

Copyright  
by  
Duehee Lee  
2015

The Dissertation Committee for Duehee Lee  
certifies that this is the approved version of the following dissertation:

**Wind Power Forecasting and Its Applications to the  
Power System**

Committee:

---

Ross Baldick, Supervisor

---

Surya Santoso

---

Aristotle Arapostathis

---

Michael Webber

---

David Morton

**Wind Power Forecasting and Its Applications to the  
Power System**

by

**Duehee Lee, B.S.; M.S.E.**

**DISSERTATION**

Presented to the Faculty of the Graduate School of  
The University of Texas at Austin  
in Partial Fulfillment  
of the Requirements  
for the Degree of

**DOCTOR OF PHILOSOPHY**

THE UNIVERSITY OF TEXAS AT AUSTIN

May 2015

Dedicated to my wife Jain Hong, my son James Lee, and my daughter  
Jacqueline Lee.

## Acknowledgments

First, I would like to express my deepest gratitude to my supervisor, Dr. Ross Baldick for his support, encouragement, and suggestions. Under his guidance, I have gained the courage to tackle any research topic. When I had no place to go, he led me toward studying the power system. He has taught me the power system protocol, how to behave in the academic field, and how to endure hard times.

I would like to thank my wife for enduring poverty, parenting, and uncertainty with me. I promise that I will be her strong supporter when she passes the same thorny path. My gratitude also goes to my other family members: my father, sister, grandfather, mother-in-law, and father-in-law.

I am also indebted to my dissertation committee members, Dr. Aristotle Arapostathis, Dr. David Morton, Dr. Michael Webber, and Dr. Surya Santoso. In particular, I would also like to thank my former supervisor Surya Santoso for his invaluable comments, support, and help in studying wind power. Furthermore, without the support of Carey King, Mack Grady, and Melanie Gulick, this dissertation would not have been written. I should not forget my English teacher, Meg Downing. She has been a friend with whom to talk about life in Austin, an editor, and a colleague with whom to discuss various research topics. I hope that she and her family are healthy, happy, and prosperous.

I wish to thank to all my Korean friends in our energy system, Joonhyun Kim, Youngsung Kyun, Myungchin Kim, Cheolhee Cho, Hunyoung Shin, Sungwoo Bae, Seunghyun Chun, Jin hur, Wonjin Cho, Seounghoon Jung, Byungchul Lee, Myungkwan Kim, Joohyun Jin, Han Kang, Kyungwoo Min, Wanki Cho, and Kwangmin Choi. I also owe my deepest gratitude to my older brothers and sisters at UT, Jinho Lee, Youngsuck Yoo, Peter Son, Mijung Park, Woori Kim, Mooryong Choi, Il Memming Park, Manho Jung, Hur Kyun, Yeajoon Kim, Youngchoon Kim, and Larkkwon Choi. I also wish to thank all my alumni friends, Byungchul Lee, Insoo Hwang, Jaewon Kim, Solkun Jee, Insun Cho, Jaehyun Bae, Jaewook Lee, Sungmin Oak, and Hoo Kim.

Friends in my research group are my comrades. We have dreamed the same dream, studied the same topics, and corresponded a lot. I will never forget Bowen Wa, Hector Chavez, Yezhou Wang, Yen-yu Lee, David Tuttle, Thuy Huynh, Deepjyoti Deka, Sambuddha Chakrabarti, Tong Jang, Juan Andrade, Mohammad Majidi, Mansoureh Peydayesh, and Michael Legatt.

Importantly, I would like to express my gratitude to other friends, Kijung Yoon, Juhun Lee, Gerad Garrison, Mohit Singh, Alicia Allen, Min Lwin, Swagata Das, Anamika Dubey, Pisitpol Chirapongsananurak, David Orn, Twoone Ngo, Quan, Suma Jothibas, Jules Campbell, Marcel Nassar, Kyungjin Kim, Neeraj Karnik, Yaidy Viswanadan, Hose Caso, Fernando Ochoa, Kai Roach, and Michael Pinnel. Finally, I want to deliver my gratitude to all my friends who know me but who are not mentioned on this page. Although they are not mentioned here, we have been and will be together.

# Wind Power Forecasting and Its Applications to the Power System

Publication No. \_\_\_\_\_

Duehee Lee, Ph.D.

The University of Texas at Austin, 2015

Supervisor: Ross Baldick

The goal of research in this dissertation is to bring more wind resources into the power grid by mitigating the uncertainty of the current wind power, by developing a new algorithm to respond to the fluctuation of the future wind power, and by building additional transmission lines to bring more wind resources from a remote area to the load center. First, in order to overcome the wind power uncertainty, the probabilistic and ensemble wind power forecasting is proposed to increase the forecasting accuracy and to deliver the probability density function of the uncertainty. Accurate wind power forecasting reduces the amounts and cost of ancillary services (AS). As the mismatch between the bid and actual amount of delivered energy decreases, the imbalance between supply and demand also decreases. If the forecasting ahead is increased up to 24 hours, accurate wind power forecasting can also help wind farm owners bid the exact amount of wind power in the day ahead (DA) market. Further-

more, wind power owners can use the parametric probabilistic density of error distributions for hedging the price risk and building a better offer curve.

Second, a novel algorithm to generate many wind power scenarios as a function of installed capacity of wind power is proposed based on an analysis of the power spectral density of wind power. Scenarios can be used to simulate the power system to estimate the required amount of AS to respond to the fluctuation of future wind power as the installed capacity of wind power increases. Scenarios have statistical characteristics of the future wind power that are regressed as a function of the installed capacity of wind power from the statistical characteristics of the current wind power. This algorithm can generate many possible scenarios to simulate the power system in many different situations.

Third, optimal transmission expansion by simulating the power system with the multiple load and wind power scenarios in different locations is planned to prepare the preliminary result to bring more wind resources in remote areas to the load center in Texas. In this process, the geographical smoothing effects of wind power and the stochastic correlation structure between the load and wind power are considered. Furthermore, the generalized dynamic factor model (GDFM) is used to synthesize load and wind power scenarios to keep their correlation structure. The premise of the GDFM is that a few factors can drive the correlated movements of load and wind power simultaneously, so the scenario generation process is parsimonious.



# Table of Contents

<b>Acknowledgments</b>	<b>v</b>
<b>Abstract</b>	<b>vii</b>
<b>List of Tables</b>	<b>xiii</b>
<b>List of Figures</b>	<b>xiv</b>
<b>Chapter 1. Introduction</b>	<b>1</b>
1.1 Wind Power in the Electricity Market . . . . .	2
1.1.1 Benefits of Wind Power . . . . .	2
1.1.2 Economic Incentives for Wind Power . . . . .	3
1.1.3 Wind Power As an Electricity Market Participant . . . . .	4
1.1.4 Effect on the Power System and Electricity Market . . . . .	6
1.2 Solutions to Increase the Wind Power Capacity . . . . .	10
1.2.1 Advanced Ancillary Services Procurement Process . . . . .	10
1.2.2 Storage System . . . . .	11
1.2.3 Electric Vehicles . . . . .	12
1.2.4 Wind Power Forecasting . . . . .	13
1.2.5 Transmission Expansion . . . . .	14
1.3 Motivation and Goals . . . . .	15
1.4 Organization . . . . .	21
<b>Chapter 2. Short-Term Wind Power Forecasting</b>	<b>22</b>
2.1 Introduction . . . . .	24
2.1.1 Literature Review . . . . .	25
2.1.2 Global Energy Forecasting Competition . . . . .	27
2.2 Program Architecture . . . . .	28
2.2.1 External Forecasting Process . . . . .	28

2.2.2	Internal Forecasting . . . . .	29
2.2.3	First Stage . . . . .	31
2.2.4	Second Stage . . . . .	34
2.2.5	Third Stage . . . . .	34
2.3	Feature Engineering . . . . .	35
2.3.1	Data Smoothing . . . . .	36
2.3.2	Data Transformation . . . . .	37
2.3.3	Data Expansion . . . . .	37
2.3.4	Outlier Detection . . . . .	38
2.3.5	Input Data Classification . . . . .	38
2.3.6	Pre-processing and Post-processing . . . . .	39
2.3.7	Simulation Settings for Individual Forecasting Machines	40
2.4	Forecasting Machines . . . . .	40
2.4.1	Ridge Regressions . . . . .	41
2.4.2	Neural Network . . . . .	42
2.4.3	Support Vector Machines . . . . .	45
2.4.4	Gaussian Process . . . . .	46
2.4.5	Bootstrap Aggregation (BAG) . . . . .	48
2.4.6	Random Forest . . . . .	49
2.4.7	Gradient Boosting Machines . . . . .	51
2.5	Ensemble Forecasting . . . . .	54
2.5.1	Analysis of Individual Forecasting Machines . . . . .	54
2.5.2	Ensemble Algorithms . . . . .	56
2.6	Quantile Estimation . . . . .	57
2.6.1	Pinball Loss Function . . . . .	58
2.6.2	Probabilistic Forecasting in the Ensemble Forecasting .	59
2.6.3	Parametric Approach . . . . .	59
2.6.4	Distributions . . . . .	61
2.6.5	Estimation of the Standard Deviations: Discrete Cluster Case . . . . .	64
2.6.6	Point Forecast Clustering . . . . .	65
2.6.7	Truncating Distribution . . . . .	67

2.6.8	Estimation of the Standard Deviations: Continuous Case	67
2.6.9	Simulation Results . . . . .	69
2.7	Conclusions . . . . .	71
<b>Chapter 3.</b>	<b>Total Future Wind Power Scenario Generation</b>	<b>74</b>
3.1	Introduction . . . . .	75
3.1.1	Literature Review . . . . .	76
3.1.2	Preprocessing . . . . .	79
3.2	Power Spectral Density Estimation . . . . .	80
3.2.1	Periodogram . . . . .	81
3.2.2	Multitaper Algorithm . . . . .	83
3.3	Piecewise Modeling of the PSD . . . . .	86
3.3.1	Hinges Model . . . . .	86
3.3.2	Original Hinges Model . . . . .	87
3.3.3	Modified Hinges Model . . . . .	91
3.4	Training Data Generation . . . . .	92
3.4.1	Factor Analysis . . . . .	94
3.4.2	Cluster Analysis . . . . .	97
3.5	Wind Power Ramp Modeling . . . . .	97
3.6	Analysis of Slopes . . . . .	99
3.6.1	Wind Power Fluctuation . . . . .	99
3.6.2	Slope Change Analysis . . . . .	103
3.6.3	Regression . . . . .	105
3.7	Scenario Synthesis . . . . .	109
3.7.1	PSD Forecasting . . . . .	111
3.7.2	Phase Angle Generation . . . . .	112
3.7.3	Genetic Algorithm . . . . .	113
3.7.4	Phase Angle Analysis . . . . .	115
3.8	Validation . . . . .	115
3.9	Ancillary Service Estimation . . . . .	123
3.10	Conclusion . . . . .	125
3.10.1	Summary . . . . .	125

<b>Chapter 4. Scenario Generation through GDFM and Transmission Expansion Planning</b>	<b>127</b>
4.1 Literature Review . . . . .	128
4.2 Additional Motivations and Goals . . . . .	132
4.3 Preprocessing and the GDFM . . . . .	133
4.3.1 Preprocessing Wind Power & Load Data . . . . .	134
4.3.2 Introduction to the GDFM . . . . .	134
4.3.3 Decomposition . . . . .	137
4.3.4 Estimation . . . . .	139
4.4 Load and wind power Scenario Generation . . . . .	141
4.4.1 Number of Dynamic Factors . . . . .	142
4.4.2 Scenario Generation . . . . .	142
4.4.3 Statistics & PSD Analysis . . . . .	145
4.4.4 Correlation Coefficient . . . . .	147
4.5 Generation & Transmission Upgrading Costs . . . . .	152
4.5.1 Simulation Settings . . . . .	153
4.5.2 Simulation Results . . . . .	154
4.6 Conclusion . . . . .	156
<b>Chapter 5. Conclusion</b>	<b>157</b>
5.1 Short-Term Wind Power Forecasting . . . . .	157
5.1.1 Key Results . . . . .	158
5.1.2 Future Work . . . . .	159
5.2 Long-Term Wind Power Scenario Generation . . . . .	159
5.2.1 Key Results . . . . .	160
5.2.2 Future Work . . . . .	161
5.3 Load and Wind Power Scenario Generation and Transmission Expansion Planning . . . . .	162
5.3.1 Key Results . . . . .	163
5.3.2 Future Work . . . . .	164
<b>Bibliography</b>	<b>165</b>
<b>Vita</b>	<b>195</b>

## List of Tables

1.1	Wind Power in Electricity Markets . . . . .	7
2.1	Performance of forecasting machines. . . . .	56
2.2	Performance of the quantile estimation methods . . . . .	70
3.1	Frequency-axis hinge locations . . . . .	92
3.2	Statistical characteristics of actual and synthesized wind power	124
4.1	Evaluation of Statistical Characteristics . . . . .	150
4.2	Results of Economic Dispatch and Transmission Expansion . .	156

## List of Figures

2.1	Forecasting unit . . . . .	24
2.2	Three-stage program architecture . . . . .	30
2.3	Program architecture of five for-loops . . . . .	32
2.4	Forecasting process of individual step . . . . .	33
2.5	Correlation coefficients of transformed features . . . . .	36
2.6	Example of wind power forecasting . . . . .	43
2.7	The Gaussian distribution and its quantile function . . . . .	60
2.8	The variance gamma distribution . . . . .	63
2.9	Quantile errors with respect to the STDs for different groups . . . . .	66
2.10	Data clustering . . . . .	68
2.11	A truncated VG distribution . . . . .	69
3.1	The PSD of wind power and its six slopes . . . . .	88
3.2	Wind power and its smoothed wind power . . . . .	98
3.3	The variability of un-normalized wind power . . . . .	100
3.4	Initial and final values of PSD . . . . .	101
3.5	Slope changes according to the number of wind farms . . . . .	103
3.6	Histogram of the power difference . . . . .	104
3.7	Slopes of first and second segments . . . . .	106
3.8	Slopes of third and fourth segments . . . . .	107
3.9	Slopes of fifth and sixth segments . . . . .	110
3.10	The means of the STDs of phase angles per frequency. . . . .	116
3.11	The actual wind power scenario sampled in April 2010. . . . .	117
3.12	The wind power scenario in 2010 based on the wind power in 2009 . . . . .	118
3.13	The wind power scenario in 2010 based on the wind power in 2010 . . . . .	119
3.14	The wind power scenario of 10,000 MW installed capacity in 2030120	

3.15	Actual distributions of wind power ramp rate and phase angle	121
3.16	Distributions of ramp rates and phase angles of synthesized wind power scenarios . . . . .	122
4.1	Coastal load data and common component . . . . .	143
4.2	Wind power data and common component . . . . .	144
4.3	Coastal load data and new scenario . . . . .	146
4.4	Wind power data and new scenario . . . . .	147
4.5	PSD of actual load data . . . . .	148
4.6	PSD of actual wind power data . . . . .	149
4.7	The correlation coefficients of actual waveforms . . . . .	151
4.8	The correlation coefficients of synthesized scenarios . . . . .	152

# Chapter 1

## Introduction

Electric energy comes from many primary energy resources. One of those energy sources is wind energy although strictly speaking wind energy itself is primarily due to energy from the sun. In the Electric Reliability Council of Texas (ERCOT), of the 340,033,353 MWh annual total electrical energy produced in 2014, 36,142,384 MWh, or 10.6% of annual total electrical energy, came from wind energy [65]. Furthermore, the installed wind power capacity comprised 14% of the total installed capacity of generators, and the maximum generation of wind power was 10,957 MW on December 25, 2014, which comprised 34% of total demand. The total installed wind power capacity might increase in 2015 because about 3,000 MW of additional wind power is under an interconnection agreement [64].

In comparison, in 2008, wind energy comprised 2.9% of total annual energy, and the capacity provided by wind power comprised just 7.1% of total installed capacity [61]. It is clear that wind energy has increased tremendously between 2008 and 2014. Why have many companies invested in wind resources, what are the advantages and disadvantages of bringing more wind power into our power system, and how can we mitigate its disadvantages and emphasize



the advantages? The answers for the first two questions are briefly answered in this chapter. The rest of this dissertation aims to answer the last question.

## **1.1 Wind Power in the Electricity Market**

The explosive increase of wind energy in the U.S. has been encouraged by the efforts of the federal and state governments to take advantage of the benefits of wind power. These efforts have been realized through various economic incentives in order to facilitate the appropriate market environment for wind farm owners and wind turbine generating companies.

### **1.1.1 Benefits of Wind Power**

There are five main benefits of wind energy. 1) Wind energy is sustainable as long as the wind blows. 2) Since wind energy is a domestic resource that does not need to be mined and transported, the U.S. can increase energy independence and diversify its energy portfolio in order to respond to the changeable international energy market. 3) Wind turbines generate negligible amounts of atmospheric emissions that cause greenhouse effects. Reducing the greenhouse gas emissions can delay the greenhouse effect. 4) Since good wind sites are often in remote areas, the rural area can benefit from increased property taxes and direct lease payments. Moreover, land used for wind farms can still be used for agriculture. 5) Wind turbines do not require large amounts of water, in contrast to nuclear power and coal power [128]. In generating the same amount of energy, a nuclear power plant requires 500 times more water

than a wind turbine does.

In spite of these benefits, the leveled cost of wind energy in dollars per MWh is still higher than that of other energy resources. For example, the leveled cost of offshore and onshore wind power is \$204/MWh and \$80/MWh, respectively although recent costs have declined somewhat [3]. In contrast, the leveled cost of natural gas combined cycle generator is \$64/MWh. Therefore, in order to realize the benefits of wind energy, it is necessary to facilitate the participation of wind energy in the electricity market through various economic incentives.

Before introducing the economic incentives for wind energy, it should be noted that the development of renewable energy has also been spurred by state governments by setting specific and mandatory target amounts of renewable energy, which is called the renewable portfolio standard (RPS) [94]. The RPS is defined as the target capacity of renewable energy by the target year, or the target percentage of the capacity of renewable energy with respect to the load capacity. For example, in Texas, the RPS target is 10,000 MW by 2025, and it has already been achieved. The economic incentives that will be explained below have played an important role in satisfying the RPS.

### **1.1.2 Economic Incentives for Wind Power**

There are three economic incentives for wind power. 1) The marginal cost of wind power is almost zero, so wind power is typically dispatched in preference to all other resources in an electricity market. In fact, wind farm

owners sometimes bid at negative prices due to other incentives that will be explained below. 2) The federal government gives tax incentives to promote renewable energy businesses: wind farm owners are eligible for a production tax credit (PTC) of 2.3 cents per *kWh* generated for the first ten years. Because of the PTC, nearly \$15 billion was invested in wind power between 2007 and 2014 in the U.S [114]. It should be mentioned that the PTC expired on December 31, 2013, so new wind resources built after 2014 are not currently eligible, but debate about the extension is on-going. 3) Wind farm owners can earn a renewable energy certificate (REC) by selling one MWh. The REC can be sold in two types of REC markets, the voluntary and the compliance markets, and the price is always changing according to the year, the type of REC market, and the state in which the REC is sold. In the voluntary REC market, many companies buy voluntary RECs to show that they are environmentally friendly, and in the compliance REC market, the RECs are sold to load entities in each state to satisfy the RPS.

### **1.1.3 Wind Power As an Electricity Market Participant**

As a result of the above-described incentives, wind power has expanded in the electricity market. However, it has been difficult to handle wind power in the same manner as conventional power because of wind power's uncertainty and fluctuation. Since the wind power is generally determined by wind speed, which cannot be predicted with 100% accuracy, the wind power prediction always has a natural uncertainty. In addition, since wind farm owners cannot

control the intermittent nature of wind, the wind power output always fluctuates with various ramp rates, although modern wind turbines can control the ramp rates of wind power penetration to a certain extent.

Because of these two distinguishing characteristics of wind power, as compared with other generators, relaxed or different market rules have been applied to wind power in various electricity markets in the U.S. For example, in the New York ISO (NYISO), there is a penalty for non-compliance to the dispatch point outside a 3% margin of error, and the cost of the penalty is the multiplication of the deviation amount and the regulation clearing price. However, up to 3,300 MW of installed wind capacity is exempt from under-generation penalties. In the Midwest ISO (MISO), wind power, which is considered as the dispatchable intermittent resource (DIR), follows different market rules compared to conventional generators. For example, DIRs are exempted for deviations less than 30 MWh. In ERCOT, wind producers are only penalized if wind power deviates above the expected point, but not for the deviation below. If wind power deviates more than 10% above the expected point, it will be charged based on the real-time (RT) price and power balance penalty curve, and wind power cannot ramp more than 10% of its capacity within a minute. On the contrary, conventional generators are fined when their generation outputs deviate more than 5% above or below the expected point. Moreover, wind power is not allowed to participate in the ancillary service market. The different wind power market rules in different ISOs are summarized in Table 1.1.

Generally, participation in the RT market and acceptance of the dispatch signal is mandatory in most ISOs, but participation in the day-ahead (DA) market is voluntary. Furthermore, most ISOs allow a negative price. If wind power is used to provide the capacity resource, particularly in Pennsylvania New Jersey Maryland Interconnection (PJM) market, then wind farm owners should offer the capacity resource in the DA market. In PJM, wind receives capacity credit based on an average of performance over the previous three summers. If the operation duration is less than three years, wind power receives 13 percent of nameplate capability [187]. For the unit commitment, which decides when generators should be turned on and off, wind power is just considered as a negative load, and the wind power variability is represented as the net load (load – wind) variability. Sometimes, as in MISO, PJM, and ERCOT, wind power is curtailed because of transmission constraints or the minimum generation events. In ERCOT in particular, wind power is often curtailed because of insufficient capacity of the transmission lines that deliver much of the wind energy from West Texas to the load centers. In addition, the centralized wind power forecasting has been used to dispatch wind power and determine the amounts of regulation services to be procured in various electricity markets.

#### **1.1.4 Effect on the Power System and Electricity Market**

The introduction of wind power might affect the power system in terms of the wholesale electricity price, system inertia, and amounts of AS. However,

Table 1.1: Wind Power in Electricity Markets

Electricity Market	PJM	NYISO	ERCOT
Dispatch	Must participate in the RT market.	Must participate in the RT market.	Must participate in the RT market.
Day Ahead Market	Only as a capacity resource, wind power must bid in the DA market.	Voluntary	Voluntary
Negative Price	Allowed	Allowed	Allowed
Imbalance Relaxation	Costs are charged for power imbalance.	Wind power is exempt from under-generation penalty up to 3,300 MW of installed capacity of wind power	During testing periods, if wind resource generates more than 10% above, it will be charged on real-time prices.
AS Market	Can participate only in the regulation market.	Can participate only in the regulation market.	Can participate in the frequency response without payment. Can provide regulation service.
Forecasting System	Since 2009. Long term: hourly from 48 hours ahead to 168 hours ahead. Medium term: hourly from 6 hours ahead to 48 hours ahead. Short Term: at every 10 min with forecast interval of 5 min for next 6 hours.	It has been used since 2008. The DA forecasting is provided twice in a day (4am, 4pm), and the RT forecasting is provided at every 15 min.	It has been used since 2008. Hourly wind power forecasting also provides 50% and 80% probabilities of over-forecast for 48 hours.
Ramp Forecasting	Updated every 10 minutes at 5-min intervals for next 6 hours.	No ramp forecast	Probabilistic ramp forecasting is provided at every 15 min for the upcoming 6 hours.

determining the direct impact of wind power on the wholesale price might be difficult because the wholesale prices are also affected by weather, natural gas prices, transmission constraints, load, and other factors. The wholesale price in 2012 was lower than the price in 2008. Low natural gas prices and electricity demand are considered to be two primary contributors to this decrease. However, it is true that wind power can potentially reduce wholesale electricity prices since its marginal cost is close to zero, and wind power is sometimes bid with a negative price. Therefore, the strong impact of wind power on the wholesale price has not been clearly shown, but it will contribute to a decrease in the wholesale price as the wind power penetration level increases. Additional transmission lines might decrease the wholesale price further by delivering more wind power to the grid.

