Fall 2009

Predictive engineering in wind energy: a data-mining approach

Wenyan Li
University of Iowa

PREDICTIVE ENGINEERING IN WIND ENERGY: A DATA-MINING APPROACH

by

Wenyan Li

A thesis submitted in partial fulfillment
of the requirements for the Master of Science
degree in Industrial Engineering
in the Graduate College of
The University of Iowa

December 2009

Thesis Supervisor: Professor Andrew Kusiak

Graduate College
The University of Iowa
Iowa City, Iowa

CERTIFICATE OF APPROVAL

_____

MASTER'S THESIS

_____

This is to certify that the Master's thesis of

Wenyan Li

has been approved by the Examining Committee
for the thesis requirement for the Master of Science degree
in Industrial Engineering at the December 2009 graduation.

Thesis Committee: _____
Andrew Kusiak, Thesis Supervisor

_____
Yong Chen

_____
Peter J. O'Grady

ACKNOWLEDGMENTS

I would like to express my sincere gratitude to my advisor Professor Andrew Kusiak for his support in this research. He is the most instrumental person in my academic and research achievements. He provided the motivation, encouragement, guidance and advice which have prepared me for the life challenges that lie ahead. I was exposed to real-world applications while working in the Intelligent Systems Laboratory. This invaluable experience has allowed me to balance theory and practice resulting in usable solutions of industrial problems.

I would like to thank my other Thesis committee members: Professor Yong Chen and Professor Peter J. O'Grady for providing valuable suggestions and feedback on my research.

I am also grateful for the financial support from Iowa Energy Center and MidAmerican Energy Company. The energy experts from the company and Iowa Energy Center have offered me invaluable guidance for this research.

I thank all the members of the Intelligent Systems Laboratory who have worked with me and provided advice, reviews, and suggestions. Special thanks to my colleagues: Dr. Zhe Song and Haiyang Zheng, who worked with me on challenging wind energy projects; Robert A. Hamel, who assisted me to getting access to the industrial data; Zijun Zhang and Mingyang Li, who supported me in research and study.

And finally, I also wish to thank my parents for their support of my Masters' degree studies.

ABSTRACT

The large-scale wind energy industry is relatively new and is rapidly expanding. The ability of a wind turbine to extract power from the wind is expressed with the power curve. The key parameter determining wind turbine performance is wind speed and it is normally measured by an anemometer placed at the nacelle of a turbine.

The dynamic nature of wind is a barrier that calls for applying predictive engineering. Traditional approaches based on physics and mathematical modeling are not fully handle the variable nature of the wind.

Data mining is a promising approach for modeling in wind energy, including power prediction and optimization, wind speed forecasting, power curve monitoring, and fault diagnosis. It involves a number of steps including data pre-processing, data sampling, feature selection, and dimensionality reduction. This Thesis focuses on applying data-mining to predictive engineering in wind industry. Models for prediction of wind speed and wind farm power, turbine, and fault diagnosis are built. However, the approach and methods discussed in this research are also applicable to other industrial processes.

Chapter 2 introduces a methodology for short-term wind speed prediction based on wind farm data. Chapter 3 and Chapter 4 present prediction models for wind turbine parameters. Chapter 5 proposes strategies for dynamic control of wind turbines. Chapter 6 explores the fault diagnosis and prediction using SCADA data.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER1.

INTRODUCTION

Generating electricity from the wind is environmentally friendly, socially acceptable, and economically competitive [1]. Wind power has become the dominant source of alternative energy [2] and experienced a dynamic growth in the recent years [3]. However, the operations and maintenance costs [4, 5] have created a barrier to an even more rapid expansion, which aims at a twenty-fold increase in the wind energy production by the year 2030 [3]. The fact that wind energy is considered as most preferred alternative energy source by many researches has motivated further growth of wind farms and research in wind energy.

1.1 Review of the Methodologies for Wind Speed Forecasting

The power extracted from the wind is expressed by (1.1) [6]:

$$P_a = \frac{1}{2}\rho\pi R^2 C_p(\lambda,\beta)v^3 \tag{1.1}$$

where ρ is the density of air [kg/m^3], πR^2 is the swept area of rotor [m^2], $C_p(\lambda,\beta)$ is the power coefficient, and v is the wind speed. Wind speed is the key parameter in (1.1), however, by far it is the most difficult parameter to estimate [1]. Therefore, the need for models for accurate prediction of wind speed is apparent.

A number of approaches have been used to predict wind speed on different time scales needed by different turbine subsystems. The control system of a wind turbine requires high frequency wind speed data to efficiently extract the energy from the wind. Delays associated with getting such data lead to decrease in performance of the turbine due to delayed controller actions [7-10]. Predicting wind speed on a short time scale is also important for monitoring wind turbines. A recurrent neural-network model for

prediction of wind speed over a two to three day horizon was proposed in [11]. The numerical weather prediction (NWP) models are commonly used to forecast wind speeds at hourly or longer basis [1, 12]. The challenges faced by such models, due to the stochastic nature of wind, are widely reported in the literature [13].

Data mining is a promising approach for wind speed prediction and has been proven to perform well. Models developed with various data mining techniques have been reported in the literature, including linear prediction models [9], fuzzy logic [1, 14, 15], neural networks [8, 16, 17, 18], and support vector machine models [19, 20]. Yet, another approach for wind speed prediction is based on time-series models [21, 13].

1.2 Review of the Methodologies for Power Prediction and Optimization

Knowing the power to be produced by a wind turbine at future time horizons is of interest to the rapidly expanding wind industry [22]. Wind power forecasts are used as input for various tools e.g., management of power dispatch and control of wind turbines [23]. Wind power generation depends on wind speed, which however, might be impacted by the terrain orography. Wind speed exhibits randomness leading to unpredictability and variability of the wind power generation, both becoming challenge faced by power system operators [24].

Various approaches have been studied to address prediction of power produced at short- and long-term horizons. The state-of-the-art approaches to wind power forecasting have been published in [25] with the more recent updates included in [26]. Models used for forecasting wind power are categorized as: physics-based models, statistical models, and spatial correlation models [24, 27-30]. Data-mining algorithms offer a promise to conquer the unresolved gap of handling the dynamic nature of wind [13].

The published literature on data-mining in wind power is growing, with Neural Networks (NNs) becoming widely used algorithms. NN algorithms can be used to

estimate power output as a function of wind turbine parameters (e.g., wind speed, generator torque) and time delay of the corresponding parameters (e.g., power itself, wind speed) [31, 32]. Wind speed, relative humidity, and time were used as input variables to train a NN model in power prediction applications [2, 33]. The recurrent multilayer-perception NN was applied for power prediction in [34]. Long- and short-term prediction of power using the *k*-nearest neighbor (*k*-NN) algorithm was presented in [13, 31]. Analysis and estimation of power based on cluster analysis was reported in [35, 36].

Most literature on wind turbine control has focused on maximizing power [37-43] in the cut-in and the cut-out rage of wind speed. This goal is usually achieved by controlling the generator torque so that the rotor speed producing the optimum power coefficient is attained. The wind power is maximized predominantly when the wind speed is below its rated value and the blade pitch angle is fixed. Besides the traditional control strategies (i.e., mainly feedback and adaptive-tracking based), predictive control [43, 44, 7] has been used to optimize the capture of the wind power. The model predictive control approach with blade pitch and generator torque as two control inputs was discussed in [44]. The research reported in [43] and [44] was based on simulated wind speeds in a reactive rather than predictive mode. Unlike traditional energy conversion systems, where the fuel input can be controlled, the speed of the wind cannot be controlled. However, knowing the wind speed ahead of time is useful in controlling a wind turbine. Wind speed prediction was considered in [7], where a linear wind speed time series model was applied to determine wind speed in a short-time horizon (i.e., seconds). The wind turbine power output was assumed to be a linear function of the wind speed [7]. Thus, knowing the wind speed ahead of time would lead to smooth power generation.

1.3 Review of Methodologies for Fault Diagnosis

The growth of wind power has increased interest in the operations and maintenance of wind turbines. As wind turbines are located at remote locations that may be difficult to access, their maintenance becomes an issue. As indicated in [4], a $5,000 replacement of a bearing can turn into a $250,000 project involving cranes and a service crew in addition to the loss of power generation. For a turbine with 20 years of operating life, the operations, maintenance, and part replacement costs were estimated in the past to be at least 10%-15% of the total income from the generation [5]. Thus, condition monitoring and fault diagnosis of wind turbines are of high priority.

The state-of-the-art research in wind turbine condition monitoring and fault diagnosis has been covered in the past literature [47-51] with more recent updates included in [52, 53]. Modern wind turbines are usually equipped with some form of condition monitoring systems, including system-level or subsystem-level fault detection. Subsystem-level fault detection systems are usually based on monitoring parameters such as the vibration of the wind turbine drive train [54], bearing temperature, oil particulate content, optical strain measurements [55], and so on. Some commercially available solutions include blade monitoring systems [56], Supervisory Control and Data Acquisition (SCADA) interpretation systems [57], and holistic models [58]. The system-level condition monitoring and fault diagnosis offer a challenge that has led to numerous modeling and solution approaches presented in the literature, including Petri Nets [59], physics-based models [60, 61], multi-agent framework for fault detection [62], and sensor-based network [63].

1.4 Predictive Engineering under Framework of Data Mining

Data mining is a promising approach for modeling wind energy, e.g., power prediction and optimization, wind speed forecasting, and power curve monitoring. It

involves a number of steps including data pre-processing, data sampling, feature selection, and dimension reduction. Based on the ideas discussed in this thesis, a data-driven approach is applied to build wind farm power prediction and wind speed forecasting models, and realize control optimization strategy. Besides, fault diagnosis is explored at system level of wind turbine. Data mining and computation intelligence technique are employed in this research.

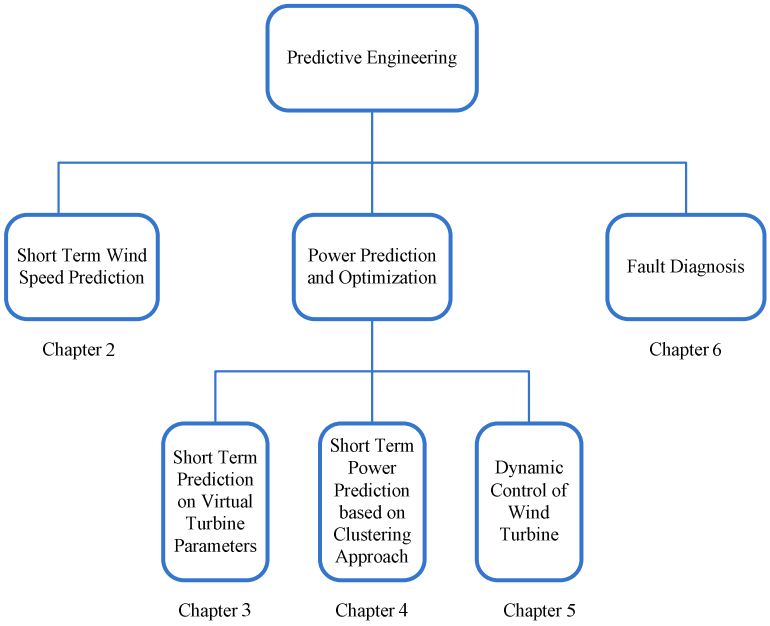

Figure 1. 1 Master Thesis Structure.

Figure 1.1 illustrates the structure of the thesis. Various data mining algorithms for regression modeling are used in Chapter 2 through Chapter 6.

In Chapter 2, a method for prediction of wind speed at a selected location based on the data collected at neighborhood locations with wind conditions is presented. The affinity of wind speeds measured at different locations is defined by Pearson's correlation

coefficient. Five turbines with similar wind conditions are selected among thirty wind turbines for in-depth analysis. The wind data from these turbines is used to predict wind speed at a selected location. A neural network ensemble is used to predict the value of wind speed at the turbine of interest. The models have been tested and the computational results are discussed. The results demonstrate that higher Pearson's correlation coefficient between the wind speed lead to better prediction accuracy for a same training and test scenario.

In Chapter 3, a data-driven methodology for the development of virtual models of a wind turbine is presented. To demonstrate the proposed methodology, two parameters of the wind turbine have been selected for modeling, power output and rotor speed. A virtual model for each of the two parameters is developed and tested with data collected at a wind farm. Both models consider controllable and non-controllable parameters of the wind turbine, as well as the delay effect of wind speed and other parameters. To mitigate data bias of each virtual model and ensure its robustness, a training set is assembled from ten randomly selected turbines. The performance of a virtual model is largely determined by the input parameters selected and the data-mining algorithms used to extract the model. Several data-mining algorithms for parameter selection and model extraction are analyzed. The research presented in the paragraph is illustrated with computational results.

In Chapter 4, a clustering approach is presented for short-term prediction of power produced by a wind turbine at low wind speeds. Increased prediction accuracy of wind power to be produced at future time periods is often bounded by the prediction model complexity and computational time involved. In this paragraph, a trade-off between the two confecting objectives is addressed. First, a set of the most relevant parameters (predictors) is selected using the underlying physics and pattern immersed in data. Five scenarios of the input space are analyzed with the k-means clustering algorithm. The most promising clustering scenario is applied to produce a model for each

clustered subspace. Computational results are compared and the benefits of cluster–specific (customized) models are discussed. The results show that the prediction accuracy is improved using fewer parameters provided the input space is properly clustered and customized prediction models are developed.

Chapter 5 presents an intelligent wind turbine control system based on models integrating the following three approaches: data mining, model predictive control, and evolutionary computation. To enhance the control strategy of the intelligent system, a multi-objective model is proposed. The model involves five different objectives with different weights controlling the wind turbine performance. These weights are adjusted in response to the variable wind conditions and operational requirements. Three control factors, wind speed, turbulence intensity, and electricity demand are considered in eight computational scenarios. The performance of each scenario is illustrated with numerical results.

Chapter 6 explores fault data provided by the supervisory control and data acquisition system and offers fault prediction at three levels: (1) fault and no-fault prediction; (2) fault category (severity); and (3) the specific fault prediction. For each level, the emerging faults are predicted 5 to 60 minutes before they occur. Various data-mining algorithms have been applied to developed models predicting possible faults. Computational results validating the models are provided. The research limitations are discussed.

CHAPTER 2.
PREDICTIVE ENGINEERING MODELS FOR SHORT TERM PREDICTION OF WIND SPEED

2.1 Introduction

A methodology for prediction of wind speed at ten second intervals is proposed. The affinity of wind speeds collected at different wind turbines is defined by the Pearson's correlation coefficient. The data from wind turbines with similar wind conditions is used to predict wind speed for a turbine of interest. Three models are proposed in this paragraph: single-predictor model, multi-predictor model, and the predictor-transformed model. The impact of wind speed measurements on the accuracy of the predicted wind speed is studied. In cases the number of inputs was excessive, a dimensionality approach was attempted. The predictive models are extracted with neural network ensembles for two training and test scenarios with each scenario including 10 experiments. That is, two training and test scenarios are compared for each of the 10 experiments in this paragraph.

2.2 Data Description and Methodology for Prediction of Wind Speed

2.2.1 Data Description

The ten second data used in this research was collected from thirty wind turbines at two different time periods. The first period covers seven days in August of 2007 and the second period covers the same length horizon in September of 2008. For both time

periods, 60481 instances were collected at each turbine. The description of the data used in this research is provided in Table 2.1.

Table 2.1. Description of the data sets.

	Start Time	End Time	Time Interval	Number of Data Points for Each Turbine
Data set 1	8/8/07 12:00 AM	8/15/07 12:00 AM	10 second	60481
Data set 2	9/22/08 12:00 AM	9/29/08 12:00 AM	10 second	60481

Two data sets provided here imply different wind characters. In data set 1, 96.75% of the wind speed values are less than 12.5 m/s, and 88.5% of the power output values are smaller than 1000 kW (out of 1500 kW). Data set 2 contains 18% of the wind speed values that are larger than 12.5 m/s, and nearly half of the power output values are higher than 1000 kW. Figure 2.1 illustrates the distribution of wind speed for both data sets.

All data used in this paragraph has been collected at 1.5MW turbines with cut-in-wind speed of 3.5 m/s, the rated speed of 12.5 m/s, and the cut-out-wind speed is 21 m/s.

2.2.2 Training and Test Data Sets

As shown in Figure 2.1, data set 1 provides a good coverage of wind speeds in the interval [3.5, 12.5] which is the focus in industrial applications. Therefore, two training and test scenarios are considered (see Table 2). In the first scenario 2/3 data from data set 1 is used for model development and the remaining 1/3 data is used to test the model. To test the robustness of the proposed models, scenario 2 is explored where the models are trained with data set 1 and are tested with data set 2.

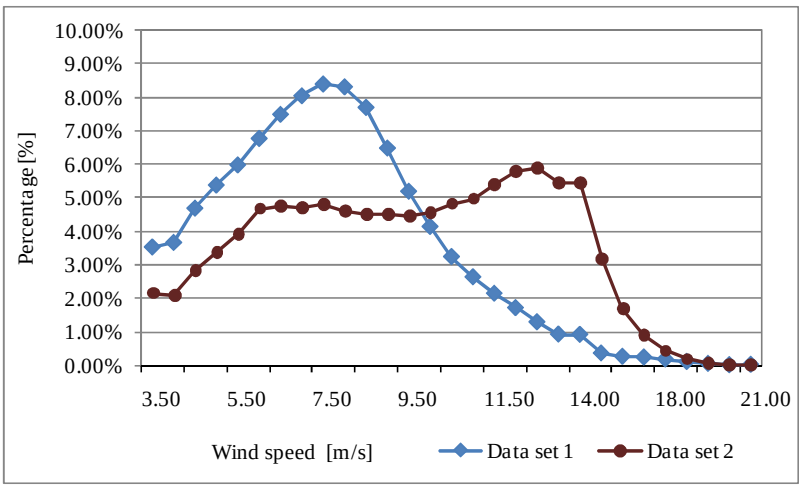

Figure 2.1. Distribution of the wind speed data.

Table 2.2. Training and test scenarios.

Scenario	Training Data	Test Data
1	Data set 1 Start Time 8/8/2007 12:00 AM End Time 08/12/2007 11:59PM	Data set 1 Start Time 8/13/2007 12:00 AM End Time 08/14/2007 11:59PM
2	Data set 1 Start Time 8/8/2007 12:00 AM End Time 08/14/2007 11:59PM	Data set 2 Start Time 9/22/2008 12:00 AM End Time 09/29/2008 11:59PM

Note that scenario 1 represents a typical case as data representing similar wind conditions is used to train and test performance of the models. Scenario 2 represents an extreme case as training and test data represent different conditions due to the data originating at a different year and a month. Using the data from similar time horizons is likely to produce more accurate results. However, the research reported in this paragraph is to demonstrate that the models developed and tested at disparate data sets produce useful results.

2.2.3 Wind Speed Similarity Metrics

The similarity of wind speed measured at different wind turbines is used as a basis for prediction of the wind speed of interest. Namely, if the wind speeds from several different turbines exhibit similar characters during a certain time period, they are used to build a model predicting wind speed at a turbine of interest. In this paragraph, Pearson's correlation coefficient defined in the next section is used to measure similarity between wind speeds at different locations.

The Pearson product-moment correlation coefficient (usually called Pearson's correlation coefficient), reflecting linear relationship between random variables, is used as a measure of correlation. The Pearson's correlation coefficient between wind speed of target turbine v_t and predictor turbine v_p is defined in (2.1) [64]:

$$r_{v_t,v_p} = \frac{\sum v_{t_i} v_{p_i} - n\overline{v_t}\,\overline{v_p}}{(n-1)s_{v_t}s_{v_p}} = \frac{n\sum v_{t_i} v_{pi} - \sum v_{t_i} \sum v_{p_i}}{\sqrt{n\sum {v_{t_i}}^2 - (\sum v_{t_i})^2}\sqrt{n\sum {v_{p_i}}^2 - (\sum v_{p_i})^2}} \tag{2.1}$$

where $\overline{v_t}$ and $\overline{v_p}$ are the sample means of v_t and v_p, S_{v_t} and S_{v_p} are the sample standard deviations of $\overline{v_t}$ and $\overline{v_p}$ and the sample number is *n*.

2.2.4. Wind Speed Prediction

In the section, a single-predictor and a multi-predictor model for wind speed estimation are discussed.

1. The Single-Predictor Model

The function $f_{estimation}(g)$ estimating wind speed $v_t(t)$ at time *t* at an interest turbine using the values of a single predictor is defined in (2.2).

$$v_t(t) = f_{estimation}(v_p(t), v_p(t-1), ..., v_p(t-k)) \quad (2.2)$$

where $f_{estimation}(g)$ is derived with the data mining algorithms (see Section 5.1), $v_p(t)$ is the wind speed of the predictor turbine at time period *t* and historical data for time periods $t-1,...,t-k$ at the turbine are used as input variables. The historical values of the wind speeds $v_t(t)$ and $v_p(t)$ of Eq. (2.2) need to satisfy the threshold inequality (2.3).

$$r_{v_t, v_p} \geq r_{threshold} \quad (2.3)$$

Equation (2.3) bounds the affinity of wind speed between the interest turbine and the predictor. The value of $r_{threhold}$ equals the maximum Pearson's correlation coefficient calculated from the training data set.

Based on Eq. (2.2), the model for prediction of wind speed at the turbine of interest at time $t+1$ is defined in (2.4).

$$v_t(t+1) = f_{estimation}(v_p^*(t+1), v_p(t), ..., v_p(t-k+1)) \quad (2.4)$$

As $v_p(t+1)$ is unknown at current time *t*, the time-series model (2.5) is used to generate the estimated value $v_p(t+1)^*$ in (2.5).

$$v_p(t+1)^* = g_{timeseriers,p}(v_p(t), ..., v_p(t-m)) \quad (2.5)$$

where *m* is the number of past time intervals. The model to predict wind speed for *q* steps ahead of time *t* can be computed from (2.6) that is derived from equations (2.4) and (2.5).

$$v_t(t+q)^* = f_{estimation}(v_p{}^*(t+q),...,v_p(t-k+q)) \quad (2.6)$$

where $v_p(t+q)^*$ is a dynamic-model shown in (2.7).

$$v_p(t+q)^* = g_{timeseries}(v_p{}^*(t-1+q),...,v_p(t-m+q)) \quad (2.7)$$

2. The Multi-Predictor Model

Function $f_{estimation,N}(g)$ estimating wind speed $v_{t,N}(t)$ at time *t* at a turbine of interest based on data collected at multiple wind turbines is defined in (2.8).

$$v_{t,N}(t) = f_{estimation,N}(v_{p1}(t),...,v_{p1}(t-k_1),v_{p2}(t),...,v_{p2}(t-k_2),...,v_{pN}(t),...,v_{pN}(t-k_N)) \quad (2.8)$$

where N is the number of wind turbines used in equation (9), k_N is the number of past time periods used for each turbine. Function $f_{estimation,N}(g)$ is built by the data mining algorithms (see Section 2.3.2). Similarly, the values of wind speeds in Eq. (2.8) satisfy inequality (2.9).

$$r_{v_t,v_{pN}} \geq r_{threshold} \quad (2.9)$$

Inequality (2.9) implies that all wind speeds used meet the similarity threshold $r_{threshold}$ which is the largest of *N* Pearson's correlation coefficients. Based on equation (2.8) and (2.9), the model for prediction of wind speed for the turbine of interest at time $t+1$ is defined as

$$\begin{aligned} v_{t,N}(t+1) = f_{estimation,N}(&v_{p1}(t+1)^*,...,v_{p1}(t-k_1+1),v_{p2}(t+1)^*,...,\\ &v_{p2}(t-k_2+1),...,v_{pN}(t+1)^*,...,v_{pN}(t-k_N+1)) \end{aligned} \quad (2.10)$$

The model for wind speed prediction using original (not transformed) data is illustrated in Figure 2.2. It involves multiple measurements of wind speeds at times $t-k_N,...,t$ from four anemometers as inputs to predict the wind speed of Turbine 9 at time t. The wind speed data from Turbine 3, 4, 13, and 26 are fed the NN model. In case when the past readings of wind speed from one anemometer only are used as one input, the model becomes a single-predictor model.

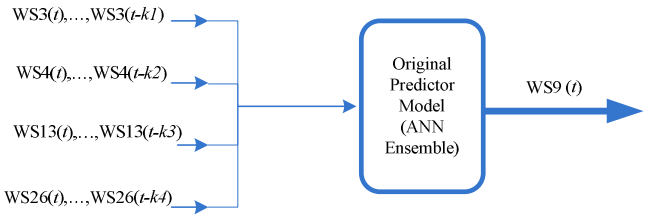

Figure 2.2 Multi-predictor wind speed model.

2.3 Industrial Case Study

2.3.1 Similarity between Wind Speeds

1. Similarity between wind speeds

In this paragraph, Turbine 9 is randomly selected from the set of 30 turbines for prediction of wind speed. The Pearson's correlation coefficient is computed between turbine of interest and other twenty nine turbines. The results of a seven day period for both data sets (see in Table 2.1) are shown in Table 2.3.

Table 2.3 Pearson's correlation coefficient between turbine of interest and twenty nine other turbines.

Turbine No.	Data set 1	Data set 2	Turbine No.	Data set 1	Data set 2
1	0.8558	0.8866	16	0.8273	0.8970
2	0.8600	0.8912	17	0.8654	0.9023
3	0.8742	0.8968	18	0.8591	0.8965
4	0.8893	0.8949	19	0.8276	0.8841
5	0.8558	0.8837	20	0.8530	0.8795
6	0.8382	0.8698	21	0.8454	0.8825
7	0.8603	0.8773	22	0.8686	0.8883
8	0.8436	0.8868	23	0.8346	0.8967
9	1.0000	1.0000	24	0.8326	0.8872
10	0.8485	0.8800	25	0.8702	0.8958
11	0.8330	0.8945	26	0.8955	0.8976
12	0.8650	0.8952	27	0.8424	0.8842
13	**0.8973**	**0.9130**	28	0.8354	0.8941
14	0.8529	0.8844	29	0.8273	0.8691
15	0.8659	0.8978	30	0.8255	0.8853

Figure 2.3 illustrates the Pearson's correlation coefficient between the turbine of interest and twenty nine remaining turbines. For data set 1, the Pearson's correlation coefficient is in the range [0.82, 0.89], while for data set 2 the Pearson's correlation coefficient is in the range [0.87, 0.91]. The Pearson's correlation coefficient corresponding to data set 2 is higher than that of data set 1. For both data sets, Turbine 13 shows strongest linear relationship with turbine of interest. Four turbines (Turbine 3, 4, 13, and 26) with the highest correlation coefficient values for data set 1 are selected for further analysis (in the circled area of Figure 2.3).

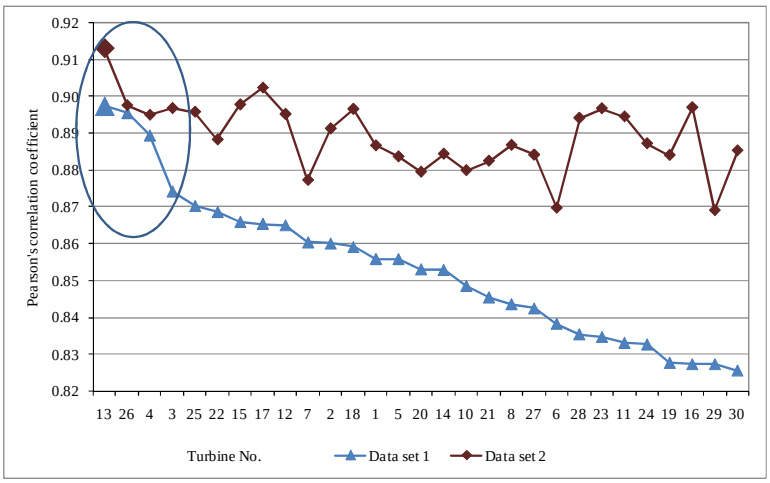

Figure 2.3 Pearson's correlation coefficient between Turbine 9 and 29 other turbines.

2. Correlation Coefficient in Time

To examine the consistency correlation between wind speeds at different locations in different time period, the Pearson's correlation coefficient of Turbine 9 is examined at seven separate days for data sets 1 and 2 of Table 2.1.

1) Data set 1

Table 2.4 shows the values of Pearson's correlation coefficient for Turbine 9 and the four selected turbines 3, 4, 13, and 26 from Aug. 8 through Aug.15 based on data set 1. The highest value of the correlation coefficient in Table 2.4 is 0.9467 and the lowest value is 0.4905.

The highest Pearson's correlation coefficient in four (Aug. 8, Aug. 11, Aug, 12, and Aug.14) out of the seven days and the highest average correlation coefficient 0.8186 across the seven days correspond to Turbine 13.

Table 2.4. Pearson's correlation coefficient for Turbine 9 and the four selected turbines over a seven day period based on data set 1.

Turbine No.	Aug. 8	Aug. 9	Aug.10	Aug.11	Aug.12	Aug.13	Aug.14	Seven Day Average
3	0.8887	0.8608	0.7554	0.9274	0.8766	0.4247	0.8055	0.7913
4	0.9019	0.8218	0.7772	0.9397	0.9180	0.3460	0.8271	0.7902
13	0.9043	0.8868	0.7692	**0.9467**	0.9268	**0.4905**	0.8058	**0.8186**
26	0.9071	0.8984	0.7766	0.932	0.897	0.4568	0.8296	0.8139
Average Across 4 Turbines	0.9005	0.867	0.7696	0.9365	0.9046	0.4295	0.8170	

2) Data set 2

For the data set 2 of Table 2.1, the strongest linear relationship (0.9378) between Turbine 13 and Turbine 9 is on Sept. 14 and the weakest linear relationship (0.5518) is on Sept. 25 as shown in bold in Table 2.5.

Table 2.5 Pearson's correlation coefficient for Turbine 9 and the four selected turbines over a seven day period based on data set 2.

Turbine No.	Sept.22	Sept.23	Sept.24	Sept.25	Sept.26	Sept.27	Sept.28	Seven Day Average
3	0.7601	0.8336	0.8922	0.6395	0.7365	0.7696	0.7839	0.7736
4	0.7197	0.8242	0.9145	0.5520	0.7300	0.8173	0.8448	0.7718
13	0.7996	0.8389	**0.9378**	**0.5518**	0.7893	0.8398	0.8666	**0.8034**
26	0.7715	0.8159	0.9344	0.5733	0.6934	0.8095	0.8187	0.7738
Average Across 4 Turbines	0.7627	0.8282	0.9197	0.5792	0.7373	0.8091	0.8285	

As illustrated in Table 2.4 and Table 2.5, the Pearson's correlation coefficient varies on daily basis. Turbine 13 and Turbine 9 exhibit most similar wind speeds for both data sets, i.e., for 10 out of 14 days (4 days in data set 1) and 6 days (all but Sept 25) in data set 2. The highest average Pearson's correlation coefficients (0.8186 for data set 1 and 0.8034 for data set 2) correspond to the same pair of turbines.

