Wind Power Forecasting of an Offshore Wind Turbine based on High-Frequency SCADA Data and Deep Learning Neural Network

Zi Lin[a], Xiaolei Liu[b1]

[a]*Offshore Engineering Institute, University of Strathclyde, Glasgow, G4 0LZ, United Kingdom*
[b]*School of Engineering, University of Glasgow, Glasgow, G12 8QQ, United Kingdom*

Abstract

Accurate wind power forecasting is essential for efficient operation and maintenance (O&M) of wind power conversion systems. Offshore wind power predictions are even more challenging due to the multifaceted systems and the harsh environment in which they are operating. In some scenarios, data from Supervisory Control and Data Acquisition (SCADA) systems are used for modern wind turbine power forecasting. In this study, a deep learning neural network was constructed to predict wind power based on a very high-frequency SCADA database with a sampling rate of 1-second. Input features were engineered based on the physical process of offshore wind turbines, while their linear and non-linear correlations were further investigated through Pearson product-moment correlation coefficients and the deep learning algorithm, respectively. Initially, eleven features were used in the predictive model, which are four wind speeds at different heights, three measured pitch angles of each blade, average blade pitch angle, nacelle orientation, yaw error, and ambient temperature. A comparison between different features shown that nacelle orientation, yaw error, and ambient temperature can be reduced in the deep learning model. The simulation results showed that the proposed approach can reduce the computational cost and time in wind power forecasting while retaining high accuracy.

Keywords: Wind power forecasting; Deep learning; Feature engineering; Offshore wind turbines; SCADA data.

Nomenclature:

Latin symbols

B	Air pressure at hub height
bj	Bias associated with neuron j

[1] Corresponding author, E-mail: Xiaolei.Liu@glasgow.ac.uk (XL)

CP	Power coefficient, denoting power captured by the turbine in percentage
Hi	Net input of neuron j in the output or deeper hidden layer
h	Output of neuron j
m	Number of tensors
max$x$	Maximum value in the span
min(x)	Minimum value in the span
n	Number of data points
P	Wind power
$Pmeasured$	Measured value from the SCADA database
$Ppredicted$	Predicted wind power from deep learning modelling
Pw	Vapour pressure
$(Pmeasured)k$	Measured value of the kth data point from the SCADA database
$(Ppredicted)k$	Predicted wind power of the kth data point from deep learning modelling
$(Qmeasured)i$	Measured value of the ith tensor from the SCADA database
$(Qpredicted)i$	Predicted value of the ith tensor from deep learning modelling
R	Rotor radius
$R0$	Gas constant of dry air
Rw	Gas constant of water vapour
T	Absolute air temperature
u	Wind speed
$uref$	Wind speed at the reference height
Var	Variance

wij	Weights that linked neuron i and j
$Xscaled$	Normalized value
xi	Input of neuron j
xp	Initial value
z	Height
$zref$	Reference height
$z0$	roughness length at the current wind direction

Greek symbols

⌀	Relative humidity
γ	Yaw error
θ	Nacelle position
ϑ	Wind direction
ρ	Air density

ABBREVIATION:

ANN	Artificial Neural Network
AWNN	Adaptive Wavelet Neural Network
BPNN	Back-propagation Neural Network
CGNN	Conjugated Gradient Neural Network
EVS	Explained Variance Score
LSSVM	Least Squares Support Vector Machine
LSTM	Long Short-term Memory
MAE	Mean Absolute Error
MAPE	Mean Absolute Percentage Error
MLP	Multilayer Perceptron
MSE	Mean Square Error
MSLE	Mean Squared Logarithmic Error

NMAE	Normalized Mean Absolute Error
NRMSE	Normalized Root Mean Square Error
NWP	Numerical Weather Prediction
O&M	Operation and Maintenance
PCC	Pearson Correlation Coefficient
RBFNN	Radial Basis Function Neural Network
ReLU	Rectified Linear Unit
RMSE	Root Mean Square Error
SCADA	Supervisory Control and Data Acquisition

1. Introduction

Renewable energies are playing an increasingly significant role in reducing global carbon footprint [1]. Among them, wind energy is considered as a great alternative to conventional fossil fuels [2,3]. For instance, European countries have highlighted a marked increase in newly installed offshore wind farms. More specifically, 80% of the world's newly installed offshore wind was from EU countries at the end of 2017 [1]. Compared with onshore wind farms, offshore wind farms have the advantage of containing plenty of wind sources, lavish construction sites and larger capacity of wind generations [4]. Therefore, the wind turbine industry has seen a continuous move from onshore wind turbines to offshore ones. Meanwhile, due to the uncertain environment that they are locating in and malfunctions of offshore wind turbines, there is an ever-increasing attention on optimizing the performance of offshore wind turbines. The aim is to lower the cost [5,6] and improve the efficiency of energy captured from newly installed renewable energy sources [7]. Accurate power forecasting is a challenging task but essential to wind turbines as they are capable of reducing the operational cost [8], which is crucial for wind farms moving from onshore to offshore [9].

Recently, it has been demonstrated that Artificial Neural Network (ANN) can effectively predict wind power while the physical process of wind turbines is too complex to be explained. A considerable literature has grown up around the theme of wind power forecasting [10–18]. Zhao *et al.* applied a Kalman filter along with Numerical Weather Prediction (NWP) in ANN models to increase the accuracy of wind power forecasting, where a monthly averaged Normalized Root Mean Square Error (NRMSE) of 16.47 % was reached [19]. Liu *et al.* developed short-term wind power forecasting models based on different algorithms of Adaptive Neuro-fuzzy Inference System (ANFIS), Back-propagation Neural Network (BPNN), and Least Squares Support Vector Machine (LSSVM), in which the data pre-processing method of Pearson Correlation Coefficient (PCC) was used to enhance the accuracy of ANNs [20]. Singh *et al.* concluded that wind speed and wind

direction were the top two influence factors on wind power prediction through Multilayer Perceptron (MLP) networks [21]. However, the influence of wind direction on power generation is much less than wind speed. The authors also claimed that a balance between modelling speed and accuracy could be achieved through training the neural network for individual wind turbine instead of wind farms, which can effectively decrease the size and complexity of the used networks. Carolin Mable and Fernandez studied wind power predictions utilizing ANN based on a 3-year database containing wind speed, relative humidity and generation hours [22]. The authors concluded that wind speed has a direct effect on power generation. Besides, it seems like wind speeds that are higher than the rated wind speed were essential for high power generation. Jafarian and Ranjbar studied annual power forecasting based on hourly recorded wind speeds from 25 different stations in Netherland by applying fuzzy modelling and ANN [23]. In this investigation, average wind speeds, standard derivation of wind speeds, and air density were selected as input features [23]. Peng et al. compared the algorithm of ANN and a hybrid strategy based on physical/statistical models in wind power predictions [24]. The authors concluded that the ANN model could provide the prediction results quickly with a relatively low accuracy while the hybrid predicting method operated slowly with high accuracy. Zameer *et al*. developed an integrated model using both ANN and genetic programming for short-term power forecasting based on an hourly sampled database from five wind farms in Europe [25]. The authors concluded that an average root mean squared error of 0.117575 is reached [25]. Zhang et al. performed a short-term wind power prediction and uncertainty analysis based on Long short-term memory (LSTM) [26]. The datasets were based on a wind farm locates in North China and recorded in the first quarter of 2010 with a 15-min sampling. The input features of the LSTM algorithm included wind speed, output power and NWP data.

The mentioned neural network algorithms, input features, and accuracies in the studies above are summarized in **Table 1**. Even though various ANN algorithms have been applied, all authors were sharing a few types of input features. All the presented studies in **Table 1** can be classified to short (a few hours), medium (days) and long term (one month ~ a few months) based on their potential predictable time.

Table 1 – Neural network algorithms and used features in wind power forecasting from previous studies.