Another effect that wind energy might have on the power system is regarding the system inertia. The system inertia decreases as more wind farms are integrated into the power system. The inertia is an index to describe the sum of all kinetic energy in the on-line synchronous generators that are running with the same 60 Hz frequency. By following Newton's first law, the inertia delays the frequency deviation from 60 Hz when there is a power imbalance between the supply and demand [139]. On the contrary, most wind turbines, which are inverter-based generators or asynchronous induction generators, are not synchronized with other generators at 60 Hz, so they cannot directly provide the inertia to the power system. With reduced inertia, the system frequency drops faster after a generator contingency than the system

frequency with normal system inertia, so more responsive services are required to delay the frequency drop in the system with less system inertia. Although wind turbines can also provide synthetic inertia, since they require a control action triggered by the system frequency drop, a market-based procurement process would likely need to be established in order for wind farm owners to provide this service.

In addition to affecting the system inertia, increased wind capacity might also increase the amount of regulation services, which are deployed to compensate for the power imbalance within each five minute dispatch interval, because the wind power fluctuation within the five minutes can increase the net load fluctuation. However, the effect of increased wind capacity on the amounts of ancillary services (AS) has not been shown clearly yet because ISOs have changed their regulation procurement methodologies as the penetration level of wind power has increased. PJM has reported that there have been no serious impacts of wind power on the procured amount of regulation services. Although ERCOT's current regulation procurement methodology is also adequate to procure sufficient regulation services [195], ERCOT has increased the procured amounts of regulation services with respect to the installed capacity of wind power by following the guidelines in [195].

Furthermore, wind power forecasting has been used for calculating the amounts of non-spinning reserve services more accurately. The load and wind power forecast uncertainties are used to calculate the net load uncertainty and thus the amount of hourly non-spinning reserve services. Therefore, in-



creased wind capacity might increase the amount of regulation services, and the improved performance of wind power forecasting is required to estimate the amounts of non-spinning reserve services more accurately.

## **1.2 Solutions to Increase the Wind Power Capacity**

Three obstacles to increase the penetration level of wind power are wind power fluctuation preventing participation in the DA market where prices are less volatile and higher than the RT market, the increasing cost of AS, and limited transmission resources. There are three key solutions to these obstacles: promoting participation of wind power providers in the AS and DA markets by improving the wind power forecasting and developing the storage system, developing smart AS procurement methodologies, and building additional transmission lines.

### **1.2.1 Advanced Ancillary Services Procurement Process**

As the wind capacity increases, the optimal amounts of AS, responsive services, regulation services, and non-spinning services can be procured based on the installed wind capacity, the inertia, governor response ramp rate, droop-characteristic, and net load variability [42]. In order to compensate for the decreasing inertial response capability, a new process to procure the responsive service can be determined by considering governor response ramp rates and system inertia so that the responsive service stabilizes the frequency when a generator trips [43]. The amounts and ramping capability of regu-

lation services can also be set with respect to the net load variability while satisfying the Control Performance Standard (CPS) [44]. As the wind penetration level increases, the absolute magnitude of net load variability also increases, so the required regulation service is also expected to increase. Although wind turbines can only provide down regulation, curtailment of wind power to provide the down regulation might erode the low-cost advantage of wind power. The non-spinning service, which compensates for net load variability and recovers the 60Hz frequency after the resources are deployed, can be upgraded to procure resources by considering the ramp rates of resources. For example, ERCOT plans to substitute the current non-spinning service for the supplemental reserve (SR) service. It should be noted that ERCOT's new AS system includes six different services: synchronous inertial response (SIR), fast frequency response (FFR), primary frequency response, up and down regulating reserve, contingency reserve, and SR services [63].

### **1.2.2 Storage System**

Participation in the DA and AS markets can be expanded through the use of the storage systems, which will mitigate the effect of wind power fluctuation and reduce the exposure of wind farm owners to the volatile price in the RT market. The storage system can also be used to relieve the transmission congestion of wind power, mitigate wind power fluctuation, and provide AS. A large capacity storage system, such as the compressed air energy storage (CAES), can relieve the transmission congestion due to wind power if it is

installed around the transmission lines and wind farms. Furthermore, by limiting the ramp rate of wind power, the storage system can mitigate the wind power fluctuation. For example, Xtreme Power installed a storage system of 10 MW power rating and 20MWh storage size on a 21 MW wind farm in Maui, Hawaii to limit the ramp rates of wind power. The ramp rate limit was  $\pm 1$  MW / min. Xtreme Power's storage system can also provide the regulation service because it can respond to the regulation signal very quickly [172]. ERCOT has a plan to introduce the storage system as the fast responding regulation service (FRRS). This storage system might not substitute for the inertia completely, but energy that discharges quickly from the storage system can delay the time when the frequency reaches the frequency nadir. In addition, the storage system can be used to increase the profits of wind farm owners by helping them participate in the DA market and hedge the price difference between the RT and DA markets. Since the energy price in the DA market is statistically higher than the price in the RT market [58], participating in the DA market can increase the profit. If wind farms do not generate the bid amount, they should pay back the cost corresponding to deviation amount at the RT price. However, if the RT price is much higher than the DA price, they would lose a lot of money.

### **1.2.3 Electric Vehicles**

Electric vehicles (EVs) can be used to mitigate not only the daily fluctuations of wind power by consuming wind energy at night, which is on average

the time of peak wind production in ERCOT, but also the instantaneous fluctuations of wind power by providing regulation services [179]. With respect to the storage application mentioned above, EVs are a good application of a storage system within the power system since the high cost of the batteries can be distributed to EV owners. However, in order to provide the regulation services in the AS market, the minimum capacity requirement should be satisfied, so an aggregator that coordinates EVs and distributes the regulation signal to the EVs is needed [86]. However, to implement EVs in the AS market, it is necessary to estimate the size of the AS market with EVs because the introduction of fast response devices might decrease the required quantity of AS procurement and reduce the regulation clearing price. Furthermore, it is necessary to develop algorithms to dispatch signals and optimal charging schedules for EVs. Proper incentives for EV owners should be determined to make EVs follow the charging schedules and the dispatch signal. In addition, the communication between the aggregator and EVs and metering technology need to be developed to measure the regulation performance. Moreover, regulation policies for interconnection and settlement must be established.

#### **1.2.4 Wind Power Forecasting**

Accurate wind power forecasting is beneficial to system operators because it reduces the imbalance of supply and demand by reducing the net load variability. It is also beneficial to wind farm owners because it increases the profits by helping them participate in the DA market, as mentioned above.

Wind power forecasting in the DA market can reduce the imbalance between the bidding and delivering amounts. Furthermore, accurate short-term wind power forecasting is also beneficial to rate payers: the more accurately the wind farm owners bid, the lower the cost of AS to the rate payers. Wind power forecasting has been advanced in two different ways. First, ensemble forecasting, which combines multiple forecasting machines or combines predictions from multiple numerical weather prediction (NWP) scenarios, can increase the forecasting performance [157]. Second, probabilistic wind power forecasting, which provides the error distribution of point-prediction can be used to statistically build the optimal decision in the electricity market [196]. For wind farm owners, probabilistic wind power forecasting can also provide opportunities to hedge their losses based on a statistical decision when they participate in the DA market. For independent system operators (ISOs), probabilistic wind power forecasting can be used in the stochastic unit commitment and economic dispatch.

### **1.2.5 Transmission Expansion**

In order to utilize the benefits of wind power more, the penetration level of wind power should be increased by building additional transmission lines. Particularly, in ERCOT, since there are ample wind resources in western Texas, new transmission lines have been built through the Competitive Renewable Energy Zones (CREZ) project [83]. Furthermore, in the Texas panhandle, additional transmission expansion, which is called the Panhandle

Renewable Energy Zone, is ongoing [160]. The motivation behind transmission expansion planning is to ensure a balance between future generation and load and to decrease the total generation costs by delivering more wind power and relieving congestion while maintaining the balance between load and generation and system reliability. The goal of expansion planning is to determine the locations of transmission lines to be enhanced or newly built. Economically optimal transmission expansion occurs when the sum of investments in transmission lines and operating costs is minimized.

### **1.3 Motivation and Goals**

As mentioned previously, the first problem of the current and high penetration level of wind power is the wind power uncertainty, which inhibits wind farm owners from bidding larger amounts of energy in the DA market, and the second problem is the wind power fluctuation, which might increase the required amounts of AS, because the power imbalance between the supply and demand within each dispatch interval increases the net load fluctuation.

Then, the first key solution to increase the penetration level of wind power is to increase the accuracy of wind power forecasting, and this will lead to the expanded participation of wind power in the DA market and to the reduction of the exposure to the price variability in the RT market. Furthermore, more accurate bidding amounts might reduce the amount of AS. Although novel methods to estimate the error distribution and to merge multiple forecasting machines are well developed in the literature, there is some room to

upgrade these forecasting methods. The framework should also include various forecasting techniques as a module so that users can select forecasting techniques with respect to application cases. A new framework for providing the point forecast and its distribution should be established by using parallel programming and high computing resources when wind power from multiple wind farms is forecasted based on a large amount of historical wind power data. Recently, off-the-shelf forecasting machines, particularly tree-based forecasting machines, have been well developed. The performance of ensemble forecasting machines, including these off-the-shelf machines, should be tested.

Therefore, the first goal is to develop a new framework for probabilistic and ensemble wind power forecasting in order to provide the accurate probabilistic and point forecasting in an efficient way within a short computation time. A new framework that consists of multiple modules must be developed so that users can build case-sensitive forecasting machines by selecting modules. The modules include the feature engineering techniques, preprocessing techniques, and various forecasting machines. The framework will provide the averaged predictions from various forecasting machines and their error distributions by sharing the cross validations of distribution estimating and prediction averaging processes. The shared cross validation can be implemented through a parallel computing environment in order to increase the computation speed. In this study, the parametric approach, which presents the distribution in a closed form and compresses the distribution information, is used to estimate the error distribution. The parametric approach assumes that the

error distribution will follow the VG distribution and generates the continuous function of the conditional error distribution. The proposed framework will also enable ensemble forecasting by averaging predictions from multiple forecasting units after multiplying weight factors, which are the inverse of the performance. In ensemble forecasting, seven forecasting machines are used: ridge regression, support vector machine, gradient boosting machine (GBM), neural network (NN), random forest (NN), Gaussian process, and bootstrapped aggregation (BAG).

The second key solution to increase the penetration level of wind power is to develop a smart AS procurement methodology. The preliminary step in determining the new rules for procuring AS is to synthesize many different scenarios of future wind power by forecasting the statistical characteristics of future wind power. These scenarios will be used to simulate the future power system with different simulation settings including the future installed capacity of wind power, seasonal trends, and different geographical dispersions. Furthermore, wind power from integrated wind farms has less fluctuation relative to the capacity than wind power from a single wind turbine. This is called the geographical smoothing effect. As more wind farms are built across distributed areas, the increased geographical smoothing effect should be considered when future sample paths of wind power are synthesized.

Therefore, the second goal of this dissertation is to synthesize the future monthly total wind power scenarios that are representative of the future diurnal trends, short-term fluctuations, and long-term fluctuations simultane-



ously. Therefore, scenarios will be synthesized by considering the deterministic and stochastic portions. In this process, it is assumed that the deterministic portion of the future wind power, the 24h daily cycle, will change in linear proportion to the total capacity. For example, wind power in Texas has a strong 24h daily cycle that peaks at 2:00 a.m. The 24h daily cycle is measured from the actual wind power and designed under the assumption that it follows the periodic waveform. On the contrary, the stochastic portion is assumed to be changed with respect to the power laws of the power spectral density (PSD). PSDs in a logarithmic plot are approximated by piecewise affine functions through the modified hinges model, and slopes of affine functions are forecasted as the total wind power capacity increases. The first and last PSD values are also forecasted with respect to the total capacity. The first PSD value corresponds to the minimum frequency  $32 \text{ days}^{-1}$ , and the last PSD value corresponds to the maximum frequency  $2 \text{ minutes}^{-1}$ . Then, the wind power scenario is synthesized by converting the forecasted PSD through the inverse discrete Fourier transform (DFT). In this process, phase angles are searched through the genetic algorithm (GA) to satisfy the distribution of wind power variability. Furthermore, scenarios must be scalable with respect to the future installed capacity of wind power while satisfying the forecasted statistical characteristics of future wind power. In short, works to accomplish this goal will be to design the 24h daily cycle separately, to model the stochastic portion by analyzing power laws of the PSD while satisfying the regressed statistical characteristics, and to reflect the forecasted future statistical characteristics.

Third, physical transmission lines should be built to bring more wind power into the grid. Transmission planning can be performed by simulating the power system with future individual scenarios of wind power and load in order to consider the net load uncertainty properly. Wind power causes an additional uncertainty to be considered because wind power is often not generated during peak load times, and wind power is generated differently at different locations. Analyzing the total wind power scenarios is not sufficient to capture the spatial and temporal wind power fluctuation of all wind farms, so sample paths of all wind farms should be considered simultaneously. Accordingly, sample paths of wind power from individual wind farms and load from individual weather zones are synthesized simultaneously while maintaining the stochastic geographical structure of wind power and load.

Therefore, the third goal is to synthesize individual load and wind power scenarios by considering the geographical characteristics of load and wind farms and the correlation structure between wind power and load. This goal can be accomplished by using the generalized dynamic factor model (GDFM) that models multiple sample paths as a linear combination of a filter and dynamic factors. The filter represents geographical characteristics of load and wind power. Dynamic factors that represent a few streams of wind speed can also be represented as the linear combination of dynamic shocks, which represent the randomness of wind power. The dynamic shocks are assumed to be uniformly distributed white noises. By changing dynamic shocks and future nameplate capacities, an infinite number of sample paths of future load and

wind power scenarios can be synthesized. In this process, the dimension of dynamic shocks is less than the dimension of the observation data, so it is expected that the number of variables will be reduced. In order to detect the dynamic factors and redesign the dynamic shocks, the GDFM decomposes observation data into a common component, which is generated by the dynamic movement of dynamic factors, and the idiosyncratic noise component. Then, the GDFM detects the dynamic factors by applying the dynamic principal component analysis to the correlation function between common components of different time series, so it is called a dynamic model. In addition, because the GDFM relaxes the orthogonality constraint in the idiosyncratic noise component by allowing weak correlations among noise components, it is called a generalized model. The detected dynamic factors are modeled using the vector autoregressive (VAR) process of dynamic shocks. Furthermore, transmission expansion planning is formulated to minimize investment costs for new transmission lines, enhancement costs for existing transmission lines, and operating costs. The transmission planning problem is a mixed integer program, and it is decomposed into a two-stage problem. In the first stage, investments in transmission lines are optimized, and in the second stage, operating costs of the given network from the first stage are optimized, given the realized wind and load.

## 1.4 Organization

The organization of this dissertation is as follows. The first key solution is to increase the performance of wind power forecasting and provide more information than the point forecast through the probabilistic wind power forecasting. Therefore, in Chapter 2, a new framework for short-term probabilistic and ensemble wind power forecasting is proposed with various forecasting techniques. The second key solution, which is to develop a new AS procurement methodology, can be facilitated by simulating power system with many different wind power scenarios. Therefore, in Chapter 3, future scenarios of total wind power in ERCOT are synthesized by forecasting and converting the PSD of wind power. The third key solution is to plan additional optimal transmission lines to bring more wind power into the grid via multiple simulations with different scenarios. To this end, Chapter 4 introduces a GDFM-based algorithm to synthesize individual load and wind power scenarios. The usefulness of scenarios is also verified by calculating the total generation and transmission upgrading costs on the IEEE 300-bus benchmark. Finally, Chapter 5 summarizes the dissertation and suggests the future work.

## Chapter 2

### Short-Term Wind Power Forecasting

In this chapter, a novel probabilistic and ensemble forecasting algorithm is suggested to predict the hourly wind power output and its probabilistic distribution for a month for ten wind farms. The probabilistic distribution is given as 1% quantiles. Probabilistic wind power forecasting can provide a solution to reduce the risk to system operations [23] and an opportunity to maximize wind farm owners' profits [154]. Furthermore, ensemble forecasting, which combines multiple forecasting machines with different settings [88] or combines the forecasted wind power of multiple numerical weather prediction (NWP) scenarios, can increase the forecasting performance [157]. The framework and techniques are verified using data from the 2014 Global Energy Forecasting Competition (GEFCom).

The probabilistic and ensemble forecasting algorithm is implemented in a new framework, which consists of multiple modules so that users can build case-sensitive forecasting machines by selecting modules. The modules include the feature engineering techniques, preprocessing techniques, and various forecasting machines with different simulation settings. The framework will perform the probabilistic forecasting and ensemble forecasting simultane-

ously by sharing the cross validation. For the probabilistic forecasting, two different parametric approaches to estimate the error distribution are proposed. It should be noted that approaches to estimate an error distribution can be classified into the non-parametric approach, in which the prior information about the error distribution is not assumed, and the parametric approach, in which the error distribution is assumed to follow distributions in a closed form. In this study, the parametric approach assumes that the error distribution will follow the approximated variance gamma (VG) distribution and generates different distributions with respect to different prediction levels. The continuous function of the conditional error distribution without data classification will also be proposed. The performance of the probabilistic forecasting will be measured by the pin-ball loss. For the ensemble forecasting, the weighted averaging, in which the weight factors are estimated by using the shared cross validations, is used. A forecasting unit in the ensemble forecasting is a single combination of a forecasting machine, internal parameters, and simulation settings as described in Figure 2.1. The seven forecasting machines are used to create heterogeneity and to increase the performance. These machines are the ridge regression, neural networks (NN), support vector machine (SVM), Gaussian process (GP), gradient boosting machine (GBM), random forest (RF), and boosted aggregation (BAG). Furthermore, the optimal memory length of wind speed, best forecasting machines, and their internal parameters are also calculated based on the framework. The proposed forecasting framework is verified using the data in the 2014 Global Energy Forecasting Competition

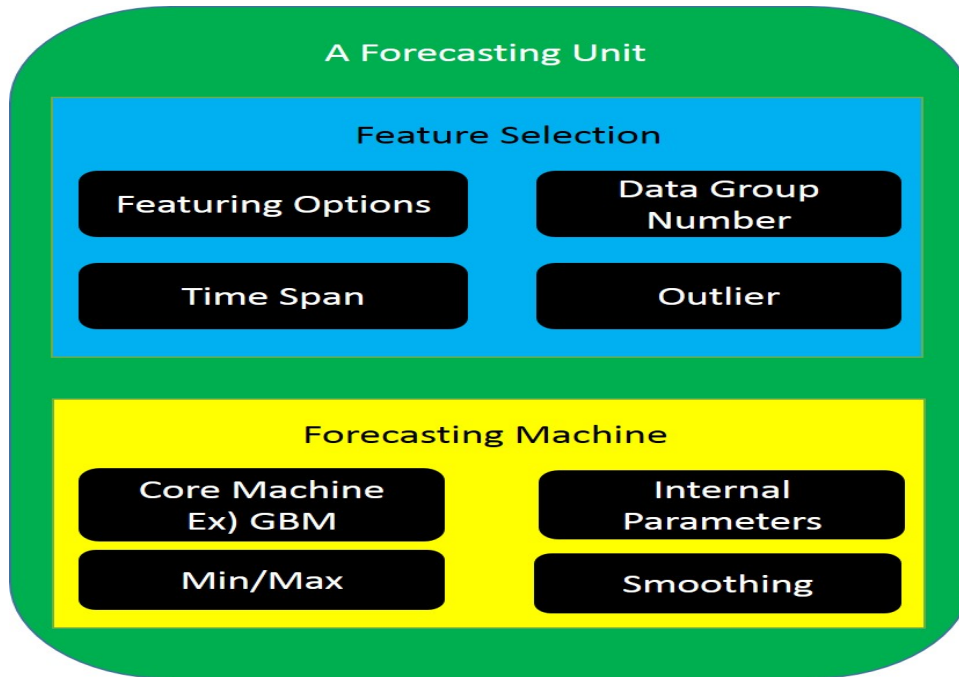


Figure 2.1: The forecasting unit consists of simulation settings, internal parameters, and forecasting machine.

(GEFCom). The best weekly ranking was second place, and the intermediate ranking was sixth among 215 participants.

## 2.1 Introduction

Much of the literature has provided non-parametric and parametric approaches to estimate the error distribution of wind power. Many algorithms of the ensemble forecasting have also been proposed.

### 2.1.1 Literature Review

To increase the performance of wind power forecasting, probabilistic and ensemble forecasting techniques have been researched for a long time. The error distribution of the wind power predictions can be estimated directly from the training data without estimating the wind power predictions by using the special structure of a forecasting machine. For example, in [106], the neural network (NN) is used to generate the prediction interval by building two-output network since one of advantages of the NN compared to other forecasting machines is to have multiple outputs simultaneously. The first output estimates the upper boundary of the given training case, and the second output estimates the lower boundary of the given training case. Weights are updated in order to minimize a special cost function, which tries to increase the probability that the prediction is inside boundaries and to decrease the width of the error distribution. Since this algorithm is only possible within the NN, it cannot use other forecasting machines having high forecasting power. Therefore, it would be beneficial to develop an algorithm to use an ensemble of multiple forecasting machines having high forecasting power compared to the NN. Furthermore, instead of estimating quantiles by stacking lower and upper boundaries, the error distribution can be globally estimated while saving computation time.

Furthermore, instead of changing the internal parameters, in [99], the error distribution is generated by changing the input wind speed. Many different wind scenarios are generated by putting random noises, which are gener-



ated by using the Monte Carlo simulation, into the vector autoregressive moving average-generalized autoregressive conditional heteroscedastic (VARMA-GARCH) model of wind speed and direction. The error distribution of wind power is estimated by converting many different wind speed scenarios to wind power scenarios using the stochastic power curve, which is based on the conditional kernel density. However, this approach might provide the general error distribution, but it cannot provide the tailored distribution with respect to the point forecast and explanatory variables.

Without assuming the parametric or non-parametric error distributions, the quantile can be estimated directly from the explanatory variables. The quantile regression can be used to estimate the error distribution [142]. In this application, developing separate models for each quantile takes a high computation time. Furthermore, better forecasting machines than the regression can be used to estimate the error distribution. The error distribution can also be determined by solving an optimization problem as described in [196]. For example, a direct optimization of both the coverage probability and sharpness is used to estimate the optimal error distribution with respect to the performance measure unit of the probabilistic forecasting.

The error distribution can also be estimated independent of forecasting machines. In [156], the adaptive re-sampling method is used to estimate the error distribution based on the classification of wind power without prior information about the error distribution. On the contrary, instead of re-sampling, it would be advantageous to get the distribution of the forecast errors through

the cross validation. However, the data clustering is also used in this application in a limited way. For the given prediction, its error distribution is truncated by the minimum and maximum of the actual values in the prediction's corresponding cluster. In this paper, an algorithm to estimate the continuous function of the conditional error distribution without data classification is proposed. Furthermore, it has already been shown that ensemble forecasting can have better forecasting power [119], but the forecasting power can be further increased by combining tree-based advanced forecasting machines, such as the random forest (RF), bootstrapped aggregation (BAG), and gradient boosted machine (GBM), with different simulation settings. Therefore, as shown in [157], combining probabilistic forecasting and ensemble forecasting might increase the performance of wind power forecasting.

### 2.1.2 Global Energy Forecasting Competition

As the training data, four explanatory variables, which are the zonal and meridional values of wind speed at 10m and 100m, are provided for 10 wind farms, so there are 40 explanatory variables. The error distribution of one-hour-ahead wind power outputs were forecasted for ten wind farms. The  $x$ -axis and  $y$ -axis values of wind speed at 10m and 100m were provided. Wind speed was also predicted based on one-hour-ahead NWP. The error distribution was represented as 99 quantiles. This competition consists of 15 competitions for 15 weeks. Every week, one hour-ahead wind power for a month was forecasted. In the next week, wind power for the next month was

forecasted, and the testing data in this week became the last month of the training data in the next week. In this study, the data from the only second week is used. The training data was measured from 1/1/2012 to 10/31/2012, and the testing part is from November, 2012.

## **2.2 Program Architecture**

In this section, the architecture of the forecasting program is introduced. In a broad sense, the forecasting program consists of three stages. The first and second stages share the same internal forecasting process with the same simulation settings.