2.3.2 Model Extraction

1. Algorithm Selection

Models derived by various data mining algorithms result in different accuracy of predicted wind speeds. To select the most appropriate algorithm, a model is built using data from Turbine 13 to predict wind speed of Turbine 9. The training data is made of 10% randomly selected wind speeds from data set 1 and the test data constitutes 5% randomly selected wind speeds from data set 2 (both from Table 1). Any model built (trained) and tested using data sets with different characteristics (see Table 1) that performs well must robust and transferable among different applications. The performance of models developed from a subset from data set 1 by the following four data mining algorithms: Random Forest Algorithm (RFA), Multiple Layer Perceptron (MLP) Neural Network (NN) Ensemble, Support Vector Machine (SVM), and the Boosting Tree Algorithm and tested on a subset from data set 2 of Table 1 is reported in Table 2.6.

The best performing algorithm is the Multiple-Layer Percepton (MLP) Ensemble, initiated with thirty neural networks (NNs). The minimum number of units in the hidden layer of the MLP NN is equals the number of input parameters and the maximum number of units in the hidden layer is set as three times that number. The best five neural networks are used to construct the neural network ensemble, i.e., for each experiment $f_{estimation}(g)$ and $f_{estimation,N}(g)$ are generated by the neural network ensemble.

Table 2.6. Performance of four data mining algorithms.

Algorithm Name	Observed Average Wind Speed [m/s]	Predicted Average Wind Speed [m/s]	Mean Absolute Error (MAE) [m/s]	STD of MAE	Mean Relative Error (MRE) [%]	STD of MRE
Random Forrest Regression	8.7524	8.7610	1.0607	0.9667	16.0072	33.2202
Multiple-Layer Percepton (Ensemble)	8.7524	8.8042	1.0661	0.9311	15.1824	23.1501
Support Vector Machine	8.7524	8.6201	1.4781	1.0000	24.6937	30.7573
Boosting Tree	8.7524	8.7667	1.0732	0.9436	17.0813	38.2927

2. Dimensionality Reduction

In the proposed models (Eqs (2.2) and (2.8)) only one parameter (wind speed) is considered as input. The input dimensionality increases with the increase number of past states of the wind speeds and the number of turbines. High dimensional data input imply expensive computation, and may decrease prediction accuracy. In this paragraph, wind speeds from multiple turbines (N = 2 to 4) measured up to 12 past states are used as input parameters for model extraction. For example, a model derived from data of two wind turbines calls for 26 (= 2 parameters $\times$(12 past states + 1 current state)) inputs; model based on data from three turbines calls for 39 inputs, and four turbines for 52 inputs. In some cases, training a model on three-turbine data has taken over 24 hours, and for four turbines more than 48 hours. Therefore, the data dimensionality reduction is considered.

Principal component analysis (PCA) is an efficient way for data dimensionality reduction. It transforms a number of possibly correlated variables into a smaller number of uncorrelated variables called principal components. The first principal component accounts for as much of the variability in the data as possible, and each succeeding component accounts for as much of the remaining variability as possible. For computational analysis, ten factors are selected to form training data for each of the three

cases (with two, three, and four turbines). The model developed with the PCA derived inputs is called here the transformed-predictor model (see Figure 2.4).

Figure 2.4 illustrates the transformed-predictor model. Data from Turbine 3, 4, 13, and 26 are transformed by PCA and the first 10 factors are used as inputs to the NN model.

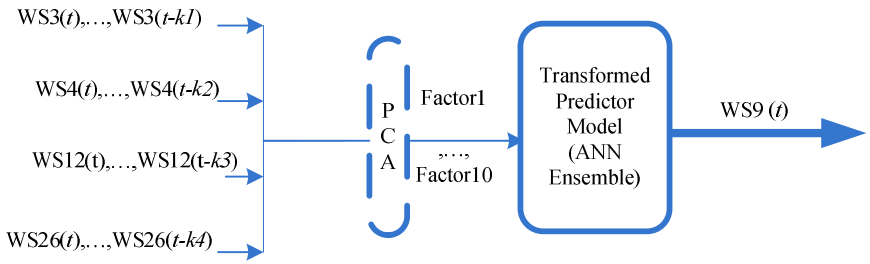

Figure 2.4 Transformed-predictor model.

2.3.3 Computational Results

To demonstrate utility of the model proposed in this paragraph, 20 different experiments have been designed in Table 2.7. In experiments 1 through 10 consequent training and test data subsets originating from data set 1 are used. The training data covers the five day period from Aug. 8 to Aug. 12, and the test data covers the two day period Aug. 13 to Aug. 14. In experiments 11 through 20, the training and test days represent different wind regimes. Data set 1 of Table 1 is used to create training data subsets, while data set 2 is used to test the models. Note that this represents an extreme situation where the training and test data sets originate at disparate data environments outlined in Table 1. Note that the numbers in the column Experiment Number of Table 2.7 are used in Table 2.8 and 2.9.

Table 2.7. Description of twenty computational experiments.

Training and Test Scenario 1				Training and Test Scenario 2			
Experiment Number	Number of Predictors	Turbine Set	PCA Used	Experiment Number	Number of Predictors	Turbine Set	PCA Used
1	1	3	No	11	1	3	No
2	1	4	No	12	1	4	No
3	1	13	No	13	1	13	No
4	1	26	No	14	1	26	No
5	2	4 & 13	No	15	2	4 & 13	No
6	2	4 & 13	Yes	16	2	4 & 13	Yes
7	3	3, 4 & 13	No	17	3	3, 4 & 13	No
8	3	3, 4 & 13	Yes	18	3	3, 4 & 13	Yes
9	4	3, 4,13 & 26	No	19	4	3, 4,13&26	No
10	4	3, 4,13&26	Yes	20	4	3, 4,13&26	Yes

Several metrics are used to measure prediction accuracy of the models involved in 20 experiments of Table 2.7, including the Mean Absolute Error (MAE), Mean Relative Error (MRE) and Standard Deviation (STD) of MAE and MRE. The Mean Absolute Error is defined in (2.11).

$$MAE = = \frac{\sum_{t \in Data} \left| v_t^*(t) - v_t(t) \right|}{|Data|} \tag{2.11}$$

Here $|Data|$ denotes the number of data points in training or test data set; $v_t^*(t)$ is the predicted wind speed at time t, and $v_t(t)$ is the observed wind speed. The Mean Relative Error is defined in (2.12).

$$MRE = \frac{\sum_{t \in Data} \left| \frac{v_t^*(t) - v_t(t)}{v_t(t)} \right| \times 100\%}{|Data|} \tag{2.12}$$

Four analyses are performed next: performance evaluation of a single-predictor model, performance evaluation of a multi-predictor model, model accuracy for extended number of future prediction horizons, and model performance on a daily basis.

1. Performance of the Single-Predictor Model

The data from each of the four turbines 3, 4, 12, and 26 (with the highest Pearson's correlation coefficient, see Figure 2.2) is used to form a single-predictor model (see equation (2.2)). For all the experiments list in this section, the number of past states wind speed $k = 12$. Experiments 1 through 4 in Table 2.8 represent the training and test scenario 1 of Table 2.2, while Experiments 11 through 14 in Table 8 present the results for training scenario 2 of Table 2.2. Note that the results in Table 2.8 use all the data in sets 1 and 2 for the turbines 3, 4, 13, and 26 of Table 2.2. The experiment number shown in Table 2.8 is quoted from Table 2.7 where each experiment is assigned a specific number.

Table 2.8 Test performance of single-predictor models derived from four different turbines.

Experiment Number	Turbine Number	Pearson's Correlation Coefficient	Observed Mean Wind Speed [m/s]	Predicted Mean Wind Speed [m/s]	Mean Absolute Error (MAE) [m/s]	STD of MAE	Mean Relative Error (MRE) [%]	STD of MRE
1	3	0.6807	8.2418	8.1299	1.0770	0.8766	14.0463	13.5380
2	4	0.6825	8.2398	8.2423	1.0336	0.8832	13.4101	13.0629
3	**13**	**0.6917**	**8.2398**	**7.9743**	**1.0374**	**0.9139**	**12.9100**	**11.6481**
4	26	0.7126	8.2398	8.2988	1.0467	0.8581	13.8296	13.5764
11	3	0.8964	8.7043	8.2642	1.2093	1.0207	17.2954	28.9083
12	4	0.8945	8.7062	8.6064	1.1747	1.0248	17.5526	35.4587
13	**13**	**0.9126**	**8.7062**	**8.6919**	**1.0878**	**0.9485**	**16.0009**	**27.6607**
14	26	0.8972	8.7035	8.5800	1.1413	0.9940	18.2277	40.9673

The results in Table 2.8 demonstrate that for both training and test scenarios of Table 2.2, the best performing model is that of Turbine 13. It turns out that the wind speed of Turbine 13 is the most highly correlated to that Turbine 9 (marked in bold in Table 2.8).

Figure 2.5 and 2.6 benchmark performance of the models built and tested in the eight experiments presented in this section. In general, the models tested in experiments 1 through 4 perform better than those of experiment 11 through 14, which implies that the training and test scenario 1 outperform scenario 2 of Table 2.2.

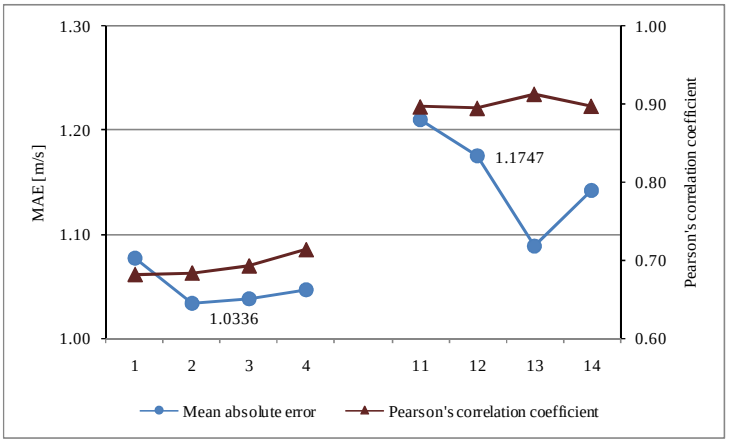

Figure 2.5. MAE and Pearson's correlation coefficients for the eight experiments of Table 2.8.

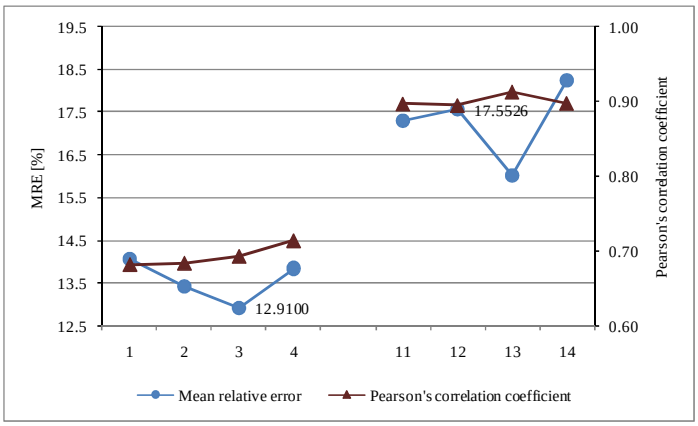

Figure 2.6 MRE and Pearson's correlation coefficients
for the eight experiments of Table 2.8.

The data used in experiments 1 to 4 represent a continuum of wind conditions. Among the first four experiments in Table 2.8, the model built in experiment 2 has the lowest MAE and best value of MRE is attained in experiment 3. Overall, best performance among the four experiments is for model of experiment 3, with the MAE value similar to experiment 2 and better MRE.

The best performing model (lowest values of MAE and MRE) among the experiments 11 through 14 (Table 2.8) was derived in experiment 13. The wind speed in this data set is most highly correlated to the wind speed of Turbine 9.

Overall, performance of all models built and tested in the eight experiments of Table 2.8 is satisfactory. The computational results show that the wind data collected at neighborhood turbines can be used to predict wind speed at any turbine. Using data from a similar wind conditions leads to better prediction performance of wind speed.

2. Performance of the Multi-Predictor Model

In this section, performance of the multi-predictor model (see equation (2.8)) tested in six experiments is discussed. The number of wind turbines *N* varies from 2 to 4. The number of past states wind speed used for each turbine equals 12. Two training and test scenarios (see Table 2.2) are applied for each set of turbines. Performance of the neural network model built and tested in six experiments is illustrated in Table 2.9. The experiment numbers used in Table 2.9 (the first column) are identical to those used in Table 2.7. The data stream used by each NN model is represented by the turbines listed in each turbine set (the second column) in Table 2.9.

Table 2.9 Performance of six multi-predictor models.

Experiment Number	Turbine Set	Observed Mean Wind Speed [m/s]	Predicted Mean Wind Speed [m/s]	Mean Absolute Error (MAE) [m/s]	STD of MAE	Mean Relative Error (MRE) [%]	STD of MRE
5	4 &13	8.2398	8.1401	0.9597	0.8387	12.1422	11.3621
7	4, 13 & 26	8.2398	8.1835	0.9408	0.8102	12.0076	11.2488
9	**3, 4, 13 & 26**	**8.2418**	**8.1147**	**0.9129**	**0.7812**	**11.5782**	**10.6893**
15	4&13	8.7062	8.7214	1.0578	0.9365	15.1655	26.3199
17	**4, 13 & 26**	**8.7043**	**8.7092**	**1.0161**	**0.9022**	**14.8205**	**26.1021**
19	*3,4,13 & 26*	*8.7043*	*8.7008*	*1.0352*	*6.9128*	*14.4829*	*64.8171*

In general, performance of the models tested in of experiments 5, 7 and 9 is better than those of experiments 15, 17 and 19. This indicates that the training and test scenario 1 provides better accuracy results than scenario 2 of Table 2.2. Based on scenario 1 data, the model of experiment 9 (shown in bold in Table 2.9) involving four predictors performs best. The model of experiment 17 is most accurate of all the models in scenario 2. In this scenario, a model using 4 predictors provided inferior MAE tend to be worse than the model with 3 predictors (marked in italics in Table 2.9).

Figure 2.7 compares the test performance of each experiment listed in Table 2.9. For the models tested in experiments 5, 7 and 9, the prediction accuracy improved with increase of number of predictors *N*. While for experiment 15, 17 and 19, the best performance is for experiment 17 (using three predictors), while in experiment 19, the MAE decreases when 4 predictors are used. This decrease in accuracy might due to the high data dimensionality.

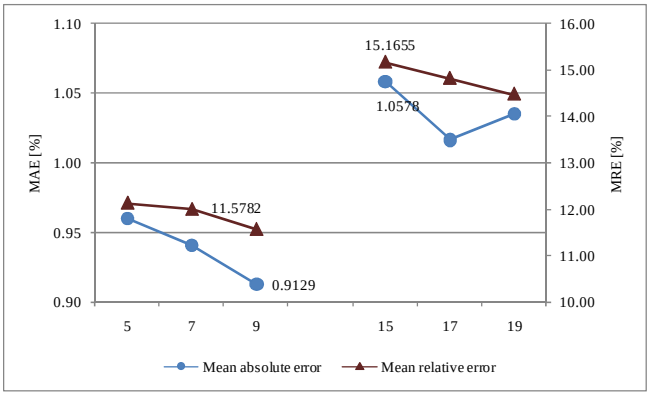

Figure 2.7 MAE/MRE vs the number of predictors.

To reduce dimensionality of the input, the transformed-predictor model (see Figure 2.4) is proposed. The original wind speed data is transformed with the PCA algorithm in 10 factors that form a new training and test data sets. Unfortunately, using multiple turbines might not show benefits in predicting wind speed if PCA analysis is used to reduce the dimension of the data. Thus, this paragraph does not the report the test performance using transformed predictor model in detail.

3. Prediction Accuracy for Different Time Horizons

To test the consistency with model accuracy with steps ahead, Turbine 13 is used as the predictor. Wind speed prediction of Turbine 13 is achieved by a dynamic-model of

Eq. (2.7) for m =12 and q =13. Table 2.10 illustrates the prediction results of wind speed at Turbine 13 at the current time t for 13 future time periods, from t + 1 to t +13.

Table 2.10 Prediction of wind speed of Turbine 9 at intervals t, t +1,.., t +13.

Time [s]	Observed Mean Wind Speed [m/s]	Predicted Mean Wind Speed [m/s]	Mean Absolute Error (MAE) [m/s]	STD of MAE	Mean Relative Error (MRE) [%]	STD of MRE
t	7.9529	7.9483	0.5205	0.4754	6.9455	7.4862
t + 1	7.9528	7.9353	0.3644	0.3440	4.7290	5.0204
t + 2	7.9528	7.9265	0.4253	0.3844	5.7230	6.8423
t + 3	7.9527	7.9176	0.4569	0.4157	6.0375	5.8765
t + 4	7.9527	7.9093	0.4713	0.4287	6.5124	10.1173
t + 5	7.9526	7.9006	0.4928	0.4449	6.5547	6.4411
t + 6	7.9526	7.8930	0.5309	0.4773	7.0513	6.8277
t + 7	7.9525	7.8849	0.5703	0.5115	7.8757	12.1101
t + 8	7.9525	7.8761	0.6031	0.5370	8.3566	13.0294
t + 9	7.9524	7.8673	0.6263	0.5572	8.4087	8.4188
t + 10	7.9523	7.8589	0.6470	0.5748	9.0531	15.2048
t + 11	7.9523	7.8503	0.6698	0.5918	9.4184	16.3600
t + 12	7.9523	7.8410	0.6866	0.6073	9.6938	17.3773
t + 13	7.9522	7.8410	0.7005	0.6227	9.9315	18.4168

The results in Table 2.10 are used in the prediction of wind speed at Turbine 9, where the predicted wind speed of Turbine 13 is incorporated in equation (2.6) to derive the wind speed of Turbine 9 for $q = 1$ to q=13 . The prediction results of 13 future periods are shown in Table 2.11.

As illustrated in Table 2.11, for each step ahead, MAE increases less than 0.01 m/s and MRE increases less than 0.1%. The values of MAE and MRE are illustrated in Figure 8. Note that at time period $t+13$ all values of the input parameters are generated by the data-mining algorithms, rather than using raw data measured at wind turbines. The

accuracy of the results is acceptable up to period t + 13, i.e., 130 seconds which is acceptable in practice.

Table 2.11 Prediction of wind speed of Turbine 9 intervals t, t +1,.., t +13.

Time [s]	Observed Mean Wind Speed [m/s]	Predicted Mean Wind Speed [m/s]	Mean Absolute Error (MAE) [m/s]	STD of MAE	Mean Relative Error (MRE) [%]	STD of MRE
t	8.2398	7.9743	1.0374	0.9139	12.9100	11.6481
t + 1	8.2397	7.9347	1.0482	0.9209	12.9924	11.5431
t + 2	8.2396	7.9165	1.0520	0.9236	13.0159	11.5285
t + 3	8.2396	7.9067	1.0602	0.9293	13.1148	11.6352
t + 4	8.2395	7.8739	1.0729	0.9391	13.2446	11.6892
t + 5	8.2394	7.8608	1.0807	0.9425	13.3329	11.7514
t + 6	8.2393	7.8419	1.0917	0.9476	13.4626	11.8093
t + 7	8.2393	7.8303	1.1007	0.9526	13.5578	11.8762
t + 8	8.2392	7.8243	1.1056	0.9553	13.6155	11.9323
t + 9	8.2392	7.8106	1.1140	0.9599	13.7016	11.9490
t + 10	8.2391	7.7947	1.1203	0.9671	13.7468	11.9613
t + 11	8.2390	7.7814	1.1266	0.9763	13.8127	12.0208
t + 12	8.2390	7.7719	1.1224	0.9732	13.7385	11.9500
t + 13	8.2389	7.7588	1.1280	0.9779	13.7889	11.9885

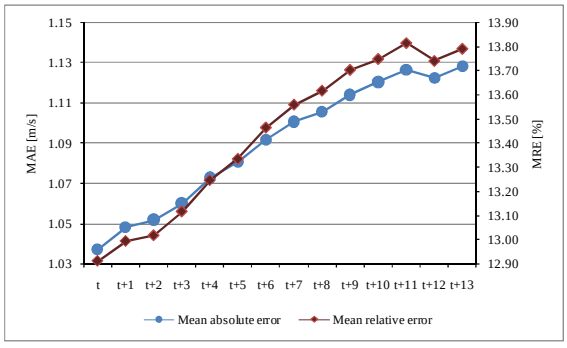

Figure 2.8 Model accuracy with 13 periods.

4. Daily Performance Analysis

The data sets from Sept. 24 to Sept. 26 are used for testing. The smallest MAE corresponds to Sept. 24 and the largest MAE to Sept. 25. The difference between the two values is 0.94 m/s. Note that the observed mean wind speed on Sept. 24 is much lower than on the other two days. This might be the reason for a lower MAE but a higher MRE on Sept. 24. Comparing the results for Sept. 25 and Sept. 26, the higher Pearson's correlation coefficient (Sept. 26) leads to higher prediction accuracy.

Table 2.12. Daily performance of Turbine 13 for test data set 2 of Table 2.2.

Date	Pearson's Correlation Coefficient	Observed Average Wind Speed [m/s]	Predicted Average Wind Speed [m/s]	Mean Absolute Error (MAE) [m/s]	STD of MAE	Mean Relative Error (MRE) [%]	STD of MRE
Sept. 24	0.9378	5.0632	4.8431	0.6957	0.5710	28.4732	63.8186
Sept. 25	0.5518	10.9253	11.4182	1.6290	1.1796	17.1483	16.0525
Sept. 26	0.7893	11.4124	11.0768	1.1952	0.8936	11.4205	11.6497

2.4 Summary

In this section, models for prediction of wind speed at a turbine of interest using wind speeds from other turbines were presented. Rather than using the wind speed data from random turbines, turbines sharing similar wind speeds were selected. The wind speed similarity was computed using Pearson's correlation coefficient. In the single-predictor model, wind speed was measured at a turbine sharing most similar wind conditions with the turbine of interest was used as predictor. In the multi-predictor models, wind speeds measured at two to four turbines are were used as predictors. To reduce the input data dimensionality, a predictor-transformed model was used. Unlike the previous two models, in this model wind speeds measured at two, three, or four turbines were the PCA technique before becoming the predictors.

The computational analysis demonstrated that the higher the Pearson's correlation coefficient, the higher the prediction accuracy for most experiments. Though the Pearson's correlation coefficient varied, the prediction accuracy remained relatively stable. The increase in the number of predictors has led to increased prediction accuracy. However, the increase in the number of predictors has led to the excessive training time without accuracy gain. The input dimensionality was reduced with the PCA technique, yet did not offer accuracy benefits often seen in other applications. All models discussed in the paragraph used only one parameter (wind speed), and therefore they are easy to apply.

CHAPTER 3.
SHORT TERM PREDICTION OF WIND TURBINE PARAMETERS

3.1 Introduction

The large-scale wind energy industry is relatively new and is rapidly expanding [22]. The ability of a wind turbine to extract power from the wind is a function of three main factors: the measured wind speed, the power curve of the turbine, and the ability of the machine to handle wind fluctuations [38]. The key parameter determining wind turbine performance is wind speed.

Wind speed is normally measured with an anemometer placed at the nacelle of a turbine. In some cases, in addition to turbine anemometers, meteorological towers are used to provide additional measurements of wind speed. However, these additional wind speed measurements are not used directly to control individual turbines, rather they are applied for assessment of wind speed.

Considering the fact that wind speeds and wind turbine performance vary across different turbine locations at a wind farm, the question arises as to whether a generalized model (called in this paragraph a virtual model) of a wind turbine could be developed. Such a virtual model has been developed based on SCADA data collected at wind turbines. As a wind turbine is a complex system, two aspects of a wind turbine are reported in this paragraph, the power output and the rotor speed. However, the methodology presented here can be applied to modeling many aspects of a wind turbine. Predicting the power output demonstrates the capability of the virtual model to improve performance of a wind turbine, while predicting the rotor speed points to the utility of the virtual model to improve the lifetime of turbine components, e.g., the gearbox.

The literature on data mining in wind energy has primarily focused on estimating and optimizing the power output. A review of the literature on forecasting wind speed

and generated power using both physical models and data-mining methods is presented in [24]. An approach to optimizing power by controlling generator torque is discussed in [38]. Optimization of power output and operational performance is reported in [39, 65].

Model building can be accomplished with a variety of learning algorithms, e.g., neural networks [11, 13]. A neural network was applied in [32] to estimate power output as a function of the time delay of wind speed and the power itself. However, none of the published paragraphs has focused on virtual models to predict any measurable parameter of interest, e.g., the power output and the rotor speed.

Data mining offers algorithms for finding patterns and relations in extensive data [38] using machine-learning algorithms. It is widely recognized that data preprocessing is a time-consuming step; for example, 80% of the time involved may be spent on data sampling, feature selection, and so on [69]. Methods such as simple random sampling and stratified sampling [70] can be used. Feature (parameter) selection is also regarded as an important task in data mining, and some algorithms have been proven to be effective [71, 72] in determining relevant parameters.

3.2 Data Description and Methodology for Short Term Prediction of Wind Turbine Parameters

3.2.1 Methodology for Developing Virtual Models

The methodology for developing virtual models to predict interest aspects of wind turbines includes three phases: data preprocessing, model extraction, and model validation (see Figure 3.1).

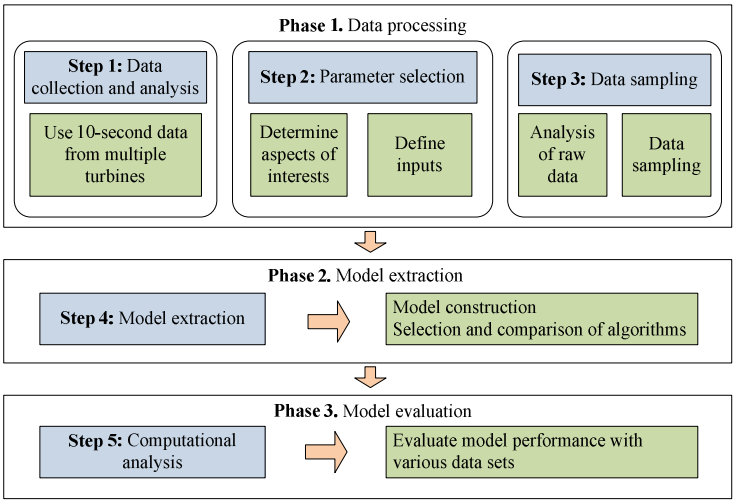

Figure 3.1 Methodology for developing virtual models.

The steps of the methodology outlined in Figure 3.1 are described next.

Step 1: Data collection and analysis

First, it is necessary to explore the content of the raw data as it is used in modeling. For example, data formats and frequency need to be preprocessed for uniformity. Any data that is incomplete, in error, or missing needs to be dealt with.

Step 2: Feature selection

Feature selection is considered from two perspectives: domain knowledge and data mining. In terms of domain knowledge, all the parameters of a wind turbine system can be classified in three categories:

- Controllable parameters, e.g., blade pitch angle, generator torque
- Non-controllable parameters, e.g., wind speed
- Turbine performance parameters, e.g., power output, rotor speed.

Controllable parameters and non-controllable parameters constitute inputs to the virtual models, while the performance parameters create the outputs predicted by the models.

The impact of input parameters on the output (performance) parameter varies. Some insignificant parameters are easy to eliminate based on the domain knowledge. The impact of input parameters on the performance parameter is ranked ordered by data-mining algorithms.

Step 3: Data sampling

Data sampling is a commonly used approach for selecting a subset of data from a large volume of data. In this paragraph, data sampling is performed according to the range of wind speed, which is the only non-controllable parameter available in the data set. This sampling strategy leads to a data sample that is representative across different wind speed ranges.

Step 4: Model extraction

Different data-mining algorithms are used to extract models. The model which performs the best is selected.

Step 5: Computational analysis

In this paragraph, three types of datasets with different characteristics are used to evaluate the performance of the model extracted by Step 4.

3.2.2 Data Collection and Analysis

Wind turbines are equipped with sensors providing various measurements, including wind speed, power output, generator torque, and so on. Some of these measurements can be used to control and monitor the performance of wind turbines. In this paragraph, data from 30 turbines generated at 10-second intervals from two time periods is used. One time period provides data for high wind speed, and the other provides data for low wind speed.

The analysis performed in the research shows that the data collected on the same parameters across different turbines of the same wind farm exhibit different

characteristics. To illustrate the data variability, four random turbines of the same type have been selected. Figure 3.2 shows the wind speed recorded by the Supervisory Control and Data Acquisition (SCADA) system of these turbines. It is clear that the range of wind speed for turbine 2 is significantly different from the other three turbines. The wind speed of turbine 1, turbine 3, and turbine 4 is higher than the cut-in speed of 3.5 m/s for this turbine type. The wind speed of turbine 2 is below the cut-in speed, which indicates that this turbine could not produce power, as opposed the other three turbines.

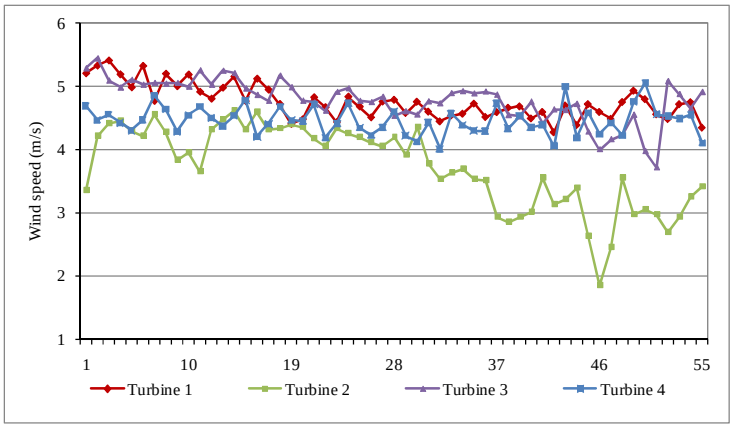

Figure 3.2 Comparison of wind speeds of four different wind turbines.

Most turbine sensors provide high frequency measurements that are usually averaged into higher frequency data, e.g., 10-minute average data. Some users set the SCADA system to store higher frequency data, e.g., 2 seconds.

Figure 3.2 shows the power curve for the four turbines, which not only looks different from the ideal power curves, but also shows distinct characteristics. It is obvious that the negative power output of turbine 2 indicates that this turbine is consuming (e.g., power electronics) rather than producing energy. The power of the remaining three turbines differs in ranges and shapes.

