Short–Term Wind Power Forecasting Using Artificial Neural Networks-Based Ensemble Model

Prepared by:

Qin Chen

CHNQIN002

Department of Electrical Engineering
University of Cape Town

Supervisor:

Komla Folly

Department of Electrical Engineering
University of Cape Town

February 2020

Submitted to the Department of Electrical Engineering at the University of Cape Town in partial fulfilment of the academic requirements for a Master of Science degree in Electrical Engineering.

Key Words: Artificial intelligence, artificial neural networks, ensemble model, particle swarm optimization, spatial correlation, wind speed forecasting, wind power forecasting

Declaration

I know the meaning of plagiarism and declare that all the work in the document, save for that which is properly acknowledged, is my own. This thesis/dissertation has been submitted to the Turnitin module (or equivalent similarity and originality checking software) and I confirm that my supervisor has seen my report and any concerns revealed by such have been resolved with my supervisor.

Name: **Qin Chen**

Signature: Signed by candidate **Date:** 2020/10/02

Acknowledgements

First and foremost, I would like to express my sincere gratitude to my supervisor, Professor K.A. Folly, for his useful advice, comments, and support throughout the learning process of this dissertation.

Many thanks to all staff at the University of Cape Town for the roles they played throughout my postgraduate education.

Furthermore, I would like to thank the National Research Foundation (NRF) for proving the financial assistance towards this research.

I would like to thank Sotavento Galicia Wind Farm, National Renewable Energy Laboratory and Wind Atlas of South Africa for providing me with the very useful and quality wind power and meteorological data.

Last but not least, I would like to thank my friends and family, particularly my parents, for their encouragement and unconditional support.

The financial assistance of the National Research Foundation (NRF) towards this research is hereby acknowledged. This research is partially supported financially in part by the NRF. Opinions expressed and conclusions arrived at, are those of the author and are not necessarily to be attributed to the NRF.

Abstract

Short-term wind power forecasting is crucial for the efficient operation of power systems with high wind power penetration. Many forecasting approaches have been developed in the past to forecast short-term wind power. In recent years, artificial neural network-based approaches (ANNs) have been one of the most effective and popular approaches for short-term wind power forecasting because of the availability of large amounts of historical data and strong computational power. Although ANNs usually perform well for short-term wind power forecasting, further improvement can be obtained by selecting suitable input features, model parameters, and using forecasting techniques like spatial correlation and ensemble for ANNs.

In this research, the effect of input features, model parameters, spatial correlation and ensemble techniques on short-term wind power forecasting performance of the ANNs models was evaluated. Pearson correlation coefficients between wind speed and other meteorological variables, together with a basic ANN model, were used to determine the impact of different input features on the forecasting performance of the ANNs. The effect of training sample resolution and training sample size on the forecasting performance was also investigated. To separately investigate the impact of the number of hidden layers and the number of hidden neurons on short-term wind power forecasting and to keep a single variable for each experiment, the same number of hidden neurons was used in each hidden layer. The ANNs with a total of 20 hidden neurons are shown to be sufficient for the nonlinear multivariate wind power forecasting problems faced in this dissertation. The ANNs with two hidden layers performed better than the one with a single hidden layer because additional hidden layer adds nonlinearity to the model. However, the ANNs with more than two hidden layers have the same or worse forecasting performance than the one with two hidden layers. ANNs with too many hidden layers and hidden neurons can overfit the training data. Spatial correlation technique was used to include meteorological variables from highly correlated neighbouring stations as input features to provide more surrounding information to the ANNs. The advantages of input features, model parameters, and spatial correlation and ensemble techniques were combined to form an ANN-based ensemble model to further enhance the forecasting performance from an individual ANN model.

The simulation results show that all the available meteorological variables have different levels of impact on forecasting performance. Wind speed has the most significant impact on both short-term wind speed and wind power forecasting, whereas air temperature, barometric pressure, and air density have the smallest effects. The ANNs perform better with a higher data resolution and a significantly larger training sample size. However, one requires more computational power and a longer training time to train the model with a higher data resolution and a larger training sample size. Using the meteorological variables from highly related neighbouring stations do significantly improve the forecasting accuracy of target stations. It is shown that an ANNs-based ensemble model can further enhance the forecasting performance of an individual ANN by obtaining a large amount of surrounding meteorological information in parallel without encountering the overfitting issue faced by a single ANN model.

Table of contents

List of figures

List of tables

Glossary

Abbreviations

ACF	Autocorrelation Function
AD	Air Density
ANFIS	Adaptive Network-Based Fuzzy Inference System
ANNs	Artificial Neural Networks
AR	Autoregressive
ARIMA	Autoregressive Integrated Moving Average
ARMA	Autoregressive Moving Average
GA	Genetic Algorithm
HIRLAM	High-Resolution Limited Area Model
KNN	K-Nearest Neighbours
LM	Levenberg-Marquardt
LSTM	Long Short-Term Memory
MA	Moving average
MAE	Mean Absolute Error
MAPE	Mean Absolute Percentage Error
MM5	Fifth-Generation Penn State University and the National Centre for Atmospheric Research Mesoscale Model
MOS	Model Output Statistics
MSE	Mean Square Error
NMAE	Normalised Mean Absolute Error
NREL	National Renewable Energy Laboratory
NRMSE	Normalised Root Mean Square Error
NWP	Numerical Weather Prediction
P	Barometric Pressure
PACF	Partial Autocorrelation Function
PI	Prediction Interval
PSO	Particle Swarm Optimization
R	Correlation Coefficient
RBF	Radial Basis Function
REI4P	Renewable Energy Independent Power Producers Procurement Programme of South Africa
RH	Relative Humidity
RMSE	Root Mean Square Error
SDE	Standard Deviation of Error
SGWF	Sotavento Galicia Wind Farm
SVM	Support Vector Machine
SVR	Support Vector Regression
T	Air Temperature
TG	Temperature Gradient
TI	Time Indicator
WASA	Wind Atlas of South Africa
WD	Wind Direction
WP	Wind Power
WRF	Weather Research and Forecasting
WS	Wind Speed

WT	Wavelet Transform

Mathematical Notation and Symbols

A	Sweep area of turbine blades
α_i	Coefficients of autoregressive
b_i	Bias
β_j	Coefficients of moving average
c	Scalar multiplier
c_1	Acceleration constant
c_2	Acceleration constant
C_p	Power coefficient
D_p	Particle dimension
e	Network errors
e_w	White noise
f_i	Activation function
GW	Gigawatt
I	Identity matrix
J	Jacobian matrix
kW	Kilowatt
MW	Megawatt
μ	Damping factor
μ_x	Mean of X
N	Number of observations
P_{av}	Available wind power
P_{pr}	Practical wind power
ρ	Air density
p_{gd}	The best position of a global particle
p_i	The best position of an individual particle
R_d	Gas constant
r_1	Randomly generated values ranging from 0 to 1
r_2	Randomly generated values ranging from 0 to 1
$\hat{\sigma}_h$	Estimation of the standard deviation of the h-step forecast distribution
σ_x	Standard deviation of X
T_{max}	Maximum iteration
Δt	Time horizon
V	Wind Speed
v_i	Particle velocity
W	Watt
W^{ij}	Connection weight
x_i	Current particle position
ω	Inertia weight

1. Introduction

This chapter presents an overview of the research background and motivation; the research objectives; the research questions that need to be investigated; and the limitations to this research.

1.1 Research background and motivation

With the depletion of conventional energy resources, fast-growing energy demand and degradation of the environment in recent decades, renewables are becoming one of the most critical energy sources [1]. In recent years, the conventional fossil fuels which have severe impacts on the global ecological systems have been gradually replaced by renewable energy [2]. Although nuclear power plants can produce a tremendous amount of power and emit relatively little carbon dioxide compares to fossil fuels, some countries are eliminating nuclear power from their energy mix portfolio due to some nuclear power plant incidents and disasters (i.e., Fukushima Daiichi and Chernobyl nuclear disaster) and their long-term effects on the ecology environment [3]. Other forms of renewables such as solar, wind energy, hydroelectric, geothermal, ocean, hydrogen, and biomass are safer than nuclear power and cleaner than fossil fuels. Among all the renewables, wind power is one of the most available, affordable, and efficient renewable sources.

The share of renewable energy in the global energy mix has increased rapidly in recent years. By the end of June 2019, worldwide nameplate capacity of the installed wind turbines has reached 597 GW which is eight times as many as the 74.2 GW installed by the end of 2006 [4], [5]. In 2018 alone, 50.1GW nameplate capacity of wind turbines was installed [5]. Wind power has become an alternative to fossil fuels in various countries [6]. The South African government also aims to increase the share of renewable energy in its energy mix. The decommissioning of the existing ageing coal power plants provides space for renewable energy into the current energy mix. Renewable Energy Independent Power Producers Procurement Programme (REI4P) of South Africa aims to install 17.8 GW of electricity generation capacity from renewables over the period 2012–2030 [7]. Already, the REI4P has allocated 6.36 GW of generation capacity to onshore wind [8].

With high wind power penetration, wind power will provide not only remarkable economic and environmental benefits but also will bring many challenges to the power system due to its stochastic and intermittent nature [4], [9], [10]. Power systems with high wind penetration need to deal with many challenges such as real-time grid operations, competitive market designs, stability and reliability of power system, and et cetera. [11]. Variations in wind power generation will reduce operational efficiency and increase regulation requirement if there is a mismatch between wind power generation and load demand [4]. As an example, in Denmark in January 2005, an extremely high wind speed that was not accurately forecasted caused a power imbalance of more than 2000 MW between the actual and forecasted wind generation which resulted in a significant economic loss [2]. More accurate wind power forecasting models can be used to reduce the power imbalance between the actual wind generation and forecasted wind generation. Accurate short-term wind power forecasting can be utilised to identify the wind power fluctuation in advance to mitigate the impact of wind intermittency on the power system [11], [12].

Improving the accuracy of wind speed forecasting has significant technological and economic impacts on wind farms and power grids planning [12]. Various wind power forecasting approaches have been proposed to improve short-term wind power forecasting accuracy in the literature. These approaches can be classified into four main groups, namely physical approaches, statistical approaches, artificial intelligence approaches and hybrid approaches. Physical approaches are mainly used for long-term

forecasting, while statistical and artificial intelligence are suitable for short-term forecasting. The main focus of the study in this dissertation is on short-term wind speed and wind power forecasting. Hence, statistical approaches and artificial intelligence approaches are considered. Statistical approaches like autoregressive integrated moving average (ARIMA), autoregressive moving average (ARMA), and Kalman filters are good at mapping the linear relationship between inputs and outputs. However, the relationships between wind speed and the historical wind speed or other meteorological variables are nonlinear. Therefore, statistical approaches are not the best option for nonlinear time series. On the contrary, artificial intelligence approaches can determine the nonlinear relationship between inputs and outputs. With the improvement of computational power and availability of historical wind data, artificial intelligence models have become very popular choices for short-term wind power and speed forecasting in recent years.

Some artificial intelligence-based wind power forecasting approaches have been used for various wind farms around the globe to forecast short-term wind power generation. Artificial neural network-based models (ANNs) are one of the most efficient and popular models for nonlinear time series problems. It has been shown that ANNs perform better than ARIMA. However, there is still a high level of inaccuracy within the forecasting results of ANNs. Many factors such as model structure, model parameter, data size and resolution, and input features can influence the forecasting performance of ANNs. The motivation of this study is to further improve the forecasting performance of ANNs by determining the influential factors of ANNs and utilising these factors to form an ensemble model.

The contribution of this dissertation is to investigate the impact of training data size and resolution, input features from target station and neighbouring stations, model structure and parameters on the short-term wind power forecasting performance of ANNs. The most important factors and configurations are then combined to form different types of ensemble models. The short-term wind speed and wind power forecasting performance of each experiment is evaluated by using three commonly used evaluation metrics; root mean square error (RMSE), mean absolute error (MAE) and mean absolute percentage error (MAPE). The ensemble model with the best forecasting performance is selected as the final model.

1.2 Objectives of this study

1.2.1 Research objectives

The focus of this research is to enhance the short-term wind speed and wind power forecasting performance of the artificial neuron networks (ANNs) by determining the effect of different factors on the forecasting performance of ANNs. Hence, the objectives of this research are:

- To investigate how the factors like model parameters, input features, training sample size and data resolution etc. affect the forecasting performance of ANNs.

- To carry out a spatial correlation analysis to determine if input features from neighbouring sites have impacts on the forecasting performance of ANNs of a target site.

- To find out if ANN-based ensemble models can further enhance the short-term wind speed and wind power forecasting performance from the best individual ANN model.

1.2.2 Research question to be investigated

- How do input features, training sample size, data resolution, and forecasting model configurations affect the forecasting performance of ANNs?

- Can input features from neighbouring sites enhance the short-term wind speed and wind power forecasting performance of the ANNs of a target site?

- Can an ANN-based ensemble model improve the short-term wind speed and wind power forecasting performance from the best individual ANN model? Which types of ANN-based ensemble models have the most significant impact on forecasting performance?

1.3 Scope and limitations

The scope of this study is to investigate the impact of some important factors on the short-term wind speed and wind power forecasting performance of ANN-based models. Evaluating the forecasting performance of feedforward neural network-based models and ensemble models are part of this research. Wind data collection, processing and analysis are within the scope of this study. Other types of computational intelligence models can be are out of the scope for this research. This research is only focused on some of the important factors and techniques that can potentially enhance the forecasting performance of the ANN-based model.

The limitation of this dissertation is the lack of actual wind power generation data from South African wind farms. The dataset of Wind Atlas of South Africa (WASA) contains many meteorological variables but without real wind power generation data. On the other hand, the dataset of Sotavento Galicia wind farm of Spain contains wind power generation data for the entire wind power. Still, it lacks the wind power data of individual wind turbine and some of the meteorological variables such as temperature, atmospheric pressure and relative humidity. The wind power data provided by the National Renewable Energy Laboratory (NREL) are simulated data from weather research and forecasting (WRF) model.

1.4 Plan of development

The rest of this dissertation is organized as follows:

Chapter 2 presents literature reviews on the fundamental knowledge required for wind power forecasting. Time-scale classification, data pre-processing, commonly used evaluation metrics and short-term wind forecasting methods are reviewed in this chapter.

Chapter 3 describes the implementation steps of short-term wind speed and wind power forecasting in details. Data collection and pre-processing are carried out before modelling several experiments to investigate the impact of different factors on the forecasting performance of the models.

Chapter 4 presents and evaluates the simulation results of the single model approach using three commonly used evaluation metrics, root mean square error (RMSE), mean absolute error (MAE) and mean absolute percentage error (MAPE).

Chapter 5 shows and discusses the simulation results of the input-feature-based ensemble model and the model-parameter-based ensemble model.

Chapter 6 summarises the factors and techniques used in this dissertation to enhance the short-term wind speed and wind power forecasting performance of ANNs.

Chapter 7 provides several recommendations for further studies.

2. Literature review

The purpose of this chapter is to provide a review of relevant literature to short-term wind power forecasting. The first section of this chapter presents the key basis of short-term wind power forecasting like time-scale classification, training data pre-processing and wind speed to wind power conversion. An overview of commonly used forecasting performance evaluation metric is presented along with the benchmark model, persistence model, in the first section. The section concludes with a summary of different types of forecasting. The second part of the chapter presents an overview of the existing forecasting models, which can be classified based on input features and forecasting time horizons. The remaining sections of the chapter discuss a model from each class of wind power forecasting model.

2.1 Basis of wind power forecasting

Wind power forecasting can generally consist of several steps. The first step is to classify and select the time horizon for wind power forecasting. Data collection and pre-processing can be carried out after the forecasting time horizon and methods are chosen. The impact of each input feature on the performance of forecasting models can be determined by using correlation analysis and sensitivity analysis in the pre-processing data stage. The model parameters and structure can be selected after the input features are decided. The next step is to convert the forecasted wind speed to wind power. Finally, the performance of forecasting models can be evaluated and compared to other forecasting models by using evaluation metrics.

2.1.1 Time-scale classification

Different authors classify the timescale of wind power forecasting differently. Many authors suggested that timescale for the operation of an electrical system should be divided into four categories, namely very-short-term ranging from few seconds to 30 minutes ahead, short-term from 30 minutes to six hours ahead, medium-term from six hours to one day ahead and long-term from more than one day ahead [6], [11], [13]. Although many authors have agreed with the above time-scale classification, some authors have different opinions about the range of each time horizon [1], [2], [14]. For example, authors in [2] suggested that very-short-term should range from few minutes to one hour ahead, short-term from one hour to several hours ahead, medium-term from several hours to one week ahead and long-term is defined as more than one week ahead. Authors in [1], [14] on the other hand suggested that short-term should range from 30 minutes to 48 hours ahead, medium-term from 48 hours to one week ahead and long-term from more than one week ahead. Based on the opinions of the above-mentioned authors, the time-scale classification adopted in this dissertation is summarised in Table 2.1.

Table 2.1 summaries applications and time range corresponding to each time horizon [11], [15], [16], [17], [18]. As shown in Table 2.1, different forecasting time horizons have different time ranges and corresponding implications on the operation of electrical systems [19]. Load increment and decrement decisions need to be made in a short-term time horizon. It is expected that if accurate wind power forecasting models are used, cost of wind power and risk of the unreliability of electricity supply can be considerably reduced [17], [18]. With greater wind power penetration, it is crucial that accurate short-term wind speed and wind power forecasting be used by potential decision-makers such as wind power producer, utilities, market analysts and traders to make better load increment/decrement decisions which could improve energy market efficiency, maintain the system security and reduce the financial risk [16], [20], [21]. Short-term forecasting models and tools can also be utilised to predict other renewable energy sources to achieve intelligent management of a microgrid system [22]. Short-term wind power forecasting will be the main focus of this research.

Table 2.1: Summary of different ranges and applications of different forecasting time horizons

Time horizon	Range	Applications
Very short-term	Few seconds to 1 hour ahead	- Regulation actions - Electricity market clearing - Real-time grid operation (turbine control and load tracking)
Short-term	1 hour to 24 hours ahead	- Load increment/decrement decisions - Economic load dispatch planning - Operational security in the electricity market
Medium-term	24 hours to 1 week ahead	- Generator online/offline decisions - Unit commitment decisions - Reserve requirement decisions - Power system management
Long-term	More than one week ahead	- Maintenance scheduling to obtain optimal operating cost -Operation management -Feasibility study for the design of wind farms

2.1.2 Data pre-processing

The data pre-processing step has been utilised in some of the wind speed and wind power forecasting models to decrease the forecasting errors [23]. One way to improve the generalization capability and forecasting performance of the forecasting model is to pre-process the input data [11]. Authors in [24] classified data pre-processing techniques into five main categories which are: normalization [25], [26], discretization, feature extraction and selection, feature indexers and encoders, and text mining. Other data pre-processing techniques such as deseasonalization (i.e. centred moving average), denoising and correlation analysis are also widely used by different authors [4] [24]. Popular data pre-processing methods suitable for wind power forecasting analysis are presented below.

i. ***Non-stationarity of time series***

The wind power and wind speed time series have substantial non-stationarity property due to a high percentage of sudden changes and seasonality [4]. Linear and non-linear forecasting models such as autoregressive integrated moving average (ARIMA), ANNs, and other models cannot handle non-stationary input data without data pre-processing [27]. Deseasonalization is one of the techniques that can be used to overcome the non-stationarity of time series.

ii. ***Normalization***

Normalization technique can minimize the impact of one variable relative to another by scaling the values of different input variables into the same range, usually for 0 to 1 or -1 to 1. Most of the normalization methods involve subtraction of offset and scaling over the desired range [25]. Linear mapping over a specified range is the most commonly used method, which is defined as following [25], [26]:

$$X_i' = \left(\frac{X_i - X_{min}}{X_{max} - X_{min}}\right)(X'_{max} - X'_{min}) + X'_{min} \qquad (2.1)$$

where:

$X_i, X_{min}, X_{max}, X_i', X_{min}', X_{max}'$ are the mean, minimum and maximum of the actual input data and the normalized target values, respectively [26].

The output results must be de-normalised to obtain the actual forecast values [28].

iii. ***Feature selection***

Input feature selection has a significant influence on the performance of forecasting models [16]. Only the most influential input features should be selected to increase the generalization capability of forecasting models and to reduce computational cost and training time [13], [23], [24], [29]. A careful selection of input features can improve the performance of forecasting models considerably [30], [31].

Many variables can be used as input features to forecasting short-term wind speed and wind power. However, not all the variables have the same level of effect on the performance of forecasting models. Using all the available variables as input features for forecasting will increase computation complexity. Moreover, machine learning algorithms may overfit forecasting models or unable to recognize patterns from many input variables [30], [32]. The number of attributes in datasets can be reduced by using feature selection.

Feature selection methods can be classified into three groups, filter, wrapper, hybrid methods [13] [23]. Filter methods use statistical tools (i.e. autocorrelation, cross-correlation) to rank every feature. Correlation analysis and data decomposition are popular feature selection techniques. The correlation between an input feature and a target feature must be high. In contrast, the correlation between an input feature and other already selected features must be small to qualify as an essential input feature [29], [32]. Wrapper methods use a search strategy to add and remove features from datasets iteratively to find the most suitable features [13]. The hybrid method uses both filter and wrapper techniques. The hybrid feature selection method was used in this dissertation.

iv. ***Missing value, outliers, noisy data***

Missing values, outliers, noisy data also need to be solved in pre-processing data stage. Instances of missing values should not be discarded easily. Valuable information may be lost, which can bring bias in the training process [24]. Data polishing methods and noise filters like wavelet transform can be used to deal with noisy data.

2.1.3 Wind speed to wind power conversion

Wind turbine output has a cubic relationship with wind speed, which changes with the time of day, day of the year, regional landscape and weather patterns [11]. The theoretical power available (P_{av}) in the wind and the practical power (P_{pr}) that can be extracted from the wind can be estimated using Eq. (2.2) and Eq. (2.3), respectively [11], [33], [34].

$$P_{av}(v) = \frac{1}{2}\rho(t)Av^3 \qquad (2.2)$$

$$P_{pr}(v) = \frac{1}{2}\rho(t)Av^3C_p(v) \qquad (2.3)$$

where P_{av} is the theoretical power available in the wind and P_{pr} is the practical amount of power that can be generated by the wind turbine in watt (W), A is the total sweep area of the turbine blades in m^2, and v is the wind speed in m/s. C_p is the power coefficient which determines the amount of practical power that can be obtained from available theoretical power. The blade pitch angle and tip speed ratio

influence the power coefficient value, which has an upper limit, also known as the *Betz limit*. The value of the Betz limit is equal to $C_{p,max} = \frac{16}{27} \approx 0.593$, which cannot be achieved in practice. The practical value of C_p is normally around 0.5 [33].

The time-varying air density can be estimated by using ideal gas law under dry conditions [28].

$$\rho = \frac{P}{TR_d} \tag{2.4}$$

where $\rho(t)$ is site air density which depends on the site temperature T and atmospheric pressure P. R_d is the gas constant, which is equal to 0.287053 $kPaK^{-1}m^3kg^{-1}$.

As shown in Eq. (2.2), wind power and wind speed are related through a cubic relationship. Therefore, a small error in wind speed forecast will result in a sizeable error in wind power [11], [16], [34]. The forecasted wind speed can also be converted to the corresponding wind power by using power curves [35]. However, manufacturer certified power curves do not always generate accurate conversions [11]. Studies indicated that the model using the power curve derived from measured wind speed and power has better forecasting performance than the model using manufacture power curves only [11], [16], [36]. Many conventional and advanced statistical techniques such as polynomial regression, non-linear hyperbolic functions, ANN, fuzzy logic, and et cetera can also be used to model the nonlinear relation between wind turbine output and wind speed [16]. Both manufacturer power curve and the power curve derived from measured wind speed and power were used to convert the forecasted wind speed to wind power in this study.

2.1.4 Performance evaluations metrics

Uncertainty is one of the inherent properties of wind power forecasting. Therefore, the forecasting performance must be properly evaluated. It is also very crucial to quantify the forecasting performance on the untrained data to get a fair result [14]. Different evaluation methods provide different information for forecasting models. RMSE is more sensitive to large errors when compared to MAE. In RMSE, the errors are squared before they are averaged [37]. Therefore, RMSE has the benefit of penalising undesirable large errors and outliers more [10], [37], [38]. The results of RMSE will always be greater or equal to the results of MAE [10]. It is easier to interpret the MAE values than the RMSE values [37]. MAPE is another widely used evaluation metric which provides a relative overall fit measurement [10]. MAPE is preferred when the cost of the forecast error has a higher correlation with percentage error than numerical size error [25]. Commonly used forecasting performance validation methods are listed in Table 2.2 [4], [10], [14], [16], [27], [30], [34], [39], [40]- [52].

Errors can be both negative and positive, which can cancel each other out. It should be noted that all the evaluation metrics listed in Table 2.2 using either the absolute or squared value of errors to avoid the cancellation of errors.

Table 2.2: Summary of the commonly used forecasting performance evaluation metrics

Name	Equation
Root Mean Square Error (RMSE)	$RMSE = \sqrt{\frac{1}{N}\sum_{t=1}^{N}(e_t)^2}$
Normalised Root Mean Square Error (NRMSE)	$NRMSE = \frac{100}{P_{rate}}\sqrt{\frac{1}{N}\sum_{t=1}^{N}(e_t)^2}$
Mean Absolute Error (MAE)	$MAE = \frac{1}{N}\sum_{t=1}^{N}\lvert e_t\rvert$
Normalised Mean Absolute Error (NMAE)	$NMAE = \frac{100}{N}\sum_{t=1}^{N}\frac{\lvert e_t\rvert}{P_{rate}}$
Mean Absolute Percentage Error (MAPE)	$MAPE = \frac{100}{N}\sum_{t=1}^{N}\left\lvert\frac{e_{lt}}{P_t}\right\rvert$
Sum Squared Errors (SSE)	$SSE = \sum_{t=1}^{N}(e_t - \bar{e})^2$
Standard Deviation of Error (SDE)	$\text{SDE} = \sqrt{\frac{1}{N-1}\sum_{t=1}^{N}(e_t - \bar{e})^2}$
Mean Square Error (MSE)	$MSE = \frac{1}{N}\sum_{t=1}^{N}(e_t)^2$

Eq. (2.5) shows the formula of MAPE.

$$MAPE = \frac{100}{N}\sum_{t=1}^{N}\left|\frac{e_t}{P_t}\right| \qquad (2.5)$$

As can be seen in Eq. (2.5), P_t cannot be zero. However, wind turbine power output is equal to zero when cut-in wind speed is not achieved. Authors in [30], [42], [51], [53] have used Eq. (2.6) to avoid the zero term in the denominator.

$$MAPE = \frac{100}{N}\sum_{t=1}^{N}\left|\frac{e_t}{\bar{P}}\right| \qquad (2.6)$$

where

$$\bar{P} = \frac{1}{N}\sum_{t=1}^{N}P_t \qquad (2.7)$$

MSE has been often used in neural network training to improve the performance of the network due to its ability to magnify more significant forecasting errors [48]. MSE was used in this research to train ANNs.

The scaling of errors by the rated power is a standard way of scaling wind power predictions. Madsen et al. [54] recommended that the normalized RMSE, normalized MAE, and normalised bias should be used as a minimum set for standardised performance measure. RMSE, NRMSE, MAE, NMAE and MAPE (Eq. (2.7)) was used in this dissertation to evaluate forecasting performance.

2.1.5 Benchmark model: persistence model

The persistence model is a popular benchmark model [16]. Newly developed methods should be tested against the persistence method to determine the usefulness of the new methods [23], [42]. Persistence method is the simplest ARMA model which assumes that wind speed or wind power at time *'t'* will be the same as it was at the time *'t+Δt'* where *Δt* is forecasting time horizon [14], [18], [21]. If wind speed at the time 't' is *v(t)*, then wind speed at the time *'t+Δt'* can be formulated as:

$$v(t) = v(t+\Delta t) \qquad (2.8)$$

where v(t) is the wind speed at the time 't' and v(t+Δt) is the wind speed at the time 't+Δt'. As an example, *Δt* is equal to 2 hours for 2 hours ahead forecasting.

The persistence method is the most straightforward wind speed or power forecasting method due to its easy implementation [18]. However, the persistence method is mainly used for very-short-term forecasts [14], [21].

Chang [18] found that persistence method performs better than other forecasting methods in the very short-term forecasting. The accuracy of this method degrades quickly with increases of forecasting time horizon [55]. Authors in [2] mentioned that the persistence method should only be used as a benchmark model due to its inaccuracy for more than one hour ahead forecasting. Many authors have used persistence model as a benchmark model to evaluate the newly developed model. However, authors in [54] indicated that using persistence model as a reference model brings slightly overoptimistic conclusions about the newly developed methods. Therefore, the persistence model together with RMSE, MAE, and MAPE, will be used to evaluate the forecasting performance of forecasting models in this research.

2.1.6 Type of forecasting

i. *Deterministic vs probabilistic forecasting*

Generally, prediction can be separated into two main groups, namely deterministic (point forecasting) and probabilistic (interval forecasting). A deterministic forecast contains only the exact value of prediction at a future time. In contrast, probabilistic forecasting provides not only the point forecast but also the confidence interval of point forecasts [56], [57]. Most of the existing wind power forecasting methods are deterministic forecasts which cannot handle the uncertainties associated with datasets correctly [48]. The probabilistic forecasting method has become one of the recent areas of research due to its ability to reflect the uncertainty information about wind power generation [2], [43]. Probabilistic forecasts can be utilised to find an optimal reserve level to compensate for the imbalance caused by the variability of wind and solar power [52], [56]. Both point forecasting and interval forecasting were used in this dissertation.

ii. *Direct vs indirect forecasting*

In a direct wind power forecasting, wind power generation is calculated directly without using a manufacturer's power curve or the empirical curve of wind turbines to convert forecasted wind speed to wind power. In an indirect wind power forecasting, the first step is to predict wind speeds at wind farms and then turn the predicted wind speed to wind power using the power curve of wind turbines of Eq. (2.3) [58]. Authors in [57] suggested that converting wind speed forecasts to get wind power output is better than predicting wind power output directly because the spatial correlation of wind makes wind speed forecasting more accurate than wind power forecasting [57]. It is also more difficult to predict wind power than wind speed due to the discontinuity of wind power before cut-in and after cut-off [59]. In this research, wind speed is forecasted first and then converted to wind power.

iii. ***Independent and iterative method for multi-steps forecasting***

There are two ways to forecast multi-steps-ahead wind power, which are independent value prediction and iterative prediction. Independent value prediction model predicts multi-steps ahead wind power directly once to get a final output. One disadvantage of independent value prediction is the time gap between the input sample and the target sample of training data is significant. The iterative method iteratively trains the forecasting model one-step ahead at a time until achieves multi-steps ahead results. Forecasting error of each small step is small. However, one disadvantage of the iterative approach is that errors in the earlier stages might be transferred to the later stages [22]. It also takes longer to train the iterative model.

2.2 Overview of forecasting methods

Based on the input, most authors classified wind forecasting methods into three major categories, namely physical approaches, statistical approaches and hybrid approaches [2], [6], [9], [30], [36], [46], [60]. In addition to these categories, authors in [18], [61] also classified the persistence method, spatial correlation method and artificial intelligence methods such as artificial neural networks (ANNs) into separated classes. Authors in [21] classified the spatial correlation approach and artificial intelligence as additional categories to the physical, statistical and hybrid approaches. Authors in [43] classified ANNs as a learning approach instead of a statistical approach. Appendix C shows a summary of some of the existing forecasting methods.

The forecasting methods are selected based on the available information and the forecasting time horizon. Physical approaches do well in long-term prediction, while statistical approaches have advantages in short-term prediction [21], [22], [62], [63].

The statistical approaches establish the relationship between all the relevant historical input data and observation data without considering the physics of the system [6], [16], [20], [46], [64], [65]. They are based on patterns instead of any predefined mathematical equations [61]. In general, statistical models are easier and cheaper to develop compared with physical models [16], [35]. Moreover, statistical models have the option to use the outputs of NWP while physical models must use NWP models [10]. However, unlike physical approaches, statistical approaches require a large amount of historical data to train forecasting models [16]. In general, the forecasting performance of statistical models decreases as the forecasting time horizon increases [46].

Conventional statistical models are unable to track complex and non-linear wind speed patterns precisely [3]. Newly developed computational intelligence approaches such as artificial neural networks (ANNs), fuzzy logic methods, support vector machine (SVM), and et cetera are capable of handling non-linear and complicated relationships between input and target data [23]. ANNs are one of the most popular computational intelligence methods for short-term wind speed and wind power forecasting [16], [21]. ANN structure mimics the brain in two ways. First, knowledge is obtained from its environment by a learning process. The interneuron connection strength, also known as synaptic weights, store the obtained knowledge [3]. The main advantage of ANNs over statistical methods is that they can model a complicated non-linear relationship and extract the dependence between inputs and outputs through the learning process [3], [9], [16], [66]. Authors in [23], [37], [83] concluded that ANNs-based methods generally perform better than the conventional statistical models for wind speed forecasting. ANNs can also learn from imprecise and incomplete data if enough amount of data is available [3], [67].

2.3 Numerical weather prediction

The physical approach provides high accuracy for long-term prediction. Numerical weather prediction (NWP) models are often used to effectively predict the wind speed over big geographic areas for time horizons ranging from 3 to 72h. The size of geographic prediction areas and the length of time horizons are depended on the spatial and temporal resolutions of NWP models [6]. It is a challenging task to downscale the mesoscale condition of NWP to the local site [52]. Physical approaches are commonly used to refine the NWP models by using the downscaling method. Detailed physical descriptions, such as terrains, surface roughness, orography, obstacles, pressures, temperature, and other meteorological and topological information are used to find the refined wind speed data at hub height [2], [6], [10], [16], [21], [22], [63]. Then the refined wind speed data is converted to the wind power production by using the corresponding power curve [16], [35], [44]. Most of the existing wind power forecasting models are based on NWP models such as WPPT, German Previento, GH forecaster from the UK, eWind model of the USA, the Danish Predictor, Rapid Update Cycle, North American Mesoscale and et cetera [43], [44]. Models without using numerical weather prediction (NWP) might produce acceptable accuracy for the short-term horizon, but the forecasting performance is gradually degraded in longer forecasting horizons [6].

The main drawbacks of physical approaches are acquiring physical data, solving the complex mathematical equations and requiring high computational cost [2], [16], [22]. Due to the difficulty of obtaining information in a short time, NWP models are usually run once or twice per day, which limits their forecasting ranges [11]. The forecasting accuracy of NWP models also depends on the weather condition. Better forecasting performance is obtained when atmospheric conditions are stable [11].

2.4 ARIMA

Some examples of statistical models are autoregressive (AR), autoregressive moving average (ARMA), autoregressive integrated moving average (ARIMA), Bayesian approach, and Gray predictions [61]. Among all the statistical approaches, ARMA and ARIMA are very popular for short-term wind speed and wind power forecasting.

ARIMA models were popularized by George Box and Gwilym Jenkins in 1970s are one of the most robust, simplest, and widely used statistical model in wind speed and power forecasting [39], [68], [69], [70]. Autoregressive (AR), autoregressive moving average (ARMA), factional ARIMA (f-ARIMA) [19], and seasonal ARIMA (sARIMA) are in the same family as ARIMA. ARIMA models are good at capturing the behaviour of linear time series by using a lagged series [41], [68]. They can perform a linear mapping between inputs and outputs [6], [70]. Identification of a suitable model, estimation of model parameters, diagnostic and verification check are the three main steps used to build ARIMA [21], [34], [46]. ARMA can be used to predict wind power if the input time series is stationary. The formula of ARMA is defined as [2]:

$$P(t) = \sum_{i=1}^{p} \alpha_i P(t-i) + \sum_{j=1}^{p} \beta_j e_w(t-j) \qquad (2.8)$$

where α_i and β_j are the corresponding coefficients of AR and moving average (MA), respectively; e_w is the white noise, and P is the output of the model [2]. The number of significant lags in the partial autocorrelation function (PACF) and the autocorrelation function (ACF) plots indicate the order of AR and MA, respectively [34].

2.5 Artificial neural networks

2.5.1 Structure and parameters of ANNs

The ANN models generally consist of an input layer, one or more hidden layers and an output layer. The input layer communicates with the external environment, and the output layer presents a pattern to the external environment [71]. The hidden layer is an intermediate layer between the input layer and the output layer. It consists of hidden neurons which have activation function applied to them. The hidden layer is not required if the training data is linearly separable. Neurons in each layer are independent and connect to the neurons in previous layers. A weight and a transfer function are assigned to each connection [21]. The selection of hidden neuron number for ANNs is one of the major problems faced by many researchers [72]. Many attempts have been made by different researchers to determine a fixed number of hidden neurons in the layer that provides minimal error and highest accuracy. Still, none was accurate for all the applications [71]. The network may overfit the data if there are too many hidden neurons [62]. The overfitting causes ANNs to overestimate the complexity of the target, which leads to the degradation of the network's generalization ability [73]. On the other hand, if there are too few hidden neurons, the network will not be flexible enough to model the data well [62]. Therefore, it is crucial to determine the proper number of hidden neurons for forecasting models to perform well.

Authors in [74] suggested that an ANN with a little number of hidden neurons is more powerful for out-of-sample forecasting than the one with many hidden neurons. Many researchers have tried their best to find an optimal number of hidden neurons. Still, nobody has yet succeeded in finding the formula for determining the optimal number of hidden neurons [71].

Trial and error methods are used by many authors to determine hidden neuron number [27], [62], [75]. Cross-validation and early-stopping are the two prevalent methods used by researchers to choose a suitable hidden neuron size [73]. Better results in less training time can be obtained if a suitable number of hidden neurons is selected [73]. Rule-of-thumb methods are also used by many authors to determine a suitable number of hidden neurons [71]. Some of the methods stated as following: [76], [71], [77]

- The hidden neuron number should be equal to 70-90% of the input layer nodes.
- The number of hidden neurons should be less than twice the input node number.
- The number of the hidden neurons should be in between the input nodes size and the output neuron size.

Practically, the rule-of-thumb methods do not apply to most problems. As the methods mentioned above only considered input and output neurons [78]. The complexity of activation function, type of training algorithm, amount of noise in the target data, number of training pairs, and quality of training sample also influence the hidden neuron number selection [28], [71], [73], [78]. Therefore, trial and error methods are applied in this dissertation.