References	*Training algorithm*	*Input features*	*Sampling rate*	*Accuracy*
Pelletier et al., 2016 [10]	MLP ANN	wind speed, wind direction, air density, turbulence intensity, wind shear, yaw error	10-min	Mean absolute error (MAE) = 15.3 ~ 15.9 kW
Giorgi et al., 2011 [11]	MLP ANN	wind speed, pressure, temperature, relative humidity	10-min	Normalised absolute average error = 0.1098 ~ 0.1550
Xu and Mao, 2016 [14]	Elman neural network	wind speed, wind direction, pressure, temperature, relative	15-min	Mean square error (MSE) = 16.55%, MAE = 10.52%

		humidity		
Bilal et al., 2018 [15]	ANN	wind speed, wind direction, solar radiation, temperature and relative humidity	10-min	Fitting rate = 98.56 %
Li et al., 2016 [16]	Conjugated gradient neural network (CGNN)	wind speed, wind direction, temperature, pressure, relative humidity	15-min	MSE = 0.002 ~ 0.004
Zhao et al., 2012 [19]	MLP ANN	wind speed, wind direction, temperature, pressure, relative humidity	6-hour	NRMSE = 0.1647
Liu et al., 2017 [20]	BPNN, RBFNN [27], and LSSVM	weighted mean wind speed, weighted wind direction, and weighted temperature	15-min	Mean absolute percentage error (MAPE) = 6.70 ~ 27.40%; Normalized mean absolute error (NMAE) = 1.01 ~ 6.35%; NRMSE = 2.37 ~ 9.45%
Singh et al., 2007 [21]	MLP ANN	wind speed, wind direction, air density	10-min	Percentage difference between measured and predicted results = 0.303 ~ 1.082 %
Mabel and Fernandez, 2008 [22]	MLP ANN	wind speed, relative humidity, generation hours	1-month	MSE = 0.0065; MAE = 0.0586
Jafarian and Ranjbar, 2010 [23]	Takagi-Sugeno modelling technique, radial basis network and Generalized Regression Network	wind speed, standard deviation of wind speed, air density	1-hour	Root Mean Square Error (RMSE) = 1.31×10^5 ~ 1.62×10^5
Peng et al., 2013 [24]	ANN and hybrid strategy based on physical and statistical methods	wind speed, wind direction, temperature	10-min	RMSE = 0.0201
Zameer et al., 2017 [25]	ANN and genetic programming	wind speed, wind direction,	1-hour	RMSE = 0.117575
Jyothi and Rao, 2016 [17]	Adaptive wavelet neural network (AWNN)	wind speed, wind direction, air density, temperature	10-min	NRMSE = 0.1647

The statistical description of how features were selected in the previous studies summarised in **Fig. 1**, which is based on the information from **Table 1**. As presented in **Fig.1**, most investigations involved a selective number of meteorology-related features in their wind power predictive model, such as wind speed, wind direction and influence factors of air density.

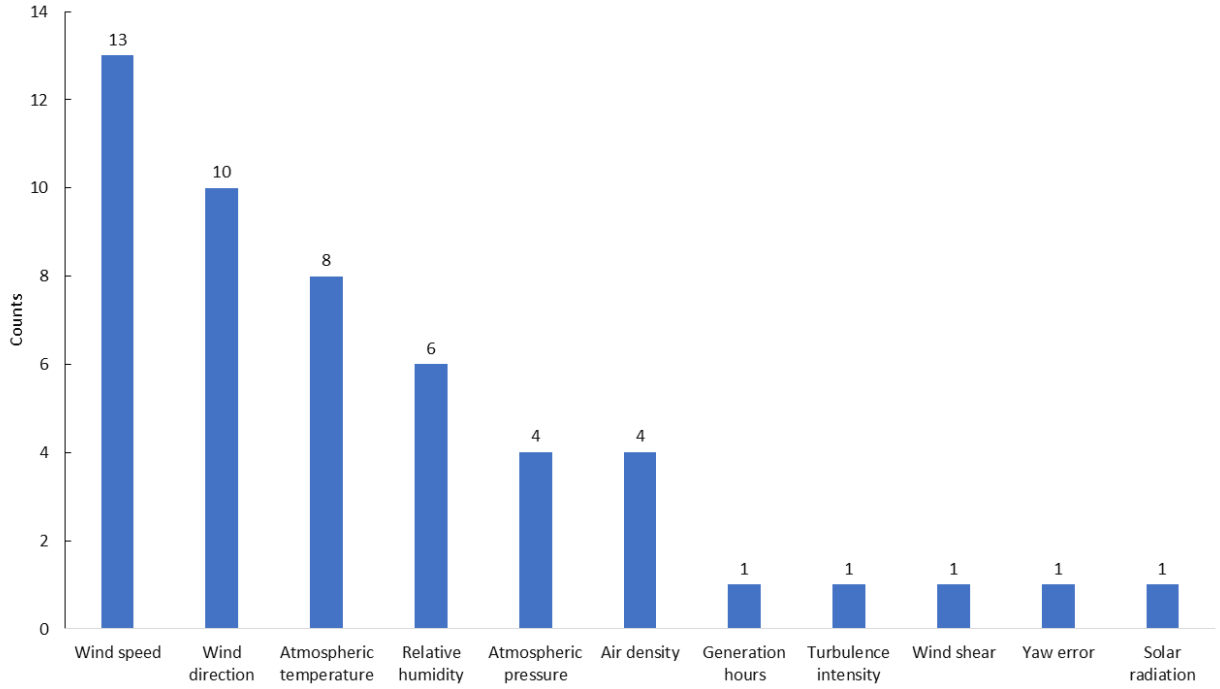

Fig. 1 – Statistics of neural network features in wind power forecasting from reviewed literature.

The offshore wind energy outputs are more unpredictable and more complex because of the harsh ocean environment in which the wind turbines are operating [2]. In recent years, ANN has been considered as a great alternative in wind power forecasting to conventional predictive methodologies while the physical process of wind production is too complicated [28,29]. The general approach of machine learning in wind power prediction is based on building relationships between power outputs and selected features that can influence wind turbine conversion systems. Thus, this method is highly dependent on the suitability of features and the size of training datasets. In our study, multiple features from the SCADA system were carefully selected to predict the active power of an offshore wind turbine located at Levenmouth, Fife, Scotland, UK. The key contribution of this paper to the current knowledge gap can be summarised as follows:

a. The most commonly used feature in wind power prediction is the wind speed at the hub height. However, wind power forecasting is also affected by wind shear, which is rarely considered in previous studies (see **Table 1**). As well known, wind speed profiles cause changed wind speeds along with the blades from the ground to the top [30]. Wind shear can establish a relatively large bending moment in the shaft of a turbine, which in turn influences the wind turbine operation [31]. The BSI Standards of IEC 61400-12-1 has recommended having as many measurement heights as possible to minimise wind speed uncertainty [32]. In this study, four wind speed measurements over a range of heights are involved in the designed deep learning neural network to take into

account the influences from wind shear, including the wind speed at the hub height of 110.6 m and three additional wind speed measurements at the heights of 25 m, 67 m, and 110 m, respectively.

b. As essential technologies of modern wind turbines, pitch-control and yaw-control have been widely applied in wind farms. However, these inherent features within the wind energy conversion system are not often considered in previous studies. A complete wind power predictive model shall take the influence of blade pitch angle into account, which controls the safe and stable operations of wind power productions when wind speeds are above the rated values. Also, wind turbines are normally operated by a yaw control system to follow wind direction for optimizing power harvesting [33]. In this paper, both features of yaw error and blade pitch angle (three measured pitch angles of each blade and average blade pitch angle) are involved in the designed deep learning neural network for wind power forecasting.

c. In supervised machine learning, it is widely accepted that too-small training dataset size results in poor predictions. It is significant to train a deep learning neural network with multiple impact factors and a reasonable large training dataset. In previous studies (see **Table 1**), sampling rates of 10 ~ 15 mins in short-term or long-term wind power predictions are often applied, resulting in using relatively small training datasets. Unlike previous investigations, the designed deep learning predictive model in this paper was based on a very high-frequency SCADA database with a sampling rate of 1-second. Also, several studies have claimed that the trends of wind speed and direction variations are similar in different years [34,35]. Therefore, in this study, the most recent completed one-year high-frequency database was extracted from the SCADA measurements, where data points were collected in every second from 01/07/2018 to 30/06/2019.

d. The power transmission systems of wind turbines have the feature of high nonlinearity and are difficult to be represented using simple models. There have been few quantitative analyses of nonlinear relationships between input features and the target output, including their correlations. To this end, this paper applied the deep learning neural network to explore nonlinear correlations between various features and power generations. Based on the identified nonlinear correlations, the deep learning predictive model was further boosted with respect to feature dimension reduction, and ultimately providing a useful tool in developing wind turbine reduced-order deep learning models. Traditional reduced-order models highly rely on linearization of the wind turbine system, including a holistic modelling of all the structural components, drivetrain system, generator, converter and link to shore, etc. which is sometimes difficult or even impossible to realise. This study presents a methodology to develop reduced-order neural networks without the linearization process, which is able to

predict wind power with a lower computational cost, and therefore more easily for scaling up reduced-order deep learning models for a wind farm. A comprehensive comparison is displayed between the reduced and non-reduced predictive models, including their computing accuracies and processing efficiencies.

The remainder of this paper is organized as follows. Section 2 presents how features were engineered in this study on predicting wind power through deep learning neural networks. Section 3 describes the SCADA database used in this paper, including data pre-processing and data correlation. Section 4 introduces the deep learning neural network configuration that was designed for this study, including how the predictive model was trained, tested and validated. Results and discussions based on the established deep learning predictive model are showed in section 5. To sum up, a series of key conclusions are performed in section 6.

2. Feature engineering

The major purposes of feature engineering are to accurately reduce dimensionality and effectively increase the computational performance of the designed wind power predictive model. In this study, the following features were characterised based on the physical process of wind energy conversion systems.