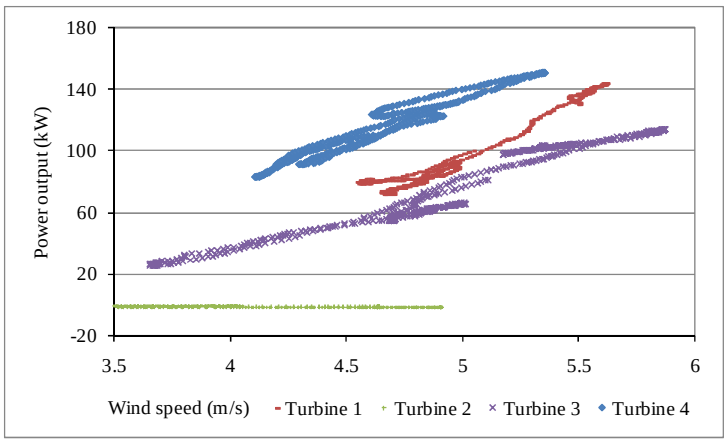

Figure 3.3 Power outputs of four turbines based on 10-minute data.

Figure 3.4 illustrates the power output generated from 10-second data. Note that the data used in Figure 3.3 was obtained by averaging the data used in Figure 3.4. Although the power ranges of turbine 1, turbine 3, and turbine 4 differ, they share similar shapes.

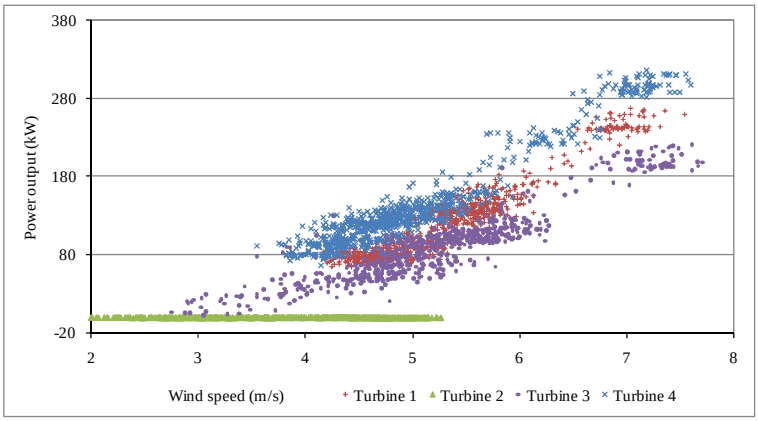

Figure 3.4 Power outputs of four turbines based on 10-second data.

To ensure that the behavior of the turbines is adequately reflected in the model, the 10-second data is used for analysis. Of the 30 turbines considered in this research, the data from 10 randomly selected turbines constitutes a training set, and the data from all 30 turbines is used to test the proposed methodology.

3.3 Industrial Case Study on Virtual Models for Prediction of Wind Turbine Parameters

3.3.1 Parameter Selection

The data provided from the SCADA system spans over 100 parameters. The SCADA parameters used in this research are listed in Table 3.1. Note that the relationships between some parameters on this list are well defined (e.g., power output, torque, and speed), while others may not be obvious or may not even exist. The selection (ranking) of parameters will be performed by algorithms rather than using different principles.

Table 3.1. Selected SCADA parameters.

Parameter Name	Abbreviation
Power output	PO
Generator torque	GT
Generator speed	GS
Wind speed	WS
Generator bearing A temperature	GBAT
Generator bearing B temperature	GBBT
Drive train acceleration	DTA
Blade pitch angle	BPA
Nacelle position	NP
Rotor speed	RS

To illustrate the methodology presented in Section 3.2.1 (see Figure 3.1), the following three performance parameters have been selected: power output, rotor speed, and generator speed. As generator speed and rotor speed are linearly dependent, only the rotor speed is selected for modeling. The virtual models are then built for the power output and the rotor speed.

After deletion of any low quality data of turbine 1, e.g., negative power outputs, two algorithms (boosting tree and neural network) were used to rank order the parameters that could be potentially used to predict power output and rotor speed (see Table 3.2 and Table 3.3). Note that both tables include parameters that can be controlled (e.g., blade pitch angle) and those that cannot be controlled (e.g., wind speed).

Table 3.2. Parameters' importance in predicting the power output.

Boosting Tree		Neural Network	
Input Parameter	Importance	Input Parameter	Importance
Generator torque	1.00	Generator torque	172.09
Rotor speed	0.96	Wind speed	5.04
Wind speed	0.90	Rotor speed	4.31
Drive train acceleration	0.64	Blade pitch angle	1.32
Blade pitch angle	0.47	Generator bearing B temperature	1.24
Generator bearing A temperature	0.29	Generator bearing A temperature	1.13
Generator bearing B temperature	0.27	Drive train acceleration	1.01
Nacelle position	0.09	Nacelle position	1.00

Based on the value of importance in Table 3.2 and Table 3.3 and the control logic of a wind turbine, two controllable parameters are selected: the generator torque and the blade pitch angle. The drive train acceleration is not selected, as it is not directly

controlled, and it is determined by the generator torque and blade pitch angle. The only non-controllable parameter, wind speed, is also selected. Table 3.4 lists the parameters used for building a virtual model.

Table 3.3. Parameters' importance in predicting the rotor speed.

Boosting Tree		Neural Network	
Input Parameter	Importance	Input Parameter	Importance
Power output	1.00	Power output	496.67
Generator torque	0.98	Generator torque	96.47
Wind speed	0.84	Drive train acceleration	1.10
Drive train acceleration	0.72	Wind speed	1.02
Generator bearing A temperature	0.33	Generator bearing A temperature	1.01
Generator bearing B temperature	0.27	Blade pitch angle	1.00
Blade pitch angle	0.20	Nacelle position	1.00
Nacelle position	0.12	Generator bearing B temperature	1.00

Table 3.4. Wind turbine parameters of the virtual model.

Parameter	Name	Unit
v	Wind speed (WS)	m/s
x_1	Blade pitch angle (BPA)	°
x_2	Generator torque (GT)	Nm
y_1	Wind turbine power output (PO)	kW
y_2	Rotor speed (RS)	rpm

The wind turbine manufacturer data shows that the maximum generator speed is 1600 rpm, the maximum rotor speed is 23 rpm, the maximum power output is 1600 kW,

the generator torque is limited to 10090 Nm, and the maximum generator torque change rate is 4500 Nm/s.

1. Impact of the past states

Due to the dynamic nature of the wind energy conversion process, it is necessary to consider the time-based values of input parameters discussed next.

1) Impact of the past values of non-controllable parameters

The only non-controllable parameter considered in this research is wind speed. The boosting tree and the neural network algorithms are used to determine the significance of different past states of the wind speed v, i.e., v at time t, t - 1, t - 2, until t - 9, in predicting the power output and rotor speed. The importance scores are shown in Table 3.5 and Table 3.6.

Table 3.5. Wind speed importance in predicting the power output.

Boosting Tree		Neural Network	
Input Parameter	Importance	Input Parameter	Importance
$v(t)$	1.00	$v(t)$	2.23
$v(t\text{-}1)$	0.96	$v(t\text{-}1)$	1.29
$v(t\text{-}2)$	0.94	$v(t\text{-}2)$	1.05
$v(t\text{-}3)$	0.92	$v(t\text{-}3)$	1.01
$v(t\text{-}4)$	0.90	$v(t\text{-}4)$	1.01
$v(t\text{-}5)$	0.89	$v(t\text{-}6)$	1.00
$v(t\text{-}6)$	0.87	$v(t\text{-}5)$	1.00
$v(t\text{-}7)$	0.87	$v(t\text{-}7)$	1.00
$v(t\text{-}8)$	0.86	$v(t\text{-}8)$	1.00
$v(t\text{-}9)$	0.85	$v(t\text{-}9)$	1.00

Table 3.6. Wind speed importance in predicting the rotor speed.

Boosting Tree		Neural Network	
Input Parameter	Importance	Input Parameter	Importance
$v\,(t\text{-}4)$	1.00	$v\,(t\text{-}8)$	1.00
$v\,(t\text{-}9)$	0.99	$v\,(t\text{-}9)$	1.00
$v\,(t\text{-}8)$	0.99	$v\,(t\text{-}2)$	1.00
$v\,(t\text{-}6)$	0.99	$v\,(t\text{-}1)$	1.00
$v\,(t\text{-}1)$	0.99	$v\,(t\text{-}7)$	1.00
$v\,(t\text{-}2)$	0.98	$v\,(t)$	1.00
$v\ (t)$	0.96	$v\,(t\text{-}6)$	1.00
$v\,(t\text{-}7)$	0.95	$v\,(t\text{-}3)$	1.00
$v\,(t\text{-}5)$	0.95	$v\,(t\text{-}5)$	1.00
$v\,(t\text{-}3)$	0.94	$v\,(t\text{-}4)$	1.00

When predicting the wind turbine power, the importance of the wind speed at the previous states is arranged in time sequence (see Table 3.5). However, when predicting the rotor speed, the order of importance deviates from the time sequence (see Table 3.6). The two algorithms, the boosting tree and the neural network, produced different sequences; however, the importance scores do not significantly differ by either algorithm. Therefore, the values v (t), v (t-1), v (t-2) and v (t-6) (in essence the wind speed at the prior minute measured at 10-second intervals) are selected.

2) Impact of the past values of controllable and performance parameters

The impact of the input parameters measured at past intervals on the future state of the turbine was shown in [12]. The model governing the relationship between the past and future parameters is not known. In this paragraph, the values of controllable parameters at time intervals t - 2, t - 1, and t and the controllable parameters at two past intervals, t - 2, t - 1, are used to predict the performance parameter at time t (see Figure 3.1).

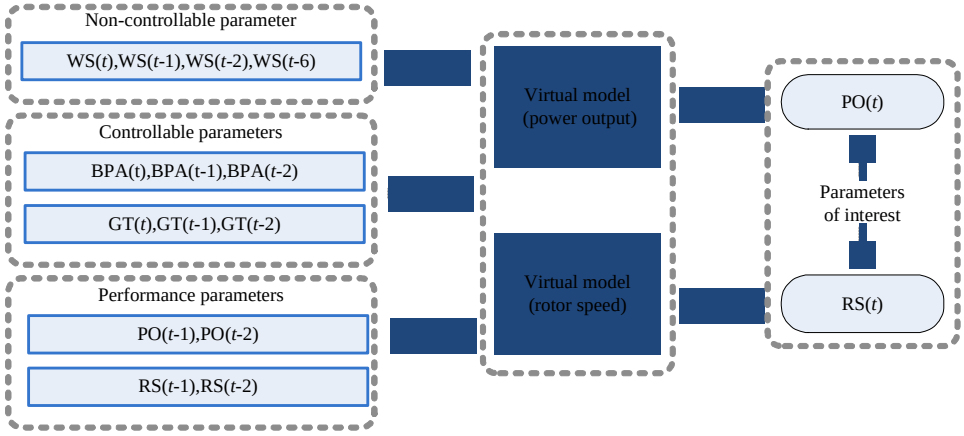

Figure 3.5 Parameter selection for virtual models (see Table 3.1 for definitions of the abbreviated parameters).

Figure 3.5 shows the final feature selection for the interest aspects: power output and rotor speed. Three kinds of input parameters and their past states' data are included:

Wind speed is the only non-controllable parameter considered in this paragraph and v (t), v (t-1) and v (t-2) and v (t-6) are used in virtual models.

Two controllable parameters, blade pitch angle and generator torque, and their two past states are selected.

The two past states of performance parameters, power output and rotor speed, are selected.

3.3.2 Data Sampling

In this section, the statistical properties of the data sets used in this research are summarized. The cumulative distributions of the wind speed, the power output, and the rotor speed are presented in Figure 3.6 through Figure 3.8. For low wind speed, 96.75% of the wind speed values are less than 12.5 m/s, and 88.5% of the power outputs are less than 1000 kW. For high wind speed, almost 18% of the wind speeds are larger than 12.5m/s, and nearly half of the power outputs are higher than 1000 kW. The rotor speed

for high wind speed values is higher than 10 rpm, while 16% of the rotor speeds for low wind speeds are less than 10 rpm.

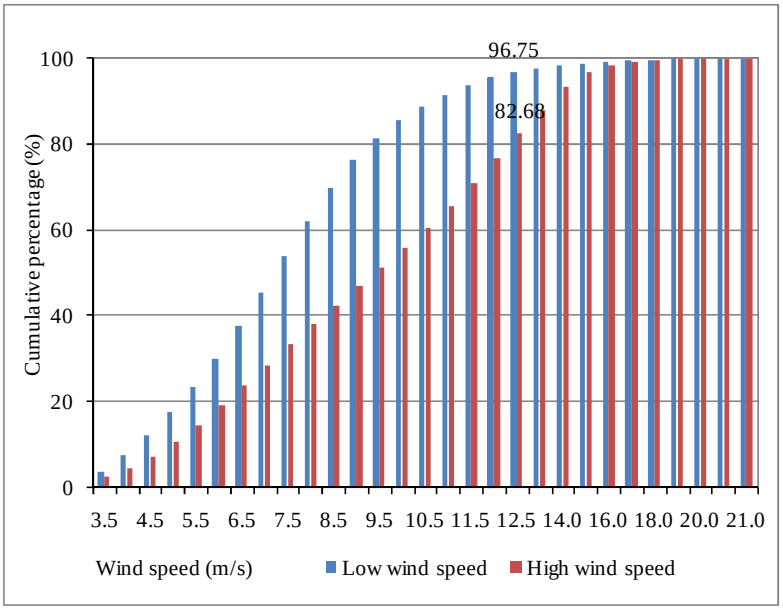

Figure 3.6 Comparison of wind speed distributions.

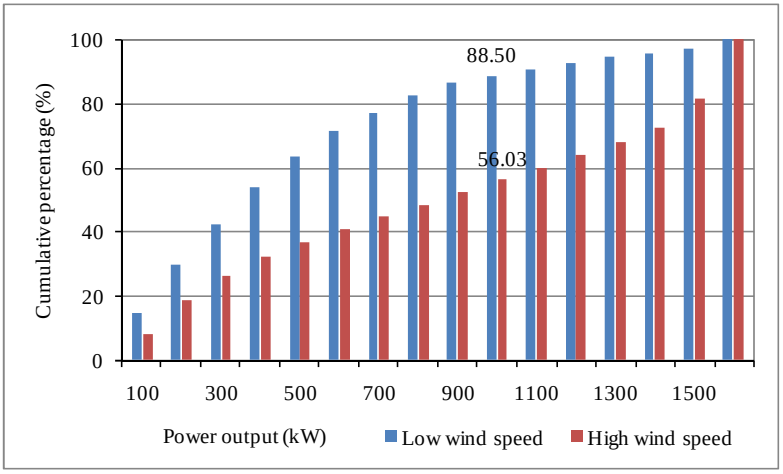

Figure 3.7 Comparison of power output distributions.

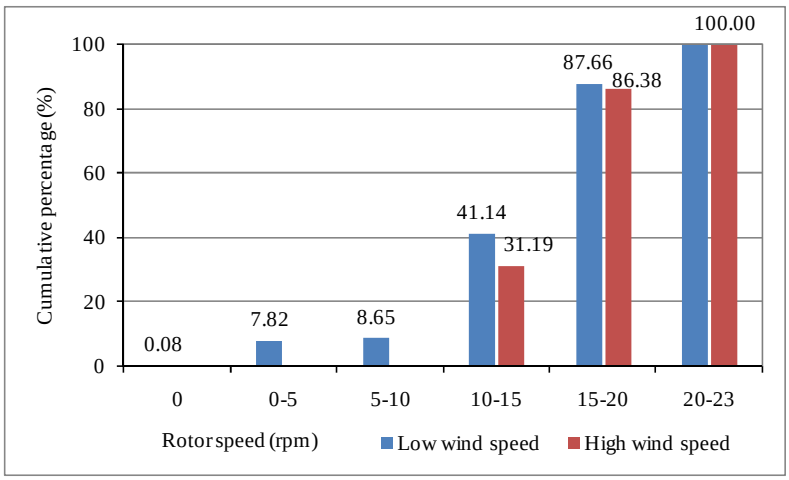

Figure 3.8 Comparison of the rotor speed distributions.

The charts in Figure 3.6 through Figure 3.8 demonstrate that the low wind speed distributions cover wider ranges than those for the high wind speed, especially for the rotor speed. Specifically, the low wind speed data contains 8.65% rotor speed data that is smaller than 10 rpm, while the 100% rotor speed data for high wind speed is higher than 10 rpm. Thus, the low wind speed data is selected to form a training set.

As the wind speed in the interval [3.5-13] m/s is studied, 1500 data points were randomly selected in each category of the wind speed data to form a training data. This way the training dataset is balanced across all categories. The data from turbines 1 through 10 was used to assemble the training data set. Table 3.7 shows the percentage of the data points selected in each class.

Table 3.7. Percentage of data selected in different classes of wind speeds.

No.	Range	Total Number of Data Points (1-10)	Number of Selected Data Points	Percentage (%)
1	3.5-4	17041	1500	8.80
2	4-4.5	21077	1500	7.12
3	4.5-5	23647	1500	6.34
4	5-5.5	24930	1500	6.02
5	5.5-6	27532	1500	5.45
6	6-6.5	30948	1500	4.85
7	6.5-7	34833	1500	4.31
8	7-7.5	37448	1500	4.01
9	7.5-8	38339	1500	3.91
10	8-8.5	37532	1500	4.00
11	8.5-9	31878	1500	4.71
12	9-9.5	25184	1500	5.96
13	9.5-10	20472	1500	7.33
14	10-10.5	15939	1500	9.41
15	10.5-11	12429	1500	12.07
16	11-11.5	10097	1500	14.86
17	11.5-12	8318	1500	18.03
18	12-12.5	6857	1500	21.88
19	12.5-13	5355	1500	28.01

3.3.3 Model Extraction

The models for predicting the power output and the rotor speed are expressed in (3.1) and (3.2), respectively.

$$y_1(t) = f_1(y_1(t-1), y_1(t-2), y_2(t-1), y_2(t-2), x_1(t), x_1(t-1), x_1(t-2)x_2(t), x_2(t-1), x_2(t-2)v(t), v(t-1)v(t-2)v(t-6)) \quad (3.1)$$

$$y_2(t) = f_2(y_1(t-1), y_1(t-2), y_2(t-1), y_2(t-2)x_1(t), x_1(t-1), x_1(t-2)x_2(t), x_2(t-1), x_2(t-2), v(t), v(t-1)v(t-2)v(t-6)) \quad (3.2)$$

The performance of the models (3.1) and (3.2) built by six different algorithms, specifically, random forest, neural network, boosting tree, support vector machine, generalized additive model, and the k-nearest neighbors, are reported in Table 3.8.

Table 3.8. Performance comparison for models extracted by six different algorithms.

	Power Output Prediction				Rotor Speed Prediction		
Algorithm	Average Observed Power Output	Average Predicted Power Output	Mean Absolute Error (kW)	Mean Relative Error (%)	Average Observed Rotor Speed	Average Predicted Rotor Speed	Mean Absolute Error (rpm)
Random Forest	573.90	573.90	28.60	21.07	15.66	15.67	0.91
Neural Network	576.03	576.03	8.03	4.95	15.64	15.64	0.18
Boosting Tree	575.65	575.75	34.55	54.71	15.62	15.62	0.27
Support Vector Machine	574.47	588.82	23.71	50.82	15.76	13.81	2.67
Generalized Additive	576.03	576.03	11.13	20.21	15.64	15.64	0.19
k-Nearest Neighbors	574.47	573.97	28.94	9.59	15.76	15.75	0.54

The Absolute Error (AE) and Relative Error (RE) used in Table 8 and all other tables are defined in (3.3) and (3.4):

$$\text{Absolute error} = \left|\hat{y}(t) - y(t)\right| \tag{3.3}$$

$$\text{Relative error} = \left|\frac{\hat{y}(i) - y(t)}{y(t)}\right| \times 100\% \tag{3.4}$$

Based on the data in Table 3.8, the neural network performed best among the six algorithms tested. The neural network algorithm is used to train a virtual model to predict

power output and rotor speed. Here, 30 different neural networks were trained, and the best model was selected to be a virtual model.

3.3.4 Analysis of Computational Results

In this section, the virtual model is evaluated using three types of industrial datasets. The nature of each dataset of the first type is similar to the training dataset. In fact, the training dataset is a subset of the combined set of data from 30 turbines. Therefore, the values of the non-controllable parameter (wind speed), controllable parameters (blade pitch angle and generator torque) and performance parameters (power output and rotor speed) share the same characteristics.

The nature of each dataset for the second type varies with the training dataset because each dataset is collected for different wind speed values. The training dataset itself includes data at low wind speeds, while each test dataset corresponds to high wind speed. Thus, the values of non-controllable parameters do not share the same characteristics of the training dataset.

In the third dataset, the values of controllable and non-controllable parameters have been randomly selected for a turbine and are much different than those in the training dataset.

1. Power Output Prediction Results

The data set collected from 30 turbines varied in quality. The data of turbine 6 and turbine 21 was removed from the test data set due to its low quality.

Table 3.9 presents statistics from the 28 turbines.

Table 3.9. Power output prediction.

Turbine No.	Average Observed Power Output (kW)	Average Predicted Power Output (kW)	Mean Absolute Error (kW)	Standard Deviation of Absolute Error	Mean Relative Error (%)	STD of Relative Error
1	474.45	476.36	6.58	6.02	2.91	121.33
2	489.6	486.5	6.61	6.22	2.32	7.12
3	379.36	378.43	5.9	5.85	3.70	116.60
4	494.98	494.85	9.89	8.23	8.65	134.68
5	474.98	473.16	6.05	5.64	3.97	69.24
7	447.35	445.80	6.59	6.56	3.85	24.76
8	462.87	467.1	7.53	6.50	4.74	86.27
9	501.49	504.58	10.36	9.04	5.5	42.13
10	423.1	421.76	5.79	5.22	11.41	768.97
11	467.04	467.13	6.63	12.27	2.56	12.44
12	460.66	459.92	6.46	6.11	5.98	172.36
13	495.85	497.24	5.92	5.05	7.18	814.55
14	448.23	448.29	6.43	6.93	3.95	156.37
15	453.11	456.38	6.25	5.20	3.42	19.54
16	387.48	394.19	10.41	11.80	3.06	5.91
17	472.72	475.38	5.79	5.09	3.7	30.40
18	487.92	490.38	6.3	7.15	3.11	92.01
19	430.67	431.29	6.01	5.84	2.39	10.18
20	429.59	430.03	6.02	6.11	2.53	7.91
22	462.14	464.44	6.35	5.96	3.56	40.24
23	446.34	448.3	7.29	16.13	3.11	17.03
24	468.82	470.45	5.59	4.83	3.66	34.36
25	481.56	482.47	6.36	11.46	2.69	14.93
26	535.22	534.37	6.95	12.74	4.17	174.39
27	459.42	460.54	6.35	12.14	2.52	12.20
28	485.73	486.58	5.57	4.94	2.34	7.96
29	434.28	434.27	5.9	5.61	2.58	14.02
30	462.01	462.58	6.18	13.26	3.11	31.57

The data in Table 3.9 indicates that the smallest mean absolute error is 5.57 kW (for turbine 28), and the smallest relative error is 2.32% (for turbine 2). The largest mean absolute error is 10.41 kW (for turbine 16), and the largest relative error is 11.41% (for turbine 10). Note that the rated power of each turbine is 1.5 MW. Thus, these four

turbines are selected for further analysis. To provide a broader context, the results for turbine 1 and turbine 26 are also included in the results discussed next.

1) Minimum absolute and relative errors

Table 3.10 shows the observed power, predicted power, and relative error data when the minimum absolute error is attained for each at the six selected turbines.

Table 3.10 Statistics for the six selected turbines representing the minimum absolute error.

Turbine No.	Minimum Absolute Error (kW)	Observed Power Output (kW)	Predicted Power Output (kW)	Relative Error (%)
1	0.00	680.71	680.71	0.00
2	0.00	122.95	122.95	0.00
10	0.00	137.81	137.81	0.00
16	0.00	47.42	47.42	0.00
26	0.00	572.19	572.19	0.00
28	0.00	199.87	199.87	0.00

2) Maximum absolute error

The observed power, predicted power, and relative error statistics in Table 3.11 correspond to the maximum absolute error for each at the six selected turbines.

For the turbines in Table 3.11, which represent the worst case prediction outcomes of the 28 turbines tested, some errors are acceptable. For example, for turbine 1, the maximum absolute error is 81.28 kW, yet the relative one is only 8.18%. The maximum absolute error for turbines 2 and 28 is similar in magnitude to that of turbine 1.

Table 3.11 Statistics for the six selected turbines representing the maximum absolute error.

Turbine No.	Maximum Absolute Error (kW)	Observed Power Output (kW)	Predicted Power Output (kW)	Relative Error (%)
1	81.28	993.75	912.47	8.18
2	66.57	1201.81	1135.24	5.54
10	159.94	974.38	1134.32	16.41
16	114.16	477.49	363.33	23.91
26	363.55	101.00	464.55	359.95
28	52.68	1251.67	1198.99	4.21

The prediction result of turbine 26, however, is not accurate. Therefore, it is necessary to analyze the error distribution over all data points for the turbines of Table 11. The distribution of points for different ranges of absolute errors, i.e., (0kW, 1KW), though greater than 200KW, is shown in Table 3.12

The results in Table 3.12 indicate that nearly 80% of the absolute errors for each turbine are smaller than 10 kW, and nearly 99% of the absolute errors are smaller than 50 kW. This implies that most of the time power output is accurately predicted. Figure 3.9 shows the absolute error distribution for turbine 16 and turbine 28. Turbine 28 shows the best results, and turbine 16 shows the worst results, among the six selected turbines. The results for the remaining four turbines are between turbine 28 and turbine 16.

Table 3.12 Absolute error distribution.

Turbine No.		Error (0, 1)	Error (1, 5)	Error (5, 10)	Error (10, 50)	Error (50, 100)	Error (100, 200)	Error >200	Total
1	Number of Points	5084	18030	13796	9689	23	1	0	46623
	Percentage (%)	10.90	38.67	29.59	20.78	0.05	0.00	0	100
	Cumulative Percentage (%)	10.90	49.58	79.17	99.95	100.00	100.00	100.00	
2	Number of Points	5263	19056	14036	9886	16	6	0	48263
	Percentage (%)	10.90	39.48	29.08	20.48	0.03	0.01	0.00	100
	Cumulative Percentage (%)	10.90	50.39	79.47	99.95	99.99	100.00	100.00	
10	Number of Points	5061	17551	11940	6532	26	7	0	41117
	Percentage (%)	12.31	42.69	29.04	15.89	0.06	0.02	0.00	100
	Cumulative Percentage (%)	12.31	54.99	84.03	99.92	99.98	100.00	100.00	
16	Number of Points	4480	15728	11472	14266	529	1	0	46476
	Percentage (%)	9.64	33.84	24.68	30.70	1.14	0.00	0.00	100
	Cumulative Percentage (%)	9.64	43.48	68.16	98.86	100.00	100.00	100.00	
26	Number of Points	5543	19400	14746	10059	22	0	81	49851
	Percentage (%)	11.12	38.92	29.58	20.18	0.04	0.00	0.16	100
	Cumulative Percentage (%)	11.12	50.04	79.62	99.79	99.84	99.84	100.00	
28	Number of Points	6255	20941	13993	7307	2	0	0	48498
	Percentage (%)	12.90	43.18	28.85	15.07	0.00	0.00	0.00	100
	Cumulative Percentage (%)	12.90	56.08	84.93	100.00	100.00	100.00	100.00	

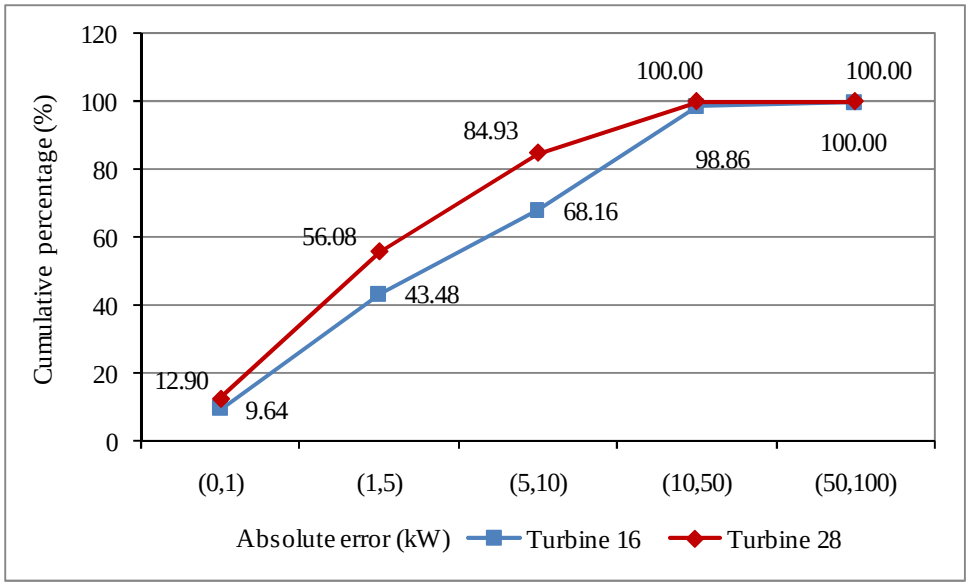

Figure 3.9 Absolute error distribution for turbines 16 and 28.

3) Maximum relative error

Figure 3.9 shows the observed and predicted power, as well as absolute error data, when the maximum relative error is attained for each at the six selected turbines.

Table 3.13 Statistics for the six selected turbines representing the maximum relative error.

Turbine No.	Maximum Relative Error (%)	Observed Power Output (kW)	Predicted Power Output (kW)	Absolute Error (kW)
1	11231.33	0.07	7.72	7.65
2	1018.55	3.09	34.61	31.51
10	38624.42	0.01	3.71	3.70
16	638.84	1.10	8.15	7.05
26	36908.10	0.04	-15.81	15.85
28	748.90	0.89	7.59	6.69

The results in Table 3.14 indicate that almost 90% of the relative errors are less than 5%, and 97% of the relative errors' percentages are not greater than 10%. This confirms that the model accurately predicts power output for the vast majority of turbines. Figure 3.10 shows the relative error distribution for turbine 16 and turbine 28. Turbine 28 shows the best results, and turbine 16 shows the worst results. The prediction results for all other four turbines fall between those of turbine 28 and turbine 16.

Table 3.14 Relative error distribution.