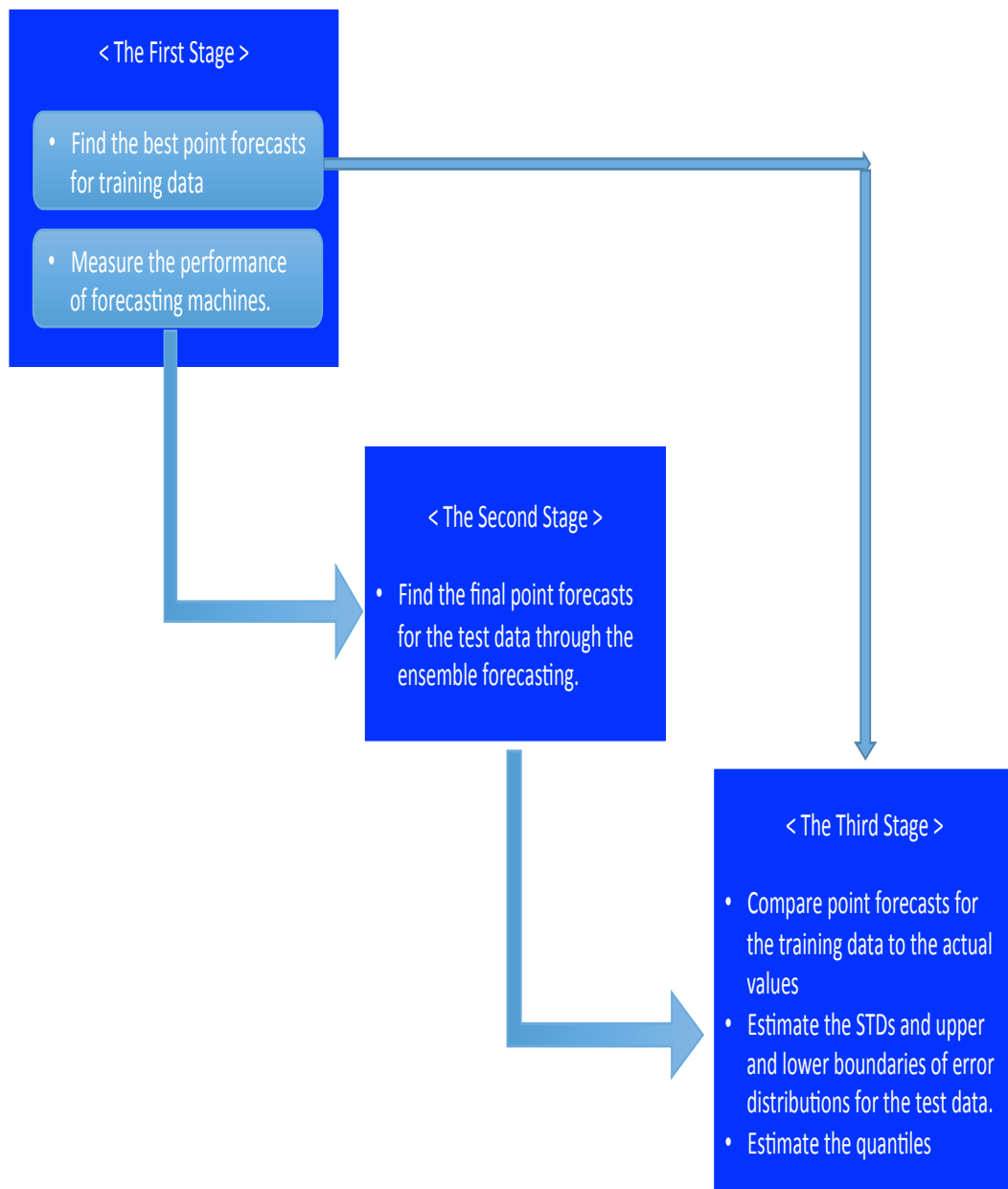
### **2.2.1 External Forecasting Process**

In the first stage, the performances of forecasting units are measured using the root mean square error (RMSE) through the external five-fold cross validation where the training data is classified into the actual training data and validation data. In this application, 20% of the data is selected as validation data recursively. Furthermore, point forecasts for the validation data are estimated through the ensemble forecasting. In the second stage, the final point forecasts for the test data are estimated through the weighted ensemble forecasting, in which weight factors are set as inversely proportional to the forecasting machines' RMSEs, which are estimated in the first stage. In the third stage, the standard deviations (STD) of the error distribution of the test data are estimated by comparing the point forecasts for the validation data in

the first stage to their actual values. The step-wise process of three stages is shown in Figure. 2.2.

### **2.2.2 Internal Forecasting**

As mentioned above, the wind power is actually forecasted by the forecasting units in the second and third stages. The forecasting units in both stages are built according to the same various layers of simulation settings, which is described in Figure 2.4. Many combinations of different simulation settings can be structured in a tree, and the tree structure is implemented in this program through five for-loops that are layered. Since each loop has multiple options, the total number of simulations is the multiplication of the number of options in each layer. The five layers consist of system parameter settings, data transformation, time expansion, outlier detection, and individual forecasting machines. The structure is shown in Figure 2.3. In the first sequence, different fixed parameters of forecasting machines are set. Then, the training data is expanded by adding the quadratic, cubic, and square of root terms in order to extract the nonlinear relationship between wind speed and wind power. Furthermore, the wind speed on the target day is expanded by adding new features that include time information and forecasted weather data on the day before and the day after the target day. Outliers are also defined and removed. In the preprocessing, each feature of each class has a zero mean and uniform standard deviation (STD). For the given actual training data, the internal cross validation is used to determine the optimal parameters in the



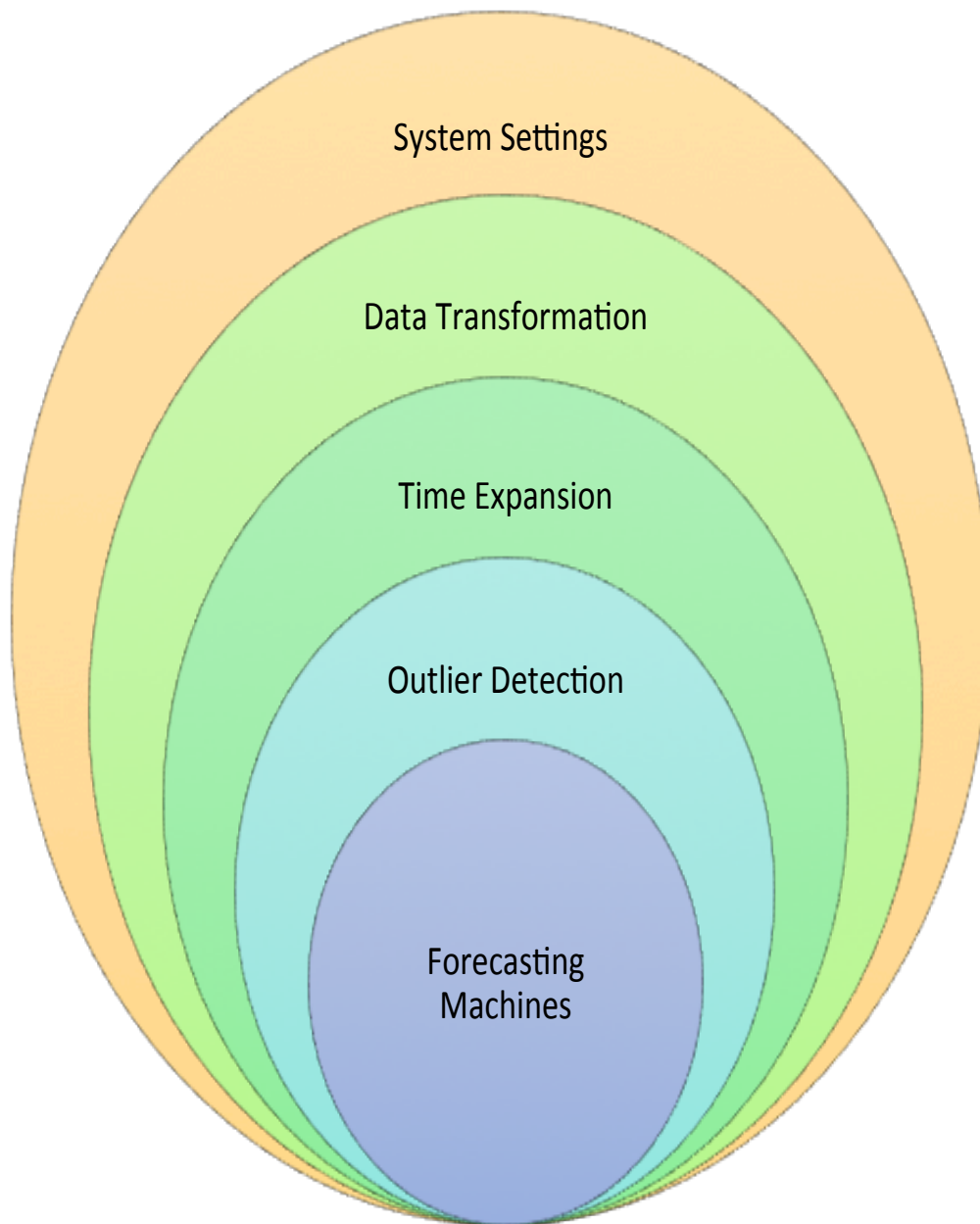
*Figure 2.2: The three-stage program architecture is introduced. The best forecasting units, STDs, minimum, and maximum of quantile for each output groups are delivered from the first stage to the second stage.*

forecasting machines. Twenty percent of the training data is set aside to verify the internal parameters. Finally, the predictions are smoothed, the minimum level of predictions is limited to zero, and the maximum level of predictions is limited to the maximum value of training data.

### **2.2.3 First Stage**

In the first stage, point forecasts for the training data and the performance of the forecasting units are estimated. The external five-fold cross validation is used to measure the performance of the forecasting units. However, in the internal cross validation, when the internal parameters of forecasting machines are estimated, different validation data is selected from the actual training data to select the optimal internal parameters of forecasting machines. The point forecasts for the training data are generated through the ensemble forecasting using the predictions of forecasting units. The strategy with ensemble forecasting is to make as many different input or feature matrices as possible and run them with different forecasting machines.

Since all forecasting units are evaluated through the same validation data, the performances are compared fairly. Furthermore, since the validation set is alternatively built through the external five-fold cross validation, the point forecasts for all training data can be estimated. Then, the point forecasts of training data are estimated by averaging the predictions of forecasting units with weight factors. Therefore, in the first stage, the performances of the forecasting units are estimated and used to estimate the point forecasts



*Figure 2.3: The MATLAB code is organized in this way.*

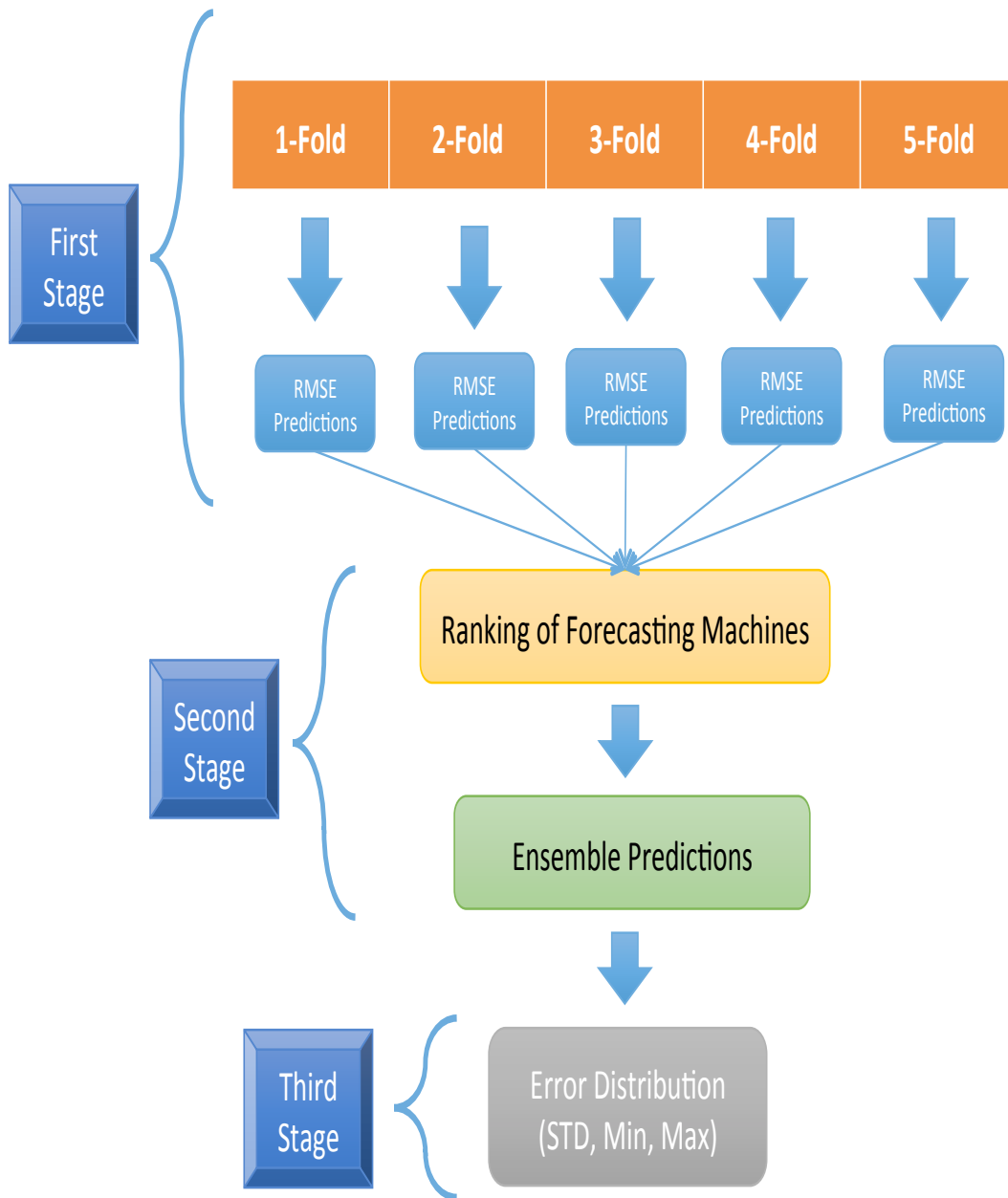


Figure 2.4: The overall forecasting process of each stage is introduced.



for the training data. The performances are also used to estimate the point forecasts for the test data in the second stage. In addition, in the third stage, point forecasts for all training data are used to estimate the STDs of error distributions, lower boundaries, and upper boundaries by comparing them with the actual values.

#### **2.2.4 Second Stage**

In the second stage, all training data is considered as the actual training data, and the final point forecasts for the test data are estimated through the ensemble forecasting by using the RMSEs in the first stage. In other words, forecasting units are re-trained using all training data. If the number of forecasting units is high, or if the performances of a few forecasting units are much lower than those of other forecasting units, a few best combinations in the first stage are used in the second stage. It should be noted that the computation time of the second stage is less than that of the first stage since it does not need to perform the external cross validation.

#### **2.2.5 Third Stage**

The point forecast of validation data are compared to the actual validation data in order to estimate the standard deviation, maximum, and minimum values of the error distribution under the assumption that the error distribution follows the approximated VG distribution. In the non-parametric approach with the data classification, the point forecast in a different group

is compared to the validation data in the same group, and each group has different standard deviation, maximum, and minimum values of the error distribution. The optimal STD is determined when a candidate STD minimizes the sum of the quantile errors of the point forecasts in each group. By increasing the STD from zero, the sum of measured quantile errors is calculated with respect to the STDs. On the contrary, in the non-parametric approach without the data classification, the optimal STDs of each point forecast of the test case is estimated with respect to the point forecast of the test data through another forecasting machine. In order to train this forecasting machine, the optimal STDs of the training data are used as the target data, and the weather data and point forecasts of the training data are used as training data for this forecasting machine. In short, this forecasting machine calculates the optimal STDs as a function of weather data and point forecasts. However, in the non-parametric approach, the data classification is also used to determine the maximum and minimum values of the error distributions. Therefore, for the given point forecast of the test data, the minimum and maximum values of wind power in the cluster that the point forecast belongs to are used to limit the error distribution of that point forecast.

### **2.3 Feature Engineering**

The goal of feature engineering is to design the number of rows and columns of the input data matrix to generate the training data. In this section, feature engineering methods to tailor the input data matrix are explained

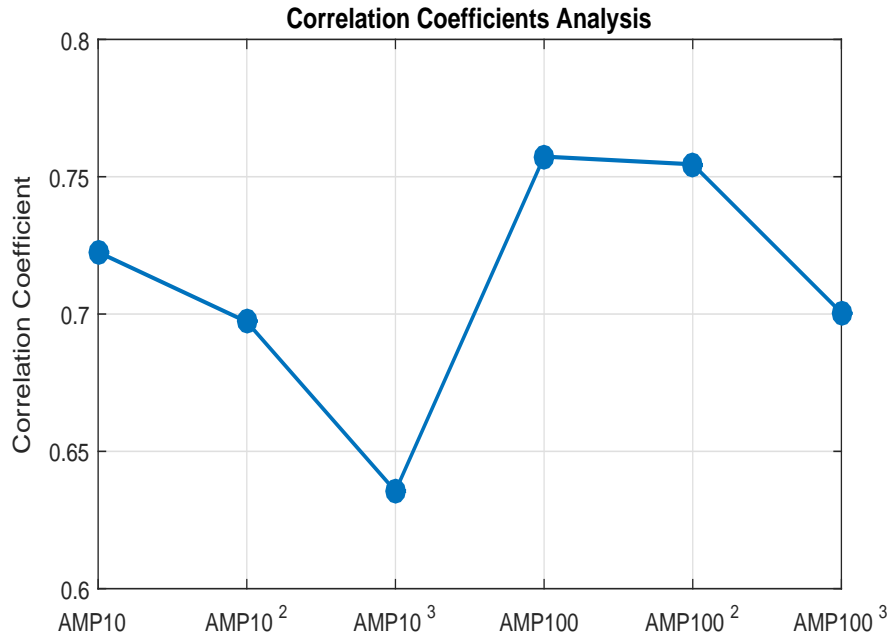


Figure 2.5: Correlation coefficients of transformed amplitudes of wind speed at 10m and 100m are plotted. The linear, quadratic, and cubic of amplitudes of wind speed at 10m and 100m are shown respectively.

individually.

### 2.3.1 Data Smoothing

The absolute value of raw wind speed data is smoothed in order to represent the dynamic movement of wind speed since NWP data that is generally sampled at a fixed interval cannot represent the time series characteristics of wind speed. In the competition, the moving average is used, and the optimal length of the moving window per wind farm is also decided.

### 2.3.2 Data Transformation

The input features are transformed into the quadratic and cubic in order to extract the nonlinear relationship between wind speed and wind power. The transformed input features include the zonal component, meridional component, and amplitude of wind speed at 10m and 100m. For example, the correlations of amplitudes of wind speed at 10m and 100m are compared to the correlations of the squared amplitudes and cubic amplitudes in Figure 2.5. It is observed that the wind speed at 100m is more highly correlated than the wind speed at 10m, although it is widely known that the cubic of wind speed is more highly correlated than the linear wind speed. It should be noted that the interaction terms of multiplication among features are not considered since including the interaction terms increases the number of variables exponentially.

### 2.3.3 Data Expansion

Since wind speed at any given moment is highly correlated with the wind speed an hour before and an hour after the given moment, in order to consider the dynamic nature of wind speed, all predictors from  $t - k$  to  $t + k$  are used to predict wind power at  $t$ . This data expansion technique can also control the number of variables, so it can also control over-fitting. The length of the timespan could differ based on the training data, but generally  $k$  is set at one or two. However, a three-hour timespan deteriorates the forecasting performance. In the ensemble forecasting, different timespans are used for

multiple forecasting units, and a few best forecasting units are selected. Furthermore, month, day, days of the year from one to 365, and year are added in order to extract the seasonal trends.

#### **2.3.4 Outlier Detection**

Outliers are defined and removed from the feature space. Predictors are ranked according to the difference between the target value and the prediction value forecasted by the ridge regression. The errors are measured as the absolute of this difference, and errors are sorted in descending order. It should be noted that each wind farm has its own outliers. Furthermore, the fraction between the outliers and training data should be searched before the simulation. The top 5% of errors are generally defined as outliers and removed. It is also interesting to observe that every forecasting machine has a different ability to resist outliers. For example, the NN seems to be weak for outliers, but the SVM is resistant to outliers. Furthermore, it is known that the GBM and RF are typically resistant to outliers [97], but they are susceptible to outliers in this simulation.

#### **2.3.5 Input Data Classification**

Wind power data can be classified into three groups with respect to the amplitude of wind speed. Generally, the power curve of the wind turbine can be classified into three ranges: wind speed less than the cut-in speed, wind speed between the cut-in speed and cut-out speed, and wind speed more

than the cut-out speed. However, since the performance of a single group is better than that of three groups, the input data classification is not used in this application.

### **2.3.6 Pre-processing and Post-processing**

After the simulations, 18 variables have shown the best performance. They are the zonal component of wind speed at 10m, zonal component of wind speed at 100m, meridional component of wind speed at 10m, meridional component of wind speed at 100m, amplitude of wind speed at 10m, amplitude of wind speed at 100m, square of zonal component at 10m, square of zonal component at 100m, square of meridional component at 10m, square of meridional component at 100m, cubic of meridional component at 10m, cubic of meridional component at 100m, square of amplitude at 10m, square of amplitude at 100m, cubic of amplitude at 10m, cubic of amplitude at 100m, angle of the cylindrical coordinate of wind speed at 10m, and angle of the cylindrical coordinate of wind speed at 100m. If predictors in the previous hour and next hour are included, the number of variables increases to 54. Then, predictions from ten wind farms are aggregated, so the number of variables increases to 544 by including day, month, day of the year, and year. However, since some features are shared between wind farms, the final number of variables is 528.

However, according to the input data generation method, the number of total features is changed. Therefore, many options are offered to control the number of features. For example, the first input data option has only raw

observation data, the second input data option has linear and quadratic terms of input data, and options become more complex after that.

Then, training data has a zero mean and has one as the standard deviation before it enters into the forecasting machines. After wind power is forecasted, the final prediction is smoothed. In addition, the prediction per wind farm is limited by the minimum and the maximum of each wind farm.

### **2.3.7 Simulation Settings for Individual Forecasting Machines**

Different simulation settings are fixed before running the forecasting. Simulation settings include the ratio of the number of outliers, the outlier detecting machine, and some fixed internal parameters in forecasting machines. Some simulation settings that can outperform other settings are set before the training data enters the loops in order to increase the system efficiency. Therefore, if the best settings are detected, they are enumerated in the code. It should be noted that the prediction horizon is not considered, and the ensembles of NWP with different initial settings are not used in this case study. If the prediction horizon needs to be considered, it would be preferably to have different forecasting machines for each prediction horizon.

## **2.4 Forecasting Machines**

In this section, seven forecasting machines are introduced: ridge regression, NN, SVM, GP, GBM, RF, and BAG. The forecasting machines used in this study can be classified into regression models, kernel-based models,

and tree-based models. The methods of controlling the internal parameters to avoid over-fitting are briefly introduced. In this forecasting, the  $n$ th target value  $t_n$  is estimated as the prediction  $y_n$ , which can be represented as a linear combination of weights  $\mathbf{w}$  and kernel functions of the  $n$ th predictor  $\mathbf{x}_n$  as

$$y_n = \mathbf{w}\phi(\mathbf{x}_n), \quad (2.1)$$

where  $\mathbf{x}_n = \{x_{n1}, x_{n2}, \dots, x_{nD}\}$ ,  $\mathbf{w} = \{w_1, w_2, \dots, w_M\}$ , and  $\phi = \{\phi_1, \phi_2, \dots, \phi_M\}$ . Training data consists of a set of target values and predictors.

