Machine Learning for Wind Power Prediction

by

Yiqian Liu

Bachelor of Science, Shandong University, 2013

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

Master of Computer Science

In the Graduate Academic Unit of Faculty of Computer Science

Supervisor:	Huajie Zhang, Ph.D, Faculty of Computer Science
Examining Board:	Weichang Du, Ph.D, Faculty of Computer Science, Chair
	Bruce Spencer, Ph.D, Faculty of Computer Science
	Donglei Du, Ph.D, Faculty of Business Administration

This thesis is accepted

Dean of Graduate Studies

THE UNIVERSITY OF NEW BRUNSWICK

June, 2016

Abstract

Wind power prediction is of great importance in the utilization of renewable wind power. A lot of research has been done attempting to improve the accuracy of wind power predictions and has achieved desirable performance. However, there was no complete performance evaluation of machine learning methods. This thesis presented an extensive empirical study of machine learning methods for wind power predictions. Nine various models were considered in this study, which also included the application and evaluation of deep learning techniques. The experimental data consisted of seven datasets based on wind farms in Ontario, Canada. The results indicated that SVM, followed by ANN, had the best overall performance, and that k-NN method was suitable for longer ahead predictions. Despite the findings that deep learning failed to give improvement in basic predictions, it showed the potential for more abstract tasks, such as spatial correlation predictions.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

ANN	artificial neuron network
DNN	deep neuron network
k-NN	k-nearest neighbors
LR	linear regression
MAE	mean absolute error
MLP	multilayer perceptron
MW	megawatts
NMAE	normalized mean absolute error
NRMSE	normalized root mean squared error
NWP	numerical weather prediction
Pers.	persistence model
RBF	radial basis function
RBFN	RBF network
REP	reduced-error pruning
RMSE	root mean squared error
RNN	recurrent neural network
SVM	support vector machine

Chapter 1

Introduction

We only have one Earth but we need 1.6 Earth to support our activities [1]. We are using the natural resources in an unsustainable way. To give a concrete example, August 13 is the Earth Overshoot Day of 2015 [2], after which the resources consumed by human activities exceed what the Earth is able to regenerate in a year. This simply means after that day we are using the resources from the future. What is more, the Earth Overshoot Day has been becoming earlier every single year [2].

Unsustainable use gives rise to imbalanced ecology. Fossil fuels (e.g. coal, petroleum and natural gas) have played an important role in human history. They have always been the main energy source of human; industrialization cannot be realized without the consumption of fossil fuels; they are the foundation of modern transportation [4]. On the other hand, the combustion of

fossil fuels globally results in excessive amount of greenhouse gas that endangers the health of our environment. The abnormal raise of carbon dioxide level, global warming and acid rain can be all attributed to burning fossil fuels [4]. Serve air pollution in northern China recent years is another striking example of the negative impact of fossil fuels [5].

It is good to see that most of the countries in the world are aware of climate change and resources shortage. The most recent United Nations Climate Change Conference, namely COP 21, is held in Paris, France in December 2015 for two weeks. 195 countries in total attended this conference [3]. The main achievement of this conference is the Paris Agreement, which sets a strict goal of controlling global warming to be below 2 degree Celsius. The Paris Agreement also states the requirement of zero greenhouse gas emission during 2051 - 2100 [3]. This clearly shows that countries around the world need a new structure of energy consumption and that renewable energy is indispensable.

1.1 About Wind

Wind energy is one of the clean and renewable resources that can be utilized to support industrial activities. It is also one of the oldest energy used by human, such as in windmills and on sailboats.

Simply speaking, wind is moving air. The movement of air forms current that transfers heat from one place to another. Although we can feel wind everyday and think it is very close to us, the scale of air circulation is actually global [7]. This global circulation itself has important meaning to the Earth's ecosystem.

Wind has its own characteristics. One of the characteristics is that wind is intermittent. Sometimes there is almost no sign of the wind, while sometimes people can easily tell that it is windy just by hearing. Wind is affected by terrain a lot. Wind is usually stronger near the shores, because of the uneven air temperature between land and water. Higher wind speed around some valleys is the result of a similar condition: the heating difference, between mountain and valley in this case. The profiles of the wind also vary from season to season and stronger wind is observed more frequently during spring in North America [7].

1.2 Wind Power

Since wind is air in motion, it must contain energy within it. More formally, a moving object with mass has kinetic energy [9]. Although air is very light, its motion still produces energy. Intuitively, higher wind speed generates

more energy. To show this intuition is correct, a few simple equations are needed.

The first equation is about the kinetic energy, as given by

$$\textit{Kinetic Energy} = \frac{1}{2}mv^2, \tag{1.1}$$

where m and v are the mass and velocity of the moving object, respectively [7]. Then, let us assume that we only focus on a small period of time t. Knowing the fact that mass is a product of density(ρ) and volume(V), we get an equation for wind energy as follows

$$\begin{aligned} \textit{Wind Energy} &= \frac{1}{2} \cdot \rho V \cdot v^2 \\ &= \frac{1}{2}\rho \cdot Avt \cdot v^2 \\ &= \frac{1}{2}\rho A t v^3, \end{aligned} \tag{1.2}$$

where A denotes the area that the wind passes through [7]. To emphasize, given a period of time t this equation gives the kinetic energy of the wind, and the volume of the wind is defined by the product of the area of interest and the distance[1] the wind traveled during time t. To go one step further, dividing the equation by the time t gives the rate of wind energy, namely

[1]Since time t is short, wind speed v can be seen as a fixed value. So $\textit{Distance} = vt$ is an acceptable approximation.

wind power

$$P = \frac{1}{2}\rho A v^3. \tag{1.3}$$

Equation (1.3) shows that the intuition about wind speed is correct. And more precisely, wind power is proportional to the *cube* of wind speed. Besides, wind power is also dependent on the density of air and the area intercepting the wind.

One important point to note about Equation (1.3) is that this equation gives the total power within the wind. It is ideal that all such amount of power can be converted into other useful forms. In reality, however, only part of the power can be harvested.

To capture the energy in the wind, human has designed many tools, such as traditional sails and modern wind turbines. The most commonly seen wind turbine consists of a tower and a rotor with three blades. When the wind is strong enough, it rotates the rotor so that wind energy can be converted to electricity. The size of a wind turbine can be described by the diameter of its rotor, which ranges from only half meter to over 100 meters [7]. The rotor's size determines the wind turbine's swept area, which corresponds to the area A in Equation (1.3). A wind turbine's size sometimes also refers to its wind power capacity, usually given in kilowatts(kW) or megawatts(MW). Again as suggested by (1.3), the power capacity of a wind turbine is closely

related to its swept area.

In early years, wind turbines were mainly used to generate electricity at remote sites where central power grid has not yet reached [7]. As wind technology improves and the size of wind turbines increases, commercial wind farms emerge. Most wind farms are connected to the power network to supply electricity along with other types of power plants. In some places, wind energy is also used for heating and pumping water [7].

Wind power has become more influential since the early 1970s. For North America, the 1980s is the time when the wind power industry found its way to the market, especially in California [7]. Global wind power capacity has the highest growth rate comparing to other types of power generation in last few decades. It even grows more rapidly than experts' expectation. In a 2007 book [8], the author predicted the global wind capacity to reach 200,000 MW by 2013; the actual capacity in 2013 has past 300,000 MW [17]. As of 2014, global wind power generation accounts for 5% of overall electricity demand [16]. This percentage rate is the so-called *wind penetration rate*. In Denmark, as a country with the highest wind penetration rate, this percentage is 39.1 [16].

While currently most wind farms around the world are on the land, the future of wind energy industry is more likely dependent on the development of

offshore wind farms. Since there is little obstacle over the sea, the offshore wind speed is higher and has less variance. And the space limit does not exist for offshore wind turbines so that both the scale of wind farms and the size of wind turbines can be very large. For example, a single wind farm called Horns Rev in Denmark was able to generate 2% electricity demand of the country [21]. These advantages of offshore wind power show that it has great potential in the near future.

1.3 Wind Power Integration

One essential task that any power system needs to handle is to match the supply with the demand. This is a dynamic balance. For the electricity demand, an amount is estimated according to past patterns. This demand estimation will help the system operator to decide how much power to supply in order to match the demand or reduce waste [23]. Apparently, the demand varies over time, which requires the supply to be adjusted accordingly. The adjustment in power production is a much easier task for traditional fossil-fired power stations than for wind power ones. For thermal power stations, the output can be scheduled to follow the demand changes. But just as most other types of renewable energy, wind power has fluctuations that can be hardly controlled [8]. Unlike natural gas power stations, wind power is not dispatchable [18]. The maximal wind power production is mostly subject to

the wind condition at the moment. Therefore, it is necessary to take extra care to ensure that the power system is able to absorb the wind power integrated.

The key issue of wind power integration is related to the variance of wind power. One way to significantly reduce the variance is to aggregate the power generation from multiple wind farms because the wind conditions are less likely correlated between different geographical locations. Using energy storage is another solution to this issue as the surplus energy generated by the wind can be used later, such as during the period of peak demand. Last but not least, wind power prediction can be extremely helpful in that it gives a quantified wind power output for estimating overall supply.

1.4 Wind Power Predictions

The role of wind power prediction is becoming increasingly important while the wind penetration rate is constantly growing. Every power system has reasonable capability to adapt demand changes as the demand estimation has never been perfectly accurate. When the penetration rate is relatively low, power systems do not have to pay too much attention to the variance of wind power supply. This is because, for example, a drop in wind speed resulting in lower wind power supply is very similar to an increase of demand,

which will be within the control of the power system [8]. However, when the wind penetration reaches a certain level, wind power prediction becomes a necessary tool. Additionally, wind power prediction can help in avoiding imbalance charge [21] and improve the stability and efficiency of the power systems [8].

The general steps to predict wind power are as follows: wind speed is usually predicted first by an appropriate model; then the predicted wind speed is used to determine the expected wind power output for a specific wind farm; and the prediction result of a wind farm may be further used to forecast regional output [20]. Sometimes the first two steps are combined by some prediction models.

Wind power prediction can be categorized to different groups in mainly two ways. One way is to consider the time scale that a method predicts. Some methods predict wind power for next minute, while others predict several hours ahead. Wind power prediction methods can also be categorized according to whether they are physical or statistical.

There is actually no absolutely strict classification rule for wind power prediction methods with different time scales. Nevertheless, prediction methods are typically categorized to four groups: very short-term, short-term, medium-term and long-term prediction. Prediction methods with time scale shorter

than 6 hours are usually considered to be short-term and methods longer than 24 hours to be long-term; Very short-term predictions focus on wind power generation within next minute [19]. Prediction methods with different time scales have different application in practice. My work in this thesis falls into the range of short-term wind power prediction, which is usually related to power dispatch planning [19].

We can also categorize wind power prediction by the actual approach it takes. A physical approach involves building specific physical models of different wind farms. This approach takes into account the detailed information of wind turbines and wind farm terrain [20]. The other one is the statistical approach that includes time-series models and machine learning models. The statistical approach only uses historical data to build a model to predict future values of interest. In some cases, Numerical Weather Prediction (NWP) is used to provide various meteorological input, such as wind speed and wind direction [19]. There are also many hybrid methods available for wind power prediction.