Turbine No.		0-5%	5%-10%	10%-20%	20%-100%	100%-1000%	> 1000%	Total
1	Number of points	42031	3135	987	418	50	2	46623
	Percentage (%)	90.15	6.72	2.12	0.90	0.11	0.00	100.00
	Cumulative percentage (%)	90.15	96.87	98.99	99.89	100.00	100.00	
2	Number of points	43795	3176	1016	262	14	1	48264
	Percentage (%)	90.74	6.58	2.11	0.54	0.03	0.00	100.00
	Cumulative percentage (%)	90.74	97.32	99.43	99.97	100.00	100.00	
10	Number of points	36037	2920	1131	724	261	44	41117
	Percentage (%)	87.65	7.10	2.75	1.76	0.63	0.11	100.00
	Cumulative percentage (%)	87.65	94.75	97.50	99.26	99.89	100.00	
16	Number of points	39202	6109	890	246	29	0	46476
	Percentage (%)	84.35	13.14	1.91	0.53	0.06	0.00	100.00
	Cumulative percentage (%)	84.35	97.49	99.41	99.94	100.00	100.00	
26	Number of points	45688	2717	1013	305	114	14	49851
	Percentage (%)	91.65	5.45	2.03	0.61	0.23	0.03	100.00
	Cumulative percentage (%)	91.65	97.10	99.13	99.74	99.97	100.00	
28	Number of points	44140	2872	1058	400	28	0	48498
	Percentage (%)	91.01	5.92	2.18	0.82	0.06	0.00	100.00
	Cumulative percentage (%)	91.01	96.94	99.12	99.94	100.00	100.00	

2. Results for the rotor speed prediction

The average observed and predicted values of the rotor speed for all 28 turbines are shown in Table 3.15.

The data in Table 3.15 illustrate that the mean absolute errors are between 0.1 rpm and 0.2 rpm. Turbine 28 shows the smallest mean absolute error, and turbine 18 shows the largest. Thus, the turbines 1, 18, 26, and 28 are selected for further analysis.

As a significant percentage of rotor speed is zero or close to zero, the corresponding relative error is large. As the results are meaningless, they are not presented here.

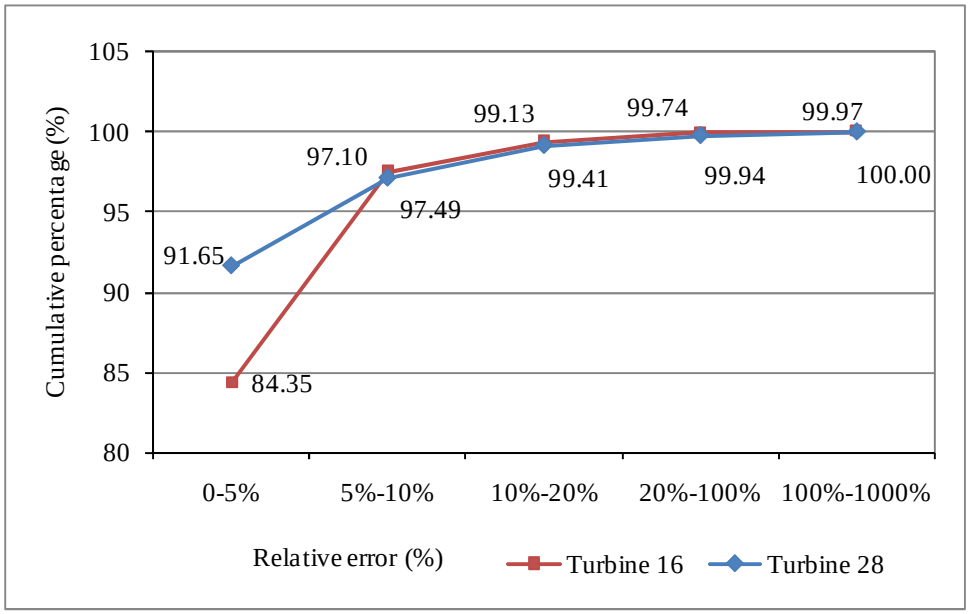

Figure 3.10 Relative error distribution for turbine 16 and turbine 28.

1) Minimum absolute error

Table 3.16 shows the observed and predicted rotor speed data for the four turbines when the minimum absolute error is attained.

Table 3.15 Statistics for the four turbines representing the minimum absolute error.

Turbine No.	Min Absolute Error (rpm)	Observed Rotor Speed (rpm)	Predicted Rotor speed (rpm)
1	0.00	0.14	0.14
18	0.00	14.02	14.02
26	0.00	0.03	0.03
28	0.00	20.25	20.25

Table 3.16 Average results for the rotor speed prediction.

Turbine No.	Average Observed Rotor Speed (rpm)	Average Predicted Rotor Speed (rpm)	Mean Absolute Error (rpm)	STD of Absolute Error
1	15.36	15.37	0.21	1.24
2	15.90	15.91	0.20	1.21
3	15.46	15.47	0.20	1.19
4	15.47	15.47	0.16	0.27
…	…	…	…	…
27	15.22	15.22	0.19	0.32
28	15.73	15.73	0.13	0.21
29	13.54	13.54	0.17	0.95
30	15.76	15.74	0.16	0.33

2) Maximum absolute error

Table 3.17 shows the observed and predicted rotor speed data for the four turbines corresponding to the maximum absolute prediction error.

Table 3.17 Statistics for the four turbines representing the maximum absolute error.

Turbine No.	Max Absolute Error (rpm)	Observed Rotor Speed (rpm)	Predicted Rotor Speed (rpm)
1	120.47	19.88	140.34
18	135.49	0.00	135.49
26	120.90	19.70	140.61
28	3.09	10.02	13.11

For maximum absolute error, some predictions are obvious wrong; however, some are acceptable, and turbine 28 is such an example. If the percentage of wrong prediction is small, then the prediction model is acceptable.

3) Distribution of absolute error

The results in Table 3.18 indicate that most of the absolute errors are less than 1 rpm. Turbine 1 represents the worst case scenario, where the absolute error is less than 1 rpm for 62.54% of the instances tested.

3. Prediction of power output and rotor speed for high wind speed

For the same turbines, data sets for high wind speed are also provided. To indicate that the model is also suitable in the high wind speed situation, turbines 1, 16, and 28 have been randomly selected to test the accuracy of the prediction models.

1) Average prediction results

The statistics for the prediction of the power rotor speed for the three selected turbines at high wind speed are shown in Table 3.19 and Table 3.20, respectively.

Table 3.18 Distribution of absolute errors for the four turbines.

Turbine No.		Error (0,0.2)	Error (0.2-1)	Error (1,2)	Error (2-5)	Error (5-10)	Error (10-100)	Error >100	Total
1	Number of points	13359	15801	6334	7048	2709	1202	170	46623
	Percentage (%)	28.65	33.89	13.59	15.12	5.81	2.58	0.36	100.00
	Cumulative percentage (%)	28.65	62.54	76.13	91.25	97.06	99.64	100.00	
18	Number of points	41041	10116	493	42	0	519	8	52219
	Percentage (%)	78.59	19.37	0.94	0.08	0.00	0.99	0.02	100.00
	Cumulative percentage (%)	78.59	97.97	98.91	98.99	98.99	99.98	100.00	
26	Number of points	36527	12195	920	117	75	13	4	49851
	Percentage (%)	73.27	24.46	1.85	0.23	0.15	0.03	0.01	100.00
	Cumulative percentage (%)	73.27	97.74	99.58	99.82	99.97	99.99	100.00	
28	Number of points	38652	9412	393	41	0	0	0	48498
	Percentage (%)	79.70	19.41	0.81	0.08	0.00	0.00	0.00	100.00
	Cumulative percentage (%)	79.70	99.11	99.92	100.00	100.00	100.00	100.00	

Table 3.19 Power prediction results for the three selected turbines at high wind speed.

Turbine No	Average Observed Power Output (kW)	Average Predicted Power Output (kW)	Mean Absolute Error (kW)	STD of Absolute Error (kW)	Average Relative Error (%)	STD of Relative Error (%)
1	735.06	742.42	10.18	8.42	2.67	28.04
16	745.96	746.73	7.24	6.94	1.63	3.63
28	804.77	807.19	6.97	6.23	2.23	9.07

2) Distribution of the results

The results in Table 3.21 indicate over 99% of the absolute errors are less than 50 kW. As this data set represents high wind speeds, the average power output is high, and the result accuracy is acceptable. Table 3.22 shows the distribution of relative error of the

selected turbines. Over 97% of the relative errors are less than 10%, and nearly 99% of the relative errors are less than 20%.

Table 3.20 Speed prediction results for the three selected turbines at high wind speed.

Turbine No	Average Observed Rotor Speed (rpm)	Average Predicted Rotor Speed (rpm)	Mean Absolute Error (rpm)	STD of Absolute Error (rpm)
1	16.72	16.71	0.24	0.34
16	16.89	16.89	0.23	0.33
28	17.04	17.03	0.15	0.22

Table 3.21 Distribution of the absolute error of the power output prediction at high wind speed.

Turbine No.		Error (0,1)	Error (1,5)	Error (5,10)	Error (10-50)	Error (50-100)	Error (100-200)	Error > 200	Total
1	Number of points	2812	10264	10684	16520	55	0	0	40335
	Percentage (%)	6.97	25.45	26.49	40.96	0.14	0.00	0.00	100.00
	Cumulative percentage (%)	6.97	32.42	58.91	99.86	100.00	100.00	100.00	
16	Number of points	4609	16369	12003	10646	46	1	0	43674
	Percentage (%)	10.55	37.48	27.48	24.38	0.11	0.00	0.00	100.00
	Cumulative percentage (%)	10.55	48.03	75.52	99.89	100.00	100.00	100.00	
18	Number of points	4672	16646	12889	10499	11	0	0	44717
	Percentage (%)	10.45	37.23	28.82	23.48	0.02	0.00	0.00	100.00
	Cumulative percentage (%)	10.45	47.67	76.50	99.98	100.00	100.00	100.00	

Table 3.22 Distribution of the relative error of the power output prediction at high wind speed.

Turbine No.		0-5%	5%-10%	10%-20%	20%-100%	100%-1000%	> 1000%	Total
1	Number of points	36828	2760	489	229	27	2	40335
	Percentage (%)	91.22	6.84	1.21	0.57	0.07	0.00	99.91
	Cumulative percentage (%)	91.31	98.15	99.36	99.93	100.00	100.00	
16	Number of points	41543	1505	448	174	4	0	43674
	Percentage (%)	95.12	3.45	1.03	0.40	0.01	0.00	100.00
	Cumulative percentage (%)	95.12	98.57	99.59	99.99	100.00	100.00	
28	Number of points	41435	1926	797	502	57	0	44717
	Percentage (%)	92.66	4.31	1.78	1.12	0.13	0.00	100.00
	Cumulative percentage (%)	92.66	96.97	98.75	99.87	100.00	100.00	

Table 3.23 Distribution of the absolute prediction error of the rotor speed at high wind speed.

Turbine No.		Error (0,0.2)	Error (0.2-1)	Error (1,2)	Error (2-5)	Error (5-10)	Error (10-100)	Error > 100	Total
1	Number of points	26633	11995	1575	132	0	0	0	40335
	Percentage (%)	66.03	29.74	3.90	0.33	0.00	0.00	0.00	100
	Cumulative percentage (%)	66.03	95.77	99.67	100.00	100.00	100.00	100.00	
16	Number of points	28872	13075	1615	112	0	0	0	43674
	Percentage (%)	66.11	29.94	3.70	0.26	0.00	0.00	0.00	100.00
	Cumulative percentage (%)	66.11	96.05	99.74	100.00	100.00	100.00	100.00	
28	Number of points	33960	10239	500	18	0	0	0	44717
	Percentage (%)	75.94	22.90	1.12	0.04	0.00	0.00	0.00	100
	Cumulative percentage (%)	75.94	98.84	99.96	100.00	100.00	100.00	100.00	

Table 3.22 illustrates the distribution of absolute errors for rotor speed at high wind speed. Over 99.5% of the errors are less than 2 rpm.

4 Prediction of power output and rotor speed using independent datasets

In this section, another 10-second dataset randomly selected from the same wind farm is used to test the virtual models. After denoising, 32725 data points are provided. The distribution of wind speed, power output, and rotor speed is shown in Figure 3.11 through Figure 3.13.

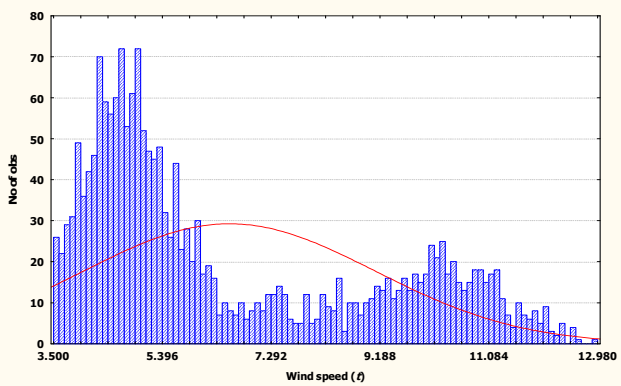

Figure 3.11 Wind speed distribution.

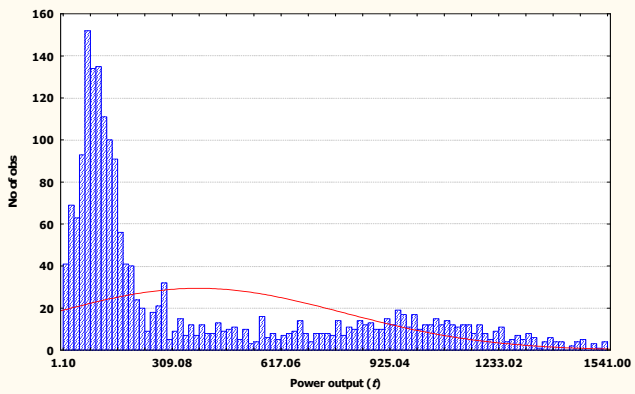

Figure 3.12 Power output distribution.

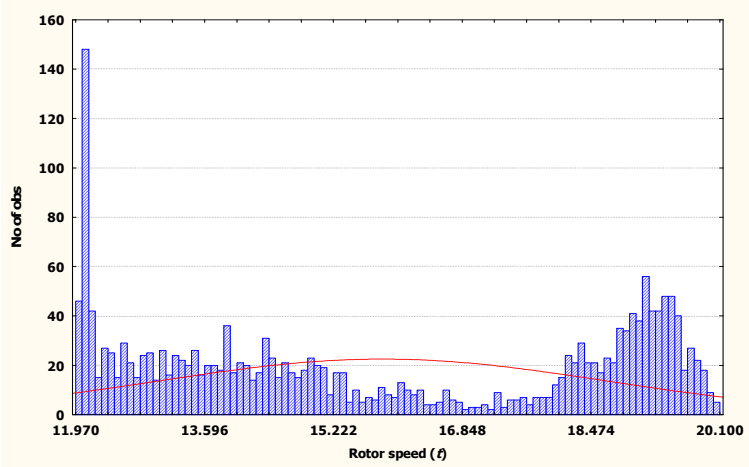

Figure 3.13 Rotor speed distribution.

The average prediction results for the power and rotor speed are shown in Table 3.24 and Table 3.25. The results indicate that the power output and rotor speed are accurately predicted. Note that the test data represents different measurements taken at different turbines. The high prediction accuracy of the power output and rotor speed is reinforced in Figures 3.14 and 3.15 for 71 consecutive wind speeds.

Table 3.24 Prediction results of the power output.

Average Observed Power Output (kW)	Average Predicted Power Output (kW)	Mean Absolute Error (kW)	STD of Absolute Error (kW)	Average Relative Error (%)	STD of Relative Error (%)
418.62	420.08	5.59	15.85	2.59	59.93

Table 3.25 Prediction results of the rotor speed.

Average Observed Rotor Speed (rpm)	Average Predicted Rotor Speed (rpm)	Mean Absolute Error (rpm)	STD of Absolute Error
14.87	14.88	0.14	0.22

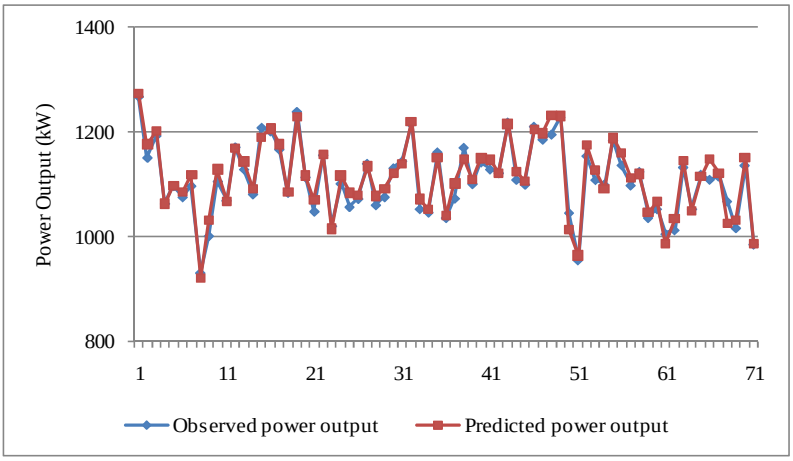

Figure 3.14 Prediction of the power output.

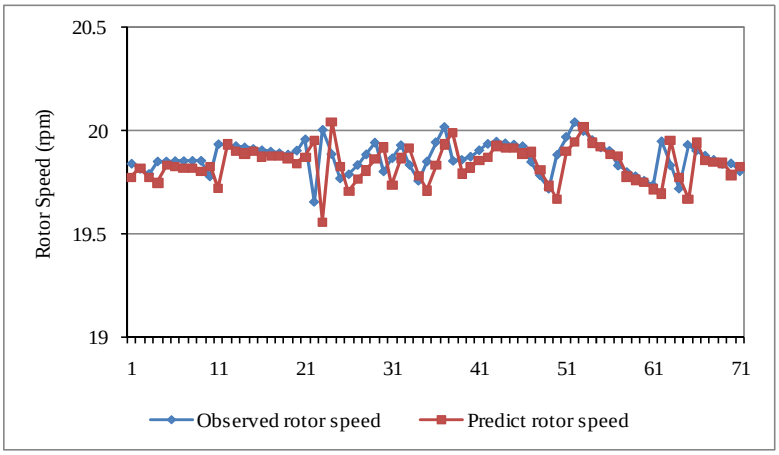

Figure 3.15 Prediction of the rotor speed.

3.5 Summary

In this section, a methodology for building virtual models for the prediction of the parameters of a wind turbine was presented. The proposed methodology involved three steps: data preprocessing, model extraction, and model validation. In the first step, after analyzing the raw data, the controllable parameters and non-controllable parameters, as well as their past states, are considered for feature selection. In order to eliminate data bias, a stratified sampling is performed based on the wind speed.

Two parameters were selected to test the proposed methodology: power output and rotor speed. The models were extracted by six different algorithms: random forest, neural network, boosting tree, support vector machine, generalized additive approach, and the *k*-nearest neighbor algorithm. The neural network showed the best performance and was selected for extraction of the models for parameter prediction.

The models developed in this paragraph were validated by three datasets of different characteristics, including the wind speed range, the time period, and the source. The first dataset included data corresponding to low wind speeds, the second dataset was generated at high wind speeds, and the final data was randomly selected from a turbine at the same wind farm. Although the test datasets share different characteristics, the parameters predicted by the virtual models were accurate. This implies that the virtual models can be used to predict the power output and rotor speed for a turbine of interest using the data collected at other turbines.

CHAPTER 4.
SHORT TERM POWER PREDICTION BASED ON CLUSTERING APPROACH

4.1 Introduction

In this chapter, data-mining algorithms are applied for short-term prediction of generated wind power. The previous research indicated that the accuracy of short-term power prediction deteriorates at low levels of the observed power. In some cases, the prediction error could be large and this makes the prediction results meaningless. The reason behind this poor performance might due to the algorithms not being able to model certain ranges of data. To address this problem, a clustering-based method for power prediction is proposed. The goal is to develop models (called here customized models) for situations that share certain data characteristics defined by the data clusters.

In the quest of maximizing prediction accuracy at low wind speeds, first, the input space (data used by data-mining algorithms) is indentified by selecting parameters using physics-based equations and data-mining algorithms. Second, the input space is clustered into several mutually separable subspaces by associating data with the most characteristics (clusters). Third, data-mining algorithms produce models for each clustered sub-space. Fourth, computational results are reported, compared, and discussed.

4.2 Data Description and Methodology for Short Term Prediction of Power with Clustering Approach

4.2.1 Data Description and Parameter Selection

Wind turbine data is usually collected by a Supervisory Control and Data Acquisition (SCADA) system. Though the data sampling frequency may be relatively

high (e.g., 20Hz), the data is averaged into time intervals, e.g., 10-s, 30-s, or 10-min, that are suitable for various applications. The data used in this paragraph was collected at 10-s intervals (called 10-s data) collected at a 1.5 MW wind turbine (randomly selected) for a period of seven days. For the selected wind turbine, the cut-in speed is 3.5 m/s, the rated speed is 12.5 m/s, and cut-out speed is 21 m/s. From view of turbine operations, wind speed in the range [3.5m/s, 12.5m/s] is of interest to industry. Thus, the data with wind speed lower than 3.5 m/s or higher than 12.5 m/s have been excluded from analysis in the research reported in this paragraph. Data points with the negative power output have been also deleted. The data from the first five days (approximately 2/3 of all data) was used to extract models and the data from the remaining two days (approximately 1/3 of all data) was used for test and validate models. The data set used in this research is characterized in Table 4.1.

Table 4.1 Description of the training and test data.

Data Set	Start Time	End Time	Time Interval	Number of Data Points
Training	8/8/07 12:00 AM	8/12/07 12:00 AM	10-s	30354
Test	8/13/07 12:00 AM	8/15/07 12:00 AM	10-s	15860

The data available for this research included numerous parameters of a wind turbine. Some of these parameters could have a potential impact on the prediction accuracy of wind power to be generated at 10-s intervals. These parameters include: the Power Output (PO), Generator Torque (GT), Generator Speed (GS), Wind Speed (WS), Generator Bearing A Temperature (GBAT), Generator Bearing B Temperature (GBBT),

Drive Train Acceleration (DTA), Blade Pitch Angle (BPA), Nacelle Position (NP), and the Rotor Speed (RS).

These ten parameters can be categorized into three classes:

- Controllable parameters (the parameters of a wind turbine that can be adjusted), e.g., Blade Pitch Angle (BPA), Generator Torque (GT).
- Non-controllable parameters (those that cannot be adjusted (controlled)), e.g., Wind Speed (WS).
- System performance parameters (the ones that are to be predicted), e.g., Power Output (PO), Rotor Speed (RS).

Both, the science and the experience, indicate that not all of these ten parameters have equal impact on the short-term power prediction, which leads to parameter selection. The parameter selection can be accomplished in three ways: expert knowledge, physics-based equations, and data-mining algorithms. In this paragraph, physics-based equations are used first for parameter selection as explained next.

The power extracted from the wind is expressed by (1.1). The power coefficient $C_p(\lambda, \beta)$ of expression (1.1) contains the blade pitch angle β, and the tip-speed ratio λ defined in (4.1) [22]:

$$\lambda = \frac{\omega_r R}{v} \tag{4.1}$$

where ω_r is the rotor speed, R is the rotor radius, v is the wind speed. For a given blade pitch angle β, any change in the wind speed implies change in the tip-speed ratio λ of (4.1), thus leading to the variation of the power coefficient $C_p(\lambda, \beta)$ and the generated power output P_a.

The mechanical power is also be expressed by the equation in (4.2) [22]:

$$P_a = \omega_r T \tag{4.2}$$

where T is the aerodynamic torque and ω_r is the rotor speed.

Based on the equations (1.1) through (4.2), five parameters are selected as candidates, including wind speed v, blade pitch angle β, generator torque T, and rotor speed ω_r. Considering the fact that the system inertia could be significant, the power output P_a is also included. The air density ρ and rotor radius R are regarded here as constants. The initially selected parameters are listed in Table 4.2.

Table 4.2 List of parameters selected for wind speed estimation.

Parameter Type	Parameter Name	Abbreviation	Symbol	Unit
Non-controllable	Wind speed	WS	v	m/s
Controllable	Blade pitch angle	BPA	x_1	°
	Generator torque	GT	x_2	Nm
Performance	Power output	PO	y_1	kW
	Rotor speed	RS	y_2	rpm

The five parameters in Table 4.2 have been selected based on the equations (1.1) through (4.2). They include a non-controllable parameter, the Wind Speed (WS); two controllable parameters, Blade Pitch Angle (BPA) and Generator Torque (GT); and two performance parameters, the Power Output (PO) and the Rotor Speed (RS). Since time delay is considered as having impact on the model accuracy, it also considered and thus further parameter selection is accomplished with data-mining algorithms.

4.2.2 Proposed Methodology

The proposed methodology is outlined in Figure 4.1. The input data (controllable, non-controllable, and performance parameters) representing the input space undergo parameter selection and clustering. Based on the data in each cluster, a model is produced

with data-mining algorithms. The number of input parameters n (dimension) and the number of instances N define the input space IS_n. For each parameter in the input space, there are $N = 30354$ training instances (see in Table 4.2), and $N = 15860$ test instances (see in Table 4.2). Each of the models $1,...,k$ predicts power output $PO(t)$ at time t.

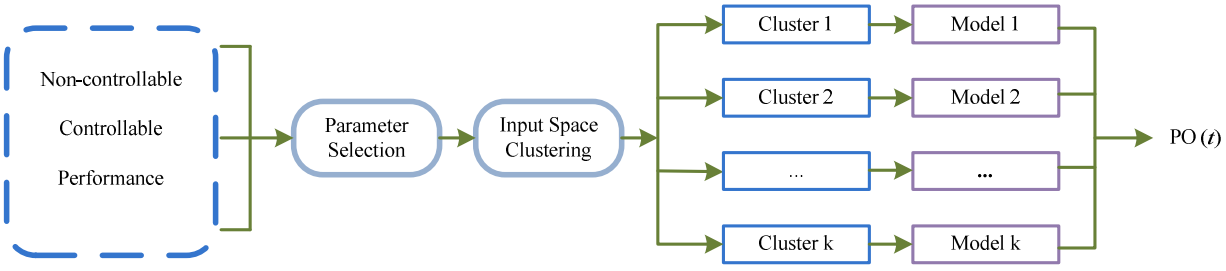

Figure 4.1 The methodology for clustering-based power prediction.

The steps of the methodology presented in Figure 4.1 are discussed next.

1. Parameter Selection

The five parameters listed in Table 4.2 partially describe the input space. The impact of past states of non-controllable, controllable, and performance parameters needs to be reflected in the proposed model. The input space IS_n for the m past states of n parameters listed in Table 4.2 is defined by vector (4.3).

$$IS_n = [v(t-1),...v(t-m), x_1(t),...,x_1(t-m), x_2(t),...,x_2(t-m), y_1(t-1),...,y_1(t-m), y_2(t-1),...,y_2(t-m)] \quad (4.3)$$

Of all n parameters, the most significant are selected with a NN algorithm.

2. Clustering Input Space

The input training data set is clustered by the *k*-means algorithm using five different data processing scenarios. Then, each instance of test data is assigned to the nearest cluster center (centroid) of the training data to identify the most similar cluster.

1) Clustering training data set

In this section, the training data set is clustered into *k* subspaces. Based on parameters selected for clustering, the following five data processing scenarios are considered:

a) Clustering wind speed estimated by its one past state
b) Clustering wind speed estimated by the time series model
c) Clustering generator torque $x_2(t)$
d) Clustering generator torque $x_2(t)$ and rotor speed $y_2(t-1)$
e) Clustering generator torque $x_2(t)$, rotor speed $y_2(t-1)$, and the power output $y_1(t-1)$

The first two scenarios explore the impact of wind speed on the accuracy of power output predictions at 10-s intervals. As the wind speed $v(t)$ at future time *t* is not known, two estimation methods have been applied. The first one (item (a) above) uses wind speed at one past state $v(t-1)$, and the second one is based on the time series model to estimate. Prediction of wind speed with the time series models has been proven to be accurate (see [13]).

The final three scenarios of clustering the input space originate in predictors' importance. The first three most significant parameter states determined by the data-mining algorithms are the generator torque $x_2(t)$, rotor speed $y_2(t-1)$, and power output $y_1(t-1)$. All parameters are studied for impact of clustering the input space.

2) Clustering the test date set

For the *n* dimensional space, the center (centroid) of the i^{th} cluster of training data is denoted as $[x^i_1, x^i_2, ..., x^i_n]$, where i is the number of the cluster satisfying $1 \le i \le k$. Note that values of $x^i_1, x^i_2, ..., x^i_n$ have been normalized according to (4.4) to balance the bias due to the variability of the input data:

$$x^{i}_{n,\ \text{normalized}} = x^{i}_{n} / x_{n,\max} \tag{4.4}$$

where $x^{i}_{n,\ \text{normalized}}$ is the normalized value and in the range [0, 1].

Each instance in test data is denoted as $[z_1, z_2, ..., z_n]$ and is normalized according to (4.5):

$$z_{n,\ \text{normalized}} = z_n / x_{n,\max} \tag{4.5}$$

Thus, the distance from a normalized instance $[z_{1,\ \text{normalized}}, ..., z_{n,\ \text{normalized}}]$ to the i^{th} cluster centriod of training data is defined in (4.6).

$$D^i = \sqrt{(z_{1,\ \text{normalized}} - x^{i}_{1,\ \text{normalized}})^2 + ,..., + (z_{n,\ \text{normalized}} - x^{i}_{n,\ \text{normalized}})^2} \tag{4.6}$$

The aim is to find for each data instance, a cluster with the minimum distance between the instance and the cluster centriods. The clustering algorithm of the test data set is shown in Table 4.3.

Table 4. 3 Algorithm for clustering test instances.

Begin
For $i = 1$ to k
$D^{\min} = 1$
$D^i = \sqrt{(z_{1,\ \text{normalized}} - x^{i}_{1,\ \text{normalized}})^2 + ,..., + (z_{n,\ \text{normalized}} - x^{i}_{n,\ \text{normalized}})^2}$
If $D^i \le D^{\min}$
Let $D^{\min} = D^i$, $[z_{1,\ \text{normalized}}, ..., z_{n,\ \text{normalized}}] = [z^{i}_{1,\ \text{normalized}}, ..., z^{i}_{n,\ \text{normalized}}]$
Else
Next i
End

As illustrated in Table 3.3, $D^{\min}$ initially set to 1 and then it is replaced with a shorter distance found. In this way, each instance from the test data set is assigned to the closest cluster.