#### 2.4.1 Ridge Regressions

The ridge regression is a regularized regression that is used to avoid over-fitting. It minimizes the sum of the squared errors and regularized coefficients, which is

$$\sum_{i=1}^N \left( y_i - \beta_0 - \sum_{j=1}^D \beta_j x_{ij} \right)^2 + \lambda \sum_{j=1}^D |\beta_j|^q. \quad (2.2)$$

In (2.2),  $q$  is 2 for the ridge regression. In this study, the shrinkage penalty  $\lambda$  is selected among different lambda values in the ridge regression through the cross validation. Furthermore, predictors are standardized since the effect of the shrinkage penalty might be different with respect to the scale of the predictors. The (2.2) is minimized by converting it to the quadratic optimization



problem below:

$$\min_{\boldsymbol{\beta}} \sum_{i=1}^N \left( y_i - \beta_0 - \sum_{j=1}^D \beta_j x_{ij} \right)^2 \quad (2.3)$$

$$\text{subject to} \quad \sum_{j=1}^D |\beta_j|^q \leq s. \quad (2.4)$$

For the given *budget* limit  $s$ , the shrinkage penalty  $\lambda$  becomes a Lagrangian multiplier of the inequality in (2.4). By reducing  $s$ , so  $\lambda$  can be increased.

The advantage of the ridge regression is the computation time; thus, it is very useful for checking the integrity of the forecasting program. In the ridge regression, it is difficult to suppress the coefficients for less correlated predictors with the nonnegative coefficients  $\boldsymbol{\beta}$ . The wind power of the first wind farm is forecasted using the ridge regression in Figure 2.6.

### 2.4.2 Neural Network

The NN is introduced by focusing on a fast training algorithm, internal parameter settings, and future modifications. Detailed information about the NN is well summarized in [90]. In the implementation used here, the NN is trained by the scaled conjugate gradient (SCG) [135]. The SCG is the fastest CG, and it consumes less memory than other algorithms. If the NN is designed with internal parameters  $\boldsymbol{w}$  to forecast the observation  $y$ , the error function of the NN can be a function of  $\boldsymbol{w}$ . This error function can be approximated by a quadratic function through the Taylor expansion at the point  $\hat{\boldsymbol{w}}$  as

$$E(\hat{\boldsymbol{w}} + \boldsymbol{w}) \simeq E(\hat{\boldsymbol{w}}) + b^T \boldsymbol{w} + \frac{1}{2} (\boldsymbol{w})^T H \boldsymbol{w}. \quad (2.5)$$

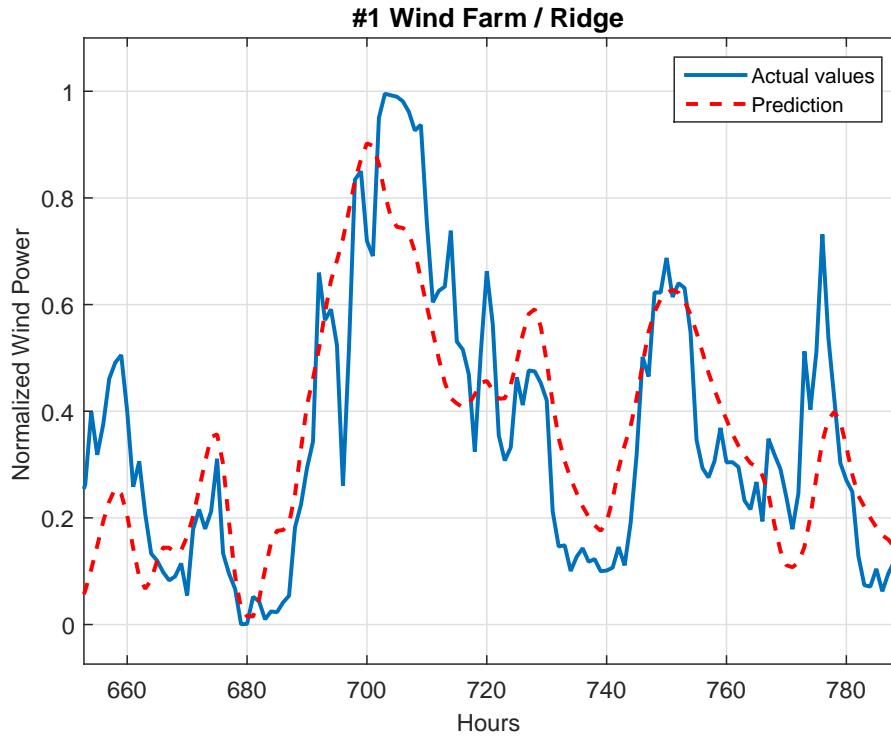


Figure 2.6: The red dotted line is the wind power predictions of the ridge regression, and the blue line is the actual wind power.

Then, the optimal  $\mathbf{w}$  is defined as

$$\mathbf{w}^* = H^{-1}\mathbf{b}. \quad (2.6)$$

However, estimating  $H$  is computationally burdensome, so  $\mathbf{w}$  is recursively estimated as

$$\mathbf{w}_{k+1} = \mathbf{w}_k + \alpha_k p_k. \quad (2.7)$$

In the CG, the step size  $\alpha_k$  is defined through a line search algorithm as

$$\alpha_k = -\frac{\mathbf{p}_k^T \mathbf{b}}{\mathbf{p}_k^T H \mathbf{p}_k}. \quad (2.8)$$

This formulation is computationally very burdensome because of the Hessian  $\mathbf{H}$ . On the contrary, in the SCG, since the scaling mechanism is used in estimating  $\alpha_k$ , the SCG avoids a time consuming line search algorithm, and it is faster than other CG based algorithms.

Three internal parameters in the NN are set to avoid over-fitting: the number of hidden neurons, the percentage of total data that is set aside as the validation data, and the number of validation data violations. First, the number of hidden neurons is strongly related to the number of parameters in the NN. In this simulation, the multiplication of the number of input features and the number of hidden neurons is fixed at 6,500. Furthermore, the percentage of total data that is set aside as validation data and the number of validation data violations are set to perform the early termination to avoid the over-fitting. Twenty percent of the training data is used as the validation data. When the weights in the neurons are updated, if the errors of the validation data increase continuously for 30 iterations, the training is stopped. Moreover, in order to avoid the local minimum, ten different NNs are estimated, and their results are averaged [124].

In future work, better training algorithms, such as the regularized NN and radial basis function NN can be used. If these training algorithms are simulated with combinations of different types of input parameters, and if additional types of parameters are introduced, the forecasting power can be increased.

### 2.4.3 Support Vector Machines

The kernel function can be used to change the feature space and the dimensionality of the feature space. The SVM and GP are kernel-based forecasting machines. The SVM for the regression minimizes the sum of the upper and lower deviations between the predictions and target values. The upper deviation  $\xi$  becomes positive when the target value is bigger than the prediction, and the lower deviation  $\zeta$  becomes positive when the target value is smaller than the prediction.

The SVM takes advantage of the sparsity by setting the upper and lower deviations in an insensitive region to zeros. The insensitive region is represented as the  $\epsilon$ -sensitive error function below.

$$E(y_n) = \begin{cases} 0 & \text{if } |y_n - t_n| < \epsilon \\ |y_n - t_n| - \epsilon & \text{otherwise} \end{cases}. \quad (2.9)$$

Targets on the outside of the insensitive regions are called support vectors, and they support targets inside the insensitive regions. In short, targets on the outside of the insensitive regions support the prediction so that the prediction is located inside the insensitive regions. This process can be represented as one minimization problem:

$$\underset{\xi_n, \zeta_n, \mathbf{W}}{\text{minimize}} \quad C \sum_{n=1}^N (\xi_n + \zeta_n) + \frac{1}{2} \|\mathbf{W}\| \quad (2.10)$$

$$\text{subject to} \quad y_n - \epsilon - \zeta_n \leq t_n \leq y_n + \epsilon + \xi_n, \quad (2.11)$$

$$0 \leq \zeta_n, \xi_n. \quad (2.12)$$

The quadratic optimization problem in (2.10) is solved by the sequential minimal optimization (SMO) algorithm [41]. In (2.10), the kernel function  $\phi$ , cost

of deviation  $C$ , tolerance  $v$  for optimization, and size of insensitive region  $\epsilon$  should be determined. In this application, the radial basis function (RBF) is used as a kernel function whose scale length is set as one over the number of input features.

$C$ ,  $v$ , and  $\epsilon$  are selected from many combinations of different values through the cross validation. After splitting the data into training and validation sets, the SVM is simulated with all possible combinations. Then, the best combination of minimum error for the validation set is selected. Generally,  $C = 0.2$ ,  $v = 0.0001$ ,  $\epsilon = 0.0001$  is the best combination. It should be noted that the trade off between the performance and computation time can be controlled by changing the iteration tolerance. The computation time to run one SVM problem is 5 seconds.

One advantages of the SVM is that it is very resistant to outliers with the help of support vectors and loss functions. Without detecting the outliers, the SVM achieves 37th, which outperforms other forecasting machines. Furthermore, its computation time is very short. However, the problem of the SVM is that it has many internal parameters that should be determined through the cross validation. Besides, the optimal kernel functions should be decided before training.

#### **2.4.4 Gaussian Process**

The GP also finds the coefficient  $\mathbf{w}$  in (2.1) using a kernel approach. In the GP, the prior distribution over  $\mathbf{w}$  is assumed to be Gaussian. Since  $\mathbf{y}$  is a

linear combination of  $\mathbf{w}$ ,  $\mathbf{y}$  is also Gaussian. Then, for  $N$  samples,  $\mathbf{y}$  becomes an  $N$  dimensional multivariate Gaussian distribution whose mean is zero, and whose covariance is a Gram matrix of  $\phi(\mathbf{x})$ . In short, the GP defines the probability of distribution over sample functions that pass near  $N$  observation points [162]. Therefore,  $\mathbf{y}$  follows the multivariate Gaussian distribution

$$\mathbf{y} \sim \mathcal{N}(0, \mathbf{K}), \quad (2.13)$$

where  $\mathbf{K}$  is a covariance matrix whose component is given as:

$$K(x_p, x_q) = \theta_1^2 \exp\left(-\frac{\|x_p - x_q\|^2}{2\theta_2^2}\right). \quad (2.14)$$

For the new observation  $\mathbf{x}^*$ , (2.13) can be represented as:

$$\begin{bmatrix} \mathbf{y} \\ y^* \end{bmatrix} \sim N\left(0, \begin{bmatrix} \mathbf{K} & \mathbf{K}^* \\ \mathbf{K}^* & k \end{bmatrix}\right), \quad (2.15)$$

where  $\mathbf{K}^*$  is the covariance vector between the training data and the predictors, so its element is represented as  $K(\mathbf{x}_n, \mathbf{x}^*)$  for  $n = 1, \dots, N$ . It should be noted that the  $\beta$  variance of measurement errors is not described here. Finally, if the conditional distribution of the multivariate Gaussian distribution is found [24], the mean of the prediction becomes

$$m(y^* | \mathbf{y}) = (\mathbf{K}^*)^T (\mathbf{K})^{-1} \mathbf{y}, \quad (2.16)$$

In order to avoid the over-fitting, internal parameters  $\theta_2$  and  $\theta_1$  in (2.14) are estimated by maximizing the log marginal likelihood

$$\log p(\mathbf{y} | \boldsymbol{\theta}) = -\frac{1}{2} \mathbf{y}^T \mathbf{K}^{-1} \mathbf{y} - \frac{1}{2} \log |\mathbf{K}| - \frac{N}{2} \log 2\pi, \quad (2.17)$$

where  $\theta$  includes  $\theta_2$  and  $\theta_1$ . This equation can be directly derived from (2.13). Since the computation time to maximize (2.17) is lengthy, 20% of the training data is randomly selected.

One advantage of the GP is that it can provide the probability distribution of the prediction. Furthermore, the GP is potentially the most customizable forecasting machine by selecting suitable kernels that represent the characteristics of the target signal. On the contrary, the GP also has many disadvantages. First, a large number of internal parameters that should be optimized depends on the kernel function, which should be selected before the training. Furthermore, optimizing the internal parameters is computationally burdensome. Although the GP is almost identical to the linear regression, its performance is better than the ridge regression because of its nonlinear kernels [161]. Therefore, the performance of the GP eventually depends on the selection of the kernels and parameters. In this study, the sum of the Gaussian, linear, and a quadratic ratio kernels is used. Thus, in order to make the GP a powerful tool, it is necessary to develop a method to address the model selection problem.

#### **2.4.5 Bootstrap Aggregation (BAG)**

The BAG, RF, and GBM forecasting machines are based on a regression tree, in which the training data is partitioned into multiple rectangles. Each rectangle corresponds to a node in a tree, and the sides of the rectangles, which are determined by splitting the variables and their values, correspond

to a junction in a tree. The trees are partitioned into two branches by a binary question at every junction [111]. A leaf, which is the terminal node, has a value that is the average of the training data corresponding to the leaf. The regression tree has a few advantages. *a)* the tree is built fast; *b)* the averaged trees can be resistive to outliers and irrelevant explanatory variables; and *c)* tree based machines are non-parametric, so the number of features can be bigger than that of observations. One disadvantage of a regression tree is that it suffers from high variance and over-fitting.

A better idea is to bootstrap the training data, to train trees with different subgroups of training data, and to average the results of the trees as

$$f(\mathbf{x}) = \frac{1}{B} \sum_{b=1}^B f_b(\mathbf{x}), \quad (2.18)$$

where  $B$  is the number of trees and  $f_b$  represents the tree. This idea is known as the bootstrapped aggregation (BAG). In the BAG, other machines can be used as weak learners, but the regression tree is the most widely used. Generally, a shallow tree, which has few branches, is preferred because it is resistant to over-fitting and is easy to interpret.

#### 2.4.6 Random Forest

A random forest is comprised of regression trees whose features are randomly selected and whose predictions are averaged to create one prediction for the forest. These randomly selected  $m$  features are different combinations of correlated and uncorrelated features and result in de-correlated trees. On



the contrary, in the BAG, all features are used to split branches, and then its predictions are driven by a few highly correlated features. However, averaging the predictions of de-correlated trees can increase the forecasting power and further reduce the variance of predictions at the expense of a small increase in bias. The key algorithms of the RF are introduced in [88].

There are three internal parameters in RF. They are  $m$ , node size, and the number of trees. The variable  $m$  is set as  $m = p/3$ , but the optimal  $m$  per wind farm is determined through the cross validation. The node size determines the number of leaves and hence the size of the tree. In this simulation, the node size is set as seven. The third parameter is the number of trees. Since the RF is not easily over-fitted, increasing this parameter will guarantee better performance. However, many trees will require high computation time. The number of trees is set between 200 and 500.

One of advantage of the RF is that it is not over-fitted. An additional advantage of the RF is that it can be used to decide the importance of the variables. The disadvantage of the RF is that it cannot forecast outputs beyond the range of the training data. Furthermore, the values of nodes are determined by averaging the outputs of the training data on the target nodes. This usually makes the forecasts under-predicted for bigger values and over-predicted for smaller values. It should be noted that the performance of the RF depends on the number correlated features and the  $m$  parameter. For example, when there are many features but few highly correlated features, the RF performs poorly with the small  $m$  because features that have strong forecasting power

are seldom selected.

### 2.4.7 Gradient Boosting Machines

The GBM builds a regression function  $\mathbf{f}$  by summing multiple weak learners. If regression trees are used, the GBM is also called the multiple additive regression trees (MART). Compared to other tree-based forecasting machines, the GBM does not change individual trees but builds  $\mathbf{f}_m$  recursively by adding different trees with different coefficients that are changed at every iteration. Generally, a shallow tree is used in the GBM. On the contrary, a deep and leafy tree is used in the BAG. The key idea of the GBM is to train the next tree using residuals between the target values and outputs of  $\mathbf{f}_{m-1}$ , so the GBM has a feedback system to improve the prediction power above samples that are not well trained by the  $\mathbf{f}_{m-1}$ . The prediction power increases by assigning bigger weight factors for untrained samples and assigning smaller weight factors for well-trained samples.

The ultimate goal of the GMB is to minimize the loss function  $L$ , which is the quadratic function in this application. In order to minimize the loss function,  $\mathbf{f}$  is updated by the steepest descent  $-\rho_m \mathbf{g}_m$ , where  $\mathbf{g}_m$  is the vector of gradients for  $\mathbf{f}_{m-1}$  at every input data, and  $\rho_m$  is the step length, which is estimated by solving

$$\rho_m = \arg \min L(\mathbf{f}_{m-1} - \rho \mathbf{g}_m). \quad (2.19)$$

Then, the  $\mathbf{f}$  is updated as

$$f_m = f_{m-1} - v \cdot \rho_m \mathbf{g}_m. \quad (2.20)$$

Unfortunately, the gradient is only calculated at values of training data, and the GBM is a non-parametric machine. Therefore, a continuous function is required to have similar prediction power for new data.

The trick in the GBM is to build a tree  $T_m$ , whose outputs are gradients, at every iteration so that new data corresponding to a certain node will have the same gradients of input features in that node.

At the  $m$ th iteration, the element of  $\mathbf{g}_m$  is given at  $n$ th input data as

$$g_{nm} = - \left[ \frac{dL(t_n, f(\mathbf{x}_n))}{df(\mathbf{x}_n)} \right]_{f=f_{m-1}}. \quad (2.21)$$

Then,  $T$  partitions input features into  $J$  nodes that have similar gradients. For the  $j$ th node, the steepest descent  $\gamma_{jm}$ , which corresponds to  $\rho_m \mathbf{g}_m$ , is estimated by minimizing the sum of the loss function for the training data in that node at  $\mathbf{f}_{m-1}$ . In short, optimal offsets of the  $\mathbf{f}_m$  outputs of the training data, whose gradients are similar, are estimated. The basic algorithm of the GBM is summarized in Algorithm 1. The detailed derivation of the equations used in this algorithm is introduced in [88].

Four internal parameters in the GBM, shrinkage  $v$ , the number of iterations  $m$ , the sampling rate, and the minimum node sizes, are controlled to avoid over-fitting. First, the GBM is not easily over-fitted when the  $v$  is low even with the big  $m$ . Therefore, a low  $v$  and high  $m$  is used in order to increase

---

**Algorithm 1** Gradient Tree Boosting

---

- 1: Initialize  $f_0(x) = \arg \min_{\gamma} \sum_N^{i=1} L(y_i, \gamma)$
- 2: **for**  $m = 1$  to  $M$  **do**

1. For  $i = 1, 2, \dots, N$ , compute

$$g_{im} = - \left[ \frac{dL(t_i, f(x_i))}{df(\mathbf{x}_i)} \right]_{f=f_{m-1}}.$$

2. Fit a regression tree to learn  $g_{im}$  and to give terminal regions  $R_{jm}$ ,  $j = 1, 2, \dots, J_m$ .
3. For  $j = 1, 2, \dots, J_m$ , compute

$$\gamma_{jm} = \arg \min_{\gamma} \sum_{x \in R_{jm}} L(y_i, f_{m-1}(x_i) + \gamma).$$

4. Update

$$f_m(\mathbf{x}) = f_{m-1}(\mathbf{x}) + v \cdot \sum_{j=1}^{J_m} \gamma_{jm} I(\mathbf{x} \in R_{jm}).$$

- 3: **end for**

- 4: Output  $\hat{f}(\mathbf{x}) = f_M(\mathbf{x})$ .
-

the forecasting power while avoiding the over-fitting. In this application, the  $v$  is set between 0.01 and 0.05, and the number of iterations is set between 2,000 and 4,000 depending on the quality of the input data. Bad quality data requires more iterations. It should be noted that the higher the  $v$ , the more computation time is required. Third, the sampling rate is set at 0.5. In each generation, only a small number of sampled training cases are used to train the tree. Using a fraction of the training data can reduce the computation time and improve the performance by reducing the variance [76]. Finally, the minimum node size of the regression tree is set at five. It should be noted that the early termination can be used through the cross validation in the training process, but it is not used in this application.

## **2.5 Ensemble Forecasting**

In this section, the performances of individual forecasting machines and the ensemble forecasting algorithms are introduced. Furthermore, the performances of simple averaging ensemble and weighted averaging ensemble are compared, and this leads to the performance comparison between a single forecasting machine and an ensemble of forecasting machines. It is worth noting that the results could differ according to different internal parameters.

### **2.5.1 Analysis of Individual Forecasting Machines**

The performances of the forecasting machines are compared by evaluating the prediction using the wind power from the first wind farm. The

training data of the second week in the competition is used in order to increase the computational speed. The performance of each forecasting machine is compared under the same simulation conditions, and the performance is shown in Table 2.1. The 528 features that are mentioned in above section are used, feature selections and data classification are not used, and the ratio of the number of outliers is 5%. In addition, five-fold cross validation is used to measure the performance of each forecasting machine. It should be noted that the performance difference coming from the different configurations of the training data will not be discussed in this dissertation.

The GBM shows the best performance of all the configurations of the training data. The second best is the BAG. Furthermore, the NN, RF, and SVM perform similarly, but the NN and RF are better than the SVM. As mentioned above, since the performance of the RF is affected by the number of correlated variables in the feature space, having the better  $m$  can improve the performance. The worst performing forecasting machine is the ridge regression. Among the three best models, the GBM requires the most computation time, which is around two hours. On the contrary, the ridge regression is the fastest machine.

Forecasting machines can be further researched. The ridge regression strongly depends on the length of training data; thus the ten-fold internal cross validation is used to select the optimal regularization coefficient in order to maintain the data length. Furthermore, the performance of the ridge regression does not vary with respect to the configurations of the training data. In the

Table 2.1: Performance of forecasting machines. The wind power from the first wind farm is used.

Number	Method	Computation Time (s)	RMSE
1	Ridge	67	0.1593
2	SVM	244	0.1561
3	NN	136	0.1546
4	RF	189	0.1553
5	GBM	821	0.1537
6	GP	887	0.1600
7	BAG	301	0.1535
7	Simple Averaging	0.1556	
7	Weighted Averaging		0.1554

GP, the best kernel is the sum of the Gaussian isotropic, linear, and constant kernels. The GP is very sensitive to the number of features. Furthermore, the SVM and NN are resistant to outliers. However, the NN is easily over-fitted without the early termination.

### 2.5.2 Ensemble Algorithms

Ensemble algorithms can be classified into four methods. The first one is the BAG algorithm, in which many weak learners are trained by different subgroups of randomly selected training data. Then, predictions from many weak learners are simply averaged. The second method is the boosting. As the number of iterations increases, the next learner is trained by focusing on

the cases that were not learned well by the previous learner. Generally, the regression tree is used as a weak learner in the first and second ensemble methods. Third, the predictions are simply averaged. Fourth, the predictions are averaged after multiplying weight factors that are the inverse of their error rates. In this dissertation, the third and fourth method are compared since the first method is implemented in the BAG, and since the second method is implemented in the GBM. The performance of ensemble algorithms is also compared in Table 2.1. The performance of weighted averaging ensemble is better than the performance of simple averaging ensemble. Moreover, it is clearly shown that the performance of the combination of multiple forecasting machines is always better than the performance of a single forecasting machine.

In addition, the forecasting power can be increased by using the  $k$ -fold cross validation, where  $k$  is bigger than five, on multiple combinations, but the computation time increases exponentially, so it is difficult to forecast the hourly wind power for real-time application. Therefore, users should consider the trade-off between the weighted averaging method that uses the few best forecasting machines and the simple averaging method that uses all possible forecasting machines.

## 2.6 Quantile Estimation

In this section, error distributions for every point forecast are estimated and represented as 1% quantiles under the assumption that the error distribution follows a parametric distribution in a closed form. The performance of



the error distribution, which is calculated by the pinball loss function, is measured with respect to four different distributions: uniform, Gaussian, Laplace, and approximated VG distribution. In order to narrow the width of the error distribution, different error distributions per cluster of point forecast or the continuous function between the point forecast and the STD of the error distribution can be estimated. The performances of these different methods are compared. In all methods, the error distribution per cluster is truncated by the minimum and maximum of actual values of each cluster.