1.5 Machine Learning

Machine learning is a field of computer science that focuses on improving the performance of the program by itself with experience. In machine learning,

the machine is not told how to solve the problem explicitly; rather, past experience is given to the machine as input and the output is typically a model that can solve future problems of the same kind. Machine learning can also be seen as a multidisciplinary field, which applies knowledge from artificial intelligence, statistics, neuroscience, etc. [11]. Nowadays, the application of machine learning is around us in everyday life. Speech recognition and online personalized advertisement are two of the most common ones. The computer program AlphaGo that beats the Go world champion is another example of the power of machine learning.

The complete procedure of machine learning includes several steps. First, past experience is usually gathered for the training in the later stage. Then, the form of an abstract target function is determined, which describes the relation between existing input and desired output. After that, a machine learning model is selected to approximate the target function. In the end, an appropriate algorithm is used to build the model from the training examples [11].

In the case of wind power prediction, the target function usually maps from weather data to future wind power. And the relevant machine learning models will be discussed in next section.

1.5.1 Machine Learning Models

Among simpler machine learning models, decision trees and k-nearest neighbors (k-NN) are two effective methods used in many application. As its name shows, a decision tree is a model with a tree structure that can be built by training examples. At each node of the tree, a decision is made depending on one of the input attributes. After several decisions one of the leaves will be reached, which corresponds to the target function output. The k-NN method can be explained by its name, too. Given a set of training examples, the k-NN produces the result by calculating the average of k training examples that are closest to the test sample. The distance is typically Euclidean distance and it can have a more general definition.

In the scope of wind power prediction, two commonly used machine learning models are artificial neural network (ANN) and support vector machine (SVM) [20]. An ANN consists of an input layer, some hidden layers and an output layer. Each layer is formed by a number of neurons. It is the neuron that has the capability to learn the relationship between input and output. Input data passes through hidden layers to produce prediction in the output layer. Many variants of the ANN are used in practice, among which are multilayer perceptron (MLP) and recurrent ANN [20]. SVM is a relatively new algorithm in the field of machine learning and it is a kernel-based learning method [13]. The mechanism behind the SVM is rather complicated. To put it in a simple way, an SVM tries to find a divider called hyperplane with

maximal margins to both sets of training examples. The resulting hyperplane is then used to do the prediction.

There are other machine learning models that are applied to predict wind power such as fuzzy systems, etc. [20]. Hybrid approaches that combine different machine learning techniques also exist [19].

1.5.2 Deep Learning

Deep learning consists of methods that can effectively and efficiently train deep neural networks. Comparing to regular ANNs, a deep neural network (DNN) has multiple hidden layers which makes it much harder to train [14]. In return, the model complexity is much higher and so is its learning ability. With enhanced learning ability, a deep neural network is able to learn more abstract conception through its layers of neurons.

Convolutional neural network belongs to the family of deep neural networks. Because of its special ability to discover spatial relation, it has very good performance in image recognition [14]. Recurrent neural network is another model that works well with deep learning. Because it has the property of saving earlier states of neurons, it is a suitable model for temporal predictions [23].

Because of the increased complexity of deep neural networks, the problem

of overfitting is more prevalent in deep learning. Overfitting is an undesired situation that may occur during the stage of training, where the model being trained fits so well to the training data that it becomes lack of generalization. If overfitting occurs, the model will perform almost perfectly on training examples and has little prediction ability. Overfitted models behave as if they have memorized the training data rather than learned from it. To tackle the problem of overfitting in deep learning, many methods including dataset expansion and dropout technique are developed [14].

The motivation for this study mainly comes from two considerations. For one thing, in the research of machine learning methods for the wind power prediction, there is a blank for an extensive empirical study. The literature review, which will be given in Chapter 2, suggests more clearly the necessity of such study. For another, deep learning is a young subfield of machine learning and it has been proven to be powerful in various problem solving settings. Since only very limited work has been done with respect to deep learning in the wind power prediction, it is of great interest to us to see how well it can perform in this field.

The remaining chapters of this thesis are structured as follows: Chapter 2 reviews related work in wind power prediction; Chapter 3 describes the data and the experiment procedure in this work; Chapter 4 presents and discusses the results of the experiment; and the last chapter concludes this thesis.

Chapter 2

Literature Review

There is lots of work that has been done in wind speed and power predictions since the year of 2000. In recent work, artificial intelligence techniques and machine learning models have been the focuses of most researchers [26]. As shown in last chapter, wind power predictions can be grouped by their time scales into four categories: very short-term (few seconds - 30 min), short-term (30 min - 6 h), medium-term (6 h - 1 day) and long-term (longer than 1 day) [24]. In this chapter, related work using machine learning methods will be reviewed for each group, with emphasis on the short-term predictions.

2.1 Very Short-term Wind Power Predictions

The time scale of very short-term prediction usually ranges from 1 minute to half an hour, but it also can be as short as few seconds. The prediction

results can be used for wind turbine control [24], real-time grid operation [25], etc.

In [34], a Kalman filter was used on the input data before feeding it into a support vector machine (SVM), which predicted wind speed with an ahead time of 10 minutes. Reference [35] presented a hybrid wind power model based on wavelet transformation, particle swarm optimization and fuzzy inference system that had a time scale of 15 minutes. This method also had a relatively short computation time. Potter et al. [36] proposed an adaptive neuro fuzzy inference system model with multiple layers that uses Gaussian activation functions, which reduced the mean absolute error (MAE) of 2.5-minute-ahead prediction to 4%. In [37], a three-layered recurrent neural network (RNN) model combined with moving average technique was used to predict wind speed in 15 minutes. In 2014, a novel ν-support vector regression model with augmented Lagrange method was used to forecast wind power by taking only wind speed as input [38].

2.2 Short-term Wind Power Predictions

Wind power predictions with time scales between 30 minutes and 6 hours belong to short-term prediction. Methods in this category mainly serve to plan load dispatch in order to make wind power economically more competitive [25].

In [43], a fuzzy model was used to predict wind farm power generation. The main input was wind speed and direction. This model also took advantage of spatial correlation which uses not only one wind farm data, but its nearby sites' information as well. Monfared [44] proposed a hybrid model combining fuzzy logic and ANN for wind speed forecasting using historical wind speed data from 2005. The ANN was trained by the traditional back-propagation algorithm and used as a 30-minute-ahead predictor.

Ramirez-Rosado et al. [39] introduced a multilayer perceptron (MLP) network using Kalman filter as data preprocessing. This two-layered neural network had the prediction interval of 30 minutes, the input of which contained wind speed, historical wind power and wind farm's real-time data. Eventually 14.1% root mean squared error (RMSE) was achieved. In Reference [45], a back propagation ANN designed for short-term wind power prediction was described. The number of neurons in the hidden layer was determined by trial and error, which showed that too many neurons led to reduced accuracy. Sideratos & Hatziargyriou [41] proposed an optimized radial basis function (RBF) neural network for probabilistic wind power forecasting. This model's time horizon was 1 hour and it took into consideration wind power, wind speed, wind direction and terrain effect when predicting wind power. The improvement of this model comparing to persistence model was over 50%.

In Reference [40], a method combining wavelet transformation and SVM was proposed to predict wind speed 30 minutes ahead of time. This SVM had a radial basis function kernel and used temperature and wind speed as input. Fan et al. [42] divided wind power prediction into two stages: firstly a Bayesian method was used to cluster the input data, including wind direction, wind gust speed, etc.; then a SVM used the processed data for 10 minutes ahead power prediction. This method combined unsupervised and supervised learning and its RMSE improvement over persistence model was around 37%. In [32], an epsilon-SVM was used for short-term prediction. It was discovered that too large a training set can result in undesired learning effect of the SVM and that the prediction became unstable when wind speed changed abruptly. Mao et al. [30] described a new process for wind power prediction. In addition to the steps in the traditional process, an error evaluation was added before the final prediction. ANN and SVM were used to verify the effectiveness of this process and results showed obvious improvements for both cases.

In 2009, Kusiak et al. [28] conducted a experiment comparing several data mining approaches for short-term wind power prediction, including regression trees, MLP and SVM. The input data, mainly wind speed, was collected for the entire month of January, 2006. The feature selection was done by a boosting tree algorithm. The results indicated that the MLP model per-

formed the best among others for hourly predictions.

2.3 Medium- and Long-term Wind Power Predictions

Predictions between 6 hours and 1 day and longer than 1 day are classified as medium-term and long-term, respectively. These predictions can be helpful in the decisions of unit commitment, reserve requirement, maintenance, etc. [25]. It is worth noting that research done for long-term prediction is less than that for shorter time horizons.

In Reference [47], an advanced structure of prediction model with multiple RBF networks was presented. It was showed that both MAE and RMSE were improved by 40%, even for forecasts that are longer than 10 hours. Togelou el al. [46] studied an unusual situation where historical data for training is limited. Complex terrain along with regular wind data was used as input for two RBF network models, of which the prediction interval was 6 hours. Rzui & Eremia [31] compared the performance of a fuzzy inference system and an ANN for medium-term wind power prediction using hourly mean wind power. The results indicated that the fuzzy inference system is slightly better than the ANN and large training set was needed for both. Tao et al. [48] introduced a deep belief network with restricted Boltzmann machine to learn historical wind speed patterns. Regular ANN and SVM

were compared to the proposed model and it was shown that, especially for longer ahead predictions, the new model outperformed the others. Amjady et al. [49] used the particle swarm optimization technique for a hybrid neural network model. This model took into account the air humidity and made prediction 1 to 2 days ahead of time. Finally 1.59 normalized MAE was achieved. In [50], a hybrid model based on ANN was used for 2-day-ahead prediction, taking only wind speed as input. Barbounis et al. [51] presented three RNNs for long-term wind speed and power prediction. Two optimized online learning algorithm were applied in this study. Air temperature and pressure were considered as auxiliary input when making prediction.

From this review, a few issues was found, being listed as follows:

- Most datasets used in different studies were not large enough to be representative. For example, some datasets only had historical data for one month, which is unable to demonstrate the seasonal characteristics of the wind.
- The evaluation metrics were not uniform and the comparison studies were not complete.
- Majority of the models were trained by data containing only wind speed and wind power; more attributes including temperature and humidity should be considered.
- Deep learning for wind power predictions is in its early stage of development.

As stated in Chapter 1, the existence of these issues is part of the motivation of this empirical study. We will attempt to address all of the above issues in following chapters.

Chapter 3

Methodology

Considering the incompleteness of current wind power prediction in the field of machine learning, We collected multiple datasets and designed different types of experiments involving various machine learning models, which focused on predicting short-term wind power. Since little research of deep learning has been done for wind power prediction, a deep neural network model is included and compared to other models. The experimental data and models will be described first in this chapter, followed by experiment detail and evaluation approaches.