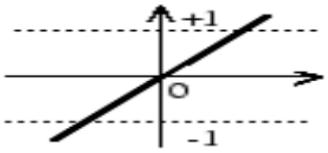
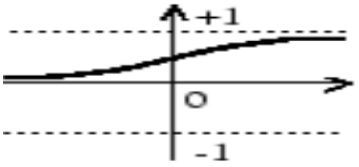
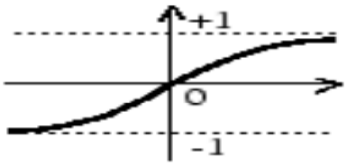
The structure of ANNs is classified into two classes, namely feedforward topology and recurrent topology. In feedforward topology, information flows in one direction only, from input to output. In recurrent topology, information flows in both directions [16], [42]. The feedforward multilayer perceptron is the best known and most widely used kind of ANNs for non-linear time series forecasting problems [44]. Therefore, multilayer feedforward ANN was used in this dissertation. There are three types of learning, supervised learning, unsupervised learning, and reinforcement learning. In supervised learning, ANN weights are set based on sample input-output pairs. In unsupervised learning, ANN weights are set based on the input data with labelled responses. In reinforcement learning, ANN weights are set by interactions with the environment [16]. Supervised learning is commonly used for time series forecasting. Therefore, supervised learning was used in this dissertation.

A bigger number of hidden layers theoretically enhances the network mapping ability [50]. However, network complexity, computational requirement and training time is also increased many times [50]. Accuracy of the network can be achieved up to a great extent when the number of hidden layers is increasing up to three layers [71]. Too many hidden neurons or layers will also cause overfitting problem [71]. One hidden layer with the suitable number of neurons will be enough for approximating any complex non-linear continuous function [44], [50], [68], [71]. Multiple hidden layers are required when the training data contain discontinuities [71].

2.5.2 Transfer function

The types of decision ANNs take made by activation functions which have a big impact on the performance of ANN models [3]. Activation functions are used by ANNs to map the relationship between input data and output targets [50]. Table 2.3 summaries the commonly used activation functions [3], [13], [28].

Table 2.3: Summary of commonly used activation functions of ANNs

Activation function	Mathematical equation	Graphical representation
Linear function (output layer)	$f(x) = x$	
Logistic sigmoid function	$f(x) = \frac{1}{(1+e^{-x})}$	
Hyperbolic tangent (tan-sigmoid) function	$f(x) = \frac{1-e^{-2x}}{1+e^{2x}}$	

where *x* is the weighted input of each layer.

At least one non-linear activation function needs to be used in the network structure to maintain the ANNs' non-linear problems solving ability [3]. The logistic sigmoid function was used in this dissertation because the range of normalised variables is also from 0 to 1.

2.5.3 Training algorithm

It is also essential to use a suitable training algorithm to train ANNs. The backpropagation algorithm (BP) is one of the most commonly used algorithms for ANNs. Around 80% to 90% of ANNs use the backpropagation algorithm or its transformations [66]. As shown in Figure 2.1 [10], the back-propagation algorithm is consisting of two stages. In the forward propagation stage, learning samples passes from the input layer to the output layer. In the back-propagation stage, the error between the output and the target passes back from the output layer to the input layer through the original connection [3], [66]. The connection weight is adjusted layer by layer to minimize the MSE error by using a gradient algorithm [3], [66]. However, authors in [79] mentioned that network with the standard backpropagation algorithm converges slowly.

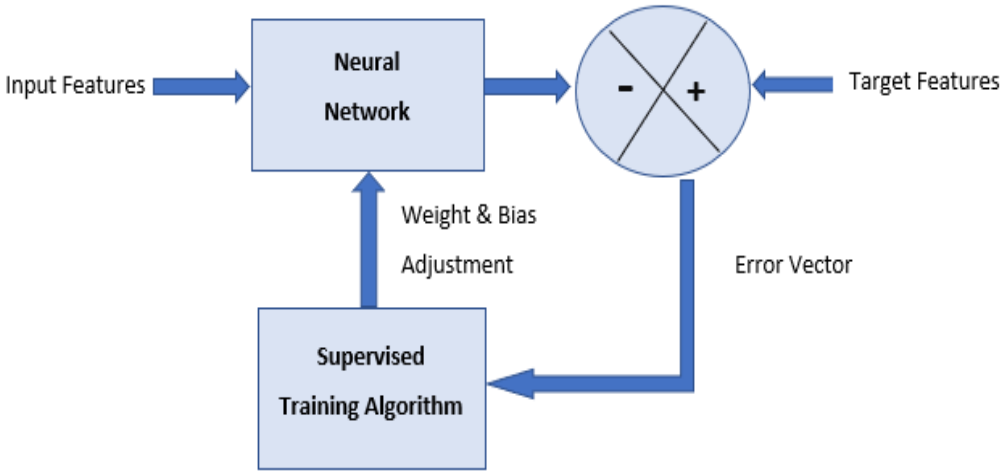


Figure 2.1: Block diagram of the backpropagation supervised training algorithm

ANN model that uses the Levenberg-Marquardt (LM) back-propagation algorithm for training is much faster than the one that uses a standard backpropagation algorithm [62]. LM is very robust and efficient because it has the speed advantage of the Gauss-Newton method and stability advantage of the steepest descent method [27], [30], [79]. Levenberg-Marquardt (LM) back-propagation algorithm was used in this research.

Many authors have also used other computational intelligence methods such as genetic algorithm (GA) and particle swarm optimization (PSO) to train the weights of ANNs. In particular, PSO has been used to train ANNs for time series forecasting by many researchers [80]- [82]. Kennedy and Eberhart first introduced particle swarm optimization in 1995 as a stochastic, population-based optimisation technique to optimize different functions [76], [83]. The author in [82] used a PSO-based ANNs to determine outcomes of a construction claim. The simulation results show that the PSO-based ANNs performed better than the BP-based ANNs for predicting outcomes of a construction claim. One of the most crucial PSO construction steps is to select suitable PSO parameters [84], [85]. A small variation of PSO parameters will result in a significant change in the algorithm performance [83], [86].

To effectively use the PSO algorithm, some critical parameters need to be appropriately defined, such as the number of particles, acceleration constants, inertia weights, the maximum number of iterations and computed precision [84], [85]. Inertia weight controls the effect of the previous velocity on the current velocity, which determines the global and local search ability of PSO [83]- [85], [87]. The impact of the particle population, acceleration constants, and inertia weights on the performance of PSO are investigated in this research.

2.6 Spatial correlation model

Future wind speed is related to previous and current wind speed in time and space. In a spatial correlation model, historical wind speeds of the predicted point and its neighbouring sites together with wind turbines power curve are used for wind power forecasting [21], [88]. Using the historical data from neighbouring sites do improve forecasting performance over the flat terrain due to the similarity between downstream wind and upstream wind. However, the forecasting performance of spatial correlation model becomes worse when the terrain is complex [89]. Alexiadis et al. [88] and Focken et al. [90] found that a spatial correlation predictor outperformed the model using the historical data of a single site. Authors in [91] developed a spatial correlation and ANN-based model, which can forecast

the wind speed of the target wind station using only the wind speed of the reference station. In this research, spatial correlation technique was used to provide surrounding meteorological information to forecasting models of a target station.

2.7 Hybrid and ensemble approaches

Despite all the advantages ANNs can provide, they still have some deficiencies such as generalization issues, overtraining, local minima problem and they are susceptible to large and sudden spikes in input data and initial value of weights and biases [2]. The overtraining and undertraining are also prevalent problems for ANNs when too many or too little parameters to be estimated from the training data [30]. Another challenge in using ANNs is to adjust different parameters to obtain an optimum ANN configuration [23]. Different structures and model parameters can produce different forecasting performance for the same data in terms of evaluation criteria [92]. Ensemble models can take advantages of each model. For that reason, ensemble methods can be an effective way of overcoming the above issues [30].

2.7.1 Hybrid approaches

The objective of using hybrid and ensemble approaches is to benefit and integrate the advantages and information of each individual model to enhance forecasting accuracy [55], [93]. Hybrid approaches and ensemble approaches are very similar. Hybrid approaches combine different approaches or incorporate superior features of various forecasting models to obtain more accurate forecasting results [11], [92]. Ensemble approaches use different configurations of the same model or different models to obtain an aggregated or weighted final output.

Combinations of physical and statistical approaches have been implemented by some researchers to improve forecasting accuracy [44], [63]. Physical approaches provide initial analysis, and their output can be used by statistical approaches to improve forecasting performance [21]. Hybrid approaches are usually constructed by using one linear and one non-linear approach to gain advantages of each approach and to enhance the prediction performance further [16], [92]. The linear approach is used for the linear component prediction, and the non-linear approach is used for the non-linear component prediction [70], [92]. Forecasting results of member models do not always carry the same weight. A weighting coefficient can be allocated to each model based on their forecasting accuracy [92].

Authors in [62] combined a Levenberg-Marquardt trained feedforward ANN and wavelet transform to form a wavelet transform and artificial neural network hybrid model (WT-ANN) which performs much better than the persistence and ARIMA approaches, and slightly better than the ANN for short-term wind power forecasting in Portugal. Authors in [94] utilised the k-nearest neighbour (KNN) method to search for suitable parameters for ANNs. They concluded that the ANNs with the k nearest neighbour search method outperforms the ANNs without it for short-term prediction [94]. Authors in [95] concluded that the hybrid models of ANN and ARIMA perform better than the individual model for wind speed forecasting.

However, the hybrid approaches do not always produce better forecasting performance than the best individual method at all the circumstances [70]. It is shown in [96] that there is no significant improvement from ARIMA to ARIMA-SVM and ARIMA- ANN hybrid models.

2.7.2 Ensemble approaches

Ensemble method combines multiple prediction models or the same model with different configurations to obtain an aggregated result which is better than any of the individual models [30], [97]. Ensemble methods can increase the chance of eliminating poor predictors [13]. Ensemble methods can be

classified into two classes, namely competitive ensemble and cooperative ensemble. In a competitive ensemble, forecasting models are trained with different datasets or the same dataset with different model parameters. The final output of the competitive ensemble is equal to the aggregated result of all individual models [96]. Results of each predictor need to be diverse to enhance the performance of ensemble models [96], [98]. In a cooperative ensemble forecast, several predictors are working together to produce a final output [96].

Authors in [49] tested several ANN methods and indicated that the ensemble model has better performance than the best-performing individual model. Authors in [67] used a multi-NWP and a multi-scheme together with an ANN model to form an ensemble model which provides better wind power forecasting performance than the individual model. Authors in [99] proposed a homogeneous ensemble model consisting of six Long short-term memory neural networks (LSTMs) with different numbers of hidden layers and hidden neurons for 10 minutes and 1 hour ahead wind speed forecasting. The simulation results showed that the ensemble model performs slightly better than each of six LSTMs. Ensemble technique was used in this research to provide better forecasting results than the best individual member model.

2.8 Discussions and summary

Wind power forecasting is usually classified into four forecasting time horizons, namely very-short-term, short-term, medium-term, and long-term. Different forecasting approaches are suitable for different forecasting time horizons. Statistical approaches and artificial intelligence approaches are suitable for very-short-term and short-term wind speed and wind power forecasting. In contrast, physical approaches are better for medium-term and long-term forecasting. The main focus of this research is on short-term wind speed and wind power forecasting. Therefore, only statistical and artificial intelligence and ensemble techniques are considered in this research. ARIMA is one of the most popular statistical approaches which are good at modelling linear time series. However, wind speed and wind power are the nonlinear time series. As a result, artificial neural networks have been used by many authors to forecast short-term wind speed and wind power in recent years. ANNs are good at modelling the nonlinear time series like wind speed and wind power. However, it is difficult to select suitable parameters for ANNs which might face overfitting problem if model parameters and input features are not chosen carefully. Input feature selection techniques and spatial correlation analysis together with data-processing steps can largely enhance the performance of forecasting models. Furthermore, hybrid and ensemble approaches can be used to obtain advantages from several models to further enhance forecasting performance from the best individual model. Different forecasting types can be divided into deterministic forecasting, probabilistic forecasting, direct forecasting, indirect forecasting, independent multi-steps forecasting, and iterative multi-steps forecasting. In indirect forecasting, the forecasted wind is converted to forecasted wind power by using a suitable wind turbine power curve. Finally, the performance of forecasting models is usually evaluated by using commonly used evaluation metric like RMSE, MAE, and MAPE, along with the benchmark model, persistence model.

3. Methodology and modelling

The purpose of this chapter is to explain the methods used in this dissertation. The chapter starts with Section 3.1 describing data source, data analysis, and data pre-processing steps. In Section 3.2, input features selection and spatial correlation analysis are addressed. The model structure and parameters of ANNs are selected in Section 3.3. Section 3.4 explains the construction steps of particle swan optimization, which is used as a training algorithm to determine weights of ANNs. Section 3.5 gives an overview of the structure of different ensemble models. The chapter finishes with Section 3.6 explaining wind speed to wind power conversion and interval forecasting methods.

3.1 Data

3.1.1 Data source

Several approaches can be used to obtain wind speed and power data, such as observations, data mining and numerical weather simulations. The most straightforward and reliable way to collect wind data is through on-site observation [10]. Three sets of data were collected and used in this research. They are from Sotavento Galicia wind farm in Spain (SGWF) [100], National Renewable Energy Laboratory of the United States (NREL) [101], and Wind Atlas of South Africa (WASA) [102]. Table 3.1 summaries the information of three data sources.

Table 3.1: Information on the three data sources

	SGWF	**NREL**	**WASA**
Data length	2004-2019	2008-2012	2010-2019
Number of stations	1	126,693	19
Data resolution	10 min, 1 hour	5 min, 1 hour	10 min
Availability of wind power data	Yes	Yes	No
Data collection approach	On-site observation	Numerical weather simulation	On-site observation
Variables	WP, WS, WD	WP, WS, WD, T, P, RH, AD	WS, WD, T, TG, P, RH

In the above Table, WP = wind power, WS = wind speed, WD = wind direction. T = temperature, P = barometric pressure, RH = relative humidity, AD = air density, TG = temperature gradient

The reason for using three data sources for this research is that each data source has its advantages and disadvantages for this research. Table 3.2 summaries the advantages and disadvantages of each data source.

In our opinion, the quality and quantity of all three data sources are good. There are 24 wind turbines with nominal nameplate capacity ranging from 0.6 MW to 1.32 MW installed in SGWF. The total installed power capacity of the wind farm is 17.56 MW. The dataset from SGWF contains three variables, which are the total generated wind power of the entire wind farm, wind speed and wind direction. However, the SGWF data are lacking wind power generation data of each wind turbine and other meteorological variables such as temperature, barometric pressure, relative humidity, air density. Table 3.3 summaries the descriptive statistics of the SGWF dataset.

Table 3.2: Summary of advantages and disadvantages of each data source

Data Source	Advantages	Disadvantages
SGWF	• Contain real wind power generation data which reflect the practical amount of wind power that can be generated under real operation conditions	• No real wind power data for each wind turbine • Only contains two meteorological variables, i.e., wind speed and wind direction • No data from neighbouring sites are available for spatial correlation analysis
NREL	• Contain meteorological data from 126, 693 sites which are suitable for spatial correlation analysis • Contain simulated wind power for all 126, 693 sites	• Wind power data are not real wind power. They are simulated from the weather research and forecasting model (WRF)
WASA	• Contain real meteorological data from 19 weather stations • All the weather stations are located in South Africa	• No real wind power generation data • Distances between each station are far away from each other (average distance is 482 km)

Table 3.3: Summary of descriptive statistics of SGWF variables

Variables	Unit	Minimum-maximum	Mean	Standard deviation
Wind power	MW	0.0-17.3	3.3	3.8
Wind speed	m/s	0.4-39.9	6.5	3.3
Wind direction	°TN	0.0-360.0	170.9	97.0

The maximum power measured in the dataset is 17.29 MW which is lower than the total installed power capacity of the entire wind farm (i.e., 17.56MW). Although the maximum wind speed is very high, it is lower than the highest wind speed recorded (i.e., 44.4 m/s) in the same region as the wind farm [103]. The mean wind speed value of the whole dataset is reasonable. Figure 3.1 shows the geographical location of the Sotavento Galicia wind farm.

Figure 3.1: Location of Sotavento Galicia wind farm (SGWF)

Sotavento Galicia wind farm is located in the northwest region of Spain. It has a Mediterranean-style climate: dry and warm summers and wet and cool winters which is similar to the climate in Western Cape Province of South Africa.

The dataset of NREL consists of wind power, wind speed, wind direction, temperature, barometric pressure, air density. The data of NREL were generated from the Weather Research and Forecasting (WRF) model version 3.4.1. NREL provides datasets for more than 126, 000 sites across the United States of America. A large number of sites is good for spatial correlation analysis. Table 3.4 shows the descriptive statistics of NREL site ID61118.

Table 3.4: Summary of descriptive statistics of NREL: ID61118 variables

Variables	Unit	Minimum-maximum	Mean	Standard deviation
Wind power	MW	0.0-16.0	6.4	5.5
Wind speed	m/s	0.0-23.9	6.7	3.3
Wind direction	°TN	0.0-360.0	220.4	89.5
Air temperature	°C	-2.6-43.0	14.3	7.0
Barometric pressure	hPa	976.8-1032.1	1012.8	5.3
Air density	kg/m^3	1.1-1.3	1.2	0.0

The power rating of each wind turbine located in site ID 61118 is 2 MW, and the total capacity of a site containing eight 2 MW wind turbines is 16 MW. The maximum and minimum values of the wind power is within the realistic range. Mean wind power of 6.4 MW makes the power capacity factor of the site equals to 0.4, which is a reasonable number. Meteorological variables are also within the realistic range. Some of the sites within the big red square shown in Figure 3.2 below were used in this research.

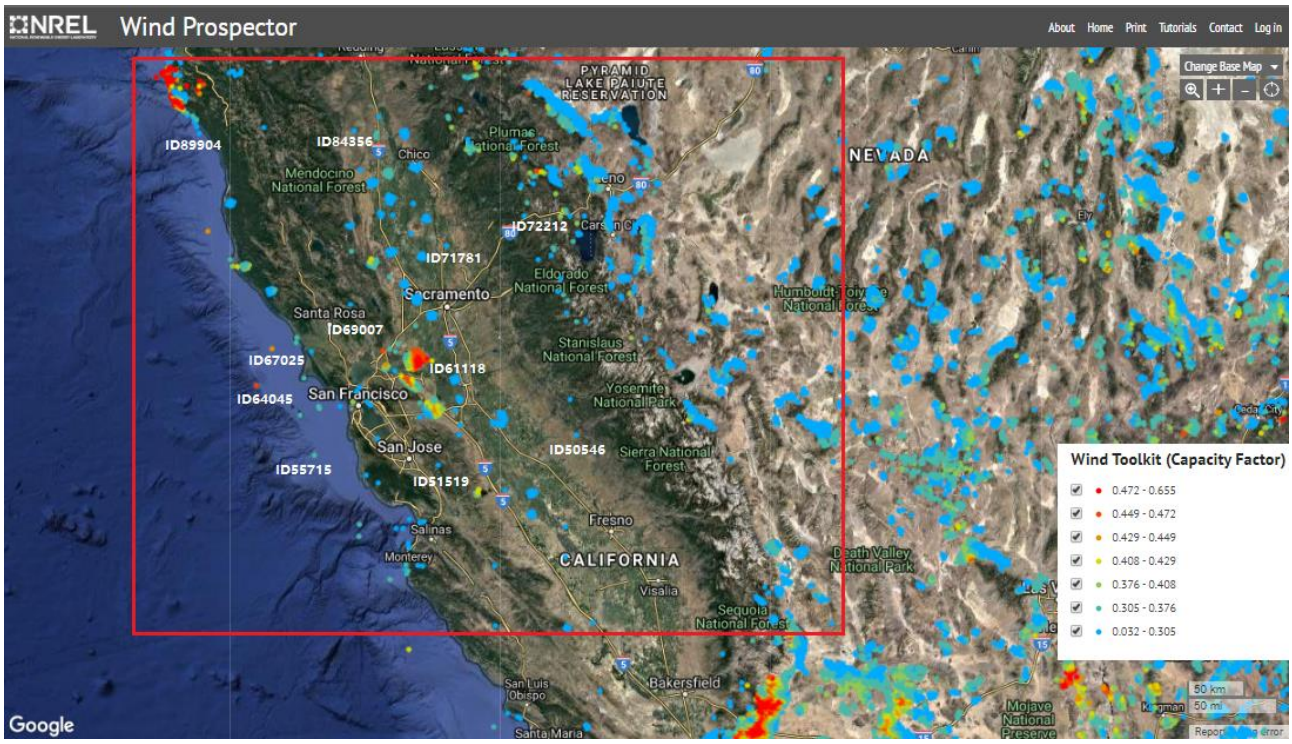


Figure 3.2 Location of the NREL sites used in this dissertation

As shown in Figure 3.2, many sites are available for this research. Circle dots represent the location of sites, and the colour of the dots represent the capacity factor of each site. Blue colour indicates a low capacity factor and red colour indicates a high capacity factor.

The dataset of Wind Atlas of South Africa (WASA) consists of the observations of six meteorological variables, which are wind speed, wind direction, temperature, temperature gradient, pressure and relative humidity. The WASA dataset contains many useful meteorological variables, but it does not provide wind power generation data. Therefore, the wind power data simulated from the actual wind speed of the WASA dataset were used in this dissertation. The datasets of WASA can be used to determine the effect of different meteorological variables on the short-term wind speed and wind power forecasting performance. Table 3.5 summarises the descriptive statistics of the meteorological variables from the weather station WM03 of WASA.

Table 3.5: Summary of descriptive statistics of meteorological variables of WASA: WM03

Meteorological variables	Unit	Minimum-maximum	Mean	Standard deviation
Wind speed	m/s	0.2-23.9	7.2	3.4
Wind direction	°TN	0.0-360.0	175.4	84.7
Air temperature	°C	3.6-41.3	17.2	5.4
Air temperature gradient	°C	-2.2-13.6	0.14	0.0
Barometric pressure	hPa	973.3-1006.3	987.9	4.7
Relative humidity	%	2.8-100.0	63.2	24.8

The ranges of meteorological variables shown in Table 3.5 are practical. Figure 3.3 shows the location of 19 WASA weather stations.

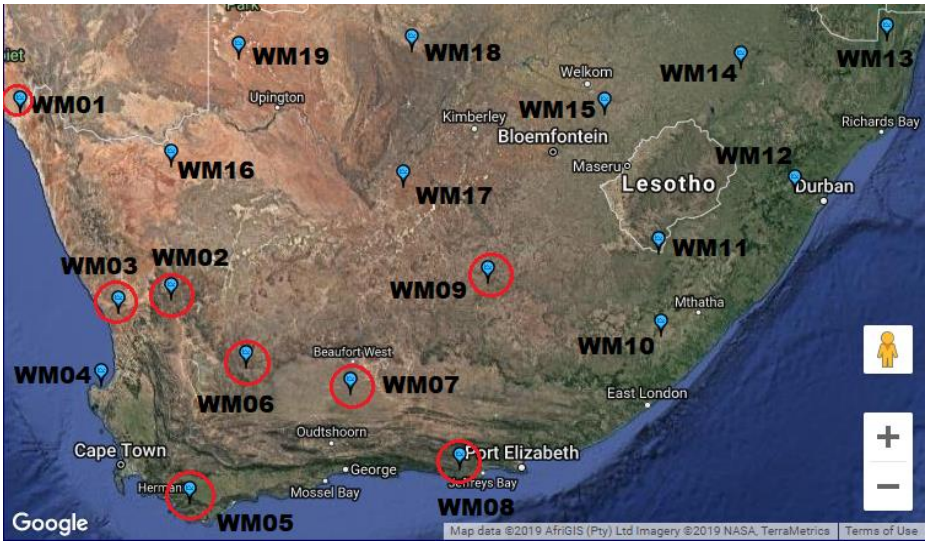


Figure 3.3: Location of the WASA weather stations

As can be seen in Figure 3.3, 19 weather stations spread out evenly across most areas of South Africa. High availability of datasets from the neighbouring sites is beneficial for spatial correlation analysis. Weather station WM11 to weather station WM19 are installed in recent years, so the sample size from these stations is not large enough to use in this research. Therefore, the data from these stations are not included in this research. Only the data from the weather stations circled in red (WM01 to WM09 except for WM04) were used in this research. Although WM04 and WM10 were installed in the same period as these stations circled in red, the data from WM04 and WM10 are incomplete. As a result, WM04 and WM10 are not considered in this dissertation. Table 3.6 summaries the distances between each WASA weather station.

Table 3.6: Distance between eight available WASA stations in kilometres.

	WM01	WM02	WM03	WM05	WM06	WM07	WM08	WM09	Total Distance	Mean
WM01	0	417	388	728	586	746	964	859	4688	781.3
WM02	417	0	92	344	170	341	561	539	1925	320.8
WM03	388	92	0	341	234	413	628	629	2096	349.3
WM05	728	344	341	0	247	322	446	623	2428	404.7
WM06	586	170	234	247	0	180	395	435	1812	302.0
WM07	746	413	413	322	180	0	222	301	2296	382.7
WM08	964	561	628	446	395	222	0	321	3216	536.0
WM09	859	539	629	623	435	301	321	0	3707	617.8

The mean distance between two stations is measured by using google map measuring tool.

3.1.2 Data quality control

Several commonly used data quality-control methods are discussed below [28]:

1. Find missing values in the dataset. Fill the missing values based on the values of adjacent rows.

2. Detect and remove or replace the outlier values by comparing the observed values with the instrument measurement range and history measurements from other sources.

3. Check gradient of meteorological variables to determine if changes are with a reasonable range.

After analysing all the datasets, it was concluded that the quality of all the datasets is good. Only a few missing values and no outliers were found in the data obtained from SGWF, NREL and WASA. Those missing values are filled based on the values of the adjacent rows. For example, the missing values at hour 2 can be estimated by calculating the average of values at hour 1 and hour 3. The quality of the meteorological variables and wind power data of the NREL datasets are validated by NREL in [104] and [105], respectively. Table 3.7 shows the summary of the datasheets of instruments used in WASA.

Table 3.7: Summary of datasheets of the instruments used in WASA

Name	Manufacturer	Model	Range	Tolerance
Cup anemometers	Windsensor, Denmark	P2546A-OPR	0-75 m/s	0.28 % (over 4-16 m/s)
Potentiometer wind vanes	Vector Instruments, United Kingdom	W200P-01	360° mechanical angle	±2° obtainable in steady winds over 5ms-1
Humidity and air temperature probes	Vaisala Oyj, Finland	HMP60	0%- 100 %RH -40°C- 60°C	±3% RH (over 0%- 90 % RH) ±0.5 °C (over 10° to 30°C range)
Temperature gradient probes	Campbell Scientific	CS109	-50°C- 70°C	±0.2°C (over 0° to 70°C range)
Barometers	Vaisala Oyj, Finland	PTB110	500-1100 hPa	±0.2 hPa at +20°C

By comparing the values in Table 3.5 and Table 3.7, it can be seen that all the variables values of the WASA datasets are within the instrument range. Data pre-processing is carried out after the data collection and data quality control are done.

3.1.3 Data pre-processing

i. *Wind speed patterns*

Wind power is an intermittent and volatile renewable resource. The patterns in wind speed time series are not obvious. Figure 3.4 shows the graph of hourly wind speed against the time of day for the first ten days of January 2011.

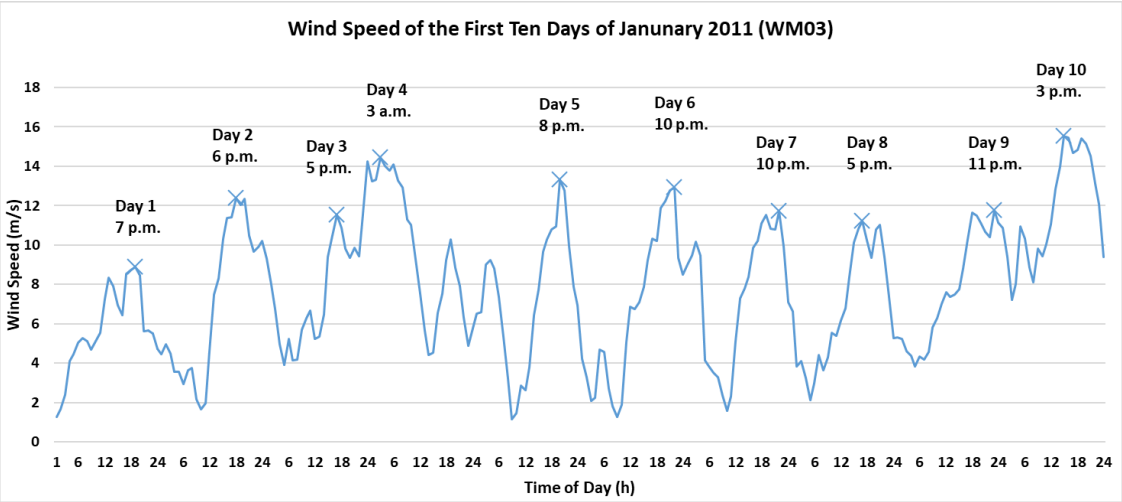


Figure 3.4: Time series of wind speed

As can be seen in Figure 3.4, the highest wind speed of each day occurs at various times of day, which indicates that the hourly wind speed time series contains a cyclic characteristic. The difference between the seasonal and cyclic pattern is the frequency of the fluctuation. In a seasonal pattern, the fluctuations occur at a fixed frequency. In a cyclic pattern, the fluctuations occur at a non-fixed frequency [106]. There is no apparent long-term increase or decrease in the wind speed time series, so the hourly wind speed time series contains no trend and seasonality in the short period.

It can be seen that the highest wind speed of a day usually occurs between late afternoon and midnight. Because the temperature on the surface of the earth is the highest in the afternoon, and the large air temperature difference between the lower atmosphere and upper atmosphere destabilizes the atmosphere. Higher wind speed from the cooler upper atmosphere moves down to the warmer earth surface between afternoon and mid-night. It is essential to label the input data with time of day so the forecasting models can recognize the daily cycle within the wind speed time series.

The hourly wind speed time series and the monthly wind speed time series contain different characteristics. Figure 3.5 shows the mean monthly wind speed of station WM03 from January 2011 to December 2019.

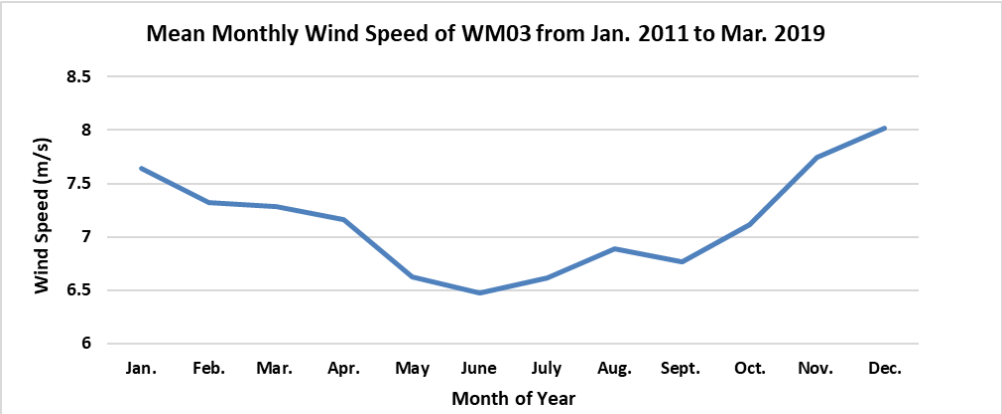


Figure 3.5: Mean monthly wind speed of station WM03 from January 2011 to December 2019.

As can be seen in Figure 3.5, the wind speed in the Western Cape, South Africa, is usually higher in the summer months (from December to February) than the winter months (from June to August). Figure 3.6 shows monthly mean wind speed from January 2011 to March 2019. 'X' in the figure indicates the month with the highest monthly mean wind speed of each calendar year.

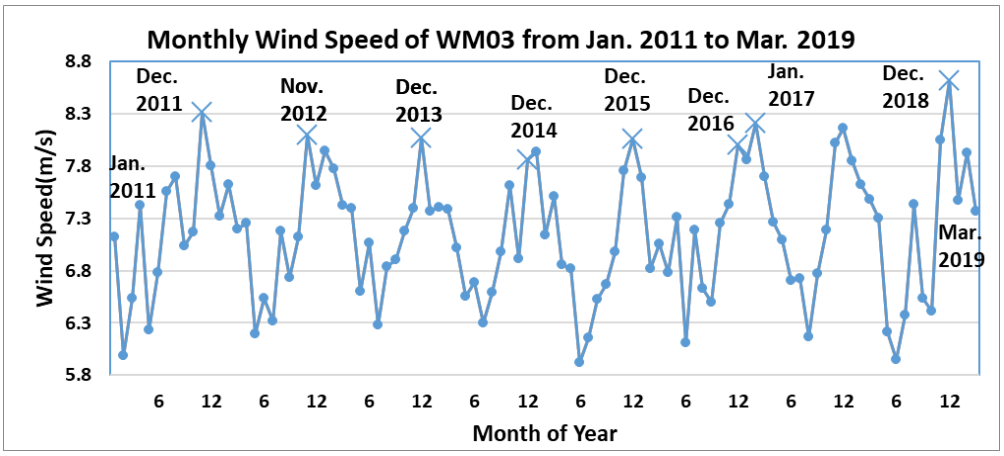


Figure 3.6: Monthly mean wind speed of station WM03 from January 2011 to March 2019

There is no long-term increase or decrease, as shown in Figure 3.6. It can be seen that the highest monthly mean wind speed usually occurs in December. The frequency of the fluctuation is fixed. It can be concluded that the monthly mean wind speed time series contains a seasonal characteristic only.

Three approaches are tried in this research to capture the daily cycle and seasonality patterns of the time series. The first approach is to simply include a time indicator as an input feature to ANNs and the second approach is to minimise the variations of wind speed between morning and afternoon by using a moving average. The third approach is to use two forecasting models to forecast morning and afternoon wind speed separately.

A basic feedforward ANN with a single hidden layer and ten hidden neurons was used to determine a suitable approach that can be used to extract the cyclic pattern of wind speed. Table 3.8 shows the experiment configurations.

Table 3.8: Summary of the configurations utilised to determine a suitable method to tackle the cyclic pattern of wind speed of WASA: WM03

Forecasting model	Feedforward ANN with a single hidden layer
Number of hidden neurons	10
Training algorithms	Levenberg-Marquardt
Forecasting horizon	2h-ahead wind speed forecasting
Data source	WASA: WM03
Sample resolution, size	10min, 300, 000
Input feature	Wind speed, time of day

Note that the Levenberg-Marquardt is a type of back-propagation training algorithm.

Figure 3.7 shows the construction steps of the moving average approaches.

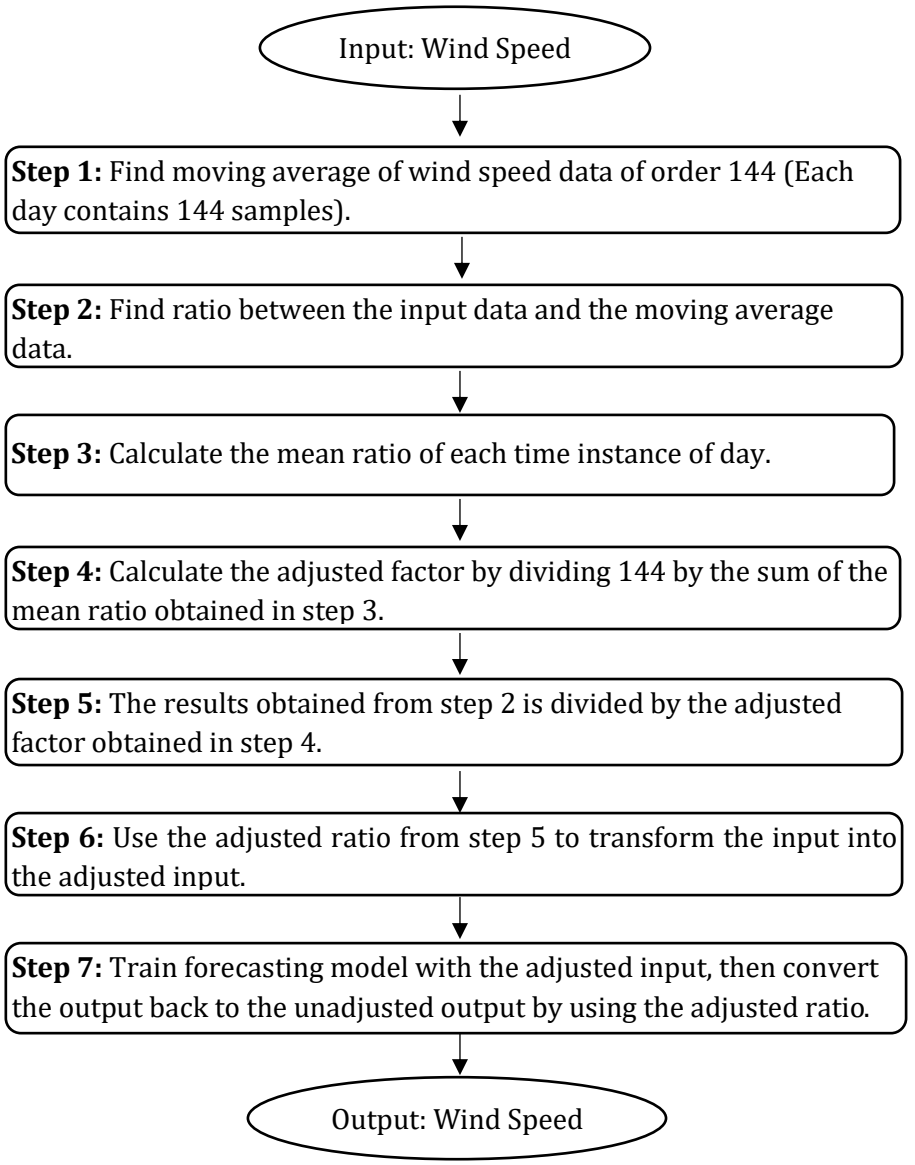


Figure 3.7 Block diagram of moving average approach

As shown in Figure 3.7, the first step is to take the moving average of the wind speed data. The wind data has a daily cyclic pattern, so a central moving average of order 144 (total number of samples per day) is carried out. The second step is to find the ratio between the original wind speed and the wind speed after the moving average operation. The third step is to calculate the mean ratio of each time

instance of the day. The fourth step is to calculate the adjusted factor by dividing 144 by the sum of the mean ratio found in step 3. In the fifth step, the ratio found in the second step is adjusted by dividing the ratio obtained in step 2 by the adjusted factor obtained in step 4. The sixth step is to transform the original wind speed to the adjusted wind speed using the adjusted factor found in the fifth step. The final step is to use the adjusted wind speed to train the forecasting model, and the output of the forecasting model is transformed back to unadjusted outputs for comparison with the actual wind speed. Figure 3.8 shows the block diagram of the two-model approach.