### 2.6.1 Pinball Loss Function

The pinball loss function is defined as

$$L(q_b, y) = \begin{cases} (1 - \frac{b}{100}) \times (q_b - y) & \text{if } y < q_b \\ \frac{b}{100} \times (y - q_b) & \text{if } y \geq q_b \end{cases}, \quad (2.22)$$

where  $n$  represents a quantile percentage between 1 and 99,  $q_b$  represents the quantile prediction, and  $y$  represents the target value. The quantile value  $q_b$  represents the value of the random variable when its accumulated probability occupies  $b$  percentage. The Gaussian distribution and its quantile function are plotted in Figure 2.7. The (2.22) tries to set the actual value at the mean of error distribution. For the over-forecasted quantile, which corresponds to the upper equation,  $b$ , which is greater than 50, can minimize the (2.22). On the contrary, for the under-forecasted quantile, which corresponds to the lower equation,  $b$ , which is less than 50, can minimize the (2.22). In short, when  $b$  is greater than 50, the over-forecasted quantile has a lower loss, and when  $b$  is less than 50, the under-forecasted quantile has a lower loss. Consequently,

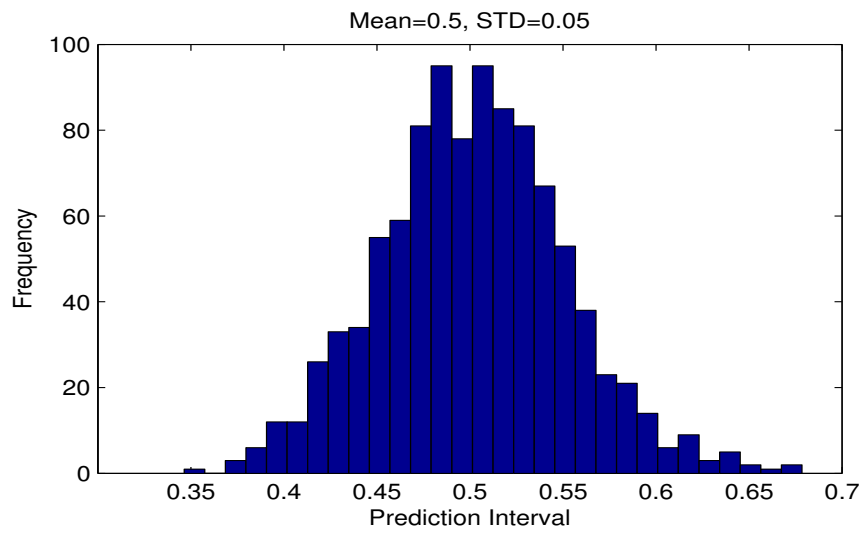
slightly bigger quantiles have better performance with respect to the testing data.

### **2.6.2 Probabilistic Forecasting in the Ensemble Forecasting**

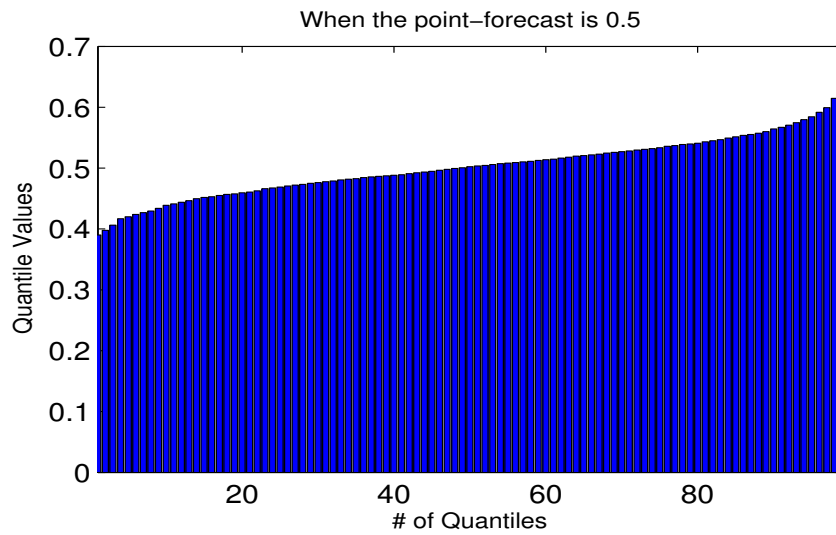
Forecasting machines that are based on Bayesian statistics or neural networks generally provide the distribution of predicted values. However, in the ensemble forecasting, predictions of multiple forecasting machines are averaged, it is hard to find the meaning and usage of error distribution that is estimated based on Bayesian statistics or neural networks. Instead of using error distribution that is estimated based on the structure of the individual forecasting machine, estimating the error distribution through the comparison between the point forecasts and actual values can be estimated by taking advantage of high forecasting performance of the ensemble forecasting.

### **2.6.3 Parametric Approach**

The error distribution of the point forecast can be estimated by finding the STD of the error distribution, which follows the parametric distribution in a closed form. This is called the parametric approach. In this dissertation, the candidates of the parametric distribution are the uniform, Gaussian, Laplace, and approximated VG distributions. The performance of these distributions are compared, and it is shown that the approximated VG distribution has the minimum loss function.



(a)



(b)

Figure 2.7: The Gaussian distribution and its quantile function are plotted. (a) Gaussian distribution (b) Quantiles function of the Gaussian distribution

#### 2.6.4 Distributions

The characteristics of the distributions mentioned in the previous subsection are explained in this subsection. First, the probability distribution function of the uniform distribution is given as

$$f(x) = \begin{cases} \frac{1}{b-a} & a \leq x \leq b \\ 0 & \text{otherwise} \end{cases}, \quad (2.23)$$

The variance of the uniform distribution is

$$V[X] = \frac{1}{12}(b - a)^2 \quad (2.24)$$

and the mean is given as

$$\mathbb{E}[X] = \frac{1}{2}(a + b). \quad (2.25)$$

The mean is the same as the point forecast, so

$$\hat{y} = \frac{1}{2}(a + b). \quad (2.26)$$

For the given  $V[X]$  and  $\hat{y}$ , the  $a$  and  $b$  in (2.23) can be estimated by using (2.26) and (2.24). Second, the PDF of the Gaussian distribution is

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad (2.27)$$

where  $\mu$  is the mean of the Laplace distribution. Suppose the mean is the same as the point forecast  $\mu = \hat{y}$ . Then, for the given STD, the distribution and its quantiles can also be estimated. Third, the PDF of the Laplace distribution is given as

$$f(x) = \frac{1}{2b} \left( -\frac{|x - \mu|}{b} \right), \quad (2.28)$$

where  $\mu$  is the mean of the Laplace distribution. The variance of the Laplace distribution is given as

$$V[X] = 2b^2. \quad (2.29)$$

For the given STD, the  $b$  can be estimated, and the distribution and its quantiles can also be estimated.

The PDF of the VG distribution, which is introduced in [169], is described here as

$$f(x) = \frac{(\alpha^2 - \beta^2)^\lambda}{\sqrt{\pi}\Gamma(\lambda)(2\alpha)^{\lambda-0.5}} |x - \mu|^{\lambda-0.5} K_{\lambda-0.5}(\alpha |x - \mu|) \exp(\beta(x - \mu)), \quad (2.30)$$

where  $K(z)$  is the modified Bessel function of the second kind. Its mean is given as

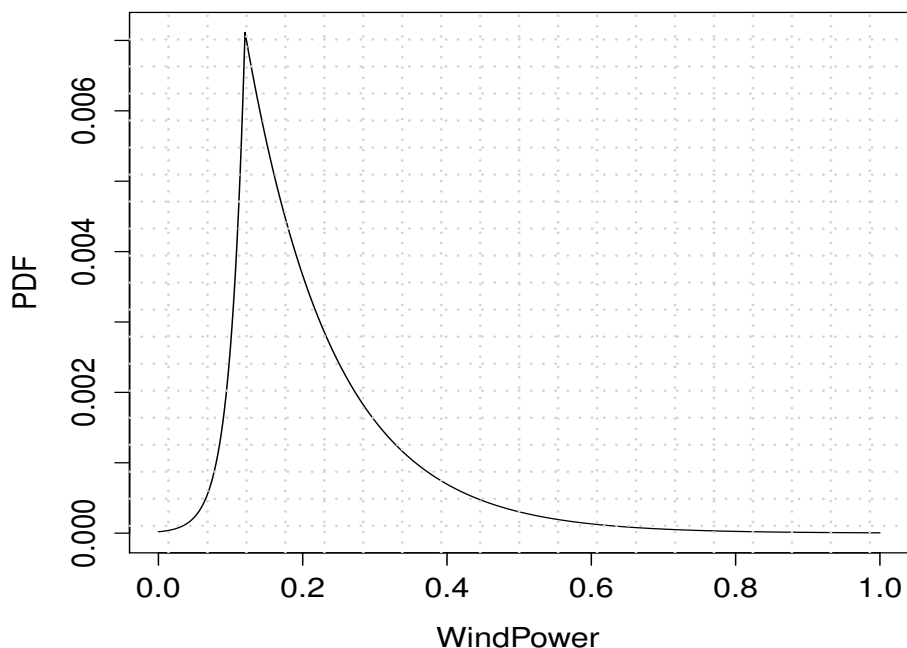
$$\mu + \frac{2\beta\lambda}{\alpha^2 - \beta^2}, \quad (2.31)$$

and the variance is given as

$$\frac{2\lambda}{\alpha^2 - \beta^2} \left( 1 + \frac{2\beta^2}{\alpha^2 - \beta^2} \right). \quad (2.32)$$

$\mu$  determines the location of the median,  $\alpha$  determines the length of tails,  $\beta$  determines the asymmetry, and  $\lambda$  determines the scale. As  $\lambda$  increases, the distribution becomes more similar to the Gaussian distribution, and as  $\lambda$  decreases, the distribution becomes more similar to the Laplace distribution. As  $\alpha$  decreases, the distribution has a fatter tail, and as  $\beta$  increases, it becomes more skewed. The VG distribution whose  $\mu$  is 0.12,  $\alpha$  is 0.07,  $\beta$  is 0.1, and  $\lambda$  is 1, is shown in Figure 2.8.

## Density of the Variance Gamma Distribution



*Figure 2.8: The VG distribution is shown. The approximated VG distribution can be represented as the combination of two exponential distributions with different standard deviations.*

Since the VG distribution is designed through multiple internal parameters, it is hard to find the optimal combination of multiple internal parameters, which minimize the pinball loss function. Therefore, in this dissertation, the VG distribution is approximated with two exponential distributions with different STDs, which is called the asymmetric Laplace distribution [108]. Then, the STD of each exponential distribution needs to be estimated. Furthermore, without loss of generality, the  $\mu$  is set at the prediction value. However, it should be noted that the asymmetric Laplace distribution does not exactly

match to summation of two exponential distributions with different STDs. Since each exponential distribution with a different STD covers 0.5 probability, the asymmetric Laplace distribution cannot be fully implemented by adding two exponential distributions with different STDs. Although the quantiles of two different exponential distributions are attached at the mean value in this dissertation, building the asymmetric Laplace distribution by attaching two different exponential distributions with different STDs should be addressed in detail in the future work.

### **2.6.5 Estimation of the Standard Deviations: Discrete Cluster Case**

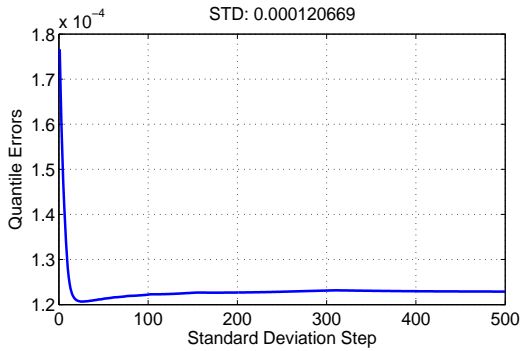
For each exponential function, the optimal STD of the exponential function, which minimize the sum of total pinball losses, is found by checking various STDs from a small value to a large STD with a fixed step. For the given candidate STD, the sum of total pinball losses was measured by comparing the point forecast to the actual values. It should be emphasized again that the point forecast for each actual value is estimated through the external cross validation. By comparing the sum of total losses, the optimal STD for each side is estimated. The sum of total pinball losses is plotted against different candidate STDs in Figure 2.9. In order to plot in a limited space, among ten clusters, only six clusters are plotted. It is clearly observed that the STD of the first cluster, whose point forecasts are less than 0.1, is smaller than the STDs of the second, third, fourth, and fifth clusters, whose point forecasts are between 0.1 and 0.6. Furthermore, the STD of the third cluster, whose point

forecasts are between 0.2 and 0.3, has the biggest STD. Therefore, the STDs of error distributions are different with respect to the level of point forecasts. The loss graph is a convex function, so an advanced optimization algorithm, such as the binary search, could be used to find the optimal STD in a faster time. This should also be addressed in detail in the future work.

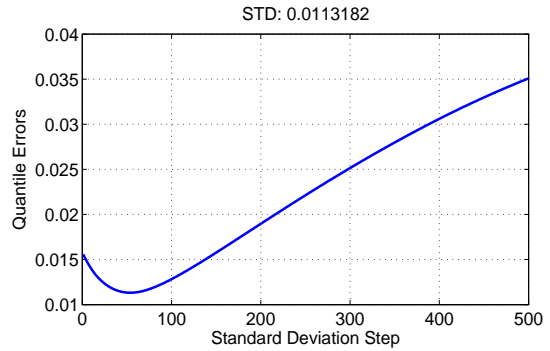
### **2.6.6 Point Forecast Clustering**

In order to narrow the ranges of the error distribution, the error distribution is estimated for each cluster of point forecasts. If a point forecast is confined, the actual value of the point forecast might be confined. The point forecasts are classified with respect to their own values, and the overall cluster process is well described in Figure 2.10. For example, suppose that the point forecasts are classified into ten clusters. The point forecasts in the first cluster are between zero and 0.1, and the point forecasts in the second cluster are between 0.1 and 0.2. Then, the maximum and minimum of actual values of point forecasts in the same cluster can be estimated, and they will be used to truncate error distribution since actual values might be confined if the point forecast is confined. Therefore, each group has its own error distribution, whose STDs are estimated by using point forecasts in the same cluster. If the error distribution is assumed to follow the uniform, Gaussian, and Laplace distributions, a single STD is estimated, but if the error distribution is assumed to follow the approximated VG distribution, two STDs are estimated. In this dissertation, one, five, ten, and 20 clusters are tested. The process of data

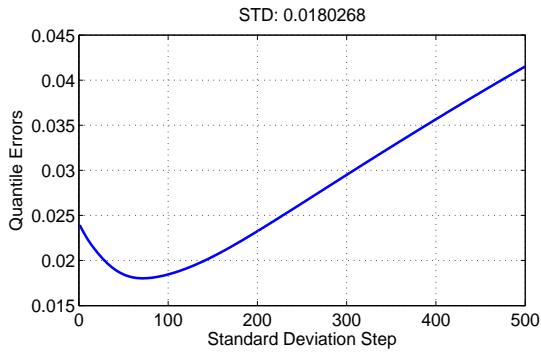




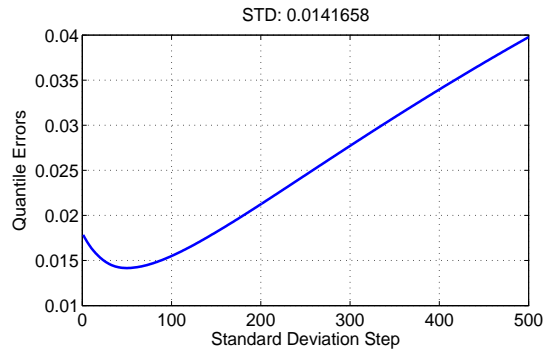
(a)



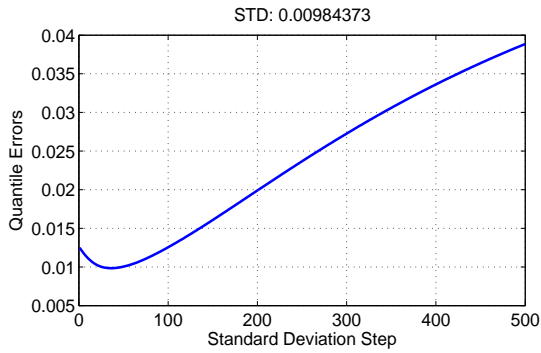
(b)



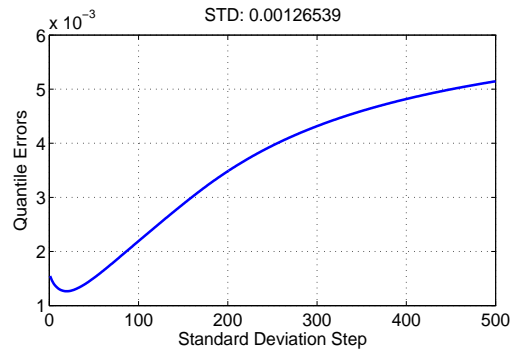
(c)



(d)



(e)



(f)

Figure 2.9: Quantile errors with respect to the STDs for different groups

clustering is briefly shown in Figure 2.10.

### **2.6.7 Truncating Distribution**

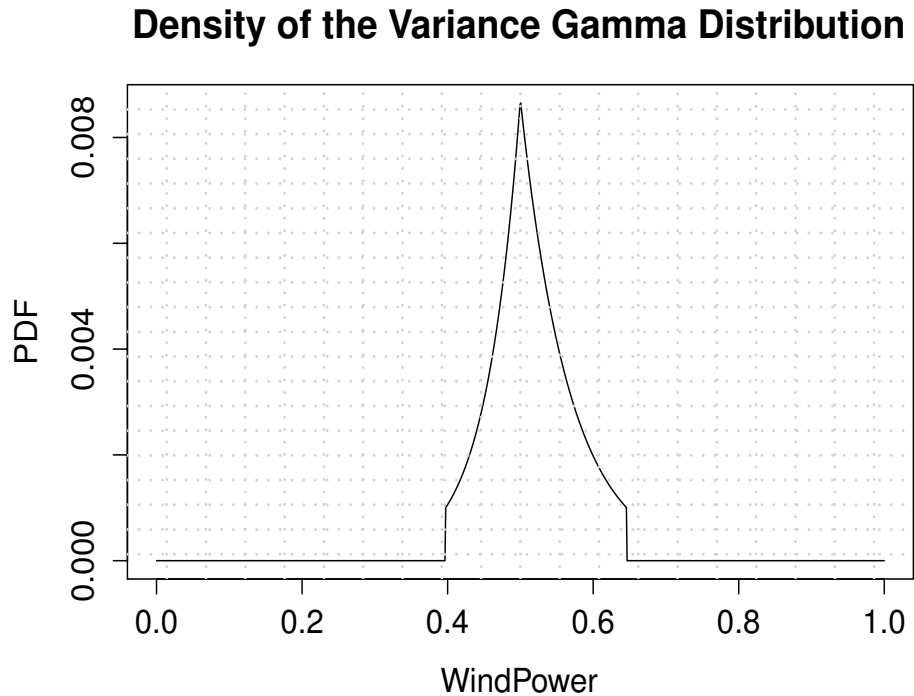
The maximum and minimum values of actual data in each cluster are measured as shown in Figure 2.10, and these values are used to truncate error distributions. For example, suppose that the actual values of point forecasts in the third cluster are 0.25, 0.27, and 0.29. The error distribution of the point forecast in the third cluster is limited by 0.25 and 0.29. Consequently, quantiles of error distributions of the point forecast in the third cluster are truncated by 0.25 and 0.29. A truncated and approximated VG distribution is plotted in Figure 2.11.

### **2.6.8 Estimation of the Standard Deviations: Continuous Case**

The optimal STD of the given point forecast can also be estimated by a continuous function. This function is trained to find the connectivity between the point forecasts and the optimal STDs, which are estimated by comparing the point forecast and its actual value. Then, the continuous function can provide the optimal STD of the point forecast without data clustering. The training data for the continuous function is the weather information and the point forecast of the validation data; the target data for the continuous function is the optimal STD of point forecast of validation data. All forecasting units can be used to train this function, but the ridge regression shows the best performance. Since the ridge regression provides the averaged STD, the



Figure 2.10: The point forecasts are clustered into ten clusters with respect to their power levels.



*Figure 2.11: A truncated VG distribution is shown.*

noise in the test data can be mitigated. The data clustering is also used in this function in a limited way. For the given point forecast, its error distribution is truncated by the minimum and maximum of the actual values in the prediction’s corresponding cluster.

### 2.6.9 Simulation Results

The performances of probabilistic forecasting algorithms are measured in Table 2.2. The algorithms for estimating quantiles can be classified with respect to the distribution and the number of clusters. The performances of the

Table 2.2: Performance of the quantile estimation methods. The performances of the five best teams in the competition is 0.03942, 0.04040, 0.04102, 0.04133, and 0.04494.

	Distribution	1	5	10	20	Continuous
1	Uniform	0.0408	0.0402	0.401	0.0400	0.0423
2	Gaussian	0.0403	0.0398	0.0397	0.0396	0.0421
3	Laplace	0.0401	0.0397	0.0396	0.0396	0.0417
4	VG	0.0400	0.0398	0.0398	0.0398	0.0406

five best teams in the competition is 0.03942, 0.04040, 0.04102, 0.04133, and 0.04494. Therefore, the performance would have been second if the quantiles had been estimated under the assumption that the error distribution follows the approximated VG distributions with ten data clusters.

The approximated VG distribution outperforms other distributions. Furthermore, it is clear that the grouping with more than ten clusters does not improve the performance. One interesting result is that the error distributions with data clustering outperforms the continuous STD estimation function between the optimal STD and point forecast. The problem of the continuous function is that it is not trained to respond to the external noise in the test data. The STD should be able to respond to the ranges of noises in the target value, but the optimal STD, which is the training data for the continuous function, is determined by only the given point forecast. On the contrary, in order to determine the range of the error distributions, multiple STDs of point forecasts should be searched. In short, the performance of the

continuous STD estimation function is not better than the performance of the error distributions with data clustering. Besides, the continuous function takes more time to estimate than the discrete STD estimation function with data clustering does.

## 2.7 Conclusions

In this dissertation, monthly future wind power outputs on ten wind farms and their error distributions are predicted. First, point forecasts of wind power are estimated through the ensemble forecasting. In order to increase the number of forecasting units, multiple combinations of different settings, internal parameters, and forecasting machines are used. Second, by comparing the point forecast and actual values, the optimal STDs of error distributions are estimated under the assumption that the error distribution follows one of four distributions: a uniform, Gaussian, Laplace, or approximated VG distribution. In order to narrow the ranges of error distributions, an error distribution is estimated for each cluster of point forecast. The point forecasts are clustered with respect to their power level in one, five, ten, or 20 clusters. Then, error distributions are truncated by the minimum and maximum of the actual values in the prediction's corresponding cluster.

Various forecasting machines have different advantages and disadvantages. The ridge regression has the lowest computation time, and its forecasting power is not far behind other forecasting machines. The tree-based forecasting machines, such as the RF, GBM, and BAG, outperform other fore-

casting machines, but they take more time to forecast than other forecasting machines do. By considering all aspects, the ridge regression, GBM, RF, SVM, and NN are the best five forecasting machines. In conclusion, the quantiles of the approximated VG distribution outperform other distributions. In this process, the performance of error distributions with ten data clusters is the best. The continuous STD estimation function does not outperform the discrete STD estimation function with data clustering since it is not designed to find the proper width of quantiles, which should be wide enough to cover measurement errors. The optimal STD of each point forecast results from observing a single training case.

In field applications, the simple ridge model might be sufficient to forecast wind power since it is fast and accurate. However, in order to use this model in a real time simulation, a few things should be updated. First, the weight factor should be multiplied to the forecasted predictors since the quality of predictors deteriorates as the forecast horizon increases. Second, the program architecture should be recursively changed to learn new training data and discard old data. Third, depending on the sampling period of wind power, the computation time should be considered. By trading off between the computation time and the performance of the forecasting machine, the prediction should be finished within the sampling period.

In conclusion, as the first step and at a shorter timescale to increase the penetration level of wind power, short-term probabilistic and ensemble wind power forecasting algorithm is proposed in this chapter to increase the perfor-

mance of wind power forecasting. Probabilistic wind power forecasting can be used to analyze the price risk and bid the wind power so that the wind farm owners bid more wind power in the AS and DA markets in order to maximize their profits. As the second step and at a longer timescale, sample paths of total wind power are synthesized. Scenarios can be used to simulate the power system so that the power system planners can determine the proper amounts of regulation service, storage size, and incentive signals for EVs. Toward this end, the next chapter discusses total wind power scenario generation.