3.1 Data Description

To reduce the occasionality of the experiments, seven datasets has been collected and prepared. These final datasets were composed of publicly available

Table 3.1: Selected Wind Farms

Wind Farm	Capacity (MW)	Location
Erieau	99	Erieau
Dillon	78	South Buxton
Spence	99	Ridgetown
Kingsbridge	40	Kingsbridge
Wolfe Island	198	Wolfe Island
Port Alma	101	Port Alma
Part Alma 2	101	Port Alma

information from two separated data sources – wind power data and meteorological data.

3.1.1 Wind Power Data

The first part of the data containing wind power generation information was collected from the Ontario's Independent Electricity System Operator (IESO) official website [52]. This dataset contains a number of wind farms' hourly power generation data that has time range from March, 2006 to November, 2015.

Seven wind farms were selected from the entire wind power dataset. Table 3.1 shows installed capacity and location information of the seven wind farms

Table 3.2: Meteorological Stations

Station Name	Wind Farm	Distance (km)	Coordinates
Erieau	Erieau	6	42.25°N, 81.90°W
	Part Alma 2	18	
Chatham Kent	Dillon	8	42.31°N, 82.08°W
	Port Alma	17	
Ridgetown RCS	Spence	9	42.45°N, 81.88°W
Goderich	Kingsbridge	12	43.77°N, 81.72°W
Kingston A	Wolfe Island	15	44.23°N, 76.60°W

which are all located in Ontario, Canada.

3.1.2 Meteorological Data

Hourly meteorological data corresponding to the above seven wind farms was then collected from the Government of Canada website [53]. The attributes in the data include temperature, pressure, humidity, wind speed, wind direction, etc. These public datasets are available in the form of meteorological stations and totally five stations closest to each of the seven wind farms were selected. Table 3.2 lists the selected meteorological stations and their details including the distances to corresponding wind farms and geographical coordinates.

Since both wind power data and meteorological data are hourly available,

Table 3.3: Experimental Datasets

No.	Dataset	Size	First Record
1	Erieau - Erieau	21933	May 1, 2013
2	Dillon - Chatham Kent	13173	May 1, 2014
3	Spence - Ridgetown RCS	28485	Aug 1, 2012
4	Kingsbridge - Goderich	16053	Jan 1, 2014
5	Wolfe Island - Kingston A	8760	Nov 1, 2014
6	Port Alma - Chatham Kent	13173	May 1, 2014
7	Part Alma 2 - Erieau	29949	Jun 1, 2012

the one-to-one mapping is automatically established between them. The final seven experimental datasets were obtained by properly joining the wind power datasets and the meteorological datasets, respectively. Their information is described in Table 3.3. The last records of the seven datasets have the same time stamp of October 31, 2015 and different first records are noted in the table.

The datasets listed in Table 3.3 are the base datasets for this study. Some derivative datasets are used in some experiments and their details will be mentioned later along with the experiments.

3.2 Models Description

A wide range of various models were selected in this study. Persistence model and linear regression are two non-machine learning models that are intended to provide some performance reference. The rest of the selected models include k-NN, two regression trees, three ANN and its variants and SVM.

3.2.1 Persistence Model

The persistence model is the most commonly used reference model in wind power prediction [21]. As the simplest model at the same time, persistence model uses merely the last one historical data point to predict the next value. So, at time t, the persistence model for time $t + \Delta t$ can be given by

$$\hat{P}_{t+\Delta t} = P_t, \tag{3.1}$$

where P denotes actual known wind power (or more generally, target attribute value) while $\hat{P}$ stands for predicted value.

Time step Δt may vary for different applications from a minute up to a day or longer. The time step used throughout this thesis will be one hour. For the persistence model, the shorter the time step, the more accurate prediction it can make. For sufficiently short time step, the performance of the persistence

model can be difficult to beat [21].

3.2.2 Linear Regression

Linear regression is a basic model in statistics. It is simple but sometimes very effective. Given a historical dataset $\{x_i, P_i\}_{i=1}^n$, a linear regression model wants to find a linear relationship

$$\hat{P} = \hat{a}x + \hat{b} \tag{3.2}$$

such that the sum of the difference between each point $x_i, i = 1, 2, ..., n$ and the 'line' $\hat{P}$ is minimized. For estimating parameter $\hat{a}$ and $\hat{b}$, ordinary least squares is the most common method. In the case of simple linear regression, these two parameters can be estimated by [10]

$$\hat{a} = \frac{\sum x_i P_i - \frac{1}{n}\sum x_i \sum P_i}{\sum x_i^2 - \frac{1}{n}\left(\sum x_i\right)^2}, \qquad \hat{b} = \bar{P} - \hat{a}\bar{x},$$

where $\bar{P}$ and $\bar{x}$ denotes the mean of x_i and $P_i, i = 1, 2, ..., n$, respectively.

3.2.3 k-Nearest Neighbors

The k-NN model is one of the instance-based learning methods. It considers all instances (or training examples) as being in a multidimensional space with the dimension equal to the number of attributes of an instance. The Euclidean distance is usually used to choose k neighbors that are closest to

the query instance. The predicted value is then determined based on these k neighbors [11]. For example, assuming there is a training set

$$\{x_i, y_i, P_i\}_{i=1}^{n}$$

where x and y denote two attributes of a instance, the distance between two instances $\mathbf{e_i} = (x_i, y_i),\ \mathbf{e_j} = (x_j, y_j)$ is defined as

$$d(\mathbf{e_i}, \mathbf{e_j}) \equiv \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2}. \tag{3.3}$$

When a query instance $\mathbf{e_q}$ is given for prediction, the k nearest instances denoted by $\mathbf{e_1}, \mathbf{e_2}, ..., \mathbf{e_k}$ can be found according to the above definition. Finally, the predicted value $\hat{P}_q$ is determined by

$$\hat{P}_q = \frac{\sum_{i=1}^{k} P_i}{k}. \tag{3.4}$$

Note that P_i is the target attribute value of a training instance $\mathbf{e_i}$.

In this study, the model parameter k was set to 10 and one further improvement has been made by adding weights to the k nearest neighbors according to the reverse of the distance, i.e. $w_i = 1/d(\mathbf{e_i}, \mathbf{e_q}),\ i = 1, 2, ..., k$. Therefore, nearer neighbors has greater weights and the refined model is given by

$$\hat{P}_q = \frac{\sum_{i=1}^{k} w_i P_i}{\sum_{i=1}^{k} w_i}. \tag{3.5}$$

3.2.4 REP Trees

REP tree stands for reduced-error pruning tree [11]. This type of regression tree is firstly constructed without paying attention to the global error. At a leaf node l with a subset of training instances, a local model is built which simply takes the mean of the target attribute values of these instances, as given by

$$\hat{P}_l = \frac{1}{\|E(l)\|} \sum_{\substack{i \\ \forall \mathbf{e_i} \in E(l)}} P_i, \tag{3.6}$$

where $E(l)$ is the set of the instances locally associated with leaf l and $\|E(l)\|$ denotes the size of the set.

After the tree is built, the possibility of pruning is examined at each node starting from the leaves. The default error metric used for REP regression tree is the sum of squared error

$$S = \sum_{l \in L} \sum_{\substack{i \\ \forall \mathbf{e_i} \in E(l)}} (\hat{P}_l - P_i)^2, \tag{3.7}$$

where L is the set of the leaf nodes of this REP tree and $\hat{P}_l$ is given by Equation (3.6). The subtree will be pruned and replaced by a leaf if doing so does not increase the global error S. The value of S is also used to stop further partitions during the construction of the tree in order to address the issue of overfitting [12].

3.2.5 M5P Trees

M5P tree is a combination of decision trees and linear regression. As with REP trees, an M5P tree is built recursively. At each tree node, the algorithm checks if it is the base case; it implies a base case when either the number of training examples associated with this node or their standard deviation is less than some threshold. A linear regression model is then built based on the remaining instances. For recursive cases, the training examples are split at this node based on the outcome of a test for one of the attributes [33].

To reduce overfitting, pruning technique was used and the minimal number of instances at a leaf was set to 4.

3.2.6 Multilayer Perceptron

MLP belongs to the family of ANNs and its structure is shown in Figure 3.1. As its name suggests, an MLP consists of multiple layers of perceptrons – a special kind of artificial neuron, which is the smallest computation unit in an ANN.

Let $x_1, x_2, ..., x_m$ be m real-valued input variables of a perceptron. The output is determined by the linear combination of the input variables, as given by $z = w_0 + \sum_{i=1}^{m} w_i x_i$, where w_i, $i = 0, 1, ..., m$, are the weights to be learned. Further, the output, also known as the activation, of the percep-

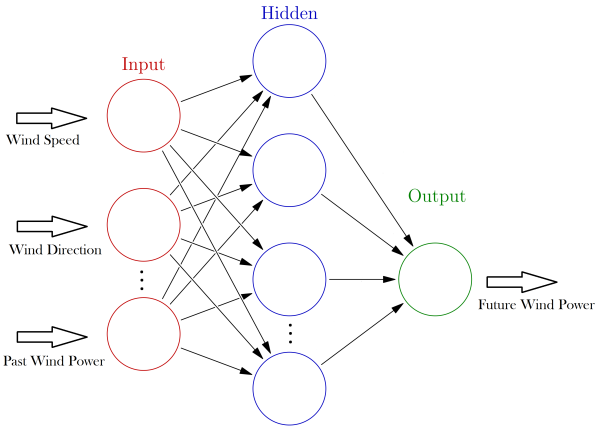


Figure 3.1: Artificial Neural Network [6]

tron $o(z) = \text{sgn}(z)$ [1]. Alternatively, the activation function can be a sigmoid function

$$o(z) = \sigma(z) \equiv \frac{1}{1 + e^{-z}}, \tag{3.8}$$

which yields continuous values between 0 and 1 [11]. Then, the output $o(z)$ of a neuron feeds forwards as an input variable to other neurons in the next layer until the output layer is reached.

The training task for an MLP is to find out the proper values of the weights w for all neurons so that the prediction error is ideally minimized. This is done by an iterative approach that considers the difference between the output

[1] sign function $\text{sgn}(z) = \begin{cases} 1, & \text{if } z > 0 \\ -1, & \text{otherwise} \end{cases}$

and target value, which is quantified by a quadratic cost function

$$C(w) = \frac{1}{2n} \sum_{i=1}^{n} (\hat{P}_i - P_i)^2, \tag{3.9}$$

where w represents all weights in this MLP and n is the number of training examples as in earlier sections.

To reduce the cost in (3.9), teh gradient descent approach is applied to update the weights in each iteration. For each weight w_k, the updating rule is [14]

$$w_k \leftarrow w_k - \eta \frac{\partial C}{\partial w_k} \tag{3.10}$$

where η denotes learning rate. The learning rate η has to be an appropriate value to avoid either slow learning or oscillation around the optimal value. In order to compute $\frac{\partial C}{\partial w_k}$, the backpropagation algorithm is used. The number of perceptrons in the hidden layer was set to 10 and all attributes were normalized during the training.