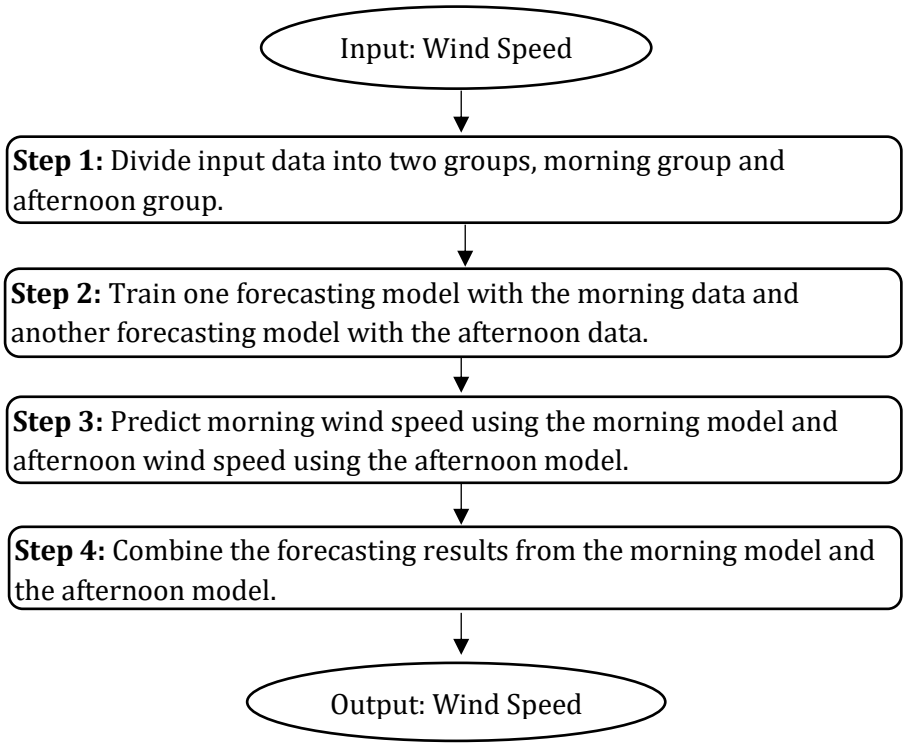


Figure 3.8: Block diagram of the two-model approach

As can be seen in Figure 3.8, the first step is to divide the wind speed into two sections, morning section and afternoon section. Two forecasting models with the same configuration are separately trained with the morning data and the afternoon data in the second step. In the third step, the model trained with the morning data is used to predict morning wind speed, and the model trained with the afternoon data is used to forecast afternoon wind speed. The final step is to combine the outputs from the morning model and afternoon model. The combined results are then compared to the results produced from a single forecasting model.

ii. *Normalisation*

The range of each input variable is different from each other. As an example, the air temperature of the WASA dataset is ranging from 3.6-41.3 °C, whereas barometric pressure is ranging from 973.3-1006.3 hPa. Variables with larger values might suppress the impact of the one with smaller values. A popular min-max normalisation technique is implemented in this research to mitigate this issue. As mentioned in Section 2.1.2, the min-max normalisation is defined as Eq. (2.1). In this study, X'_{min} is set to 0 and X'_{max} is set to 1 because the logistic sigmoid function of ANNs has the same range. Therefore, Eq. (2.1) becomes Eq. (3.1).

$$X'_i = \left(\frac{X_i - X_{min}}{X_{max} - X_{min}}\right) \qquad (3.1)$$

3.1.4 *Training, validation, and testing datasets*

All the pre-processing data approaches mentioned in the above sections are implemented on the training, validation and testing datasets. A total of 300000 samples from each data sources were used in this dissertation, 170000 samples were used for model training, 30000 samples were used for model validation, and 100000 samples were used for model testing.

i. *The time interval (data resolution) and forecasting horizon*

The time interval between samples denotes forecast interval (data resolution), and the forecasting horizon is the length of time into the future for which forecasts are to be prepared. Three datasets contain a total of three types of data resolution, which are one sample per every five minutes, one sample per every ten minutes and one sample per hour. For example, a forecasting model needs to perform a12-out-samples forecasting (12 x 10 min = 120 min =2 h) to forecast wind power two hours in advance when the one sample per ten minutes dataset is used. On the other hand, when one sample per hour dataset is used, forecasting models only need to perform 2-out-samples forecasting (2 x 1 hour = 2h). For other forecasting time horizons, the gap between the input and the target is changed accordingly.

Table 3.9 and Table 3.10 shows the selection of input and target training pairs with one sample per ten minutes resolution and one sample per hour resolution, respectively (2h-ahead wind speed forecasting).

Table 3.9: Input and target training pair selection for the data with one sample per ten minutes resolution (2h-ahead wind speed forecasting)

Step	Date	Speed (m/s)	
1	01/01/2018 00:10	6.22	Input1
2	01/01/2018 00:20	5.6	Input2
3	01/01/2018 00:30	8.11	Input3
4	01/01/2018 00:40	7.87	
5	01/01/2018 00:50	7.96	
6	01/01/2018 01:00	7.41	
7	01/01/2018 01:10	7.56	
8	01/01/2018 01:20	9.08	
9	01/01/2018 01:30	8.98	
10	01/01/2018 01:40	9.28	
11	01/01/2018 01:50	8.08	
12	01/01/2018 02:00	7.16	
13	01/01/2018 02:10	8.48	Target1
14	01/01/2018 02:20	7.6	Target2
15	01/01/2018 02:30	8.05	Target3

Table 3.10: Input and target training pair selection for the data with one sample per hour resolution (2h-ahead wind speed forecasting)

Step	Date	Speed (m/s)	
1	01/01/2018 01:00	7.2	Input1
2	01/01/2018 02:00	8.36	Input2
3	01/01/2018 03:00	8.25	Target1
4	01/01/2018 04:00	8.48	Target2
5	01/01/2018 05:00	10.63	
6	01/01/2018 06:00	9.87	

As can be seen in both Table 3.9 and Table 3.10, the gap between the input data and the target data is two hours for both cases. The performance of the forecasting model using different data resolutions are

tested in this research. Table 3.11 summarises the model configuration and training sample sizes used for data resolution analysis.

Table 3.11: Summary of the configurations used for data resolution analysis

Forecasting model	Feedforward ANN with a single hidden layer
Number of hidden neurons	10
Training algorithms	Levenberg-Marquardt
Forecasting horizon	2h & 6h-ahead wind speed forecasting
Data source	SGWF, NREL: ID61118, WASA: WM03
Sample size	5min = 600, 000, 10min = 300, 000, 1h = 50, 000, 2h = 25, 000
Input feature	All the available meteorological variables and time indicator (see Table 3.1)

A basic feedforward ANN with Levenberg-Marquardt training algorithm was used to determine the impact of the data resolution on the performance of 2h and 6h-ahead wind speed forecasting.

ii. *Training sample size selection*

The training sample size affects the training time and model accuracy. Sample size selection is also crucial for forecasting models. In general, larger training sample requires stronger computational power and longer training time. Therefore, a suitable number of sample size is needed to obtain good forecasting results at the same time to reduce training time. For comparison, all the variables, such as input features, model architecture, model configurations, and forecasting time horizon (2h-ahead) stay the same except the sample size. A basic ANN model, as defined in Table 3.12, is used in this research to determine a suitable number of input training samples. Table 3.12 summarises the basic ANN model configurations and the input features used to determine a suitable number of training pairs.

Table 3.12: Summary of the basic ANN model configurations and the input features used to determine a suitable number of training pairs

Model	Feedforward neural network
Number of hidden neurons	10
Number of hidden layers	1
Training algorithm	Levenberg-Marquardt
Input features (WASA: WM03)	WS, WD, T, TG, P, RH, TI (see Table 3.1)
Transfer function	Logistic sigmoid function
Variables	Number of samples

TI: time indicator

The initial training sample size is set to 1000 pairs. The number of training samples is then increased by 1000 each time until the model stops improving.

3.2 Input features selection and spatial correlation

The SGWF, NREL and WASA datasets contain a total of seven different meteorological variables. However, not all seven meteorological variables have a significant influence on the performance of the forecasting models. The relationship between wind speed and other meteorological variables was determined by using the Pearson correlation coefficient. Pearson correlation is defined as:

$$\rho(x,y) = \frac{1}{N-1}\sum_{i=1}^{N}\left(\frac{x_i-\mu_x}{\sigma_x}\right)\left(\frac{y_i-\mu_y}{\sigma_y}\right) \quad (3.2)$$

where μ_x and μ_y are the mean of x and y, respectively, σ_x and σ_y are the standard deviation of x and y, respectively. N is the total number of observations.

Pearson correlation can measure the linear dependence between two variables. A basic ANN model is also used to map the non-linear relationship between two variables. Table 3.13 summaries the parameters of the basic ANN.

Table 3.13: Summary of the basic ANN model configurations used to determine the impact of different input features on the forecasting performance of ANNs

Model	Feedforward artificial neural network
Number of hidden neurons	10
Number of hidden layers	1
Training algorithm	Levenberg-Marquardt
Transfer function	Logistic sigmoid function
Training sample Size	170, 000
Validation sample size	30, 000
Testing sample Size	100, 000
Variables	Input features

3.2.1 Time indicator

As mention earlier, wind speed has a daily cyclic pattern. Wind speed is usually higher in the late afternoon due to big air temperature difference on the earth surface. Therefore, it is crucial to let the training model know the time of day. Time indicator can be used to tell the training model the time of day. Other than meteorological variables, dummy variable like time of day is also used as an input feature. Different data resolutions will have different time indicators. Table 3.14 shows time indicators of different data resolutions.

Table 3.14: Time of day indicators of different data resolutions

	Time Indicator		
Time	**A sample/ hour**	**A sample/ 10 minutes**	**A sample/ 5 minutes**
01:00	1	6	12
02:00	2	12	24
03:00	3	18	36
…	…	…	…
22:00	22	132	264
23:00	23	138	276
00:00	24	144	288

As an example, for one sample per hour data, using 1 to 24 to indicate the time of day is sufficient; for one sample per ten minutes data, a total of 144 dummy variables 144 x 10 minutes = 1440 minutes = 24 hours) are needed to indicate one day of data.

The Pearson correlation can also be used to determine the wind speed correlation between different weather stations. In this dissertation, the distance and correlation coefficient between the target stations and neighbouring stations were used as a guideline to determine which stations to use as a target station. Datasets from NREL and WASA were used for spatial correlation analysis because both data sources contain multiple sites.

3.2.2 Spatial correlation case study one (WASA)

Table 3.15 summaries the distances between the available weather stations and three nearest neighbouring stations. The stations in bold rows (WM02, WM03, WM06, WM07) are selected for spatial correlation analysis.

Table 3.15: Distance between the target stations and neighbouring stations (WASA)

Target station	Nearest station	Second nearest station	Third nearest station	Mean distance (km)
WM01	WM03	WM02	WM06	464
WM02	**WN03**	**WM06**	**WM07**	**201**
WM03	**WM02**	**WM06**	**WM05**	**238**
WM05	WM06	WM07	WM03	303
WM06	**WM02**	**WM07**	**WM03**	**195**
WM07	**WM06**	**WM08**	**WM09**	**234**
WM08	WM07	WM09	WM06	313
WM09	WM07	WM08	WM06	352

The mean distances between station WM02, WM03, WM06 and WM07 and their corresponding three nearest neighbouring stations are relatively shorter compare to other stations (see bold in Table 3.15). Wind speeds from one station usually have a relatively strong correlation with wind speeds from the adjacent stations. However, the correlation between stations does not only depend on the distance between the stations but also depends on wake effects, geostrophic wind and ground roughness. Instead of using a mathematical model to determine the spatial correlation between different weather stations, a feedforward artificial neural network together with the Pearson correlation were used in this dissertation. Table 3.16 shows the wind speed correlation coefficients between the WASA stations.

Table 3.16: Wind speed Pearson correlation coefficients between the WASA stations

	WM01	**WM02**	**WM03**	**WM05**	**WM06**	**WM07**	**WM08**	**WM09**
WM01	1.00	0.21	0.57	0.04	0.04	0.08	0.12	0.07
WM02	0.21	1.00	0.40	0.22	0.47	0.29	-0.03	0.21
WM03	0.57	0.40	1.00	0.16	0.21	0.20	0.07	0.12
WM05	0.04	0.22	0.16	1.00	0.38	0.39	0.26	0.21
WM06	0.04	0.47	0.21	0.38	1.00	0.51	0.09	0.45
WM07	0.08	0.29	0.20	0.39	0.51	1.00	0.15	0.34
WM08	0.12	-0.03	0.07	0.26	0.09	0.15	1.00	0.09
WM09	0.07	0.21	0.12	0.21	0.45	0.34	0.09	1.00
Total	2.12	**2.76**	**2.73**	2.66	**3.16**	**2.96**	1.75	2.49

The correlation coefficients between the wind speed of neighbouring stations are obtained by using the Pearson correlation. Wind speed is the most important factor for wind power forecasting. Therefore,

only the correlation coefficient of wind speed was used to determine which sites to use as neighbouring sites. ANNs can learn the relationship between wind speed of different sites.

Table 3.17 shows the wind speed Pearson correlation between the target stations and their three most related neighbouring stations. The range of the correlation coefficient is between -1 to 1. The correlation coefficient of +1 indicates a perfect positive correlation (i.e., positive relationship); the coefficient of -1 indicates a perfect negative correlation (i.e., negative relationship). The coefficient closes to zero means that the relationship between the variables is weak.

Table 3.17: Mean wind speed correlation between the target stations and the corresponding three most correlated neighbouring stations

Target station	Most correlated	Second most correlated	Third most correlated	Mean correlation coefficient
WM01	WM03	WM02	WM08	0.30
WM02	**WN06**	**WM03**	**WM07**	**0.39**
WM03	**WM01**	**WM02**	**WM06**	**0.40**
WM05	WM07	WM06	WM08	0.34
WM06	**WM07**	**WM02**	**WM09**	**0.48**
WM07	**WM06**	**WM05**	**WM09**	**0.42**
WM08	WM05	WM07	WM01	0.18
WM09	WM06	WM07	WM02	0.34

As can be seen in Table 3.17, the mean correlation coefficient of WM06 and its three most related neighbouring stations is the highest. WM06 has a stronger correlation with its second-closest neighbouring station (WM07) instead of the nearest station (WM02). The results are consistent with what was discussed previously that the distance between two stations is not the only factor that determines the wind speed correlation between two stations. Based on the distance and correlation figures, four stations (WM02, WM03, WM06, WM07) are selected as target stations in four experiments in this case study because they have a relatively high correlation with their neighbouring stations and shorter mean distance away from their neighbouring stations. These four target stations are WM02, WM03, WM06 and WM07. Figure 3.9 shows the structure of the ANN-based spatial correlation model.

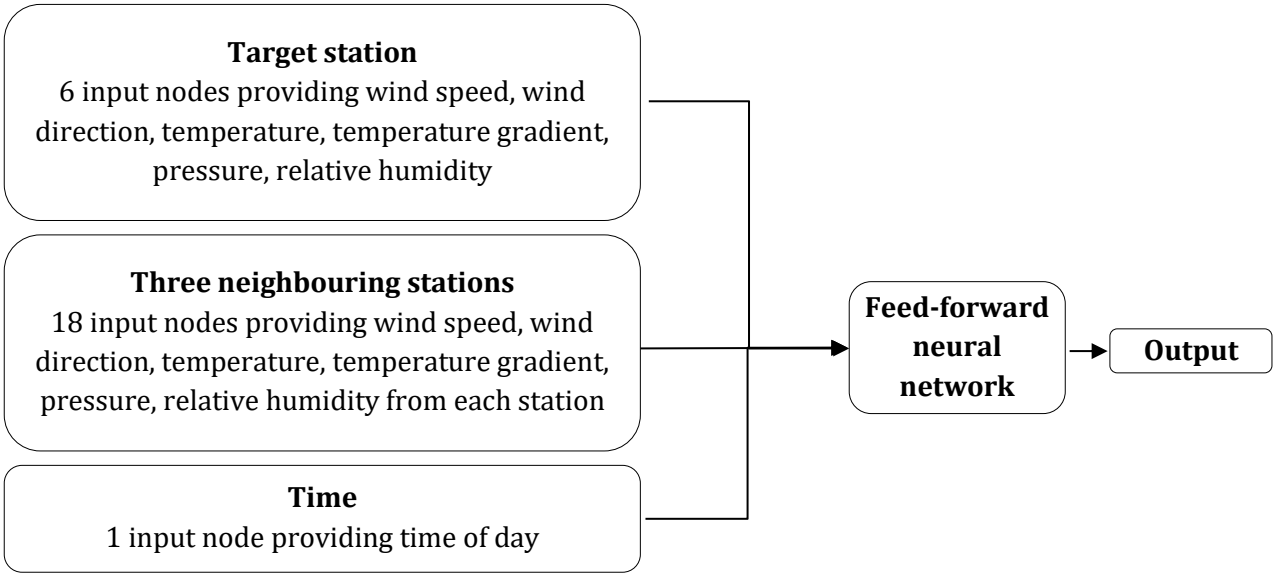


Figure 3.9: Structure of ANN-based spatial correlation model

All six meteorological variables from each station together with the time of day were used as input features to train the feedforward artificial neuron network with two hidden layers. A total of 25 input features are utilised in this research. Figure 3.10 shows the structure of the forecasting model used for spatial correlation analysis in this case study.

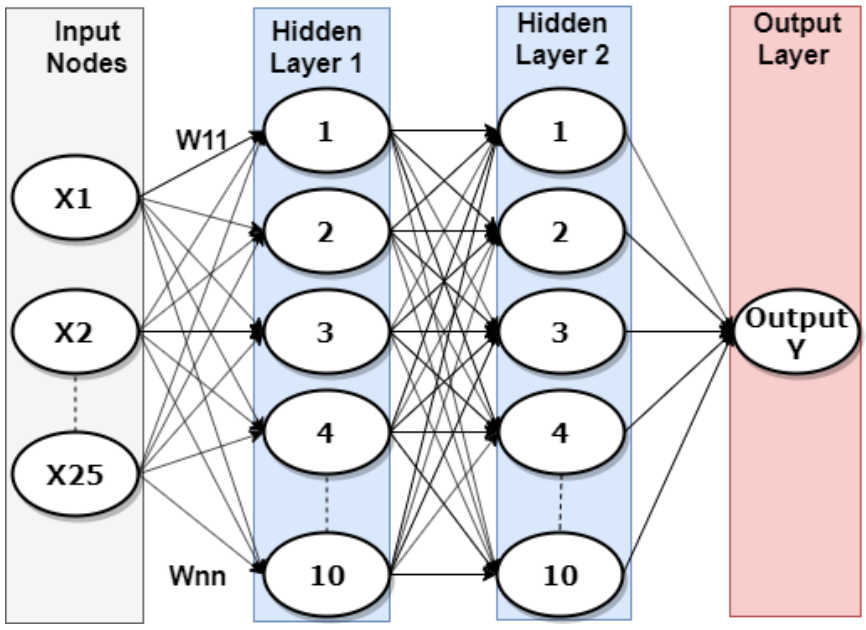


Figure 3.10: The structure of the forecasting model used for spatial correlation analysis in this case study

The ANNs with two hidden layers is used because input features from neighbouring sites complicated the forecasting task. The output of the forecasting model is the wind speed of the target station.

3.2.3 Spatial correlation case study two (NREL)

As shown in Figure 3.2 in Section 3.1.1, many available sites from NREL can be used for spatial correlation analysis. Site ID61118 was used as the target station, and its surrounding sites were used as neighbouring sites. The first step is to use the variables of the closest adjacent stations as the input features to the forecasting model of the target site. The second step is to include more sites that are slightly far away from the target site (ID61118) until the performance of the forecasting model stop improving or the training time become too long. Figure 3.11 shows the structure of the forecasting model with input features from multiple neighbouring sites.

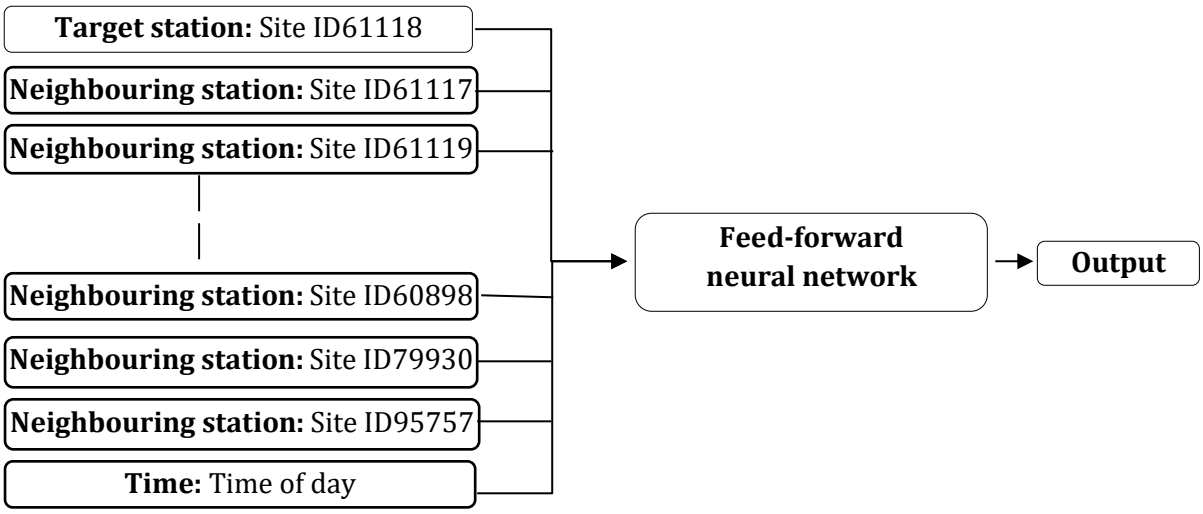


Figure 3.11: Structure of the forecasting model with the input features from multiple neighbouring stations

As can be seen in Figure 3.11, the meteorological variables of many neighbouring stations together with the time of day were used as input features to the forecasting model. A total of 30 stations were used. The meteorological variables from highly correlated neighbouring stations can enhance the forecasting performance of the target station. Although there are many sites from NREL that can be used as neighbouring sites, including too many of them might overfit the forecasting model. A total of 29 neighbouring sites of NREL were used for the spatial correlation analysis.

3.3 Artificial neural networks (ANNs)

It has been shown that different input features, model structures, training algorithms, learning rates and other adjustable parameters can directly affect forecasting models' performance [16]. One of the objectives of this research is to find suitable parameters for ANNs. The following steps were used to build the feedforward ANN in MATLAB R2019a:

The first step is to determine the suitable number of samples and input features for the model. A total of 300000 samples from each data sources were used in this dissertation. The data is divided into three groups which are a training group, a validation group, and a testing group. Training data contains 170000 samples, validation data contains 30000 samples, and testing data contains 100000 unseen samples.

The next step is to find a suitable ANNs structure and configurations based on the sample size, input feature and problem complexity. The structure of ANNs is mostly depended on the problem. The number of input neurons is equal to the number of input features, and the number of output neurons is equal to the size of the output data. In this research, the output of the network is either predicted wind speed or wind power. The number of hidden layers and hidden neurons is depended on many factors such as the number of inputs, the number of outputs, problem complexity and et cetera.

The forecasting performance of ANNs is sensitive to the number of hidden neural layers and the number of hidden neurons. In this research, the number of hidden layer and hidden neurons are selected by using the rule-of-thumb and trial and error approach. Different combinations of hidden layers and hidden neurons are tested, and the best performer is selected for the final model configuration. Examples of some combinations of numbers of hidden layers and hidden neurons that have been considered in this dissertation are listed in Table 3.18. To focus on the investigation of the effect of the number of hidden layers on the forecasting performance of the model and to keep a single variable for each experiment, the same number of hidden neurons was used in each hidden layer. Different numbers of hidden neurons in different hidden layers will be investigated in further research.

Table 3.18: Some combinations of different numbers of hidden layers and hidden neurons tested in this research

Number of hidden layers	Number of hidden neurons
1	5
1	10
1	15
1	20
1	25
2	[5, 5]
2	[10, 10]
2	[15, 15]
3	[5,5,5]

3	[10,10,10]
3	[15,15,15]
4	[5, 5, 5, 5]
4	[10, 10, 10, 10]
4	[15, 15, 15, 15]

The only variables in this test are the number of hidden layers and hidden neurons. The initial number of hidden neurons is set to 5, and the number was increased by five each time until the model stops improving. Figure 3.12 shows the structure of a feedforward ANN with two hidden layers.

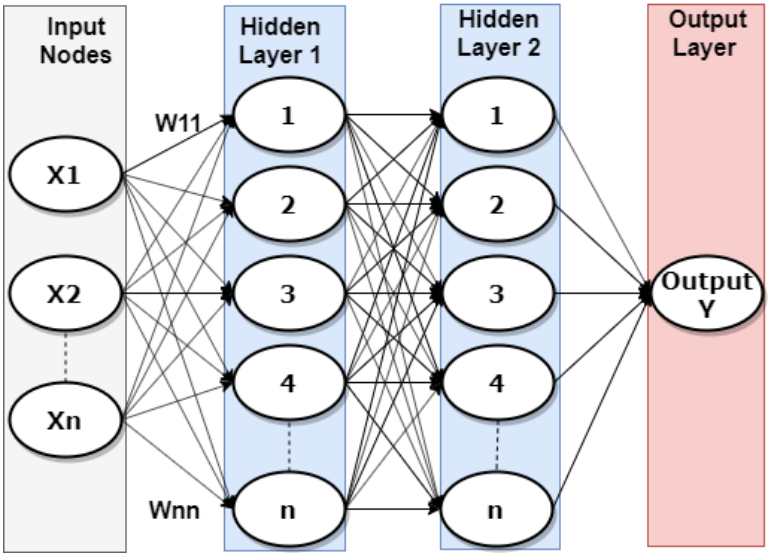


Figure 3.12: Structure of a feedforward ANN model with two hidden layers

As can be seen in Figure 3.12, the feedforward ANN contains two hidden layers, and each hidden layer contains *n* hidden neurons. The relationship between the inputs and output of each node is defined as [10]:

$$Y_i = f_i(\sum_{j=1}^{n} W_{ij}.X_j + b_i) \qquad (3.3)$$

where X_j is the jth input, Y_i is the output, W^{ij} is the connection weight, b_i is the bias, and f_i is the activation function [10].

The next step is to find a suitable training algorithm. The backpropagation algorithm is one of the most commonly used algorithms for ANNs. Levenberg-Marquardt (LM) back-propagation algorithm is a popular backpropagation algorithm. It is the combination of the Gauss-Newton method and the steepest descent method. LM is defined as:

$$X_{K+1} = X_K - [J^T J + \mu I]^{-1} J^T e \qquad (3.4)$$

where *x* is the training weight, μ is the damping factor, *J* is the Jacobian matrix of the first derivatives of the network errors with respect to the weights and biases, *I* is the identity matrix, and *e* is a vector of network errors. When μ is zero, LM becomes Newton's method. When μ is large, LM becomes gradient descent with a small step size [107]. LM has the speed advantage of the Gauss-Newton method and stability advantage of the steepest descent method. Therefore, Levenberg-Marquardt back-propagation algorithm is utilised in this research. The performance of the PSO-trained ANNs is also investigated in this research to see whether ANN(PSO) can enhance the forecasting performance of LM trained ANNs.

The initial weights and bias are then generated randomly. The training process starts after all the parameters are set. The mean square error between the input data and the target data is the objective function of all ANNs used in this dissertation. Simulation of ANNs stops when any one of the following conditions is met:

- The number of epochs is reached (i.e. 1000 epochs).
- The mean square error is reduced to zero.
- The gradient is lower than 10^{-8}.
- The μ exceeds 10^{9}.
- The mean square error (MSE) of the validation set increased six times (recommended by the ANN toolbox to avoid overfitting) since it last time deceased.

Each simulation of ANNs has different initial conditions that slightly affect the performance of the forecasting model. The evaluation results will be slightly different for each simulation. Repeating the searching process of ANNs several times is the most effective way to evaluate the performance of an ANNs configuration. Therefore, each test performed in this dissertation is simulated five times to minimize the random error, and the final results are the aggregate of these five simulations [48].

3.4 Particle swarm optimization (PSO)

Particle velocity and position of PSO are updated using the following equations [84]:

$$v_i(t+1) = \omega v_i(t) + c_1 r_1\big(p_i - x_i(t)\big) + c_2 r_2\left(p_{gd} - x_i(t)\right) \qquad (3.5)$$

$$x_i(t+1) = x_i(t) + v_i(t+1) \qquad (3.6)$$

where v_i is particle velocity, x_i is current particle position; ω is inertia weight; c_1 and c_2 are acceleration constant, r_1 and r_2 are randomly generated values ranging from 0 to 1 for good state-space exploration [108]; p_i and p_{gd} are the best position of an individual particle and a global particle, respectively. As shown in Eq. (3.5), particle velocity is determined by three components, namely the momentum components, the cognitive components and the social components [84].

The fitness function needs to be determined before running the optimization process. The fitness function in this experiment is mean square error (MSE) between input data and target data of ANNs. Figure 3.13 shows the flowchart of the particle swarm optimization algorithm.

The first step is to determine suitable parameters for PSO. The values of PSO parameters are set based on the literature recommendation and experiments. Different authors have different opinions on setting the values of c_1, c_2. The values are usually ranging from 0.5 to 2, and most authors set the same value for both c_1 and c_2 [83], [85], [87]. Different combinations of acceleration constants are tested in this dissertation. The dimension of particles depends on the optimization problem. In this research, the dimension of the particle is equal to the total number of ANNs weights and bias. The next step is to determine the inertia weight. The bigger the inertia weight, the larger the flying velocity, which results in a more significant exploration area [84]. Usually, inertia weight has a bigger value in the early stage and decrease linearly toward a small value in the later stage [83]. Therefore, linear decreasing inertia weight defined as Eq. (3.7) is used in this dissertation [51], [84].

$$\omega = \omega_{max} - \left(\frac{\omega_{max} - \omega_{min}}{T_{max}}\right)t \qquad (3.7)$$

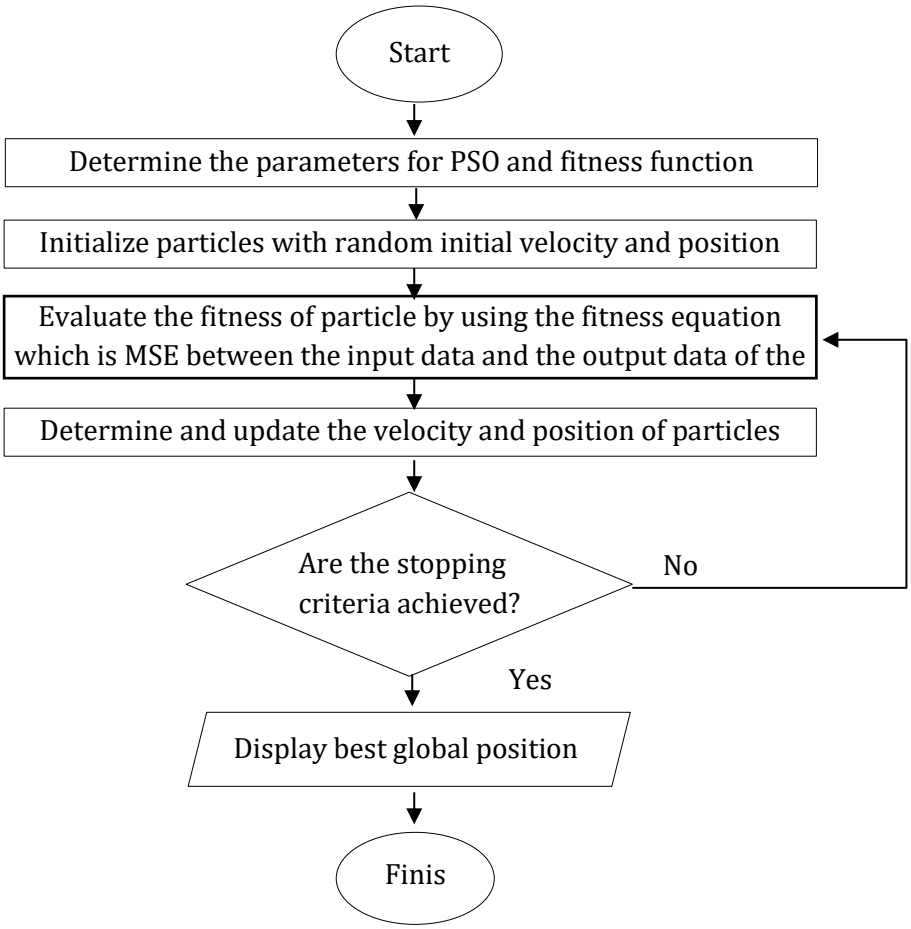


Figure 3.13: Flowchart of particle swarm optimization algorithm

where ω_{max} and ω_{min} are the maximum and minimum inertia weights, respectively; T_{max} is the maximum number of iterations and t is the current iteration number. Typical values for ω_{max} and ω_{min} are 0.9 and 0.4, respectively [84]. Authors in [85] concluded that the PSO algorithm hardly converges when ω is smaller than 0.4 or ω is greater than 0.85. The algorithm has a high convergence rate when ω is between 0.5 and 0.75. The number of swarm particles also affect the performance of PSO. Too many particles will increase the training time, and too few particles might weaken the global search ability of PSO [83], [85]. The number of particles is usually proportional to the dimension of the problem [83]. Different numbers of particle population are tested in this research. Table 3.19 summarizes the range of PSO parameters tested in this research.

Table 3.19: Parameters of PSO tested in this research

Parameter Name	Range
Acceleration constant c_1, c_2	0.5 to 2
Particle dimension D_p	D_p = number of ANNs weights
Inertia weight ω	ω_{max} = 0.25 to 1, ω_{min}= 0 to 0.5
Maximum iteration T_{max}	T_{max} = 1000
Particle population size	5 to 90
Upper and lower bounds for particle	Upper Bound = 0.5; Lower Bound =-0.5

The maximum iteration number was fixed to 1000. Each optimization experiment was run five times with random initial values of *x* and *v*. The range of each parameter shown in Table 3.19 is decided based on the recommendations provided in the literature.

3.5 Ensemble

One of the most common problems encountered by many authors is to select suitable model parameters and input features for forecasting models. Ensemble model can take advantages of different forecasting models to produce better final results. In a homogeneous ensemble, the outputs of a single type of forecasting model with different model parameters or different input features are used to obtain a final output. Two types of the homogeneous ensemble are tested in this dissertation, namely model-parameter-based ensemble and input-feature-based ensemble. The model-parameter-based ensemble uses outputs from the ANN model with different model parameters to form the final output. The input-feature-based ensemble uses the outputs of the same model with different combinations of input features to determine the final output. Figure 3.14 shows the structure of the model-parameters ensemble used in two experiments in this research.

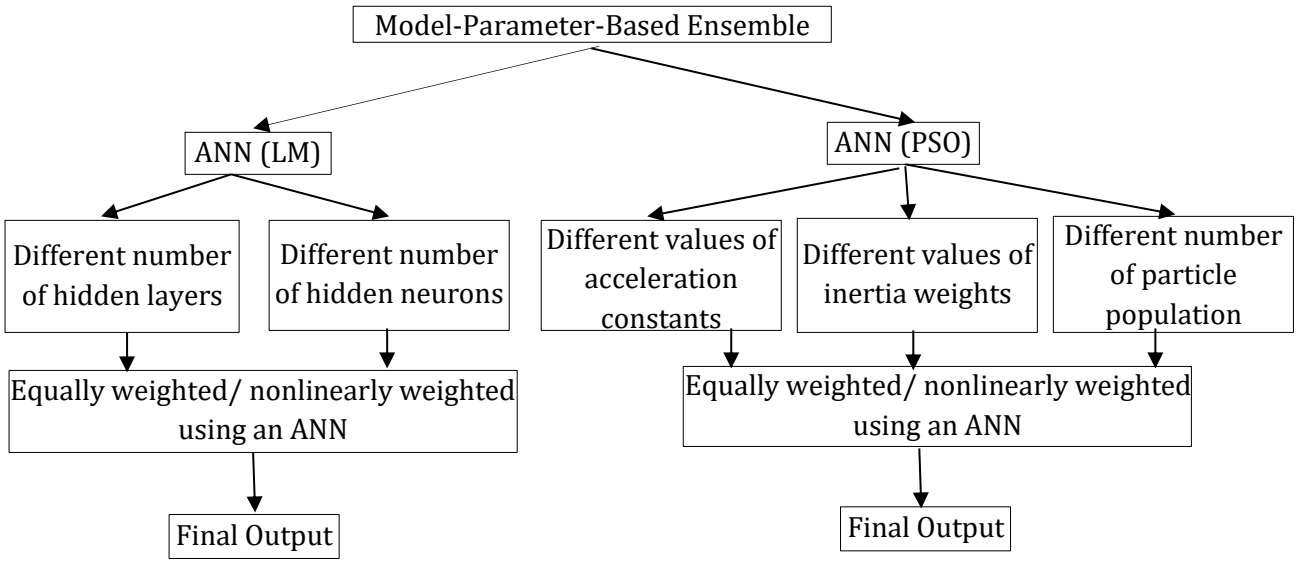


Figure 3.14: Block diagram of the model-parameter-based ensemble methods implemented in this research

Two different forecasting models (ANN(LM) and ANN(PSO)) were used in two separate experiments to check if model-parameter can enhance the forecasting performance of the best individual model of the previous sections. Two crucial parameters for ANN(LM) are the number of hidden layers and the number of hidden neurons. Therefore, the outputs of the ANN(LM) with different numbers of hidden layers and hidden neurons are equally weighted, or nonlinearly weighted by using another ANN to get the final output. Three essential parameters for PSO are acceleration constants, inertia weights and particle population number. Different values of these parameters were used to produce different outputs which were then equally or nonlinearly weighted to get a final output.