3) Evaluation of clustering effectiveness

The reason for clustering the input space is to achieve high prediction accuracy of the customized model with a few input parameters. To demonstrate the effectiveness of clustering, NN models are constructed for five different clustering scenarios. A clustering scenario with the highest prediction accuracy is used for the final model extraction and validation.

3. Model Extraction and Validation

After all instances from the training and test data sets have been assigned to the corresponding clusters, data-mining models are constructed and validated. For each training data set of a given cluster, models are extracted with data-mining algorithms. Every model is tested using the corresponding test data set.

The structure of the cluster-based power prediction model is presented in (4.7).

$$y_1^*(t) = \begin{cases} 0 & if \ \ v(t) < 3.5 \ m/s \\ f_1(g) & if \ \ IS_n \in Cluster_1 \\ f_2(g) & if \ \ IS_n \in Cluster_2 \\ \ldots & \ldots \\ f_k(g) & if \ \ IS_n \in Cluster_k \\ 1500 & if \ \ v(t) > 12.5 \ m/s \end{cases} \tag{4.7}$$

The predicted power output $y_1^*(t)$ is assumed to be 0 when wind speed is less than 3.5 m/s and it equals 1500 kW for the wind speed higher than 12.5 m/s. For the wind speed in [3.5, 12.5] m/s, the input space is clustered into k subspaces. The prediction accuracy for the test data in each subspace is discussed in Sections 4.3.

4.3 Industrial Case Study on Short Term Prediction of Power based on Clustering Approach

4.3.1 Parameter Selection

Different algorithms, e.g., the Boosting Tree Algorithm (BTA), Neural Network (NN), and the Random Forrest Algorithm (RFA), can be used to determine importance of parameters and their past states (predictors). As the NN algorithm has been used numerous times in this paragraph, it is also applied to compute absolute importance, based on which, relative importance of parameters is generated according to (4.8).

$$\text{RelativeImportance} = \frac{\text{AbsoluteImportance}\,(PS_j)}{\sum_{j=1}^{k}\text{AbsoluteImportance}\,(PS_j)} \tag{4.8}$$

where PS_j is the j^{th} parameter $(1 \le j \le n)$ and n is total number of predictors.

Next, models with five different number of inputs (predictors $n = 27, 10, 3, 2, 1$) are discussed.

1. Twenty Seven Inputs

To reflect the importance of parameters used for building of the power prediction model and their past states, the initial value *m* of the past states is set to 5 (50 seconds = 5 past values measured at 10-s intervals). Thus, the total number of predictors (input states) n equals 27 (5 WS past states + 5 PO past states + 5 RS past states + 5 GT past states + 5 BPA past states + 1 current GT state + 1 current BPA state). The input space is represented as vector (4.9).

$$\begin{aligned} IS_{27} = [&v(t-1),\ldots,v(t-5), x_1(t),\ldots,x_1(t-5), x_2(t),\ldots,x_2(t-5),\\ &y_1(t-1),\ldots,y_1(t-5), y_2(t-1),\ldots,y_2(t-5)] \end{aligned} \tag{4.9}$$

A NN algorithm is applied to determine the importance of predictors. The first ten most important predictors among all the 27 states in Eq. (4.9) are shown in Table 4.4

Table 4. 4 The 10 most important predictors of Eq. (4.9).

Input No.	Parameter State	Importance	Percentage (%)
1	GT(t)	781.48	94.28
2	RS(t -1)	11.53	1.39
3	RS(t -2)	3.82	0.46
4	PO(t -1)	3.30	0.40
5	RS(t -3)	1.99	0.24
6	GT(t -1)	1.87	0.23
7	PO(t -2)	1.86	0.22
8	RS(t -4)	1.53	0.18
9	PO(t -3)	1.48	0.18
10	PO(t -5)	1.48	0.18

As illustrated in Table 4.4, the ten most important predictors amount to 97.76% of the total importance score of all 27 states. To reduce the computational effort, the remaining 17 states contributing 2.34% of the overall importance and not considered for further analysis.

2. Ten Inputs

Here, $n = 10$ parameters listed in Table 4 are used as inputs to the NN model. The input space is represented as the vector of predictors (4.10).

$$IS_{10} = [y_1(t-1), ..., y_1(t-3), y_1(t-5), y_2(t-1), ..., y_2(t-4), x_2(t), x_2(t-1)] \qquad (4.10)$$

The importance of predictors is computed by the NN algorithm as shown in Table 4.5.

Table 4.5 Ranked importance of predictors of Eq. (4.10).

No. of Inputs	Parameter State	Importance	Percentage
1	**GT(*t*)**	**865.32**	**95.61**
2	**RS(*t* -1)**	**25.15**	**2.78**
3	**PO(*t* -1)**	**2.76**	**0.31**
4	RS(*t* -2)	2.61	0.29
5	GT(*t* -1)	2.27	0.25
6	PO(*t* -2)	1.86	0.21
7	RS(*t* -4)	1.50	0.17
8	RS(*t* -3)	1.34	0.15
9	PO(*t* -3)	1.15	0.13
10	PO(*t* -5)	1.07	0.12

The sequence of predictors in Table 4.4 with Table 4.5 differs, yet GT(*t*) and RS(*t*-1) remain on the top of both lists. The third predictor of interest is PO(*t*-1) rated as important in both tables, comparable to RS (*t*-2). As illustrated in Table 4.5, the importance of the three predictors amounts to 98.70% (95.61%+2.78%+0.31%). The importance of the remaining seven predictors that are not considered for further analysis amounts to 1.3%.

3. One, Two, and Three Inputs

For $n = 3$, the input space is represented as the vector of predictors (4.11).

$$IS_3 = [y_1(t-1), y_2(t-1), x_2(t)] \tag{4.11}$$

The importance of predictors recomputed with a NN algorithm is shown in Table 4.6.

Table 4.6 Importance of predictors of Eq. (4.11).

No. of Inputs	Predictor	Importance	Percentage
1	GT(t)	824.83	96.14
2	RS(t-1)	30.48	3.55
3	PO(t-1)	2.63	0.31

Based on (4.11), the predictor vectors for n = 2 and n = 1 are presented in (4.12) and (4.13), respectively.

$$IS_2 = [y_2(t-1), x_2(t)] \tag{4.12}$$

$$IS_1 = [x_2(t)] \tag{4.13}$$

4. Comparison of Prediction Performance

To evaluate prediction performance of the models developed in this paragraph the following metrics are used: the Mean Absolute Error (MAE), Mean Relative Error (MRE), Standard Deviation of MAE, and MRE. The MAE and MRE that serve as the basis for formulating other two metrics are defined in (4.14) and (4.15).

$$\text{MAE} = \frac{\sum_N \left| y^*_1(t) - y_1(t) \right|}{N} \tag{4.14}$$

$$\text{MRE} = \frac{\sum_N \left| \frac{y^*_1(t) - y_1(t)}{y_1(t)} \right|}{N} \tag{4.15}$$

where $y^*_1(t)$ is the predicted power output of Eq. (4.7) and $y_1(t)$ is the observed value provided in the data, N is the number of instances in the corresponding cluster.

To compare performance of models with different number of predictors (inputs), here n = 27, 10, 3, 2, 1, the prediction accuracy of both training and test data are summarized in Table 7. For each n, 30 NNs are constructed and the best performing NN is selected for comparative study.

Table 4.7 Prediction results for different number of predictors n.

Number of Predictors	Data Set	Average Observed Power Output [kW]	Average Predicted Power Output [kW]	Mean Absolute Error [kW]	STD of MAE	Mean Relative Error [%]	STD of MRE
27	Training	442.75	442.74	8.48	7.11	4.25	20.08
	Test	*605.56*	*605.56*	*10.72*	*9.01*	*2.20*	*5.84*
10	Training	442.75	442.73	8.60	7.43	4.04	14.44
	Test	605.56	605.34	10.78	9.56	2.23	7.14
3	Training	442.75	442.68	8.83	7.65	4.89	62.14
	Test	**605.56**	**605.17**	**10.98**	**9.78**	**2.25**	**5.73**
2	Training	442.75	442.74	8.86	7.66	4.33	19.84
	Test	605.56	605.38	10.97	9.77	2.26	6.50
1	Training	442.75	442.66	9.42	7.91	4.48	16.54
	Test	605.56	604.88	11.40	9.89	2.34	5.20

The best test prediction performance reported in Table 4.7 was attained for $n = 27$ predictors (the row in italics) on the four metrics (MAE, MRE, STD of MAE, and MRE). A reduced number of predictors has resulted in decreased prediction accuracy for the test data. Compared with $n = 27$, for $n = 3$, the MAE increases by 0.26 kW; and it increases by 0.68kW for $n = 1$.

A larger value of n (number of predictors) produced better quality results, however, at an increased computational time. The NN training time on a standard desktop computer for different number of predictors n is listed in Table 4.8. For each n, 30 NNs were built and the best performing NN was used to generate results.

Table 4.8 NN Training time for different number of predictors n.

Number of Predictors	Computational Time [hour]
27	13.67
10	3.58
3	1.33
2	1.23
1	1.16

As illustrated in Table 4.8, the computational time was reduced over 10 hours when n was reduced from 27 to 10, and by 2.2 hours when n was reduced from 10 to 3. For n smaller than 3, the computational time did not get significantly reduced. To balance prediction accuracy and computational time, the vector of predictors (4.11) has been selected for further analysis. The three parameter states in Eq. (4.11) include generator torque $x_1(t)$, rotor speed $y_2(t-1)$, and power output $y_1(t-1)$ with the corresponding base performance marked in bold in Table 4.7.

The input space is partitioned using the k-means clustering algorithm according to five different clustering scenarios. The clustering scenario with the best performance is selected for model extraction.

4.3.2 Clustering Input Space

In this section, five clustering scenarios are studied. Each scenario involves different set of predictors.

First, the training data set is clustered into k clusters by the k-means algorithm. The number of clusters k varies from 2 to 5. The number k is not expected to be large due to concerns of the model complexity. Instances of the test data are assigned to the nearest cluster centriod of the training dataset with the algorithm shown in Table 3.

Second, 30 NNs are built for each cluster and the best performing model is selected. Third, prediction results for each cluster are compared.

Computational experience with the five clustering scenarios is discussed in the next five sections.

1. Clustering Based on a Single Past State of Wind Speed

The generated wind power is impacted by the wind speed $v(t)$, therefore here and in the scenario to follow, it is used to cluster the input space. After the training data has been clustered on $v(t)$ with the *k*-means algorithm into k clusters, the corresponding value of the generator torque $x_1(t)$, rotor speed $y_2(t-1)$, and power output $y_1(t-1)$ are assigned to each instance in every cluster. Note that the wind speed $v(t)$ is used only to clustering, without being used to train a NN model. The value of wind speed $v(t)$ is estimated by two methods: one using one past state of the wind speed only, i.e., $v^*(t) = v(t-1)$, and the other (see Eq. (4.16)) in the form of a time series model.

For the training data set clustered on wind speed $v(t-1)$ by the *k*-means algorithm and number of clusters k = 2,…, 5. After the best performing NN has been selected (out of 30 NNs produced for each cluster), the prediction results for the test data are shown in Table 4.9.

Table 4. 9 Prediction performance based on the data clustered on the past state of wind speed $v(t-1)$.

Input Space Clustered by	Number of Clusters	Observed Average Power Output [kW]	Predicted Average Power Output [kW]	Mean Absolute Error [kW]	STD of MAE	Mean Relative Error [%]	STD of MRE
WS(*t*-1)	2	605.56	605.29	10.90	9.64	2.29	8.63
WS(*t*-1)	3	605.56	605.22	10.98	10.45	2.29	7.90
WS(*t*-1)	**4**	**605.56**	**605.32**	**10.86**	**9.54**	**2.27**	**6.04**
WS(*t*-1)	5	605.56	605.40	10.89	9.69	2.33	10.80

As illustrated in Table 4.9, the best prediction accuracy is attained for $k = 4$ (shown in bold). However, it is not significantly better than the base performance for vector space (4.10) with MAE = 10.98 kW, STD of MAE = 9.78 kW, MRE = 2.25 %, and STD of MRE = 5.73% (see in bold in Table 4.7).

2. Clustering on the Wind Speed Estimated by a Time Series Model

Using a time series model is another way to estimate the wind speed $v(t)$ based on model in (4.16).

$$v(t) \approx v^*(t) = f(v(t-1),...,v(t-5)) \qquad (4.16)$$

Here, the five past values of wind speed $v(t-1),...,v(t-5)$ serve as inputs to a NN algorithm. The estimated wind speed $v^*(t)$ is used for clustering by the *k*-means algorithm, for k = 2,…, 5. The corresponding values of the generator torque $x_1(t)$, rotor speed $y_2(t-1)$, and power output $y_1(t-1)$ are used to build 30 NN models for each cluster. The best performing NN model is selected and the prediction results for the test data are shown in Table 4.10.

Table 4.10 Prediction performance based on the data clustered on the wind speed $v^*(t)$ of (4.16).

Input Space Clustered by	Number Of Clusters	Observed Average Power Output [kW]	Predicted Average Power Output [kW]	Mean Absolute Error [kW]	STD of MAE	Mean Relative Error [%]	STD of MRE
WS*(*t*)	**2**	**605.56**	**605.27**	**10.99**	**10.41**	**2.25**	**4.87**
WS*(*t*)	3	605.56	605.14	11.00	9.90	2.38	12.80
WS*(*t*)	4	605.56	605.40	11.03	10.62	2.38	11.39
WS*(*t*)	5	605.56	605.05	11.27	10.44	2.39	8.97

The best prediction accuracy it attained for $k = 2$ as shown in bold in Table 10. However, this prediction accuracy is worse than the base performance for the vector space (4.11) reported in bold in Table 4.7 with MAE = 10.98 kW, STD of MAE = 9.78 kW, MRE = 2.25 kW, and STD of MRE = 5.73 kW.

Though the wind speed directly contributes to the power output generation, clustering input space by the wind speed does not benefits prediction accuracy. This might be because the current wind speed $v(t)$ is not known and both estimations here carry error impacting the prediction accuracy.

3. Clustering on Generator Torque $x_2(t)$

Generator torque $x_2(t)$ is the most significant predictor with the importance of 94.28% among all 27 predictors (see Table 4.4). Therefore, it is used to cluster the input space with the corresponding rotor speed $y_2(t-1)$ and the power output $y_1(t-1)$ assigned to each cluster. The test data is categorized according to the nearest cluster centroid. The performance of the best NN models (out of 30 for each cluster) is shown in Table 4.11.

Table 4.11 Prediction performance based on the data clustered on generator torque $x_2(t)$.

Input Space Clustered by	Number of Clusters	Observed Average Power Output [kW]	Predicted Average Power Output [kW]	Mean Absolute Error [kW]	STD of MAE	Mean Relative Error [%]	STD of MRE
GT(t)	2	605.56	605.29	10.74	9.44	2.23	6.56
GT(t)	3	605.54	605.37	10.79	9.50	2.23	5.96
GT(t)	**4**	**605.56**	**605.14**	**10.58**	**9.11**	**2.16**	**5.48**
GT(t)	5	605.56	605.25	10.58	9.24	2.15	4.47

The best prediction accuracy is attained for $k = 4$ as shown in bold in Table 11. This prediction accuracy is significantly improved compared with the vector space of Eq. (12) with MAE = 10.98kW, STD of MAE = 9.78 kW, MRE = 2.25 kW, and STD of

MRE = 5.73 % (see in bold in Table 7). This prediction accuracy is also improved over the vector space of Eq. (10) with MAE = 10.72 kW, STD of MAE = 9.01kW, MRE = 2.20 %, and STD of MRE = 5.84% (see in italic in Table 4.7).

4. Clustering on Generator Torque $x_2(t)$ and Rotor Speed $y_2(t-1)$

Generator torque $x_2(t)$ and rotor speed $y_2(t-1)$ have a combined importance of 95.67% (94.28%+1.39%) of all the provided 27 predictors (see Table 4.4). In this scenario, they are used to cluster the input space with the corresponding power output $y_1(t-1)$ assigned to each cluster. The test data is categorized according to the nearest cluster centroid. The performance of the best NN model (out of 30 for each cluster) is shown in Table 4.12.

Table 4.12 Prediction performance for clustering by GT (t) and RS (t-1).

Input Space Clustered by	Number Of Clusters	Observed Average Power Output [kW]	Predicted Average Power Output [kW]	Mean Absolute Error [kW]	STD of MAE	Mean Relative Error [%]	STD of MRE
(GT(*t*), RS(*t*-1))	2	605.56	605.36	10.83	9.79	2.22	5.80
(GT(t), RS(t-1))	*3*	*605.56*	*604.10*	*11.17*	*10.07*	*2.25*	*5.13*
(GT(*t*), RS(*t*-1))	**4**	**605.56**	**605.05**	**10.78**	**9.70**	**2.21**	**5.76**
(GT(*t*), RS(*t*-1))	5	605.56	605.10	10.80	9.63	2.25	6.51

The best prediction accuracy occurs for $k = 4$ as shown in bold in Table 4.12. This prediction accuracy is significantly improved over the vector space of Eq. (12) with MAE = 10.98 kW, STD of MAE = 9.78 kW, MRE = 2.25 kW, and STD of MRE = 5.73 % (see in bold in Table 4.7). However, the improvement of prediction accuracy is not stable as when number of clusters $k = 3$, the prediction accuracy (see in italics in Table 12) are

worse than the vector space for Eq. (12) with MAE = 10.98 kW, STD of MAE = 9.78 kW, MRE = 2.25 kW, and STD of MRE = 5.73 % (see in bold in Table 4.7). Note that, best prediction accuracy (see in bold in Table 4.12) is no better than the vector space for Eq. (10) with MAE = 10.72 kW, STD of MAE = 9.01kW, MRE = 2.20 % and STD of MRE = 5.84% (see in italic in Table 4.7).

5. Clustering on generator torque $x_2(t)$, rotor speed $y_2(t-1)$ and power output $y_1(t-1)$

Generator torque $x_2(t)$, rotor speed $y_2(t-1)$, and the power output $y_1(t-1)$ have a cumulative importance of 96.13% (94.28%+1.39%+0.46%) of all 27 predictors (see Table 4.4). In this scenario, they are used to cluster the input space. Each test instance is labeled with the nearest cluster centroid. The number of clusters varies from 2 to 6. The performance of the best NN models (out of 30 for each cluster) is shown in Table 4.13.

Table 4.13 Prediction performance for clustering by GT(t), RS(t-1) and PO(t-1).

Input Space Clustered by	Number of Clusters	Observed Average Power Output [kW]	Predicted Average Power Output [kW]	Mean Absolute Error [kW]	STD of MAE	Mean Relative Error [%]	STD of MRE
(GT(*t*), RS(*t*-1), PO(*t*-1))	2	605.56	605.10	10.83	9.65	2.24	6.74
(GT(*t*), RS(*t*-1), PO(*t*-1))	3	605.56	605.04	10.73	9.30	2.16	4.71
(GT(*t*), RS(*t*-1), PO(*t*-1))	**4**	**605.56**	**605.30**	**10.72**	**9.33**	**2.19**	**5.17**
(GT(*t*), RS(*t*-1), PO(*t*-1))	5	605.56	605.22	10.76	9.55	2.21	5.69
(GT(t), RS(t-1), PO(t-1))	6	605.56	605.41	10.79	9.28	2.31	4.89

The best prediction accuracy for $k = 4$ is shown in bold in Table 4.13. This prediction accuracy is improved compared with vector space for Eq. (4.11) with MAE = 10.98 kW, STD of MAE = 9.78 kW, MRE = 2.25 kW, and STD of MRE = 5.73 % (see in

bold in Table 4.7). However, this prediction accuracy (see in bold in Table 13) is no better than the one for the vector in Eq. (4.9) with MAE = 10.72 kW, STD of MAE = 9.01kW, MRE = 2.20 % and STD of MRE = 5.84% (see in italic in Table 4.7).

6. Description of the Clustered Input Space

The most promising scenario for clustering the input space is based on the generator torque $x_2(t)$. The number of clusters $k = 4$ produces the best prediction accuracy among k = 2,…, 5. The centriods for the training data sets are shown in Table 4.14.

Table 4.14 Clustering centriods for training data set.

Cluster No.	GT(t)	PO(t)	PO(t-1)	RS(t-1)	Average Distance to the Centroid	Number of Instances	Instance Percentage (%)
1	13.89	112.84	115.13	12.17	0.05	8534	28.12
2	31.36	342.67	344.72	14.67	0.04	11993	39.51
3	48.28	658.93	665.12	17.99	0.06	6683	22.02
4	84.56	1258.93	1240.31	19.78	0.09	3143	10.35

Four clusters of the training data are illustrated in Table 4.14. The generator torque $x_2(t)$ is used to cluster input space by *k*-means algorithm. The clustering results represent four levels of the generator torque $x_2(t)$ from low to high with the corresponding observed power output $y_1(t)$. Clusters are arranged by the observed power output $y_1(t)$ from low to high, e.g., cluster 1 contains data with the lowest generator torque $x_2(t)$ and lowest observed power output $y_1(t)$; while cluster 4 includes data with the highest values of the generator torque $x_2(t)$ and the observed power output $y_1(t)$.

Instances of test data are categorized according to the cluster centriods of the training data using the algorithm in Table 4.3. After all instances are categorized, the

cluster centriods of the test data are measured by average values through each cluster for each parameter state (see equation (4.17)).

$$\text{TestClusterCentriod}_{i,j} = \frac{\sum_{1}^{N_i} PS_j}{N_i} \qquad (4.17)$$

where PS_j is the j^{th} parameter state, $1 \le j \le 3$; N_i is the number of instances of the i^{th} cluster, $1 \le i \le 4$.

Table 4.15 illustrates four cluster centriods of the test data set.

Table 4.15 Clustering centriods for the test data set.

Cluster No.	GT(t)	PO(t)	PO(t-1)	RS(t-1)	Average Distance to Centroid	Number of Instances	Instance Percentage (%)
1	17.55	148.87	164.4	12.4	0.05	715	4.51
2	32.21	357.51	362.6	14.93	0.04	6107	38.51
3	49.95	693.19	706.28	18.32	0.05	6910	43.57
4	79.79	1186.54	1132.69	19.72	0.09	2127	13.41

Four clusters of the test data set are illustrated in Table 4.15 for the parameters of Eq. (4.11), the observed power output $y_1(t)$, average distance to centriods, and the number of instances in each cluster. The similarity between training and test clusters is measured by the relative distance between the cluster centriods of the training and test data for each parameter state. A shorter relative distance between cluster centriods implies greater similarity between them. The relative distance metric is defined in (4.18):

$$\text{Relative Distance} = \frac{|\,\text{TrainingClusterCentriod} - \text{TestClusterCentriod}\,|}{\text{TrainingClusterCentriod}} \times 100\% \qquad (4.18)$$

The relative distances between cluster centriods of the training and test data sets for each predictor are shown in Table 4.16.

Table 4. 16 Relative distances between cluster centriods for four predictors.

Cluster No.	GT(*t*)	PO(*t*)	PO(*t*-1)	RS(*t*-1)
1	26.35	31.93	42.80	1.89
2	2.71	4.33	5.19	1.77
3	3.46	5.20	6.19	1.83
4	5.64	5.75	8.68	0.30

The relative distances between the cluster centriods of the training and test data sets are shown in Table 4.16. The distance values corresponding to clusters 2, 3 and 4 are smaller than that of cluster 1. Higher similarity (smaller distance) between the training and test data sets is a good indicator of a better prediction performance.

4.3.3 Model Extraction

1. Cluster One

In this section, test results by five data-mining algorithms are discussed. The tested algorithms include: the Random Forest Algorithm (RFA), Boosting Tree Algorithm (BTA), Support Vector Machine (SVM), Neural Network (NN), and a Neural Network Ensemble. The test performance of each algorithm for cluster 1 (of Table 4.15) is shown in Table 4.17.

Table 4.17 Test performance of five data-mining algorithms for cluster 1 data.

Algorithm	Observed Average Power Output [kW]	Predicted Average Power Output [kW]	Mean Absolute Error [kW]	STD of MAE	Mean Relative Error [%]	STD of MRE
RFA	148.87	148.40	9.62	9.84	6.46	88.75
BTA	**148.87**	**148.86**	**6.12**	**4.77**	**5.51**	**11.54**
SVM	148.87	164.40	6.48	4.93	4.35	24.46
NN (Ensemble)	148.87	148.63	6.25	4.95	4.20	25.62
NN	148.87	148.57	6.21	5.06	5.88	24.21

The average observed power output for cluster 1 in Table 4.17 is 148.87 kW. The previous research has shown that predicting power output at low levels with a NN resulted in large errors, e.g., when the observed power output was less than 10 kW, the relative error was as high as 300%, or even 1000%. Note that the errors reported in Table 4.17 for low values of power are smaller.

Of all algorithms in Table 4.17, the Boosting Tree Algorithm (BTA) outperformed than both the NN and NN (Ensemble) on the four metrics (MAE, MRE, STD of MAE, and MRE) as indicated in bold in Table 4.17. Therefore the BTA is selected for constructing the model based on cluster 1 data. The learning rate of $\alpha = 0.3$ has produced the best prediction accuracy.

2. Cluster Two

The test performance for the five data-mining algorithms for cluster 2 data (see Table 4.15) is shown in Table 4.18. As illustrated in Table 4.18, the average observed power output for cluster 2 is 357.51 kW. The best performance (see in bold) is attained by the NN selected of 30 NNs.

Table 4.18 Test performance of the five data-mining algorithms for cluster 2 data.

Algorithm	Average Observed Power Output [kW]	Average Predicted Power Output [kW]	Mean Absolute Error [kW]	STD of MAE	Mean Relative Error [%]	STD of MRE
RFA	357.51	356.86	24.40	22.39	6.45	5.82
BTA	357.51	357.53	11.16	10.07	3.25	2.96
SVM	357.51	362.60	9.75	7.92	2.83	2.28
NN (Ensemble)	357.51	357.29	8.98	7.70	2.59	2.17
NN	**357.51**	**357.24**	**8.94**	**7.62**	**2.58**	**2.14**

3. Cluster Three

The test results of the five data-mining algorithms for cluster 3 data (see Table 4.15) is shown in Table 4.19.

Table 4.19 Test performance of the five data-mining algorithms for cluster 3 data.

Algorithm	Observed Average Power Output [kW]	Predicted Average Power Output [kW]	Mean Absolute Error [kW]	STD of MAE	Mean Relative Error [%]	STD of MRE
RFA	693.19	693.60	38.79	43.69	5.22	5.57
BT	693.19	692.96	12.04	10.48	1.76	1.58
SVM	693.19	706.28	13.12	10.40	1.95	1.62
NN (Ensemble)	693.19	692.42	11.87	10.22	1.75	1.54
NN	**693.19**	**692.35**	**11.82**	**10.14**	**1.74**	**1.52**

The average observed power output of cluster 3 is 693.19 kW (Table 19). The best prediction accuracy is accomplished by at NN algorithm (the best one among 30 NNs tested).

4. Cluster Four

The test performance of the five data-mining algorithms on cluster 4 data (see in Table 4.15) is provided in Table 4.20.

Table 4.20 Test performance of the data-mining algorithms for cluster 3 data.

Algorithm	Observed Average Power Output [kW]	Predicted Average Power Output [kW]	Mean Absolute Error [kW]	STD of MAE	Mean Relative Error [%]	STD of MRE
RFA	1186.54	1185.95	21.87	20.99	1.82	1.66
BT	1186.54	1186.25	13.96	11.75	1.19	1.02
SVM	1186.54	1132.69	14.07	11.03	1.21	0.99
NN (Ensemble)	**1186.54**	**1187.14**	**12.90**	**10.77**	**1.11**	**0.96**
NN	1186.54	1187.03	12.93	10.77	1.11	0.96

Table 4.20 shows the average observed power output of 1186.54 kW. The best performance has been produced by the NN Ensemble algorithm with 5 (out of 30) best performing NNs.

5. Overall Results

The best prediction accuracy results for the four clusters are summarized in Table 4.21.

The average prediction accuracy reported in Table 4.21 has been significantly improved compared with performance based on the space vector in Eq. (4.11) with MAE = 10.98 kW, STD of MAE = 9.78 kW, MRE = 2.25 kW, and STD of MRE = 5.73 % (see in bold in Table 4.7). The average prediction accuracy is also higher than the one for Eq. (10) shown in Table 4.7 (MAE = 10.72 kW, STD of MAE = 9.01kW, MRE = 2.20 % and STD of MRE = 5.84%).

Table 4.21 Best performing clusters.

Cluster No.	Average Observed Power Output [kW]	Average Predicted Power Output [kW]	Mean Absolute Error [kW]	STD of MAE	Mean Relative Error [%]	STD of MRE	Algorithm
Cluster 1	148.87	148.86	6.12	4.77	5.51	11.54	BT
Cluster 2	357.51	357.24	8.94	7.62	2.58	2.14	NN
Cluster 3	693.19	692.35	11.82	10.14	1.74	1.52	NN
Cluster 4	1186.54	1187.14	12.90	10.77	1.11	0.96	NN (Ensemble)
Average	**605.56**	**605.14**	**10.57**	**9.08**	**2.14**	**3.08**	

4.3.4 Comparative Study

1. Comparison of Four Typical Scenarios

Four typical scenarios from Sections 4, 5, and 6 are compared.

Scenario A: 3 model inputs (generator torque $x_2(t)$, rotor speed $y_2(t-1)$ and power output $y_1(t-1)$) are used by NN model. The best performing NN out of 30 NNs is selected.

Scenario B: 27 model inputs (see Eq. (10)) are used by a NN model. The best performing NN out of 30 NNs is selected.

Scenario C: 3 model inputs (generator torque $x_2(t)$, rotor speed $y_2(t-1)$ and power output $y_1(t-1)$) are used by a NN model. Input space is clustered by the generator torque $x_2(t)$ with the *k*-means algorithm. A NN algorithm is used to extract prediction model for each cluster. For each extraction, 30 NNs are produced and the one with best performance is selected.

Scenario D: 3 model inputs (generator torque $x_2(t)$, rotor speed $y_2(t-1)$ and power output $y_1(t-1)$) are used by data-mining models. Input space is clustered by

generator torque $x_2(t)$ using the *k*-means algorithm. For each cluster, an algorithm with the best prediction performance is selected to extract the model (see in Table 21).

Table 4.22 lists the prediction accuracy results of test data for the four scenarios.

Table 4.22 Comparison of test results.