3.2.7 Radial Basis Function Networks

An RBF network is one kind of ANNs that shares some similarities with the k-NN model. The RBF network is named by the fact that the activation function of its neurons is a radial basis function ϕ whose value is only dependent on the distance between the input and a predefined 'center'. Generally,

given m centers $\{\mathbf{e_i}\}_{i=1}^{m}$, the approximation that an RBF network can yield for $\mathbf{e_q}$ is expressed as [11]

$$\hat{P}_q = w_0 + \sum_{i=1}^{m} w_i \phi(d(\mathbf{e_i}, \mathbf{e_q})), \tag{3.11}$$

where d typically gives the Euclidean distance as in Equation (3.3) and ϕ is usually the Gaussian function, i.e. $\phi(d) = \exp(-\frac{d^2}{2\sigma^2})$. Equation (3.11) implies that the RBF network has an very similar structure to an MLP with a single neuron in the output layer.

Training an RBF network is usually divided to two stages. The first stage involves calculating the centers $\{\mathbf{e_i}\}_{i=1}^{m}$. This is done by applying a k-mean clustering algorithm on the training examples. The number of centers, m, was set to 100, so there were 100 clusters randomly initialized. At the second stage, the weights $\{w_i\}_{i=1}^{m}$ are trained while the radial basis functions are fixed.

3.2.8 Support Vector Machines

At first sight, an SVM appears to be only able to represent a linear relationship between input and output because it looks for support vectors to determine a linear hyperplane. However, original input data can be transformed to a more complex form often with higher dimensions. In this new feature space, the linear relationship can be found which can be described

by

$$\hat{P}_q = w_0 + \sum_{i=1}^{m} w_i \phi_i(\mathbf{e_q}) \tag{3.12}$$

where $\{\phi_i\}_{i=1}^{m}$ is a set of non-linear mapping to the new feature space. When transferred back to the original feature space, the non-linear relationship between input $\mathbf{e_q}$ and prediction $\hat{P}_q$ is obtained accordingly [13].

In this study, the ϵ-SVM was used. The ϵ is a parameter that indicates the width of the regression band inside which the loss is considered zero. The loss is represented by slack variables ξ_i or $\xi_i^*,\ i = 1, 2, ..., n$, depending on the directions of deviation. Knowing the fact that smaller weights usually lead to a simpler model, training the ϵ-SVM is equivalent to

$$\begin{aligned} \text{Min} \quad & \frac{1}{2}\sum_{i=1}^{m} w_i^2 + C\sum_{i=1}^{n}(\xi_i + \xi_i^*) \\ \text{s.t.} \quad & P_i - \hat{P}_i \leq \epsilon + \xi_i,\ i = 1, 2, ..., n \\ & \hat{P}_i - P_i \leq \epsilon + \xi_i^*,\ i = 1, 2, ..., n \\ & \xi_i, \xi_i^* \geq 0,\ i = 1, 2, ..., n \end{aligned} \tag{3.13}$$

where C indicates model complexity. The above quadratic programming problem can be solved by solving its dual problem and the solution is given by [13]

$$\hat{P}_q = \sum_{i=1}^{n_{SV}} (\alpha_i - \alpha_i^*) K(\mathbf{e_i}, \mathbf{e_q}), \qquad \text{s.t.} \quad 0 \leq \alpha_i, \alpha_i^* \leq C, \tag{3.14}$$

where n_{SV} is the number of support vectors and the kernel $K(\mathbf{e_i}, \mathbf{e_q}) = \sum_{j=1}^{m} \phi_j(\mathbf{e_i})\phi_j(\mathbf{e_q})$. In the experiments, the value of parameter C was 1000 and ϵ was set to 1.0. The specific kernel applied was Gaussian RBF as given by

$$K(\mathbf{e_i}, \mathbf{e_j}) = \exp\left(-\frac{d^2(\mathbf{e_i}, \mathbf{e_j})}{\sigma^2}\right). \tag{3.15}$$

3.2.9 Deep Neural Networks

The ANN is one of the machine models that take effort to train; in the case of DNN that has multiple hidden layers, greater challenge is faced in the learning process. Nevertheless, the gain is the ability to learn more complex and abstract relationships.

In this study, a DNN with the regular structure was used, consisting of rectified linear units whose activation function is

$$o(z) = \max(0, z). \tag{3.16}$$

In comparison to the sigmoid activation function, the rectifier given by Equation (3.16) will not suffer neuron saturation at $\pm\infty$ which impairs the learning efficiency.

As a model with higher complexity, the DNN tends to overfit and avoiding this issue is of great importance. One of the techniques for reducing overfit-

ting used in the experiments is dropout. Dropout is a randomized approach which at the start of each training iteration randomly selects a portion of neurons to be turned off. The deactivated neurons will be temporarily ignored in this iteration. Then, the weights of the modified and simplified network are updated and the deactivated neurons are resumed before the next iteration. This technique simulates the effect of voting by multiple networks [14]. Another regularization technique called weight decay was also used in this study. Sharing the same idea as in (3.13), this technique has the bias in favor of smaller weights. With weight decay, the original cost function (3.9), noted by C_0 here, gained an extra term as shown by

$$C = C_0 + \frac{\lambda}{2n} \sum_w w^2, \tag{3.17}$$

where n is the number of training examples as before and λ is the weight penalty parameter that was set to 1.0e$-$9. The default DNN model in this study had three hidden layers with 100 neurons in each layer.

3.3 Experiment Procedure

3.3.1 Auxiliary Platform

In the course of experiments, Weka API was used as an auxiliary tool. Weka is open-source software developed by University of Waikato and it is written in Java [54]. Weka provides many useful tools for machine learning and data

mining. The specific version of Weka API used in this study was 3.7.13.

The use of Weka API greatly improved the efficiency of the experiments. As a result, various experiment settings were able to be realized in this study. The uniformity of the experiment outcomes provided a convenience for the comparison and analysis, which in turn improved the consistency of different experiments.

3.3.2 Feature Selection & Parameter Tuning

The datasets' feature selection and models' parameter tuning were mostly achieved by the try-and-error method. In some appropriate cases, feature selection used the experience from other papers.

After applying the try-and-error method, nine attributes were kept in the final experimental datasets, which included Year, Month, Temperature, Dew Point, Relative Humidity, Wind Direction, Wind Speed, Station Pressure and Wind Power. All of these attributes contained numeric values. According to the wind power formula shown in Equation (1.3), an artificial attribute called Wind Speed Cube that had the value of cube of the Wind Speed attribute was also considered for some cases. Based on the experience in [28], the Wind Power attributes included historical data up to three hours back, namely P_{t-1}, P_{t-2} and P_{t-3}.

The parameters for all models used in this study were tuned using the same dataset that was a subset of the Erieau - Erieau dataset. Overly tuning the parameters of a model may lead to a type of overfitting that is different from the one appears during the training process [14]. To avoid this type of more general overfitting, most parameters were adjusted to integer or half orders of magnitude, e.g. 5.0e−9, 1.0e−4, 0.5, 1000, etc.

3.3.3 Experiment Setups

In order to obtain a preferably complete analysis of machine learning for wind power prediction, various experiments have been designed and conducted using the prepared datasets and models.

All of the seven datasets containing wind power information from different wind farms, as listed in Table 3.3, were utilized by all selected models, which formed the core component of this study. The seven wind farms are all geographically separated and therefore have their own wind power characteristics. By building and training models given these wind farms' data, the generalization abilities of the experimental models could be illustrated and compared. For this experiment, all datasets were modified to only contains one entire year data from November, 2014 to October, 2015. The target attribute was next hour wind power.

Different prediction time horizons, from one hour ahead to six hours ahead,

were considered in this study. Each experimental model was trained from scratch for six times corresponding to the six time horizons. As a result, the further ahead predictions did not rely on former predicted values. The training and testing data used for this experiment was based on the Erieau - Erieau dataset. The target attribute was slightly adjusted for each case.

Larger datasets tend to alleviate the issues of overfitting for non-linear models [23]. To verify this against different selected models, the Spence - Ridgetown RCS dataset were further used to conduct the dataset size experiment. All models were built based on these datasets among which the only difference was the number of their instances.

Spatial correlations between wind farms were also briefly examined by a experiment, where the wind power of Erieau and Port Alma wind farm was predicted by nearby Spence and Erieau wind farm, respectively.

3.3.4 Deep Learning for Wind Power Predictions

As an attempt to investigate and apply deep learning techniques in the field of wind power prediction, several designated experiments were included in this study.

To keep the experiment environment in line with that of preceding ones, the truncated datasets that contain one year data were firstly used. The per-

formance of DNNs with different numbers of neurons were examined in two ways. Three DNNs that had three hidden layers with 50, 100 and 150 neurons were trained and observed, respectively. Next, by keeping the number of neurons in each hidden layer constant at 100, four DNNs with one to four hidden layers were investigated. After that, the total number of neurons were fixed at 300 for two experiments focusing on the structure of DNNs. The first experiment was about the arrangement of hidden layers, which consisted of one 200-neuron hidden layer and two 50-neuron ones. The 200-neuron layer was put to be the first, second and third hidden layer in turn and the DNN were trained for each case. Then, four more DNNs were built which had one hidden layer with 300 neurons, two layers with 150 neurons, three with 100 and four with 75, respectively. Default DNNs that had different dropout rates were also tested.

Since larger datasets are in favor of deep learning, all of the above experiments were conducted again using the complete Spence - Ridgetown RCS dataset that has historical wind power records from August, 2012 to October 2015. The aforementioned artificial attribute Wind Speed Cube was also added into this dataset.

3.4 Evaluation

Cross-validation technique was adopted to evaluate the experimental results. A k-fold cross-validation first evenly divides the entire dataset to k subsets. Each of the k subsets is to be selected as a validation set in turn. The remaining part of the dataset is used to train models while the validation set functions as a testing set. After k repetitions, the test results are averaged and reported [11]. In this study, k was set to 10. The 10-fold cross-validation was applied to train and test each model for 10 times – each time with a different random seed for partitioning the datasets.

Mean absolute error (MAE) and root mean squared error (RMSE) were used as the evaluation metrics for wind power predictions. As before, let $\hat{P}$ and P denote the predicted value and the actual value, respectively. These two evaluation metrics can be given as follows:

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}|\hat{P}_i - P_i| \tag{3.18}$$

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{P}_i - P_i)^2} \tag{3.19}$$

where n denotes the size of testing set. As shown in the above equations, the MAE is related to the prediction error's first moment while the RMSE is associated with the second moment (e.g. variance of the error). Therefore, RMSE will be affected more by larger errors [27].

Since the wind farms have different capacities, errors for smaller wind farms will definitely be less than that for larger ones. To overcome this bias and have a easier comparison between wind farms, normalized MAE and RMSE were also used, which are given by

$$\begin{aligned} \text{NMAE} &= \frac{1}{n}\sum_{i=1}^{n}\frac{|\hat{P}_i - P_i|}{M} \\ &= \frac{1}{nM}\sum_{i=1}^{n}|\hat{P}_i - P_i|, \end{aligned} \tag{3.20}$$

and

$$\begin{aligned} \text{NRMSE} &= \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\frac{\hat{P}_i - P_i}{M})^2} \\ &= \frac{1}{M}\sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{P}_i - P_i)^2}, \end{aligned} \tag{3.21}$$

where M represents the installed capacities of the wind farms, as given in Table 3.1 on page 23.