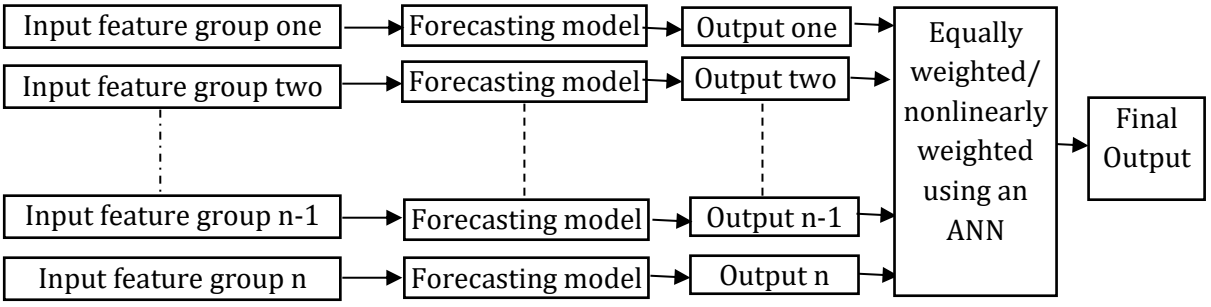


Figure 3.15: Block diagram of the input-feature-based ensemble model

The input-feature-based ensemble model uses multiple ANNs to reduce the number of input features of each member model to avoid overfitting issue. The parallel structure of the input-feature-based ensembles mimics the structure of ANNs which allows them to use parallel processing method to run the similar process at the same time to improve the computation speed and robustness of the model [3], [66]. Figure 3.15 shows the block diagram of the structure of the input-feature-based ensemble.

As shown in Figure 3.15, the input-feature-based ensemble uses multiple same types of forecasting models to take different input feature combinations from the target site or neighbouring sites to produce initial outputs (see Output one to Output n in Figure 3.15). The initial outputs are then either equally weighted or non-linearly weighted based on the forecasting performance of the member models to get a final output.

3.6 Wind speed to wind power conversion

The most crucial factor of wind power generation is wind speed. The indirect forecasting method is utilised in this research. The first step of the indirect forecasting is to forecast future wind speed. The second step is to convert the forecasted wind speed to wind power by using a suitable power curve. The datasets from both SGWF and NREL contain wind power data; the power curves of both data sources are determined using the measured wind speed and wind power data. The WASA datasets do not contain any wind power generation data. Therefore, a suitable power curve is used to simulate the wind power from the actual and predicted wind speed. The wind turbine selection procedure for the WASA datasets is shown in the following sections.

3.6.1 Wind turbine selection

Table 3.20 summarises the commonly used wind turbines in South Africa wind farms [109].

Table 3.20: Summary of the commonly used wind turbines in South Africa wind farms

Turbine type		Rating (MV)	Number of turbines	Total rating (MV)	Brand total (MV)	Percentage
	Vestas V112	3.0	66	198		
	Vestas V100/2000	2.0	44	88		
	Vestas V100/1800	1.8	78	140.4	563.6	31.6%
	Vestas V90/2000	2.0	47	94		
Vestas	Vestas V90/1800	1.8	24	43.2		
United Power	United Power UPC 86	1.5	163	244.5	244.5	13.7%
Suzlon	Suzlon S88	2.1	66	138.6	138.6	7.8%
Sinovel	Sinovel SL3000	3.0	18	54	54	3.0%
Siemens	Siemens 2.3MW	2.3	217	219.3	219.3	12.3%
	NordexN90	2.5	32	80		
	NordexN117/3000	3.0	37	111	425.4	23.9%
	Nordex N100	2.5	40	100		
Nordex	Nordex N117	2.4	56	134.4		
Acciona	Acciona	3.0	46	138	138	7.7%

As indicated in Table 3.20, Vesta, United Power, Siemens and Nordex are trendy brands for South Africa wind farms. The power curve of the Vesta V90/2000 wind turbine is selected in this project due to its

popularity in South Africa wind farms. Table 3.21 summaries the technical information about Vesta V90/2000 [110].

Table 3.21: Datasheet of the wind turbine Vesta V90

Turbine model: Vesta V90	
Rated power	2000 kW
Cut-in wind speed	4 m/s
Rated wind speed	13 m/s
Cut-out wind speed	25 m/s
Rotor diameter	90 m
Swept area	6362 m2
Hub height	80/95/105 m

Figure 3.16 shows the power curve of the wind turbine Vesta V90.

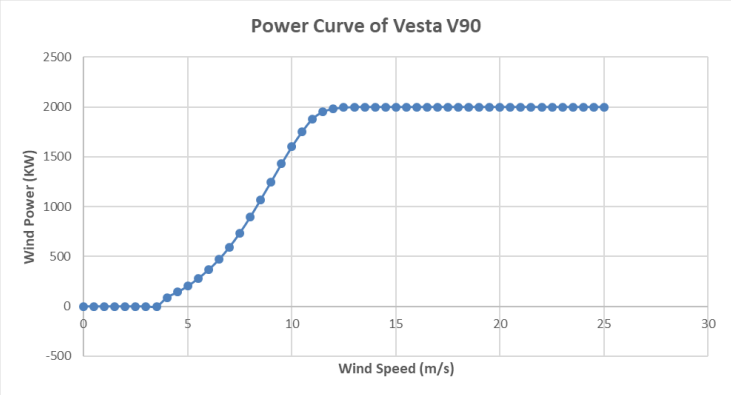


Figure 3.16: Power curve of Vesta V90

As shown in Figure 3.16 [110], the relationship between the available wind power and actual power is nonlinear. There is no power output when wind speed is less than the cut-in wind speed (4 m/s). The rapid growth of power output can be seen between the cut-in wind speed and rated wind speed(13m/s). The rated power is achieved when the wind speed is higher than the rated wind speed, and before the cut-off speed is attained (25 m/s) [43]. The curve between the cut-in wind speed and rated wind speed is estimated by Eq. (2.3).

3.6.2 Prediction interval

A prediction interval provides a specific probability that the forecasted output will fall within the interval. The prediction interval can be written as

$$\hat{y}_{t+h|t} \pm c\hat{\sigma}_h \qquad (3.8)$$

where $\hat{\boldsymbol{y}}_{\boldsymbol{t+h|t}}$ is the h-step ahead forecast, $\hat{\boldsymbol{\sigma}}_{\boldsymbol{h}}$ is an estimate of the standard deviation of the h-step forecast distribution, c is the multiplier which depends on the coverage probability. The value of the multiplier and the corresponding percentage interval is shown in Table 3.22.

The value of prediction intervals provides the uncertainty in the forecasts. The higher the confidence level, the bigger the multiplier which results in a more significant uncertainty range of the forecasting results.

Table 3.22: The value of the multiplier and the corresponding probability percentage

Percentage	Multiplier
50	0.67
55	0.76
60	0.84
65	0.93
70	1.04
75	1.15
80	1.28
85	1.44
90	1.64
95	1.96
99	2.58

3.7 Discussions and summary

Datasets from SGWF, NREL, and WASA are collected and used in this research. All three data sources provide good quality and large quantity datasets. Different input variables such as wind speed, wind direction, temperature, pressure, relative humidity, air density have a different value range. The input variable with a more significant range can affect the one with a smaller range. A popular min-max normalisation technique was used to standardize the range of each input variable. The normalized data were then divided into three groups, namely training set, validation set and testing set. Two experiments were set up to determine the impact of input data resolution and input sample size on the forecasting performance of a basic ANN which contains a hidden layer and ten hidden neurons. Pearson correlation, together with a basic ANN, was used to determine the importance of each input feature to the performance of short-term wind speed and wind power forecasting. Different model parameters and model structures of ANNs were used to check the impact of the number of hidden layers, the number of hidden neurons and the types of training algorithm on short-term wind speed forecasting performance. Two ensemble models, namely model-parameter-based ensemble and input-feature-based ensemble, were set up to take advantages of each member model to mitigate the chance of selecting a weak predictor and avoiding overfitting issue encountered by a single model. A popular wind turbine was picked to convert forecasted wind speed to forecasted wind power for the WASA dataset. Finally, a prediction interval was used to find the uncertainty of forecasted wind power.

4. Simulation results and discussions of the single model approach

The purpose of this chapter is to show and discuss the simulation results of the single model approach. The chapter starts with Section 4.1, showing the impact of input features on short-term wind speed and wind power forecasting performance of the ANN model. In Section 4.2, the simulation results of the impact of the data resolution and sample size on the forecasting model are shown and discussed. The effect of model parameters on the performance of the ANN model is shown and discussed in Section 4.3. Section 4.4 shows the simulation results of the ANN (PSO), the effect of different parameters of the PSO on the forecasting performance of the ANN (PSO) is also discussed in this section. The chapter finishes with Section 4.5, showing the impact of input features of neighbouring stations on the forecasting performance of the ANN model of the target station.

All the simulations were performed on a computer with Intel Core i7-7500U CPU @ 2.70GHz and 8GB of RAM using MATLAB R2019a. Each simulation is repeated five times to validate the repeatability of the algorithm and provide quantitative and convincing results [48].

4.1 Feature selection

This section shows the simulation results of the input feature selection.

4.1.1 Feature selection for wind speed forecasting

Table 4.1 summarises the Pearson correlation coefficients between wind speed and meteorological variables of the WASA and NREL datasets.

Table 4.1: Summary of Pearson correlation coefficients between wind speed and other meteorological variables

Data source	Input feature	Correlation coefficients between wind speed and other meteorological variables (no time lag)
WASA	Wind speed	1.00
	Wind direction	0.11
	Temperature gradient	-0.24
	Temperature	0.24
	Relative humidity	-0.10
	Barometric pressure	-0.10
NREL	Air density	-0.07

The first column of Table 4.1 shows the data sources used for the input feature selection analysis. The reason to use two data sources is that the NREL dataset provides additional meteorological variable such as air density, for input feature analysis. The third column of Table 4.1 display the Pearson correlation coefficients between the wind speed and the meteorological variables at the same time instance. Wind speed has a low correlation with other meteorological variables, especially with air density. Figure 4.1 shows the correlation between wind speed and other meteorological variables and their lag terms.

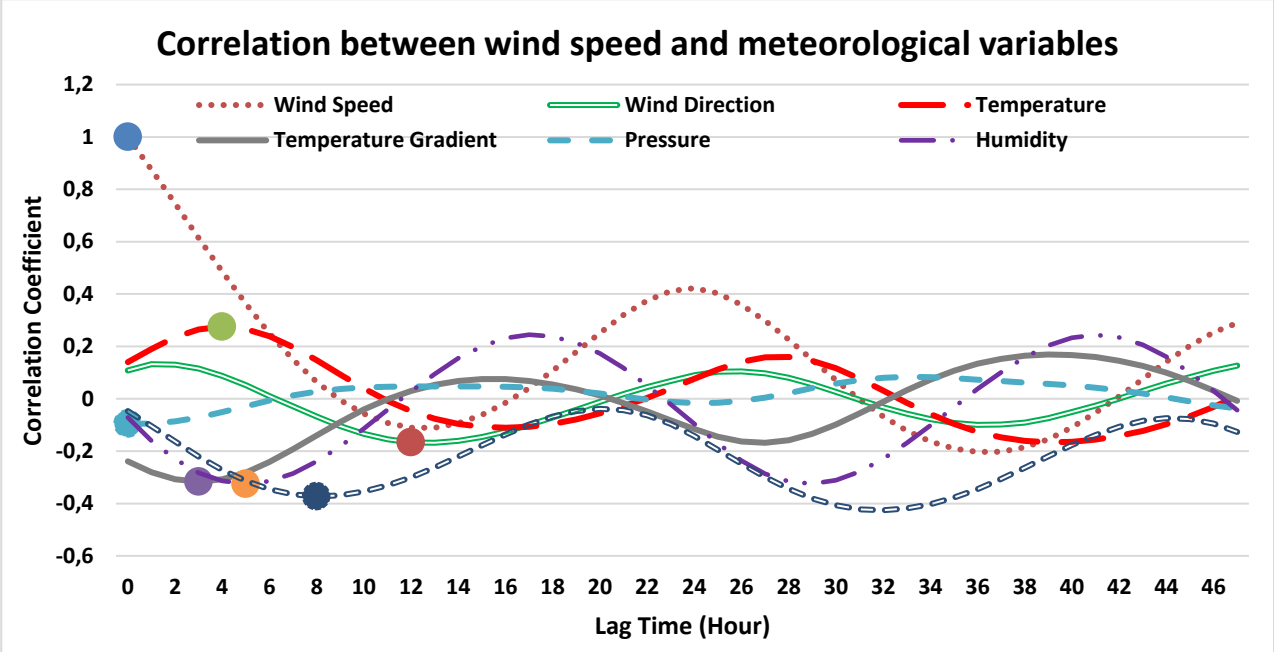


Figure 4.1: Correlation between wind speed and meteorological variables with lag time from 0 hour to 48 hours.

Circles in Figure 4.1 indicate the lagging time when the highest correlation between wind speed and other variables occurs. As an example, the light green circle in Figure 4.1 indicates that wind speed has a higher correlation with 4 hours ago temperature. The present wind speed usually has higher correction coefficients with lagged meteorological variables because it requires some time for the changes in these meteorological variables to take effect on the wind speed. Therefore, the lagged term of meteorological variables can also provide valuable information to the forecasting model. The correlation coefficients alone are not enough to determine which input features to use to train the forecasting model because two highly correlated variables might share similar characteristics which do not provide additional information to the forecasting model.

As can be seen in Figure 4.1, wind speed is the most important input feature for wind speed forecasting, as the current wind speed has a strong correlation with the previously measured wind speeds. Wind speed has a relatively higher correlation with temperature, relative humidity and temperature gradient than atmospheric pressure and wind direction. The correlation between wind speed and pressure is oscillating around zero, which indicates that the correlation between wind speed and pressure is weak for all the time instances. The cyclic patterns in Figure 4.1 suggest that observations from the same hour of the previous day can be useful input features to forecasting models.

Table 4.2 shows the influence of each meteorological variable on the short-term wind speed forecasting performance of the basic ANN model. The second column of the table shows the input feature used to train the basic ANN model. The third to fifth columns indicate the mapping error between wind speed and the meteorological variables. The last three columns show the correlation coefficient (R) for the training, validation and testing of the basic ANN.

Table 4.2: Summary of the nonlinear mapping results between wind speed and other input variables (WASA)

Test no.	Input feature	NRMSE	NMAE	MAPE	Training R	Validation R	Test R	Ranking
4.1	WS	0%	0%	0%	1	1	1	1
4.2	WD	14.0%	11.3%	34.6%	0.484	0.459	0.446	2
4.3	TI	14.0%	11.3%	34.5%	0.449	0.452	0.448	3

4.4	TG	15.2%	12.5%	38.1%	0.286	0.273	0.280	4
4.5	P	15.3%	12.6%	38.6%	0.246	0.249	0.249	5
4.6	AD* (NREL)	15.1%	12.3%	38.2%	0.192	0.191	0.179	6
4.7	T	15.5%	12.8%	39.3%	0.169	0.168	0.161	7
4.8	RH	15.6%	13.0%	39.7%	0.133	0.130	0.133	8

The numbers in the first column of Table 4.2 represent the test number for reference. For example, Test 4.1 uses wind speed only as an input feature to train the ANN model. Test 4.3 uses the ANN model to map time of day as input to train the ANN model. Based on the simulation results of Test 4.2 and Test 4.3, it can be seen that other than wind speed itself, wind direction and time of day are highly correlated with wind speed. Correlation between wind speed and other meteorological variables such as temperature, relative humidity, barometric pressure at the same time instance is weak. The last column of the table ranks the variables according to the results of three evaluation metrics (i.e., NRMSE, NMAE and MAPE) and correlation coefficients (R) of training, validation and testing. Wind speed is the most influential input feature for wind speed forecasting. It is inaccurate to simply combine several meteorological variables that have a high correlation with wind speed as input features. This is because two highly correlated variables might provide similar information to the forecasting model and therefore, may not necessarily improve the forecasting accuracy.

The following table (Table 4.3, Table 4.4 and Table 4.5) summarise the impact of different input variables combinations on the forecasting performance of a basic ANN model. Table 4.3 shows the simulation results of 2-hours-ahead wind speed forecasting results of the basic ANN model with wind speed and another variable as input features.

Table 4.3: Summary of the performance of the basic ANN model with wind speed and another variable as input features used to determine the second most important input feature for short-term wind speed forecasting (WASA)

Test no.	Input feature	NRMSE	NMAE	MAPE	% of NRMSE improvement over Test 4.9	Ranking
4.9	WS	10.5%	8.3%	25.5%	-	-
4.10	WS, TI	9.7%	7.7%	23.1%	7.6%	1
4.11	WS, TG	9.7%	7.6%	23.2%	7.6%	1
4.12	WS, RH	10.2%	8.0%	24.5%	2.9%	2
4.13	WS, T	10.2%	8.0%	24.6%	2.9%	2
4.14	WS, WD,	10.2%	8.1%	24.7%	2.9%	2
4.15	WS, P	10.5%	8.3%	25.5%	0%	3

As shown in Table 4.3, time of day (TI) and temperature gradient (TG) are the second most influential variables for wind speed forecasting (Test 4.10 and Test 4.11). The forecasting performance of the model improved by 7.6% by adding TI or TG to WS as input features. Even though wind direction has a relatively high correlation with wind speed (see Test 4.2 in Table 4.2), the input combination of wind speed and wind direction does not yield a better result than Test 4.10 and Test 4.11. The forecasting model does not improve a lot when both wind speed and wind direction are used (see Test 4.14) because wind speed shares some common information with the wind direction. The simulation results of Test 4.8, Test 4.9 and Test 4.12 indicate that the more influential input feature (WS) has a bigger impact on the forecasting results than the less influential input feature (RH). Addling less influential input feature the model does not degrade the forecasting performance because weight training process allows ANNs to assign bigger weights to more influential input features. Although temperature, barometric pressure,

and relative humidity have shared different information with wind speed, they also provide less information to the forecasting model due to the low correlation with wind speed.

Based on the results shown in Table 4.2 and Table 4.3, it can be seen that wind speed (WS), temperature gradient (TG), and time of day (TI) are the three most influential input features for short-term wind speed forecasting. Results shown in Table 4.4 is used to determine the fourth most important input feature. As can be seen in Table 4.4, wind direction is the fourth most important input feature for short-term wind speed forecasting. Relative humidity, air temperature and barometric pressure have a small impact on the forecasting performance of the basic ANN. This is verified by the results shown in Table 4.5.

Table 4.4: Summary of the performance of the basic ANN model with wind speed, time of day, temperature gradient and another variable as input features to determine the fourth most important input feature for short-term wind speed forecasting (WASA)

Test no.	Input feature	NRMSE	NMAE	MAPE	% of NRMSE improvement over Test 4.16	Ranking
4.16	WS, TI, TG	9.5%	7.4%	22.7%	-	-
4.17	WS, TI, TG, WD	9.0%	7.0%	21.3%	5.3%	1
4.18	WS, TI, TG, RH	9.2%	7.2%	22.0%	3.5%	2
4.19	WS, TI, TG, T	9.4%	7.3%	22.4%	1.1%	3
4.20	WS, TI, TG, P	9.4%	7.3%	22.3%	1.1%	3

Table 4.5: Summary of the performance of the basic ANN model with different input feature combinations (WASA)

Test no.	Input feature	NRMSE	NMAE	MAPE	% of NRMSE improvement over Test 4.21	Ranking
4.21	WS, TI, TG, WD	9.0%	7.0%	21.3%	-	-
4.22	WS, TI, TG, WD, T, P RH	8.8%	7.2%	22.0%	2.2%	1
4.23	WS, TI, TG, WD, P	8.9%	6.9%	21.0%	1.1%	2
4.24	WS, TI, TG, WD, T	8.9%	6.9%	21.1%	1.1%	3
4.25	WS, TI, TG, WD, RH	8.9%	6.9%	21.2%	1.1%	4

The simulation results shown in Table 4.5 indicate that relative humidity, air temperature and barometric pressure have a minimal impact on the performance of the forecasting model because the model does not yield a significant improvement (see Test 4.23 to Test 4.25) with these input variables. However, including all the available meteorological variables can slightly enhance forecasting accuracy, as shown in Test 4.22. If computational power is strong enough and the number of training samples is big enough, all the available meteorological variables should be used as input features to the forecasting model. The same input feature selection analysis is carried out for wind power forecasting in the next section.

4.1.2 Feature selection for direct wind power forecasting

Data from NREL were used to determine the impact of each input feature on the performance of the short-term wind power forecasting. The reason to use NREL dataset is that the dataset contains both wind power generation data and several different meteorological variables like wind direction, temperature, relative humidity and air density. The data from the other two data sources are not used for direct wind power forecasting due to lack of either wind power or meteorological variables.

Table 4.6: Summary of Pearson correlation coefficients between the wind power and meteorological variables from NREL

Data Source	Input feature	Correlation coefficients between wind power and meteorological variables (no time lag)
NREL	Wind power	1.00
	Wind speed	0.95
	Wind direction	0.15
	Temperature	0.00
	Relative humidity	-0.34
	Air density	-0.05

As shown in Table 4.6, wind power has a high correlation with wind speed. The present wind power has the highest correlation with the present wind speed. One the other hand, the present wind power has a higher correlation with the 11-hours lagged air temperature instead of the current air temperature. The correlation coefficient between the 10-hours lagged air density and the present wind power is much stronger than the present air density (see Figure 4.2). Therefore, it is essential to include lagged meteorological variables as input features to the forecasting model if the forecasting horizon is longer than 11 hours. Figure 4.2 shows the correlation coefficient between wind power and lag terms of meteorological variables.

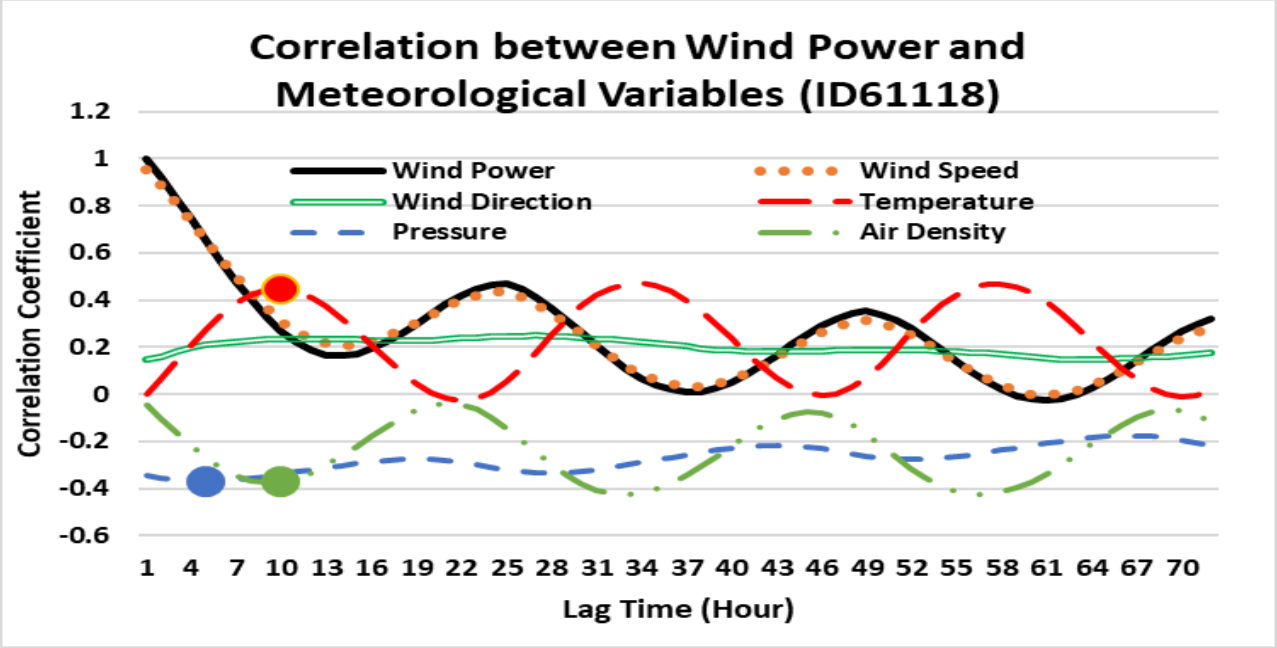


Figure 4.2: Correlation between wind power and meteorological variables from site ID61118

As can be seen in Figure 4.2, wind power has the strongest correlation with wind power and wind speed. The result is reasonable, as wind speed is the main factor that affects the wind turbine power output. Wind power also has a strong correlation with the 10-hours lagged temperature, and the 11-hours lagged air density because it requires some time for the changes in these meteorological variables to take effect on the wind speed which has a significant impact on wind power. The correlation between

wind power and the lag term of wind direction is around 0.2. Table 4.7 summaries the mapping results between wind power and input variables from NREL.

Table 4.7: Summary of the mapping results between wind power and other input variables (NREL)

No.	Input feature	NRMSE	NMAE	MAPE	Training R	Validation R	Test R	Ranking
4.26	WP	0%	0%	0%	1.000	1.000	1.000	1
4.27	WS	0.8%	0.5%	1.2%	0.999	0.999	0.999	2
4.28	WD	29.5%	24.0%	59.5%	0.549	0.546	0.553	3
4.29	TI	32.1%	27.6%	68.5%	0.357	0.364	0.365	4
4.30	P	32.8%	28.1%	69.8%	0.374	0.372	0.373	5
4.31	T	33.6%	29.6%	73.5%	0.212	0.206	0.206	6
4.32	AD	33.6%	29.7%	73.6%	0.207	0.199	0.209	6

As shown in Table 4.7, correlation coefficients (R) have a strong influence on the mapping performance of the basic ANN model. In general, the higher the correlation coefficient between the two variables, the better the mapping results. Historical wind power and wind speed data can map wind power very well. The result indicates that historical wind power and wind speed data are crucial for wind power forecasting. The simulation results of 2-hours-ahead wind power forecasting of a basic ANN model for input feature analysis is shown in Table 4.8 to validate the above observation.

Table 4.8: Experiment results of the basic ANN model uses wind power as the base input feature (NREL)

Test no.	Input feature	NRMSE	NMAE	MAPE	% of NRMSE Improvement over Test 4.33	Ranking
4.33	WP	19.5%	14.0%	34.8%	-	-
4.34	WS	19.3%	13.8%	34.2%	1.0%	-
4.35	WP, TI	17.7%	12.2%	30.1%	9.2%	1
4.36	WP, T	18.5%	13.0%	32.3%	5.1%	2
4.37	WP, AD	18.8%	13.2%	32.8%	3.6%	3
4.38	WP, WS	18.9%	13.3%	33.0%	3.1%	4
4.39	WP, WD,	19.3%	13.7%	34.0%	1.0%	5
4.40	WP, P	19.3%	13.7%	34.0%	1.0%	5

The results of Test 4.33 and Test 4.34 shown in Table 4.8 indicates that both historical wind speed and wind power are essential input features for wind power forecasting. The results also indicated that using the historical wind speed only to forecast wind power yield slightly better results than using the historical wind power only to forecast wind power. The wind power value becomes zero when the wind speed is lower than the cut-in wind speed and when the wind speed is higher than the cut-off wind speed. The discontinuity of the historical wind power time series makes it harder for the forecasting model to predict future wind power accurately.

The combination of the two most correlated input features (wind power and wind speed- see Test 4.38) does not outperform other combinations shown in Table 4.8. The is due to the common characteristics shared by wind power and wind speed. On the other hand, Test 4.35 and Test 4.36 improved more than 5% from Test 4.33. Even though TI and T have slightly weaker correction with wind power compare to wind speed, the diverse information provided by TI and T enhanced the forecasting model performance.

The results in Table 4.8 shows that using wind speed as the input to forecast wind power directly can also be very effective. Table 4.9 shows the 2-hours ahead forecasting performance of the basic ANN model with wind speed as the base input feature.

Table 4.9: Experiment results of the basic ANN model uses wind speed as the base input feature (NREL)

Test no.	Input feature	NRMSE	NMAE	MAPE	% of NRMSE improvement over corresponding results in Table 4.8
4.41	WS	19.3%	13.8%	34.2%	-
4.42	WS, WD	18.9%	13.3%	33.0%	0.5%
4.43	WS, T	18.5%	13.0%	32.1%	0.0%
4.44	WS, P	19.2%	13.6%	33.6%	0.5%
4.45	WS, AD	18.8%	13.1%	32.6%	0.0%
4.46	WS, TI	17.7%	12.2%	30.3%	0.0%

The results from Table 4.8 and Table 4.9 are very similar, which indicates that historical wind speed and wind power data have a similar effect on the wind power forecasting performance. Therefore, the indirect wind power forecasting method can be used to forecasting wind power when the historical wind power data are unavailable.

Table 4.10: Experiment results of the basic ANN for determining the third most influential input feature for wind power forecasting (NREL)

Test no.	Input feature	NRMSE	NMAE	MAPE	% of NRMSE improvement from Test 4.47	Ranking
4.47	WP, WS	18.9%	13.3%	33.0%	-	-
4.48	WP, WS, TI	17.6%	12.1%	29.9%	6.7%	1
4.49	WP, WS, T	18.4%	12.9%	32.0%	2.7%	2
4.50	WP, WS, WD	18.7%	13.1%	32.5%	1.1%	3
4.51	WP, WS, AD	18.8%	13.2%	32.6%	0.5%	4
4.52	WP, WS, P	18.9%	13.2%	32.7%	0%	5

The results in Table 4.10 shows that time of day (TI) is the third most influential input feature after wind speed and wind power for short-term wind power forecasting. Wind speed, wind power, and time of day were used as base input features to determine the fourth most important input feature for wind power forecasting. Table 4.11 shows the experiment results for determining the fourth most important input feature.

Table 4.11: Experiment results of the basic ANN for determining the fourth most influential input feature for wind power forecasting (NREL)

Test no.	Input feature	NRMSE	NMAE	MAPE	% of NRMSE improvement from Test 4.53	Ranking
4.53	WP, WS, TI	17.6%	12.1%	29.9%	-	-
4.54	WP, WS, TI, WD, T, P, AD	16.7%	11.5%	28.5%	5.1%	1
4.55	WP, WS, IT, WD	17.3%	12.0%	29.7%	1.7%	2
4.56	WP, WS, IT, T	17.4%	11.8%	29.4%	1.1%	3
4.57	WP, WS, IT, P	17.5%	11.9%	29.5%	0.5%	4
4.58	WP, WS, IT, AD	17.5%	11.9%	29.6%	0.5%	4

The simulation results shown in Table 4.11 indicate that wind direction (see Test 4.55) is the fourth most important input feature for short-term wind power forecasting. Although air temperature, barometric pressure, and air density have a small effect on the performance of short-term wind power forecasting, including all the available input features do slightly enhance the forecasting performance of the basic ANN model (see Test 4.54). Table 4.12 summaries the input feature rankings for both short-term wind speed and wind power forecasting.

Table 4.12: Input features ranking for both wind speed forecasting and wind power forecasting based on the simulation results of the above Tables

Ranking	Variable name	
	Wind speed forecasting	**Wind power forecasting**
1	wind speed (WS)	wind speed (WS), wind power (WP)
2	time indicator (TI), temperature gradient (TG)	time indicator (TI)
3	wind direction (WD)	wind direction (WD)
4	relative humidity (RH)	temperature (T)
5	temperature (T), pressure (P)	pressure (P), air density (AD)

As can be seen in Table 4.12, wind speed and time indicator have significant impacts on the forecasting performance of short-term wind speed and wind power. It is suggested in the literature that barometric pressure and temperature have a small to no impact on the short-term wind speed and wind power forecasting performance. The simulation results verify that barometric pressure and air temperature have small impacts on forecasting performance. Some authors did not use barometric pressure and air temperature to forecast short-term wind speed and wind power [6]. However, including barometric pressure and air temperature as input features can slightly enhance the forecasting performance if the training sample size is big enough. Authors in [16] and [45] also included temperature and pressure as input features to train forecasting models. Pearson correlation coefficient can effectively indicate the importance of each input variable to the forecasting model when only one variable is used as input. Combining most highly correlated variables as input features to the model does not always yield a better forecasting performance, because highly correlated variables might share some common information to the model. In general, all the available input features should be used to train the forecasting model if there are no computational constraints and sample size limitation.

4.2 Data resolution and sample size

4.2.1 Date resolution

The data obtained from three data sources contain different data resolutions. The effect of the data resolution on the short-term wind speed forecasting performance of the basic ANN is shown in this section. Table 4.13 summarises the evaluation results of 2h-ahead wind speed forecasting of three types of data resolution.

Table 4.13: Evaluation results of 2h-ahead wind speed forecasting of the basic ANN for data resolution analysis

Data resolution	SGWF			NREL			WASA		
	NRMSE	NMAE	MAPE	NRMSE	NMAE	MAPE	NRMSE	NMAE	MAPE
5 min	-	-	-	6.8%	4.7%	15.5%	-	-	-
10 min	4.3%	3.1%	21.1%	6.8%	4.7%	15.6%	8.9%	6.9%	21.3%
1 hour	4.5%	3.3%	21.2%	7.3%	5.1%	15.9%	9.4%	7.3%	21.5%

As shown in Table 4.13, higher training data resolution usually result in better forecasting performance. Although the training model requires to forecast more steps ahead with high training data resolution than the lower data resolution (i.e. ten minutely data requires to predict 12 steps ahead to achieve 2h-ahead forecasting horizon, whereas hourly data requires to predict two steps ahead to achieve 2h-ahead forecasting horizon), the forecasting model performs better with high-resolution data. One advantage of using low-resolution data to train the model is that it requires less time to train the forecasting model. The results in Table 4.13 shows that the same forecasting model produces different results with different training data resource, which is one of the main barriers to compare forecasting model performance with other researchers accurately. The same analysis was carried out for 6h-ahead wind speed forecasting to verify the above observations. Table 4.14 shows the evaluation results of 6h-ahead wind speed forecasting.

Table 4.14: Evaluation results of 6h-ahead wind speed forecasting of the basic ANN for data resolution analysis

Data resolution	SGWF			NREL			WASA		
	NRMSE	NMAE	MAPE	NRMSE	NMAE	MAPE	NRMSE	NMAE	MAPE
5 min	-	-	-	9.7%	7.0%	23.1%	-	-	-
10 min	4.3%	3.1%	21.1%	9.6%	6.9%	22.9%	12.1%	9.6%	29.4%
1 hour	5.6%	4.2%	27.0%	10.3%	7.4%	23.2%	12.7%	10.0%	29.6%

As shown in Table 4.14, a similar conclusion made for 2h-ahead wind speed forecasting can be made for 6h-ahead wind speed forecasting. The forecasting model performed slightly better with the higher training data resolution than the lower data resolution. Therefore, the data with a sample per ten minutes resolution was used in this dissertation.

4.2.2 Sample size

Figure 4.3 shows the RMSE values of the forecasting model against the training sample size.

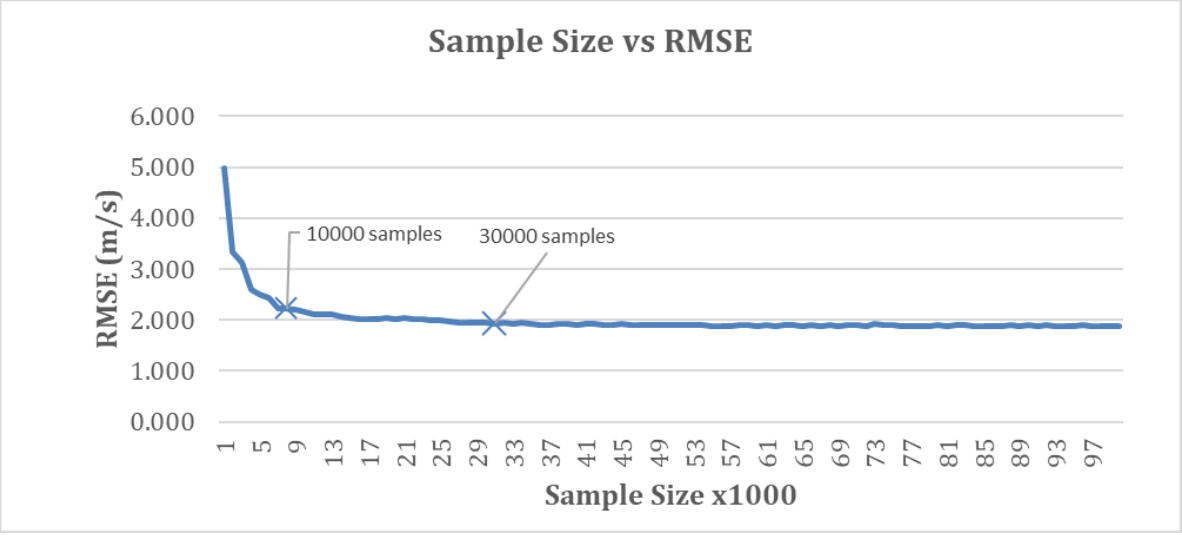


Figure 4.3: RMSE values of the forecasting model against the training sample size

As can be seen in Figure 4.3, the forecasting model performance is improved with an increased number of training samples. The performance of the model improves rapidly when the sample size is increased from 1000 samples to 10000 samples. However, there is no noticeable performance improvement when the sample size is over 30000. Figure 4.4 shows a graph of RMSE against sample size between 30000 to 100000 to provide a clear view.