## Chapter 3

### Total Future Wind Power Scenario Generation

This chapter describes the synthesis of the future total wind power scenarios with respect to the future installed capacity of wind power. These scenarios can be used to forecast future wind production trends, design new AS procurement methodologies, estimate the required amounts of ASs, and determine the incentives for EV owners. First, the wind power is normalized by its capacity, and the 24h daily cycle is subtracted. Then, the PSD of the normalized wind power is estimated by using the multitaper algorithm. The estimated PSDs are approximated as six piecewise affine functions through the modified hinges model. Approximated PSDs are used to design training data that will be used to analyze the wind power fluctuation and the trends of the PSDs. Training data is generated by successively omitting wind farms from a pool of all 70 wind farms in the 2010 ERCOT interconnection. As each wind farm is omitted, for the given number of wind farms, 100 cases of different combinations of wind farms are generated. Consequently, wind power outputs for almost 6,000 combinations of farms and their PSDs are generated. Simultaneously, the statistical characteristics are measured and regressed for the target installed capacity of wind power. Then, the statistical characteristics and wind power fluctuation are analyzed with respect to the geographical

smoothing effects in the frequency domain. Then, the power law in the PSD is analyzed with respect to frequencies as wind capacity increases. Finally, the future PSD is forecasted by forecasting the slopes of the affine functions. By converting the forecasted PSD through the inverse DFT, three wind power scenarios are synthesized to verify the suggested scenario generation method in this dissertation. First, a wind power scenario in April 2010 is synthesized using actual wind power in April 2010. Second, a wind power scenario in April 2010 is synthesized using actual wind power in April 2009. Synthesized scenarios are compared to the actual wind power in 2010. Third, the future wind power at 10,000 MW wind power capacity is synthesized. In this process, the genetic algorithm (GA) is used to find phase angles that make the distribution of wind power variability follow the empirically observed Laplace distribution.

It should be noted that “fluctuation” is used to represent the general variation of wind power and that “variability” is used as a fluctuation index to represent the wind power fluctuation quantitatively. The definition of wind power variability in [195] is adopted, where variability is defined as the standard deviation of the difference between wind power outputs over successive one minute intervals.

### **3.1 Introduction**

To simulate the power system with increased wind capacity, much of the literature has generated current or future total wind power scenarios by considering the current or future wind power variability. These scenarios have

been used in stochastic optimization problems for power system operations and planning.

### 3.1.1 Literature Review

Much of the literature has modeled the wind power variability and synthesized wind power scenario by considering these following models of wind power variability [188]. First, short-term scenarios have been modeled via a stochastic process based on the Monte Carlo simulation [132]. For example, Tuohy et al. [192] generated wind power and load scenarios through Monte Carlo simulations of the forecast errors through a time series analysis. However, formulations in time series analysis can only model the characteristics of a short-term period. The spinning reserve requirement is estimated in [27,145], under the assumption that both the wind power and its prediction errors follow the normal distribution. However, wind power is not normally distributed, and prediction errors have a heavy tail [78]. Furthermore, scenarios in the far future are typically forecasted by multiplying empirical time series by a constant. In this process, the variability of the scenario can be modeled via the Markov Chain based on the patterns of the forecasting errors of past data [32]. However, wind power forecasting based on the meso-scale model with a coarse geographical and temporal resolution does not provide the fluctuation within a short time, and these methods do not consider the variability of future aggregated wind farms distributed geographically [110].

The future fluctuation of wind power is also important to estimate the

required amounts of regulation and other ancillary services due to the increased wind power. Currently, the ERCOT protocols specify that ERCOT must procure regulation capacity equal to 98.8 percent of the net load (wind-load) fluctuation, and in order to consider increased wind capacity, ERCOT added additional amounts that are specified in the table in [62] to the regulation capacity. The historical net load forecast error is used to calculate the non-spinning reserve requirements, which should be enough to cover 95 percent of the forecast error in the historical net load. This process is a kind of brute force method, and the fluctuation is not modeled quantitatively [188]. Therefore, there is a need to accurately synthesize wind power scenarios for future wind power considering the variability of wind power.

Not only wind power scenarios, but also wind fluctuation and wind power fluctuation have been analyzed in the frequency domain by focusing on finding the power law in the PSD of wind and wind power. Power laws have been observed in the spectral density in the frequency domain in several applications. For example,  $1/f$  frequency scaling of spectral density has been observed in electroencephalograms of brain signals [20]. A power law has also been observed in wind power. Nanahara et al. in [141] observed the smoothing effects of wind power from a large number of wind turbines through the PSD analysis and analyzed theoretically the reduction of power level by aggregating distant wind farms. Moreover, they analyzed the cross spectral density between two wind speed outputs in the far distance and found that the cross spectral density of wind speed has a power law that depends on

the distance between measurement points. The power law in the PSD of wind power has also been analyzed on the wind turbine levels. Joaquin and Angel in [140] researched the PSD slope changes with respect to the number of interconnected wind turbines. When  $N$  wind turbines are interconnected, they also found that the distribution of PSD values of uncorrelated wind turbines has a  $\sqrt{N}$  times smaller standard deviation than the distribution of PSD values of correlated wind turbines. They also found that the distribution of PSD magnitudes follows the Rayleigh distribution.

Apt [8] estimated the PSD of wind power and found that the slope of a line fitted to the PSD in a logarithmic plot follows a Kolmogorov slope  $-5/3$  over a period of 30 seconds to 2.6 days [107]. The Kolmogorov slope represents the ratio of the energy dissipation of wind turbulence with respect to the turbulence size. However, Apt's research was limited to analyzing the slope of the PSD using a few wind turbines and wind farms, making it difficult to fully characterize the PSD of the total wind power. Furthermore, Katzenstein [104] discovered that the PSD slope becomes more negative as *a)* correlation coefficients corresponding to the 15 minutes period increase, *b)* the number of interconnected wind farms increases, and *c)* the ratio between the capacity of wind farms increases. However, in order to fully follow the trend of the PSD slope, the number of samples should be increased: the 400 or fewer samples used in [8, 104] is an insufficient amount.

Summarizing, a new algorithm to synthesize the total wind power scenario should model the 24h daily cycle, the short-term fluctuation, and long-

term fluctuations. In order to develop the new algorithm, wind power scenarios are generated based on the PSD analysis. Furthermore, by using the PSD analysis, the variability can be analyzed quantitatively through the magnitudes of coefficients in the PSD.

### **3.1.2 Preprocessing**

Wind power from all wind farms in the ERCOT interconnection is used to verify the approach. Wind power for 2009 and 2010 is sampled at every minute. In 2009, 73 wind farms were in ERCOT, and in 2010, 81 wind farms were installed in ERCOT. However, the data measured only for a year is not able to represent variability over multiple years, so wind power outputs with different capacities are generated by adding actual wind power outputs from randomly selected farms. The wind power output in the training data is normalized by its capacity, which is the sum of the capacities of corresponding wind farms. Furthermore, under the assumption that there is only one deterministic signal for a 24-hour period, the 24h daily cycle of normalized wind power is calculated by averaging daily wind power at a given minute over a month. Then, wind power is deseasonalized by subtracting this 24h daily cycle from the normalized wind power. The 24h daily cycle are estimated from the monthly wind power data that is sampled at every minute. A 24h daily cycle is estimated by averaging wind power for a given minute in the day over all days in the month. Therefore, there are 1,440 means that are calculated for each month of data. A more concrete explanation follows for a month with 30

days. The vector of monthly wind power  $X \in \mathbb{R}^{24 \cdot 60 \cdot 30}$  is reshaped into the matrix  $Y \in \mathbb{R}^{24 \cdot 60 \times 30}$ . The minute-by-minute mean of  $Y$  is defined as  $Z \in \mathbb{R}^{24 \cdot 60}$  which is averaged over 30 days. The future 24h daily cycle is generated under the assumption that the 24h daily cycle follows the sinusoidal waveform. The amplitude and mean of the future 24h daily cycle are regressed at the target installed capacity of wind power, and the synthesized future 24h daily cycle is added to the synthesized normalized waveform.

Statistical information of each wind power output in the training data is also measured from the actual wind power. Statistical characteristics include the variability, standard deviation (STD), maximum wind power, minimum wind power, mean, monthly ramp size, monthly ramp rate, amplitude of the daily cycle, and mean of the daily cycle. The 24h daily cycle is assumed to be a sinusoidal waveform having forecasted amplitude and mean, and it is added to the synthesized waveform.

### 3.2 Power Spectral Density Estimation

Spectral density is the spectrum of power of periodic signals with respect to frequencies. When a discrete signal is represented as a superposition of the sinusoidal components of the DFT, spectral density represents the square of amplitudes of the sinusoidal components. The PSD of zero-mean stationary sequence  $x[n]$  can be estimated by the DFT of the autocorrelation function as

defined in [57]:

$$P(e^{j\omega}) = \sum_{q=-\infty}^{\infty} r[q] \exp[-j\omega q], \quad (3.1)$$

where  $r[q]$  is the sample autocorrelation sequence of  $x[n]$ . For a finite sequence, it can be redefined as

$$\hat{r}[q] = \begin{cases} \frac{1}{N} \sum_{n=0}^{N-q-1} x[n+q] x[n] & 0 \leq q \leq N-1 \\ \hat{r}[-q] & -(N-1) \leq q \leq 0 \\ 0 & \text{elsewhere,} \end{cases} \quad (3.2)$$

where  $N$  is the data length. Then, the PSD of the finite duration sequence  $x[n]$  can be estimated as

$$\hat{P}(e^{j\omega}) = \sum_{q=-(N-1)}^{N-1} \hat{r}[q] \exp[-j\omega q]. \quad (3.3)$$

### 3.2.1 Periodogram

The PSD in (3.3) can be directly estimated by  $x[n]$ . Since the autocorrelation sequence  $\hat{r}$  is the convolution of  $x[n]$  [89], the PSD can also be estimated by squaring the magnitude of the DFT of the signal; this is called the periodogram. The equation (3.2) can be rewritten as

$$\hat{r}[q] = \frac{1}{N} \sum_{n=-\infty}^{\infty} x[n+q] x[n] = \frac{1}{N} x[q] * x[-q], \quad (3.4)$$

under the assumption that

$$x[n] = \begin{cases} x[n] & 0 \leq n \leq N-1 \\ 0 & \text{elsewhere.} \end{cases} \quad (3.5)$$

Then, (3.3) can be rewritten as

$$\hat{P}(e^{j\omega}) = \frac{1}{N} X(e^{j\omega}) X(e^{-j\omega}) = \frac{1}{N} |X(e^{j\omega})|^2, \quad (3.6)$$



where

$$X(e^{j\omega}) = \sum_{n=0}^{N-1} x[n] \exp[-j\omega n]. \quad (3.7)$$

In order to reduce leakage and increase the resolution, windows are used as

$$x_{\omega}[n] = x[n] w[n], \quad (3.8)$$

where  $w$  is the window. Furthermore, when angular frequency  $\omega$  is discretized as

$$\omega_k = \frac{2\pi k}{N}, \quad (3.9)$$

where  $0 < k < N - 1$ ; then, the PSD of the finite-duration sequence can be represented as the square of magnitude of the DFT, which is a sequence corresponding to equally sampled frequencies.

$$\hat{P}(f_m) = \frac{1}{NF_s} \left| \sum_{n=0}^{N-1} x_{\omega}[n] \exp\left(-j2\pi n \frac{m}{N}\right) \right|^2. \quad (3.10)$$

Since frequency is also discretized as

$$f_m = m \frac{F_s}{N}, \quad (3.11)$$

where  $F_s$  is the sampling frequency, (3.10) can also be represented as

$$\hat{P}(f_m) = \frac{1}{NF_s} \left| \sum_{n=0}^{N-1} x_{\omega}[n] \exp\left(-j2\pi n \frac{f_m}{F_s}\right) \right|^2. \quad (3.12)$$

The periodogram is an asymptotically unbiased estimator of the PSD. However, since the variance of the periodogram is not reduced as the length

of the data sequence increases, the periodogram is not a consistent estimation [57]. There are some algorithms to reduce the variance and bias of estimated PSDs. For example, Welch's algorithm can reduce the variance of the PSD by partitioning a signal into equal segments and averaging PSDs of each segment. However, in the next section, the multitaper algorithm is used instead because of its superior performance.

### **3.2.2 Multitaper Algorithm**

The multitaper algorithm is used to estimate the PSD of wind power output in order to maintain information in the low frequency ranges and reduce variance [191]. This contrasts with several previous approaches to estimating the PSD of wind power. For example, Apt [8] used Welch's algorithm to reduce the variance and bias of the PSD. However, even though Welch's method can reduce the variance of the estimate, since it partitions a signal into equal segments with windows having the same size and form, it loses spectrum information on the low frequency range. Since low frequency issues are also important, it is necessary to preserve low frequency information. Furthermore, Katzenstein [104] used the Lomb periodogram [123] to compensate for irregularly sampled data. However, it is hard to convert the PSD from the Lomb periodogram into a signal since it is based on the least squares of sinusoids. In contrast, the multitaper can maintain the signal information of the low frequency while reducing the variance at the same time.

The multitaper estimates multiple periodograms using different orthog-

onal tapers (which correspond to windows in Welch’s algorithm) on the full length data, and then averages periodograms to find the multitaper PSD. If properly designed orthogonal windows known as tapers are used, spectral density estimates at every frequency become uncorrelated to each other [191]. Averaging the uncorrelated PSD estimates using those tapers can reduce the variance, bias, and power leakage [190]. In particular, the variance of uncorrelated spectral densities is reduced by a factor  $1/K$ , where  $K$  is the number of tapers. Since the full length of each window is used, the multitaper algorithm can maintain the signal information in the low frequency ranges. Tapers have different frequency responses with various mainlobe widths  $W$  with respect to the taper number  $k$ . For the given  $k$ , the mainlobe bandwidth  $W$  is given as

$$W = \frac{F_s}{2N}(k + 1), \quad (3.13)$$

where  $F_s$  is the sampling frequency, and  $N$  is the data length. A wider mainlobe can help to suppress sidelobes and reduce the power leakage, but its resolution is lower. It is worth noting that a higher taper number will smooth out the PSD when averaging periodograms because of the lower resolution. In this application,  $k = 9$  tapers are used. Furthermore, tapers are orthogonal, so they can generate independent periodograms at all frequency ranges, so averaging them reduces variance while keeping low frequency information.

Thomson in [191] suggested discrete prolate spheroidal sequences (DPSS) as tapers. In spite of the absence of a closed form, the  $k$ th taper  $w_k$  is defined

in [190] as

$$w_k[n] = \frac{1}{\lambda_k} \sum_{m=0}^{N-1} \frac{\sin 2\pi W(n-m)}{\pi(n-m)} w_k(m), \quad (3.14)$$

where each  $\lambda_k$  is an eigenvalue of a matrix  $\Lambda$ . The eigenvalue  $\lambda_k$  can be calculated from the matrix  $\Lambda$ , which is saved as a table. The  $\Lambda$  is defined in [82] as

$$\{\Lambda(N, W)\}_{i,j} = \begin{cases} \frac{1}{2}i(N-i) & \text{if } j = i-1 \\ \left(\frac{N-1}{2} - i\right)^2 \cos(2\pi W) & \text{if } j = 1 \\ \frac{1}{2}(i+1)(N-1-i) & \text{if } j = i+1 \\ 0 & \text{otherwise,} \end{cases} \quad (3.15)$$

where  $i, j = 0, \dots, N-1$ . The eigenvalues can be estimated using the inverse iteration. Further detail is explained in [82]. Furthermore, different tapers are orthogonal, so that:

$$\sum_{n=0}^{N-1} w_k[n] w_l[n] = \begin{cases} 1, & \text{if } k = l \\ 0, & \text{if } k \neq l, \end{cases} \quad (3.16)$$

Then, the PSD of the discrete wind power sequence  $x_w[n]$  in (3.8) is estimated in [89] as

$$\hat{P}_k(f_m) = \frac{1}{NF_s} \left| \sum_{n=0}^{N-1} x_w[n] \exp\left(-j2\pi n \frac{f_m}{F_s}\right) \right|^2. \quad (3.17)$$

Furthermore, the  $x_k[n]$  is the product of sequence  $x[n]$  and a  $k$ th taper  $w_k[n]$  and defined as

$$x_k[n] = x[n] w_k[n]. \quad (3.18)$$

Samples outside of the finite duration sequence are set to be zero. With  $K$  tapers, the multitaper PSD is calculated by averaging (3.17) over the  $K$  tapers

and is defined in [57] as

$$\hat{P}_{MT}(f) = \frac{1}{K} \sum_{k=1}^K \hat{P}_k(f). \quad (3.19)$$

The characteristics of the multitaper PSD, which have non-smoothed waveforms and less variation in the low frequency range, are shown in Figure 3.1 for a particular wind farm. According to (3.19), the unit of the PSD should be  $\text{Unit}^2/\text{Hz}$ . Since wind power is a “signal” in this application, and its unit is MW, the unit of the PSD becomes  $\text{MW}^2/\text{Hz}$ . For the normalized signal,  $(\text{Normalized Power})^2/\text{Hz}$  is used. It is assumed that wind power data is wide-sense stationary and ergodic in the mean, since the autocovariance of wind power converges to zero quickly with increasing data length and depends only on the time difference.

### 3.3 Piecewise Modeling of the PSD

The modified hinges model based on the original hinges model in [167] is used to approximate the PSD with six piecewise affine functions on the logarithmic-scaled axes.

#### 3.3.1 Hinges Model

The hinge  $H_n(k_n, h_n)$  is the point where the slope in the frequency domain changes angle. The value  $k_n$ , which is the frequency-axis location of hinge  $H_n$  is called a knot and is associated with the frequency interval of the discrete Fourier transform (DFT). The value  $h_n$ , which is the spectral-

density axis location of hinge  $H_n$ , is the spectrum density associated with a given knot  $k_n$ , and it is calculated to minimize the MSE compared to the PSD of two linear pieces between hinges  $H_{n-1}$  and  $H_n$  and between hinges  $H_n$  and  $H_{n+1}$  respectively. Since the power law in the PSD in Figure 3.1 can explain the hidden connection between wind power fluctuation and frequencies, power laws are represented as various slopes of affine functions, and the PSD can be approximated with those affine functions. Seven hinges including the first and last hinges (with fixed frequency locations of  $32^{-1}\text{day}^{-1}$  and  $2^{-1}\text{m}^{-1}$ , respectively) are used. The original hinges model determines the number of hinges and the frequency-axis location of hinges in every PSD. For the given number and frequency-axis location of hinges, the modified hinges model fits every PSD as six piecewise affine functions by finding the spectral-density axis location of hinges by solving an unconstrained quadratic problem.

### 3.3.2 Original Hinges Model

The original hinges model consists of four steps.

1) **Smoothing Step:** The PSD is smoothed by the moving average in order to reduce the PSD variation.

2) **Growing Step:** All samples are considered as hinges. For  $N$  samples, there are  $N - 1$  piecewise functions and  $N$  hinges.

3) **Pruning Step:** In the pruning step, a hinge is deleted one by one from the first one to the last one. For every deleted hinge, the total mean square errors (MSE) between the hinges model and actual data is measured.

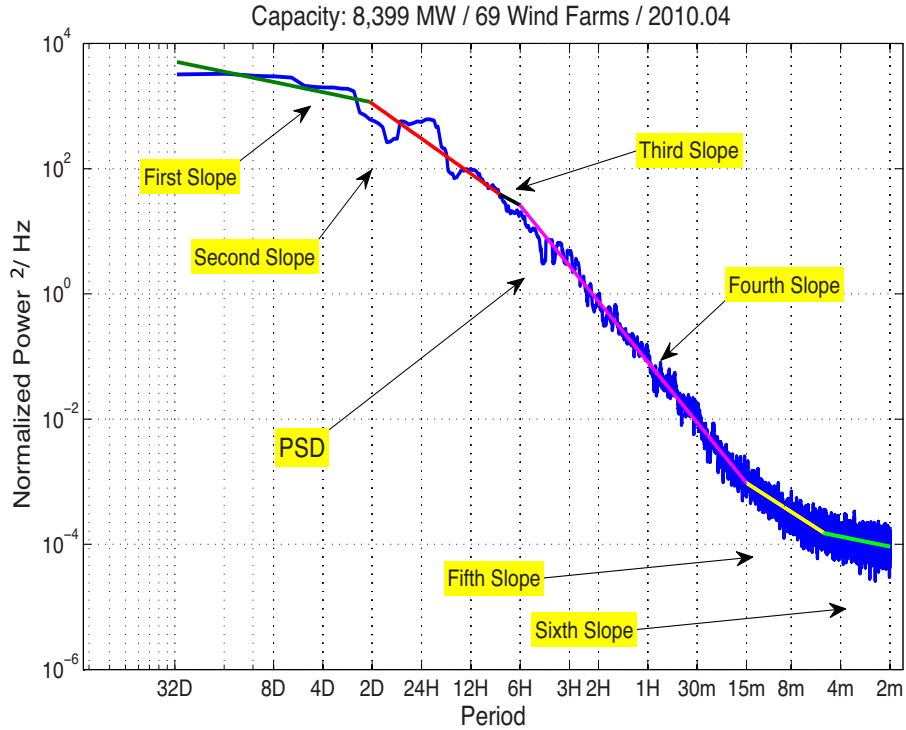


Figure 3.1: PSD of wind power and six slopes fitted to the PSD are plotted.

In this process, the hinge that increases the total mean square errors (MSE) minimally is deleted until the first and last hinges remain. After this process is finished, the number of hinges is set to be the number of hinges when the MSE increases drastically as hinges are deleted. The more hinges, the better the model fits, but many hinges will over-fit the data, so the number of hinges should be limited to 10 to reduce the computational complexity.

4) **Refitting Step:** Frequency-axis and spectral-axis location of the hinges are recursively found. For a given PSD, the spectral-density axis location  $h_n$  is determined to minimize the MSE between the fitted line and

scatter points, which are located between two adjacent hinges. Let's assume that there are three consecutive hinges  $H_{m-1}(k_{m-1}, h_{m-1})$ ,  $H_m(k_m, h_m)$ , and  $H_{m+1}(k_{m+1}, h_{m+1})$ . For the  $m$ th hinge  $H_m(k_m, h_m)$ ,  $k_m$  is the frequency-axis location of the hinge and  $h_m$  is the spectral density-axis location of the hinge with respect to  $k_m$ . Suppose that the location of this hinge needs to be decided. First, locations of  $(m-1)$ th and  $(m+1)$ th hinges are fixed. Second, the original hinges model searches the  $k_m$  among samples between  $(m-1)$ th hinge and  $(m+1)$ th hinge so that it minimizes the MSE. Then,  $h_i$  is calculated by solving an unconstrained quadratic optimization to minimize the MSE of the PSD and samples between hinges  $H_{m-1}(k_{m-1}, h_{m-1})$  and  $H_{m+1}(k_{m+1}, h_{m+1})$ . The MSE of the fitted line among three hinges is defined as

$$\begin{aligned} \text{MSE} = & \sum_{i=1}^{N_l} [(y_{l,i} - h_{n-1}) + a_{l,i} (h_n - h_{n-1})]^2 \\ & + \sum_{i=1}^{N_r} [(y_{r,i} - h_{n+1}) + a_{r,i} (h_n - h_{n+1})]^2, \end{aligned} \quad (3.20)$$

where

$$\begin{aligned} a_{l,i} &= -\frac{x_{l,i} - k_{n-1}}{k_n - k_{n-1}}, \\ a_{r,i} &= -\frac{x_{r,i} - k_{n+1}}{k_n - k_{n+1}}. \end{aligned} \quad (3.21)$$

The  $x_{l,i} \in X_l$  for

$$X_l = \{x_i | k_{n-1} < x_i < k_n, \forall i = 1, \dots, N_l\} \quad (3.22)$$

are the horizontal frequency axis locations of points between hinges  $H_{n-1}$  and  $H_n$ , where  $N_l$  is the number of points, and  $y_{l,i}$  are their vertical axis locations.



The  $x_{r,i} \in X_r$  for

$$X_r = \{x_i | k_n < x_i < k_{n+1}, \forall i = 1, \dots, N_r\} \quad (3.23)$$

are the horizontal axis locations of points between hinges  $H_n$  and  $H_{n+1}$ , where  $N_r$  is the number of points, and  $y_{r,i}$  are their vertical axis locations.