In order to give a more direct comparison between the performance of different models, an improvement score was used in the evaluation. This score is in percentage and illustrates the skill of a machine learning model by providing the gain relative to a reference model. In our case, the persistence model was used as reference and RMSE value was selected for comparison. Thus, this

improvement score is given by

$$\text{Impr.} = \frac{\text{RMSE}_{Pers.} - \text{RMSE}}{\text{RMSE}_{Pers.}} \times 100\%. \tag{3.22}$$

Paired t-test was also adopted in the evaluation process to test for statistical significance. The t-test requires that the sample follow a normal distribution [10], which the MAE satisfies. As the random seeds used for the cross-validation were the same, the MAEs obtained from it followed the same normal distribution. Therefore, the MAEs of two different models were treated as paired examples.

Chapter 4

Results & Discussion

In this chapter, the results of different experiments will be shown and discussed. The various experiment setups have been described in Section 3.3.3 and Section 3.3.4 and the results and discussion will be presented in the same order. Most tables shown in this chapter are tables that only contain aggregate results from all experimental datasets. As the original results of the experiments can take up too much space, they are available in the appendix.

4.1 Different Datasets

The performance of various models across different datasets will be shown in this section, which is the basic result of this study.

The original MAE for all selected models and for all available datasets can

Table 4.1: Original MAE for Different Wind Farms (MW)

	Erieau	Dillon	Spence	Kings	Wolfe	Alma	Alma2	*Total*
Persistence	6.0621	4.9211	7.7494	2.1740	11.4417	7.2857	7.7258	47.3598
Linear Regression	6.0093	4.8723	8.0046	2.2179	11.4400	7.4403	8.0330	48.0175
k-NN	6.7178	5.5645	8.4774	2.4208	12.5225	8.1581	8.5400	52.4012
REP Tree	6.4147	5.2307	8.2926	2.3503	12.1669	7.8062	8.2935	50.5548
M5P Tree	5.9862	4.8741	7.9321	2.2157	11.4169	7.4306	7.9867	47.8423
MLP	5.9629	4.8609	7.9067	2.2031	11.2198	7.4445	7.9360	47.5339
RBF Network	6.3288	5.1212	7.9969	2.3036	12.1910	7.4828	8.0362	49.4604
SVM	5.7302	4.6196	7.4564	2.1450	10.6792	6.9854	7.4743	**45.0900**
DNN	6.0413	4.8594	7.8899	2.2123	11.3075	7.4189	8.0209	47.7501

Table 4.2: Aggregate MAE for Different Wind Farms

	MAE (MW)	**NMAE** (%)			
	sum	min.	max.	μ	σ
Pers.	47.3598	5.44	7.83	6.62	0.87
LR	48.0175	5.54	8.09	6.72	0.98
k-NN	52.4012	6.05	8.56	7.34	0.95
REP	50.5548	5.88	8.38	7.07	0.94
M5P	47.8423	5.54	8.01	6.70	0.96
MLP	47.5339	5.51	7.99	6.66	0.97
RBFN	49.4604	5.76	8.08	6.90	**0.84**
SVM	**45.0900**	5.36	7.53	**6.33**	0.86
DNN	47.7501	5.53	7.97	6.69	0.96

be found in Table 4.1. Each cell in this table gives the averaged MAE of 10 cross-validations with different random seeds. The middle columns show results based on different datasets and the last column is the sum of the MAEs. Since the MAE is a prediction error metric, smaller values are preferable. The bold font shows that SVMs have the best overall performance with a total 45.09 MAE, followed by Persistence, MLP and DNNs. The SVM model has an apparent advantage over other models in terms of MAE. To eliminate the bias between the columns caused by different wind farm capacities, MAEs are further normalized by dividing corresponding capacities. The tables for normalized errors can be found in the appendix.

The aggregate MAE and NMAE are shown in Table 4.2. It can be seen that

Table 4.3: Aggregate RMSE for Different Wind Farms

	RMSE (MW)	**NRMSE** (%)		**Impr.** (%)	
	sum	μ	σ	μ	σ
Pers.	79.9891	11.13	1.70	-	-
LR	75.8118	10.56	1.64	5.22	0.60
k-NN	82.1926	11.46	1.50	-3.30	2.73
REP	79.6958	11.10	1.55	0.14	1.30
M5P	75.6148	10.53	1.60	5.42	0.56
MLP	**75.0579**	**10.47**	1.61	**6.03**	0.79
RBFN	76.9435	10.69	1.49	3.80	1.50
SVM	75.4896	10.51	1.69	5.71	0.92
DNN	75.1301	10.48	1.60	5.90	0.76

the NMAE also indicates that SVMs have the best prediction accuracy while RBF networks have the lowest standard deviation. However, the best predictor in terms of RMSE is the MLP, as shown in Table 4.3. The MLP model has the lowest sum of RMSEs, 75.0579, and also lowest average NRMSE and highest average improvement score. The accuracy of DNNs follows closely at the second place. Combining the results from Table 4.2 and 4.3, it can be concluded that the SVM model has the best overall performance while the MLP is better at avoiding large prediction errors. The performance of the DNN model is slightly worse than, but consistent with that of the MLP.

It also can be observed from Table 4.1 that the SVM has not only the lowest total MAE but also lowest individual error for each of the seven wind farms.

Table 4.4: Paired t-test Result for SVM

		Closest		
Dataset	Best Model	Model	p-value	μ_d
Erieau	SVM	MLP	6.05e-12	-0.2327
Dillon	SVM	MLP	3.69e-13	-0.2413
Spence	SVM	DNN	7.71e-13	-0.4335
Kingsbridge	SVM	DNN	1.54e-09	-0.0674
Wolfe Island	SVM	MLP	1.72e-11	-0.5406
Port Alma	SVM	MLP	2.46e-13	-0.4591
Port Alma 2	SVM	MLP	1.34e-12	-0.4617

To further support this observation, paired t-test has been conducted and the result is given in Table 4.4. The paired t-test shows that for all seven datasets, the MAEs produced by SVMs are significantly less. As shown in the table, the p-values are so small that it almost impossible that this observation is incorrect. The rightmost column gives the difference between means.

4.2 Prediction Time Horizons

It is not surprising at all that the longer the prediction time horizons are, the greater error will be produced. The experimental results of six hours prediction time horizons support this assertion. As shown in Figure 4.1, all six models[1] have MAEs clustered around 6 MW for one-hour-ahead predictions and their accuracy drops as the time horizon increases. Another noticeable

[1]Only representative models are presented in figures to improve readability.

Figure 4.1: Prediction Errors over Time Horizons

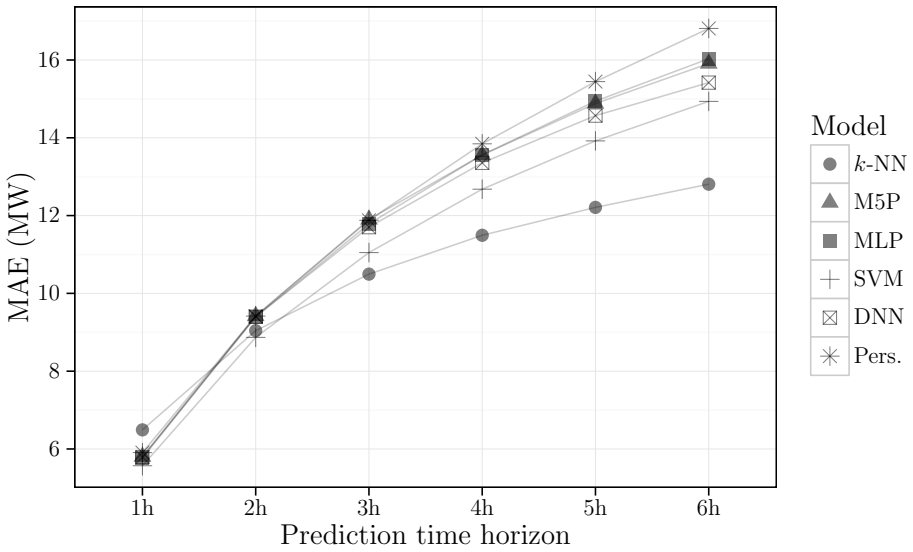
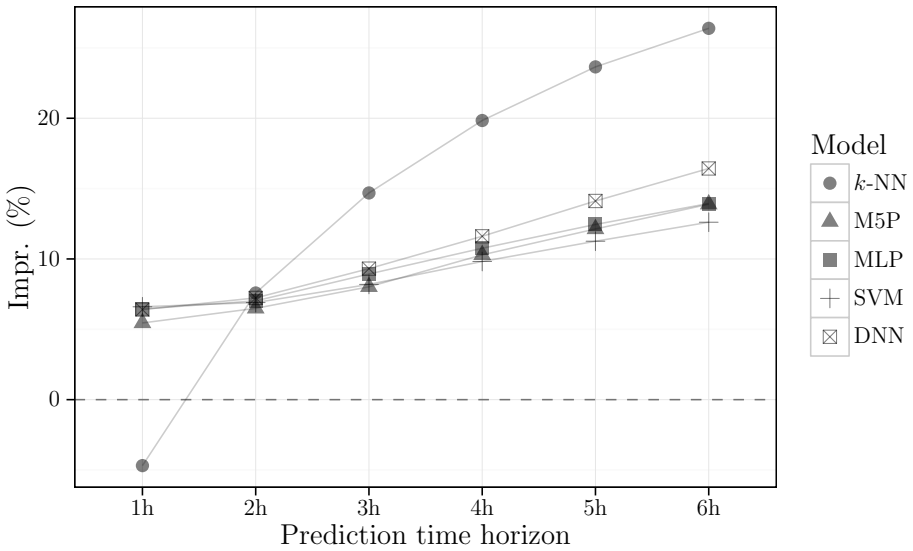


Figure 4.2: Prediction Improvements over Time Horizons



trend in this figure is that the rate at which the MAE increases actually decreases over time. It is noted that the MAEs increase the most between 1h and 2h and as prediction time becomes longer the error measure does not increase as much. Also, the performance of different models has a larger spread to the right of the time axis.

Last but not least, the k-NN model demonstrates a quite different performance pattern comparing to the other models. Again from Figure 4.1, one can notice that k-NN has the worst MAE at 1h. As the time horizon increases, however, the k-NN model produces relatively lower prediction error and has the largest lead among others at 6h. This special property of the k-NN can be more easily observed in the plotting of the improvement score shown in Figure 4.2. At 1h, the improvement score of the k-NN is blow 0%, which means it has a lower accuracy than the persistence model. In contrast, the improvement of the k-NN model for six-hour-ahead prediction is well above 20% and all other models stay closely around 15%. A different improvement curve of the k-NN is observed in the figure while the curves of other models follow a similar pattern.