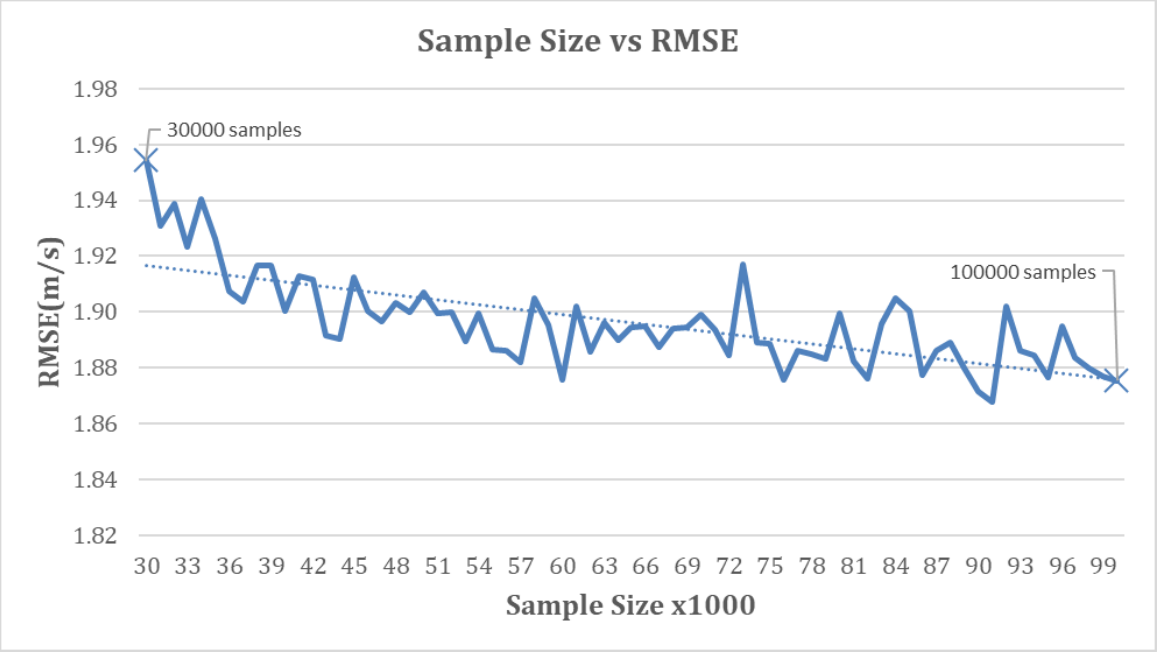


Figure 4.4: Training sample size (30000-100000) against the forecasting model performance

As can be seen in Figure 4.4, the forecasting model with 100000 training samples performs slightly better than the model with 30000 training samples. Based on Figure 4.3 and the trend lines in Figure 4.4, it can be concluded that the model with a larger number of training samples generally performs better than the one with a smaller number of training samples. However, a larger number of training samples require more time to train. Figure 4.5 shows the graph of the training sample size against the training time. The straight line in Figure 4.5 is the trend line, and the nonlinear curve is the number of sample size against the training time.

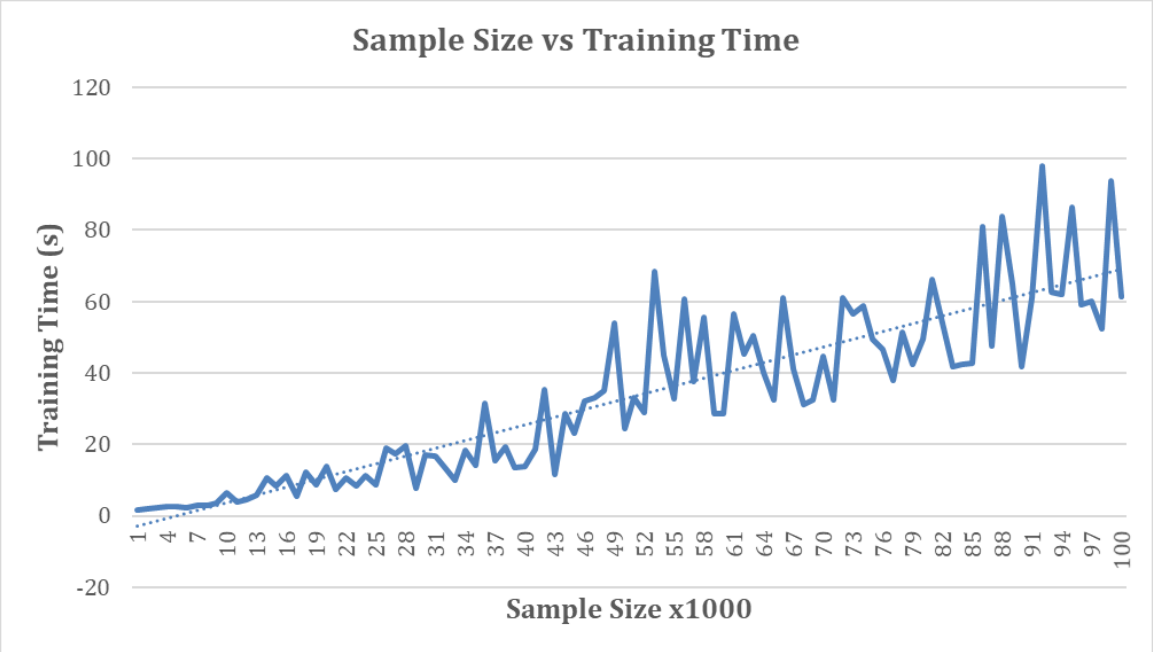


Figure 4.5: Training time against training sample size

As shown in Figure 4.5, it usually takes a longer time for the model to train a larger number of training samples. The trade-off between training time and forecasting performance can be determined based on Figure 4.4 and Figure 4.5. The forecasting performance is more important than the training time in this dissertation because the forecasting model is trained offline. Authors in [67] suggested that at least one

year of training data is needed to include the seasonal variation. Therefore, a larger number of training samples is preferred in this dissertation.

4.2.3 The cyclic pattern of wind speed

As mentioned earlier in Section 3.1.3, daily wind speed contains a cyclic pattern. Wind speed is usually higher in the afternoon than in the morning. Three approaches are proposed in this research to combat the daily cyclic pattern of wind speed. The first approach includes time indicator as an input feature to tell the forecasting model the time of day. The second approach uses moving average to mitigate wind speed difference between morning and afternoon. The third approach uses two models to predict morning wind speed and afternoon wind speed separately. Table 4.15 shows the simulation results of the three approaches of WASA: WM03 dataset.

Table 4.15: Simulation results of three approaches used to tackle the cyclic pattern of wind speed (WASA: WM03)

Test no.	Model	RMSE (m/s)	MAE (m/s)	MAPE	% of RMSE imp. with respect to Test 4.59
4.59	Basic ANN	2.23	1.77	25.5%	-
4.60	Basic ANN + time indicator	2.06	1.61	23.2%	7.6%
4.61	Moving average approach	2.10	1.63	23.7%	5.8%
4.62	Morning model	2.22	1.71	30.5%	-5.4%
	Afternoon Model	2.48	1.95	23.8%	
	Mean (Morning, Afternoon)	2.35	1.84	27.2%	

Test 4.59 directly forecasts the wind speed without using any approaches to mitigate the effect of the daily cyclic pattern of wind speed. As can be seen in Table 4.17, both time indicator approach and moving average approach can help the forecasting model to mitigate the effect of the cyclic pattern of wind speed. However, the average evaluation results of the morning model and the afternoon model (Test 4.62) are worse than the basic ANN model (Test 4.59). The morning model has better RMSE and MAE values than the afternoon model because of the variation of wind speed in the morning is much calmer than the one in the afternoon. However, the calmer morning wind speed makes the denominator of MAPE equation smaller, which result in a more significant MAPE value in the morning. The morning model has a similar performance as the whole day model, but the aggregate results of the morning model and afternoon model are worse than the whole day model due to the bad performance of the afternoon model. The discontinuity of the two-models approach also contributes to the worse short-term wind speed forecasting results. Based on the results shown in Table 4.15, the first approach (use time indicator as an input feature) is utilised in the final model to capture the daily cycle and seasonality patterns of the wind speed time series.

4.3 Artificial neural networks

4.3.1 Number of hidden neurons

As indicated earlier, the number of hidden neurons in the hidden layer is depended on many factors. Different problems require different numbers of hidden neurons to obtain an optimal solution. However, there is no mathematical equation that can be used to determine an optimal number of hidden neural neurons for all the problems. Therefore, the trial and error method was used to determine a suitable number of hidden neurons in this research, as other authors did in [27], [62], [75]. The experiment configuration for determining a suitable number of hidden neurons is shown in Table 4.16. All three data sources were used to check the generalization ability of the forecasting model.

Table 4.16: Summary of the model configurations utilised to determine a suitable number of hidden neurons for short-term wind speed and wind power forecasting

Forecasting model	Feedforward neural network with a single hidden layer
Number of hidden layers	1
Training algorithm	LM
Type of forecasting	2h-ahead wind speed forecasting
Input features (SGWF)	WP, WS, WD, TI
Input features (WASA: WM03)	WS, WD, T, TG, P, RH, TI
Input features (NREL: ID61118)	WP, WS, WD, T, P, AD, TI
Data resolution and sample size	Resolution: 10min, sample size: 300, 000
Variable parameter	Number of hidden neurons

The purpose of this experiment is to determine a suitable number of hidden neurons. Therefore, the only variable for this experiment is the number of hidden neurons. The forecasting horizon is fixed to 2 hours ahead for all three data sources. The simulation results of SGWF is shown in Table 4.17.

Table 4.17: Simulation results of 2h-ahead wind speed forecasting of the feedforward ANN with a single hidden layer and different numbers of hidden neurons (SGWF)

Test No.	Number of hidden neurons	RMSE	MAE	MAPE	NRMSE	NMAE
4.63	1	1.63	1.22	21.2%	4.1%	3.1%
4.64	5	1.62	1.21	21.0%	4.1%	3.0%
4.65	10	1.61	1.20	20.9%	4.0%	3.0%
4.66	15	1.60	1.19	20.8%	4.0%	3.0%
4.67	20	1.61	1.29	20.8%	4.1%	3.0%
4.68	25	1.60	1.19	20.8%	4.0%	3.0%
4.69	30	1.60	1.19	20.8%	4.0%	3.0%

As can be seen in Table 4.17, the number of hidden neurons does not have a significant impact on the forecasting performance of the model. The performance of the forecasting model stops improving when the number of hidden neurons is increased above 15. The SGWF dataset contains only three meteorological variables (WS, WD, TI). Therefore, the relationship between the target variable (WS) and input variables (WS, WD, TI) is relatively simpler compared to the one with more input variables. The ANN model requires a smaller number of hidden neurons to determine a less complicated relationship. Therefore, a higher number of hidden neurons does not have a significant impact on the model forecasting performance (see Test4.66-Test 4.69). Table 4.18 shows the simulation results of the same experiment for the NREL: ID61118 dataset.

The number of hidden neurons has a more significant impact on the NREL dataset than the SGWF because there are more input features from the NREL dataset than the SGWF dataset. It required more hidden neurons to train more complex problems. The model with 20 or more hidden neurons yields a better result than the one with lesser number of hidden neurons. Table 4.19 shows the simulation results of the same experiment for the WASA: WM03 dataset.

Table 4.18: Simulation results of 2h-ahead wind speed forecasting of the feedforward neural network with a single hidden layer and different numbers of hidden neurons (NREL: ID61118)

Test no.	Number of hidden neurons	RMSE	MAE	MAPE	NRMSE	NMAE
4.70	1	1.66	1.18	17.38%	7.36%	5.23%
4.71	5	1.56	1.08	15.90%	6.90%	4.79%
4.72	10	1.53	1.06	15.58%	6.78%	4.69%
4.73	15	1.53	1.05	15.44%	6.75%	4.65%
4.74	20	1.52	1.04	15.31%	6.72%	4.61%
4.75	25	1.52	1.04	15.30%	6.72%	4.61%
4.76	30	1.52	1.04	15.24%	6.71%	4.59%

Table 4.19: Simulation results of 2h-ahead wind speed forecasting of the feedforward neural network with a single hidden layer and different numbers of hidden neurons (WASA: WM03)

Test no.	Number of hidden neurons	RMSE	MAE	MAPE	NRMSE	NMAE
4.77	1	2.05	1.61	23.28%	9.66%	7.61%
4.78	5	1.93	1.50	21.69%	9.11%	7.09%
4.79	10	1.90	1.47	21.28%	8.95%	6.95%
4.80	15	1.89	1.46	21.10%	8.89%	6.89%
4.81	20	1.88	1.46	21.06%	8.87%	6.88%
4.82	25	1.88	1.46	21.03%	8.87%	6.87%
4.83	30	1.87	1.45	20.96%	8.84%	6.85%
4.84	35	1.87	1.45	20.94%	8.84%	6.84%
4.85	40	1.87	1.44	20.85%	8.81%	6.81%
4.86	45	1.88	1.45	20.95%	8.85%	6.84%

The forecasting performance of the model has a noticeable improvement when the number of hidden neurons is increased from 1 to 15. Based on the simulation results of all three data sources, it can be concluded that the feedforward ANN with 15 or more hidden neurons is sufficient to determine the nonlinear relationship between the input variables (different meteorological variables) and the target variable (WS). The RMSE values of the forecasting results of three data sources are plotted in Figure 4.6.

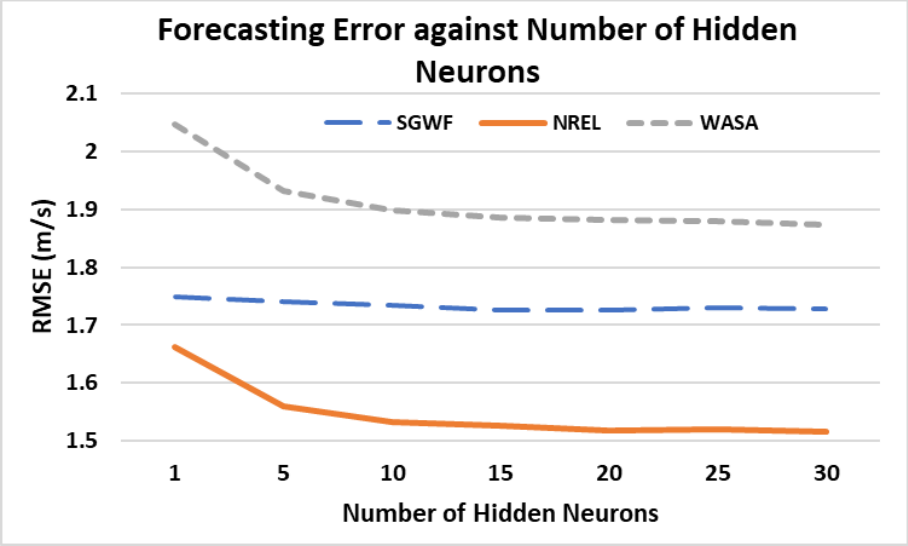


Figure 4.6: RMSE of the forecasting results of three data sources against the number of hidden neurons

As shown in Figure 4.6, the RMSE values of the forecasting results of the NREL and WASA datasets are decreasing with an increasing number of hidden neurons. The SGWF dataset contains only three

variables, wind speed, wind direction and time indicator. The increasing number of hidden neurons do not have a noticeable effect on the forecasting performance of the SGWF dataset. Therefore, the forecasting model does not need more hidden neurons to learn the relationship between input and output of the SGWF data.

Although the same forecasting model is used to train the data from three data sources, the RMSE values of the forecasting are different. The site with fewer ramp events and lower average wind speed will have lower RMSE values.

4.3.2 Number of hidden layers

Usually, an ANN model with a single hidden layer is enough to tackle non-linear problems effectively. The effect of the number of hidden layers on the forecasting performance of the feedforward ANN was determined in this section. The training configurations used to determine a suitable number of hidden layers are shown in Table 4.20.

Table 4.20: Summary of the configurations utilised to determine a suitable number of hidden layers for short-term wind speed and wind power forecasting.

Forecasting model	Feedforward neural network
Training algorithm	LM
Type of forecasting	2h-ahead wind speed forecasting
Input features (SGWF)	WS, WD, TI
Input features (WASA: WM03)	WS, WD, T, P, TG, RH, TI
Input features (NREL: ID61118)	WS, WD, T, P, AD, TI
Data resolution and sample size	Resolution: 10min, sample size: 300, 000
Variable parameter	Number of hidden neurons and hidden layers

The only variables in this experiment are the number of hidden neurons and hidden layers. Table 4.21 shows the simulation results of 2h-ahead wind speed forecasting of a feedforward neural network of SGWF with different numbers of hidden layers. As shown in Table 4.21, a larger number of hidden layers does not have a significant impact on the forecasting performance of the model when the total number of hidden neurons is bigger than 10. The SGWF dataset contains fewer variables than the other two data sources. Therefore, the model with one to two hidden layers is strong enough to handle the SGWF dataset.

Table 4.21: Simulation results of 2h-ahead wind speed forecasting of a feedforward neural network with different numbers of hidden layers (SGWF)

Test no.	**Number of hidden neurons**	**Number of hidden layers**	**RMSE**	**MAE**	**MAPE**	**NRMSE**	**NMAE**
4.87	5	1	1.62	1.21	21.0%	4.1%	3.0%
4.88	5, 5	2	1.61	1.19	21.0%	4.1%	3.0%
4.89	5, 5, 5	3	1.63	1.21	21.0%	4.1%	3.0%
4.90	5, 5, 5, 5	4	1.62	1.21	21.1%	4.1%	3.0%
4.91	10	1	1.61	1.20	20.9%	4.0%	3.0%
4.92	**10, 10**	**2**	**1.59**	**1.19**	**20.7%**	**4.0%**	**3.0%**
4.93	10, 10, 10	3	1.60	1.19	20.7%	4.4%	3.0%
4.94	10, 10,10,10	4	1.61	1.19	20.8%	4.0%	3.0%

4.95	15	1	1.60	1.19	20.8%	4.0%	3.0%
4.96	15, 15	2	1.60	1.19	20.8%	4.0%	3. 0%
4.97	15, 15, 15	3	1.60	1.19	20.8%	4.0%	3.0%
4.98	15, 15, 15, 15	4	1.60	1.19	20.8%	4.0%	3.11%

Table 4.22 shows the simulation results of 2h-ahead wind speed forecasting of a feedforward neural network of NREL: ID61118 with different numbers of hidden layers.

Table 4.22: Simulation results of 2h-ahead wind speed forecasting of a feedforward neural network with different numbers of hidden layers (NREL: ID61118)

Test no.	Number of Hidden neurons	Number of hidden layers	RMSE	MAE	MAPE	NRMSE	NMAE
4.99	5	1	1.56	1.08	15.90%	6.90%	4.79%
4.100	5, 5	2	1.54	1.06	15.57%	6.83%	4.69%
4.101	5, 5, 5	3	1.54	1.06	15.58%	6.81%	4.69%
4.102	5, 5, 5, 5	4	1.54	1.06	15.58%	6.80%	4.69%
4.103	10	1	1.53	1.06	15.58%	6.78%	4.69%
4.104	**10, 10**	**2**	**1.51**	**1.03**	**15.18%**	**6.69%**	**4.57%**
4.105	10, 10, 10	3	1.53	1.04	15.25%	6.78%	4.59%
4.106	10, 10,10,10	4	1.57	1.05	15.40%	6.94%	4.64%
4.107	15	1	1.53	1.05	15.44%	6.75%	4.65%
4.108	15, 15	2	1.53	1.03	15.16%	6.75%	4.57%
4.109	15, 15, 15	3	1.54	1.04	15.34%	6.83%	4.62%
4.110	15, 15, 15, 15	4	1.60	1.07	15.66%	7.09%	4.71%

As shown in Table 4.22, the model with two hidden layers performs slightly better than the model with a single hidden layer. The forecasting performance of the model with four hidden layers degrades noticeably because the model is overfitted with the training data. The same experiment is carried out for the WASA: WM03 dataset and the simulation results are shown in Table 4.23.

Table 4.23: Simulation results of 2h-ahead wind speed forecasting of a feedforward neural network with different numbers of hidden layers (WASA: WM03)

Test no.	Number of hidden neurons	Number of hidden layers	RMSE	MAE	MAPE	NRMSE	NMAE
4.111	5	1	1.93	1.50	21.69%	9.11%	7.09%
4.112	5, 5	2	1.89	1.47	21.23%	8.93%	6.94%
4.113	5, 5, 5	3	1.89	1.47	21.16%	8.91%	6.91%
4.114	5, 5, 5, 5	4	1.90	1.47	21.29%	8.95%	6.95%
4.115	10	1	1.90	1.47	21.28%	8.95%	6.95%
4.116	**10, 10**	**2**	**1.87**	**1.45**	**20.61%**	**8.83%**	**6.84%**
4.117	10, 10, 10	3	1.88	1.45	20.99%	8.85%	6.86%
4.118	10, 10,10,10	4	1.88	1.45	20.88%	8.85%	6.82%
4.119	15	1	1.89	1.46	21.10%	8.89%	6.89%
4.120	15, 15	2	1.89	1.45	20.99%	8.90%	6.86%
4.121	15, 15, 15	3	1.88	1.45	20.96%	8.88%	6.85%

4.122	15, 15, 15, 15	4	1.89	1.46	21.05%	8.92%	6.88%

The simulation results indicate that the model with two hidden layers and ten hidden neurons in each hidden layer has the best performance among all the combinations shown in Table 4.23. Based on the simulation results of the data sources, it can be concluded that the model with two hidden layers and ten hidden neurons in each hidden layer is more suitable for the short-term wind speed forecasting in this research.

4.4 ANN-PSO

The training configurations used to determine suitable parameters for PSO are shown in Table 4.24. The impacts of three important parameters of PSO on the forecasting performance of the ANN are investigated in this section. As shown in Appendix D, the ANN (PSO) model with five hidden neurons has a better short-term wind speed forecasting performance than the ANN (PSO) with ten hidden neurons. As a result, five hidden neurons were used in this experiment.

Table 4.24: Summary of the configurations utilised to determine the suitable model parameters for short-term wind speed and wind power forecasting

Forecasting model	PSO trained feedforward ANN
Number of hidden layers	1
Number of hidden neurons	5
Training algorithm	PSO
Type of forecasting	2h-ahead wind speed forecasting
Maximum iteration T_{max}	T_{max} = 1000
Input features (SGWF)	WP, WS, WD, TI
Input features (WASA: WM03)	WS, WD, T, TG, P, RH, TI
Input features (NREL: ID61118)	WP, WS, WD, T, P, AD, TI
Data resolution and sample Size	Resolution: 10min, sample size: 300, 000
Variable parameters	Acceleration constant c_1, c_2 Inertia weight ω Particle population size

Table 4.25 shows the simulation results of 2h-ahead wind speed forecasting of the ANN (PSO) for acceleration constants analysis. The only variable in this experiment is the acceleration constants, c_1, c_2.
The maximum inertia weight ω_{max} is set to 0.9 and the minimum inertia weight ω_{min} is set to 0.4. The number of particle population is set to 20 based on the size of the ANNs.

Table 4.25: Simulation results of 2h-ahead wind speed forecasting of ANN-PSO for acceleration constants analysis

Test no.	**Acceleration constant**		**SGWF**			**NREL: ID61118**			**WASA: WM03**		
	c1	c2	RMSE (m/s)	MAE (m/s)	MAPE	RMSE (m/s)	MAE (m/s)	MAPE	RMSE (m/s)	MAE (m/s)	MAPE
4.123	0.50	0.50	1.80	1.37	23.1%	1.81	1.33	19.6%	2.11	1.68	23.5%
4.124	0.75	0.75	1.68	1.25	21.7%	1.70	1.21	17.9%	2.06	1.62	22.9%
4.125	1.00	1.00	1.87	1.19	20.8%	1.78	1.29	19.2%	2.23	1.76	25.1%
4.126	1.25	1.25	1.68	1.26	21.8%	1.79	1.30	19.3%	**2.03**	**1.60**	**22.4%**
4.127	1.50	1.50	1.71	1.28	21.9%	1.80	1.32	19.5%	2.04	1.60	22.5%

4.128	1.75	1.75	1.68	1.25	21.5%	**1.67**	**1.15**	**16.9%**	2.04	1.60	22.8%
4.129	2.00	2.00	**1.66**	**1.24**	**21.5%**	1.70	1.26	18.4%	2.05	1.62	22.7%
4.130	2.25	2.25	1.70	1.27	21.9%	1.76	1.25	19.0%	2.08	1.64	23.7%
4.131	2.50	2.50	1.74	1.31	22.3%	1.69	1.20	17.8%	2.15	1.69	23.8%
4.132	3.00	3.00	1.71	1.29	22.0%	1.88	1.39	20.5%	2.16	1.67	24.4%
4.133	0.75	1.00	1.96	1.48	25.1%	1.78	1.30	19.3%	2.12	1.68	23.8%
4.134	1.00	1.25	1.70	1.28	21.8%	1.79	1.31	19.3%	2.03	1.60	22.5%
4.135	1.25	1.00	1.68	1.25	21.5%	1.78	1.28	19.0%	2.08	1.64	21.5%
4.136	1.50	2.00	1.76	1.33	22.4%	1.73	1.25	18.2%	2.08	1.64	23.2%
4.137	1.75	2.00	1.76	1.33	22.7%	1.73	1.24	18.5%	2.14	1.70	24.0%

As shown in Table 4.25, the values of acceleration constant have a significant influence on the performance of the forecasting model. The RMSE values difference between the best performer and the worst performer are 0.3 m/s, 0.21m/s and 0.2 m/s for SGWF, NREL: ID61118 and WASA: WM03, respectively. The model usually performs well when the acceleration constant, c_1 and c_2 are equal. The results are very similar to the one concluded in the literature.

Table 4.26 shows the simulation results of 2h-ahead wind speed forecasting of the ANN (PSO) for determining suitable inertia weights. The only variables in this experiment are the maximum and the minimum inertia weight, ω_{max} , ω_{min} . The number of particle population is set to 20 and the acceleration constant, c_1 and c_2, are both set to 2.

Table 4.26: Simulation results of 2h-ahead wind speed forecasting of the ANN-PSO for determining suitable inertia weights

Test	**Inertia weight**		**SGWF**			**NREL: ID61118**			**WASA: WM03**		
	ω_{max}	ω_{min}	RMSE (m/s)	MAE (m/s)	MAPE	RMSE (m/s)	MAE (m/s)	MAPE	RMSE (m/s)	MAE (m/s)	MAPE
4.138	0.25	0.00	1.70	1.27	21.9%	1.69	1.20	17.7%	2.08	1.64	23.5%
4.139	0.50	0.00	1.70	1.27	21.9%	1.67	1.18	17.4%	2.05	1.62	22.5%
4.140	0.50	0.25	1.69	1.26	21.7%	1.68	1.20	17.6%	**2.03**	**1.60**	**22.7%**
4.141	0.75	0.50	1.68	1.26	21.9%	**1.68**	**1.17**	**17.2%**	2.06	1.62	23.4%
4.142	0.80	0.40	1.68	1.26	21.6%	1.77	1.29	19.1%	2.07	1.64	23.0%
4.143	0.85	0.40	1.70	1.28	22.0%	1.73	1.28	19.1%	2.06	1.61	22.7%
4.144	0.90	0.30	1.71	1.28	22.0%	1.70	1.23	18.2%	2.04	1.61	22.5%
4.145	**0.90**	**0.40**	**1.66**	**1.24**	**21.5%**	1.70	1.26	18.4%	2.05	1.62	22.7%
4.146	0.90	0.45	1.76	1.34	22.9%	1.69	1.23	17.9%	2.15	1.70	24.0%
4.147	0.90	0.50	1.77	1.34	22.8%	1.67	1.19	17.7%	2.12	1.68	23.4%
4.148	0.90	0.60	1.68	1.26	21.6%	1.71	1.26	18.2%	2.08	1.65	22.6%
4.149	0.90	0.70	1.75	1.32	22.6%	1.71	1.23	18.1%	2.05	1.61	22.5%
4.150	1.00	0.50	1.74	1.31	22.0%	1.72	1.26	18.7%	2.07	1.64	22.8%

The impact of the inertia weights on the performance of the forecasting model is smaller compared to the acceleration constants. The RMSE values difference between the best performers and the worst performers are 0.09 m/s, 0.09 m/s and 0.12 m/s for SGWF, NREL: ID61118 and WASA: WM03, respectively. The forecasting model of SGWF dataset has the best performance when the maximum inertia weight is set to 0.9, and a minimum inertia weight is set 0.4 which are the same as the one

suggested in [84], [85]. The maximum and minimum inertia weights of the NREL and WASA datasets are also very similar to the one suggested in [84], [85].

Table 4.27 shows the simulation results of 2h-ahead wind speed forecasting of the ANN (PSO) for finding a suitable number of particle population. The only variable in this experiment is the number of particle population. The maximum inertia weight ω_{max} is set to 0.9 and the minimum inertia weight ω_{min} is set to 0.4. The acceleration constant, c_1 and c_2, are both set to 2.

Table 4.27: Simulation results of 2h-ahead wind speed forecasting of ANN-PSO for determining a suitable number of the particle population

Test no.	Population	SGWF			NREL: ID61118			WASA: WM03		
		RMSE (m/s)	MAE (m/s)	MAPE	RMSE (m/s)	MAE (m/s)	MAPE	RMSE (m/s)	MAE (m/s)	MAPE
4.151	5	1.69	1.27	21.89%	1.74	1.27	19.02%	2.07	1.64	23.38%
4.152	10	1.85	1.41	23.65%	1.96	1.45	21.64%	2.13	1.68	23.76%
4.153	20	1.66	1.24	21.49%	1.70	1.26	18.36%	2.05	1.62	22.71%
4.154	30	1.66	1.24	21.43%	1.75	1.24	18.36%	2.07	1.64	22.84%
4.155	40	1.76	1.33	23.00%	1.77	1.27	18.57%	2.04	1.60	22.45%
4.156	50	1.69	1.26	21.74%	1.71	1.22	18.20%	2.07	1.63	22.66%
4.157	60	1.66	1.24	21.51%	1.69	1.18	17.57%	2.09	1.65	22.93%
4.158	70	1.69	1.27	21.73%	1.67	1.18	17.53%	**2.04**	**1.60**	**22.08%**
4.159	80	1.68	1.26	21.85%	**1.66**	**1.18**	**17.33%**	2.05	1.61	22.31%
4.160	90	**1.66**	**1.23**	**21.37%**	1.69	1.18	17.53%	2.05	1.61	22.39%

As can be seen in Table 4.27, the ANN (PSO) usually performs better with a bigger number of particle population. Although it takes longer to train the ANN with a larger number of particle population, the model with bigger particle population can explore more combinations and also reduce the chance of trapped in a local minimum. Therefore, the number of particle population are selected based on the bold values shown in Table 4.27 for each data source.

Simulation results of Table 4.25, Table 4.26, and Table 4.27 show the impact of the acceleration constant, inertia weight and particle population on the forecasting performance of the ANN (PSO) model, respectively. Each experiment has a variable parameter and two fixed parameters, whose values are set according to the literature. In this research, the effect of each parameter on the forecasting performance of the ANN (PSO) is individually analysed. In further research, combined scenarios with different c_1, c_2 ω_{max}, ω_{min}, and population sizes will be considered.

4.5 Spatial correlation

4.5.1 WASA

i. *Wind speed forecasting*

Table 4.28 shows the evaluation results of 2h-ahead wind speed forecasting of four WASA target stations. As can be seen in Table 4.28, the wind speed forecasting performances of all four target stations are improved when the meteorological variables of three most correlated neighbouring stations are included as input features to the ANN-based spatial correlation model.

Table 4.28: Evaluation results of 2h-ahead wind speed forecasting of the ANN-based spatial correlation model of four WASA target stations

Target Station	Without Neighbouring Station			With Neighbouring Stations			% of RMSE Improvement from the one without neighbouring station
	RMSE (m/s)	*MAE (m/s)*	*MAPE*	*RMSE (m/s)*	*MAE (m/s)*	*MAPE*	
WM02	1.93	1.45	23.7%	1.78	1.34	21.9%	7.8%
WM03	1.82	1.41	19.9%	1.72	1.32	18.7%	5.5%
WM06	1.97	1.50	20.6%	1.87	1.43	19.4%	5.1%
WM07	1.80	1.38	20.6%	1.68	1.27	18.6%	6.7%

Four target stations used in this research are located in different parts of South Africa with different altitudes and terrains. Although the mean distances between the target stations and the corresponding neighbouring stations are long, and the wind speed correlations between the target stations and neighbouring stations are not strong enough, the results still indicate that using the meteorological variables from neighbouring stations can enhance the model performance of the target stations. It is expected that the forecasting model's performance of the target stations will be able to improve further if closer and more highly related neighbouring stations are available and used in the forecasting. Figure 4.7 shows the actual wind speed against 2h-ahead forecasted wind speed of station WM03 and a 95% prediction interval which is obtained by using the method shown in Section 3.6.2.

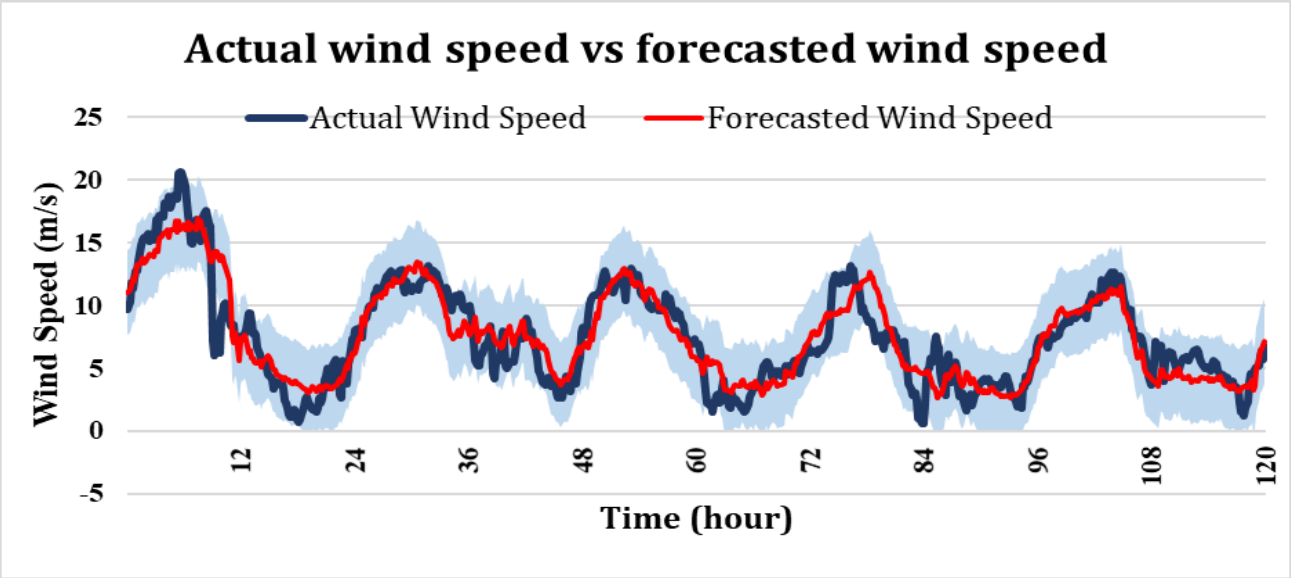


Figure 4.7: The actual wind speed against the forecasted 2h-ahead wind speed of the ANN-based spatial correlation model of station WM03

As can be seen in Figure 4.7, the forecasted wind speed can closely track the actual wind speed for most of the samples. The gaps between the actual wind speed and forecasting wind speed from hour 1 to hour 12 and from hour 72 to hour 84 are relatively bigger compared to other samples. The large gaps are due to the sudden increase and decrease in wind speed. The model was trained based on a large amount of historical data which contains less wind ramp events than the typical wind speed patterns. Therefore, the model can forecast the common wind speed patterns better than the rare wind speed ramp events. Using meteorological variables from more neighbouring stations and ramp event indicators can further improve the forecasting model.

ii. *Wind power forecasting*

The actual wind speed and forecasted wind speed are then converted to simulated wind power and forecasted wind power by using the power curve shown in Figure 3.16. Table 4.29 shows the evaluation results of 2h-ahead wind power forecasting of four case studies.

Table 4.29 Evaluation results of 2h-ahead wind power forecasting of the ANN-based spatial correlation model (WASA).

Target station	Without neighbouring station		With neighbouring station		% NRMSE improvement
	NRMSE	*NMAE*	*NRMSE*	*NMAE*	
WM02	20.6%	13.3%	19.3%	12.4%	6.3%
WM03	20.0%	13.4%	19.2%	12.7%	4.0%
WM06	20.8%	14.1%	19.5%	13.2%	6.4%
WM07	20.3%	14.0%	18.8%	12.8%	7.4%

As shown in Table 4.29, the forecasting model performs better when the meteorological variables of neighbouring stations are used as input features. The percentage of RMSE improvement shown in the last column of Table 4.29 is ranging from 4% to 7.4% for four target stations (WM02, WM03, WM06, WM07). The wind power forecasting performance should be able to improve further if more highly related neighbouring stations are available. Figure 4.8 shows the simulated wind power against the forecasted 2h-ahead wind power of WM03. The simulated wind power is converted from the actual wind speed by using the power curve shown in Figure 3.16. Forecasted wind power is obtained by converting the forecasted wind speed through the same power curve.

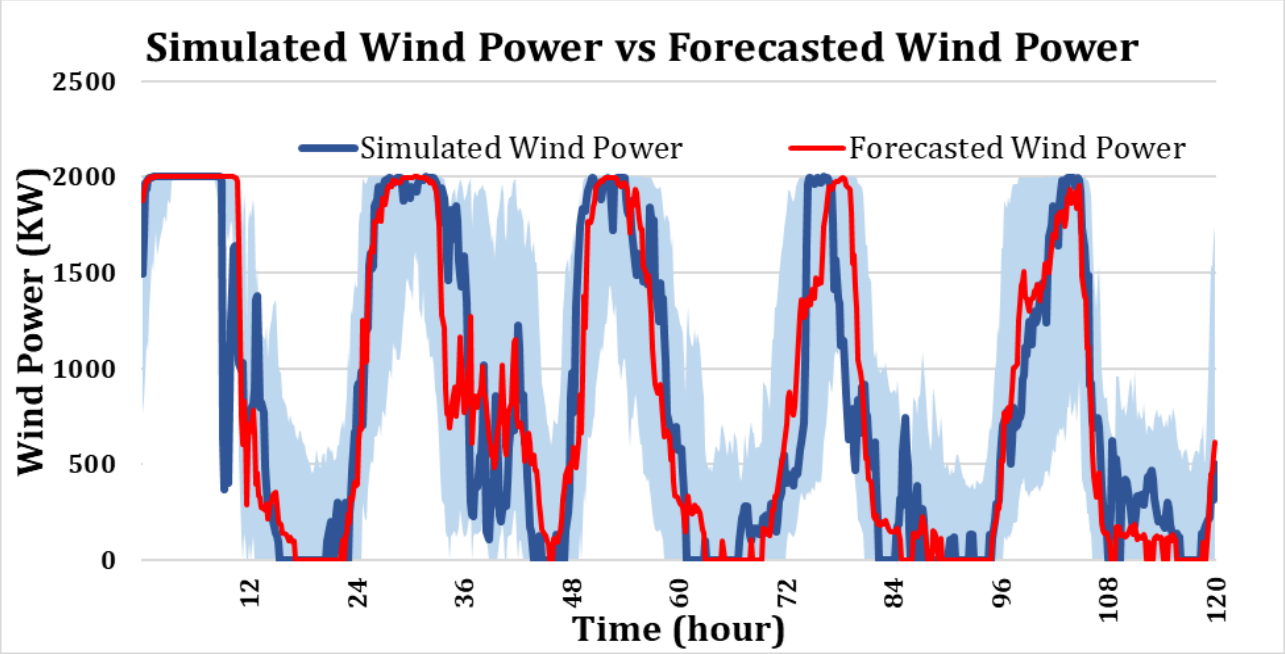


Figure 4.8: The simulated wind power against the forecasted 2h-ahead wind power of the spatial correlation- ANN of WM03

As shown in Figure 4.8, the forecasted wind power can track the simulated wind power nicely for most of the samples. The 95% prediction confidence interval covers most of the simulated wind power. The peak between hour 72 and hour 84 is narrow due to the sudden wind speed changes. The unexpected sudden changes in wind speed are difficult to forecast accurately. Wind ramp events need to be considered in further studies to improve the forecasting performance further.