Wind speed and wind shear

Generally, the higher the wind speed is, the more power can be generated. Theoretically, wind power can be evaluated by the following equation [36]:

$$P=12\rho\pi R2CPu3 \tag{1}$$

As displayed in **Eq. (1)**, the performance of a wind turbine is directly shaped by wind speed, since wind power is proportional to the cube of wind speed. Furthermore, wind shear is also playing a significant role in wind power extraction. The fact that the wind profile is trending towards a relatively lower wind speed as the height is closer to ground level, is called wind shear. The wind shear formula can be used to calculate wind speeds at different heights through the logarithmic wind profile law, which can be expressed as [37]:

$$u= urefln\left(\frac{z}{z0}\right)\ln\left(\frac{zref}{z0}\right) \tag{2}$$

In this study, four wind speeds from SCADA measurements at different heights were used in the designed deep learning neural network to present the influences of wind speed and wind shear on wind power forecasting, including the wind speed at the hub height of 110.6 m and the three additional wind speed measurements at the heights of 25 m, 67 m, and 110 m, respectively.

Wind direction

Wind direction is also one of the most widely used features in wind power predictive models. However, compared with wind speed, it has less impact on power generation because all wind turbines are designed to face into the wind during operating time. Under identical wind speeds, there is no obvious difference in wind power generations from different wind directions. The wind direction can be derived from the position of the nacelle and yaw error, which can be expressed as [38]:

$$\vartheta = \theta + \gamma \quad (3)$$

In this study, the features of nacelle orientation and yaw error were selected as inputs to represent the influence of wind directions.

Blade pitch angle

In this paper, parameters regarding blade pitch angle are also considered as input features. Pitch angles adjust blades of a wind turbine to control them so that they use the proper fraction of the available wind to obtain the regulated power generation while making sure the turbine itself does not exceed its rated power. When wind speed is over or close to its rated value, blade pitch angle may play a significant role in power predictions. In the target wind power conversion system, blades can be pitched individually. Therefore, in this investigation, four features regarding blade pitch angle are used in the designed deep learning neural network for wind power forecasting, including the average blade pitch angle, the measured pitch angle of blade 1, the measured pitch angle of blade 2, and the measured pitch angle of blade 3.

Air density

As showed **Eq. (1)**, wind power is also linearly proportional to air density. Therefore, as air density is changing during day and night, the produced active power will also vary accordingly. The air density could be governed by air temperature, air pressure and relative humidity, which can be expressed in **Eq. (4)** [32]:

$$\rho = 1TBR0 - \varnothing Pw1R0 - 1Rw \quad (4)$$

Due to measuring errors and availability issues, most studies were not using all the four parameters as features in their neural networks to represent the influence of air density. Actually, the impact of relative humidity on air density is minor, international standards allow to use a constant of 50% to represent it in when it is not measured [32]. In this paper, air temperature is used as the feature in our deep learning model to characterise the influence of air density.

In summary, eleven critical features were used in this investigation based on the actual physical process of wind operations, which are the wind speed at the hub height of 110.6 m, the wind speeds at heights of 25 m, 67 m and 110 m,

respectively, the average blade pitch angle, the measured pitch angle of blade 1, the measured pitch angle of blade 2, the measured pitch angle of blade 3, nacelle orientation, yaw error, and ambient temperature.

3. Data description

The investigated SCADA database was recorded from a demonstration offshore wind turbine, which is owned by the ORE Catapult [39]. **Fig. 2** displays the major properties of the Levenmouth offshore wind turbine, which is a three-bladed upwind turbine with a rated power of 7 MW. The support structure is a jacket type and the length from hub height to sea level is 110.6 m (see **Fig. 2**). More details about wind turbine dimensions can be found in **Table 2**.

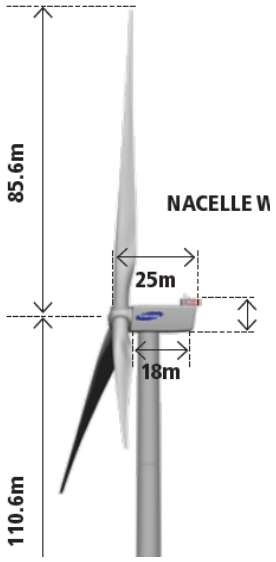

Fig. 2 – Layout of the 7 MW Levenmouth offshore wind turbine [39].

Table 2 – Major properties of Levenmouth offshore wind turbine [39].

Properties	*Value*
Wind class	IEC class 1A
Rotor diameter	171.2m
Capacity	7 MW
Hub height	110.6m
Blade length	83.5m
Generator	Medium (3.3kV), PMG
Converter	Full power conversion
Drivetrain	400rpm
Rated frequency	50Hz
Rotor speed	5.9-10.6rpm
Wind speed	3.5-25m/s
Rated wind speed	10.9m/s
Design life	25years
Certification	DNV

3.1 Data pre-processing

Before processing any data into wind power predictions, outliers that deviate from normal observations were detected and removed from the SCADA database, representing measuring variability or error detections. In this paper, the algorithm

of isolation forest is used to detect data points that diverge from the overall pattern in wind power measurements, which is a type of tree ensemble methods that are based on decision trees. Isolation forest has been identified as one of the most effective algorithms in wind power prediction [40]. The outlier fraction is identified as 5% in the isolation forest, which kept 95% of what is reflected as normal data. A comparison was presented in **Fig. 3** based on wind power curves, where the SCADA databases before (**Fig. 3a**) and after (**Fig. 3b**) isolation forest filtering are displayed, respectively. The distinguish between the two patterns were noticeable. For the scenario before using isolation forest, three types of operating issues can be observed in **Fig. 3a**:

- Outlier 1: This type of anomalies is located in a horizontally dense cluster at the bottom of the power curve that was caused by turbine downtime, where wind speeds are larger than the cut-in wind speed (3.5 m/s) but the corresponding active power is near null.
- Outlier 2: This group of outliers are represented by a dense cluster that is located at the middle of the power curve, where the wind turbine performance is constrained. Wind curtailment can be triggered by the operators for several reasons, including the grid supply limitations, lack of demand at given times, or the difficulty in storing large capacity wind power.
- Outlier 3: This type of outliers is randomly scattered surrounding the power curve, which can be caused by sensor malfunction (over or under measured wind speeds) or potential noise in signal processing.

For the case after isolation forest filtering (see **Fig. 3b**), most detected outliers, which were located at the boundaries of the pattern, have been automatically discarded. In the following sessions, the SCADA dataset filtered by the isolation forest was used as the target database in the deep learning model.

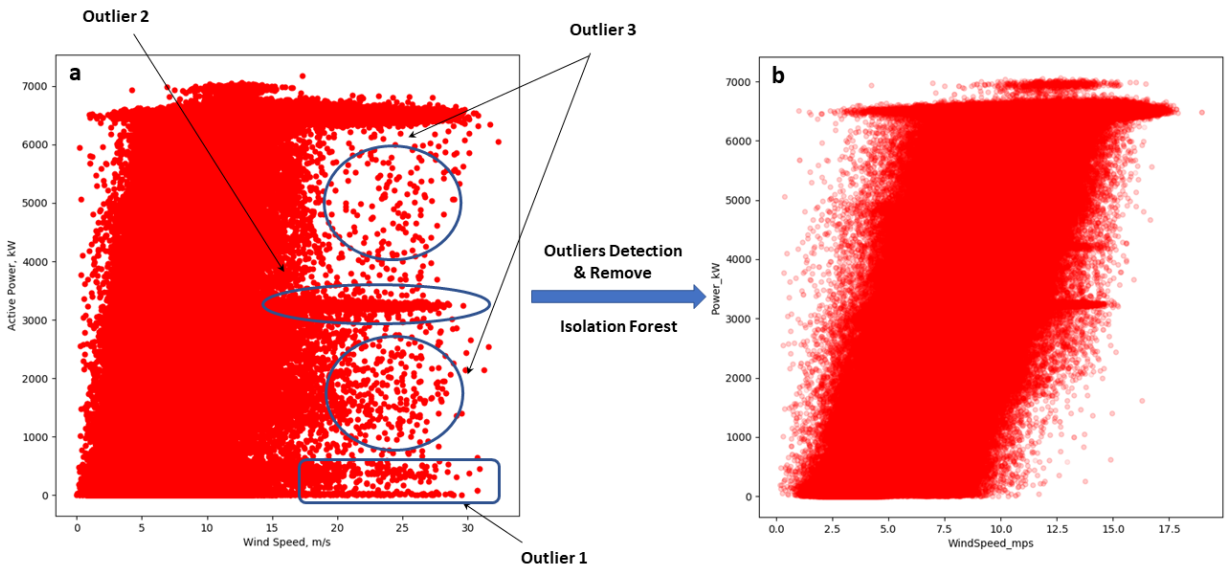

Fig. 3 – Wind power curves before (a) and after (b) isolation forest filtering.