The evaluation metric sums of all six time horizons are presented in Table 4.5. As with the figures, this table also shows that the k-NN model has the best overall performance in terms of all evaluation measures, including MAE, RMSE and improvement. In addition, the DNN and SVM are two models

Table 4.5: Sum of Metrics of All Time Horizons

	MAE (MW)	RMSE (MW)	Impr. (%)
Pers.	73.3018	109.3661	-
LR	73.5612	99.7886	48.57
k-NN	**62.5368**	**90.2427**	**87.46**
REP	72.1362	101.3048	37.53
M5P	71.4665	98.2360	56.28
MLP	71.4882	97.7236	59.54
RBFN	78.6587	105.4921	2.39
SVM	67.0202	98.6434	55.38
DNN	70.2383	96.4703	65.10

that are second to the k-NN. It also can be found in this table that the performance of the DNNs for longer prediction time is slightly better than that of the simpler MLP models.

4.3 Training Set Sizes

On top of the basic experiment using one year data, one extra experiment regarding dataset size was run and the result is presented in Table 4.6, where n is the number of records. The first column, where $n = 8760$, shows the case of one year data being used.

One can see from the table that all models, including two reference models

Table 4.6: RMSE for Different Dataset Sizes (MW)

	$n = 8760$	$n = 14242$	$n = 28485$
Pers.	13.4997	13.0176	11.7928
LR	12.7140	12.2176	11.1327
k-NN	13.3619	12.8442	11.7172
REP	13.1765	12.7273	11.4797
M5P	12.6421	12.1495	10.9832
MLP	12.5797	12.0486	10.9104
RBFN	12.6353	12.1376	11.7863
SVM	12.7932	12.2471	11.0138
DNN	12.5814	12.0537	10.9170

gain prediction accuracy when there are more training examples fed into them. The RMSE of the MLP keeps being the lowest for all three cases. It is worth noting that as a model with more complexity, the DNN does not benefit more from the increase of dataset size than other models do. This can be probably attributed to the fact that dataset size is no longer a bottleneck for DNNs at this level.

4.4 Spatial Correlation

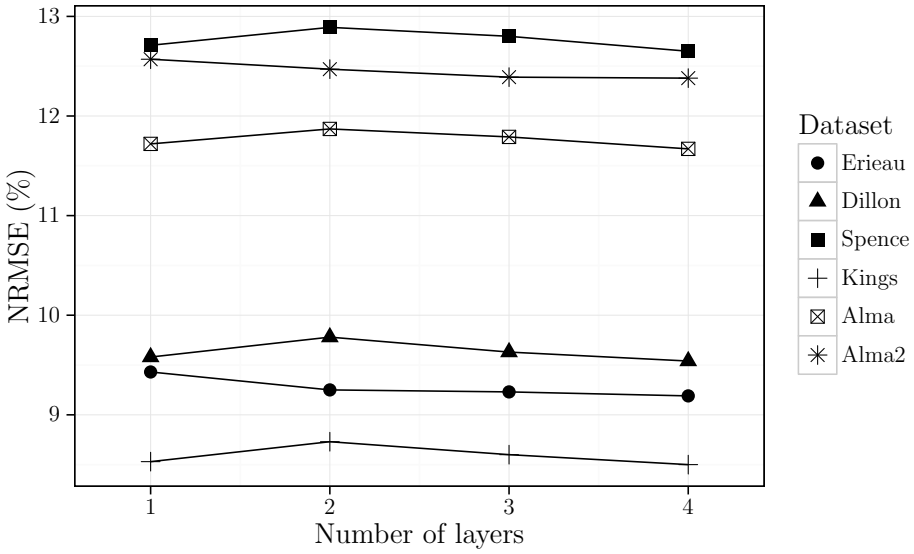
The spatial correlation experiment result is given in Table 4.7. It is shown in the table that for the normal case the SVM model and the RBFN model are the best predictor in terms of MAE and RMSE, respectively. When it

Table 4.7: Spatial Correlation Result

	MAE (MW)		**RMSE** (MW)	
	itself	neighbor	itself	neighbor
Erieau (Spence)				
LR	6.2179	10.5927	9.4200	15.0757
k-NN	6.3581	**7.5423**	9.7772	11.6918
REP	6.4774	8.7862	9.8350	13.6348
M5P	6.2179	8.2160	9.4217	12.5031
MLP	6.2103	9.0906	9.4010	13.4889
RBFN	6.2161	10.2265	**9.3853**	14.4573
SVM	**5.9796**	8.1068	9.4261	12.9468
DNN	6.3534	7.7317	9.4624	**11.5964**
Port Alma (Erieau)				
LR	7.4897	9.7311	11.9617	15.5990
k-NN	7.8406	**7.7947**	12.4775	**12.6935**
REP	7.8046	8.7558	12.3967	14.4521
M5P	7.4784	8.7091	11.9368	14.1686
MLP	7.5282	8.8801	11.9455	14.1687
RBFN	7.5233	9.8701	**11.8766**	15.4140
SVM	**7.0699**	8.3274	12.0112	14.9459
DNN	7.5425	8.8842	11.9414	14.3024

comes to predicting wind power by neighbor sites, however, the k-NN model shows its competence. In the case of predicting Port Alma from Erieau, the k-NN model has the MAE of 7.7947 that is even lower than its MAE of the regular prediction, 7.8406. Based on the limited experiment, the DNN model also shows the potential ability for spatial correlation prediction. This is in accordance with the fact that the DNN is suitable for finding more abstract

Figure 4.3: Prediction Error vs. Number of Layers



relationships which is what spatial correlation predictions likely possess.

4.5 DNN-related Results

In this section, the results of designated experiments for DNNs are presented and discussed.

The relation between prediction error and the number of hidden layers of the DNN is depicted in Figure 4.3. Focusing on the two-, three- and four-layered, the performance of DNNs on all datasets improves as the number of layers increases. However, the improvement gained from extra hidden layers is not significant. It is worth considering the tradeoff between the training time

Table 4.8: Aggregate RMSE for Different-sized DNNs

	RMSE (MW)	**NRMSE** (%)		**Impr.** (%)	
	sum	μ	σ	μ	σ
Pers.	60.4202	11.34	1.75	-	-
DNN:					
$n = 5$	57.4875	10.78	1.70	4.96	1.26
$n = 10$	57.1141	10.72	1.64	5.44	0.65
$n = 20$	57.0450	10.71	1.64	5.56	0.65
$n = 50$	56.9829	10.70	1.64	5.65	0.62
$n = 100$	56.9606	10.69	1.63	5.69	0.60
$n = 150$	56.9215	10.69	1.64	5.75	0.62

increased by hidden layers and the prediction accuracy.

Table 4.8 shows the RMSE measures for the DNNs with different number of neurons on their hidden layers. The number of hidden layers of these DNNs are fixed at three and they only differ in the number of the neurons on each layer, as indicated by n in the table. Similar to the effect of adding layers, increasing the number of neurons each layer also improves the performance of DNNs, but only to a limited extent. Note that the NRMSE does not change when adding 50 more neurons on top of 100. Meanwhile, the computational time dramatically increases with the size of DNNs.

Next, the performance of several DNNs that consist of 300 neurons is pre-

Table 4.9: Aggregate RMSE for DNNs with 300 Neurons

	RMSE (MW)	**NRMSE** (%)		**Impr.** (%)	
	sum	μ	σ	μ	σ
Pers.	60.4202	11.34	1.75	-	-
DNN:					
300	57.8939	10.88	1.63	4.03	0.73
150-150	57.2474	10.75	1.65	5.21	0.64
100-100-100	56.9606	10.69	1.63	5.69	0.60
75-75-75-75	**56.6800**	**10.64**	1.62	**6.13**	0.52
200-50-50	56.9382	10.69	1.63	5.73	0.62
50-200-50	56.9610	10.70	1.63	5.68	0.61
50-50-200	56.9862	10.70	1.63	5.63	0.62

sented in Table 4.9. While keeping DNNs' total number of neurons constant, the increase of the number of hidden layers results in improved performance. As shown in this table, the 4-layered DDN with 75 neurons on each layer yields the best outcomes for all measures. As for the four 3-layered 300-neuron DNNs in the table, the different placement of neurons does not appear to affect the performance. Nevertheless, putting more neurons in the first layer should be considered since the DNN with 200 neurons on its first layer (200-50-50) does achieve the lowest sum of RMSEs, 56.9382, among other three 3-layered DNNs.

The paired t-test was run for these DNNs on each individual wind farm dataset. The result shows that for all but Spence and Kingsbridge datasets,

Table 4.10: Influence of Extra Examples & Artificial Attribute

unit: MW

	Spence		*+ examples*		*+ v^3*	
	MAE	RMSE	MAE	RMSE	MAE	RMSE
Pers.	7.7494	13.4997	6.9889	11.7928	6.9889	11.7928
DNN:						
300	8.0806	12.7617	7.1198	11.1324	7.1203	11.1314
150-150	7.9637	12.6585	7.0091	11.0125	7.0153	11.0104
100-100-100	7.8899	12.5814	6.9221	10.9141	6.9296	10.9218
75-75-75-75	7.8888	12.5402	6.8540	10.8407	6.8613	10.8487
200-50-50	7.8906	12.5758	6.9235	10.9167	6.9194	10.9085
50-200-50	7.8937	12.5810	6.9473	10.9384	6.9380	10.9298
50-50-200	7.9021	12.5809	6.9382	10.9251	6.9388	10.9302

the 4-layered DNN has the significantly less MAE. In the cases of Spence and Kingsbridge, the last four DNNs in Table 4.9 are believed having the same performance in terms of MAE.

Table 4.10 presents the results when extra training examples and extra attributes were added. Adding extra training examples is already shown to be favorable to better accuracy in the earlier section and what is to note in this table is that the improvement of all DNN models is about the same in terms of both MAE and RMSE. The fact that all MAEs are reduced by about 1 MW and RMSEs by 1.6 implies that no specific structure of DNNs benefits more from the added training examples. On top of extra training examples,

Table 4.11: DNN Hidden Layer Dropout Rates

	MAE (MW)	**NMAE** (%)			
Dropout	sum	min.	max.	μ	σ
0.0	47.7501	5.53	7.97	6.69	0.96
0.01	47.7281	5.55	7.98	6.69	0.96
0.1	47.9628	5.64	8.00	6.73	0.95
0.5	51.0024	5.64	8.47	7.14	1.04
0.9	156.1770	12.08	28.53	23.62	5.01

adding the Wind Speed Cube attribute does not make any meaningful effect, as shown by the last two columns in Table 4.10. This is probably because the DNN has the ability to learn such simple non-linear relationships between attributes, so similar feature engineering is not necessary.

The MAE result regarding different hidden layer dropout rates is given in Table 4.11. It can be easily found that small dropout rates, less than 0.1 for example, are preferred for wind power predictions.

Chapter 5

Conclusion

Wind power prediction is a key component in wind power industry which is in high growth rate. Accurate predictions can greatly help in lowering the cost and improving the stability of wind power. Various prediction approaches have been studied in research and applied in practice, among which the machine learning models are currently the most promising ones. Although there has been a fair amount of research for machine learning models, such as ANNs and SVMs, in wind power prediction, the different research context settings made it difficult to evaluate the relative performance of these models. In addition, as a new technique deep learning is likely to exhibit its strength in this field, yet little work has been done regarding it.

