The time series is from 04/01 to 04/28. Using the train-test split percentage of 0.8-0.2, the used dataset (672 points) is split into two parts: a training set (540 points) and a testing set (132 points). This time series is pre-processed using the same method as above. Applying the same model building process, we present the model selection parameters and corresponding prediction results as follows.

As for SARIMA model selection, SARIMA(2,1,2)$(0,0,2)_3$, SARIMA(2,0,3)$(1,0,0)_{24}$, SARIMA(1,0,1)$(1,0,2)_{24}$ and SARIMA(1,0,5)$(1,0,0)_{12}$ are selected for cA, cD, cA2 and cD2 component, respectively. Their corresponding AIC and BIC value are 9699.120&9730.586, 7850.897&7882.193, 8754.836&8781.434, and 8438.105&8474.031. For prediction results at level 1 decomposition, the forecasting accuracies are 96.75% for cA and 88.91% for cD. For the prediction results of level 2 decomposition, the accuracy of cA2 and cD2 is 97.95% and 94.58% respectively. The combined model from level 1 decomposition (cD+cA) shows an accuracy of 96.49% and the accuracy of the combined model from level 2 decomposition (cD+cA2+cD2) achieved 98.32%. This indicates an increase of accuracy (~2%) by using level 2 decomposition.

As for LSTM model selection, a four-layer deep learning LSTM (neuron number of 20, 50, 15 and 1 in each layer) was selected for cD; a five-layer deep learning LSTM (neuron numbers of 15, 50, 50, 15, 1 in each layer) was selected for cD2; a four-layer deep learning LSTM (neuron numbers of 10, 20, 50, 1 in each layer) was selected for cA2. The activation function and Optimizer for all three models are set as the same as in previous cases. The prediction accuracy of cD residual, cA2 residual and cD2 residual is 95.13%, 96.95% and 96.85% respectively.

The prediction of each decomposed component was obtained by an additive combination of forecasting from SARIMA and the corresponding residual forecasting from LSTM, like in previous cases. The prediction accuracy for cD, cA2 and cD2 was achieved at 99.46%, 99.94% and 99.83%, respectively. It shows that with the assistant of LSTM modelling, prediction accuracy is enhanced with an increase of 10.55% for cD, 1.99% for cA2 and 5.25% for cD2.

The final prediction is shown in **Fig. 18**. The accuracy of the SARIMA model at level 2 decomposition (98.32%) is higher than that at level 1 decomposition (96.49%). With the assistant of LSTM models, the prediction accuracy is up to 99.92%, which indicates a further increase of 1.6%. Compared with the accuracy using the naïve model counterpart (86.7%), there is an increase of 13.2%. The prediction performance of using time series in other weather conditions (**Table 8**) further proves the integrity of the proposed hybrid model.