3.2 Correlations

The histogram of each feature is presented in **Fig. 4**. In the current database, the rated wind speed of the target offshore wind turbine is 10.9 m/s, while the mean and median of the recorded wind speeds are 10.9 and 10.7 m/s, respectively (see the histogram of wind speed in **Fig.4b**). It indicated that the generated active powers are close to the rated power (7 MW) at most of the operating time (see the histogram of active power in **Fig.4a**). Similarly, the blade pitch angles were mainly varying on the range of 3 ~ 4 deg (see the histogram of blade pitch angles in **Fig.4f**, **g**, **h**, **and i**), where the mean value of average blade pitch angles is around 3.36 deg. Also, the dispersals of wind speeds at different heights, ambient temperature, and yaw error followed normal distributions (see the histogram of wind speeds at different heights, yaw error, and ambient temperature in **Fig.4b**, **c**, **d**, **e**, **j** and **l**). The histogram of nacelle orientation goes along with a bimodal distribution (**Fig. 4k**), revealing local wind directions can be roughly classified into two different clusters.

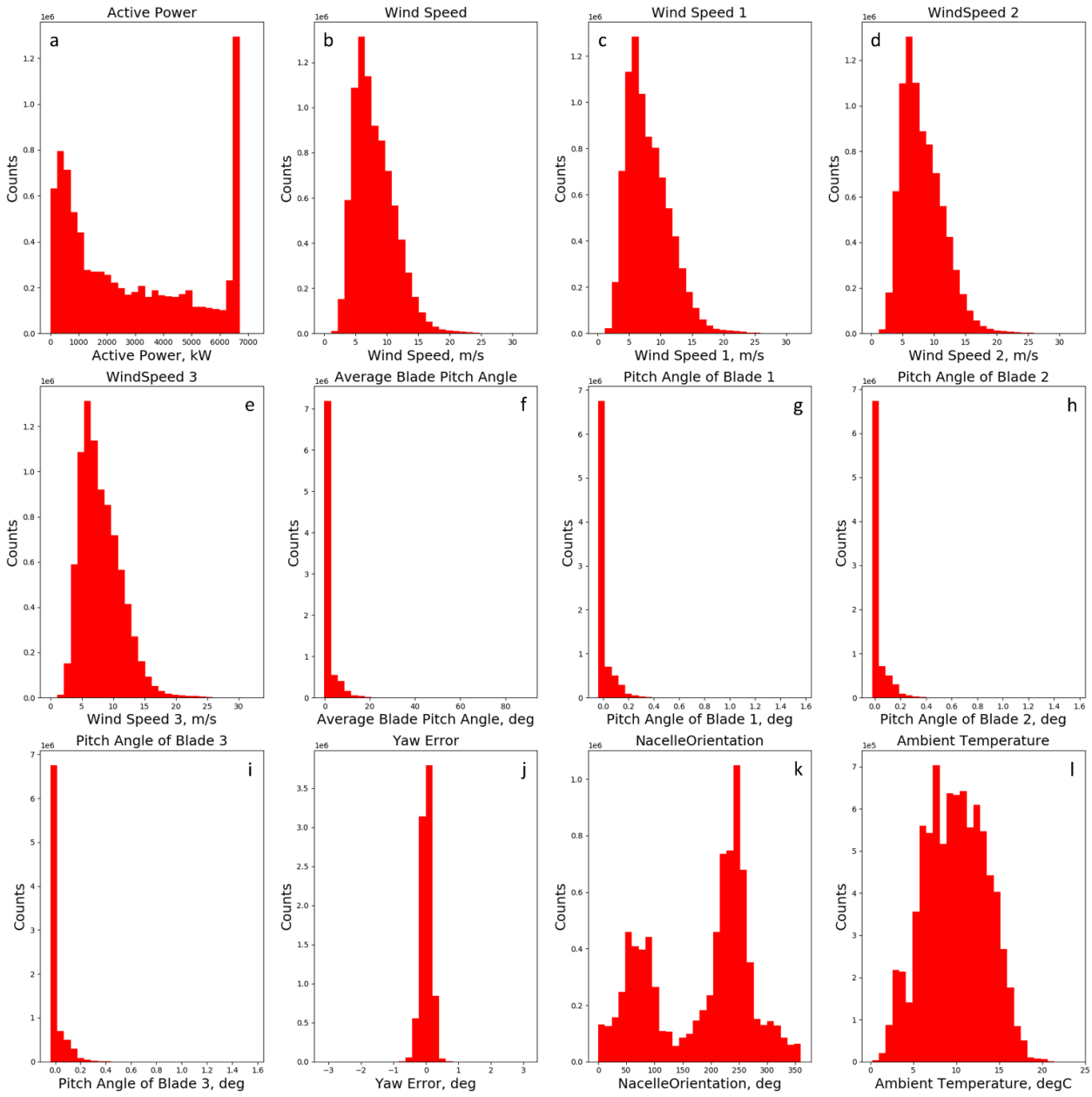

Fig. 4 – Scatter matrix of selected features, including wind speeds, blade pitch angles, nacelle orientation, ambient temperature, and yaw error.

It is also essential to discover and quantify the degree to which parameters in the SCADA database have greater influences on wind power generations than others. Noted that, it is difficult to identify these relationships directly from the SCADA database when as many factors have impacts on power generation simultaneously. A graphical representation of correlation coefficients of all input features to active powers is shown in **Fig. 5** in the form of a heat map, where the individual coefficient contained in a matrix are represented by colours. The Pearson product-moment correlation coefficients were marked for each feature in the displayed heat map, which is one of the most common measurements of the strength in a linear relationship between any two variables. This type of correlation is defined as the covariance of the target variables

divided by the product of the corresponding standard deviations. It provided values between +1 and −1, in which +1 is representing a completely positive linear correlation, 0 is indicating no linear correlation, and −1 is characterizing a completely negative linear correlation. As presented in the dashed bordered rectangle of **Fig. 5**, correlations of wind speeds and blade pitch angles to active power are positive, indicating that the values of these three variables are highly correlated and growing in the same direction. On the other hand, correlations of other parameters (ambient temperature, nacelle orientation and yaw error) to active power are closer to zero, meaning that those variables are not strongly linear-related to the generated power. Note that, correlation coefficients only considered linear relationships. In another word, this method may completely ignore non-linear relationships and could only be used for preliminary evaluations. A more accurate methodology to measure how much each feature correlates with active powers will be proposed in session 5.2 through deep learning neural networks.

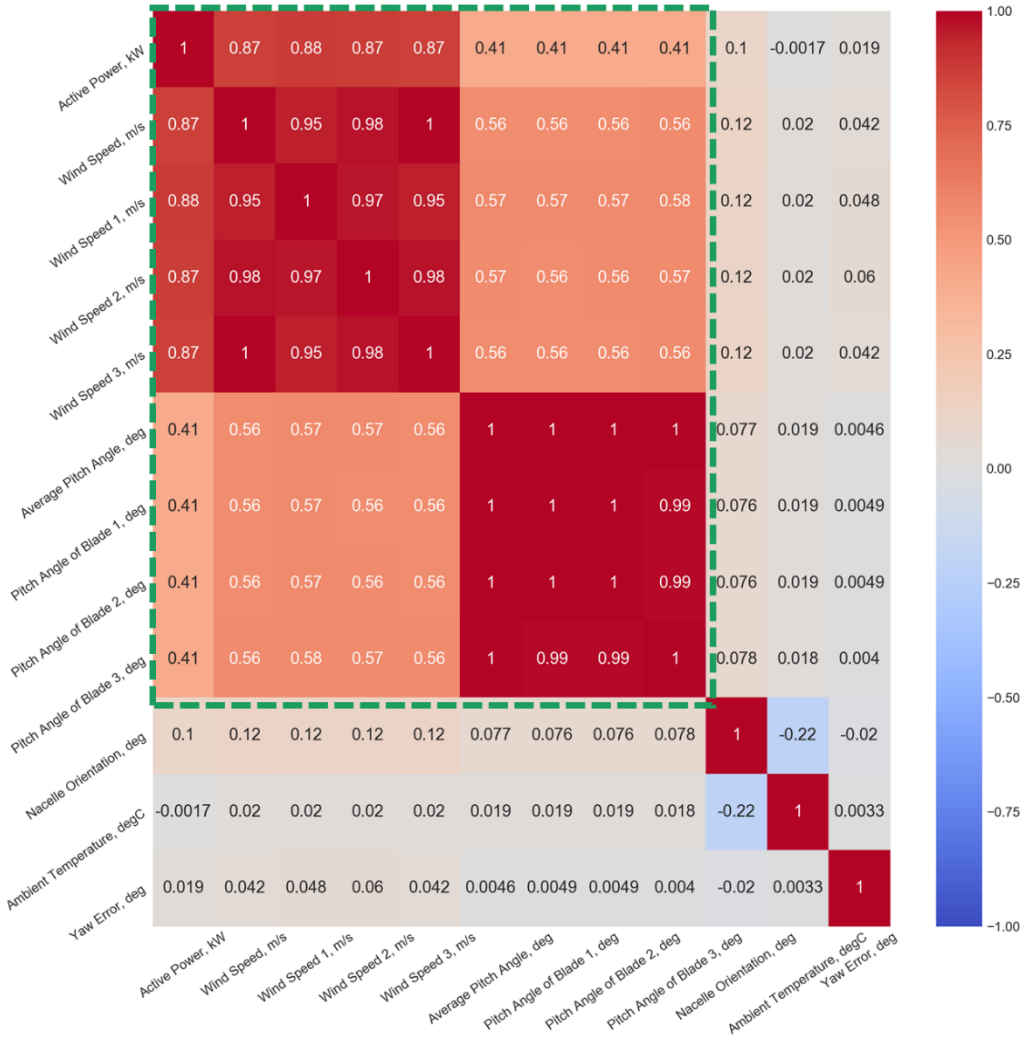

Fig. 5 – Heat map of input features against active power.