In this study, we carefully collected the experimental data and ensured that the datasets contain as many wind characteristics as possible. This compre-

hensive empirical study considered nine models, seven of which are representative machine learning models, namely k-NN, REP tree, M5P tree, MLP, RBF network, SVM and DNN. Among these models, the SVM demonstrated the best overall performance and the k-NN showed its unique advantage for wind power predictions with longer time horizons. The MLP had the accuracy just second to the SVM while the DNN is slightly outperformed by the MLP. Nonetheless, the DNN exhibited its potential for more abstract predictions, such as 6-hour-ahead and spatial correlation predictions. For all models selected in this study, more training data resulted in better performance.

The above conclusion could be seen as reliable reference for future researchers and practitioners. The new findings regarding the DNN in this study indicates its limitations and potential, which provides guidance for prospective research.

We believe that the future work should mainly focus on the application of deep learning in the field of wind power prediction. In this study, the performance of DNNs improved as the prediction time horizon increased. Meanwhile, it is widely accepted that the NWP can significantly assist in longer term predictions [21]. Therefore, the combination of these two approaches deserves more attention. Additionally, based on the findings of this study and the fact that convolutional DNNs are good at revealing spatial topology

structures, the convolutional DNN should be able to unleash its power in spatial correlation predictions in the future.

Bibliography

[1] World Footprint: Do we fit on the planet? (2016). Retrieved March 18, 2016, from Global World Footprint website: http://www.footprintnetwork.org/en/index.php/GFN/

[2] Past Earth Overshoot Days. (2015). Retrieved February 15, 2016, from Earth Overshoot Day website: http://www.overshootday.org/newsroom/past-earth-overshoot-days/

[3] This Is A Turning Point: Three Things You Need to Know about The Paris Agreement. (2015, December 12). Retrieved March 4, 2016, from The Climate Reality Project webstie: https://www.climaterealityproject.org/blog/cop21-paris-agreement

[4] Fossil Fuel. (2016, March 3). In *Wikipedia*. Retrieved March 16, 2016, from https://en.wikipedia.org/wiki/Fossil_fuel

[5] The Cause of the Crisis (n.d.). In *Air Pollution in China*. Retrieve from http://sites.dartmouth.edu/anth491/causes/

[6] Artificial Neural Network. (2016, April 16). In *Wikipedia*. Retrieved April 20, 2016, from
https://en.wikipedia.org/wiki/Artificial_neural_network

[7] Gipe, P. (2004). *Wind Power: Renewable Energy for Home, Farm, and Business*. White River Junction, Vermont: Chelsea Green Publishing Company.

[8] Fox, B., Flynn, D., Bryans, L., Jenkins, N., Milborrow, D., O'Malley, M., Watson, R., & Anaya-Lara, O. (2007). *Wind Power Integration: Connection and system operational aspects*. London, UK: The Institution of Engineering and Technology.

[9] Jain, M. C. (2009). *Textbook Of Engineering Physics*. PHI Learning Pvt. Ltd.

[10] Freedman, D. A. (2009). *Statistical Models: Theory and Practice*. Cambridge, UK: Cambridge University Press.

[11] Mitchell, T. M. (1997). *Machine Learning*. Singapore: McGraw-Hill Companies, Inc.

[12] Hand, D. J., Mannila, H. & Smyth, P. (2001). *Principles of Data Mining*. Boston, USA: MIT Press.

[13] Cristianini, N. & Shawe-Taylor, J. (2000). *An Introduction to Support Vector Machines and Other Kernel-based Learning Methods*. Cambridge, UK: Cambridge University Press.

[14] Nielsen, M. (2016). *Neural Networks and Deep Learning*. Retrieved from http://neuralnetworksanddeeplearning.com/index.html

[15] Shelquist, R. (2016, January 14). An Introduction to Air Density and Density Altitude Calculations. Retrieved March 19, 2016, from Shelquist Engineering website: http://wahiduddin.net/calc/density_altitude.htm

[16] *IEA Wind 2014 Annual Report*. (2015, August). Retrieved from IEA Wind website: http://www.ieawind.org/annual_reports_PDF/2014.html

[17] *IEA Wind 2013 Annual Report*. (2014, August). Retrieved from IEA Wind website: http://www.ieawind.org/annual_reports_PDF/2013.html

[18] Dispatchable generation. (2016, January 22). In *Wikipedia*. Retrieved February 18, 2016, from https://en.wikipedia.org/wiki/Dispatchable_generation

[19] Soman, S. S., Zareipour, H., Malik, O. & Mandal, P. (2010). A review of wind power and wind speed forecasting methods with different time horizons. *North American Power Symposium*, 1-8.

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

[21] Kariniotakis, G. et al. (2004) The State of the Art in Short-term Prediction of Wind Power - From an Offshore Perspective. In *Proc. of 2004 SeaTechWeek*

[22] Jung, J. & Broadwater, R. P. (2013). Current status and future advances for wind speed and power forecasting. *Renewable and Sustainable Energy Reviews, 31*, 762-777.

[23] Busseti, E., Osband, I. & Wong, S. (2012, December 14). *Deep Learning for Time Series Modeling* (Project report). Retrieved from `http://cs229.stanford.edu/`

[24] Colak, I., Sagiroglu, S. & Yesilbudak, M. (2012). Data mining and wind power prediction: A literature review. *Renewable Energy, 46*, 241-247.

[25] Chang, W.-Y. (2014). A Literature Review of Wind Forecasting Methods. *Journal of Power and Energy Engineering, 2*, 161-168. `http://dx.doi.org/10.4236/jpee.2014.24023`

[26] Saroha, S. & Aggrawal, S. K. (2015). A Review and Evaluation of Current Wind Power Prediction Technologies. *WSEAS Transactions on Power Systems, 10*.

[27] Madsen, H., Pinson, P., Kariniotakis, G., Nielsen, H. A. & Nielsen, T. S. (2005). A protocol for standardizing the performance evaluation of short term wind power prediction models. *Wind Engineering, 29*(6), 475-489.

[28] Kusiak, A., Zheng, H. & Song, Z. (2009). Short-Term Prediction of Wind Farm Power: A Data Mining Approach. *IEEE Transactions on Energy Conversion, 24*(1), 125-136.

[29] Ernst, B., Oakleaf, B., Ahlstrom, M. L., Lange, M., Moehrlen, C., Lange, B., Focken, U. & Rohrig, K. (2007). Predicting the wind. *IEEE power & energy magazine*, 78-89.

[30] Mao, M., Cao, Y. & Chang, L. (2013). Improved Fast Short-term Wind Power Prediction Model Based on Superposition of Predicted Error. *Power Electronics for Distributed Generation Systems (PEDG), 2013 4th IEEE International Symposium on*, Rogers, AR, 2013, pp. 1-6.

[31] Rzui, P. C. & Eremia, M. (2011). Prediction of wind power by artificial intelligence techniques. *Intelligent System Application to Power Systems (ISAP), 2011 16th International Conference on*, Hersonissos, 2011, pp. 1-6.

[32] Zeng, J. & Qiao, W. (2011). Support vector machine-based short-term wind power forecasting. *Power Systems Conference and Exposition (PSCE)*, Phoenix, AZ, pp. 1-8.

[33] Quinlan, J. R. (1992). Learning with continuous classes. In *Proc. of AI'92 (Adams & Sterling, Eds)*, 343-348, Singapore

[34] Chen, K. & Yu, J. (2014). Short-term wind speed prediction using an unscented Kalman filter based state-space support vector regression approach. *Applied Energy, 113*, 690-705.

[35] Catalao, J. P. S., Pousinho, H. M. I. & Mendes, V. M. F. (2011). Hybrid wavelet-PSO-ANFIS approach for short-term wind power forecasting in Portugal. *IEEE Transactions on Sustainable Energy, 2*(1), 50-59.

[36] Potter, C. W. & Negnevitsky, M. (2006). Very short-term wind forecasting for Tasmanian power generation. *IEEE Transactions on Power Systems, 21*(2), 965-972.

[37] Cao, Q., Ewing. B. T. & Thompson, M. A. (2012). Forecasting wind speed with recurrent neural networks. *European Journal of Operational Research, 211*, 148-154.

[38] Hu, Q., Zhang, S., Xie, Z., Mi, J. & Wan, J. (2014). Noise model based -support vector regression with its application to short-term wind speed forecasting. *Neural Networks, 57*, 1-11.

[39] Ramirez-Rosado, I. J., Fernandez-Jimenez, L. A., Monteiro, C., Sousa, J. & Bessa, R. (2009). Comparison of two new short-term wind-power forecasting systems.*Renewable Energy, 34*, 1848-1854.

[40] Liu, D., Niu, D., Wang, H. & Fan, L. (2014). Short-term wind speed forecasting using wavelet transform and support vector machines optimized by genetic algorithm. *Renewable Energy, 62*, 592-597.

[41] Sideratos, G. & Hatziargyriou, N. D. (2012). Probabilistic wind power forecasting using radial basis function neural networks. *IEEE Transactions on Power Systems*, 1-9.

[42] Fan, S., Liao, J. R., Yokoyama, R., Chen. L. & Lee, W.-J. (2009). Forecasting the wind generation using a two-stage network based on meteorological information. *IEEE Transactions on Energy Conversion, 24*(2), 474-482.

[43] Damousis, I. G., Alexiadis, M. C., Theocharis, J. B. & Dokopoulos, P. S. (2004). A fuzzy model for wind speed prediction and power generation in wind parks using spatial correlation. *IEEE Transactions on Energy Conversion, 19*(2), 352-361.

[44] Monfared, M., Rastergar, H. & Kojabadi, H. M. (2009). A new strategy for wind speed forecasting using artificial intelligent methods. *Renewable Energy, 34*, 845-848.

[45] Chang, W.-Y. (2013). Application of Back Propagation Neural Network for Wind Power Generation Forecasting. *International Journal of Digital Content Technology and its Application, 7*, 502-509.

[46] Togelou, A., Sideratos, G. & Hatziargyriou, N. D. (2012). Wind power forecasting in the absence of historical data. *IEEE Transactions on Sustainable Energy, 3*(3), 416-421.

[47] Sideratos, G. & Hatziargyriou, N. D. (2007). An advanced statistical method for wind power forecasting. *IEEE Transactions on Power Systems, 22*(1), 258-265.

[48] Tao, Y., Chen, H. & Qiu, C. (2014). Wind power prediction and pattern feature based on deep learning method. *Power and Energy Engineering Conference (APPEEC), 2014 IEEE PES Asia-Pacific*, Hong Kong, pp. 1-4.

[49] Amjady, N., Keynia, F. & Zareipour, H. (2011). Wind power prediction by a new forecast engine composed of modified hybrid neural network and enhanced particle swarm optimization. *IEEE Transactions on Sustainable Energy, 2*(3), 265-276.

[50] Cadenas, E. & Riversa, W. (2010). Wind speed forecasting in three different regions of Mexico, using a hybrid ARIMA-ANN model. *Renewable Energy, 35*, 2732-2738.

[51] Barbounis, T. G., Theocharis, J. B., Alexiadis, M. C. & Dokopoulos, P. S. (2006). Long-term wind speed and power forecasting using recurrent neural network models. *IEEE Transactions on Energy Conversion, 21*(1), 273-284.