The spectral density-axis location  $h_m$  for a fixed frequency-axis location  $k_m$  is calculated as

$$h_m = - \frac{\sum_{n=1}^{N_{l_m}} a_{l,n} b_{l,n} + \sum_{n=1}^{N_{r_m}} a_{r,n} b_{r,n}}{\sum_{n=1}^{N_{l_m}} a_{l,n}^2 + \sum_{n=1}^{N_{r_m}} a_{r,n}^2} \quad (3.24)$$

where

$$b_{l,n} = \left( \frac{x_{l,n} - k_{m-1}}{k_m - k_{m-1}} \right) h_{i-1} - h_{m-1} + y_{l,n}, \quad (3.25)$$

$$b_{r,n} = \left( \frac{x_{r,n} - k_{m+1}}{k_m - k_{m+1}} \right) h_{i+1} - h_{m+1} + y_{r,n}. \quad (3.26)$$

In (3.21),  $l_m$  represents the PSD values on the left side of hinge  $H_m$  up to the previous hinge  $H_{m-1}$ , and  $r_m$  represents the PSD values on the right side of hinge  $H_m$  up to the next hinge  $H_{m+1}$ . The refitting step is repeated for all hinges by fixing two adjacent hinges until the difference between the previous and present total MSE becomes less than 0.001. It is worth noting that although each PSD could have a different optimal number of hinges, the number of hinge is fixed at the mean of optimal number of hinges, which is seven including the first and last hinges.

### 3.3.3 Modified Hinges Model

In order to compare slopes of all PSDs, the number and frequency-axis location of hinges should be fixed to be able to compare slopes for given frequency-axis locations. To unify the number of hinges over all PSDs, the means of every frequency-axis location of hinges for all PSDs are set as the frequency-axis location of hinges. Then, frequency-axis locations are adjusted to well reflect the trends of the PSD. Using the fixed number and frequency-axis location of hinges, every PSD as six affine functions is approximated. These frequency-axis values are held constant and used for the rest of the dissertation. They are shown in Table 3.1. For the fixed number and frequency-axis location of hinges, the spectral density-axis location can be estimated by solving the unconstrained quadratic problem below:

$$\min_{h_m} \sum_{m=1}^{M-1} \sum_{n=1}^{N_m} \left( h_{m+1} + \frac{h_m - h_{m+1}}{k_m - k_{m+1}} (x_n^m - k_{m+1}) - y_n^m \right)^2, \quad (3.27)$$

where  $m$  is the tag of data between hinges,  $M$  is the number of hinges, and  $N_m$  is the number of samples between the  $m$ th hinge and  $(m + 1)$ th hinge. In addition,  $h_m$  is the spectral density-axis location of the  $m$ th hinge. Furthermore,  $x_n^m$  is the  $x$ -axis location of  $n$ th data between the  $m$ th hinge and  $(m + 1)$ th hinge, and  $y_n^m$  is the spectral density axis location of the  $n$ th hinge between the  $m$ th hinge and  $(m + 1)$ th hinge. In Figure 3.1, the line on the log scaled axis is indicative of a power-law.

The hinges model in this section is used to approximate the PSD. In order to estimate the PSDs of future wind power, the trends of slopes of affine

Table 3.1: Frequency-axis hinge locations

Hinges	Frequency (Hz)	Period	Hinges	Frequency (Hz)	Period
1st	3.6169e-7	32 days	2nd	5.7870e-6	2 days
3rd	3.4722e-5	8 hours	4th	4.6296e-5	6 hours
5th	0.0011	15 min	6th	0.0033	5 min
7th	0.0083	2 min			

functions in the hinges model should be observed for the PSDs of various wind power outputs with different installed capacity of wind power. The next section introduces how various wind power outputs are generated with respect to different installed capacity.

### 3.4 Training Data Generation

Training data provides trends of the statistical characteristics of wind power as the total wind capacity increases, and trends are used to forecast the statistics of future wind power scenarios that are used to estimate the required amounts of future AS. Increasing penetration level of wind power might change the wind power fluctuation because of the geographical smoothing effects, so it might affect the amounts of regulation services. Therefore, tracking the changes of the statistical characteristics of wind power is very important for synthesizing future wind power scenarios. The basic idea is to omit wind farms from a pool of all wind farms in the 2010 ERCOT interconnection. Wind farms are clustered based on their similar fluctuations through a hierarchical cluster

algorithm, and wind farms are omitted from clusters. The exclusion order is decided by the capacity size and cluster tag. The number of wind farms that should be omitted is equally allocated to clusters. If the number of omitted wind farms is not a multiple of the number of clusters, additional wind farms are omitted from clusters with smaller tags. The larger the number of wind farms in a cluster, the smaller the given cluster tag. When wind farms are omitted from a certain cluster, the smaller the capacity size, the earlier it is subtracted. Seventy wind farms are used to generate training data. For each number of wind farms, 100 cases of wind farm configurations are used to reduce the computational complexity. Among the 100 cases, wind farms are selected to maximize the geographical smoothing effects for the given number of wind farms, since geographical smoothing effects increase as the capacity increases. When 100 combinations are made, the number of wind farms to be omitted increases.

This clustering will maintain the correlations between wind power fluctuation in low frequency ranges but does not affect the wind power fluctuation in high frequency ranges since all wind farms have different fluctuations for high frequency signals. Similar patterns of wind power fluctuation are measured by factor loadings estimated through factor analysis. Suppose that there are a few air masses in Texas and that they are sources of the main wind streams. Since wind moves with a high inertia [87], if local weather events are neglected, It can be assumed that the speed and direction of wind streams do not change much over a long period of time. If wind streams sweep a group

of wind farms simultaneously, those wind farms generate similar wind power. The movement of wind streams is not observed directly through the data, but their existence can be inferred by the co-movements of wind power data. Therefore, wind streams can be represented as time-varying factors. Wind farms distributed in Texas will be affected by those streams to varying degrees according to their location relative to wind streams. In addition, factor loadings are fixed coefficients of factors and represent geographical characteristics [164]. The key ideas of factor analysis are introduced in the subsequent subsections.

### 3.4.1 Factor Analysis

The emphasis in this subsection is on the definitions of factor analysis and estimation of factors and factor loadings. Suppose the observation data  $\mathbf{x} \in \mathbb{R}^{N \times T}$  have an  $N$  cross-sectional dimension and  $T$  time dimension. Each data in  $\mathbf{X}$  can be represented as

$$\mathbf{X} = \{x_{it} | i = 1, \dots, N, t = 1, \dots, T\} \quad (3.28)$$

The  $\mathbf{x}$  can be partitioned into common components  $\boldsymbol{\chi} \in \mathbb{R}^{N \times T}$  and idiosyncratic noise components  $\mathbf{E} \in \mathbb{R}^{N \times T}$ . At the given time  $t$ , (3.28) can be rewritten as

$$\mathbf{X}_t = \boldsymbol{\chi}_t + \mathbf{E}_t, \quad (3.29)$$

where  $\mathbf{X}_t = (x_{1t}, \dots, x_{Nt})'$ ,  $\boldsymbol{\chi}_t = (\chi_{1t}, \dots, \chi_{Nt})'$ , and  $\mathbf{E}_t = (e_{1t}, \dots, e_{Nt})'$ . If  $\boldsymbol{\chi}_t$  is represented by a linear combination of factor loadings  $\boldsymbol{\Lambda} \in \mathbb{R}^{N \times P}$  and

factors  $\mathbf{F} \in \mathbb{R}^{P \times T}$ , (3.29) becomes

$$\mathbf{X}_t = \mathbf{\Lambda} \times \mathbf{F}_t + \mathbf{E}_t, \quad (3.30)$$

where  $\mathbf{F}_t = (f_{1t}, \dots, f_{Pt})'$ , and where  $P$  is the number of factors.

Factor analysis accounts for the covariance matrix of observation data through the specific variance and factor loadings. The auto covariance of  $\mathbf{X}$  is given as

$$\text{cov}(\mathbf{X}) = \text{cov}(\mathbf{\Lambda}\mathbf{F} + \mathbf{E}) = \text{cov}(\mathbf{\Lambda}\mathbf{F}) + \text{cov}(\mathbf{E}) \quad (3.31)$$

$$= \mathbf{\Lambda} \text{cov}(\mathbf{F})\mathbf{\Lambda}' + \mathbf{\Psi} = \mathbf{\Lambda}\mathbf{\Lambda}' + \mathbf{\Psi}, \quad (3.32)$$

where  $\mathbf{\Psi} = \text{cov}(\mathbf{E})$ , and  $\text{cov}(\mathbf{F}) = \mathbf{I}_P$ .

The factor loadings capture the covariances between observation variables, and the specific variance  $\mathbf{\Psi}$  captures the variance of independent noise for each variable [24]. Therefore, the multiplication of the factor loadings and their transpose explains most of the covariance of observation data.

Factors, factor loadings, and specific variance can be estimated through the principal factor method iteratively. At the  $n$ th iteration, the approximated covariance matrix is given as

$$\mathbf{\Sigma}_n = \mathbf{\Lambda}_n \mathbf{\Lambda}_n' + \mathbf{\Psi}_n. \quad (3.33)$$

Suppose that the approximated covariance matrix  $\mathbf{\Sigma}_n$  is the same as the covariance matrix  $\mathbf{S}$ . Then, the multiplication of the loading factor and its transpose is defined as

$$\mathbf{\Lambda}_n \mathbf{\Lambda}_n' = \mathbf{S} - \mathbf{\Psi}_n. \quad (3.34)$$

Since the estimation starts from the initial  $\Psi_n$ ,  $\Psi_{n+1}$  should be redefined as

$$\Psi_{n+1} = \text{diag}(\mathbf{S} - \mathbf{\Lambda}_n \mathbf{\Lambda}_n'), \quad (3.35)$$

where  $\text{diag}$  represents the diagonal matrix. Factor loading is estimated from (3.34) using the eigenvectors.

In (3.33), an infinite number of sets of factor loadings is possible, since multiplying an orthogonal matrix by a factor loading  $\mathbf{\Lambda}$  does not change the covariance matrix. Suppose that the  $\mathbf{\Lambda}$  is the un-rotated factor loading. Then, the rotated factor loading  $\mathbf{\Lambda}^*$  by the transformation matrix  $\mathbf{T}$  is given by

$$\mathbf{\Lambda}^* = \mathbf{\Lambda} \mathbf{T}. \quad (3.36)$$

Factor loadings are rotated so that they have the maximum variance of squared loadings. This will make them distinctive, so it would be easy to cluster wind farms based on factor loadings. The promax rotation [102] is used to maximize the variance among factor loadings. Since squared factor loadings lie between 0 and 1, pushing them into one of these two boundary numbers will maximize the variance of squared factor loadings [116]. Then, elements of rotated factor loadings are exponentiated by a coefficient, which is generally four, to make them become more bipolar. Factors that are estimated through the Factor Analysis in this subsection are clustered through the Cluster Analysis that are explained in the subsequent subsection.

### 3.4.2 Cluster Analysis

The basic concept of the agglomerative hierarchical cluster analysis is that it finds the two closest lower clusters and combines them as new upper clusters. First, the pair with the least dissimilarity comprises the first cluster. Then, the dissimilarities between clusters is calculated. In this process, the distance between clusters are called the linkage, which is calculated from the dissimilarities between pairs of all observation points, so the clustering algorithm is called the “agglomerative” algorithm [96]. The dissimilarity calculates the distance between factor loadings, and the linkage calculates the distance between clusters. In this dissertation, the dissimilarity is defined as the Minkowski distance of order  $p$ , which is measured by

$$d(\mathbf{x}, \mathbf{y}) = \left( \sum_{j=1}^P |x_j - y_j|^p \right)^{1/p}, \quad (3.37)$$

where  $P$  is the number of factors, and where  $\mathbf{x}$  and  $\mathbf{y}$  are a vector of the factor loading of each wind farm. In addition, the linkage is defined as the minimum distance between observation  $\mathbf{x}$  in one cluster  $X$  and the other cluster  $Y$

$$D(X, Y) = \min\{d(\mathbf{x}, \mathbf{y}) ; \mathbf{x} \in X \text{ and } \mathbf{y} \in Y\} \quad (3.38)$$

Two clusters having minimum single linkages are merged and become a new cluster.

## 3.5 Wind Power Ramp Modeling

In order to reflect the long-term wind power fluctuation in synthesized waveforms, the statistical information about ramp events is analyzed and fore-



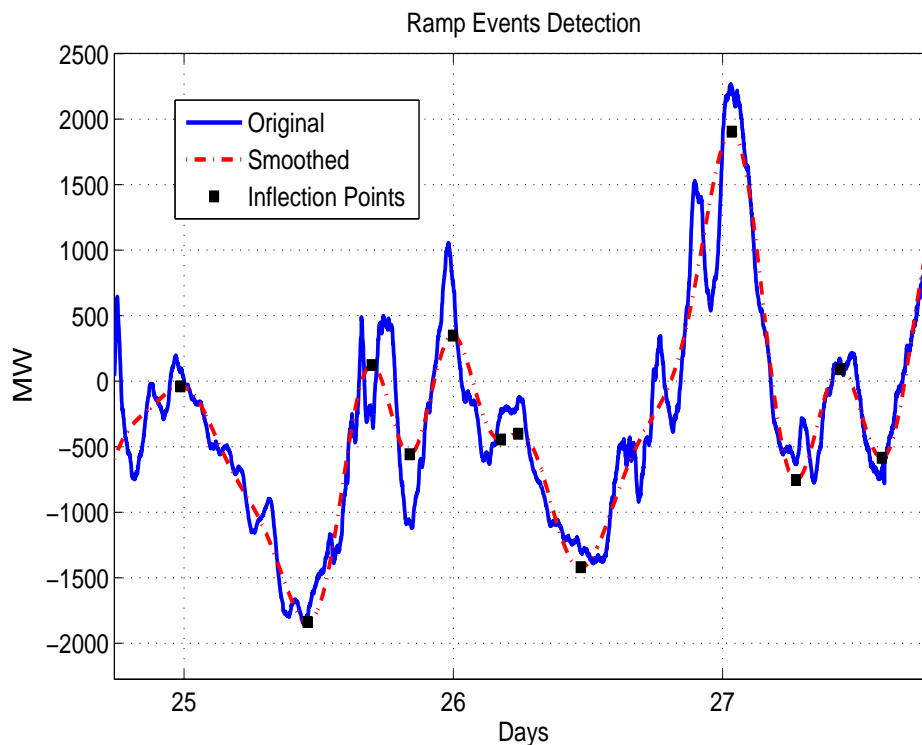


Figure 3.2: Wind power and its smoothed wind power are plotted. Red dots represent the inflection point where the direction of ramps changes.

casted. In [51], a ramp event was defined as wind power fluctuation whose amplitude is bigger than 75% of the capacity and whose duration is longer than three hours. In this application, a ramp event is defined as changes between two adjacent inflection points of smoothed wind power. Wind power is smoothed by passing through a low-pass filter, following [25]. The cut-off frequency is set as three hours. The filtered wind power and its inflection points are shown in Figure 3.2. Ramp size is defined as the power difference between inflection points, ramp duration is defined as the time difference be-

tween inflection points, and the ramp rate is defined as the ramp size over the ramp duration. By following the definition of the ramp event in [51], only ramp events whose ramp duration is longer than three hours are selected. The STDs of ramp size and ramp rate are measured and, and wind power scenarios that are satisfying the STDs are found through the GA. Therefore, in the wind power synthesizing process, signal components in the low frequency are modeled by analyzing ramp events, signal components in the middle frequency are modeled through the PSD analysis, and signal components in the high frequency are considered through the variability analysis.

### **3.6 Analysis of Slopes**

In this section, the wind power fluctuation is analyzed with respect to the changes of overall PSD slopes. Then, the slopes of affine functions are analyzed and regressed at the target wind capacity.

#### **3.6.1 Wind Power Fluctuation**

In this section, the fluctuation of un-normalized wind power with respect to the geographical distribution of wind farms is analyzed in the frequency domain. The variability of un-normalized wind power is well-fitted to a quadratic function with a small negative quadratic coefficient as shown in Figure 3.3. This can be interpreted to mean that the fluctuation of wind power with respect to the wind capacity decreases relatively as the wind capacity increases.

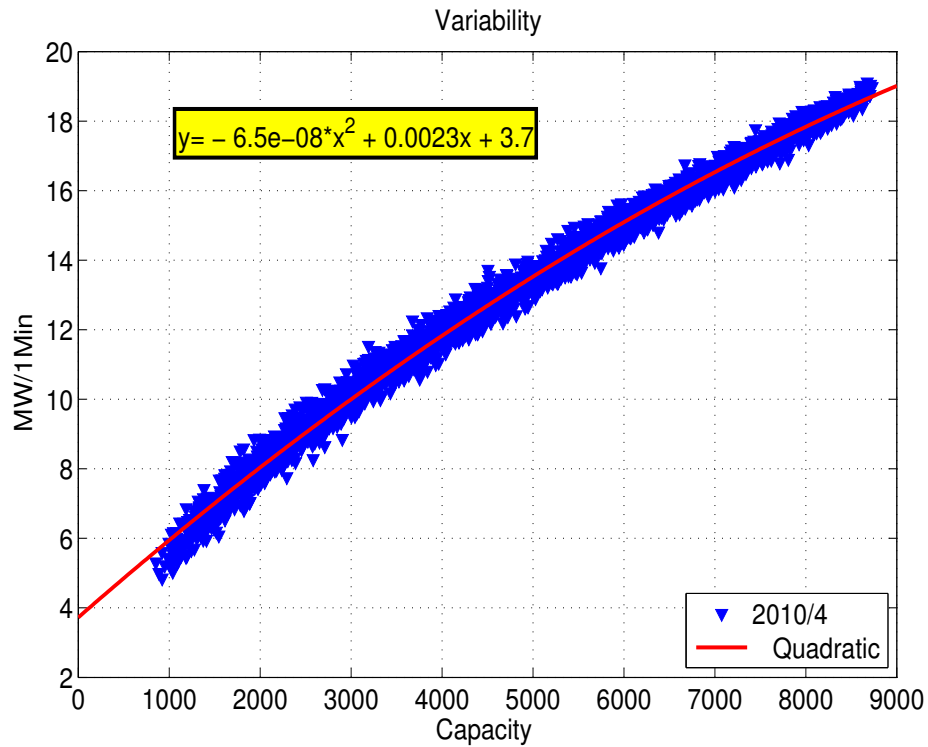
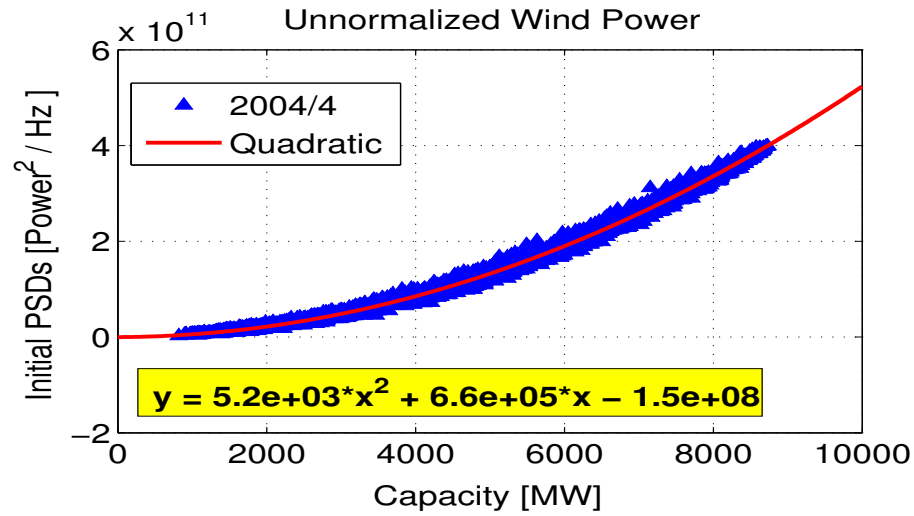
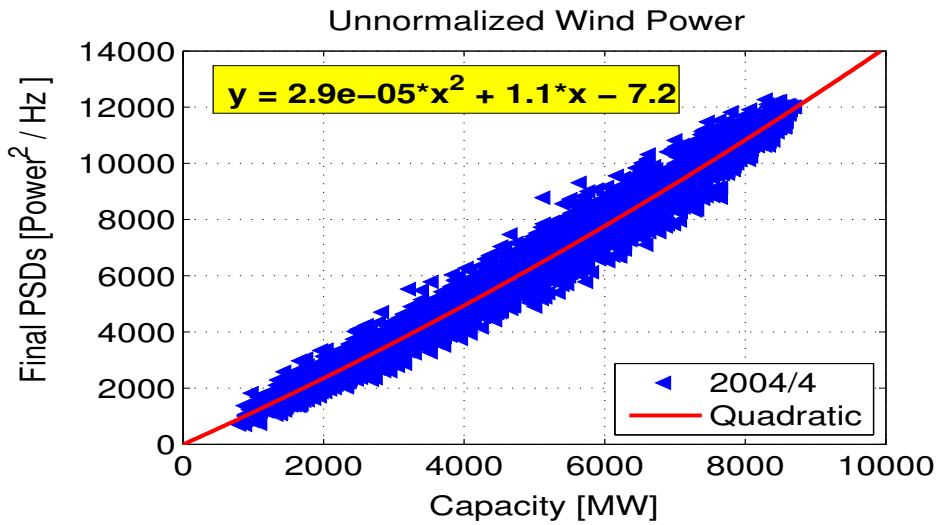


Figure 3.3: The variability of un-normalized wind power as the wind power capacity increases is plotted. It follows a square root function.

Wind farms in a larger area as well as wind farms in a smaller area are under the same effects of long-term weather phenomena, such as fronts, convective outflow, and air masses. Wind farms under these weather phenomena show similar wind power patterns in the low frequency since those phenomena last a long period of time. As wind farms are integrated, sinusoidal components of the DFT of wind power in the low frequency are highly positively superposed with similar phase angles since long-term weather phenomena start in a wider area at a similar time. Since sinusoidal components are positively



(a)



(b)

Figure 3.4: Initial and final values of PSD. (a) The initial value of PSD is plotted against the total wind power capacity. The initial PSD follows a quadratic function. (b) The final PSD is plotted against the total wind power capacity. The final PSD is approximately linearly proportional to the wind power capacity.

superposed, the amplitude of sinusoidal components of aggregated wind power in the low frequency increases in proportion to the total wind capacity. Considering that the square of an amplitude is the PSD level, the PSD level of un-normalized wind power increases in linear proportion to the square of the total wind capacity as shown in Figure 3.4(a), so the first PSD values (that is, for period 32 day) of normalized wind power are almost constant.

Wind farms are also affected by local weather phenomena, such as gusts, turbulence, and the wake effect. Since they happen in a short time, they will affect the sinusoidal components of the DFT in high frequency ranges. Since they happen independently in a local area, they start at different times at different wind farms, so phase angles of sinusoidal components in high frequency ranges are different. Then, when wind farms are integrated, sinusoidal components are less positively superposed, so the increasing speed of amplitudes is less than the increasing speed of capacity. Amplitudes of sinusoidal components in high frequency ranges increase approximately in linear proportion to the square root of the total wind capacity as shown in Figure 3.4(b). This means that the last PSD values (that is, for period 2 min) of un-normalized wind power increase in linear proportion to the total wind capacity, so the first PSD values of normalized wind power are inversely proportional to the wind capacity.