4. Deep learning configuration

In this paper, TensorFlow, which was developed and supported by Google, was used as the platform to create a deep learning structure, where Python3 was employed as the major programming language. In TensorFlow, large datasets with certain individual attributes could be smoothly handled, such as multi-dimensional arrays. These multi-dimensional arrays can also be named as tensors. Graphically, tensors flow from one layer to the other in neural networks. To increase the accuracy of wind power predictions, several structures of deep learning neural networks were critically tested while assessments were carried out for different layer amounts and neuron numbers in each layer. Subsequently, a five-layer feedforward neural network was selected to shape the relationship between the input and the output tensors. In the designed neural network, the eleven input features are the four wind speeds at different heights, the three measured pitch angles of each blade, the average blade pitch angle, nacelle orientation, yaw error, and ambient temperature while the output feature is active power. Before input features (or tensors) flowing into deep learning layers, they were shrunk into a range between 0 and 1 by the Min-Max scaler. The used correlation can be expressed as:

$$Xscaled=xp-\min(x)\max x-\min(x) \quad (5)$$

The visualization of how tensors flowed in the deep learning structure is displayed in **Fig.6**, in which the flow direction of tensors among different computing operations was presented as solid arrows. The deep learning networks were fully identified as a computational graph, where all layers were connected so that tensors flowed from the initial layer throughout the final one. In the presented configuration, the 1st layer has 20 neurons, while the 2nd, 3rd, and 4th layer has 50 neurons, and the final layer has 20 neurons once again. A similar internal structure was assigned to all five layers. The operation of the 3rd layer was further extended to display the inner design for inspecting how deep learning functions in each layer. As presented, there are three essential components in each layer – weights, biases, and an activation function. The algorithm of Xavier was applied for weight initializations to avoid any overlarge or too small weight values [41]. On the other hand, bias initializations were achieved by the built-in initializer within TensorFlow to initialize bias values of each neuron as zero. The flow directions of initializations among different neural network layers were displayed as arrows with dotted lines in **Fig. 6**. After that, a net input of neuron was identified by multiplying the weights and adding the biases. In the end, the Rectified Linear Unit (ReLU) non-linear activation functions were called in the program. More specifically, the formula that is used for fully connected layer definitions is:

$$Hi=j=1mxiwij+bj \quad (6)$$

$$h=ReLU(Hi) \quad (7)$$

When the training phase was carried out, the cost function of MSE is used as a metrics to measure the accuracy of the outputting tensors, calculating the mean of tensor elements along various dimensions of the tensor, which can be expressed as:

$$MSE= \quad (8)$$

$$1mi=1m(Qpredicted)i-(Qmeasured)i(Qmeasured)i2$$

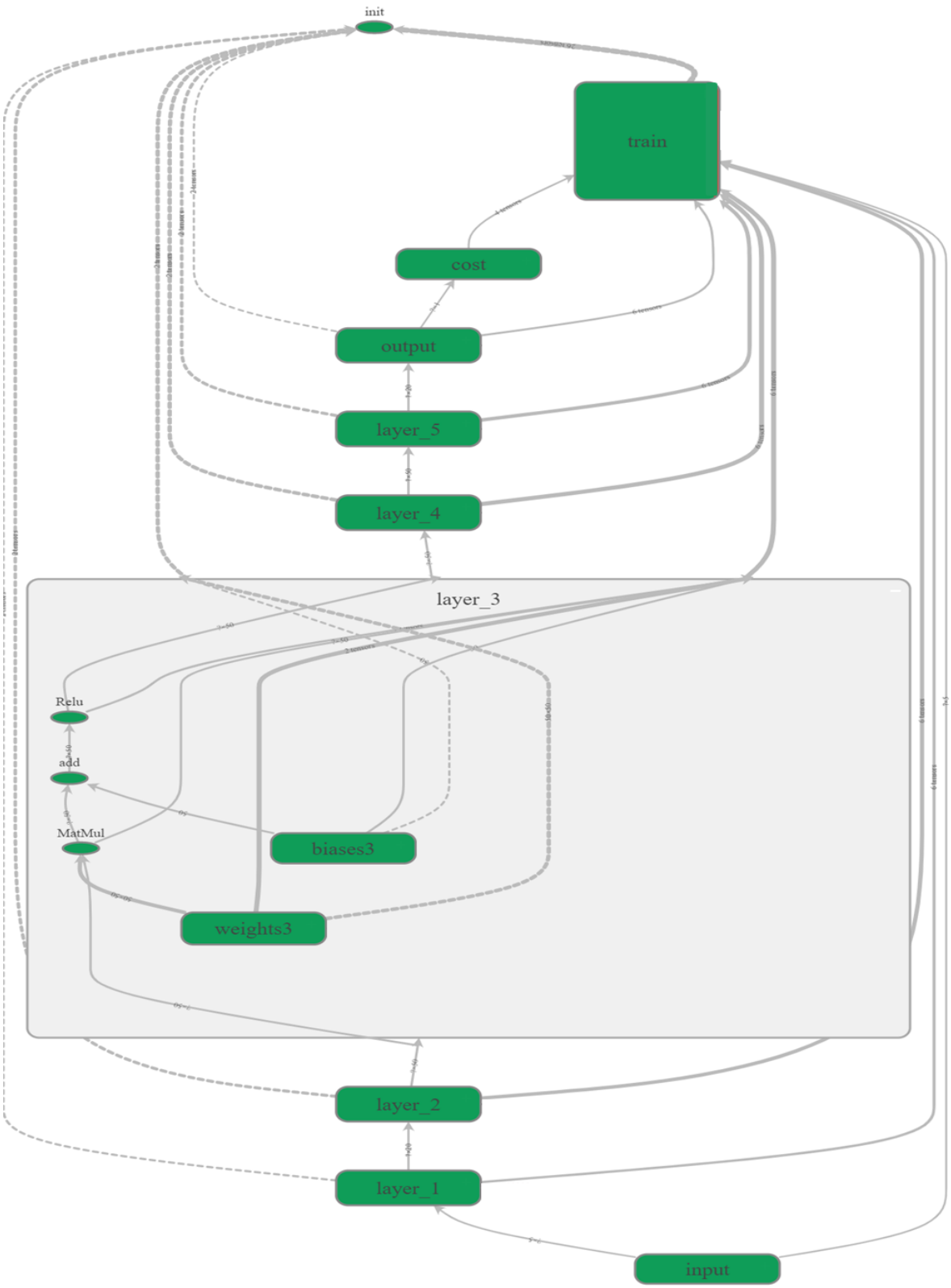

Fig. 6 – Visualization of how tensors flow in designed computational graphs within the deep learning model.

The conventional training-testing-validation workflow was obeyed in this study. The one year SCADA database was randomly divided into two groups – training group with 4.46×10^6 datasets (80%) and testing group with 1.11×10^6 datasets (20%). In the validation phase, the deep learning model was validated by a one-month SCADA dataset, which was collected in July 2019.

5. Results and discussions

5.1 Modelling results in the training and testing loops

In this neural network model, the learning rate is set as 0.01 while the running epochs were defined as 200. In the training and testing phases, the deep learning model began to converge after around 150 epochs while the final MSEs are equal to 0.004782 and 0.004780 in the last epoch for training and testing loops, respectively (see **Fig. 7**).