Scenario	Average Observed Power Output [kW]	Average Predicted Power Output [kW]	Mean Absolute Error [kW]	STD of MAE	Mean Relative Error [%]	STD of MRE
A	605.56	605.17	10.98	9.78	2.25	5.73
B	605.56	605.56	10.72	9.01	2.20	5.84
C	605.56	605.14	10.58	9.11	2.16	5.48
D	**605.56**	**605.14**	**10.57**	**9.08**	**2.14**	**3.08**

Scenario D produces the best prediction accuracy on most metrics (MAE, MRE, and STD of MRE) in Table 4.24. The STD of MAE ranks the second best and it is only slightly worse (by 0.07kW) than the ones produced in Scenario B.

The results reported in this research and partly summarized in Table 4.22 lead to the following conclusions:

(1) Prediction accuracy improves as the number of inputs increases

As indicated by Scenario A and B, prediction accuracy improves with the increase in number of predictors. For example, when n decrease from 27 to 3, MAE increases 0.26 kW based on an average observed power of 605.56 kW for 15860 test instances. The overall absolute error increased by 4123kW.

(2) Prediction accuracy improves when clustering the input data by the generator torque $x_2(t)$ is used

Clustering input space by the generator torque $x_2(t)$, reduces MAE by 0.4 kW (compared with the base performance of vector space in Eq. (4.12) shown in Table 4.7. This is for the average observed power of 605.56 kW for 15860 test instances. The overall absolute error is reduced by 6344 kW.

(3) Extracting models from partitioned data improves prediction accuracy

Each of the four models in Scenario C was extracted by a NN algorithm (one model per cluster). In Scenario D models were extracted for each cluster using the best performing algorithm. For example, when the observed power output is low, the NN algorithm generates rather large error. The BTA has resulted in a model with small MAE, MRE, and STD of MAE. The improvements in the STD of MRE are particularly impressive. For Scenario C and D, the following improvements have been accomplished: MAE by0.01 kW, MRE by 0.02%, STD of MAE by 0.03, and STD of MRE by 2.4%.

2. Benefits from the Proposed Approach

The proposed approach has a number of benefits, including:

(1) Higher prediction accuracy is achieved by using fewer parameters as model inputs.

High prediction accuracy is attained for just 3 inputs when the input space is clustered by the generator torque $x_2(t)$. The base line is the performance for vector of 27 predictors in Eq. (4.9) without clustering the input space.

(2) Shortened training time is achieved due to the following three reasons:

(a) Using smaller number of input parameters;

(b) Clustering data produces subspaces with fewer instances included in each input space;

(c) Training can be accomplished in parallel which offers computational benefits.

(3) The proposed model is more reliable.

Using fewer model inputs enhances reliability. In particular, if n parameters are selected as inputs, the probability of not getting a poor (in error) prediction is expressed in (4.19).

$$p(\text{prediction not in error}) = \prod_{i=1}^{n}(1 - p(\text{ input in error})) \qquad (4.19)$$

where $p(\text{input in error})$ is the probability of an erroneous input. Assuming the probability distribution function is uniform and equal to 1%; then for k = 3, $p(\text{prediction not in error}) = (1-0.01)^3 = 0.97$. Note that for $n = 27$, $p(\text{prediction not in error}) = (1-0.01)^{27} = 0.76$, which is a significant decrease.

(4) The proposed models are customized

Training data is clustered into mutually separable subspaces. The most suitable data-mining algorithm is applied to the data of each cluster. This makes the models specialized to the data included in the various clusters.

4.4 Summary

Predicting future production of wind power at low wind speeds is challenge. Estimates of power at low levels are extremely inaccurate. This paragraph proposed a clustering-based method for power prediction at low wind using 10-s data. Numerous data mining algorithms were developed for low level power prediction, including neural networks (NNs) that dominate in wind energy industry. The NN models produce large errors when it comes to predicting power output allow levels.

The approach proposed in the paragraph takes advantage data subspaces that led to accurate predictive models. The successful results reported in the paragraph were

accomplished in a number of steps. First, significant parameters were selected by physics-based equations and data-mining algorithms. This parameter selection showed that using larger number of model inputs resulted in a prediction accuracy gain. Second, training and test data sets were clustered according to five different criteria (scenarios). The clustering algorithm used here was the *k*-means algorithm. Third, for the data in each cluster the most suitable algorithm for building a power prediction model was identified.

The computational results reported in the paragraph demonstrated that the proposed model customization approach produced accurate prediction models using a small number of input parameters. The latter provided important side benefits, including reduced computational effort and increased reliability.

CHAPTER 5.
DYNAMIC CONTROL OF WIND TURBINES

5.1 Introduction

Reduction of operations and maintenance costs [4, 5] is key to the expansion of wind industry destine to grow in the next decades [3]. Development of control solutions [38, 77, 78, 43] is a valid approach to reduce these costs. A well designed wind turbine control system should maximize not only the energy captured from the wind but also extend the lifetime of turbine components, e.g., the gearbox. It is known that smoothing the turbine power output is important in its integration with the electricity grid. All operations and maintenance considerations have to be properly managed. For example, for a high wind speed and low electricity demand, a wind turbine can be operated so that smoothing the rotor speed and generator torque becomes a priority. Wind turbine control technology is relatively new, and opportunities exist to improve turbine performance in the presence of operations and maintenance constraints. An intelligent wind turbine control system increasing competitiveness of wind energy is needed.

In this section, an intelligent system for control of wind turbines is introduced. Unlike the research reported in the literature [7, 38-45, 69, 78], where wind turbine models are obtained from the first principles and aerodynamics, in the proposed intelligent control system, data mining algorithms extract the turbine models from the process data, i.e., SCADA (Supervisory Control and Data Acquisition) system. A time series model [9, 24, 75] is used to predict wind speed, and MPC (Model Predictive Control) optimizes the process variables of wind turbines.

The model considered in this section considers five weighted objectives. The weights are adjusted for the eight typical scenarios defined by wind conditions and operational requirements.

5.2 Data Description and Methodology for Dynamic Control of Wind Turbines

5.2.1 Data Description

Modern wind turbines are equipped with sensors for control and monitoring purposes. The data sampling frequency can be high (e.g., milliseconds), however, for specific applications, the high frequency data is usually aggregated (e.g., averaged) over a certain time period (e.g., 10 seconds, 10 minutes). At present, the 10-minute data standard is widely used in industry. Analysis of 10-minute data does not allow observing important details. Figure 5.1 through Figure 5.3 show the power curves of a 1.5 MW wind turbine when it is operating between the cut-in and the rated wind speed. The power curve shown in Figure 5.1 is scattered though the status of the wind turbine and is considered to be normal. It can be observed that the spread of the points which create the power curve increases around the rated wind speed. This increased variability translates into variable power output and results in loads that could be hazardous to the drive train components.

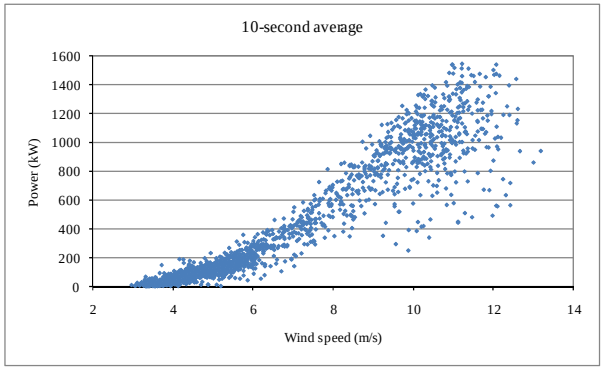

Figure 5.1 Power curve of a 1.5 MW wind turbine for 10-second average data.

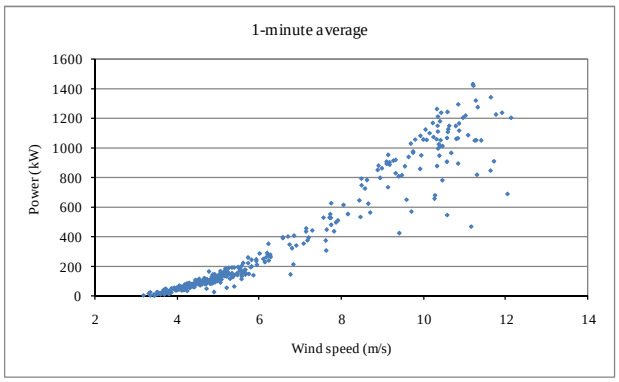

Figure 5.2 Power curve of a 1.5 MW wind turbine for 1-minute average data.

Averaging the 10-second data over the 1-minute data has resulted in a power curve with a reduced spread of the data points (see Figure 5.2). For the 10-minute average data (the industry standard), the data points follow a typically displayed pattern (see Figure 5.3). Note that the scattered graph in Figure 5.3 is due to insufficient number of data points included in the 10-second data set.

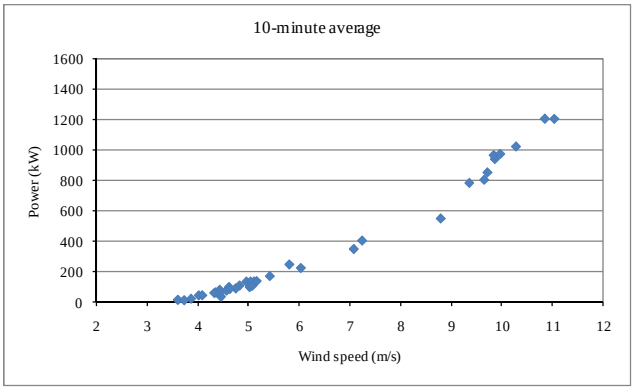

Figure 5.3 Power curve of a 1.5 MW wind turbine for the 10-minute average data.

The results presented in Figure 5.1 through Figure 5.3 indicate that the wind turbine control system considered in this research needs to be improved. An intelligent wind turbine control system is discussed and illustrated with the data from a 1.5 MW wind turbine. Table 4.2 lists the process parameters used in the study.

Two major controllable parameters (i.e., the parameters optimizing the energy conversion process) are the blade pitch angle and the generator torque. In addition, three response parameters of interest to this research are the wind speed, rotor speed, and wind turbine power output. They are important indicators of the energy conversion process. The rotor speed and the generator speed are highly correlated, and in fact they are modeled by a linear function. In this paragraph, the rotor speed is selected as a response parameter to be included in an objective function. Rapid changes in the rotor speed accelerate the failure of its mechanical components.

The data used for this case study was collected at a sampling interval of 10 seconds for a day. This data set satisfactorily represents the data generated at different turbines across many time horizons. The wind speed varied in the interval [2.97 m/s, 13.16 m/s]. After initial denoising (e.g., removing turbine down time, wind speed below the cut-in speed, and the wind above the cut-out speed), 2054 valid data points were considered. The data for the wind speed between the cut-in and the cut-out speed ranges was considered in particular because this operational region presents a major opportunity to optimize the wind turbine power generation process. The distribution of data used in this paragraph is shown in Figure 5.4.

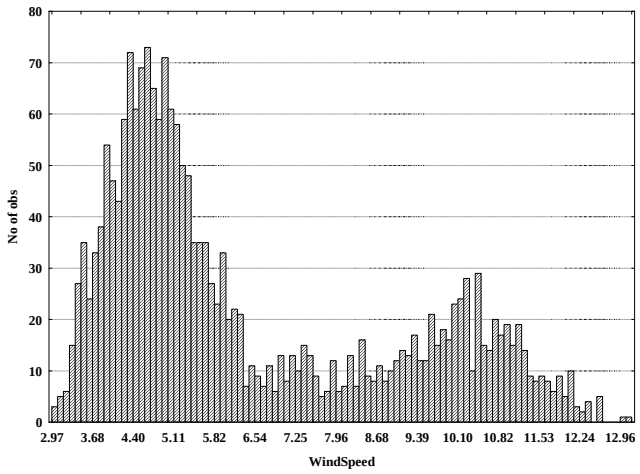

Figure 5.4 Wind speed distribution of 2054 data points.

5.2.2 Data Driven Intelligent Wind Turbine Control System

In this paragraph, an intelligent wind turbine control system is developed using data mining algorithms to optimize the wind turbine energy conversion process when the wind speed is between the cut-in and the rated speed. Figure 5.5 shows the basic components (two modules) of the control system and its information flow.

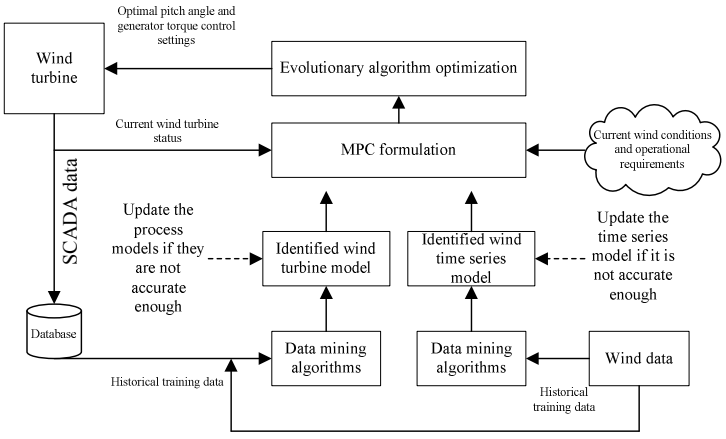

Figure 5.5 Intelligent control system of wind turbine.

The first module extracts the process and the time series models from the historical SCADA and wind data. These models are updated once their performance degrades, which is accomplished by comparing the model's predicted values with the actual measurements recorded by SCADA and anemometers.

A time series model predicts wind speed at short time intervals, which can be traced back to [13], where a linear wind speed time series model was built. Once an accurate wind speed time series model is obtained, wind turbine is optimized with a model predictive control algorithm. Although several different algorithms are compared, such as the boosting tree regression [66, 67], linear regression, and neural network [81, 90] the wind speed time series model (5.1) is built by a neural network ensemble, and the predictor is selected using the genetic wrapper approach [71, 73].

$$v(t) = g(v(t-1), v(t-2), v(t-6)) \tag{5.1}$$

The process models (5.2) and (5.3) are built by a neural network algorithm [76, 32]. Since the sampling rate is 10 seconds, and the time delay of the wind turbine control system is not longer than 20 seconds, one previous state is sufficient in building models.

$$y_1(t) = f_1(y_1(t-1), x_1(t), x_1(t-1), x_2(t), x_2(t-1), v(t), v(t-1)) \tag{5.2}$$

$$y_2(t) = f_2(y_2(t-1), x_1(t), x_1(t-1), x_2(t), x_2(t-1), v(t), v(t-1)) \tag{5.3}$$

The second module includes the MPC (model predictive control) component and an EC (evolutionary computation) solver. The MPC model is formulated based on the extracted models and the current wind turbine status as well as the wind speed. Then current wind conditions and operational requirements are used to determine the importance of the optimization objectives of the MPC model (here weights associated with the MPC objective functions). Updating process models is an important issue in

MPC control; however, this paragraph does not focus on this topic. Rather it emphasizes identification of various optimization objects according to different wind conditions and operational requirements.

Control systems in the presently available wind turbines usually follow one control strategy, i.e., when the wind speed is between the cut-in and rated one, the power output is optimized by following the maximum theoretic optimal power coefficient. Actually there are other factors to be considered besides maximization of the power output. For example, smoothing the power output or rotor speed variation is important depending on the wind conditions and operational demands. An intelligent control system should provide more options to control a wind turbine based on the wind conditions and operational requirements. Five different objectives are considered in optimizing the energy conversion process as shown in (5.4).

$$J = w_1 J_{Power} + w_2 J_{Rotor} + w_3 J_{P\_ramp} + w_4 J_{G\_ramp} + w_5 J_{Pitch\_ramp} \tag{5.4}$$

where: J_{Power} is a function to minimize the distance between the power output to its upper limit and therefore maximizing the power output,

J_{Rotor} is a function to minimize rotor speed ramp,

J_{P_ramp} is a function to minimize power output ramp,

J_{G_ramp} is a function to minimize generation torque ramp,

J_{Pitch_ramp} is a function to minimize pitch angle ramp.

The weights w_1, w_2, w_3, w_4, w_5 in (5.4) are assigned different values depending on the priority assigned to the corresponding objective. Each weight is in the interval [0, 1], and $w_1 + w_2 + w_3 + w_4 + w_5 = 1$. The MPC model is defined in (5.5):

$$\min J(w_1, w_2, w_3, w_4, w_5, J_{Power}, J_{Rotor}, J_{P\_ramp}, J_{G\_ramp}, J_{Pitch\_ramp}) \tag{5.5}$$

It is worth mentioning that minimizing $J(w_1, w_2, w_3, w_4, w_5, J_{Power}, J_{Rotor}, J_{P\_ramp}, J_{G\_ramp}, J_{Pitch\_ramp})$ is a challenge as the objective function and the constraints are nonlinear. To solve this type of optimization problem, an evolutionary strategy algorithm is proposed [74]. However, this paragraph focuses on analysis of the weights for different scenarios rather than the solution of the function.

5.3 Industrial Case Study on Dynamic Control of Wind Turbine

5.3.1 Adjusting Objectives Based on Wind Conditions and Operational Requirements

In this paragraph, the following three factors, wind speed, wind turbulence, and electricity demand, are considered in deriving the control priorities (i.e., the weights w_1, w_2, w_3, w_4, w_5).

Wind turbulence is an important metric describing the degree of variability of the wind speed. If the current wind speed is highly variable, maximizing the power capture may significantly damage the turbine‘s mechanical components. An intelligent wind turbine control system should optimize the control priorities based on wind turbulence.

1. Wind turbulence intensity

Turbulence intensity is a measure of the overall level of wind turbulence [30], and it is defined in (5.6).

$$I_t = \frac{\sigma_t}{\bar{v}_t} \tag{5.6}$$

where σ_t is the standard deviation of the wind speed variation about the mean wind speed $\bar{v}_t$ at time stamp t, and $\bar{v}_t$ is the mean wind speed over a certain interval at time t.

The mean wind speed $\bar{v}_t$ over n consecutive sampling data points is defined in (5.7):

$$\bar{v}_t = \frac{1}{n} \sum_{i=t-n+1}^{t} v(i) \tag{5.7}$$

where n is the number of consecutive data points.

The standard deviation σ_t is computed as follows (see (5.8)):

$$\sigma_t = \sqrt{\frac{1}{n-1}\sum_{i=t-n+1}^{t}(\mathrm{var}_i - \overline{\mathrm{var}})^2} \tag{5.8}$$

where var_t is the variation of wind speed at time t, and is computed from (5.9):

$$\mathrm{var}_t = v(t) - \bar{v}_t \tag{5.9}$$

where $\overline{\mathrm{var}}_t$ is the mean wind speed variation at time t as shown in (5.10):

$$\overline{\mathrm{var}}_t = \frac{1}{n}\sum_{i=t-n+1}^{t}\mathrm{var}_i \tag{5.10}$$

The wind turbulence intensity (5.6) implies that current wind turbulence intensity at time t is estimated based on historical data points. Alternatively, the wind turbulence intensity could be computed based on the data obtained from the time series prediction model.

In this paragraph, the turbulence intensity is computed over a 1-minute time horizon (i.e., 6 consecutive 10 second data points) or a 5-minute time horizon (i.e., 30 consecutive 10 second data points).

Figure 5.6 shows the distribution of turbulence intensity computed over a 1-minute time interval. According to Figure 5.6, I_t is generally smaller than 0.274 most of the time, and the mode of I_t is around 0.069.

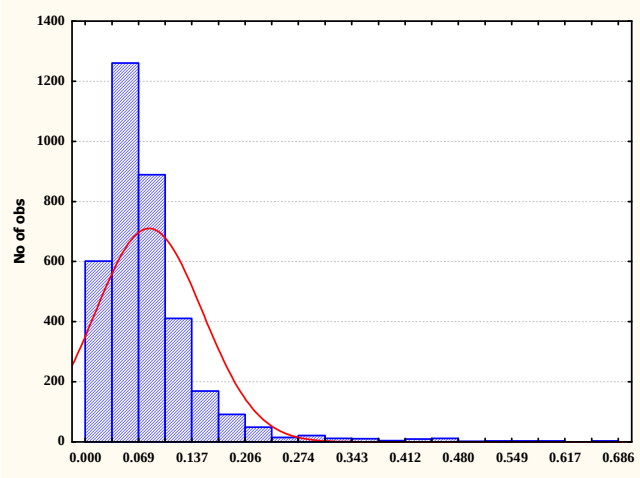

Figure 5.6 Turbulence intensity distribution (1-minute time intervals)

In comparing the distributions of turbulence intensity calculated from one minute to five minute time intervals, the common trend is obvious. For example, most turbulence intensities are between 0.03~0.36, and the highest frequency occurs around 0.1, though there are some small deviations.

For an intelligent wind turbine control system, 1-minute data reflects resolution, and it is suitable to classify the current turbulence status. In this paragraph $I_t = 0.06$ is used as the threshold to distinguish between high turbulence intensity and low turbulence intensity. Note that this threshold value was established based on the specific data sets. More research could be conducted in the future to find a better classification rule.

2. Wind conditions

Wind speed is an important factor to be considered by the intelligent control system. For control purposes, three regions are usually considered [42]:

1) Wind speed smaller than the cut-in speed.
2) Wind speed larger the rated speed.
3) Wind speed that is between cut-in and the rated speed.

To optimize the wind turbine operations, this paragraph suggests further classification of the wind speed region between the cut-in and the rated wind speed. In

this way, more precise control strategies can be implemented. The data used in this paragraph is collected from a 1.5 MW turbine with cut-in speed of $3.5\,m/s$ and rated speed of $12.5\,m/s$. The speed of $7\,m/s$ is used as an additional threshold category (see Figure 4) to define the following four wind speed scenarios, seen in Table 5.1.

Table 5.1 Classification of four wind speed scenarios.

No	W/T	Wind Speed Characteristic				Turbulence Intensity Characteristic			
		Min Wind Speed [m/s]	Ave Wind Speed [m/s]	Max Wind Speed [m/s]	Percentage of wind speed higher than 7 m/s	Min Turbulence Intensity	Ave Turbulence Intensity	Max Turbulence Intensity	Percentage of Turbulence intensity higher than 0.06
1	High/High	7.43	9.63	12.22	100.00%	0.03	0.06	0.16	54.84%
2	High/Low	7.02	9.58	11.66	100.00%	0.02	0.05	0.09	43.33%
3	Low/High	3.75	5.88	8.47	17.24%	0.04	0.16	0.32	89.66%
4	Low/Low	5.29	6.11	6.70	0.00%	0.01	0.03	0.06	4.17%

Table 5.1 illustrates the classification of wind conditions according to wind speed and turbulence intensity. The threshold of wind speed is 7m/s and threshold of turbulence intensity is 0.06. Take scenario 1 for example, as all the wind speed is higher than 7m/s, this period of wind condition is defined as high wind speed period. As 54.84% turbulence intensity is higher than 0.06 and the maximum is around 0.16, the scenario is defined as high turbulence intensity. Similarity, scenario 2 is characterized as high wind speed and

low turbulence intensity. Scenario 3 is characterized as low wind speed and high turbulence intensity and Scenario 4 is low wind speed and low turbulence intensity.

3. Electricity demand

Electricity demand is important in turbine control because when it is low, there is no need to maximize the power output, and more attention should be given to smoothing the power output. On the other hand, if the demand is high, more emphasis should be given to maximizing the power output. In this paragraph electricity demand (i.e., low or high) is also considered as a factor in classifying the operational scenarios.

4. Classification of operational scenarios

Table 5.2 Scenario classification according to wind status and electricity demand.

Scenario Number	Wind Speed	Turbulence Intensity	Demand	Weight				
				w_1	w_2	w_3	w_4	w_5
1	High (7 ~ 12.5 m/s)	High (I > 0.06)	High	0.4	0.15	0.2	0.2	0.05
2			Low	0.2	0.25	0.25	0.25	0.05
3		Low (I < 0.06)	High	0.6	0.1	0.15	0.15	0
4			Low	0.1	0.25	0.25	0.25	0.15
5	Low (3.5 ~ 7 m/s)	High (I > 0.06)	High	0.55	0.15	0.15	0.15	0
6			Low	0.2	0.25	0.25	0.25	0.05
7		Low (I < 0.06)	High	0.6	0.1	0.1	0.1	0.1
8			Low	0.2	0.2	0.2	0.2	0.2

Eight operational scenarios are discussed in this paragraph according to the turbulence intensity, wind speed, and electricity demand. Table 5.2 shows the details of various scenarios and weights used in model (5.4). Each scenario represents a combination of weights, which differentiate the importance of the five objectives, where $w_1 + w_2 + w_3 + w_4 + w_5 = 1$ and w_1, w_2, w_3, w_4, w_5 are between 0 and 1 (see Table 5.3). The weight combinations are derived based on the heuristic domain knowledge. More research is needed to algorithmically generate these weights.

5.3.2 Computational results

The computational results reported in this section are based on the dynamic equations extracted by the neural network algorithm and the wind speed time series model built from a neural network ensemble. The MPC model (5.5) is solved by an evolutionary strategy algorithm for constrained optimization problems with certain fixed parameter settings (i.e., the population size, selection pressure). For each scenario, during a fixed time period, model (5.5) is solved to find the optimal pitch angle and generator torque settings for the starting time stamp *t*. For the next sampling time $t+1$, model (5.5) is solved again based on the previously found optimal control settings at time *t* (i.e., pitch angle and generator torque). This simulation continues until the fixed time period ends. Then the optimized wind turbine status is compared with the original one for that fixed time period.

1. Summary of optimization results

Table 5.3 illustrates the optimization results of power output for computational eight scenarios. The power output has increased in each of the scenarios listed in Table 4, except scenario 4 (shown in bold), caused by the low electricity demand. STD (Standard Deviation) of power output implies the quality and smoothness of power output. Five optimized scenarios have smaller values of STD of power output than the original ones.

In scenarios 5, 7 and 8 of Table 5.4 the power quality is diminished in order to satisfy the electricity demand (shown in bold).

Table 5.3 Summary of power output generation.

Scenario Number	Original Average Power Output [kW]	Optimized Average Power Output [kW]	Original STD of Power Output [kW]	Optimized STD of Power Output [kW]
1	795.01	1278.36	171.99	153.80
2		1098.36		87.24
3	907.60	1427.70	228.90	125.61
4		**554.90**		55.69
5	200.40	372.62	139.82	**226.93**
6		271.72		153.23
7	275.00	535.55	36.97	**64.58**
8		498.89		**62.51**

Table 5.3 illustrates the smoothness of rotor speed, blade pitch angle, and generator torque. Optimization has resulted in smoother values of rotor speed in eight scenarios. Comparing with the original blade pitch angle and generator torque, the optimized values are smoother in high wind speed scenarios (1-4). At low wind speed (scenarios 5 through 8 of Table 5.5), the smoothness of blade pitch angle and generator torque (shown in bold) has diminished to benefit the power output.

The details of two illustrative operational scenarios 4 and 7 of Table 5.4 are discussed next.

Table 5.4 Summary of standard deviations for four parameters.

Scenario Number	Original STD of Rotor Speed [rpm]	Optimized STD of Rotor Speed [rpm]	Original STD of Blade Pitch Angle [°]	Optimized STD of Blade Pitch Angle [°]	Original STD of Generator Torque [Nm]	Optimized STD of Generator Torque [Nm]
1	0.25	0.24	4.39	1.91	1152.89	881.82
2		0.10		2.47		563.29
3	0.37	0.27	3.57	3.96	1470.40	747.91
4		0.09		2.15		400.58
5	1.79	1.69	4.09	**4.54**	934.38	**1614.48**
6		1.39		**4.34**		**1142.18**
7	0.82	0.30	0.00	**2.36**	207.19	**438.80**
8		0.40		**0.72**		**446.31**

2. Operational scenario 4: High wind speed, low turbulence intensity, and low electricity demand

Scenario 4 is concerned with high wind speed, low turbulence intensity and low electricity demand. As electricity is demand low, w_1 is set as $0.1, w_2$, w_3 and w_4 are set as 0.25, w_5 is set as 0.15 (see Table 3). In this case, the optimized power output is lower but much smoother than the original one. Due to low electricity demand, the increase of generation is diminished to improve the power quality. In the same way, rotor speed, generator torque and blade pitch angle are smoothed after optimization. Figure 5.7 through Figure 5.10 show the results of optimization of the power output, rotor speed, generator torque, and pitch angle.

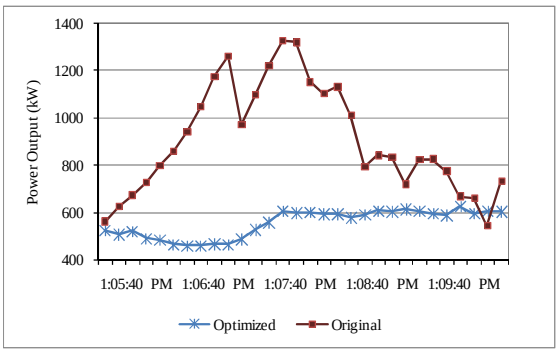

Figure 5.7 Original and optimized power output in the interval "1:05:40 PM" to "1:10:30 PM".

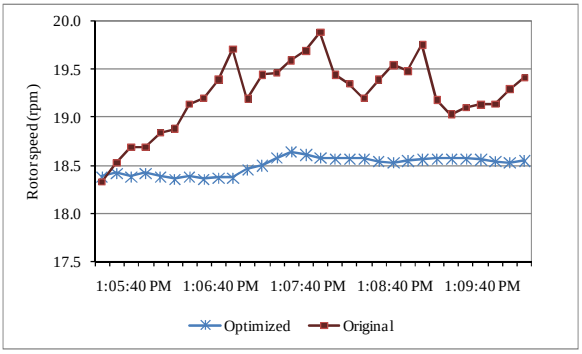

Figure 5.8 Original and optimized rotor speed in the interval "1:05:40 PM" to "1:10:30 PM".

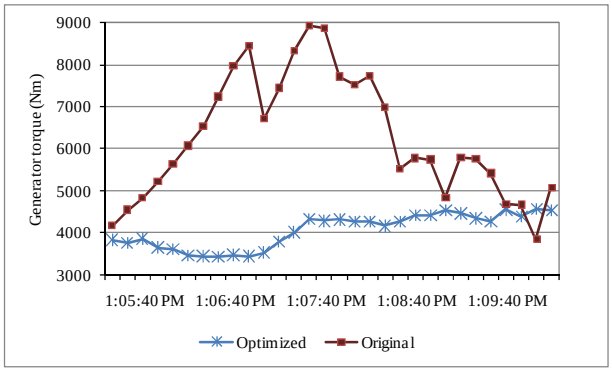

Figure 5.9 Original and optimized generator torque in the interval "1:05:40 PM" to "1:10:30 PM".