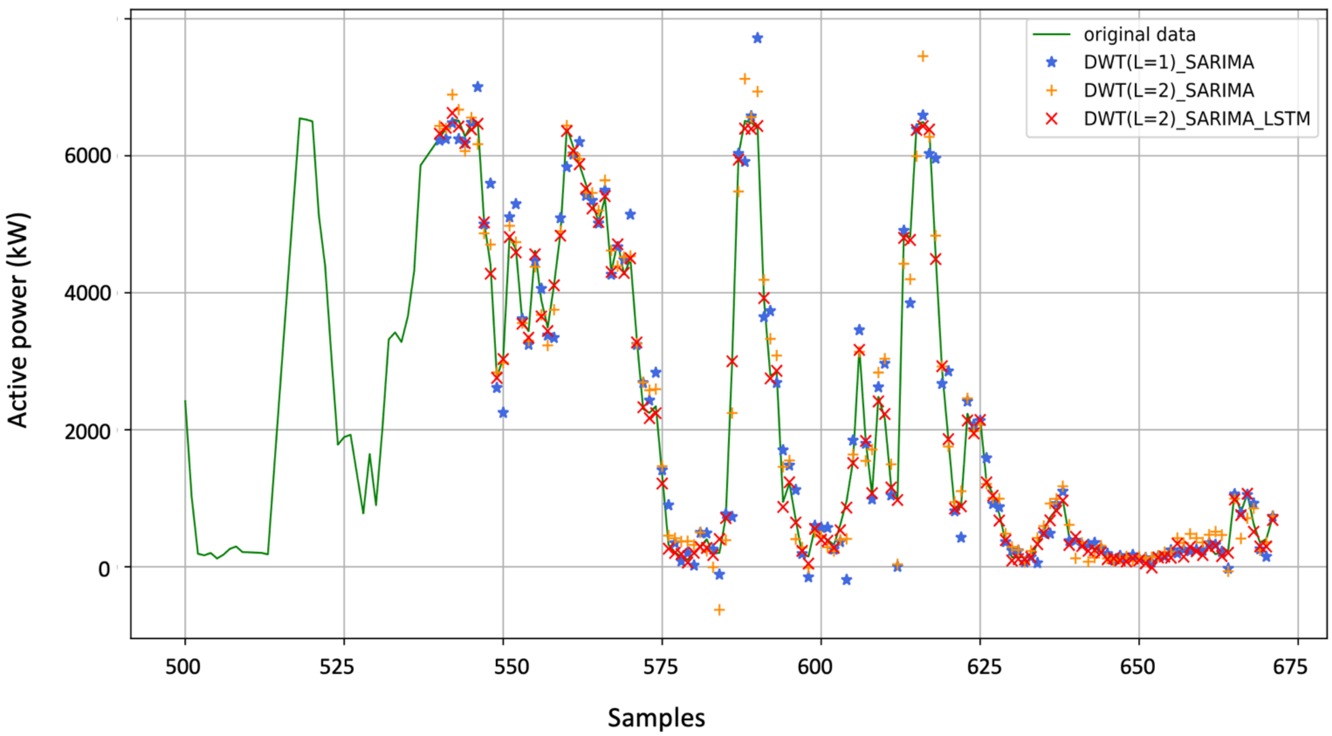

Figure 18. Comparison of prediction accuracy at different stages during model developments using the dataset for another weather condition.

Table 8. Prediction accuracy of SARIMA model at level1/2 and that of the proposed hybrid model, and the naïve model under the different weather condition

	L=1	L=2	Naïve model	Hybrid model
R2	0.965	0.983	0.867	0.999

6. Conclusions

This paper presented a novel hybrid model to predict wind power for a 7 MW offshore wind turbine in Scotland. The used datasets were collected from a high-frequency SCADA database with a 1-s sampling rate. To sum up, the following conclusions have been reached:

- In this study, data pre-processing is applied to clean the used datasets before analysing data with a prediction model. Removal of obvious outliers and anomalies from the SCADA database by using IF can improve prediction accuracy by removing abnormal points from normal points. Resampling of 1-s samples to hourly samples can mitigate the influence of turbulence. The implementation of spline interpolation can mitigate the effect of missing values, contributing to a continuous dataset and thus enhancing prediction accuracy, especially for SARIMA models with the characteristic of periodicity. This mixed pre-processing method significantly improved the quality of the used dataset.
- DWT and IDWT were used to decompose and reconstruct power signals, respectively. A proper decomposition of signals into several sub-series enables data to be more stationary and thus make further analysis with prediction models easier. The prediction accuracy of the SARIMA model is increased from 96.17% at level 1 decomposition to 98.51% at level 2 decomposition.
- Without assuming approximation is purely linear signal or detail is a purely nonlinear signal, both decomposed components are treated into linear and nonlinear models. SARIMA is used as the linear model, which can support seasonal components in time series power. LSTM with a deep learning neural network is used as a nonlinear model to dig out information in residuals from SARIMA. Prediction accuracy at decomposition level 2 is 98.51% for the SARIMA model and is enhanced to 99.46% for the proposed hybrid model.
- To further prove the integrity of the proposed hybrid wind power prediction model, data for another weather condition is considered. Compared to power prediction results in winter, the

results in spring also shows high prediction accuracy. The accuracy of the hybrid model has an increase of 13.2%, compared to that of using the naïve model (86.7%).