4.5.2 NREL

The simulation results of this section are used to determine if input features from more neighbouring sites can enhance the forecasting performance of the target site. Only short-term wind speed forecasting is presented to verify the positive impact of more neighbouring sites on the performance of the target site in this section. Many sites are available in the NREL dataset can be used for the spatial correlation analysis. Table 4.30 shows the effect of the number of neighbouring sites on the forecasting performance of the target station (ID61118).

Table 4.30: Simulation results of 2h-ahead wind speed forecasting of the ANN-based spatial correlation model of the NREL dataset.

No. of sites used	RMSE (m/s)	MAE (m/s)	MAPE	No. of sites used	RMSE (m/s)	MAE (m/s)	MAPE
1	1.51	1.04	15.3%	16	1.38	0.94	13.8%
2	1.49	1.02	14.9%	17	1.36	0.93	13.6%
3	1.46	1.00	14.7%	18	1.35	0.91	13.4%
4	1.46	1.00	14.7%	19	1.37	0.91	13.4%
5	1.44	0.98	14.4%	20	1.35	0.92	13.5%
6	1.45	0.98	14.5%	21	1.37	0.94	13.7%
7	1.44	0.97	14.5%	22	1.33	0.90	13.2%
8	1.43	0.97	14.3%	23	1.37	0.94	13.7%
9	1.43	0.97	14.3%	24	1.41	0.93	13.5%
10	1.42	0.96	14.1%	25	1.35	0.91	13.3%
11	1.43	0.97	14.3%	26	1.38	0.93	13.6%
12	1.39	0.95	14.0%	27	1.40	0.93	13.6%
13	1.39	0.95	14.0%	28	1.33	0.89	13.1%
14	1.36	0.93	13.7%	29	1.42	0.95	13.9%
15	1.37	0.93	13.9%	30	1.44	0.94	13.8%

As can be seen in Table 4.30, the forecasting performance of the target station is usually improving with an increasing number of neighbouring sites involved in the forecasting model training. Different numbers of neighbouring sites have different levels of impact on the forecasting performance of the target site. Figure 4.9 shows the graph of RMSE values of the target station against the number of neighbouring sites used.

As can be seen in Figure 4.9, the input features from the neighbouring sites do have a positive impact on the forecasting performance of the target site. However, when the number of neighbouring sites involved in the training becomes too big, the forecasting performance of the target station degrades a bit due to the model overfitting caused by a large number of input features from the neighbouring sites. It also requires a longer time to train the forecasting model with many input features from neighbouring sites. One way to solve this issue is to use an ensemble model to break the single forecasting model used in this section into multiple smaller models to avoid model overfitting.

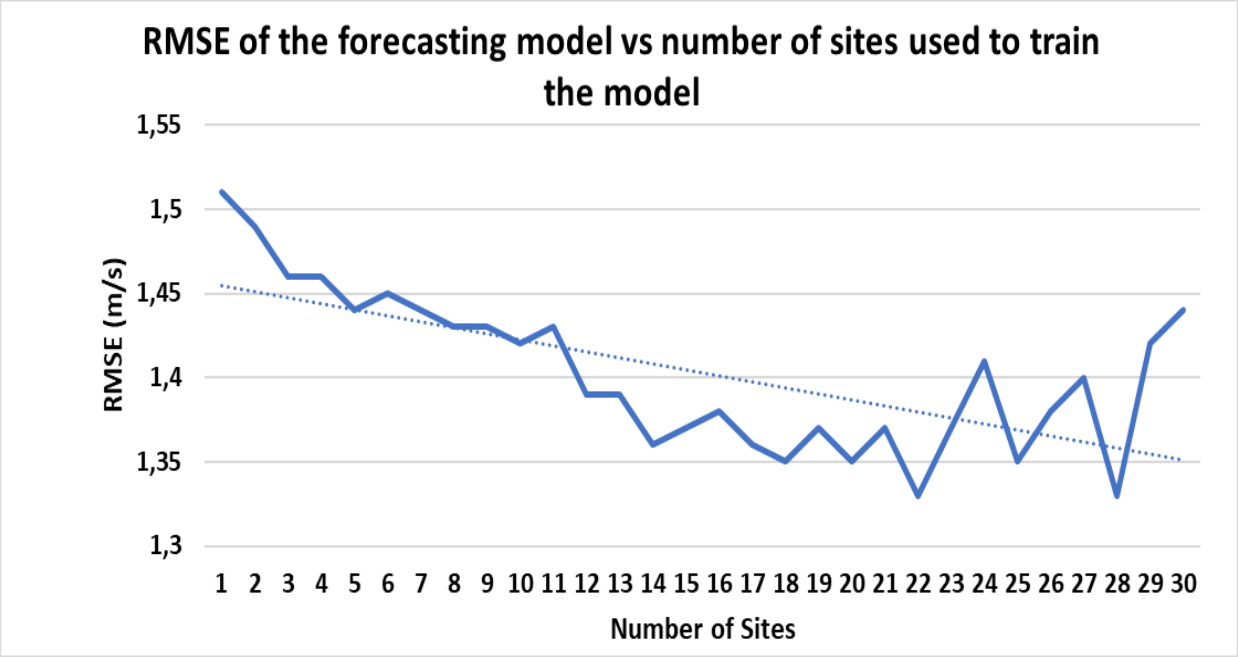


Figure 4.9: RMSE values of the target site against the number of neighbouring sites used in the forecasting model training (NREL)

Based on the simulation results of both case studies, it can be concluded that the forecasting model of the target sites perform better when input features from the relatively high correlated neighbouring sites are used. However, using too many input features from neighbouring sites will overfit the forecasting model. It requires a much longer time to train the forecasting model with a large number of input features. Ensemble model shown in the next chapter can be used to solve these two problems.

4.6 Discussions and summary

The input feature analysis suggested that wind speed and time indicator have significant impacts on forecasting performance. In contrast, barometric pressure and temperature have a small to no impact on forecasting performance. In general, all the available input features should be used to train the forecasting model if there are no computational constraints and training sample size limitation. The sample size and sample resolution simulation results suggested that the forecasting model performed slightly better with a higher training data resolution and a bigger training sample size. However, it requires a longer time to train the forecasting model with a higher sample resolution and a bigger training sample size. The trade-off between training time and forecasting performance need to be considered. Although the impact of the number of hidden layers and hidden neurons on short-term wind speed forecasting is small, the forecasting model has a slightly better performance with two hidden layers and ten hidden neurons in each hidden layer. Acceleration constants, inertia weights, and particle population size of PSO have different levels of impact on the model performance. Acceleration constants and particle population size have a more significant impact on PSO than inertia weights. The ANN(LM) performs better than the ANN (PSO) for all three data sources. There are many different parameter combinations can be selected to train both ANN and PSO. A model-parameter-based ensemble model was used to avoid selecting a weak predictor.

The simulation results of the model-parameter-based ensemble are shown in the next chapter. Usually, input features have a more significant effect on forecasting performance than the model parameters. Input features from neighbouring sites can also provide surrounding meteorological information to the forecasting model of a target site. The simulation results of the ANN-based spatial correlation model

indicated that using a suitable number of input features from neighbouring sites can enhance the forecasting performance of a target site model. However, using too many input features from neighbouring sites will overfit the ANN-based spatial correlation model. An input-feature-based ensemble model was used in this research to avoid overfitting issue faced by the ANN-based spatial correlation model. The simulation results of the input-feature-based ensemble are shown in the next chapter.

5. Simulation results and discussions of the ensemble approach

The purpose of this chapter is to show and discuss the simulation results of the ensemble model approach. The chapter starts with Section 5.1, showing the short-term wind speed and wind power forecasting performance of the model-parameter-based ensemble model. In Section 5.2, the simulation results of the input-feature-ensemble model are shown and compared to the ANN-based spatial correlation model. Section 5.3 shows the short-term wind power forecasting results of different combinations of forecasting techniques and model configurations. Five levels of forecasting configurations were used for both NREL and WASA datasets, and three levels of forecasting configurations were used for the SGWF dataset.

5.1 Model-parameter-based homogeneous ensemble

5.1.1 ANN(LM)

In a homogeneous ensemble, a single forecasting model with different parameters is used. Table 5.1 shows the simulation results of 2h-ahead wind speed forecasting of the feedforward ANN with different parameters. The second column of the table shows the forecasting model configuration. As an example, 5 HN indicates that the feedforward ANN contains a single hidden layer with five hidden neurons within the hidden layer and 10, 10 HN indicates that the feedforward ANN contains two hidden layers and each layer contains ten hidden neurons.

Ensemble (ANN) represents that the ensemble model uses a feedforward ANN to assigns weights to the outputs of the best three models (Test 5.4 to Test 5.6 in bold) to get the final output. Ensemble (Mean) takes the average of the best three forecasting models' output to obtain the final output.

Table 5.1: Simulation results of 2h-ahead wind speed forecasting of the feedforward ANN for model-parameter-based ensemble analysis (SGWF)

Test no.	**Model Config.**	**SGWF**			**NREL: ID61118**			**WASA: WM03**		
		RMSE (m/s)	MAE (m/s)	MAPE	RMSE (m/s)	MAE (m/s)	MAPE	RMSE (m/s)	MAE (m/s)	MAPE
5.1	5 HN	1.62	1.21	21.4%	1.57	1.08	15.7%	1.91	1.49	20.9%
5.2	10 HN	1.61	1.20	21.2%	1.53	1.05	15.4%	1.90	1.48	20.4%
5.3	15 HN	1.61	1.19	21.1%	1.52	1.05	15.4%	1.89	1.47	20.4%
5.4	20 HN	**1.60**	**1.19**	**21.1%**	**1.52**	**1.04**	**15.3%**	**1.87**	**1.45**	**20.3%**
5.5	25 HN	**1.60**	**1.19**	**20.8%**	**1.51**	**1.03**	**15.1%**	**1.86**	**1.45**	**20.2%**
5.6	10, 10 HN	**1.60**	**1.19**	**20.8%**	**1.51**	**1.03**	**15.1%**	**1.86**	**1.45**	**20.2%**
5.7	Ensemble (ANN)	1.59	1.19	20.9%	1.52	1.03	15.3%	1.84	1.43	20.2%
5.8	Ensemble (Mean)	1.59	1.18	20.7%	1.50	1.03	15.1%	1.84	1.43	20.1%

The ensemble models have the same or slightly better performance than the best individual model (Test 5.6) for all three data sources. Simulation results of Test 5.1 to Test 5.6 are very similar. The number of hidden layers and hidden neurons has little impacts on the forecasting performance of ANNs.

Based on the simulation results shown above, it can be concluded that the model-parameter-based ensemble models of ANNs have a small impact on improving the performance of the short-term wind speed forecasting. The output of each model is not diverse enough for the ensemble model to show its advantages. Although the forecasting improvement of the model-parameter-based ensemble is negligible, ensemble models (Test 5.8) can mitigate the chance of picking a weak predictor.

5.1.2 ANN(PSO)

The same procedure used in Section 5.1.1 is used in this section to determine whether the model-parameter-based ensemble model can enhance the forecasting performance of the ANN(PSO). Table 5.2 shows the simulation results of 2h-ahead wind speed forecasting of the ANN(PSO) for acceleration constant ensemble analysis. ANN(PSO) with six different sets of acceleration constants were used in this dissertation. The best three performers of each dataset (results in bold) were used to form model-parameter-based ensemble models.

As can be seen in Table 5.2, the ensemble models that use the feedforward ANN at the top level (ensemble (ANN)) perform better than the best individual model for all three data sources. Adding more individual model to the ensemble does not necessarily improve the forecasting performance of the ensemble model, only the one with both diverse and useful information can enhance the forecasting performance of ensemble models. Table 5.3 shows the simulation results of 2h-ahead wind speed forecasting of the ANN(PSO) with different maximum and minimum inertia weight combinations.

Table 5.2: Simulation results of 2h-ahead wind speed forecasting of the ANN(PSO) for model-parameter-based ensemble (acceleration constant) analysis

Test no.	**Acceleration constant**		**SGWF**			**NREL: ID61118**			**WASA: WM03**		
	c1	c2	RMSE (m/s)	MAE (m/s)	MAPE	RMSE (m/s)	MAE (m/s)	MAPE	RMSE (m/s)	MAE (m/s)	MAPE
5.9	0.75	0.75	**1.68**	**1.25**	**21.7%**	**1.70**	**1.21**	**17.9%**	2.06	1.62	22.9%
5.10	1.25	1.25	1.68	1.26	21.8%	1.79	1.30	19.3%	**2.03**	**1.60**	**22.4%**
5.11	1.50	1.50	1.71	1.28	21.9%	1.80	1.32	19.5%	**2.04**	**1.60**	**22.5%**
5.12	1.75	1.75	**1.68**	**1.25**	**21.5%**	**1.67**	**1.15**	**16.9%**	**2.04**	**1.60**	**22.8%**
5.13	2.00	2.00	**1.66**	**1.24**	**21.5%**	**1.70**	**1.26**	**18.4%**	2.05	1.62	22.7%
5.14	2.25	2.25	1.70	1.27	21.9%	1.76	1.25	19.0%	2.08	1.64	23.7%
5.15	Ensemble (ANN)		1.64	1.23	21.8%	1.65	1.15	16.8	1.99	1.56	22.3%
5.16	Ensemble (Mean)		1.66	1.24	21.3%	1.67	1.17	17.0	2.02	1.58	22.6%

Table 5.3: Simulation results of 2h-ahead wind speed forecasting of the ANN(PSO) for the model-parameter-based ensemble (inertia weight) analysis

Test no.	Inertia weight		SGWF			NREL: ID61118			WASA: WM03		
	ω_{max}	ω_{min}	RMSE (m/s)	MAE (m/s)	MAPE	RMSE (m/s)	MAE (m/s)	MAPE	RMSE (m/s)	MAE (m/s)	MAPE
5.17	0.50	0.00	1.70	1.27	21.9%	**1.67**	**1.18**	**17.4%**	**2.05**	**1.62**	**22.5%**
5.18	0.50	0.25	1.69	1.26	21.7%	**1.68**	**1.20**	**17.6%**	**2.03**	**1.60**	**22.7%**
5.19	0.80	0.40	**1.68**	**1.26**	**21.6%**	1.77	1.29	19.1%	2.07	1.64	23.0%
5.20	0.90	0.30	1.71	1.28	22.0%	**1.70**	**1.23**	**18.2%**	**2.04**	**1.61**	**22.5%**
5.21	0.90	0.40	**1.66**	**1.24**	**21.5%**	1.70	1.26	18.4%	2.05	1.62	22.7%
5.22	0.90	0.60	**1.68**	**1.26**	**21.6%**	1.71	1.26	18.2%	2.08	1.65	22.6%
5.23	Ensemble (ANN)		1.63	1.22	21.6%	1.66	1.17	17.1%	2.01	1.58	22.6%
5.24	Ensemble (Mean)		1.70	1.27	21.6%	1.66	1.17	17.1%	2.03	1.61	22.7%

As can be seen in Table 5.3, the ensemble models use an ANN model at the output node to nonlinearly assign the weight to the results of the member models perform better than the best individual model for all three data sources. However, the ensemble (Mean) which assign weight equally (Test 5.24) to the results of the member model does not have better performance than the best individual model. The ensemble model with the feedforward ANN at the top-level able to find advantages of the output from each model and provide a better final result. Table 5.4 shows the simulation results of 2h-ahead wind speed forecasting of the ANN(PSO) model for the analysis of the ensemble model of the different numbers of particle population.

Once again, the ensemble models with the feedforward top-level model perform better than the best individual model for all three data sources. Although the forecasting performance improvement of the Ensemble (ANN) (Test 5.31) from the best individual model (i.e., Test 5.30 for SGWF, Test 5.29 for WASA, Test 5.28 for NREL) is minimal, the ensemble models always have the same or slightly better results than the best individual model. The improvement of RMSE values of the ensemble (ANN) model with respect to the best individual model of three data sources are 1.8%, 1.2%, 1.8% for SGWF, WASA, NREL, respectively.

Based on the results shown in this section, it can be concluded that the ensemble model with a nonlinear model at output node can slightly enhance the short-term wind speed forecasting performance.

Table 5.4: Simulation results of 2h-ahead wind speed forecasting of the ANN(PSO) for the analysis of ensemble model (particle population)

Test no.	Particle population	SGWF			NREL: ID61118			WASA: WM03		
		RMSE (m/s)	MAE (m/s)	MAPE	RMSE (m/s)	MAE (m/s)	MAPE	RMSE (m/s)	MAE (m/s)	MAPE

5.25	5	1.69	1.27	21.89%	1.74	1.27	19.02%	2.07	1.64	23.38%
5.26	20	**1.66**	**1.24**	**21.49%**	1.70	1.26	18.36%	2.05	1.62	22.71%
5.27	60	**1.66**	**1.24**	**21.51%**	**1.69**	**1.18**	17.57%	2.09	1.65	22.93%
5.28	70	1.69	1.27	21.73%	**1.67**	**1.18**	17.53%	**2.04**	**1.60**	**22.08%**
5.29	80	1.68	1.26	21.85%	**1.66**	**1.18**	**17.33%**	**2.05**	**1.61**	**22.31%**
5.30	90	**1.66**	**1.23**	**21.37%**	1.69	1.18	17.53%	**2.05**	**1.61**	**22.39%**
5.31	Ensemble (ANN)	1.63	1.22	21.61%	1.64	1.14	16.72%	2.01	1.58	22.56%
5.32	Ensemble (Mean)	1.66	1.23	21.35%	1.65	1.15	16.81%	2.03	1.60	22.29%

5.2 Input-feature-based homogeneous ensemble

The model configuration shown in Table 5.5 below was used to analyse the performance of the input-feature-based ensemble of a single site. Figure 3.15 shows the structure of the input-feature-based ensemble.

Table 5.5: Summary of the configurations utilised to analyse the performance of the input-feature-based ensemble model

Forecasting model	Feedforward neural network
Number of hidden layers	2
Number of hidden neurons	10 in each layer
Training algorithm	LM
Type of forecasting	2h-ahead wind speed forecasting
Input features (SGWF)	WS, WD, TI
Input features (WASA: WM03)	WS, WD, T, TG, P, RH, TI
Input features (NREL: ID61118)	WS, WD, T, P, AD, TI
Data resolution and sample size	Resolution: 10min, sample size: 300, 000

Table 5.6 shows the simulation results of 2h-ahead wind speed forecasting of the feedforward ANN for input-feature-based ensemble analysis. The ensemble (ANN) model uses the outputs of all the member model to produce a final output. Results in Table 5.6 show that the forecasting performance of the input-feature-based ensemble models of the single site does not improve from the best individual model. The number of input features in each site is small (i.e., three input features for SGWF, six input features for NREL, seven input features for WASA), a single model is capable of modelling a small number of input features. The input-feature-based ensemble model will be useful if more input features from neighbouring sites are used as input features to train the forecasting model of the target stations.

Table 5.6: Simulation results of 2h-ahead wind speed forecasting of the feedforward ANN for the analysis of the input-feature-based ensemble of a single site

Test		**SGWF**	**NREL: ID61118**	**WASA: WM03**

no.	No. of input features used	RMSE (m/s)	MAE (m/s)	MAPE	RMSE (m/s)	MAE (m/s)	MAPE	RMSE (m/s)	MAE (m/s)	MAPE
5.33	1	1.64	1.22	21.2%	1.75	1.26	18.5%	2.22	1.76	25.0%
5.34	2	1.63	1.22	21.2%	1.68	1.19	17.4%	2.17	1.70	24.1%
5.35	3	1.61	1.19	20.8%	1.68	1.18	17.4%	2.07	1.62	22.7%
5.36	4	-	-	-	1.66	1.18	17.4%	1.90	1.48	21.0%
5.37	5	-	-	-	1.62	1.16	17.1%	1.88	1.46	20.5%
5.38	6	-	-	-	1.51	1.03	15.2%	1.86	1.45	20.3%
5.39	7	-	-	-	-	-	-	1.86	1.44	20.2%
5.40	Ensemble (ANN)	1.59	1.19	21.1%	1.51	1.04	15.2%	1.84	1.43	20.4%

The next section shows the simulation results for the analysis of the input-feature-based ensemble model with input features from multiple stations. Data from NREL and WASA were used to obtain the following simulations because both data sources contain multiple sites, whereas SGWF only has one site.

Table 5.7: Simulation results of 2h-ahead wind speed forecasting for input-features ensemble model with input features from multiple stations (target station WASA: WM03, neighbouring stations WASA: WM01, WM02, WM05-WM09)

Test no.	No. of sites	ANN-based spatial correlation model			Input-feature-based ensemble model		
		RMSE (m/s)	MAE (m/s)	MAPE	RMSE (m/s)	MAE (m/s)	MAPE
5.41	1	1.82	1.41	19.87%	1.82	1.41	20.11%
5.42	2	1.82	1.40	19.65%	1.78	1.37	19.54%
5.43	3	1.76	1.35	19.00%	1.73	1.33	18.85%
5.44	4	1.73	1.33	18.73%	1.71	1.31	18.55%
5.45	5	**1.71**	**1.32**	**18.59%**	1.69	1.30	18.37%
5.46	6	1.73	1.33	18.85%	1.68	1.29	18.26%
5.47	7	1.72	1.32	18.65%	1.67	1.29	18.23%
5.48	8	1.74	1.34	18.91%	**1.67**	**1.28**	**18.20%**

As can be seen in Table 5.7, the single ANN-based spatial correlation model with input features from five neighbouring sites has RMSE value improved by 6% compared with the one using input features from a single site (see Test 5.41 and Test 5.45). The input-feature-based ensemble model with input feature from eight neighbouring sites has RMSE value improved by 8.2% compared with the one using the input features from a single site (see Test 5.41 and Test 5.48). However, as can be seen from Test 5.45 to Test 5.48 that, the ANN-based spatial correlation model stops improving when the input features from more distant sites (WM06-WM09) are used. The input features from distant neighbouring sites make the number of input features of the forecasting model large, which might overfit the forecasting model. The

sixth to eighth columns of the table show the simulation results of the input-feature-based ensemble model (see Figure 3.15 for the structure of the input-ensemble model). The input-feature-based ensemble can mitigate the model overfitting issue by using multiple forecasting models to training different combinations of input features in parallel. Then use another basic feedforward ANN model at the top level to learn the relationship between the outputs of member models and then produce a final output. The input-feature-based ensemble model can take useful information from the better forecasting models. Figure 5.1 shows the RMSE values of the ANN-based spatial correlation model and the input-feature-based ensemble model with input features from different numbers of neighbouring sites (WASA).

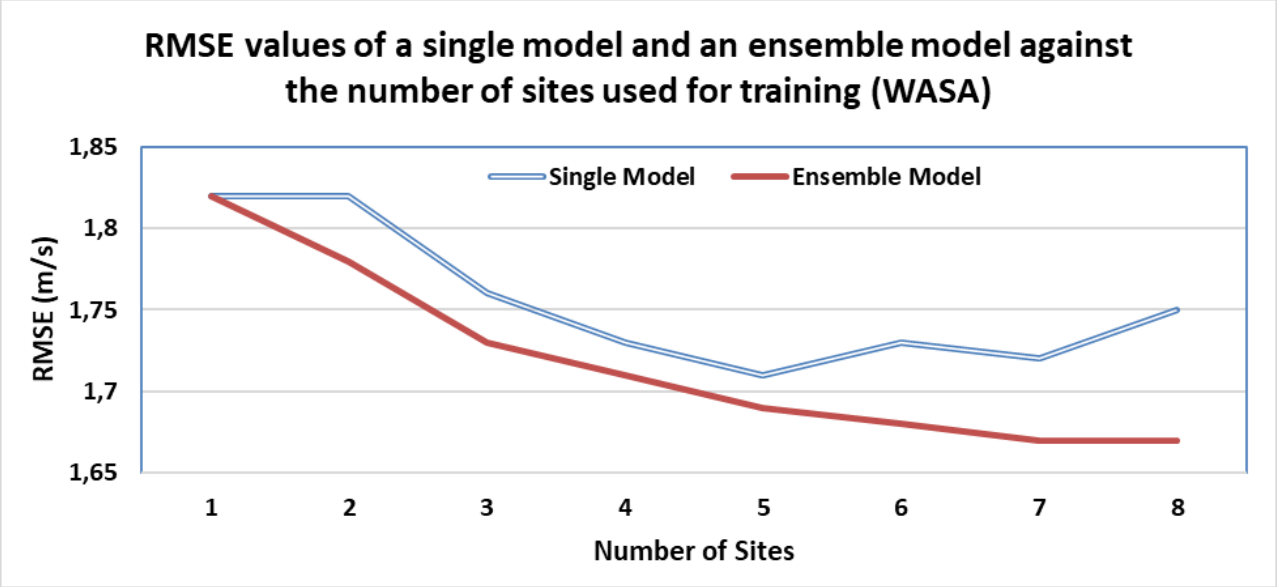


Figure 5.1: RMSE values of the single ANN-based spatial correlation model and the input-feature-based ensemble model with input features from different numbers of neighbouring sites (WASA)

As can be seen in Figure 5.1, the ANN-based spatial correlation model has lower RMSE values when input features from five neighbouring sites are used. The performance of the ANN-based spatial correlation model stops improving and becomes worse when the input features from more than six sites are used. This result is due to the reason that the input features from the distant sites (i.e. WM06 to WM09) do not provide enough useful information to enhance the performance of the forecasting model. The input-feature-based ensemble model performs better than the single ANN-based spatial correlation model, especially when the number of neighbouring sites becomes large. As an example, the input-feature-based ensemble has 4% of RMSE improvement compared with the single ANN-based spatial correlation model when input features from all eight sites are used.

The same experiment is carried out for the NREL dataset. Only the simulation found results of even number of sites are shown in Table 5.8 due to the limited space. The full table can be in Appendix B. As can be seen in Table 5.8, the input-feature-based ensemble model performs much better than the single ANN-based spatial correlation model when the number of sites used for training the forecasting model becomes large. The ensemble model can capture features and advantages from more neighbouring sites to enhance the forecasting performance of the target site. As can be seen from the third column and the fifth column of Table 5.8, the ANN-based spatial correlation model is also able to take the useful information from the input features of the neighbouring sites to enhance the performance of the target site until the number of neighbouring sites used for training the forecasting model becomes large. On the other hand, the last three columns of Table 5.8 show that input-feature-based ensemble model can overcome the overfitting issue of ANNs caused by a large number of input features by using multiple small ANNs to train smaller input feature batches in parallel.

Table 5.8: Simulation results of the ANN-based spatial correlation model and the input-feature-based ensemble model with input features from different numbers of neighbouring sites for 2h-ahead wind speed forecasting (NREL)

Test no.	No. of sites	ANN-based spatial correlation model			Input-feature-based ensemble model		
		RMSE (m/s)	MAE (m/s)	MAPE	RMSE (m/s)	MAE (m/s)	MAPE
5.49	2	1.49	1.02	15.3%	1.49	1.01	14.8%
5.50	4	1.46	1.00	14.7%	1.46	1.00	14.5%
5.51	6	1.45	0.98	14.5%	1.42	0.96	14.1%
5.52	8	1.43	0.97	14.3%	1.41	0.95	14.0%
5.53	10	1.42	0.96	14.1%	1.39	0.93	13.8%
5.54	12	1.39	0.95	14.0%	1.37	0.93	13.6%
5.55	14	1.36	0.93	13.7%	1.34	0.91	13.4%
5.56	16	1.38	0.94	13.8%	1.31	0.89	13.1%
5.57	18	1.35	0.91	13.4%	1.29	0.88	12.9%
5.58	20	1.35	0.92	13.5%	1.27	0.86	12.6%
5.59	22	1.33	0.90	13.2%	1.25	0.85	12.4%
5.60	24	1.41	0.93	13.5%	1.24	0.85	12.4%
5.61	26	1.38	0.93	13.6%	1.24	0.85	12.4%
5.62	28	**1.33**	**0.89**	**13.1%**	1.23	0.84	12.3%
5.63	30	1.44	0.94	13.8%	**1.23**	**0.83**	**12.2%**

Figure 5.2 shows the RMSE values of the ANN-based spatial correlation model and the input-feature-based ensemble model against the number of sites used for training the forecasting models for NREL.

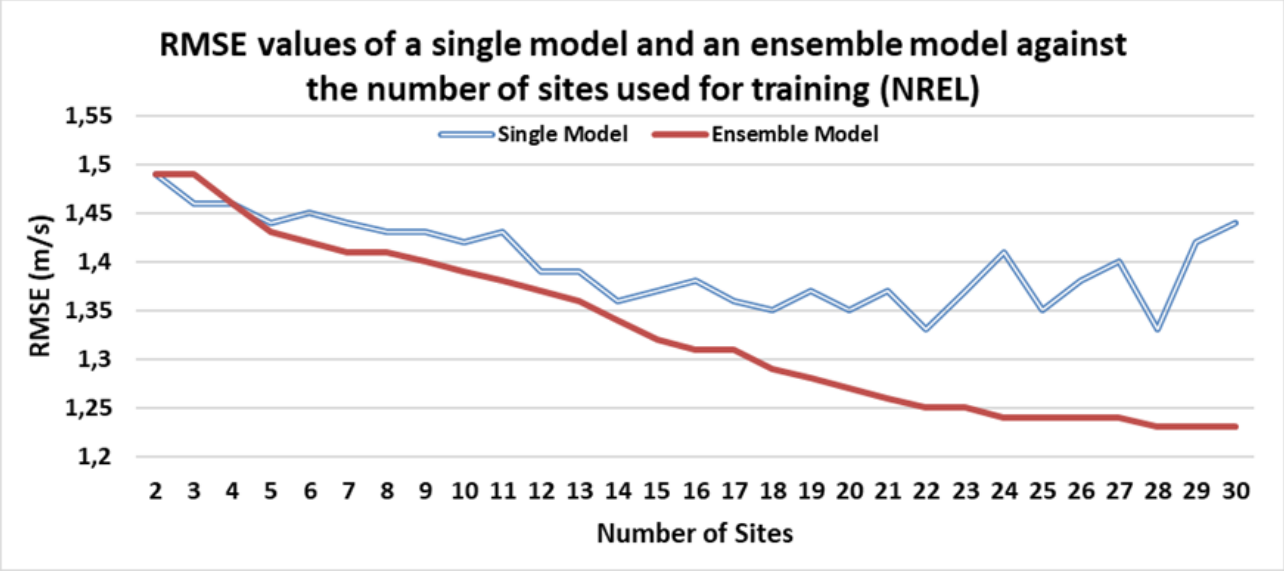


Figure 5.2: RMSE values of a single model and an ensemble model against the number of sites used for training the forecasting models (NREL)

As can be seen in Figure 5.2, the advantage of using the input-feature-based ensemble is not very obvious when a small number of neighbouring sites are involved in the training of the forecasting model. The input-feature-based ensemble model performs much better than the ANN-based spatial correlation model when the number of neighbouring sites used for training becomes larger than 15 sites. The input-

feature-based ensemble is capable of training multiple small models in parallel to reduce training time and to avoid overfitting the training model.

The simulation results of two types of ensemble models show that the input-feature-based ensemble model has a better performance than the model-parameter-based ensemble model. Input features are the most important factors that affect the forecasting performance of the ANN-based model. Forecasting models with different input features produce relatively diverse outputs. Therefore, the input-feature-based ensemble model can take advantages of the diversity of the outputs of its member models. The nonlinear top-level model is recommended to assign weights nonlinearly to the outputs of each member models of ensemble models.

5.3 Wind power forecasting

The wind speeds forecasted in the previous sections are then converted to wind power to compare with the actual wind power. In this section, the 2h-ahead wind power forecasting performance of different model configurations of all three data sources is presented.

5.3.1 SGWF

As shown in the simulation results of the previous chapters, different techniques have different impacts on the forecasting performance of the ANN-based models. The techniques that have a significant impact on forecasting performance are combined to form different forecasting configurations. Table 5.9 shows the summary of different forecasting configurations for the SGWF dataset.

Table 5.9: Summary of different short-term wind power forecasting configurations for the SGWF dataset

	Forecasting model	Input feature	No. of sites
Level 1	A feedforward ANN with a single hidden layer (10 hidden neurons)	A single input feature (WP)	1
Level 2	A feedforward ANN with a single hidden layer (10 hidden neurons)	All the available input features (WP, WS, WD, TI)	1
Level 3	A feedforward ANN with double hidden layers (10 hidden neurons in each hidden layer)	All the available input features (WP, WS, WD, TI)	1

As shown in Table 5.10, only three levels of forecasting configurations were used to obtain forecasted wind power for the SGWF dataset, because the SGWF dataset only contains data from one wind farm and the number of input features is very small. Therefore, the spatial correlation and input-feature-based ensemble techniques cannot be applied to the SGWF dataset. Configuration level 1 uses a basic feedforward ANN model with a single input feature to obtain 2h-ahead wind power forecasting results. Configuration level 2 uses a basic feedforward ANN model with all the available input features (WP, WS, WD, TI) to predict wind power. Configuration level 3 uses a feedforward ANN with double hidden layers and all the available input features to forecast wind power two hours in advance. Table 5.11 shows the simulation results of 2h-ahead wind power forecasting of different forecasting configurations using the SGWF dataset.

Table 5.10: Simulation results of 2h-ahead wind power forecasting of different forecasting approaches (SGWF)

Forecasting model	**RMSE (kW)**	**MAE (kW)**	**MAPE**	**NRMSE**	**NMAE**	% NRMSE imp. with respect to the persistence model	% NRMSE imp. with respect to level 1
Persistence	75.4	47.8	42.00%	10.32%	6.55%	0.00%	-3.39%
Level 1	72.9	49.2	41.52%	9.97%	6.73%	3.39%	0.00%
Level 2	71.7	48.4	41.19%	9.82%	6.62%	4.84%	1.45%
Level 3	**71.5**	**47.9**	**40.90%**	**9.79%**	**6.56%**	**5.13%**	**1.74%**

As can be seen in Table 5.10, the forecasting configuration level 3 has a slightly better performance than other forecasting configurations. The difference in results between configuration level 1 and configuration level 3 is minimal (i.e., 1.74% NRMSE improvement with respect to NRMSE of configuration level 1). This is due to the small configuration change from level 1 to level 3. The number of input features is small. Therefore, there is no significant performance difference between level 1 and level 2. The additional hidden layer used in configuration level 3 does not have a significant impact on forecasting performance because the number of input features is small. Therefore, a simpler ANN model (level 1) can also produce similar results compared with the model with configurations level 3. The bar chart of RMSE values of 2h-ahead wind power forecasting of different forecasting configurations is shown in Figure 5.3 to provide better visualization.

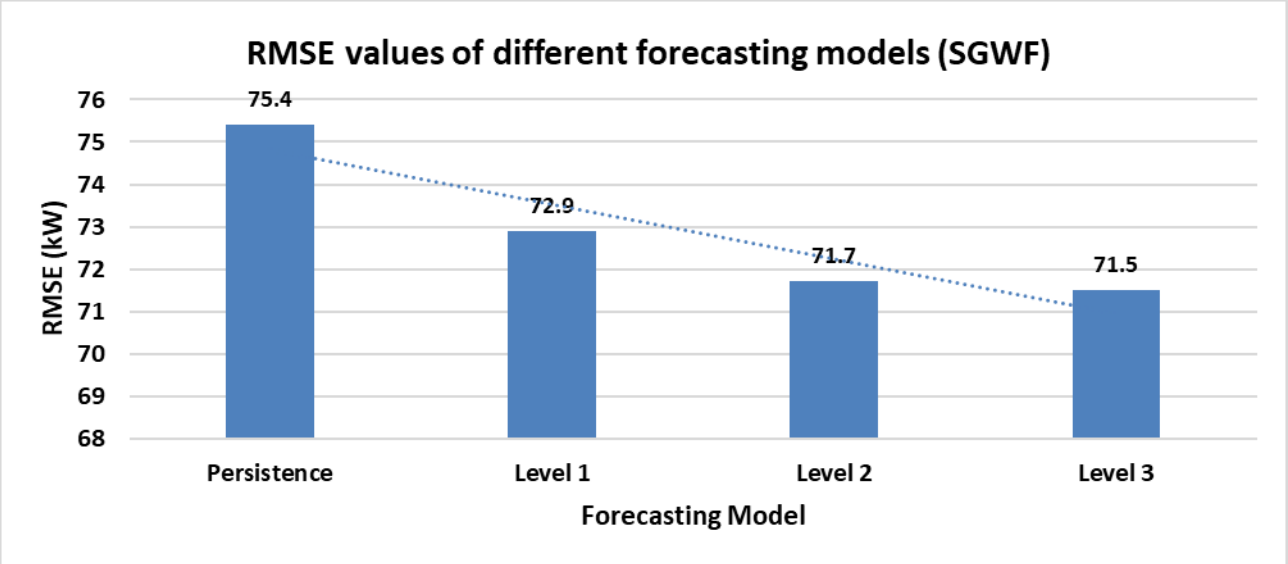


Figure 5.3: RMSE values of 2h-ahead wind power forecasting of different forecasting approaches (SGWF)

Figure 5.3 shows that the forecasting model with a higher configuration level performs slightly better than the lower one. Figure 5.4 shows the graphs of the actual wind power and the forecasted wind power of two different forecasting configurations (level 1 and level 3) of the SGWF dataset.