[52] Data Directory. Retrieved October, 2015 from IESO website: `http://www.ieso.ca/Pages/Power-Data/Data-Directory.aspx`

[53] Historical Climate Data. Retrieved October, 2015 from Government of Canada website:
http://climate.weather.gc.ca/index_e.html#access

[54] Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P. & Witten, I.H. (2009). The WEKA Data Mining Software: An Update. *SIGKDD Explorations, 11*(1).

Appendix A

Original Experimental Results

Table A.1: Original NMAE – Wind Farms (%)

	Erieau	Dillon	Spence	Kings	Wolfe	Alma	Alma2	*min.*	*max.*	μ	σ
Pers.	6.12	6.31	7.83	5.44	5.78	7.21	7.65	5.44	7.83	6.62	0.87
LR	6.07	6.25	8.09	5.54	5.78	7.37	7.95	5.54	8.09	6.72	0.98
k-NN	6.79	7.13	8.56	6.05	6.32	8.08	8.46	6.05	8.56	7.34	0.95
REP	6.48	6.71	8.38	5.88	6.14	7.73	8.21	5.88	8.38	7.07	0.94
M5P	6.05	6.25	8.01	5.54	5.77	7.36	7.91	5.54	8.01	6.70	0.96
MLP	6.02	6.23	7.99	5.51	5.67	7.37	7.86	5.51	7.99	6.66	0.97
RBFN	6.39	6.57	8.08	5.76	6.16	7.41	7.96	5.76	8.08	6.90	0.84
SVM	5.79	5.92	7.53	5.36	5.39	6.92	7.40	5.36	7.53	6.33	0.86
DNN	6.10	6.23	7.97	5.53	5.71	7.35	7.94	5.53	7.97	6.69	0.96

Table A.2: Original NRMSE – Wind Farms (%)

	Erieau	Dillon	Spence	Kings	Wolfe	Alma	Alma2	μ	σ
Pers.	9.78	10.20	13.64	9.01	9.88	12.38	13.04	11.13	1.70
LR	9.29	9.63	12.84	8.51	9.31	11.81	12.50	10.56	1.64
k-NN	10.33	11.00	13.50	9.40	10.18	12.70	13.14	11.46	1.50
REP	9.86	10.32	13.31	9.11	9.94	12.26	12.87	11.10	1.55
M5P	9.31	9.64	12.77	8.51	9.29	11.77	12.42	10.53	1.60
MLP	9.21	9.61	12.71	8.46	9.15	11.79	12.32	10.47	1.61
RBFN	9.59	9.91	12.76	8.67	9.67	11.76	12.45	10.69	1.49
SVM	9.26	9.53	12.92	8.39	9.22	11.79	12.45	10.51	1.69
DNN	9.23	9.58	12.71	8.53	9.18	11.72	12.39	10.48	1.60

Table A.3: Original MAE – Time Horizons (MW)

	1h	2h	3h	4h	5h	6h	*Total*
Pers.	5.9068	9.4155	11.8795	13.8450	15.4445	16.8105	73.3018
LR	5.9009	9.6307	12.1132	13.9535	15.3979	16.5650	73.5612
k-NN	6.4910	9.0403	10.4944	11.4931	12.2107	12.8073	62.5368
REP	6.1198	9.7498	12.0526	13.6532	14.8429	15.7178	72.1362
M5P	5.8011	9.4191	11.9004	13.5569	14.8804	15.9088	71.4665
MLP	5.7698	9.3960	11.7866	13.5626	14.9416	16.0315	71.4882
RBFN	8.2420	10.7442	12.8192	14.4299	15.7052	16.7182	78.6587
SVM	5.5687	8.8687	11.0493	12.6781	13.9200	14.9354	67.0202
DNN	5.7911	9.4021	11.7018	13.3559	14.5696	15.4179	70.2383

Table A.4: Original RMSE – Time Horizons (MW)

	1h	2h	3h	4h	5h	6h	*Total*
Pers.	9.4848	14.5570	17.8826	20.4782	22.5963	24.3672	109.3661
LR	9.0173	13.7635	16.6081	18.6745	20.2412	21.4840	99.7886
k-NN	9.9292	13.4535	15.2561	16.4148	17.2534	17.9358	90.2427
REP	9.4589	14.2025	16.9688	18.9167	20.3499	21.4080	101.3048
M5P	8.9675	13.6123	16.4488	18.3708	19.8552	20.9815	98.2360
MLP	8.8739	13.5356	16.2850	18.2749	19.7827	20.9716	97.7236
RBFN	11.7876	14.9593	17.3729	19.1774	20.5583	21.6366	105.4921
SVM	8.8588	13.5530	16.4179	18.4678	20.0512	21.2947	98.6434
DNN	8.8774	13.5052	16.2179	18.1017	19.4052	20.3629	96.4703

Table A.5: Original MAE & RMSE – Number of Hidden Layers (MW)

# of layers	Erieau	Dillon	Spence	Sp. L	Sp. V3	Kings	Alma	Alma2	*Total*
MAE									
1	6.1909	4.8594	7.8899	6.9221	7.1189	2.2123	7.4189	8.1741	50.7864
2	6.0252	4.9796	8.0838	7.1177	7.0137	2.3012	7.5220	8.0379	51.0809
3	6.0413	4.8823	7.9538	7.0166	6.9296	2.2442	7.4398	8.0209	50.5284
4	5.9756	4.8245	7.8717	6.8434	6.8554	2.2060	7.3793	7.9996	49.9554
RMSE									
1	9.3315	7.4752	12.5814	10.9141	11.1346	3.4124	11.8360	12.6957	79.3808
2	9.1552	7.6251	12.7600	11.1338	11.0113	3.4918	11.9935	12.5940	79.7647
3	9.1406	7.5140	12.6674	11.0147	10.9218	3.4384	11.9060	12.5151	79.1178
4	9.1018	7.4392	12.5206	10.8282	10.8410	3.3999	11.7861	12.5068	78.4234

Table A.6: Original MAE & RMSE – Number of Neurons (MW)

# of neurons	Erieau	Dillon	Spence	Sp. L	Sp. V3	Kings	Alma	Alma 2	*Total*
MAE									
5	6.0836	4.8688	7.9640	7.0149	7.0191	2.2364	7.4682	8.2598	50.9148
10	6.0635	4.8665	7.9240	6.9837	6.9836	2.2259	7.4513	8.0625	50.5609
20	6.0484	4.8627	7.9129	6.9589	6.9528	2.2223	7.4383	8.0463	50.4425
50	6.0363	4.8592	7.8890	6.9409	6.9529	2.2184	7.4243	8.0334	50.3544
100	6.0413	4.8594	7.8899	6.9221	6.9296	2.2123	7.4189	8.0209	50.2943
150	6.0144	4.8533	7.8907	6.9260	6.9189	2.2170	7.4101	8.0059	50.2363
RMSE									
5	9.1775	7.4918	12.6493	11.0005	11.0032	3.4218	11.8908	12.8564	79.4912
10	9.1633	7.4893	12.6098	10.9762	10.9708	3.4215	11.8694	12.5608	79.0611
20	9.1513	7.4824	12.5914	10.9452	10.9474	3.4168	11.8563	12.5467	78.9375
50	9.1393	7.4787	12.5847	10.9293	10.9421	3.4142	11.8404	12.5256	78.8542
100	9.1406	7.4752	12.5814	10.9141	10.9218	3.4124	11.8360	12.5151	78.7964
150	9.1214	7.4694	12.5758	10.9178	10.9101	3.4138	11.8333	12.5078	78.7494

Table A.7: Original MAE & RMSE – Hidden Layers Structure (MW)

Structure	Erieau	Dillon	Spence	Sp. L	Sp. V3	Kings	Alma	Alma 2	*Total*
MAE									
300	6.1882	4.9795	8.0806	7.1198	7.1203	2.3007	7.5225	8.1703	51.4818
150-150	6.0215	4.8718	7.9637	7.0091	7.0153	2.2415	7.4264	8.0213	50.5705
100-100-100	6.0413	4.8594	7.8899	6.9221	6.9296	2.2123	7.4189	8.0209	50.2943
75-75-75-75	5.9812	4.8293	7.8888	6.8540	6.8613	2.2079	7.3831	7.9483	49.9538
200-50-50	6.0302	4.8525	7.8906	6.9235	6.9194	2.2116	7.4132	8.0173	50.2583
50-200-50	6.0386	4.8566	7.8937	6.9473	6.9380	2.2206	7.4110	8.0132	50.3190
50-50-200	6.0492	4.8724	7.9021	6.9382	6.9388	2.2206	7.4299	8.0278	50.3789
RMSE									
300	9.3295	7.6237	12.7617	11.1324	11.1314	3.4916	11.9929	12.6946	80.1577
150-150	9.1546	7.5075	12.6585	11.0125	11.0104	3.4360	11.9001	12.5907	79.2702
100-100-100	9.1406	7.4752	12.5814	10.9141	10.9218	3.4124	11.8360	12.5151	78.7964
75-75-75-75	9.1016	7.4417	12.5402	10.8407	10.8487	3.3997	11.7901	12.4067	78.3694
200-50-50	9.1332	7.4704	12.5758	10.9167	10.9085	3.4124	11.8335	12.5130	78.7634
50-200-50	9.1393	7.4770	12.5810	10.9384	10.9298	3.4147	11.8345	12.5145	78.8292
50-50-200	9.1461	7.4829	12.5809	10.9251	10.9302	3.4168	11.8399	12.5197	78.8416

Table A.8: Original MAE & RMSE – Dropout Rates (MW)

Dropout	Erieau	Dillon	Spence	Sp. L	Sp. V3	Kings	Wolfe	Alma	Alma 2	Total
MAE										
0.0	6.0413	4.8594	7.8899	6.9221	6.9296	2.2123	11.3075	7.4189	8.0209	61.6018
0.01	6.0151	4.8666	7.9022	6.9151	6.9394	2.2207	11.2847	7.4203	8.0184	61.5825
0.1	6.0545	4.8871	7.9164	6.9388	6.9391	2.2554	11.3539	7.4463	8.0492	61.8407
0.5	6.5520	5.4288	8.3122	7.2988	7.3037	2.2554	11.9926	7.9049	8.5564	65.6049
0.9	26.1051	22.2549	22.3277	23.9800	23.9797	9.8476	23.9119	25.2342	26.4956	204.1366
RMSE										
0.0	9.1406	7.4752	12.5814	10.9141	10.9218	3.4124	18.1695	11.8360	12.5151	96.9660
0.01	9.1182	7.4759	12.5774	10.9033	10.9278	3.4174	18.1414	11.8393	12.5057	96.9064
0.1	9.1205	7.4777	12.5753	10.9168	10.9206	3.4460	18.1645	11.8310	12.4965	96.9487
0.5	9.3822	7.7549	12.7132	11.0906	11.0950	3.4460	18.5447	12.0112	12.7020	98.7396
0.9	30.7637	25.9386	27.5963	29.0294	29.0310	11.7726	29.4241	30.7023	31.4470	245.7049

Vita

Candidate's full name: Yiqian Liu
University attended: Shandong University, Bachelor of Science, 2013