The limitations and possible future improvements for this study are discussed as follows:

- Because the used signal is collected from real-world equipment, it is unavoidable to obtain a dataset with missing values. Although spline interpolation is used to mitigate this problem, the portion of missing value about 24% in raw data is relatively high, which may lead models to deviate from the actual scenario. Second, this study investigates the wavelet transform with db3. There are various types of wavelets such as other Daubechies wavelets i.e., db2, db4, or db7, and other types i.e., harr wavelet, coiflet wavelet, which can be used in time series prediction. One paper has proposed to mitigate the problem of selecting the proper wavelet by taking the average of several wavelets [13]. This can be a solution, but it is still interesting to investigate the effect of using different wavelets on prediction models, which can be considered as one direction for future improvement for this proposed hybrid model.
- Strong gust is an important factor affecting the performance of prediction models. Winds are least gusty offshore because of the large water surfaces while most gusty onshore is due to the rough land and near high constructions [41]. Therefore, we do not consider the factor of strong gust in this study while it will be discussed in an onshore study in future.
- The dataset used in this study is collected in Scotland. In future, more datasets in different sites, such as in other countries, will be considered. This novel idea of building the hybrid model has the potential to advance wind power prediction models worldwide.

Acknowledgement

The authors thank the Offshore Renewable Energy (ORE) Catapult for provisions of the SCADA database.

References

[1] Wind Power to dominate power sector growth | Global Wind Energy Council, (n.d.). https://gwec.net/wind-power-to-dominate-power-sector-growth/ (accessed April 19, 2021).

[2] Z. Lin, X. Liu, Wind power forecasting of an offshore wind turbine based on high-frequency SCADA data and deep learning neural network, Energy. 201 (2020) 117693. https://doi.org/10.1016/j.energy.2020.117693.

[3] Z. Lin, X. Liu, S. Lotfian, Impacts of water depth increase on offshore floating wind turbine dynamics, Ocean Eng. 224 (2021) 108697. https://doi.org/10.1016/j.oceaneng.2021.108697.

[4] Global Offshore Wind Report 2020 | Global Wind Energy Council, (n.d.). https://gwec.net/global-offshore-wind-report-2020/ (accessed May 5, 2021).

[5] Queen's Speech December 2019 - GOV.UK, (n.d.). https://www.gov.uk/government/speeches/queens-speech-december-2019 (accessed April 19, 2021).

[6] T. Ackermann, L. Söder, Wind energy technology and current status: a review, Renew. Sustain. Energy Rev. (2000). https://doi.org/10.1016/S1364-0321(00)00004-6.

[7] M.Q. Raza, A. Khosravi, A review on artificial intelligence based load demand forecasting techniques for smart grid and buildings, Renew. Sustain. Energy Rev. (2015). https://doi.org/10.1016/j.rser.2015.04.065.

[8] H. Liu, H.Q. Tian, X.F. Liang, Y.F. Li, Wind speed forecasting approach using secondary decomposition algorithm and Elman neural networks, Appl. Energy. (2015). https://doi.org/10.1016/j.apenergy.2015.08.014.

[9] M. Geurts, G.E.P. Box, G.M. Jenkins, Time Series Analysis: Forecasting and Control, J. Mark. Res. (1977). https://doi.org/10.2307/3150485.

[10] E. Yatiyana, S. Rajakaruna, A. Ghosh, Wind speed and direction forecasting for wind power generation using ARIMA model, 2017 Australas. Univ. Power Eng. Conf. AUPEC 2017. 2017-Novem (2018) 1–6. https://doi.org/10.1109/AUPEC.2017.8282494.

[11] J. Zhang, X. Jiang, X. Chen, X. Li, D. Guo, L. Cui, Wind power generation prediction based on LSTM, ACM Int. Conf. Proceeding Ser. (2019) 85–89. https://doi.org/10.1145/3325730.3325735.

[12] Z. Qian, Y. Pei, H. Zareipour, N. Chen, A review and discussion of decomposition-based hybrid models for wind energy forecasting applications, Appl. Energy. (2019). https://doi.org/10.1016/j.apenergy.2018.10.080.

[13] I. Khandelwal, R. Adhikari, G. Verma, Time series forecasting using hybrid arima and ann models based on DWT Decomposition, Procedia Comput. Sci. 48 (2015) 173–179. https://doi.org/10.1016/j.procs.2015.04.167.

[14] M. Shafaei, J. Adamowski, A. Fakheri-Fard, Y. Dinpashoh, K. Adamowski, A wavelet-SARIMA-ANN hybrid model for precipitation forecasting, J. Water L. Dev. 28 (2016) 27–36. https://doi.org/10.1515/jwld-2016-0003.