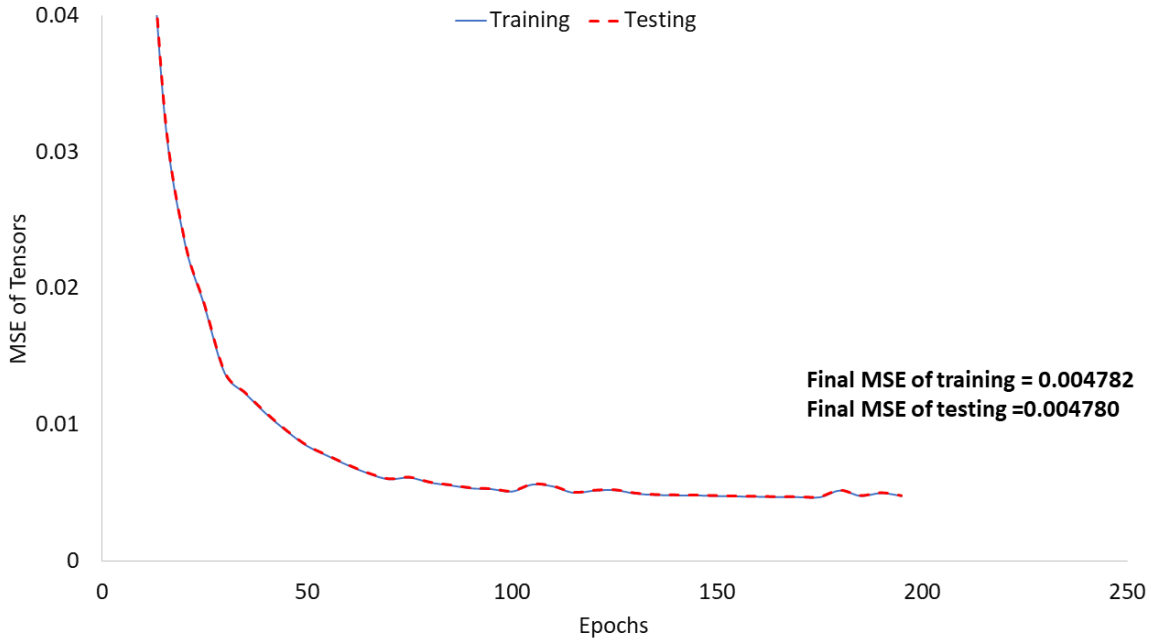

Fig. 7 – Variations of MSEs in training and testing loops along 200 epochs in the designed deep learning configuration.

5.2 Performance evaluation in the validation loop

The accuracy of the designed deep learning model in the validation loop was quantified through five different metric functions, including Root Mean Square Error (*RMSE*) [42,43], Mean Absolute Error (*MAE*) [44,45], R-square (R^2) [46], Mean Squared Logarithmic Error (*MSLE*) [47], and Explained Variance Score (*EVS*) [48], which were expressed in **Eqs. (9)**, **(10)**, **(11)**, **(12)**, and **(13)**, respectively.

RMSE is one of the most commonly used functions for measuring the differences between predicted values from a model and observed values, which can be defined as:

$$RMSE = k=1n(Ppredicted)k-(Pmeasured)k2n \quad (9)$$

The metric of *MAE* is corresponding to the estimated value of the absolute error loss, which can be expressed as:

$$MAE = 1nk=1n(Ppredicted)k-(Pmeasured)k(Ppredicted)k \quad (10)$$

R-square, which is also called the coefficient of determination, offered a sign of goodness of fit of how well the recorded data can be forecasted by a built model. The highest possible R score can reach 1.0. It can also be displayed as negative to indicate an arbitrarily worse predicting model. *R-square* can be defined as:

$$R2=1- \quad (11)$$
$$k=1n(Ppredicted)k-(Pmeasured)k2k=1n(Pmeasured)k-1nk=1n(Pmea\ ur\ ed)k2$$

The metric of *MSLE* is corresponding to the assessed value of the squared logarithmic errors, which can be stated as:

$$MSLE=1nk=0n-1(\log e(1+(Pmeasured)k)-\log e(1+(Ppredicted)k))2 \quad (12)$$

In data science, explained variation quantifies the proportion to which a predicting model accounts for the dispersion of a given dataset. The highest possible *EVS* is 1.0 in the best scenario, then lower values become worse. The *EVS* is valued as follow:

$$EVS=1-Var\{Pmeasured-Ppredicted\}Var\{Pmeasured\} \quad (13)$$

The predicting performance of the deep learning model in the validation loop is presented in **Table 3**, where high values of R2/EVS and low values of RMSE/MAE/MSLE were obtained, indicating a high accuracy in the current modelling effort.

Table 3 – Performance of the deep learning model under different metric functions in the validation loop.

Metric Functions	*RMSE*	R^2	*MAE*	*EVS*	*MSLE*
Values	517.33	0.91	374.41	0.91	0.29

The predicted wind powers by the neural network in the validation loop are also compared against the corresponding SCADA observations along with the time series of July 2019 in **Fig. 8**, where a great agreement was achieved.

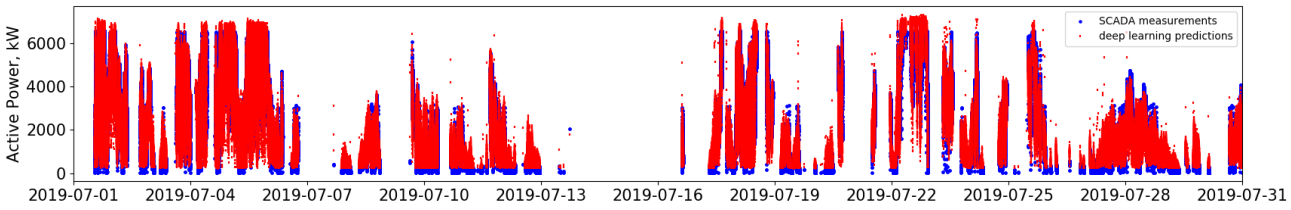

Fig. 8 – Comparison of wind power generation between deep learning model (red dotted points) and actual SCADA measurements (blue dotted points).

5.2 Non-linear correlations of each feature with active power

In this section, the non-linear correlations of input features to wind power generations were examined through deep learning neural networks. Owing to the aim of this study is to identify wind powers, variations of the final MSEs in the validation loop were used as the reference points. While the trained deep learning model was kept as it is, only one type of features was replaced by its mean value in the validation database at each trial to investigate its influence in wind power predictions. Since just one type of features is altered at one time in the validation loop, this analysis will be repeated on all the inputs one by one, including the four wind speeds at different heights, the four measured pitch angles of blades, nacelle orientation, yaw error, and ambient temperature. Variations of the last epoch of MSEs in validation loops under each type of features' changing are presented in **Fig. 9**. Comparing with the original case, the values of the final MSEs exaggeratively increased in the cases of blade pitch angle and wind speed & wind shear, while the cases of yaw error, ambient temperature, and nacelle orientation have nearly no influence to the results, indicating the level of significance of features to the current deep learning model can be ranked in such an order. As presented in **Fig. 9**, blade pitch angle and wind speed & wind shear have a strong impact on predictions, and the deep learning model could not function without these features. On the other hand, the influences from nacelle orientation, ambient temperature, and yaw error are relatively minor, the predicting model can still converge well after certain iterations. In machine learning, the phenomena occurred in the cases of blade pitch angle and wind speed & wind shear are called "overfitting". Overfitting occurs when the used machine learning algorithm fits well to the training dataset while the predicting model has a very hard time to be generalized to testing/validation data. These phenomena happened in the scenarios of low bias and high variance (see **Eq. (6)**), which sends high total errors. Comparing with correlation coefficients that were presented in session 3.2, the most related feature has been regarded as blade pitch angle instead of wind speed. As stated in session 3.2, at most of the operating time, the local wind speeds were over or close to the rated wind speed (10.9 m/s), which may reduce the influence from wind speed on power predictions. In this scenario, the blade pitch angle becomes more significant for power generation, as wind turbines often use this feature to regulate the rotation speed and the generated power. The current correlations are considered to be more precise, in which non-linear relationships were deliberated through analysing deep learning features.

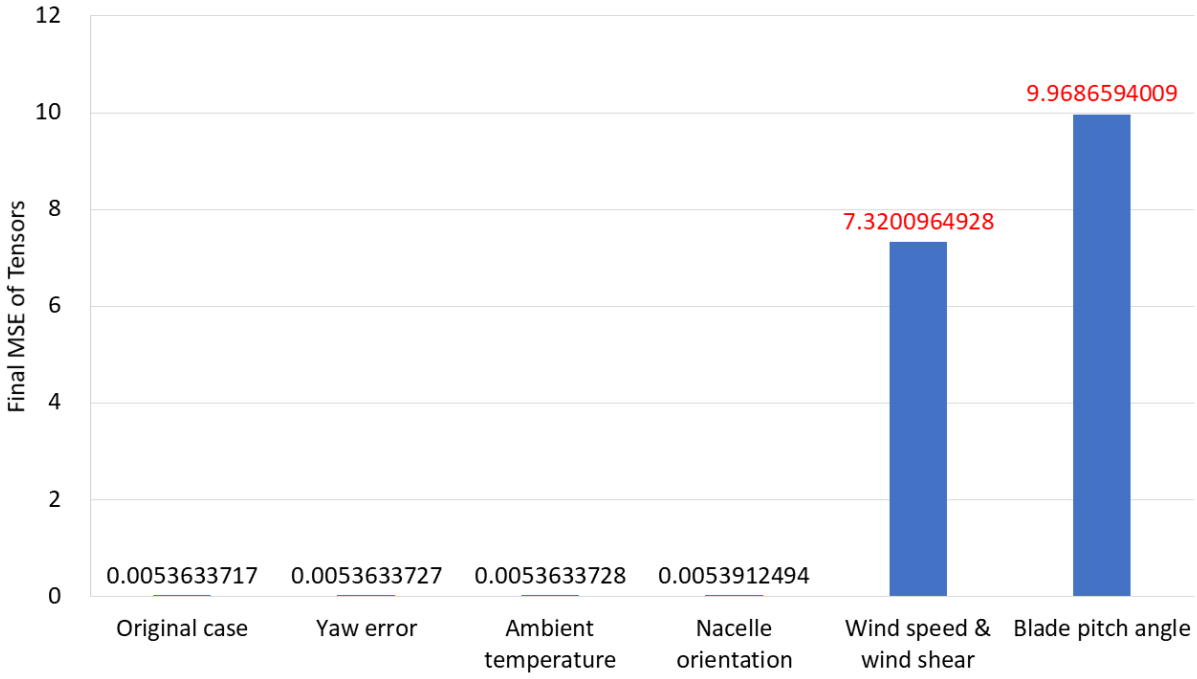

Fig. 9 – Variations of final MSEs in different feature scenarios.