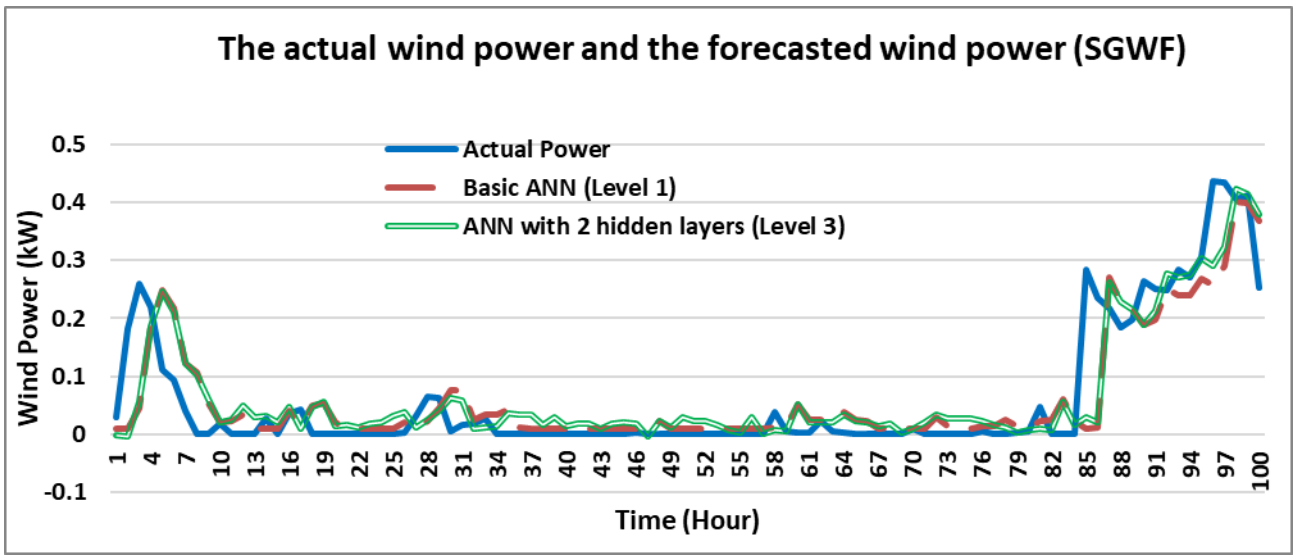


Figure 5.4: The actual wind power and the forecasted wind power of two different forecasting configurations (level 1 and level 3) of the SGWF dataset

As can be seen in Figure 5.4, forecasted 2h-ahead wind power can track the actual wind power closely for most of the samples. However, the two hours phase errors between the actual wind power and the forecasted wind power of both level 1 and level 3 configurations are very obvious. The time gap between the input features and the output target is two hours for the 2h-ahead wind power forecasting. The model can track the actual power with a phase lag of two hours because there are not enough meteorological variables such as, temperature, relative humidity, pressure, and et cetera. to provide more information for the forecasting model to narrow the gap. Figure 5.5 shows the actual wind power and the forecasted wind power with a 95% prediction interval of the model with level 3 configuration.

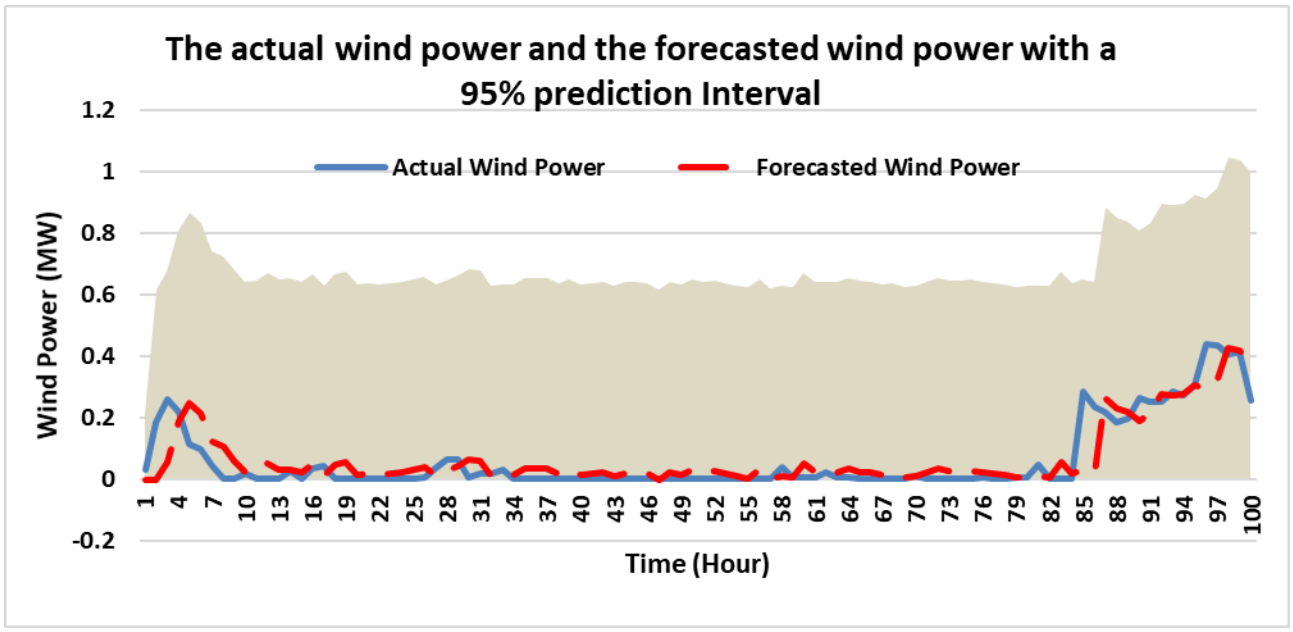


Figure 5.5: The actual wind power (solid blue line) and the forecasted wind power (red dash line) with a 95% prediction interval of a level 3 configuration (SGWF).

The grey area covers the actual wind power 95% of the time. The big prediction interval is due to the high standard deviation of the forecasted wind power. The steady wind generation allows the prediction interval to cover most of the actual wind power. A lower percentage of the forecasting confidence level (i.e. 50%) will largely reduce the prediction interval and uncertainty but at the time reduce the confident level of the forecasting results.

5.3.2 WASA

Table 5.11 shows five different forecasting configurations for the WASA dataset. The WASA dataset contains input features from more than one sites. Therefore, the spatial correlation (Level 4) and the input-feature-based ensemble (Level 5) can be used to enhance the 2h-ahead wind power forecasting performance of the ANN-based model.

Table 5.11: Summary of different short-term wind power forecasting configurations for the WASA dataset

	Forecasting model	Input feature	No. of sites
Level 1	A feedforward ANN with a single hidden layer (10 hidden neurons)	A single input features (WS)	1
Level 2	A feedforward ANN with a single hidden layer (10 hidden neurons)	All the available input features	1
Level 3	A feedforward ANN with double hidden layers (10 hidden neurons in each hidden layer)	All the available input features	1
Level 4	A feedforward ANN with double hidden layers (10 hidden neurons in each hidden layer)	All the available input features from highly correlated neighbouring sites	8
Level 5	Input-feature-based ANN ensemble	All the available input features from highly correlated neighbouring sites	8

Forecasting configuration level 4 uses input features from eight weather stations and the configuration level 5 uses an input-feature-based ensemble model on top of configuration level 4 to improve the short-term wind power performance of the ANNs-based model further. Table 5.12 shows the simulation results of 2h-ahead wind power forecasting of all five configuration levels.

Table 5.12: Simulation results of 2h-ahead wind power forecasting of different forecasting approaches (WASA)

Forecasting model	**RMSE (MW)**	**MAE (MW)**	**MAPE**	**NRMSE**	**NMAE**	**% NRMSE imp. with respect to the persistence model**	**% NRMSE imp. with respect to the configuration level 1**
Persistence	0.52	0.35	46.07%	26.02%	17.72%	0.00%	-8.15%
Level 1	0.48	0.37	48.40%	24.06%	18.70%	7.53%	0.00%
Level 2	0.40	0.30	38.07%	20.01%	14.89%	23.10%	16.83%
Level 3	0.39	0.29	36.53%	19.52%	14.30%	24.98%	18.87%
Level 4	0.38	0.27	34.85%	18.87%	13.51%	27.48%	21.57%
Level 5	**0.37**	**0.27**	**34.74%**	**18.50%**	**13.50%**	**28.90%**	**23.11%**

As can be seen in Table 5.12, the forecasting model with the configuration level 5 has the best short-term wind power forecasting performance. The RMSE value of the configuration level 5 is improved by 23.11% and 28.9% compared to the model with configuration level 1 and the persistence model, respectively. Figure 5.7 shows the bar chart of RMSE values of 2h-ahead wind power forecasting of different forecasting configurations.

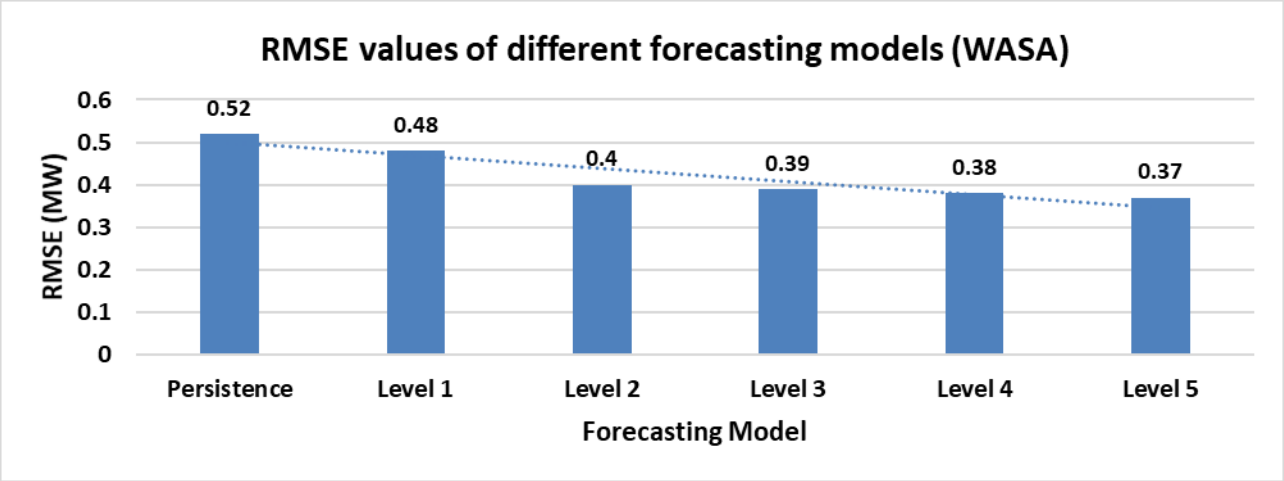


Figure 5.6: RMSE values of 2h-ahead wind power forecasting of different forecasting approaches (WASA)

The model with the configuration level 5 performs better than the one with a lower one. The most significant improvement (i.e., level 2 has a 16.7% of RMSE improvement with respect to level 1) occurs between configuration level 1 and level 2. The results indicate that using more input features can vastly enhance the forecasting performance of the forecasting model. The improvement between level 4 and level 5 is small, which is due to the small number of neighbouring sites involved in the training. An ANN-based spatial correlation (level 4) model can achieve a similar result as the input-feature-based ensemble model (level 5). Figure 5.7 shows the graphs of the simulated wind power and the forecasted wind power of the WASA dataset.

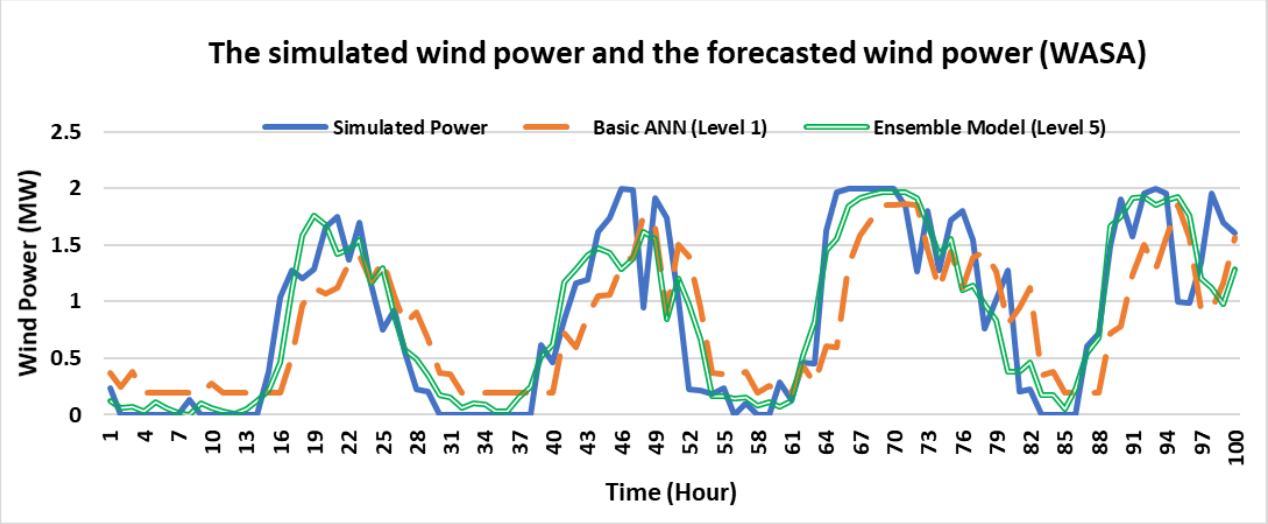


Figure 5.7: The simulated wind power (solid blue line) and the forecasted 2h-ahead wind power of the WASA datasets

As can be seen in Figure 5.7, the forecasted wind power of the ensemble model (configuration level 5) can track the simulated power closely for most of the samples. Although the forecasted wind power of the basic ANN (configuration level 1) can track the pattern of the simulated wind power, the two hours phase errors are apparent due to the small number of input features used by the basic ANN model (level 1). On the other hand, the phase errors between the simulated wind power and the forecasted wind power of the input-feature-based ensemble model are not obvious. Figure 5.8 shows the graph of the simulated wind power and the forecasted wind power with a 95% prediction interval of the input-feature-based ensemble model (level 5).

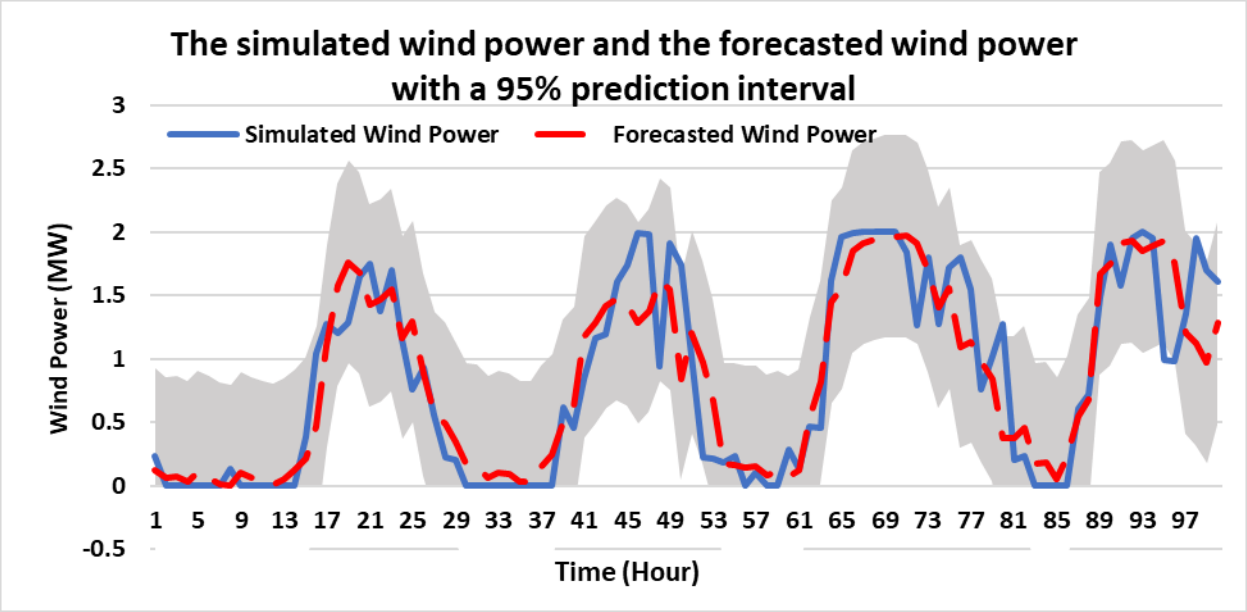


Figure 5.8: The simulated wind power (solid blue line) and the forecasted (red dash line) 2h-ahead wind power with a 95% prediction interval of the input-feature-based ensemble (WASA)

As can be seen in Figure 5.8, the grey area covers the simulated wind power with 95% probability.

5.3.3 NREL

Table 5.13 shows five different levels of forecasting configurations for the NREL dataset. The forecasting configurations for the NREL dataset and the WASA dataset are very similar because both datasets contain input features from more than one sites and each site contain similar meteorological variables. The only difference between the NREL dataset and the WASA dataset is the number of the sites used.

Table 5.13: Summary of different short-term wind power forecasting configurations for the NREL dataset

	Forecasting model	Input feature	No. of sites
Level 1	A feedforward ANN with a single hidden layer (10 hidden neurons)	A single input feature	1
Level 2	A feedforward ANN with a single hidden layer (10 hidden neurons)	All the available input features	1
Level 3	A feedforward ANN with double hidden layers (10 hidden neurons in each hidden layer)	All the available input features	1
Level 4	A feedforward ANN with double hidden layers (10 hidden neurons in each hidden layer)	All the available input features from highly correlated neighbouring sites	30
Level 5	Input-feature-based ANN ensemble	All the available input features from highly correlated neighbouring sites	30

Many sites are available from the NREL dataset. The input features of 30 sites were used in this research due to the time and computational constraints. The input features from more NREL sites will be considered in further studies. Table 5.14 shows the simulation results of 2h-ahead wind power forecasting of five different configuration levels.

Table 5.14: Simulation results of 2h-ahead wind power forecasting of different forecasting approaches (NREL)

Forecasting model	RMSE (MW)	MAE (MW)	MAPE	NRMSE	NMAE
Persistence	0.41	0.26	31.35%	20.50%	13.00%
Level 1	0.39	0.28	34.03%	19.40%	13.86%
Level 2	0.33	0.23	27.99%	16.48%	11.30%
Level 3	0.32	0.22	26.59%	16.06%	10.73%
Level 4	0.30	0.20	23.69%	15.01%	9.77%
Level 5	**0.27**	**0.18**	**21.58%**	**13.25%**	**8.72%**

As can be seen in Table 5.14, the model with the highest configuration level performs better than the one with a lower configuration level. The RMSE value of the model with a configuration level 5 improved by 30.8% and 34.1% compared with the model with a configuration level 1 and the persistence model, respectively. Figure 5.9 shows the bar chart of RMSE values of the model with different levels of configurations for the NREL dataset.

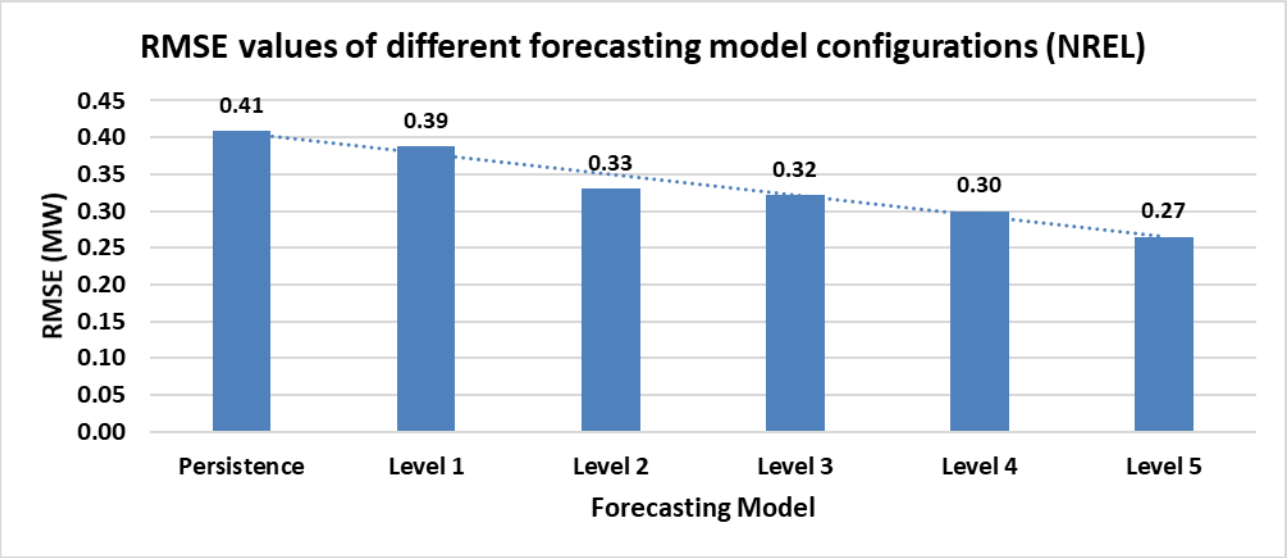


Figure 5.9: RMSE values of 2h-ahead wind power forecasting of different forecasting approaches (NREL)

As shown in Figure 5.9, the model with a higher configuration level performs better than the one with a lower configuration level. The total number of sites of the NREL dataset involved in this research is 30 It is expected that the forecasting performance of the model with the level 5 configuration can be improved further if more sites are included. Figure 5.10 shows the actual wind power and the forecasted wind power of the forecasting model with level 1 and level 5 configurations of the NREL dataset. The actual wind power is produced by NREL using the weather research and forecasting model.

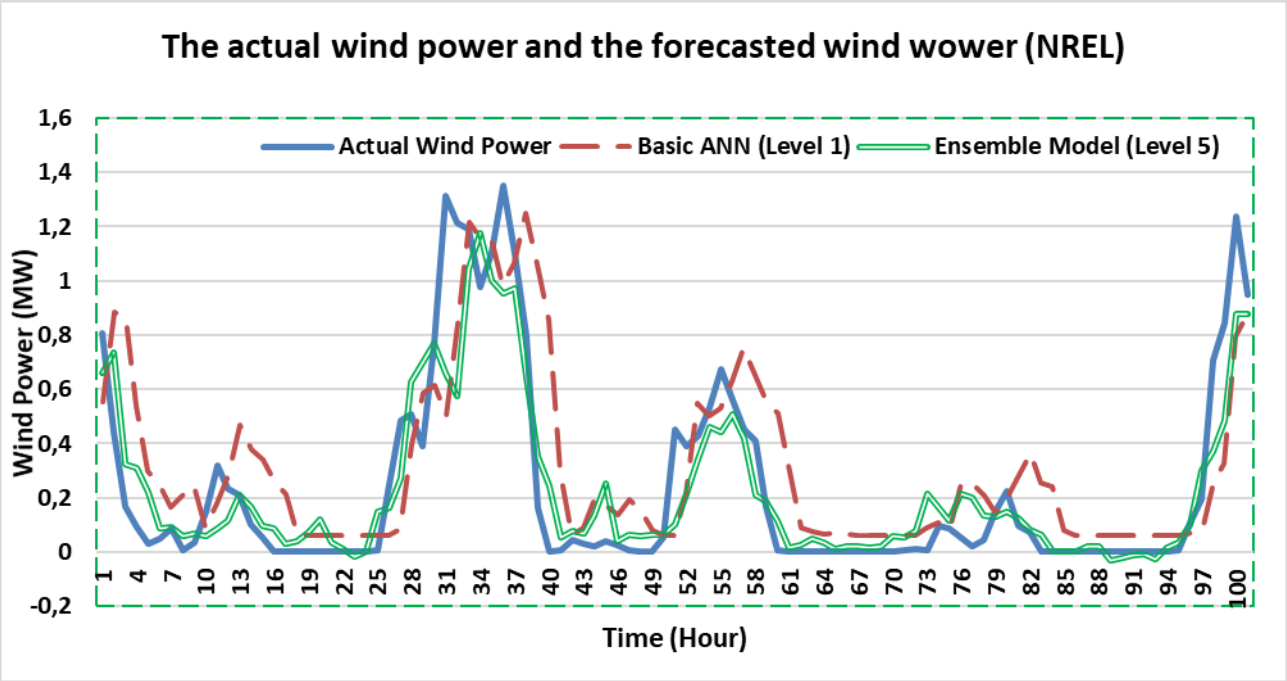


Figure 5.10: The actual wind power (solid blue line) and the forecasted wind power of forecasting model with level 1 and level 5 configurations of the NREL datasets

As can be seen in Figure 5.10, the gap between the forecasted wind power of the ensemble model (level 5) and the actual wind power is much smaller than the gap between the forecasted wind power of the basic ANN (Level 1) and the actual wind power. The ensemble model can take useful information from more neighbouring stations to enhance the forecasting performance of the target station. Figure5.11 show the actual wind power and the forecasted wind power with a 95% prediction interval.

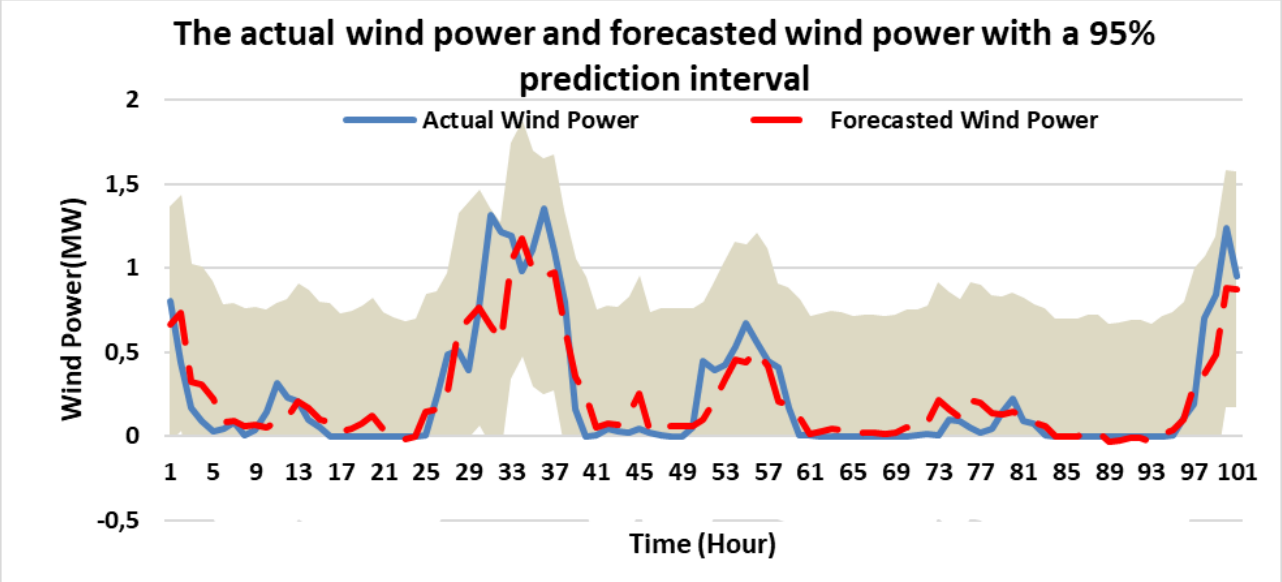


Figure 5.11: The actual wind power and the forecasted wind power with a 95% prediction interval (NREL)

As can be seen in Figure 5.11, the 95% prediction interval covers most points of the actual wind power. The prediction interval provides useful information to the short-term wind power forecasting users. Based on the prediction interval, the forecasting model users can plan ahead to encounter the uncertainty of the forecasting results.

5.4 Discussions and summary

The simulation results of the model-parameter-based ensemble model indicated that the model-parameter-based ensemble model could not significantly enhance the forecasting performance of the best individual model. The small enhancement of the model-parameter-based model is due to the small diversity between the outputs of its member models, which contribute a small amount of new information to the ensemble model. However, the model-parameter-based ensemble model still has a slightly better forecasting performance than the best member model. Using a large number of input features from neighbouring sites can provide meteorological information to the forecasting model of the target station. The simulation results of the input-feature-based ensemble model indicated that using multiple forecasting models to train a large number of input features in parallel can avoid overfitting issues. Wind farm operators should consider to obtain meteorological information around wind farms and use the surrounding meteorological information to train the input-feature-based ensemble model to enhance the forecasting performance. Wind power generation information from neighbouring wind farms can also contribute useful information to the forecasting model of a target station. Short-term wind power forecasting performance of the input-feature-based ensemble model was compared to the persistence model and four ANN-based models. It is shown that the input-feature-ensemble model has the best performance because it can use input features from many neighbouring stations and avoid the overfitting issue encountered by the ANN-based spatial correlation model. With greater wind power penetration in South Africa, wind power producers, utilities, market analysts, and energy traders need to use accurate short-term wind power and wind speed forecasting results to make better load increment/decrement decisions.

6. Conclusions

The effect of input features, model parameters, and spatial correlation and ensemble techniques on short-term wind power forecasting performance of the ANNs models have been investigated in this dissertation. The simulation results indicated that different types of input features and model parameters can influence the forecasting performance of the ANN-based models. Spatial correlation technique can be used to enhance the forecasting performance of the ANN-based models if highly correlated input features from neighbouring sites are available to be used. The input-feature-based ensemble can further improve the forecasting performance of the ANN-based spatial correlation model when the number of input features from neighbouring sites becomes large. Based on the simulations results, the following are the conclusions:

All the meteorological variables provided in three data sources have different levels of impact on the forecasting performance of the ANNs model. The simulation results indicated that wind speed has the most significant impact. In contrast, barometric pressure, air temperature, and air density have the least impact on short-term wind speed and wind power forecasting. The simulation results also indicated that time of day has a significant impact on the short-term wind forecasting performance.

The model with a higher data resolution performed slightly better than the one with a smaller data resolution. The results also showed that the forecasting performance of the model improved rapidly when the sample size is increased from 1000 samples to 10000 samples. There is no noticeable performance improvement when the sample size is over 30000. In general, the more input features are used, the larger training sample size is required to train the ANN-based model. The model with more input features should use a larger training sample size to improve the generalization of the forecasting model. A large training sample size with high data resolution should be used for training if there is no computational power constraint, and the training time is not the priority.

Some of the guideline from the literature, together with trial and error methods, have been used in this research to determine suitable parameters for the ANN models. The simulation results indicated that the number of hidden layers and hidden neurons has a minimal impact on the problem with a simple input and output relationship. The ANN models with two hidden layers, each hidden layer having ten hidden neurons have a better 2h-ahead wind speed and wind power forecasting performance than other combinations tested in this research. Different combinations of parameters of PSO are also tested in this research. The selected values of the PSO parameters are very similar to the one suggested in the literature. Trial and error methods, together with the guidelines from the literature, have been used as an effective way to determine suitable model parameters.

The effect of the meteorological variables from the neighbouring stations on the forecasting accuracy of the target station was investigated. Spatial correlation analysis is carried out for two data sources, the WASA dataset and the NREL dataset. The simulation results of both WASA dataset and NREL dataset indicated that using the input features from the highly correlated neighbouring sites can enhance the forecasting performance of the target station. However, using the input features from the distant stations not only slows down the training process but also increases the chance of overfitting the forecasting model. The performance of the forecasting models has stopped improving when too many input features from the distant stations are used to train the forecasting models.

The input-feature-based ensemble has been used to obtain useful information from many surrounding meteorological features in parallel structure to avoid overfitting issues and improve forecasting

performance effectively. Currently, these methods are not used in the Utilities because the technology is still being developed. With high penetration of wind energy, it is expected that these methods will be used in the future for accurate prediction of short-term wind power.

7. Recommendations and future work

Plenty of research has been carried out on short-term wind speed and wind power forecasting. The factors that affect the performance of artificial neural network-based forecasting models are investigated in this research. Further research is required to improve the performance of short-term wind speed and wind power forecasting further. Based on the simulations results and conclusions, several recommendations are drawn in the following.

A bigger training sample size can provide more scenarios which can enhance the generalization of the forecasting model. Although it requires more steps of out-sample forecasting for the model with a higher data resolution than the one with a lower data resolution, high-resolution data can provide more information about the transition between two states to the forecasting model. The only disadvantage of using larger sample size and a higher date resolution is that it requires a longer time to train the forecasting model.

Artificial intelligence methods: The popular artificial intelligence methods, feedforward neural networks (ANN) was considered in this research. Other artificial intelligence methods, such as the convolutional neural network (CNN), recurrent neural networks (RNN), support vector machine (SVM), fuzzy logic, and et cetera should be investigated in further research.

Meteorological variables: Only seven different meteorological variables from three separate data sources are available to use in this research. Other meteorological variables, such as precipitation, clouds coverage, solar radiation, and et cetera, need to be investigated to see whether they can improve the performance of the forecasting model.

Actual wind power data: The data obtained from Wind Atlas of South Africa contain six useful meteorological variables. However, the WASA datasets contain no actual wind power. Even though the Sotavento Galicia wind farm datasets have real wind power data for the entire wind farm, it lacks other meteorological variables data, such as temperature, pressure, relative humidity, and the real wind power data of individual wind turbine. Real wind power data of individual wind turbine together with more meteorological variables from the same site should be used in further research.

More ensemble and hybrid models: Two ensemble model are investigated in this research. There are plenty of other ensemble models (i.e., SVM-based ensemble, RNN-based ensemble) and hybrid models (i.e., ANN-SVM, ANN-RNN) need to be investigated in further research.

8. References

[1] F. Fazelpour, N. Tarashkar and M. A. Rosen, "Short-term wind speed forecasting using artificial neural networks for Tehran, Iran," *International Journal of Energy and Environment,* vol. 7, pp. 377-390, 2016.

[2] S. Chai, Z. Xu, L. L. Lai and K. P. Wong, "An overview on wind power forecasting methods," in *International Conference on Machine Learning and Cybernetics*, Guangzhou, 2015.

[3] A. Yadav and K. Sahu, "Wind forecasting using artificial neural networks: survey and taxonomy," *International Journal of Research In Science & Engineering,* vol. 3, no. 2, pp. 148-155, 2017.

[4] S. Fan, J. R.Liao, R. Yokoyama, L. Chen and W.-J. Lee, "Forecasting the wind generation using a two-stage network based on meteorological information," *IEEE Transactions on Energy Conversion,* vol. 24, no. 2, pp. 474-482, 2009.

[5] World Wind Energy Association, "World Wind Energy Association," World Wind Energy Association, 25 February 2019. [Online]. Available: https://wwindea.org/blog/2019/02/25/wind-power-capacity-worldwide-reaches-600-gw-539-gw-added-in-2018/. [Accessed 23 August 2019].

[6] M. G. De Giorgi, A. Ficarella and M. Tarantino, "Assessment of the benefits of numerical weather predictions in wind power forecasting based on statistical methods," *Energy,* vol. 36, pp. 3968-3978, 2011.

[7] D. R. Walwyn and A. C. Brent, "Renewable energy gathers steam in South Africa," *Renewable and Sustainable Energy Reviews,* vol. 41, pp. 390-401, 2015.

[8] "IRP PROJECTS," [Online]. Available: https://www.ipp-renewables.co.za/. [Accessed 27 January 2020].

[9] D. Barbosa de Alencar, C. d. M. Affonso, R. C. Limao de Oliveira, J. L. M. Rodriguez, J. C. Leite and J. C. R. Filho, "Different models for forecasting wind power generation: case study," *Energies,* vol. 10, no. 1976, pp. 1-27, 2017.

[10] A. Tesfaye, J. Zhang, D. Zheng and D. Shiferaw, "Short-term wind power forecasting using artificial neural networks for resource scheduling in microgrids," *International Journal of Science and Engineering Application,* vol. 5, no. 3, pp. 144-151, 2016.

[11] S. S. Soman, H. Zareipour, O. Malik and P. Mandal, "A review of wind power and wind speed forecasting methods with different time horizons," in *North American Power Symposium*, Arlington, TX, USA, 2010.

[12] X. Peng, D. Deng, J. Wen, L. Xiong, S. Feng and B. Wang, "A very short term wind power forecasting approach based on numerical weather prediction and correction method," in *International Conference on Electricity Distribution*, Xi'an, 2016.

[13] P. S. Kumar and D. Lopez, "Forecasting of wind speed using feature selection and neural networks," *International Journal of Renewable Energy Research,* vol. 6, no. 3, pp. 833-837, 2016.

[14] X. Zhao, S. Wang and T. Li, "Review of evaluation criteria and main methods of wind power forecasting," *Energy Procedia,* vol. 12, pp. 761-769, 2011.

[15] A. Costa, A. Crespo, J. Navarro, G. Lizcano, H. Madsen and E. Feitosa, "A review on the young history of the wind power short-term prediction," *Renewable & Sustainable Energy Reviews,* vol. 12, pp. 1725-1744, 2008.

[16] J. Jung and R. P. Broadwater, "Current status and future advances for wind speed and power forecasting," *Renewable and Sustainable Energy Reviews,* vol. 31, pp. 762-777, 2014.

[17] X. Wang, P. Guo and X. Huang, "A review of wind power forecasting models," *Energy Procedia,* vol. 12, pp. 770-778, 2011.

[18] W.-Y. Chang, "A literature review of wind forecasting methods," *Journal of Power and Energy Engineering,* vol. 2, pp. 161-168, 2014.

[19] R. G. Kavasseri and K. Seetharaman, "Day-ahead wind speed forecasting using f-ARIMA models," *Renewable Energy,* vol. 34, pp. 1388-1393, 2009.

[20] A. B. Botterud, J. Wang, V. Miranda and R. J. Bessa, "Wind power forecasting in U.S. electricity markets," *The Electricity Journal,* vol. 23, no. 3, pp. 71-82, 2010.

[21] L. Ma, S. Luan, C. Jiang, H. Liu and Y. Zhang, "A review on the forecasting of wind speed and generated power," *Renewable and Sustainable Energy Reviews,* vol. 13, pp. 915-920, 2009.

[22] B. Doucoure, K. Agbossou and A. Cardenas, "Time series prediction using artificial wavelet neural network and multi-resolution analysis: application to wind data," *Renewable Energy,* vol. 92, pp. 202-211, 2016.

[23] A. A. Abdoos, "A new intelligent method based on combination of VMD and ELM for short term wind power forecasting," *Neurocomputing,* vol. 203, pp. 111-120, 2016.

[24] S. Garcia, S. Ramirez-Gallego, J. Luengo, J. M. Benitez and F. Herrera, "Big data preprocessing: methods and prospects," *Big Data Analytics,* vol. 1, no. 9, pp. 1-22, 2016.