[15] W. Pannakkong, V.N. Huynh, A hybrid model of ARIMA and ANN with discrete wavelet transform for time series forecasting, in: Lect. Notes Comput. Sci. (Including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics), Springer Verlag, 2017: pp. 159–169. https://doi.org/10.1007/978-3-319-67422-3_14.

[16] A. Kisvari, Z. Lin, X. Liu, Wind power forecasting – A data-driven method along with gated recurrent neural network, Renew. Energy. 163 (2021) 1895–1909. https://doi.org/10.1016/j.renene.2020.10.119.

[17] L. Ziegler, E. Gonzalez, T. Rubert, U. Smolka, J.J. Melero, Lifetime extension of onshore wind turbines: A review covering Germany, Spain, Denmark, and the UK, Renew. Sustain. Energy Rev. (2018). https://doi.org/10.1016/j.rser.2017.09.100.

[18] J. Serret, C. Rodriguez, T. Tezdogan, T. Stratford, P. Thies, Code comparison of a NREL-fast model of the levenmouth wind turbine with the GH bladed commissioning results, in: Proc. Int. Conf. Offshore Mech. Arct. Eng. - OMAE, 2018. https://doi.org/10.1115/OMAE2018-77495.

[19] V. Sohoni, S.C. Gupta, R.K. Nema, A Critical Review on Wind Turbine Power Curve Modelling Techniques and Their Applications in Wind Based Energy Systems, J. Energy. 2016 (2016) 1–18. https://doi.org/10.1155/2016/8519785.

[20] How effectively a Sigmoid function curve can be fitted with a scaled Weibull Cumulative Distribution Function (CDF)? | by Neeraj Dhanraj | Medium, (n.d.). https://neerajdhanraj.medium.com/how-effectively-a-sigmoid-function-curve-can-be-fitted-with-a-scaled-weibull-cumulative-6f3a3dd7f19f (accessed April 15, 2021).

[21] Y. Zhu, C. Zhu, C. Song, Y. Li, X. Chen, B. Yong, Improvement of reliability and wind power generation based on wind turbine real-time condition assessment, Int. J. Electr. Power Energy Syst. (2019). https://doi.org/10.1016/j.ijepes.2019.05.027.

[22] T. Yuan, Z. Sun, S. Ma, Gearbox fault prediction of wind turbines based on a stacking model and change-point detection, Energies. (2019). https://doi.org/10.3390/en12224224.

[23] Z. Lin, X. Liu, M. Collu, Wind power prediction based on high-frequency SCADA data along with isolation forest and deep learning neural networks, Int. J. Electr. Power Energy Syst. 118 (2020) 105835. https://doi.org/10.1016/j.ijepes.2020.105835.

[24] L. Puggini, S. McLoone, An enhanced variable selection and Isolation Forest based methodology for anomaly detection with OES data, Eng. Appl. Artif. Intell. (2018). https://doi.org/10.1016/j.engappai.2017.09.021.

[25] E. Roslan, H. Mohamed, L. Chan, M.R. Isa, Effect of averaging period on wind resource assessment for wind turbine installation project at UNITEN, in: AIP Conf. Proc., 2018. https://doi.org/10.1063/1.5066898.

[26] Wavelet methods for time series analysis, Choice Rev. Online. 38 (2001) 38-4508-38–4508. https://doi.org/10.5860/choice.38-4508.

[27] A. Araghi, M. Mousavi Baygi, J. Adamowski, J. Malard, D. Nalley, S.M. Hasheminia, Using wavelet transforms to estimate surface temperature trends and dominant periodicities in Iran based on gridded reanalysis data, Atmos. Res. 155 (2015) 52–72. https://doi.org/10.1016/j.atmosres.2014.11.016.

[28] S. Al Wadi, M.T. Ismail, M.H. Alkhahazaleh, S.A.A. Addul Karim, Selecting wavelet transforms model in forecasting financial time series data based on ARIMA model, Appl. Math. Sci. (2011).

[29] X. Liu, Z. Lin, Z. Feng, Short-term offshore wind speed forecast by seasonal ARIMA - A comparison against GRU and LSTM, Energy. 227 (2021) 120492. https://doi.org/10.1016/j.energy.2021.120492.