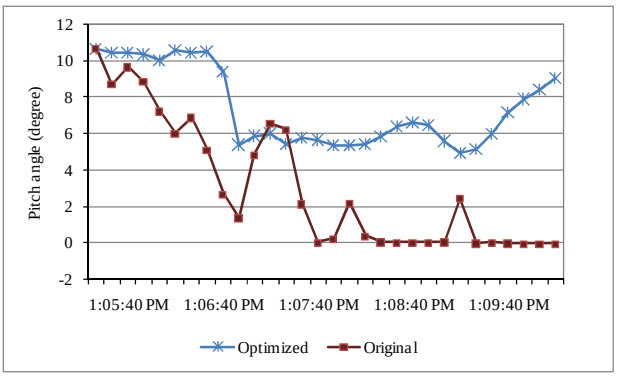

Figure 5.10 Original and optimized pitch angle in the interval "1:05:40 PM" to "1:10:30 PM".

3. Operational scenario 7: Low wind speed, low turbulence intensity, and high electricity demand

Here is the situation of low wind speed, low turbulence intensity and high electricity demand. As wind speed is low, the original low power output cannot satisfy the high electricity demand. So higher w_1 is needed. In this case, w_1 is set as 0.6, w_2, w_3, w_4 and w_5 are set to 0.1. Figure 5.11 through Figure 5.14 show the results of optimization of power output, rotor speed, pitch angle, and generator torque.

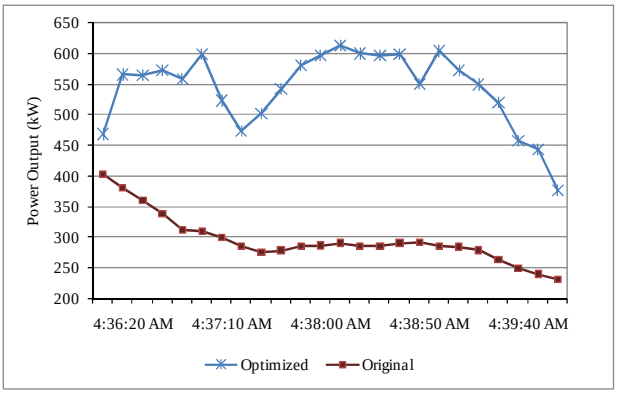

Figure 5.11 Original and optimized power output in the interval "4:36:20 AM" to "4:40:10 AM".

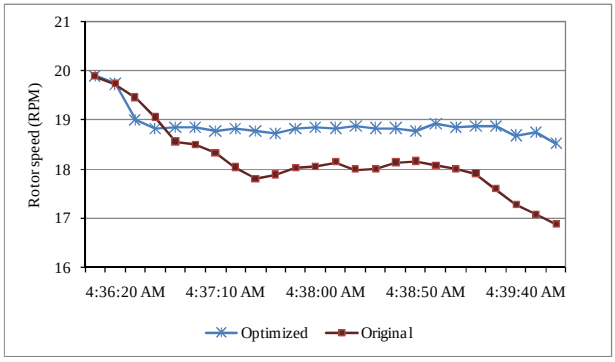

Figure 5.12 Original and optimized rotor speed in the interval "4:36:20 AM" to "4:40:10 AM".

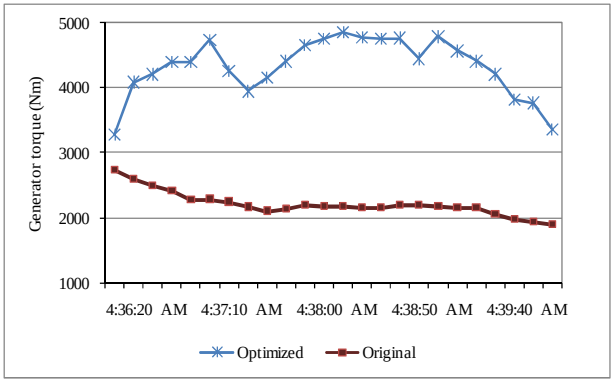

Figure 5.13 Original and optimized generator torque in the interval "4:36:20 AM" to "4:40:10 AM".

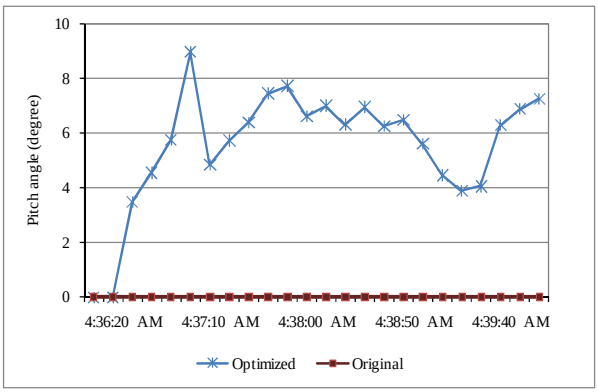

Figure 5.14 Original and optimized pitch angle in the interval "4:36:20 AM" to "4:40:10 AM".

As presented from Figure 5.11 to Figure 5.14, the generated power has increased at the expense of its quality to satisfy the electricity demand. Rotor runs smoother after optimization. The original generator torque does not vary much and the original blade pitch angle is constant for the range of wind speeds considered in this scenario. The optimized blade pitch angle and generator torque have resulted in increased the power output.

5.4 Summary

In this paragraph, an intelligent system for control of wind turbines was presented. The system integrated data mining, evolutionary computation, predictive control, and time series approaches. A time series model for prediction of wind speed was proposed. A dynamic function was built with five different weights determined by various operational scenarios. The system modifies the control objectives by observing the wind conditions and electricity demand. Wind conditions were characterized by the wind speed and wind turbulence intensity. Eight scenarios were defined based on the wind conditions and the electricity demand.

Each scenario was illustrated with computational results. The original wind turbine status and the optimized ones were compared by simulation. The results produced by the intelligent control system are better than those of the current wind turbine control system. For turbulent wind, the intelligent control system smoothed the power output, generator torque, and rotor speed without compromising the electricity demand.

Further research should focus on automating the weight generation based on wind conditions and electricity demand. Other objectives could be considered by the predictive control model to explore other aspects of the wind energy conversion process. Additional research is needed to improve the prediction accuracy of the wind speed, which is of importance in the proposed approach. Denoising techniques could be applied to enhance the data quality.

CHAPTER 6.

THE PREDICTION AND DIAGNOSIS OF WIND TURBINE FAULTS

6.1 Introduction

Wind power is key to meeting the planned targets of the carbon emission reductions and ensuring diversity of energy supply sources [82]. The growing interest in wind energy has led to the rapid expansion of wind farms [83, 84].

The growth of wind power has increased interest in the operations and maintenance of wind turbines. As wind turbines are located at remote locations that may be difficult to access, their maintenance becomes an issue. Thus, condition monitoring and fault diagnosis of wind turbines are of high priority.

This section proposes a methodology for system-level fault diagnosis in wind turbines using a data-driven approach. The fault-related data is analyzed at three levels. The existence of a status or a fault is predicted (Level 1), the category (severity) of the fault or the status is determined (Level 2), and the specific fault is predicted (Level 3).

6.2 Data Description

The data available for the research reported in this paragraph has been collected by SCADA systems at four wind turbines (Turbine 1, Turbine 2, Turbine 3, and Turbine 4). For each turbine, two separate sets of data were provided: SCADA data and status/fault data. Both data sets were collected at period of three months from 01/04/2009 to 30/06/2009. The details of the data are discussed next.

6.2.1 Data Description and Pre-processing

1. SCADA data

The SCADA data for four wind turbines was collected at 5-minute intervals. The nearly 25000 records (instances) collected for each turbine on over 60 parameters have been grouped into four categories.

1) *Wind parameters:* Wind parameters are the direct measurements of the wind (e.g., wind speed, wind direction) and derived values (e.g., wind intensity and turbulence).

2) *Energy conversion parameters:* Parameters in this category are related to the energy conversion process (e.g., power output, blade pitch angle, generator torque, rotor speed) and so on.

3) *Vibration parameters:* Vibration parameters indicate operational conditions of the turbine systems. They usually involve measurements of the drive train acceleration and tower acceleration.

4) *Temperature parameters:* This category of parameters includes the temperature measured at turbine components (e.g., bearing temperature) and the air temperature around turbine components and subsystems (e.g., nacelle interior temperature).

2 Status/Fault data

Status/fault data provides information on statuses and faults recorded by the SCADA system. A fault, in this paragraph, refers to a status that with a certain probability results in a severe consequence to the wind turbine system. For example, ignoring the status “Emergency stop nacelle/hub” or “Pitch thyristor 1 fault” might damage the wind turbine components. Other statuses, however, such as “No errors” and “Remote start” may not lead to severe consequences. Examples of status codes are illustrated in Table 6.1.

Each status code in Table 6.1 is associated with a specific abnormality of a turbine component or a subsystem. There are nearly 350 different status codes in the data considered in this research. The status text in Table 1 provides a short description of the

status, and the category denotes its severity. Category “1” implies the most severe status, and Category “4” corresponds to the least severe status.

The status/fault data has been collected by the SCADA system. Nearly 7000 occurrences of status codes have been observed at each turbine over the three-month period, including the seven parameters illustrated in Table 6.2.

Table 6.1 Sample status codes.

Status Code	Status Text	Category
1	Program start PLC	2
2	No errors	4
3	Manual stop	4
4	Remote stop	4
5	Remote start	4
6	System OK	4
9	Under-voltage	4
21	Cable twisting left	4
25	No speed reduction with primary braking	1
28	No speed reduction with secondary braking	1

Table 6.2 Parameters related to the fault information.

Parameter Name	Definition	Unit	Symbol
Fault time	Date and time of the fault occurrence		t_{fault}
Status code	Status code assigned to the fault		
Category	Category of the status code (four categories)		*Category*
Generator speed	Generator speed at the time the fault occurred	Nm	$GS(t_{fault})$
Power output	Power production at the time the fault occurred	kW	$PO(t_{fault})$
Wind speed	Wind speed at the time the fault occurred	m/s	$WS(t_{fault})$

3 Issues with status/fault data

Although the raw data contained over 7000 status/fault instances for each of the four wind turbines, some of the instances could not be considered for the following reasons:

1) Presence of wind speed measurements with unreasonably large values, as illustrated in Table 6.3.

Table 6.3 Illustration of data instances with out-of-range values of wind speed.

Date	Time	Status Code	Wind Speed	Power Output	Generator Torque
4/9/2009	4:24:10 AM	0	**-42946720**	-1	0
4/9/2009	4:34:10 AM	0	**-42946720**	-1	0
6/25/2009	1:54:54 AM	183	**32509316**	-2	0
6/25/2009	1:54:54 AM	183	**27872676**	-1	0

The wind speed measured by an anemometer should be in the range [0, cut-out speed], here [0 m/s, 21m/s].

In this case, the negative values of wind speed were assigned status code “0”, and the positive out-of-range speed was assigned status code “183” (see Table 3). However, the status code “183” is not unique to the wind speed error, as it is used to label other anomalies when the wind speed is in the feasible range. A possible reason for the multiple meaning of the same status code (here “183”) might be due to multiple errors occurring simultaneously. The status code “0” is discussed next.

2) Status code “0”

In Table 6.3 status code “0” was assigned to the out-of-range negative values of the wind speed. The same status value is assigned for the four instances in Table 4. Status code “0”, however, does not offer any useful status or fault information. Based on the data analysis, the meaningful status codes appear to be in the range of [1, 350].

Table 6.4 Fault information for status code “0”.

Date	Time	Status Code	Wind Speed	Power Output	Generator Torque
4/7/2009	3:02:43 AM	**0**	17	-3	76
4/7/2009	3:18:05 AM	**0**	16	-3	72
4/7/2009	3:18:05 AM	**0**	15	-3	47
4/7/2009	3:22:46 AM	**0**	14	0	77

3) Presence of duplicate data

Some of the data entries associated with the status code could be repeated a number of times, as illustrated in Table 6.5 for the status code “183”. The reason behind the repeated values could be in the imperfection of the SCADA software.

Table 6.5 Duplicate fault information.

Date	Time	Status Code	Wind Speed	Power Output	Generator Torque
5/24/2009	7:20:34 PM	183	5	-2	998
5/24/2009	7:20:34 PM	183	5	-2	998
5/24/2009	7:20:34 PM	183	5	-2	998
5/24/2009	7:20:34 PM	183	5	-2	998

4 Pre-processing status/fault data

As incorrect data would negatively impact the models built, all status/fault data is pre-processed for removal of the data in doubt. The number of status/fault instances after data pre-processing for each of the four turbines is shown in Table 6.6.

Table 6.6 Reduced status/fault data set.

Turbine No.	Status/Fault Instances	No of Status Codes
1	2383	65
2	2619	59
3	817	49
4	**1329**	**66**

The data set in Table 6.6 has been significantly reduced. For example, the 7000 instances of the status/fault data initially provided for Turbine 4 have led to 1329 instances covering 66 different status codes. The data collected at Turbine 4 has been selected for further analysis.

6.2.2 The Power Curve

1. Power curve based on SCADA data

The shape of the power curve determines the health of a wind turbine. A model power curve is portrayed as a sigmoid function representing the relationship between the power produced for the wind speed in the range between cut-in and cut-out speed.

A power curve built from the actual data deviates from an ideal power curve in the following: (1) some power outputs are negative; (2) there are different values of power output for identical wind speeds. The results of analysis of over 25000 instances of 5-min SCADA data collected for each of the four turbines during a three-month period are summarized in Table 6.7.

Table 6.7 Summary of SCADA data for four turbines.

Turbine	Number of Positive Power Values	Number of Negative Power Values	Number of Erroneous Data Values	Total Number of Instances
1	24892	3415	1031	29338
2	21035	3844	1030	25909
3	3504	21309	1108	25921
4	**8466**	**16359**	**1095**	**25920**

The data in Table 6.7 has been organized according to the values of the power output. Three categories of power output are considered for each wind turbine: positive values, negative values, and values in error. Positive power implies generation of electrical energy. Negative power implies that the wind turbine is consuming energy likely due to the low wind speed. The erroneous data is due to various status/fault situations. Figure 6.1 and Figure 6.2 illustrate the power curve for Turbine 4 for positive and negative power values, respectively.

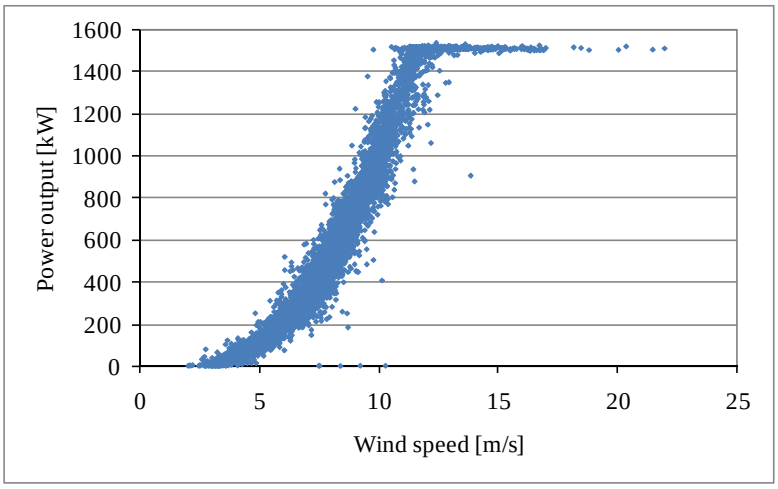

Figure 6. 1 Turbine 4 curve for positive power values.

The power curve in Figure 6.1 includes scattered points providing a basis for fitting into an ideal power curve. There are a number of reasons for the variability reflected in the power curve, including the errors caused by malfunctions of the turbine systems and components.

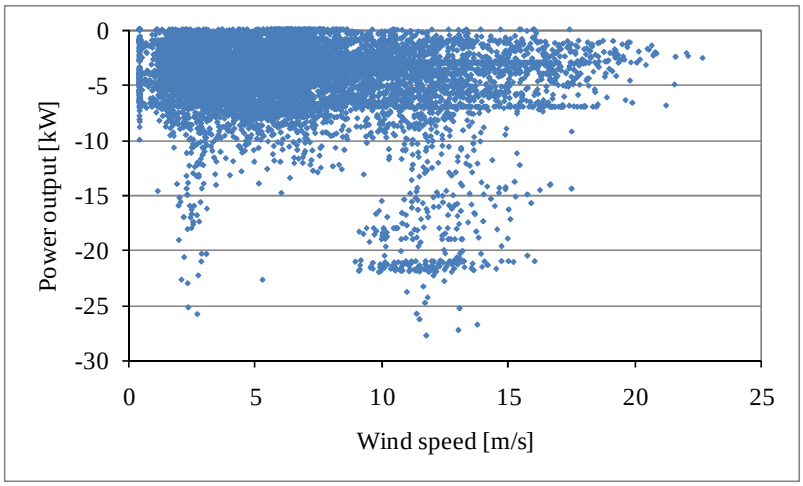

Figure 6.2 Turbine 4 curve for negative power values.

Most (97.48%) negative values of the power output are in the range of [-10kW, 0kW], and the minimum negative power is -30kW. Nearly 2/3 of the power outputs (63.11%) of Turbine 4 are negative during the period analyzed. There are two main reasons for negative power: the wind speed is lower than the cut-in speed, and there are maintenance issues with the turbine.

2. Power curve based on status/fault data

In addition to turbine operational data collected by the SCADA system, a turbine status/fault data is generated for each turbine. The status/fault data is time stamped, and therefore it can be linked with the SCADA records. The solid line in Figure 6.3 illustrates a model power curve obtained from the data used to map the power curve in Figure 6.1.

The model power curve in Figure 6.3 was built from 8466 instances representing the normal (fully functional) status of Turbine 4 by constructing 30 neural networks.

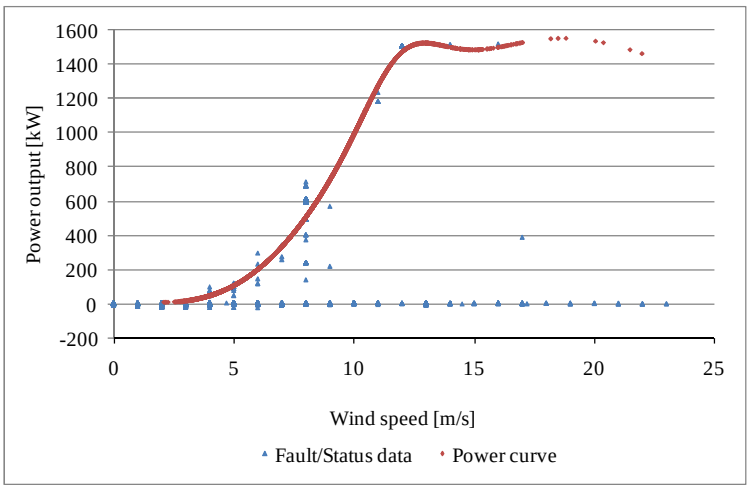

Figure 6.3 Power curve of Turbine 4 and scattered points included in the status/fault file.

The neural network with the smallest training error was selected to predict the power curve (solid line) in Figure 6.3. A similar approach to generate a power curve was used in previous research [85]. The scattered points in Figure 6.3 represent the status/faults instances collected as a separate file. As illustrated in Figure 6.3, some status/fault data points present themselves no differently than the points creating a typical power curve. The data points representing zero or negative power consumption also fall in the status/fault category.

The wind speed and power output are used in this paragraph as input variables to identify faults of a wind turbine.

3. Status Codes in the Turbine Data

1) Frequency for statuses/faults

The frequency of faults in the data set varies. For example, for Turbine 4, the status codes “180” to “184” happen hundreds of times, while the status code “1” or “5” occurs only a few times, or as rarely as once every three months. Figure 6.4 illustrates the frequency of statuses/faults for Turbine 4.

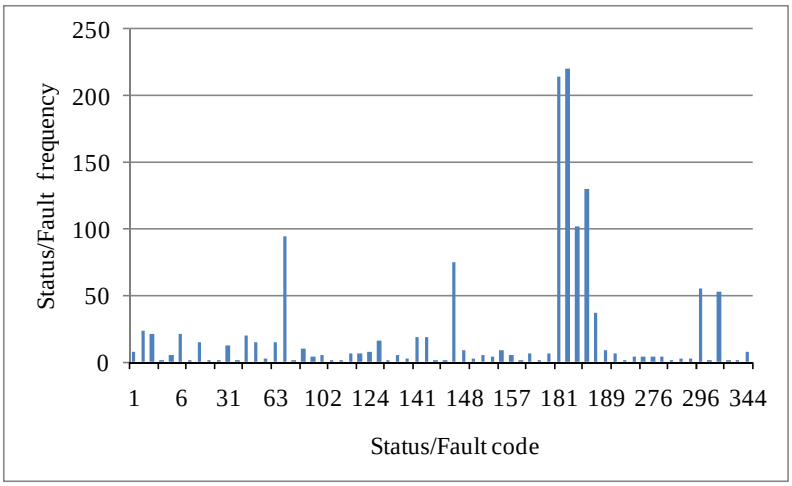

Figure 6.4 Fault frequency of Turbine 4.

The most frequent status shown in Table 6.8 is “Start-up”, which occurs 220 times in the three-month period. The other three statuses occur hundreds of times. These statuses, however, do not seriously impact the wind turbine system. The low impact status codes are of lesser interest to this research. Rather, the focus is on the severer faults.

Table 6.8 provides detailed information on five status codes, 181 to status 185. The most frequent status shown in Table 6.8 is “Start-up”, which occurs 220 times in the three-month period. The other three statuses occur hundreds of times. These statuses, however, do not seriously impact the wind turbine system. The low impact status codes are of lesser interest to this research. Rather, the focus is on the severer faults.

Table 6.8 Detailed Information from status code 181 to 185.

Status Code	Status Text	Category	Frequency
181	Idling position	4	214
182	Start-up	4	220
183	Load operation	4	102
184	Shut down	4	130
185	Manual operation of pitch	4	37

2) Fault versus status

Neither the data available in this research nor the current literature discusses the relationship between “statuses” and “faults” in wind turbines. This paragraph presents a useful approach for making such a distinction.

Each status/fault code of a wind turbine is assigned one of four categories according to its severity of impact on the wind turbine system. It is observed from the data provided that categories 1, 2 and 3 might adversely impact the wind turbine system and its components. But the status codes in Category 4 are not likely to seriously hinder the operations of a wind turbine. Statuses in categories 1, 2 and 3 are regarded as faults, and statuses in Category 4 are considered as statuses. The distribution of faults for all four turbines in each category is shown in Table 6.9.

As illustrated in Table 6.9, Category 4 statuses occur most frequently (87.33% on average for the four turbines). The most severe faults (Category 1) happen on average 1.50% of the time. The faults of Categories 2 and 3 occur more frequently than those of Category 1. The fault distribution of Turbine 4 is illustrated in Figure 6.5.

Table 6.9 Distribution of faults and statuses by category.

Turbine	Category 1		Category 2		Category 3		Category 4		Overall
	Number	Percentage	Number	Percentage	Number	Percentage	Number	Percentage	
1	40	1.68%	417	17.50%	44	1.85%	1882	78.98%	2383
2	14	0.53%	42	1.60%	36	1.37%	2527	96.49%	2619
3	11	1.35%	40	4.90%	29	3.55%	737	90.21%	817
4	**42**	**3.16%**	**131**	**9.86%**	**60**	**4.51%**	**1096**	**82.47%**	**1329**
Sum/ Average	107	1.50%	630	8.81%	169	2.36%	6242	87.33%	7148

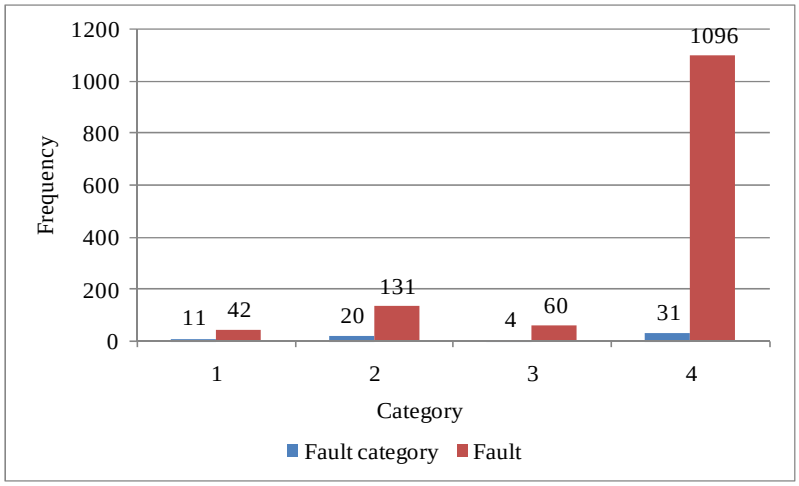

Figure 6.5 Fault distribution for Turbine 4.

There are 35 specific faults (11 in Category 1, 20 in Category 2, 4 in Category 3), and 31 different status occurrences. In total 233 (42 + 131 + 60) faults and 1096 statuses are captured during the three-month period.

3) Most frequent faults

The faults that occur relatively frequently (Fault frequency > 10) and their categories are listed in Table 6.10.

Table 6.10 Most frequent faults of Turbine 4.

No	Status Code	Status Text	Category	Fault Frequency
1	31	Timeout of yaw counter	2	12
2	45	Hydraulic pump time too high	2	20
3	52	Gearbox oil pressure too low	2	15
4	63	Safety chain	1	14
5	141	Rotor CCU collective faults	2	18
6	142	Line CCU collective faults	2	18
7	**296**	**Malfunction of diverter**	**3**	**55**

As illustrated in Table 6.10, only seven faults happen more than 10 times during the time period reflected in the data, including one fault from Category 1, five faults from Category 2, and one fault from Category 3. As the malfunction of the diverter (status code “296”) occurs most frequently, it is selected for further analysis.

6.3 Methodology for Fault Diagnosis of Wind Turbines

6.3.1 Three-level fault prediction

The proposed methodology for fault prediction of wind turbine systems involves three levels (see Figure 6.6).

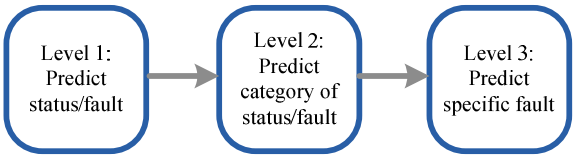

Figure 6.6 Levels for fault prediction.

Level 1: Predict status/fault

The goal of this level is to distinguish the status/fault data from the labeled SCADA data (to be discussed in Section 6.4). No differentiation is made between a status and a fault.

Level 2: Predict category of status/fault

It is not enough to recognize whether a status/fault has occurred at a certain time. At this level, the category of a status or a fault is detected.

Level 3: Predict specific fault

There are nearly 350 different status codes occurring with different frequencies for wind turbines. It is easier to detect statuses that are more frequent. At this level, fault "Malfunction of diverter" is predicted up to 60 minutes before it occurs.

For each level of fault prediction, the general process is divided into four steps: labeling SCADA data, data sampling, model extraction, and computational results analysis. The process of fault prediction is outlined in Figure 6.7 and discussed in the next sections.

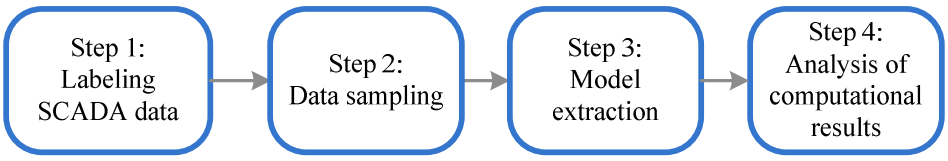

Figure 6.7 Process of fault prediction.

6.3.2 Labeling SCADA data with status/fault code and category

The data of Table 6.2 is generated whenever a status/fault occurs. The SCADA and the status/fault data is integrated by assigning status/fault codes and their categories to SCADA data according to (6.1)

$$\text{If } T_{SCADA}(t - n) < T_{fault}(t_{fault}) < T_{SCADA}(t - n + 1) \text{ Then}$$
$$\text{Status\_Code}(t - n) = \text{Status\_Code}(t_{fault})$$
$$\text{Category}(t - n) = \text{Category}(t_{fault}) \quad (6.1)$$

where Status_Code and Category are as shown in Table 6.2; n is the number of time stamps in advance of the status/fault. An attempt will be made to detect a status/fault $n \times 5$ minutes in advance. In this paragraph, n is assumed as 12; i.e., up to 60 minutes ahead of reporting the status/fault.

In the process of matching the status/fault data (see Table 6.2) with the SCADA data (unlabeled data), some status/fault data is deliberately ignored. The reason is that the status/fault data is recorded whenever the status or the fault happened, while the turbine operational data is reported at 5-min intervals. During the 5-min interval, the status code with the most severe category is considered. Table 6.11 shows a typical status code file.

Table 6.11 Ignored status code data while matching it with the SCADA data.

Date	Time	Status Code	Status Text	Category	Wind Speed	Power Output	Generator Torque
4/3/2009	11:23:27 PM	95	PC restart	4	6	-7	22
4/3/2009	11:23:27 PM	156	Repair	4	6	-7	22
4/3/2009	11:23:27 PM	292	Malfunction of cabinet heaters	3	6	-7	22
4/3/2009	11:23:27 PM	293	Malfunction of temp switch cabinet	3	6	-7	22
4/3/2009	11:23:27 PM	296	Malfunction of diverter	3	6	-7	22

As illustrated in Table 6.11, a number of status codes are reported at the same time, i.e., 11:23:27 PM. Of those, only the most severe category status code, e.g., status code “292” “Malfunction of cabinet heaters” is merged with the SCADA data. This way the 5-min record of the SCADA file corresponding to the time stamp 11:23:27 PM is assigned the Category 3 label.

After the turbine operations data have been labeled with the status/fault according to (6.1), 637 status/fault instances remain for Level 1 and Level 2 predictions. In other words, almost 50% of the status/fault infor

mation is lost. The fault “malfunction of diverter” with the status code 296 is used in this experiment. Of 55 status occurrences, 50 status/fault instances were used, and 5 instances were lost.

6.3.3 Data Sampling

An ideal training data set should be balanced with status/fault and normal operations data. The selection of the status/fault data was discussed in Section 5.2. To construct a training data set reflecting normal turbine operations, direct use of the labeled SCADA data is not acceptable, as the number of records (8466) would vastly exceed the number of instances of the status/fault data and thus cause a prediction bias. Data sampling is an effective technique to deal with this issue.

A data sample is randomly selected from the normal instances of the labeled SCADA data of Turbine 4. To create a balanced data set, the size of the data sample depends on the size of the fault data at a particular level. Thus, for Level 1 and Level 2, 650 normal instances are selected, and for Level 3, 118 instances. At each level, the fault and normal instances are combined into one file. The combined data set for Level 1 and Level 2 predictions contains 650 normal instances and 637 fault instances. For Level 3 predictions, the combined data set contains 118 normal instances and 50 fault instances.