[25] A. Azadeh, M. Sheikhalishahi, M. Tabesh and A. Negahban, "The effects of pre-processing methods on forecasting improvement of artificial neural networks," *Australian Journal of Basic and Applied Sciences,* vol. 5, no. 6, pp. 570-580, 2011.

[26] Q. Chen and K. Folly, "Effect of input features on the performance of ANN-based wind power forecasting," in *2019 SAUPEC/RobMech/PRASA Conference*, Bloemfontein, 2019.

[27] J. Adamowski and H. F. Chan, "A wavelet neural network conjunction model for groundwater level forecasting," *Journal of Hydrology,* vol. 407, pp. 28-40, 2011.

[28] B. B. Afshar, "Wind Power Forecasting Using Artificial Neural Networks with Numerical Prediciton - A Case Study for Mountainous Canada," The University of British Columbia, Vancouver, 2009.

[29] H. Zareipour, D. Huang and W. Rosehart, "Wind power ramp events classification and forecasting: A data mining approach," in *IEEE Power and Energy Society General Meeting*, Detroit, 2011.

[30] S. Li, P. Wang and L. Goel, "Wind power forecasting using neural network," *IEEE Transactions on Sustainable Energy,* vol. 6, no. 4, pp. 1447-1456, 2015.

[31] I. Dimoulkas, P. Mazidi and L. Herre, "Neural networks for GEFCom2017 probabilistic load forecasting," *International Journal of Forecasting,* pp. 1-15, 2018.

[32] H. Wimmer and L. Powell, "Principle component analysis for feature reduction and data preprocessing in data science," in *Proceedings of the Conference on Information Systems Applied Research*, Las Vegas, 2016.

[33] C. Carrillo, A. Obando Montano, J. Cidras and E. Diaz-Dorado, "Review of power curve modelling for wind turbines," *Renewable and Sustainable Energy Reviews,* vol. 21, pp. 572-581, 2013.

[34] K. R. Nair, V. Vanith and M. Jisma, "Forecasting of wind speed using ANN, ARIMA and hybrid models," in *International Conference on Intelligent Computing, Instrumentation and Control Technologies*, Kannur, India, 2017.

[35] Z.-h. Guo, J. Wu, H.-y. Lu and J.-z. Wang, "A case study on a hybrid wind speed forecasting method using BP neural network," *Knowledge-Based Systems,* vol. 24, pp. 1048-1056, 2011.

[36] A. M. Foley, P. G. Leahy, A. Marvuglia and J. E. McKeogh, "Current methods and advances in forecasting of wind power generation," *Renewable Energy,* vol. 37, pp. 1-8, 2012.

[37] H. i. a. M. World, "Human in a Machine World," Human in a Machine World, 2016 March 23. [Online]. Available: https://medium.com/human-in-a-machine-world/mae-and-rmse-which-metric-is-better-e60ac3bde13d. [Accessed 2019 August 31].

[38] B. Uniejewski, G. Marcjasz and R. Weron, "Understanding intraday electricity markets: Variable selection and very short-term price forecasting using LASSO," *International Journal of Forecasting,* vol. 35, no. 4, pp. 1533-1547, 2019.

[39] S. A. Wadi and M. T. Ismail, "Selection wavelet transforms model in forecasting financial time series data based on ARIMA model," *Applied Mathematical Sciences,* vol. 5, no. 7, pp. 315-326, 2011.

[40] H. Chitsaz, N. Amjady and H. Zareipour, "Wind power forecasting using wavelet neural network trained by improved Clonal selection algorithm," *Energy Conversion and Management,* vol. 89, pp. 588-598, 2015.

[41] C. Stolojescu, I. Railean, S. Moga, I. Alexandru and P. Lenca, "A wavelet based prediction method for time series," in *Stochastic Modeling Techniques and Data Analysis International Conference*, Chania, Crete, 2010.

[42] H. Pousinho, V. Mendes and J. Catalao, "A hybrid PSO-ANFIS approach for short-term wind power prediction in Portugal," *Energy Conversion and Management,* vol. 52, pp. 397-402, 2010.

[43] M. Layia and S. S. Kumar, "A comprehensive overview on wind power forecasting," in *Power Engineering Conference (IPEC), International*, Singapore, 2010.

[44] P. Zhao, J. Wang, J. Xia, Y. Dai, Y. Sheng and J. Yue, "Performance evaluation and accuracy enhancement of a day-ahead wind power forecasting system in China," *Renewable Energy,* vol. 43, pp. 234-241, 2012.

[45] B. Lange, K. Rohrig, B. Ernst, F. Schloegl, U. Cali, R. Jursa and J. Moradi, "Wind power prediction in Germany-recent advances and future challenges," in *European Wind Energy Conference & Exhibition (EWEC)*, Athen, 2006.

[46] J. Yu, X. Chen, K. Yu and Y. Liao, "Short-term wind power forecasting using hybrid method based on enhanced boosting algorithm," *Journal of Modern Power Systems and Clean Energy,* vol. 5, no. 1, pp. 126-133, 2017.

[47] H. Masrur, M. Nimol, M. Faisal and S. M. G. Mostafa, "Short term wind speed forecasting using artificial neural network: A case study," in *International Conference on Innovations in Science, Engineering and Technology (ICISET)*, Dhaka, Bangladesh, 2016.

[48] H. Quan, D. Srinivasan and Abbas Khosravi, "Short-term load and wind power forecasting using neural network-based prediction intervals," *IEEE Transactions on Neural Network and Learning Systems,* vol. 25, no. 2, pp. 303-314, 2014.

[49] S. Smyl and N. G. Hua, "Machine learning methods for GEFCom2017 probabilistic load forecasting," *International Journal of Forecasting,* vol. 35, no. 4, pp. 1424-1431, 2019.

[50] C. A. G. Santos and G. B. Lima Da Silva, "Daily streamflow forecasting using a wavelet transform and artificial neural network hybrid models," *Hydrological Sciences Journal,* vol. 59, no. 2, pp. 2150-3435, 2014.

[51] A. T. Eseye, J. Zhang, D. Zheng, H. Li and J. Gan, "A double-stage hierarchical hybrid PSO-ANN model for short-term wind power prediction," in *IEEE 2nd International Conference on Cloud Computing and Big Data Analysis (ICCCBDA)*, Chengdu, China, 2017.

[52] M. Mana, M. Burlando and C. Meibner, "Evaluation of two of ANN approaches for the wind power forecast in a mountainous site," *International Journal of Renewable Energy Research,* vol. 7, no. 4, pp. 1629-1638, 2017.

[53] G. Osorio, J. Matias and J. Catalao, "Short-term wind power forecasting using adaptive neuro-fuzzy inference system combined with evolutionary particle swarm optimization, wavelet transform and mutual information," *Renewable Energy,* vol. 75, pp. 301-307, 2015.

[54] H. Madsen, P. Pinson, G. Kariniotakis, H. A. Nielsen and T. S. Nielsen, "Standardizing the performance evaluation of short-term wind power prediction models," *Wind Engineering,* vol. 29, no. 6, pp. 475-489, 2005.

[55] Y.-K. Wu and J.-S. Hong, "A literature review of wind forecasting technology in the world," in *IEEE Lausanne Power Tech*, Lausanne, Switzerland, 2007.

[56] S. Alessandrini, L. D. Monache, S. Sperati and J. Nissen, "A novel application of an analog ensemble for short-term wind power forecasting," *Renewable Energy,* vol. 76, pp. 768-781, 2015.

[57] X. Zhu and M. G. Genton, "Short-term wind speed forecasting for power system operations," *International Statistical Institute,* vol. 80, no. 1, pp. 2-23, 2012.

[58] M. J. Ghadi, S. H. Gilani, H. Afrakhte and A. Baghramian, "A novel heuristic method for wind farm power prediction: A case study," *International Journal of Electrical Power and Energy Systems,* vol. 63, pp. 962-970, 2014.

[59] J. Zeng and W. Qiao, "Support vector machine-based short-term wind power forecasting," in *2011 IEEE/PES Power Systems Conference and Exposition*, Phoenix, AZ, USA, 2011.

[60] A. Khosravi, S. Nahavandi and D. Creighton, "Prediction intervals for short-term wind farm power generation forecasts," *IEEE Transactions on Sustainable Energy,* vol. 4, no. 3, pp. 602-610, 2013.

[61] M. Premkumar and R. Sowmya, "Wind power forecasting using artificial neural network," *International Journal of Pure and Applied Mathematics,* vol. 118, no. 11, pp. 145-152, 2018.

[62] J. Catalao, H. Pousinho and V. Mendes, "Short-term wind power forecasting in Portugal by neural networks and wavelet transform," *Renewable Energy,* vol. 36, pp. 1245-1251, 2011.

[63] H. Liu, H.-Q. Tian, C. Chen and Y.-f. Li, "A hybrid statistical method to predict wind speed and wind power," *Renewable Energy,* vol. 35, pp. 1857-1861, 2010.

[64] C. Potter and M. Negnevitsky, "Very short-term wind forecasting for Tasmanian power generation," *IEEE Transactions on Power Systems,* vol. 21, no. 2, pp. 965-972, 2006.

[65] M. Lange and U. Focken, "New developments in wind energy forecasting," in *IEEE Power and Energy Society General Meeting - Conversion and Delivery of Electrical Energy in the 21st Century*, Pittsburgh, PA, USA, 2008.

[66] S. Mi, "Short-term forecasting of wind speed and wind power based on BP and AdaBoost_BP," University of Wisconsin-Milwaukee, Milwaukee, 2014.

[67] U. Cali, B. Lange, J. Dobschinsi, M. Kurt, C. Moehrlen and B. Ernst, "Artificial neural network based wind power forecasting using a multi-model approach," in *7th International Workshop on Large Scale Integration of Wind Power and on Transmission Networks for Offshore Wind Farms*, Madrid, 2008.

[68] J. Jin and J. Kim, "Forecasting natural gas prices using wavelets, time series, and artificial neural networks," *Plos One,* vol. 10, no. 11, pp. 1-23, 2015.

[69] A. Vaccaro, P. Mercogliano, P. Schiano and D. Villacci, "An adaptive framework based on multi-model data fusion for one-day-ahead wind power forecasting," *Electric Power Systems Research,* vol. 81, pp. 775-782, 2011.

[70] J. Shi, J. Guo and S. Zheng, "Evaluation of hybrid forecasting approaches for wind speed and power generation time series," *Renewable and Sustainable Energy Reviews,* vol. 16, pp. 3471-3480, 2012.

[71] S. Karsoliya, "Approximating number of hidden layer neurons in multiple hidden layer BPNN architecture," *International Journal of Engineering Trends and Technology,* vol. 3, no. 6, pp. 714-717, 2012.

[72] K. G. Sheela and S. Deepa, "Review on methods to fix number of hidden neurons in neural networks," in *14th International Conference on the European Energy Market*, Dresden, Germany, 2013.

[73] Y. Liu, J. A. Starzyk and Z. Zhu, "Optimizing number of hidden neurons in neural networks," in *IASTED International Conference on Artificial Intelligence and Applications*, Innsbruck, Austria, 2007.

[74] A. K. Fard and M.-R. Akbari-Zadeh, "A hybrid method based on wavelet, ANN and ARIMA model for short-term load forecasting," *Journal of Experimental & Theoretical Artificial Intelligence,* vol. 26, no. 2, pp. 167-182, 2014.

[75] L. Wang, H. Zou, J. Su, L. Ling and S. Chaudhry, "An ARIMA-ANN hybrid model for time series forecasting," *Systems Research and Behavioral Science,* vol. 30, pp. 244-259, 2013.

[76] J. Kennedy and R. Eberhart, "Particle swarm optimization," in *Proceedings of ICNN'95 - International Conference on Neural Networks*, Perth, Australia, 1995.

[77] Z. Boger and H. Guterman, "Knowledge extraction from artificial neural network models," in *IEEE International Conference on Systems, Man, and Cybernetics. Computational Cybernetics and Simulation*, Orlando, USA, 1997.

[78] S. Xu and L. Chen, "A novel approach for determining the optimal number of hidden layer neurons for FNN's and its application in data mining," in *5th International Conference on Information Technology and Applications (ICITA 2008*, Cairns, Queensland, Australia, 2008.

[79] M. T. Hagan and M. B. Menhaj, "Training feedforward networks with the Marquardt algorithm," *IEEE Transactions on Neural Networks,* vol. 5, no. 6, pp. 989-993, 1994.

[80] J. Zhou, Z. Duan, Y. Li, J. Deng and D. Yu, "PSO-based neural network optimization and its utilization in a boring machine," *Journal of Materials Processing Technology,* vol. 178, pp. 19-23, 2006.

[81] R. Adhikari and R. Agrawal, "Effectiveness of PSO Based Neural Network for Seasonal Time Series Forecasting," in *Proceedings of the Fifth Indian International Conference on Artificial Intelligence*, Tumkur, India, 2011.

[82] K. Chau, "Application of a PSO-based neural network in analysis of outcomes of construction claims," *Automation in Construction,* vol. 16, pp. 642-646, 2007.

[83] A. R. Jordehi and J. Jasni, "Parameter selection in particle swarm optimisation: a survey," *Journal of Experimental & Theoretical Artificial Intelligence,* vol. 25, no. 4, pp. 527-542, 2013.

[84] Y. He, W. J. Ma and J. P. Zhang, "The parameters selection of PSO algorithm influencing on performance of fault diagnosis," *MATEC Web of Conferences,* vol. 63, pp. 1-5, 2016.

[85] Y. Dai, L. Liu and Y. Li, "An intelligent parameter selection method for particle swarm optimization algorithm," in *2011 Fourth International Joint Conference on Computational Sciences and Optimization*, Yunnan, China, 2011.

[86] M. E. H. Pedersen, "Good Parameters for Particle Swarm Optimization," Hvass Laboratories, 2010.

[87] T. Beielstein, K. E. Parsopoulos and M. N. Vrahatis, "Turning PSO parameters through sensitivity analysis," 2002.

[88] M. Alexiadis, P. Dokopoulos, H. Sahsamanoglou and I. Manousaridis, "Short-term forecasting of wind speed and related electrical power," *Solar Energy,* vol. 63, no. 1, pp. 61-68, 1998.

[89] I. G. Damousis, M. C. Alexiadis, J. B. Theocharis and P. S. Dokopoulos, "A fuzzy model for wind speed prediction and power generation in wind parks using spatial correlation," *IEEE Transactions on Energy Conversion,* vol. 19, no. 2, pp. 352-361, 2004.

[90] U. Focken, M. Lange, M. Kai, H.-P. Waldl, H. G. Beyer and A. Luig, "Short-term prediction of the aggregated power output of wind farms-a statistical analysis of the reduction of the prediction error by spatial smoothing effects," *Journal of Wind Engineering,* vol. 90, pp. 231-246, 2002.

[91] M. Bilgili, B. Sahin and A. Yasar, "Application of artificial neural networks for the wind speed prediction of target station using reference stations data," *Renewable Energy,* vol. 32, pp. 2350-2360, 2007.

[92] A. Tascikaraoglu and M. Uzunoglu, "A review of combined approaches for prediction of short-term wind speed and power," *Renewable and Sustainable Energy Reviews,* vol. 34, pp. 243-254, 2014.

[93] D. Zhao, Y. Zhu and X. Zhang, "Research on wind power forecasting in wind farms," in *Power Engineering and Automation Conference (PEAM)*, Wuhan, China, 2011.

[94] R. Jursa and K. Rohrig, "Short-term wind power forecasting using evolutionary algorithms for the automated specification of artificial intelligence models," *International Journal of Forecasting,* vol. 24, pp. 694-709, 2008.

[95] E. Cadenas and W. Rivera, "Wind speed forecasting in three different regions of Mexico, using a hybrid ARIMA," *Renewable Energy,* vol. 35, pp. 2732-2738, 2010.

[96] Y. Ren, P. Suganthan and N. Srikanth, "Ensemble methods for wind and solar power forecasting-A state-of-the-art review," *Renewable and Sustainable Energy Reviews,* vol. 50, pp. 82-91, 2015.

[97] D. Opitz and R. Maclin, "Popular ensemble methods: an empirical study," *Journal of Artificial Intelligence Research,* vol. 11, pp. 169-198, 1999.

[98] G. Brown, J. L. Wyatt and P. Tino, "Managing diversity in regression ensembles," *Journal of Machine Learning Research,* vol. 6, pp. 1621-1650, 2005.

[99] J. Chen, G.-Q. Zeng, W. Zhou, W. Du and K.-D. Lu, "Wind speed forecasting using nonlinear-learning ensemble of deep learning time series prediction and extremal optimization," *Energy Conversion and Management,* vol. 165, pp. 681-695, 2018.

[100] "Sotavento," Sotavento Galicia, S.A., [Online]. Available: http://www.sotaventogalicia.com/en. [Accessed 05 February 2020].

[101] "Wind Prospector," National Renewable Energy Laboratory, [Online]. Available: https://maps.nrel.gov/wind-prospector/. [Accessed 11 November 2019].

[102] "Wind Atlas for South Africa," Wind Atlas for South Africa, [Online]. Available: http://wasa.csir.co.za/web/welcome.aspx. [Accessed 14 June 2019].

[103] "Wind in Spain," Mallorca Kiteschool, 01 December 2017. [Online]. Available: https://mallorcakiteschool.com/en/winds-in-spain/. [Accessed 05 February 2020].

[104] C. Draxl, B. M. Hodge, A. Clifton and J. McCaa, "Overview and Meteorological Validation of the Wind Integration National Dataset Toolkit," National Renewable Energy Laboratory (NREL), Denver, 2015.

[105] J. King, A. Clifton and B. Hodge, "Validation of Power Output for the WIND Toolkit," National Renewable Energy Laboratory, Denver, 2014.

[106] R. J. Hyndman and G. Athanasopoulos, Forecasting: Principles and Practice, 2nd edition, Melbourne, Australia: OTexts, 2018.

[107] MathWorks, "MathWorks," The Matchworks, Inc., [Online]. Available: https://www.mathworks.com/help/deeplearning/ref/trainlm.html. [Accessed 22 October 2019].

[108] I. C. Trelea, "The particle swarm optimization algorithm convergenec analysis and parameter selection," *Information Processing Letters,* vol. 85, pp. 317-325, 2003.

[109] "South Africa Wind Farms," South Africa Wind Energy Association, [Online]. Available: https://sawea.org.za/wind-map/wind-ipp-table/. [Accessed 01 October 20].

[110] V. W. Systems, "Vestas," Vestas Group, [Online]. Available: https://www.vestas.com/en/products/2-mw-platform/v90-2_0_mw#!. [Accessed 04 December 2019].

[111] Q. Chen and K. Folly, "Comparison of three methods for short-term wind power forecasting," in *International Joint Conference on Neural Networks (IJCNN)*, Rio de Janeiro, 2018.

[112] Q. Chen and K. Folly, "Wind power forecasting," in *Control of Power and Energy Systems - 10th CPES 2018*, Tokyo, Japan, 2018.

[113] Q. Chen and K. Folly, "Short-term wind power forecasting based on spatial correlation and artificial neural network," in *Southern African Universities Power Engineering Conference/ Robotics and Mechatronics/ Pattern Recognition Association of South Africa*, Cape Town, 2020.

[114] G. Siderators and N. D. Hatziargyriou, "An advanced statistical method for wind forecasting," *IEEE Transactions on Power Systems,* vol. 22, no. 1, pp. 1-8, 2007.

9. Published papers

1. Q. Chen and K. Folly, “Comparison of three methods for short-term wind power forecasting,” in *International Joint Conference on Neural Networks (IJCNN)*, Rio de Janeiro, 2018. [111].

2. Q. Chen and K. Folly, “Wind power forecasting,” in *Control of Power and Energy Systems - 10th CPES 2018*, Tokyo, Japan, 2018. [112].

3. Q. Chen and K. Folly, "Effect of input features on the performance of the ANN-based wind power forecasting," in 2019 Southern African Universities Power Engineering Conference/Robotics and Mechatronics/Pattern Recognition Association of South Africa (SAUPEC/RobMech/PRASA), pp. 673-678 Bloemfontein, South Africa, 2019 [26].

4. Q. Chen and K. Folly, “Short-term wind power forecasting based on spatial correlation and artificial neural network,” in Southern African Universities Power Engineering Conference/ Robotics and Mechatronics/ Pattern Recognition Association of South Africa (SAUPEC/RobMech/PRASA), Cape Town, 2020 [113].

10. Appendix

10.1 Appendix A (comparison with other forecasting models)

To accurately make a quantitative comparison between two different models, the data used must be the same [15]. It is difficult to accurately compare a large number of forecasting models which uses various datasets and under different situations [42]. Although it is difficult to accurately compare the performance of the forecasting models used in this research with the forecasting models presented in the literature [99], numerical comparison with these models are shown in the following tables.

10.1.1 Comparison of different artificial intelligence methods

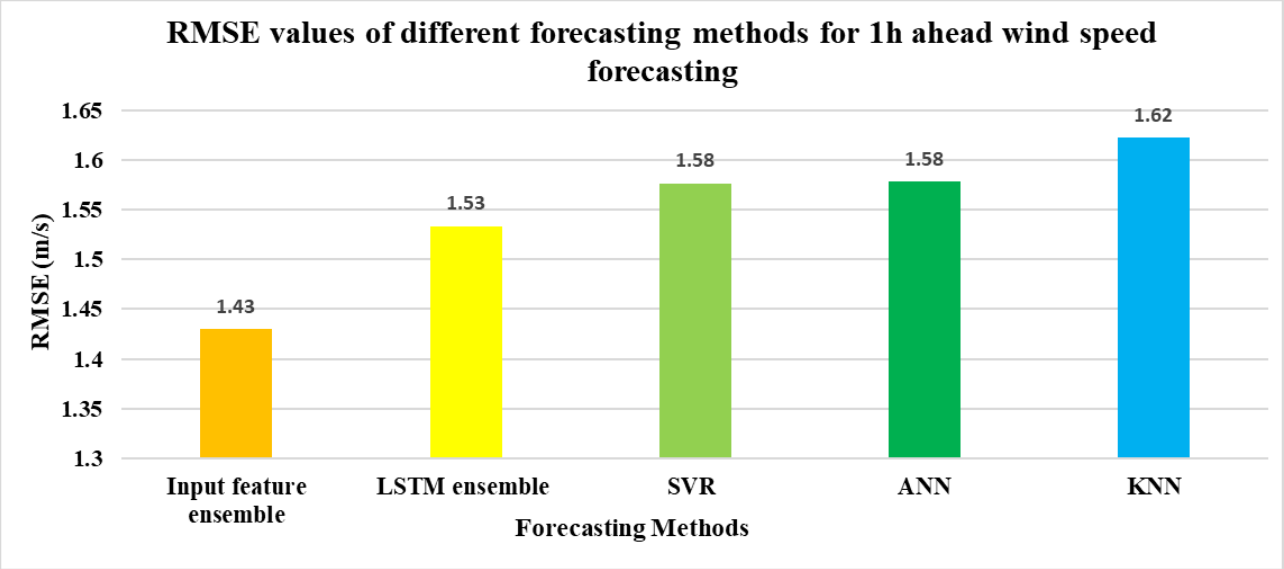


Figure 10.1: RMSE values of different forecasting methods for 1h ahead wind speed forecasting (NREL)

Note: LSTM = Long short-term memory, SVR = support vector regression, ANN = feedforward artificial neural network, KNN = K-Nearest Neighbours.

As can be seen in Figure 10.1, the input-feature-based ensemble (Level 5) of the NREL dataset has the lowest RMSE value compared to the RMSE values of other methods presented in the literature for one hour ahead wind speed forecasting. Figure 10.2 and Figure 10.3 show the MAE values and MAPE values of different forecasting methods, respectively.

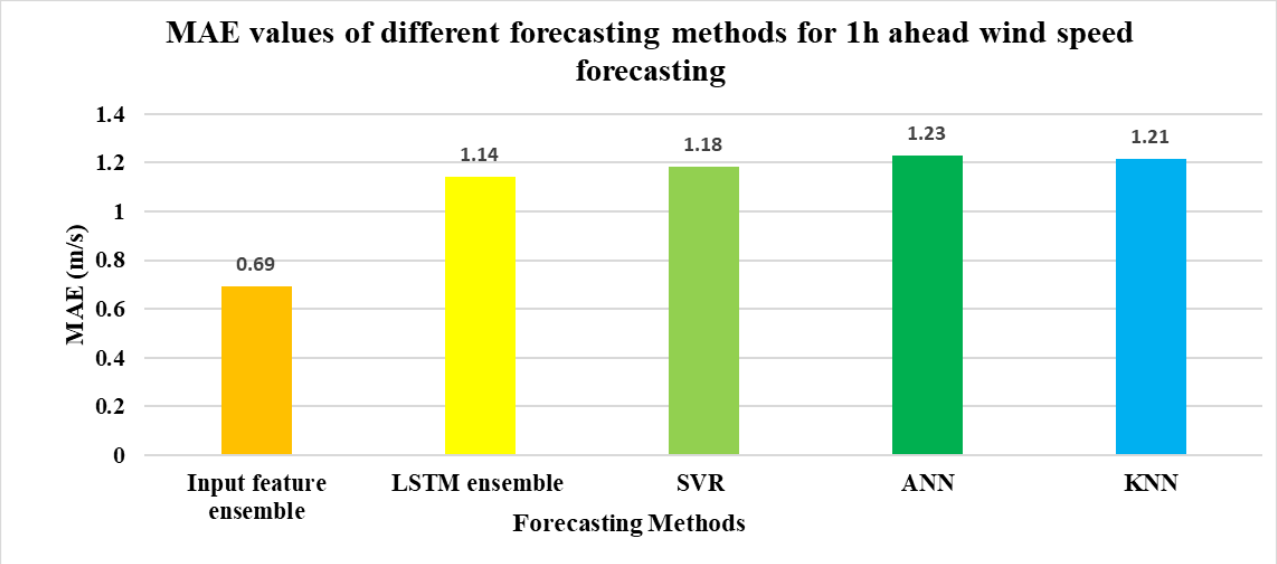


Figure 10.2: MAE values of different forecasting methods for 1h ahead wind speed forecasting (NREL)

As shown in Figure 10.2, the input-feature-based ensemble model proposed in this research has the lowest MAE value compared to the method suggested in the literature.

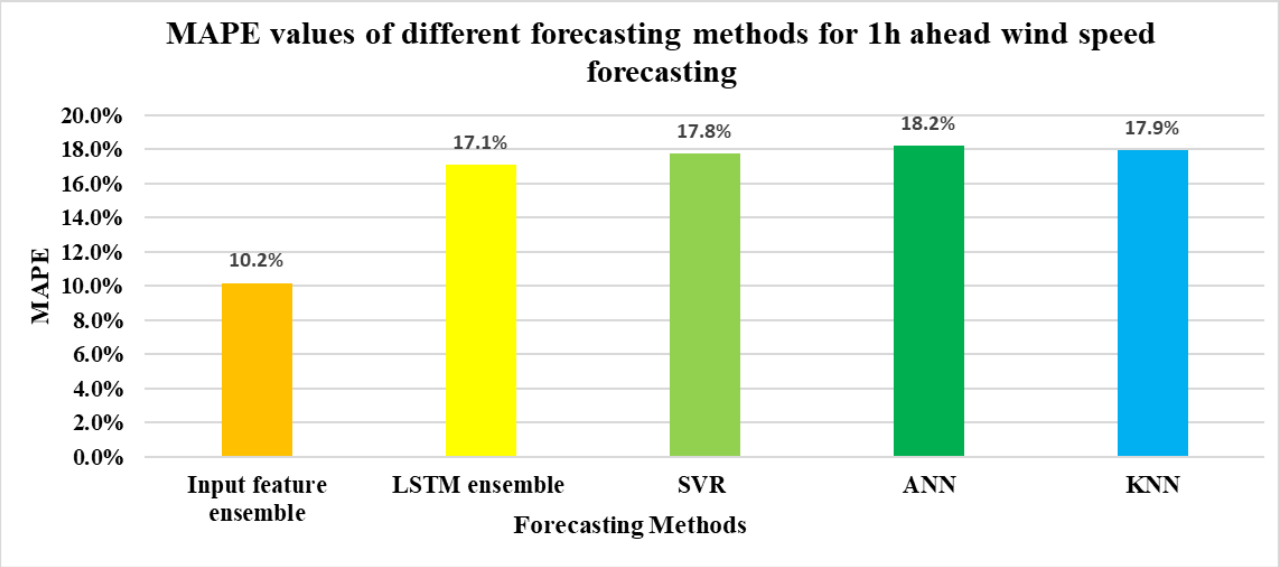


Figure 10.3: MAPE values of different forecasting methods for 1h ahead wind speed forecasting (NREL)

The conclusions made for the RMSE values and MAE values can be made for the MAPE values. The input-feature-based ensemble (Level 5) has the lowest MAPE value among other forecasting methods for one hour ahead wind speed forecasting. NMAE values of the input-feature-based ensemble model of NREL against forecasting time horizon is shown in Figure 10.4.

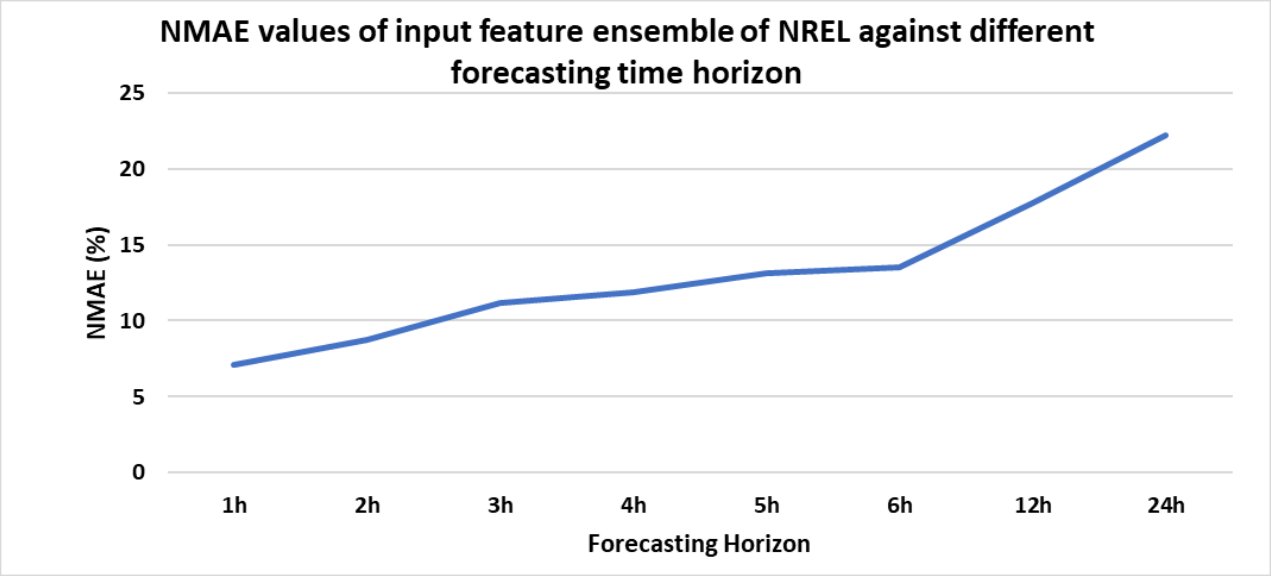


Figure 10.4: NMAE values of the input-feature-based ensemble model of NREL against different forecasting time horizon of short-term wind speed forecasting (NREL).

The NMAE values of the input-feature-based ensemble model of NREL increase gradually with a longer forecasting time horizon. In general, the forecasting accuracy degrades with increasing forecast time horizon, terrain complexity, and unstable weather condition [16] [20].

10.1.2 ARIMA

The 2h-ahead wind speed forecasting performance of the ARIMA (3, 1, 2) (which stands for the third order of autoregressive and first-order difference and second-order moving average) is compared to a basic ANN model that uses only wind speed as input. For a fair comparison, both the ARIMA (3, 1, 2) model and the basic ANN model use only wind speed as an input feature.

Table 10.1: Simulation results of the ARIMA (3, 1, 2) and a basic ANN model that uses only wind speed as the input

Forecasting model	RMSE (m/s)	MAE (m/s)	MAPE
ARIMA (3, 1, 2)	2.62	2.08	29.15%
Basic ANN uses wind speed as input feature	2.23	1.77	25.48%

The simulation results indicate that the basic ANN model performs much better than the ARIMA model. This is expected as ANNs are usually good at modelling and forecasting non-linear time series like wind speed. ARIMA models are suitable for linear time series and one-step ahead forecasting. In this dissertation, ARIMA was used to forecast 2h-ahead wind speed, which is 12 steps out of sample forecasting for the 10 minutes per sample dataset.

10.2 Appendix B (input-feature-based ensemble)

Table 10.2: Simulation results of the ANN-based spatial correlation model and the input-feature-based ensemble model with input features from different numbers of neighbouring sites for 2h-ahead wind speed forecasting (NREL)

No. of sites	ANN-based spatial correlation model			Input-feature-based ensemble model		
	RMSE(m/s)	MAE(m/s)	MAPE	RMSE(m/s)	MAE(m/s)	MAPE
2	1.49	1.02	15.3%	1.49	1.01	14.8%
3	1.46	1.00	14.7%	1.49	1.01	14.9%
4	1.46	1.00	14.7%	1.46	1.00	14.5%
5	1.44	0.98	14.4%	1.43	0.97	14.2%
6	1.45	0.98	14.5%	1.42	0.96	14.1%
7	1.44	0.97	14.3%	1.41	0.96	14.0%
8	1.43	0.97	14.3%	1.41	0.95	14.0%
9	1.43	0.97	14.3%	1.40	0.95	13.9%
10	1.42	0.96	14.1%	1.39	0.93	13.8%
11	1.43	0.97	14.3%	1.38	0.94	13.7%
12	1.39	0.95	14.0%	1.37	0.93	13.6%
13	1.39	0.95	14.0%	1.36	0.92	13.5%
14	1.36	0.93	13.7%	1.34	0.91	13.4%
15	1.37	0.93	13.7%	1.32	0.90	13.2%
16	1.38	0.94	13.8%	1.31	0.89	13.1%
17	1.36	0.93	13.6%	1.31	0.89	13.1%
18	1.35	0.91	13.4%	1.29	0.88	12.9%
19	1.37	0.91	13.4%	1.28	0.87	12.7%
20	1.35	0.92	13.5%	1.27	0.86	12.6%
21	1.37	0.94	13.7%	1.26	0.86	12.6%
22	1.33	0.90	13.2%	1.25	0.85	12.4%
23	1.37	0.94	13.7%	1.25	0.85	12.4%
24	1.41	0.93	13.5%	1.24	0.85	12.4%
25	1.35	0.91	13.3%	1.24	0.85	12.4%
26	1.38	0.93	13.6%	1.24	0.85	12.4%
27	1.40	0.93	13.6%	1.24	0.85	12.4%
28	1.37	0.95	13.9%	1.23	0.84	12.3%
29	1.42	0.95	13.9%	1.23	0.84	12.2%
30	1.44	0.94	13.8%	1.23	0.83	12.2%

10.3 Appendix C

Table 10.3 summarises some of the existing wind speed and wind power forecasting methods [2], [11], [92].

Table 10.3: Summary of some of the existing short-term wind speed and wind power forecasting methods

Forecasting method	Subclass	Examples	Advantages and disadvantages
Physical approach	-NWP	-MM5 -Prediktor -High-Resolution Limited Area Model (HIRLAM) - Weather Research and Forecasting (WRF)	-Accurate for long-term forecasting -Not suitable for the short time horizon -High computational cost
Statistical approach	-Time Series	-AR -ARX -ARMA -ARIMA -Persistence -Linear predictions -Gray predictor -Kalman filter -Exponential smoothing	-Accurate for short term prediction. -Time series models are easy to build. -Able to provide confidence intervals for predictions. -A large amount of historical data is required. -Not suitable for non-linear problems.
Computational intelligence	-ANN	-Feedforward -Multilayer Perceptron -Recurrent -Radial basis function	-Accurate for short term forecasting -Strong learning and training abilities -Usually, outperform time series models -High data error tolerance -Requires a large amount of training data
	-Other	-SVM -Fuzzy Logic -Evolutionary Computation and swarm intelligence (Genetic Algorithm GA, PSO, etc.)	-SVM: high generalization performance, need to tune parameter properly, longer training time compares to other models -Fuzzy Logic: suitable for the systems that are difficult to model; good at reasoning problems; high model complexity; requires longer training time when there are many rules.
Hybrid approach	-	-ANN+PSO -ANN+NWP -ANN+ Fuzzy (ANFIS) -ANN+ARIMA -ANN + WT -ANN + spatial correlation	-Able to incorporate the individually superior features of various forecasting models to obtain more accurate forecasting results -ANN+NWP is suitable for the medium- and long-term prediction

Ensemble approach	-	- ANNs ensemble -LSTMs ensemble -Decision tree and SVM ensemble	- Ensemble models are usually more accurate than each of its member model - Ensemble method can increase the chance of eliminating a poor predictor

10.4 Appendix D (selection of the number of hidden neurons for ANN(PSO))

Table 10.4: Summary of the configurations used to determine the suitable number of hidden neurons.

Forecasting model	PSO trained feedforward ANN
Number of hidden layers	1
Training algorithm	PSO
Type of forecasting	2h-ahead wind speed forecasting
Maximum iteration T_{max}	T_{max} = 500
Input features (NREL: ID61118)	WS, WD, T, P, AD, TI
Data resolution and sample Size	Resolution: 10min, sample size: 150, 000
Acceleration constant $c_1,\ c_2$	$c_1 = 2, c_2 = 2$
Inertia weight ω	$\omega_{min} = 0.4, \omega_{max} = 0.9$
Particle population size	10
Variable parameters	Number of hidden neurons

Table 10.5: Simulation results of 2h-ahead wind speed forecasting of the ANN (PSO) with different numbers of hidden neurons

Model Config.	Training RMSE (m/s)	Testing RMSE (m/s)	Training time (s)
5 HN	1.85	1.82	155
10 HN	2.73	2.47	175
15 HN	5.28	5.09	197

As shown in Table 10.5, selecting five hidden neurons for the ANN(PSO) model can lead to a better result by having the shortest training time, and the lowest testing RMSE value. Different combinations of hidden neurons and PSO parameters can lead to different results. To focus on the investigation of the effect of the PSO parameters on the forecasting performance of the ANN (PSO) model and to keep a single variable for each experiment, five hidden neurons were used for all the simulations of the ANN(PSO) model. In further research, combined scenarios with different c_1, c_2, $\omega_{min}, \omega_{max}$, and population sizes will be considered.