[30] C.A. Martín, J.M. Torres, R.M. Aguilar, S. Diaz, Using deep learning to predict sentiments: Case

study in tourism, Complexity. 2018 (2018). https://doi.org/10.1155/2018/7408431.

[31] P.G. Zhang, Time series forecasting using a hybrid ARIMA and neural network model, Neurocomputing. 50 (2003) 159–175. https://doi.org/10.1016/S0925-2312(01)00702-0.

[32] An Introduction to Wavelets and Other Filtering Methods in Finance and Economics - 1st Edition, (n.d.). https://www.elsevier.com/books/an-introduction-to-wavelets-and-other-filtering-methods-in-finance-and-economics/gencay/978-0-12-279670-8 (accessed April 13, 2021).

[33] W. Wang, J. Ding, Wavelet Network Model and Its Application to the Prediction of Hydrology, Nat. Sci. (2003).

[34] Critically-Sampled Wavelet Reconstruction - MATLAB & Simulink - MathWorks United Kingdom, (n.d.). https://uk.mathworks.com/help/wavelet/gs/wavelet-reconstruction.html (accessed April 13, 2021).

[35] S. Ng, P. Perron, A Note on the Selection of Time Series Models, Oxf. Bull. Econ. Stat. 67 (2005) 115–134. https://doi.org/10.1111/j.1468-0084.2005.00113.x.

[36] G. Schwarz, Estimating the Dimension of a Model, Ann. Stat. 6 (1978) 461–464. https://doi.org/10.1214/aos/1176344136.

[37] D.A. Dickey, W.A. Fuller, Distribution of the Estimators for Autoregressive Time Series with a Unit Root, J. Am. Stat. Assoc. 74 (1979) 427–431. https://doi.org/10.1080/01621459.1979.10482531.

[38] D.P. Kingma, J.L. Ba, Adam: A method for stochastic optimization, in: 3rd Int. Conf. Learn. Represent. ICLR 2015 - Conf. Track Proc., International Conference on Learning Representations, ICLR, 2015. https://arxiv.org/abs/1412.6980v9 (accessed April 14, 2021).

[39] M. Chandra Mukkamala, M. Hein, Variants of RMSProp and Adagrad with Logarithmic Regret Bounds, PMLR, 2017. http://proceedings.mlr.press/v70/mukkamala17a.html (accessed April 14, 2021).

[40] Y. Ida, Y. Fujiwara, S. Iwamura, Adaptive Learning Rate via Covariance Matrix Based Preconditioning for Deep Neural Networks, IJCAI Int. Jt. Conf. Artif. Intell. 0 (2016) 1923–1929. https://doi.org/10.24963/ijcai.2017/267.

[41] What causes wind gusts? | The Weather Guys, (n.d.). https://wxguys.ssec.wisc.edu/2012/01/09/what-causes-wind-gusts/ (accessed December 3, 2021).

[42] B. Kosovic, S.E. Haupt, D. Adriaansen, S. Alessandrini, G. Wiener, L.D. Monache, Y. Liu, S. Linden, T. Jensen, W. Cheng, M. Politovich, P. Prestopnik, A comprehensive wind power forecasting system integrating artificial intelligence and numerical weather prediction, Energies. 16 (2020). https://doi.org/10.3390/en13061372.

[43] J. Sun, N.A. Crook, Dynamical and microphysical retrieval from doppler radar observations using a cloud model and its adjoint. Part II: Retrieval experiments of an observed Florida convective storm, J. Atmos. Sci. 55 (1998) 835–852. https://doi.org/10.1175/1520-0469(1998)055<0835:DAMRFD>2.0.CO;2.

[44] A. Yamaguchi, T. Ishihara, Maximum instantaneous wind speed forecasting and performance evaluation by using numerical weather prediction and on-site measurement, Atmosphere (Basel). 12 (2021). https://doi.org/10.3390/atmos12030316.

[45] L.S. Seregina, R. Haas, K. Born, J.G. Pinto, Development of a wind gust model to estimate gust speeds and their return periods, Tellus A Dyn. Meteorol. Oceanogr. 66 (2014) 22905.

https://doi.org/10.3402/tellusa.v66.22905.