The accuracies of the built deep learning model under various scenarios are presented in **Table 4** by various metric functions, where all the functions reach a consistent agreement. The best results were kept by the original case. The worst scenario was observed in the case of blade pitch angle and wind speed & wind shear (highlighted as red in **Table 4**), where *R-square* and *EVS* were displayed as negative values. It indicated that the deep learning neural network has become an arbitrarily worse predicting model without the considerations of blade pitch angle or wind speed & wind shear. As stated in session 3.2, at most of the operating time, the active power was nearing to the rated power, indicating the pitch angle control provide a nonnegligible impact on wind power generations of the target wind turbine. On the other hand, as wind speed and wind profile governed power generations, it is self-evidently significant for wind power predictions. Furthermore, the influences from ambient temperature, nacelle orientation and yaw error on the accuracies of the deep learning predictive model could be approximately ignored. The ambient temperature in our study is mainly representing the impact of air density variations. Some investigations have proposed that the influence of air density variations to final energy output is relatively small. For instance, Jung and Schindler claimed that a wind energy yield variation of 0.7% was observed in wind resource assessment under the consideration of air density [49]. The features of nacelle orientation and yaw error also offered minor influences on the power predicting results. A combination of those two features can be considered as wind direction. The well-functioning of the yaw system of the target turbine has satisfactorily orientated the wind turbine rotor towards the wind, minimizing the influence from wind directions.

Table 4 – Performance of the deep learning model by examining different feature structures in the validation loop.

	RMSE	R^2	*MAE*	*EVS*	*MSLE*
Original case	517.33	0.91	374.41	0.91	0.29
Yaw error	517.33	0.91	374.08	0.91	0.29
Ambient temperature	517.34	0.91	374.51	0.91	0.29
Nacelle orientation	518.68	0.91	377.05	0.91	0.28
Wind speed & wind shear	19112.13	-120.28	17668.17	-16.65	8.91
Blade pitch angle	22303.30	-164.17	21586.84	-9.44	9.85

5.3 Robust with respect to feature dimension reduction

As ambient temperature, nacelle orientation and yaw error contributed minor influences on the predicting results in the designed deep learning model, it is recommended to remove these three features from the model. In this approach, the computational cost and time can be further reduced for wind power predictions while retaining high accuracy. The accuracy and CPU time of the deep learning models in the original and reduced cases (without ambient temperature, nacelle orientation and yaw error) are displayed in **Table 5**, respectively. As can be seen, even though the dimensions of input features have been reduced from 11 to 8, the reduced model performed similarly in comparison with the original model with 2% less CPU time on a Windows PC (Intel ® Core(TM) i7-8700 CPU @ 3.20GHz, 16GB RAM). Although this may not seem a substantial reduction in processing time, it has to be considered that this is relative to a single wind turbine instead of an entire wind farm. Actually, there are around 100-200 wind turbines can be placed in an offshore wind farm in the UK. This temporal advantage of processing time for one single wind turbine can be translated into a much bigger computational time reduction for a farm with hundreds of wind turbines.

Table 5 – Performance and CPU time of deep learning models with original and reduced features.

	RMSE	R^2	*MAE*	*EVS*	*MSLE*	*CPU time in second*
Original case	517.33	0.91	374.41	0.91	0.29	2425.72
Reduced case	545.28	0.90	401.05	0.91	0.33	2378.96

6. Conclusions

In this paper, a deep learning neural network model was constructed to forecast wind power generations for a 7MW offshore wind turbine, which were trained, tested and validated through a high-frequency SCADA database. Unlike conventional methods, this model used four wind speeds at different heights, three measured pitch angles of each blade, average blade pitch angle, nacelle orientation, yaw error, and ambient temperature as input features in the predictive model. Besides, we developed a novel methodology to investigate non-linear correlations between input features and wind power outputs through deep learning neural networks. The methodology applied here is general and can be utilized to other wind turbines or upscaled to wind farms. Based on the facts above, this study has the following conclusions:

- The non-linear correlations regarding how input features influence wind power forecasting can be quantitatively evaluated through the designed deep learning model. Simulation results showed, the level of significance of blade pitch angles on the predictive model is ranking as the first among all the features, which is even higher than wind speed & wind shear in our case studies. It is concluded that blade pitch angles were essential for high power generation when wind speeds are higher than the rated values.
- On the other hand, the level of significance of wind direction and air density, which was represented by nacelle orientation & yaw error and ambient temperature in inputs, was considered as the lowest among the eleven features. Therefore, it is recommended to remove these three features from the predictive model with the respect of feature dimension reductions. As a result, the computational cost and time of the predictive model were further decreased for wind power predictions while retaining high accuracy.
- For reduced deep learning models, the adoption of feature dimension reductions resulted in a slight saving of processing time (0.77 minutes) for a single wind turbine. It may not be significant reduction under a single wind turbine condition, but when considering a typical wind farm, typically consisting of 100 ~ 200 turbines in the UK, the saved simulation time can be sizeable.

Acknowledgement

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

References

[1] Caglayan DG, Ryberg DS, Heinrichs H, Linßen J, Stolten D, Robinius M. The techno-economic potential of offshore wind energy with optimized future turbine designs in Europe. Applied Energy 2019;255:113794. doi:10.1016/j.apenergy.2019.113794.

[2] Wang X, Zeng X, Yang X, Li J. Feasibility study of offshore wind turbines with hybrid monopile foundation based on centrifuge modeling. Applied Energy 2018;209:127–39. doi:10.1016/j.apenergy.2017.10.107.

[3] Dai J, Tan Y, Shen X. Investigation of energy output in mountain wind farm using multiple-units SCADA data. Applied Energy 2019;239:225–38. doi:10.1016/j.apenergy.2019.01.207.

[4] Wang X, Zeng X, Yang X, Li J. Seismic response of offshore wind turbine with hybrid monopile foundation based on centrifuge modelling. Applied Energy 2019;235:1335–50. doi:10.1016/j.apenergy.2018.11.057.

[5] Chen J, Wang F, Stelson KA. A mathematical approach to minimizing the cost of energy for large utility wind turbines. Applied Energy 2018;228:1413–22. doi:10.1016/j.apenergy.2018.06.150.

[6] Yin X, Zhao X. Big data driven multi-objective predictions for offshore wind farm based on machine learning algorithms. Energy 2019;186:115704. doi:10.1016/j.energy.2019.07.034.

[7] Castellani F, Astolfi D, Sdringola P, Proietti S, Terzi L. Analyzing wind turbine directional behavior: SCADA data mining techniques for efficiency and power assessment. Applied Energy 2017;185:1076–86. doi:10.1016/j.apenergy.2015.12.049.

[8] Wang Y, Hu Q, Meng D, Zhu P. Deterministic and probabilistic wind power forecasting using a variational Bayesian-based adaptive robust multi-kernel regression model. Applied Energy 2017;208:1097–112. doi:10.1016/j.apenergy.2017.09.043.

[9] Arun Kumar SVV, Nagababu G, Kumar R. Comparative study of offshore winds and wind energy production derived from multiple scatterometers and met buoys. Energy 2019;185:599–611. doi:10.1016/j.energy.2019.07.064.

[10] Pelletier F, Masson C, Tahan A. Wind turbine power curve modelling using artificial neural network. Renewable Energy 2016;89:207–14. doi:10.1016/j.renene.2015.11.065.