6.3.4 Test strategy

At each level of fault prediction, a training data set is created by randomly selecting 2/3 of the instances of the combined data set for each time stamp from *t* to $t-12$. Specifically, 13 training data sets are provided for prediction from current time t to the proceeding 60 minutes, i.e., $t-12$.

Two types of test data sets are provided. The first test data set uses 1/3 of the instances of the combined data for each time stamp. Normal instances and fault instances are sampled separately to avoid unbalanced fault distribution in the training and test data sets. For example, 16 faults of Category 1 are provided in the combined data sets; 10 are used for training and the other 6 are used for testing. The second test data set is created by randomly selecting 10% of the data from the labeled SCADA data. In the first test data set, the percentage of faults versus normal instances is much higher than that in the labeled SCADA data set. The second test data represents the distribution of the labeled SCADA data. The number of instances sampled for each data set (at each of the three levels) is illustrated in Table 6.12.

Table 6.12 Training and test data sets.

Level	Training Data Set					Test Data Set 1					Test Data Set 2				
Level 1	Normal	Status/Fault				Normal	Status/Fault				Normal	Status/Fault			
	433	425				217	212				2007	61			
Level 2	Normal	C1	C2	C3	C4	Normal	C1	C2	C3	C4	Normal	C1	C2	C3	C4
	433	10	22	16	375	217	4	12	9	187	2007	1	2	2	56
Level 3	No Fault 296	Fault 296				No Fault 296	Fault 296				No Fault 296	Fault 296			
	80	35				38	15								

As illustrated in Table 6.12, the number of fault instances in the second test data set is limited. There are 61 status/fault instances for Level 1 predictions. Only 5 faults (1 for Category 1, 2 for Category 2, and 2 for Category 3) are provided among the sampled data for Level 2 predictions. For Level 3 predictions, the randomly sampled test data has a very low probability of including the fault "Malfunction of diverter".

6.4 Industrial Case Study

6.4.1 Model Extraction

The model's extraction process is illustrated in Fig. 6.8.

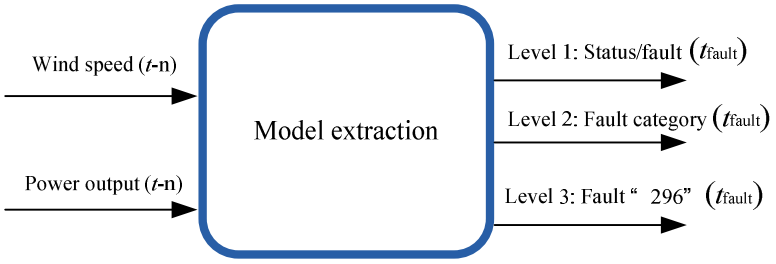

Figure 6.8 The model extraction process.

As illustrated in Fig. 6.8, input variables are wind speed ($t-n$) and power output ($t-n$). The target outputs are (1) fault-no fault at t_{fault}; (2) category of the fault at t_{fault}; and (3) fault "296" at t_{fault}. To compare the prediction results, three metrics defined in (6.2) - (6.4) are used.

$$Accuarcy = \frac{\text{Number of correctly predicted fault instances + Number of correctly predicted normal instances}}{\text{Number of fault instances + Number of normal instances}} \times 100\% \quad (6.2)$$

$$Sensitivity = \frac{\text{Number of correctly predicted fault instances}}{\text{Number of fault instances}} \times 100\% \quad (6.3)$$

$$Specification = \frac{\text{Number of correctly predicted normal instances}}{\text{Number of normal instances}} \times 100\% \quad (6.4)$$

Accuracy provides the percentage of correctly made predictions. Sensitivity expresses the percentage of correctly predicted faults, and specificity expresses the percentage of correctly predicted normal instances.

1. Model extraction at Level 1

Four data-mining algorithms have been applied to extract the models, the Neural Network (NN), the Neural Network Ensemble (NN Ensemble), the Boosting Tree Algorithm (BTA), and the Support Vector Machine (SVM). The prediction results for the test data set 1 at current time *t* are shown in Table 6.13.

Table 6.13 Performance of four algorithms predicting status/fault at time t.

Algorithm	Accuracy (%)	Sensitivity (%)	Specificity (%)
NN	74.71	81.00	68.67
NN Ensemble	**74.56**	**83.67**	**65.81**
BTA	71.27	84.66	59.50
SVM	69.64	59.97	78.92

As illustrated in Table 6.13, the NN-ensemble makes the best quality predictions, and therefore it is recommended for building Level 1 models. To construct the NN-ensemble, 30 NNs are built and the best five are selected.

2. Model extraction at Level 2

Several data-mining algorithms have been applied to extract the models, including the Neural Network (NN), the Standard Classification and Regression Tree (CART), the Boosting Tree Algorithm (BTA), and the Support Vector Machine (SVM). The prediction accuracy (%) results for test data set 1 at current time *t* are compared in Table 6.14.

Table 6.14 Performance of four algorithms for fault category predictions.

Algorithm	Prediction Accuracy for Normal	Prediction Accuracy for Category 1	Prediction Accuracy for Category 2	Prediction Accuracy for Category 3	Prediction Accuracy for Category 4
NN	76.66	0.00	0.00	12.00	74.91
BTA	41.00	22.22	83.33	0.00	72.15
CART	**96.08**	**62.50**	**52.94**	**56.00**	**95.20**
SVM	80.88	0.00	0.00	0.00	69.28

As illustrated in Table 6.14, CART exhibits the strongest potential and is selected for further predictions of fault categories.

3. Model extraction at Level 3

Several algorithms have been applied to extract the data-mining models, including the Neural Network (NN), Neural Network Ensemble (NN Ensemble), Boosting Tree

Algorithm (BTA), and Support Vector Machine (SVM). The prediction results for test data set 1 at current time t are compared in Table 6.15.

Table 6.15 Performance of four algorithms in prediction of a specific fault.

Algorithm	Accuracy (%)	Sensitivity (%)	Specification (%)
BTA	69.81	86.67	63.16
NN	72.00	66.67	70.45
NN Ensemble	68.00	82.88	66.67
SVM	70.59	47.06	82.35

As illustrated in Table 6.15, the BTA algorithm has been selected for prediction of the fault “Malfunction of diverter”. The learning rate used by this algorithm was 0.1.

6.4.2 Computational Results Analysis

1 Computational results for Level 1

1) Performance of test data set 1

In this section, the models extracted in Section 6.4.1 have been applied to test data 1 (as illustrated in Table 6.12) for Level 1 predictions. The models are extracted from current time t to time stamp t - 12 (13 prediction models). The prediction results for six models (one per time stamp) are illustrated in Table 6.16.

As illustrated in Table 6.16, the prediction accuracy is in the interval of [63%, 77%]. The sensitivity is relatively high, implying that most faults and statuses have been correctly identified. The accuracy and sensitivity at the time stamp t - 12 are lower than at other periods.

Table 6.16 Test 1 results for status/fault at six time stamps.

Time Stamp	Accuracy (%)	Sensitivity (%)	Specification (%)
t	74.56	83.67	65.81
t - 1	74.42	75.98	72.93
t - 3	75.19	85.87	65.02
t - 6	75.10	86.34	64.43
t - 9	76.03	88.38	64.40
t - 12	63.77	51.18	75.63

2) Performance of test data set 2

In this section, the models extracted in Section 6.1 have been applied to test data set 2 (as illustrated in Table 6.12) for Level 1 prediction. The prediction results obtained at time stamps are illustrated in Table 6.17.

Table 6.17 Test 2 results for status/fault prediction at six time stamps.

Time Stamp	Accuracy (%)	Sensitivity (%)	Specification (%)
t	68.63	78.69	63.88
t - 1	66.17	93.18	65.58
t - 3	65.88	83.61	65.37
t - 6	66.70	65.57	66.77
t - 9	65.39	73.77	65.37
t - 12	77.42	39.34	78.57

The results in Table 6.17 show that prediction accuracy at the time stamps t to $t-12$ is in the range of [65%, 78%]. Most statuses/faults have been correctly predicted. The percentage of correctly predicted statuses/faults is in the interval of [39%, 94%], and correctly predicted normal instances are in the range of [63%, 79%].

2. Computational results for Level 2 prediction

1. Performance of test data set 1

In this section, the models extracted in Section 6.2 have been applied to test data 1 (as illustrated in Table 6.12) for Level 2 predictions. The models are extracted from the current time t to time stamp t - 12 (13 prediction models). The prediction accuracy (%) results produced at six time stamps are illustrated in Table 6.18.

Table 6.18 Test 1 results for prediction of status/fault category at six time stamps.

Time	Accuracy	Normal instances	Category 1	Category 2	Category 3	Category 4
t	93.39	96.08	62.50	52.94	56.00	95.20
t - 1	93.39	95.79	68.75	47.06	52.00	95.91
t - 3	94.31	92.27	50.00	70.59	56.00	94.31
t - 6	92.21	94.79	56.25	67.65	68.00	92.70
t - 9	91.86	94.83	68.75	52.94	48.00	93.24
t - 12	90.72	94.24	56.25	38.24	40.00	92.88

As illustrated in Table 6.18, the prediction accuracy for normal and status instances is high. However, the percentage of correctly predicted fault instances is in the range of [40%, 71%].

2. Performance of test data set 2

In this section, the models extracted in Section 6.2 have been applied to test data 2 at Level 2. The accuracy results (%) for six time stamps are illustrated in Table 6.19.

Table 6.19 Test 2 results for prediction of status/fault category at six time stamps.

Time	Accuracy	Normal Instances	Category 1	Category 2	Category 3	Category 4
t	99.27	99.75	100.00	100.00	100.00	82.14
t - 1	99.17	98.00	100.00	50.00	100.00	84.62
t - 3	99.13	99.36	50.00		50.00	77.78
t - 6	99.08	99.90	100.00	50.00	50.00	73.21
t - 9	98.87	99.75	50		100	75.93
t - 12	98.77	99.51	100	100	50	76.79

Despite the fact that the number of status/fault categories is small, the results presented in Table 6.19 are quite impressive. The variability in accuracy seen there is due to the small number of status/fault categories. For example, if one of the two status/fault categories is predicted in error, then the accuracy decreases from 100% to 50%. The prediction accuracy for normal instances is still high. However, the accuracy for status/fault category prediction is lower compared to test data set 1.

3 Computational results for Level 3 predictions

In this section, the models extracted in Section 6.2 have been applied to test data set 1 (as illustrated in Table 6.12) for Level 3 predictions. The models are extracted from current time t to time stamp t - 12 (13 prediction models). The prediction results for six time stamps are shown in Table 6.20.

Table 6.20 Test 1 results for prediction of a specific fault at six time stamps.

Time Stamp	Accuracy (%)	Sensitivity (%)	Specification (%)
t	69.81	86.67	63.16
t - 1	64.15	66.67	63.16
t - 3	67.92	73.33	65.79
t - 6	67.92	73.33	65.79
t - 9	66.04	33.33	78.95
t - 12	49.06	24.53	34.21

As illustrated in Table 6.20, the prediction accuracy of the fault "Malfunction of diverter" is in the interval of [49%, 70%]. The percentage of correctly predicted faults is in the interval of [24%, 87%], and the correctly predicted instances without the fault "Malfunction of diverter" is in the interval of [34%, 79%].

6.5 Summary

A methodology to predict turbine faults using information provided by SCADA systems and fault files was presented. The methodology involves three steps: (1) the existence of a status/fault was identified; (2) the category (severity) of the fault was predicted; and (3) a specific fault was predicted. The computational results reported in the paragraph demonstrated that, in most cases, faults can be predicted with a reasonable accuracy 60 minutes before they occur. The prediction accuracy of the fault category is somewhat lower yet acceptable. Due to the data limitations, identifying a specific fault, though valuable, decreases accuracy.

The research reported in this paragraph was performed with industrial data collected at operating wind turbines. The major difficulty was in the low frequency of the data. The description of faults was not clear, and the number of fault occurrences was far

from sufficient. A better prediction performance would have been achieved with higher quality data.

The limitations surrounding this research are as follows:

1) The volume of fault data was limited, and therefore many faults did not appear in the data or occurred only sporadically. Such rare faults are difficult to detect by any modeling approach.

2) The 5-min interval for collecting the vast majority of data was too long. Such a long interval led to a significant loss of the history of the fault emergence.

3) In this paragraph, every status code was considered independently. The relationship between faults has not been considered largely due to the low frequency data

CHAPTER 7.

CONCLUSION

This thesis proposes a framework of predictive models under data mining technique. Chapter 1 provided a review of predictive models in wind energy with emphasis on short-term wind speed forecasting, wind power generation, optimization and condition monitoring and diagnosis. Chapter 2 introduced a methodology for short-term wind speed prediction based on wind farm layout information. Wind speeds collected from neighborhood wind turbines were used as predictors. Chapter 3 presented models for short-term prediction of wind turbine parameters, including wind power and rotor speed. A clustering-based method for power generation was proposed in Chapter 4. Chapter 5 introduced an intelligent wind turbine system and dynamic control strategies for optimization of power generation and rotor ramp rates. Fault diagnosis and prediction using SCADA data was explored in Chapter 6.

There are still other interesting research questions that should be answered in the future. Future research can be focused on specific fault detection, condition monitoring and adjustable dynamic control of the wind turbine based on SCADA data.

REFERENCES

[1] M. Monfared, H. Rastegar and H.M. Kojabadi, "A new strategy for wind speed forecasting using artificial intelligent methods," *Renewable Energy*, Vol. 34, No.7, pp. 845-848, 2009.

[2] M.C. Mabel and E. Fernandez, "Analysis of wind power generation and prediction using ANN: A case study," *Renewable Energy*, Vol. 33, No. 5, pp. 986-992, 2008.

[3] http://www.awea.org, Accessed 2nd February, 2009.

[4] C.A. Walford, Wind turbine reliability: Understanding and minimizing wind turbine operation and maintenance costs. Sandia National Laboratoires, Albuquerque, N.M., 2006, Available: www.prod.sandia.gov/cgi-bin/techlib/access-control.pl/2006/061100.pdf.

[5] R. Wiser and M. Bolinger, Annual Report on U.S. Wind Power Installation, Cost, and Performance Trends: 2006. NREL, US Department of Energy, Golden, CO, 2007. Available: http://www.nrel.gov/wind/pdfs/41435.pdf .

[6] F. Bianchi, H. Battista, and R. Mantz, "Wind Turbine Control System: Principles, Modeling and Gain Scheduling Design" , *Springer*, 2006, pp.8-28

[7] N. Nanayakkara, M. Nakamura and H. Hatazaki, "Predictive control of wind turbines in small power systems at high turbulent wind speeds," *Control Engineering Practice*, Vol. 5, No. 8, pp. 1063-1069, 1997.

[8] M.Monfared, S. Rehman and T. Halawani, "A neural networks approach for wind speed prediction," *Renewable Energy*, Vol. 13, No. 3, pp. 345-354, 1998.

[9] G. Riahy and M. Abedi, "Short term wind speed forecasting for wind turbine applications using linear prediction method," *Renewable Energy*, Vol. 33, No. 1, pp. 35-41, 2008.

[10] E. Bossanyi, "Short-term wind prediction using Kalman filters," *Wind Engineering*, Vol. 9, No. 1 pp. 1-8, 1985.

[11] T. Barbounis, and J. Theocharis, "Locally recurrent neural networks for long-term wind speed and power prediction," *Neurocomputing*, Vol. 69, No. 4-6, pp. 466-496, 2006

[12] S. Watson, L. Landberg and J. Halliday, "Application of wind speed forecasting to the integration of wind energy into a large scale power system," *IEE Proceedings: Generation, Transmission and Distribution*, Vol. 141, No. 4, pp. 357-362, 1994.

[13] A. Kusiak, H. Zheng and Z. Song, "Short-term prediction of wind farm power: A data-mining approach, *IEEE Transactions on Energy Conversion*, Vol. 24, No. 1, pp. 125-136, 2009.

[14] I. Damousis and P. Dokopoulos, "A fuzzy expert system for the forecasting of wind speed and power generation in wind farms," *In: 22nd IEEE Power*

Engineering Society International Conference on Power Industry Computer Applications, pp. 63-69, May 20–24, 2001.

[15] I. Damousis, M. Alexiadis, J. Theochairs and P. 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. 353-361, 2004.

[16] S. Sancho, P. Angel M., O. Emilio G., P. Antonio, P. Luis and C. Francisco, "Accurate short-term wind speed prediction by exploiting diversity in input data using banks of artificial neural networks," *Neurocomputing*, Vol. 72, No. 4-6, pp. 1336-1341, 2009.

[17] P. Flores, A. Tapia and G. Tapia, "Application of a control algorithm for wind speed prediction and active power generation," *Renewable Energy*, Vol. 30, No. 4, pp. 523-536, 2005.

[18] B. Mehmet, S. Besir and Y. Abdukadri, "Application of artificial neural networks for the wind speed prediction of target station using reference stations data," *Renewable Energy*, Vol. 32, No. 14, pp. 2350-2360, 2007.

[19] M. Mohandes, T. Halawani, S. Rehman, A. Hussain, "Support vector machines for wind speed prediction," *Renewable Energy*, Vol. 29, No. 6, pp. 939-947, 2004.

[20] A. Ahmed and D. Lee, "SVR-based wind speed estimation for power control of wind energy generation system," *Fourth Power Conversion Conference-NAGOYA, PCC-NAGOYA 2007 - Conference Proceedings*, pp. 1431-1436, 2007.

[21] H. Tarek, F. Ehab and M. Magdy, "One day ahead prediction of wind speed and direction," *IEEE Transactions on Energy Conversion*, Vol. 23, No. 1, pp. 191-201, 2008.

[22] G.V. Kuik, B. Ummels and R. Hendriks, “Sustainable Energy Technologies,” *Springer*: Amsterdam, The Netherlands, 2007.

[23] A. Kusiak, H. Zheng and Z. Song, “Wind Farm Power Prediction: A Data-Mining Approach”, *Wind Energy*, Vol. 12, No. 3, pp. 275-293, 2009.

[24] L. Ma, S.Luan, C.Liang, H.Liu and Y. Zhang, “A review on the forecasting of wind speed and generated power”, *Renewable and Sustainable Energy Reviews*, Vol. 13, No. 4, pp. 915-920, 2009.

[25] G. Kariniotakis, P.Pinson, N.Siebert, G.Giebel and R.Barthelmie “The State of the art in short-term prediction of wind power-from an offshore perspective”, Anemos Project Report D1.1 (Available online: http://anemos.cma.fr), 2003.

[26] C. Alexandre, C. Antonio, N. Jorge, L. Gil, M. Henrik and F. Everaldo, “A review on the young history of the wind power short-term prediction,” Renewable and Sustainable Energy Reviews, Vol. 12, No. 6, pp. 1725-1744, 2008.

[27] L. Landberg, “Short-term prediction of the power production from wind farms,” Journal of Wind Engineering and Industrial Aerodynamics, Vol. 80, No.1-2, pp. 207-220, 1999.

[28] 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.

[29] M. Negnevitsky and C.W. Potter, “Innovative short-term wind generation prediction techniques”, Proceedings of the Power Systems Conference, pp. 60-65, 2006.

[30] U. Focken, M. Lange, K. Monnich, 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 and Industrial Aerodynamics, Vol. 90, No. 3, pp. 231-246, 2002.

[31] S. Haykin,” Neural Networks: A Comprehensive Foundation”. Macmillan Publishing: New York, 1994.

[32] S. Kelouwani and K. Agbossou, "Nonlinear model identification of wind turbine with a neural network," IEEE Transactions on Energy Conversion, Vol. 19, No. 3, pp. 607-612, 2004.

[33] Y. Xiao, W. Wang and X. Huo. “Study on the time-series wind speed forecasting of the wind farm based on neural networks”, Energy Conservation Technology, Vol.25, No. 2, 2007, pp.106–109.

[34] S. Li. “Wind power prediction using recurrent multilayer perceptron neural networks”, Power Engineering Society General Meeting, Vol. 4, 2003. pp. 2325–2330.

[35] S. Lou, Z. Li and Y. Wu, “Clustering analysis of the wind power output based on similarity theory,” 3rd International Conference on Deregulation and Restructuring and Power Technologies, DRPT 2008, pp. 2815-2819, 2008

[36] U. Taner and A. Ahmet, “Wind turbine power curve estimation based on cluster center fuzzy logic modeling”, Journal of Wind Engineering and Industrial Aerodynamics, Vol. 96, No. 5 pp. 611-620, 2008.

[37] B. Boukhezzar, H. Siguerdidjane and M. Maureen Hand, "Nonlinear control of variable-speed wind turbines for generator torque limiting and power optimization," *ASME Transactions: Journal of Solar Energy Engineering*, Vol. 128, No. 4, pp. 516-530, 2006.

[38] R. Datta and V. T. Ranganathan, "A method of tracking the peak power points for a variable speed wind energy conversion system," *IEEE Transactions on Energy Conversion*, Vol. 1, No.1, pp. 163-168, 2003.

[39] S. Morimoto, H. Nakayama, M. Sanada and Y. Takeda, "Sensorless output maximization control for variable-speed wind generation system using IPMSG," *IEEE Transactions on Industry Applications*, Vol. 41, No.1, pp. 60-67, 2005.

[40] E. Muljadi and C. P. Butterfield, "Pitch-controlled variable-speed wind turbine generation," *IEEE Transactions on Industry Applications*, Vol. 37, No.1, pp. 240-246, 2001.

[41] I. Munteanu, N. A. Cutululis, A. I. Bratcu and E. Ceanga, "Optimization of variable speed wind power systems based on a LQG approach," *Control Engineering Practice*, Vol. 13, pp. 903-912, 2005.

[42] K.E. Johnson, L.Y. Pao, M.J. Balas, and L.J. Fingersh, "Control of variable-speed wind turbines: standard and adaptive techniques for maximizing energy capture," *IEEE Control Systems Magazine*, Vol. 26, No. 3, pp. 70-81, 2006.

[43] L. C. Henriksen, "Model predictive control of a wind turbine," 2007.

[44] E. B. Muhando, T. Senjyu, N. Urasaki, A. Yona and T. Funabashi, "Robust predictive control of variable-speed wind turbine generator by self-tuning regulator," in *IEEE Power Engineering Society General Meeting*, 2007, pp. 1-8.

[45] E. F. Camacho and C. Bordons, *Model Predictive Control.* London, UK: Springer, 1999.

[46] J. A. Rossiter, *Model-Based Predictive Control: A Practical Approach.* New York: CRC Press, 2003.

[47] R. Hyers, J. McGowan, K. Sullivan, J. Manwell, and B. Syrett, "Condition monitoring and prognosis of utility scale wind turbines," *Energy Materials*, Vol. 1, No. 3, 2006, pp. 187-203.

[48] Z. Hameed, Y. Hong, Y. Cho, S. Ahn, and C. K. Song, "Condition monitoring and fault detection of wind turbines and related algorithms: a review," *Renewable and Sustainable Energy Reviews*, Vol. 13, No. 1, 2009, pp. 1-39.

[49] Y. Amirat, M. Benbouzid, B. Bensaker, and R. Wamkeue, "Condition monitoring and fault diagnosis in wind energy conversion systems: a review", *in Proc. 2007 IEEE International Electric Machines and Drives Conference*, Vol. 2, May 2007, pp. 1434-1439.

[50] P. Tavner, G. W. Bussel, and F. Spinato, "Machine and converter reliabilities in wind turbines," *in Proc. 3rd IET International Conference on Power Electronics, Machines and Drives*, 2006, pp. 127-130.

[51] M. Wilkinson, F. Spinato, and P. Tavner, "Condition monitoring of generators and other subassemblies in wind turbine drive trains", *in Proc. 2007 IEEE International Symposium on Diagnostics for Electric Machines, Power Electronics and Drives*, Sep. 2007, pp. 388-392.

[52] Y. Amirat, M. Benbouzid, E. Al-Ahmar, B. Bensaker and S. Turri, "A brief status on condition monitoring and fault diagnosis in wind energy conversation systems", *Renewable and Sustainable Energy Reviews*, Vol. 13, No. 9, 2009, pp. 2629-2636.

[53] B. Lu, Y. Li, X. Wu and Z. Yang, "A review of recent advances in wind turbine condition monitoring and fault diagnosis", *in Proc. IEEE Conference* on *Power Electronics and Machines in Wind Applications*, 2009, pp. 1-7.

[54] E. Becker and P. Posta, “Keeping the blades turning: condition monitoring of wind turbine gears”, *Refocus*, Vol. 7, No. 2, 2006, pp. 26-32.

[55] Editorial, “Managing the wind: reducing kilowatt-hour costs with condition monitoring”, *Refocus*, Vol. 6, No.3, 2005, pp. 48-51.

[56] P. Caselitz and J. Giebhardt, ”Rotor condition monitoring for improved operational safety of offshore wind energy converters”, *Trans. ASME, J. Sol. Energy Eng.*, Vol. 127, No. 2, 2005, pp. 253-261.

[57] V. Leany, D. Sharpe and D. Infield, “Condition monitoring techniques for optimization of wind farm performance”, *Int. J. COMADEM*, Vol. 2, No. 1, 1992, pp. 5-13.

[58] M. Sanz-Bobi, M. Garcia, P. Del, “SIMAP: intelligent system for predictive maintenance application to the health condition monitoring of a wind turbine gearbox”, *Comput. Ind.*, Vol. 57, No. 6, 2006, pp. 552-568.

[59] L. Rodriguez, E. Garcia, F. Morant, A. Correcher and E. Quiles, “Application of latent nestling method using colored Petri nets for the fault diagnosis in the wind turbine subsets,” *in Proc. 2008 IEEE Int. Conf. Emerging Technologies and Factory Automation*, pp. 767-773.

[60] E. Echavarria, T. Tomiyama, and G. van Bussel, “Fault diagnosis approach based on a model-based reasoner and a functional designer for a wind turbine: an approach towards self-maintenance,” *Journal of Physics Conference Series*, vol. 75, 2007, 012078.

[61] E. Echavarria, T. Tomiyama, H. Huberts and G. van Bussel, “Fault diagnosis system for an offshore wind turbine using qualitative physics,” *in Proc. EWEC 2008*, Brussels, Belgium, 2008.

[62] A. Zaher and S. McArthur, “A multi-agent fault detection system for wind turbine defect recognition and diagnosis,” *in Proc.2007 IEEE Lausanne POWERTECH*, pp. 22-27.

[63] M. Whelan, K. Janoyan and Q. Tong, “Integrated monitoring of wind plant systems,” *Proc. SPIE Smart Sensor Phenomena, Technology, Networks, and Systems*, Vol. 6933, 2008, pp. 69330F.

[64] P. Ahlgren, B. Jarneving and R. Rousseau, “Requirements for a cocitation similarity measure, with special reference to Pearson's correlation coefficient”, *Journal of the American Society for Information Science and Technology*, Vol. 54, No. 6, , pp. 550-560, 2003.

[65] T. Ustuntas and A.D. Sahin, "Wind turbine power curve estimation based on cluster center fuzzy logic modeling," *Journal of Wind Engineering and Industrial Aerodynamics*, Vol. 96, No. 5 pp. 611-621, 2008.

[66] J. H. Friedman, "Stochastic gradient boosting," *Computational Statistics & Data Analysis*, Vol. 38, No. 4, pp. 367-378, 2002.

[67] J. H. Friedman, "Greedy function approximation: A gradient boosting machine," *Annals of Statistics*, Vol. 29, No. 5, pp. 1189-1232, 2001.

[68] M.L.Hambaba, "Intelligent hybrid system for data mining," *Proceedings of the IEEE/IAFE 1996 Conference on Computational Intelligence for Financial Engineering*, pp. 111, March 1996.

[69] S. Piramuthu, "Evaluating feature selection methods for learning in data mining applications," *Proceedings of the Thirty-First Hawaii International Conference on System Science*, Kohala Coast, HI, Vol. 5, pp. 294-302, 1998.

[70] P.N. Tan, M. Steinbach, V. Kumar, "Introduction to Data Mining," Addison Wesley, 2006.

[71] J. Hua, W.D. Tembe, and E. R. Dougherty ,"Performance of feature-selection methods in the classification of high-dimension data" *Pattern Recognition,* Vol. 42, No. 3, pp. 409-424, 2009.

[72] C.Tsai, "Feature selection in bankruptcy prediction," *Knowledge-Based Systems, In Press,* Available online August 14, 2008.

[73] C. Bishop, *Neural Networks for Pattern Recognition.* Oxford: University Press, 1995.

[74] A. E. Eiben and J. E. Smith, *Introduction to Evolutionary Computation.* New York: Springer, 2003.

[75] B. Ernst, B. Oakleaf, M. L. Ahlstrom, M. Lange, C. Moehrlen, B. Lange, U. Focken and K. Rohrig, "Predicting the wind," *IEEE Power & Energy Magazine,* Vol. 5, pp. 78-89, 2007.

[76] J. Espinosa, J. Vandewalle and V. Wertz, *Fuzzy Logic, Identification and Predictive Control.* London, UK: Springer, 2005.

[77] J. F. Manwell, J. G. McGowan and A. L. Rogers, *Wind Energy Explained: Theory, Design and Application.* 1st Ed., London, UK: John Wiley, 2002.

[78] Y. D. Song, B. Dhinakaran and X. Y. Bao, "Variable speed control of wind turbines using nonlinear and adaptive algorithms," *Journal of Wind Engineering and Industrial Aerodynamics,* Vol. 85, No.3, pp. 293-308, 2000.

[79] I. H. Witten and E. Frank, *Data Mining: Practical Machine Learning Tools and Techniques,* 2nd Ed. San Francisco: Morgan Kaufmann, 2005.

[80] Y. C. Zhu, *Multivariable System Identification for Process Control.* New York: Pergamon Press, 2001.

[81] R. Kohavi and G.H. John, "Wrapper for feature subset selection," *Artificial Intelligence,* Vol. 97, Nos 1-2, pp. 273-324, 1997.

[82] D. McMillan and G.Ault, "Condition monitoring benefit for onshore wind turbines sensitivity to operational parameters", *Renewable Power Generation,* Vol. 2, No. 1, 2009, pp. 60-72.

[83] "20% wind energy by 2030: increasing wind energy's contribution to U.S. electricity supply," *United States Department of Energy*, Report No. DOE/GO-102008-2567, July 2008.

[84] "Strategic research agenda: market deployment strategy from 2008 to 2030," *European Wind Energy Technology Platform*, July 2008. Available online: http://www.windplatform.eu/fileadmin/ewetp_docs/Bibliography/SRA_MDS_July_2008.pdf

[85] A. Kusiak, H.-Y. Zheng, and Z. Song, "On-line Monitoring of Power Curves", *Renewable Energy*, Vol. 34, No. 6, 2009, pp. 1487-1493.