[11] De Giorgi MG, Ficarella A, Tarantino M. Assessment of the benefits of numerical weather predictions in wind power

forecasting based on statistical methods. Energy 2011;36:3968–78. doi:10.1016/j.energy.2011.05.006.
[12] Ciulla G, D'Amico A, Di Dio V, Lo Brano V. Modelling and analysis of real-world wind turbine power curves: Assessing deviations from nominal curve by neural networks. Renewable Energy 2019;140:477–92. doi:10.1016/j.renene.2019.03.075.
[13] Kou P, Gao F, Guan X. Sparse online warped Gaussian process for wind power probabilistic forecasting. Applied Energy 2013;108:410–28. doi:10.1016/j.apenergy.2013.03.038.
[14] Xu L, Mao J. Short-term wind power forecasting based on Elman neural network with particle swarm optimization. Proceedings of the 28th Chinese Control and Decision Conference, CCDC 2016, 2016, p. 2678–81. doi:10.1109/CCDC.2016.7531436.
[15] Bilal B, Ndongo M, Adjallah KH, Sava A, Kebe CMF, Ndiaye PA, et al. Wind turbine power output prediction model design based on artificial neural networks and climatic spatiotemporal data. Proceedings of the IEEE International Conference on Industrial Technology, Institute of Electrical and Electronics Engineers Inc.; 2018, p. 1085–92. doi:10.1109/ICIT.2018.8352329.
[16] Li T, Li Y, Liao M, Wang W, Zeng C. A New Wind Power Forecasting Approach Based on Conjugated Gradient Neural Network. Mathematical Problems in Engineering 2016;2016:1–8. doi:10.1155/2016/8141790.
[17] Jyothi MN, Rao PVR. Very-short term wind power forecasting through Adaptive Wavelet Neural Network. 2016 - Biennial International Conference on Power and Energy Systems: Towards Sustainable Energy, PESTSE 2016, IEEE; 2016. doi:10.1109/PESTSE.2016.7516513.
[18] Gilbert C, Messner JW, Pinson P, Trombe PJ, Verzijlbergh R, van Dorp P, et al. Statistical post-processing of turbulence-resolving weather forecasts for offshore wind power forecasting. Wind Energy 2020;23:884–97. doi:10.1002/we.2456.
[19] Zhao P, Wang J, Xia J, Dai Y, Sheng Y, Yue J. Performance evaluation and accuracy enhancement of a day-ahead wind power forecasting system in China. Renewable Energy 2012;43:234–41. doi:10.1016/j.renene.2011.11.051.
[20] Liu J, Wang X, Lu Y. A novel hybrid methodology for short-term wind power forecasting based on adaptive neuro-fuzzy inference system. Renewable Energy 2017;103:620–9. doi:10.1016/j.renene.2016.10.074.
[21] Singh S, Bhatti TS, Kothari DP. Wind power estimation using artificial neural network. Journal of Energy Engineering 2007;133:46–52. doi:10.1061/(ASCE)0733-9402(2007)133:1(46).
[22] Carolin Mabel M, Fernandez E. Analysis of wind power generation and prediction using ANN: A case study. Renewable Energy 2008;33:986–92. doi:10.1016/j.renene.2007.06.013.
[23] Jafarian M, Ranjbar AM. Fuzzy modeling techniques and artificial neural networks to estimate annual energy output of a wind turbine. Renewable Energy 2010;35:2008–14. doi:10.1016/j.renene.2010.02.001.
[24] Peng H, Liu F, Yang X. A hybrid strategy of short term wind power prediction. Renewable Energy 2013;50:590–5. doi:10.1016/j.renene.2012.07.022.
[25] Zameer A, Arshad J, Khan A, Raja MAZ. Intelligent and robust prediction of short term wind power using genetic programming based ensemble of neural networks. Energy Conversion and Management 2017;134:361–72. doi:10.1016/j.enconman.2016.12.032.
[26] Zhang J, Yan J, Infield D, Liu Y, Lien F sang. Short-term forecasting and uncertainty analysis of wind turbine power based on long short-term memory network and Gaussian mixture model. Applied Energy 2019;241:229–44. doi:10.1016/j.apenergy.2019.03.044.
[27] Telmoudi AJ, Tlijani H, Nabli L, Ali M, M'Hiri R. A New rbf neural network for prediction in industrial control. International Journal of Information Technology and Decision Making 2012;11:749–75. doi:10.1142/S0219622012500198.
[28] Marugán AP, Márquez FPG, Perez JMP, Ruiz-Hernández D. A survey of artificial neural network in wind energy systems. Applied Energy 2018;228:1822–36. doi:10.1016/j.apenergy.2018.07.084.
[29] Hong YY, Rioflorido CLPP. A hybrid deep learning-based neural network for 24-h ahead wind power forecasting. Applied Energy 2019;250:530–9. doi:10.1016/j.apenergy.2019.05.044.
[30] Heier S. Grid Integration of Wind Energy Conversion Systems. 2nd Revise. John Wiley and Sons Ltd; 2006.
[31] Harrison R, Hau E, Snel H. Large Wind Turbines - Design and Economics. John Wiley & Sons; 2001.
[32] BSI Standards Publication. Wind power generation systems; Part 12-1: Power performance measurement of electricity producing wind turbines (IEC 61400-12-1:2017). 2017.
[33] Song D, Fan X, Yang J, Liu A, Chen S, Joo YH. Power extraction efficiency optimization of horizontal-axis wind turbines through optimizing control parameters of yaw control systems using an intelligent method. Applied Energy 2018;224:267–79. doi:10.1016/j.apenergy.2018.04.114.
[34] Shu ZR, Li QS, Chan PW. Investigation of offshore wind energy potential in Hong Kong based on Weibull distribution function. Applied Energy 2015;156:362–73. doi:10.1016/j.apenergy.2015.07.027.
[35] Shu ZR, Li QS, Chan PW. Statistical analysis of wind characteristics and wind energy potential in Hong Kong. Energy Conversion and Management 2015;101:644–57. doi:10.1016/j.enconman.2015.05.070.
[36] Lydia M, Kumar SS, Selvakumar AI, Prem Kumar GE. A comprehensive review on wind turbine power curve

modeling techniques. Renewable and Sustainable Energy Reviews 2014;30:452–60. doi:10.1016/j.rser.2013.10.030.
[37] Banuelos-Ruedas F, Angeles-Camacho C, Rios-Marcuello S. Analysis and validation of the methodology used in the extrapolation of wind speed data at different heights. Renewable and Sustainable Energy Reviews 2010;14:2383–91. doi:10.1016/j.rser.2010.05.001.
[38] Mittelmeier N, Kühn M. Determination of optimal wind turbine alignment into the wind and detection of alignment changes with SCADA data. Wind Energy Science 2018;3:395–408. doi:10.5194/wes-3-395-2018.
[39] Serret J, Rodriguez C, Tezdogan T, Stratford T, Thies P. Code comparison of a NREL-fast model of the levenmouth wind turbine with the GH bladed commissioning results. Proceedings of the International Conference on Offshore Mechanics and Arctic Engineering - OMAE, 2018. doi:10.1115/OMAE2018-77495.
[40] Lin Z, Liu X, Collu M. Wind power prediction based on high-frequency SCADA data along with isolation forest and deep learning neural networks. International Journal of Electrical Power and Energy Systems 2020;118:105835. doi:10.1016/j.ijepes.2020.105835.
[41] Aydiner C, Demir I, Yildiz E. Modeling of flux decline in crossflow microfiltration using neural networks: The case of phosphate removal. Journal of Membrane Science 2005;248:53–62. doi:10.1016/j.memsci.2004.07.036.
[42] Luo X, Sun J, Wang L, Wang W, Zhao W, Wu J, et al. Short-term wind speed forecasting via stacked extreme learning machine with generalized correntropy. IEEE Transactions on Industrial Informatics 2018;14:4963–71. doi:10.1109/TII.2018.2854549.
[43] Sideratos G, Hatziargyriou ND. An advanced statistical method for wind power forecasting. IEEE Transactions on Power Systems 2007;22:258–65. doi:10.1109/TPWRS.2006.889078.
[44] Zhao Y, Ye L, Pinson P, Tang Y, Lu P. Correlation-Constrained and Sparsity-Controlled Vector Autoregressive Model for Spatio-Temporal Wind Power Forecasting. IEEE Transactions on Power Systems 2018;33:5029–40. doi:10.1109/TPWRS.2018.2794450.
[45] Chen N, Qian Z, Nabney IT, Meng X. Wind power forecasts using gaussian processes and numerical weather prediction. IEEE Transactions on Power Systems 2014;29:656–65. doi:10.1109/TPWRS.2013.2282366.
[46] Mashaly AF, Alazba AA, Al-Awaadh AM, Mattar MA. Predictive model for assessing and optimizing solar still performance using artificial neural network under hyper arid environment. Solar Energy 2015;118:41–58. doi:10.1016/j.solener.2015.05.013.
[47] Roostaei AA, Pahlevanpour A, Behravesh SB, Jahed H. On the definition of elastic strain energy density in fatigue modelling. International Journal of Fatigue 2019;121:237–42. doi:10.1016/j.ijfatigue.2018.12.011.
[48] Huijbregts MAJ, Rombouts LJA, Hellweg S, Frischknecht R, Hendriks AJ, Van De Meent D, et al. Is cumulative fossil energy demand a useful indicator for the environmental performance of products? Environmental Science and Technology 2006;40:641–8. doi:10.1021/es051689g.
[49] Jung C, Schindler D. The role of air density in wind energy assessment - A case study from Germany. Energy 2019;171:385–92. doi:10.1016/j.energy.2019.01.041.