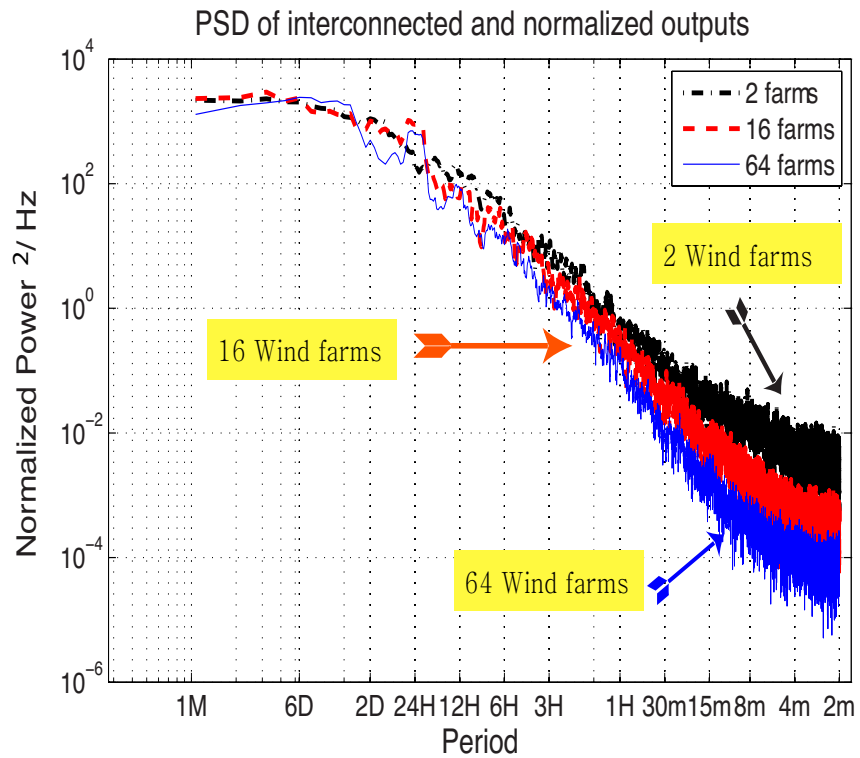
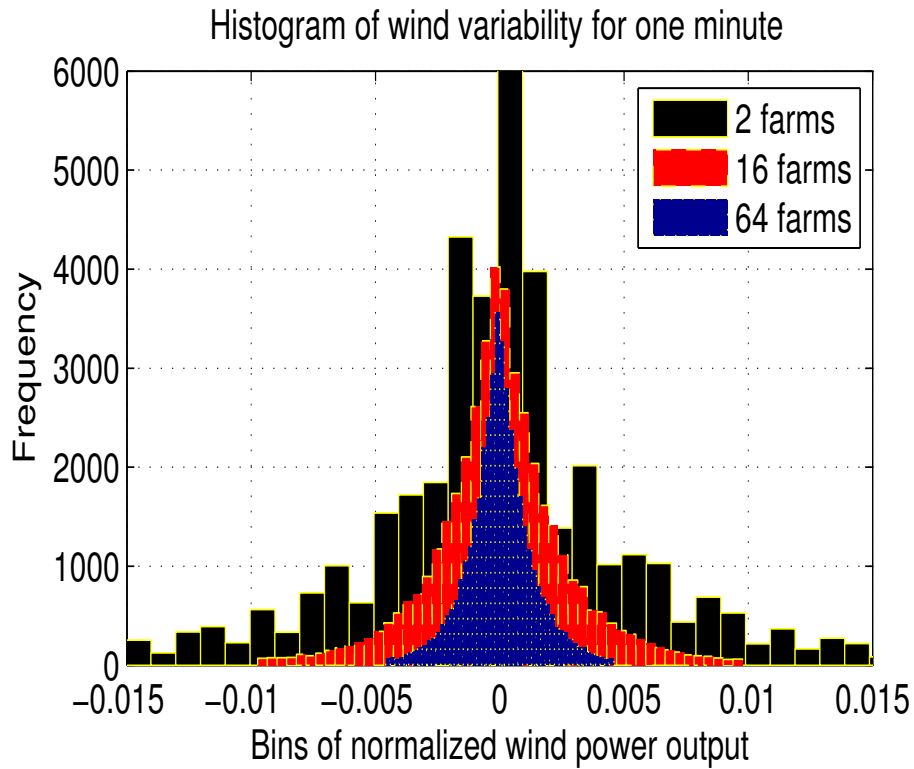


Figure 3.5: Slope becomes more negative as more wind farms are interconnected.

### 3.6.2 Slope Change Analysis

Since the first PSD value of wind power increases faster than the last PSD value of wind power, overall slopes of the PSD become more negative as more wind farms are integrated as shown in Figure 3.5. Under the assumption that phase angles are randomly distributed, if the PSD level decreases, amplitudes of sinusoidal components of the DFT decrease, and their power differences also decrease. Therefore, the normalized variability decreases as more wind farms are interconnected. Figure 3.6 shows that the standard deviation of the histograms of power differences at every one-minute time interval



*Figure 3.6: The power difference follows the Laplace distribution. The variability of normalized wind power decreases as more wind farms are interconnected.*

decreases as the number of interconnected wind farms increases. In short, the overall slopes of affine segments become more negative, and the relative variability decreases as the number of interconnected wind farms increases. This is also supported by Ernst’s empirical observation in [66] because correlation coefficients of wind power variability  $\Delta P$  decrease as time steps decrease. The goal of the below section is to reveal this relationship quantitatively.

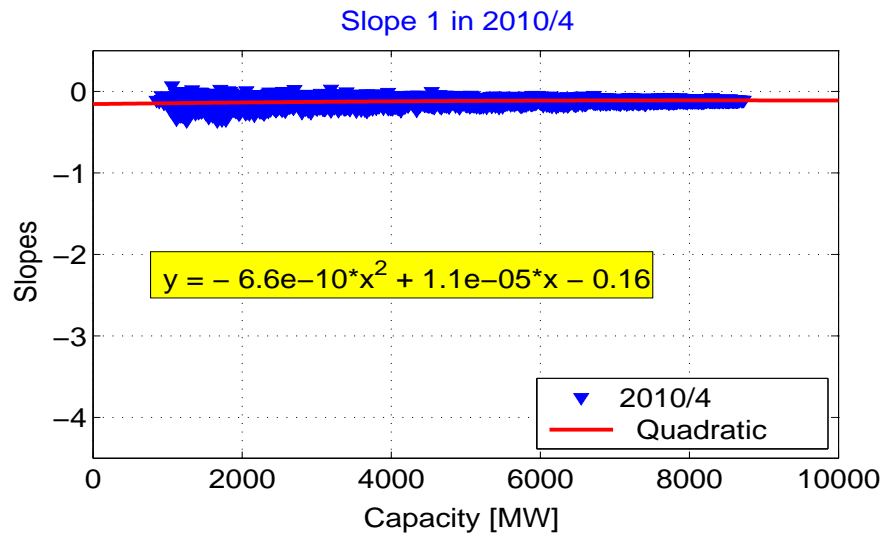
### 3.6.3 Regression

As the number of interconnected wind farms increases, sixteen response variables from the training data are modeled through the linear regression models: five slopes (not including the third slope), the first and last PSD values, the variability, the STD, the maximum wind power, minimum wind power, the mean of future output, the monthly ramp size, the monthly ramp rate, the amplitude of the daily cycle, and the mean of the daily cycle. These variables are fitted using a quadratic function, and regression coefficients are estimated by the least squares method [112]. The explanatory variable is selected as the capacity. It is assumed that wind farms in ERCOT will keep a similar configuration of turbine types and a similar configuration of wind farms in the future, although this may not be true and will affect result.

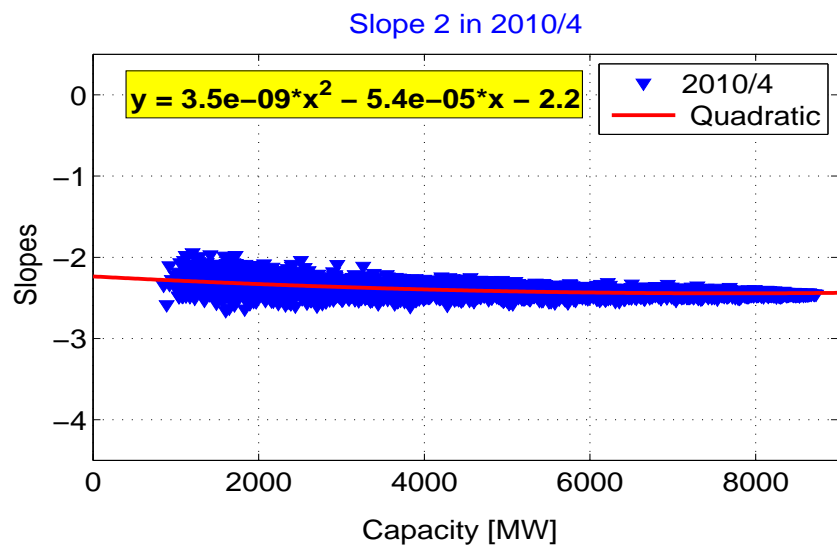
Figure 3.7(a) shows the slope of the first affine function whose frequency range has period between two days and one month. It is observed that the first slope does not change much as capacity increases, which means that the power outputs over longer periods are highly correlated with many wind farms. In Figure 3.7(b), the slope of the second affine function whose frequency range has period between eight hours and two days is shown. It is fixed around  $-2.4$  relatively independent of capacity. Since the second slope is around  $-2.4$ , power dissipation to the higher frequency starts from a period of two days.

Figure 3.8(a) shows the slope of the third affine function whose frequency range has period between six hours and eight hours. This segment works as a joint between the PSD pattern in the low and middle frequency.



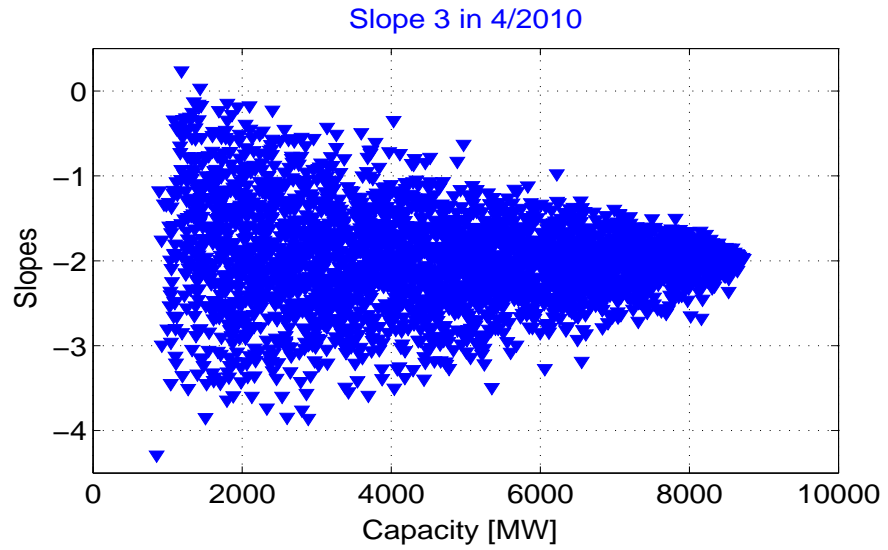


(a)

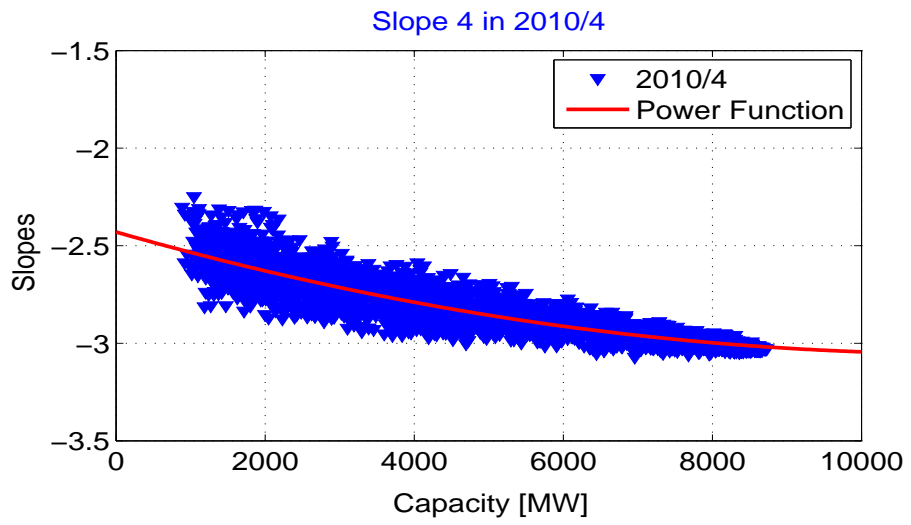


(b)

Figure 3.7: Slopes of first and second segments. (a) The slope of the first segment is plotted against the total installed capacity of wind power. (b) The slope of the second segment is plotted against the total installed capacity of wind power.



(a)



(b)

Figure 3.8: Slopes of third and fourth segments. (a) The slope of the third segment is plotted against the total installed capacity of wind power. Slopes do not follow a specific trend. (b) The slope of the fourth segment is plotted against the total installed capacity of wind power.

The first two segments start from the first PSD and have a distinct pattern. On the contrary, the last three segments start from the last PSD value and also have a distinct pattern. To satisfy two patterns at the same time, a piecewise function in a short frequency range is required to connect these two patterns. Since the third affine function will be used as a degree of freedom, patterns and fitted functions are not searched for the third slope. Figure 3.8(b) shows the slope of the fourth affine function whose frequency range has period between 15 minutes and six hours. It is observed that the fourth slope drastically decreases as capacity increases. The frequency range of the fourth segment might correspond to periods of short-term weather events in different local areas. Since those weather events are not strongly correlated, wind power fluctuation in this frequency range is canceled out when wind power in those areas is aggregated. Therefore, the power level in this frequency range is reduced. Fourth slopes appear to converge to a constant negative value, but further investigation with more wind power data is necessary. If it converges to a certain negative value, it means that wind power variability could not be reduced significantly with greatly increased total capacity. It also means that the geographical smoothing effect is saturated, so there is always a certain amount of wind power variability.

Figure 3.9(a) shows the slope of the fifth affine function whose frequency range has period between five minutes and 15 minutes. To detect trends of the fifth slope clearly, 100 wind farm integration orders are simulated, the mean of linear slopes of fifth slopes is measured. Then, it is observed that

the fifth slopes slightly increase as wind power capacity increases with a slope of 0.017 per GW. Figure 3.9(b) shows the slope of the sixth affine function whose frequency range has period between two minutes and five minutes. It should be noted that signals corresponding to the sixth slope are affected by independent local weather events. Sixth slopes are also fitted by a quadratic function.

In summary, the five slopes (not including the third slope) are forecasted through the regression analysis. The first, second, fifth, and sixth slopes are fitted to a linear function, but the fourth slope is fitted to a quadratic function. Furthermore, the initial and final PSD are fitted to a quadratic function. Other statistical characteristics are also analyzed through the regression model. Variability of actual wind power is linearly proportional to the square root of capacity. The following statistical characteristics are all linearly proportional to the capacity: the maximum value, minimum value, mean, STD of wind power, amplitude of the daily pattern, STD of the monthly ramp size, and STD of the monthly ramp rate.

### **3.7 Scenario Synthesis**

In this section, three wind power scenarios are synthesized based on the forecasted PSD at the target installed capacity to verify the suggested method. When the forecasted PSD is converted to the wind power waveform, the phase angles are estimated using a genetic algorithm (GA) with to match the distribution of the wind power differences to a Laplace distribution. Three

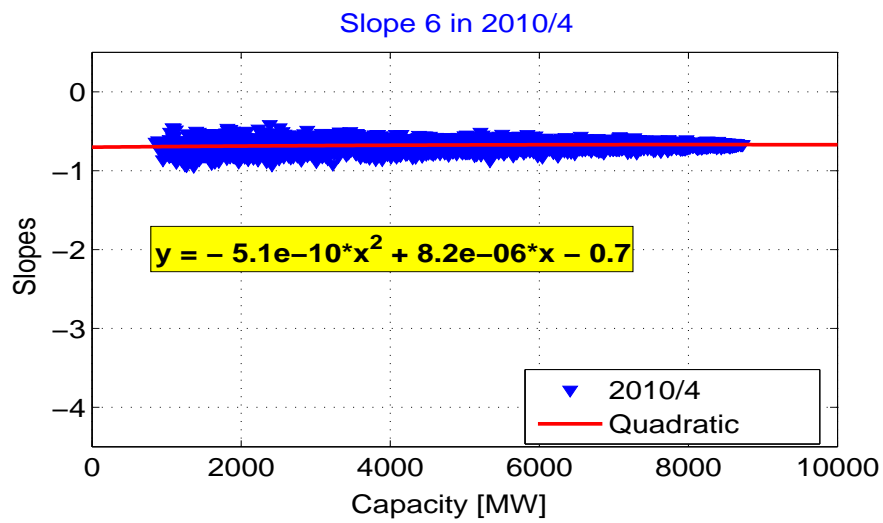
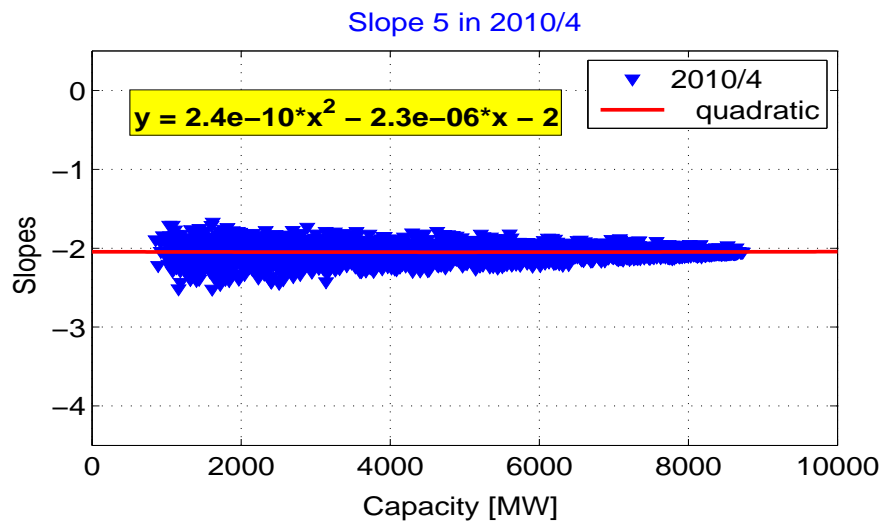


Figure 3.9: Slopes of fifth and sixth segments. (a) The slope of the fifth segment is plotted against the total installed capacity of wind power. (b) The slope of the sixth segment is plotted against the total installed capacity of wind power.

scenarios are as follows. First, a wind power scenario in April 2010 is synthesized using actual wind power in April 2010. Second, a wind power scenario in April 2010 is synthesized using actual wind power in April 2009. These two scenarios are compared to the actual wind power in 2010. Finally, the future wind power at 10,000 MW wind power capacity is synthesized.

### 3.7.1 PSD Forecasting

An intermediate sample path is synthesized from the forecasted PSD through several steps. An intermediate sample path is a waveform converted from the forecasted PSD. The process of forecasting the PSD as follows. *a)* The slopes of six segments, the first PSD value, and the last PSD value are estimated from the linear regression analysis at the target installed capacity of wind power. *b)* The first three segments are concatenated and anchored at the initial PSD value, and the last three segments are concatenated and anchored at the final PSD value. *c)* A DFT of the normalized waveform is rebuilt from the forecasted PSD. The conversion equation is as follows:

$$|X(f)| = \sqrt{\frac{\hat{P}_x(f) \times L \times F_s}{2}}, \quad (3.39)$$

where  $P_x(f)$  is the PSD and  $L$  is the length of the data, which corresponds to the total number of minutes in a month.  $F_s$  is the sampling frequency, so it is 1/60. Then, the DFT is converted to amplitudes of sinusoidal components using the inverse-FFT. *d)* Since the transformed waveform has a normalized value, it is recovered to the original scale by multiplying by the future capacity.

After these processes, the 24h daily cycle should be added back to the synthesized intermediate sample path. The amplitude and mean of the daily deterministic signal are used to synthesize the daily signal. It is well known that wind tends to blow strongly at night in Texas [195]. The mean of hourly wind power is calculated, and it is found that wind tends to blow most strongly at two a.m. Then, it is assumed that a daily cycle of a sample path also peaks at two a.m. Therefore, its phase angle is set at  $30^\circ$ , which shifts the cosine function of the daily pattern to two a.m. Finally, the negative power output is set at zero, and generated power that is bigger than the capacity is curtailed to the capacity.

### 3.7.2 Phase Angle Generation

To transform the forecasted PSD to a waveform, phase angles of the FFT are required. However, phase information is lost when the PSD is estimated. Since the phase angle is a random signal itself, if it is not modeled stochastically, one realization of phase angles should be chosen from an infinite number of ensemble phase angles satisfying the conditions of future wind power output. In this paper, the GA is used to search phase angles. The termination conditions are that the searched phase angles have the same statistical characteristics as actual phase angles, and that wind power scenarios have the forecasted statistical characteristics. Constraints to satisfy termination conditions are as follows: *a)* Phase angles are almost uniformly distributed, *b)* The histogram of ramp rates of transformed wind power should follow the

Laplace distribution, *c*) The mean of restored signals should match the forecasted mean, *d*) The maximum value of the restored signal should be less than the capacity, and finally, *e*) since the phase angle corresponding to a 24-hour period has a fixed angle of  $30^\circ$ , a 24-hour period daily deterministic signal with an angle of  $30^\circ$  will be added.

### 3.7.3 Genetic Algorithm

Searching phase angles that satisfy the conditions mentioned above is designed as an optimization problem in a GA. The objective function is designed based on the second condition, so it minimizes the MSE between two distributions: *a*) the Laplace distribution with the forecasted STD, and *b*) the minute to minute power difference of synthesized wind power. This objective function will make the power difference of wind power follow the Laplace distribution. Other conditions are described as inequality constraints and added to the objective function with Lagrange multipliers. Searching phase angles is an optimization problem to satisfy several objective functions at the same time. In this kind of problem, defining optimality and comparing local optimums is difficult. One approach to this kind of problem is to use a GA. A phase angle of a given frequency is considered as a gene, and a collection of genes with respect to all frequency ranges is considered as a chromosome. Therefore, phase angles are searched using the GA in this paper.

1) ***Initialization:*** Uniformly distributed phase angles corresponding to the length of the FFT are initially generated and considered as a population



of chromosomes. The size of the population is 1000, and their values are limited between  $-\pi$  and  $\pi$ .

2) ***Evaluation:*** Chromosomes in a pool are used to transform the PSD in the wind power, and the wind power is evaluated by the objective function. The objective function of the GA is decided by considering the constraints mentioned above. The inverse of the output of the objective function is called the fitness value of given chromosome. Since the ramp rates of wind power follow the Laplace distribution, the mean of the Laplace distribution of training data is measured. The realized waveform of a certain chromosome should follow the Laplace distribution with the same parameters as the training data. The objective function is the mean square of errors between two Laplace distributions.

3) ***Evolution:*** According to fitness values, the 20% of the population having minimum fitness values are selected. The selected chromosomes evolve through three evolutionary steps: generation, mutation, and adoption. The surviving chromosomes generate children using the crossover. Furthermore, randomly selected genes of a selected chromosome are mutated. Finally, the GA receives new randomly generated chromosomes which occupy 20% of all chromosomes and are mixed with surviving chromosomes. Mixing purely randomly generated phase angles with surviving ones can allow the randomly generated ones to survive; otherwise, they would never be able to survive in the competition against the naturally surviving ones. Mixing phase angles in this manner generates new patterns which guarantee a more accurate final

solution.

4) **Termination:** When the values of the objective function are converged to the local minimum, or when the program reaches the maximum iteration number, the program terminates. The maximum iteration number is set as 1,000, and convergence is detected when iterated values change less than a certain error.

#### 3.7.4 Phase Angle Analysis

To verify the hypothesis that there is a strong diurnal period, the means of the STDs of phase angles versus frequency are plotted in Figure 3.10. The STD of phase angles from all wind farms is calculated for every frequency in order to observe how phase angles are distributed with frequencies. This process is repeated for monthly wind power for seven years, and the mean of the STD of phase angles versus frequency is obtained. It is observed that phase angles are more dispersed as the frequency increases. It should be noted that the STD of uniformly distributed phase angles is 1.813 and that the STD is close to 1.813 for high frequencies in Figure 3.10.

### 3.8 Validation

The first scenario with 8,728 MW target capacity in April 2010 is synthesized based on wind power in April 2010. It is synthesized to validate whether the suggested method can regenerate wind power with forecasted statistics. The second scenario of 8,728 MW capacity is synthesized based on

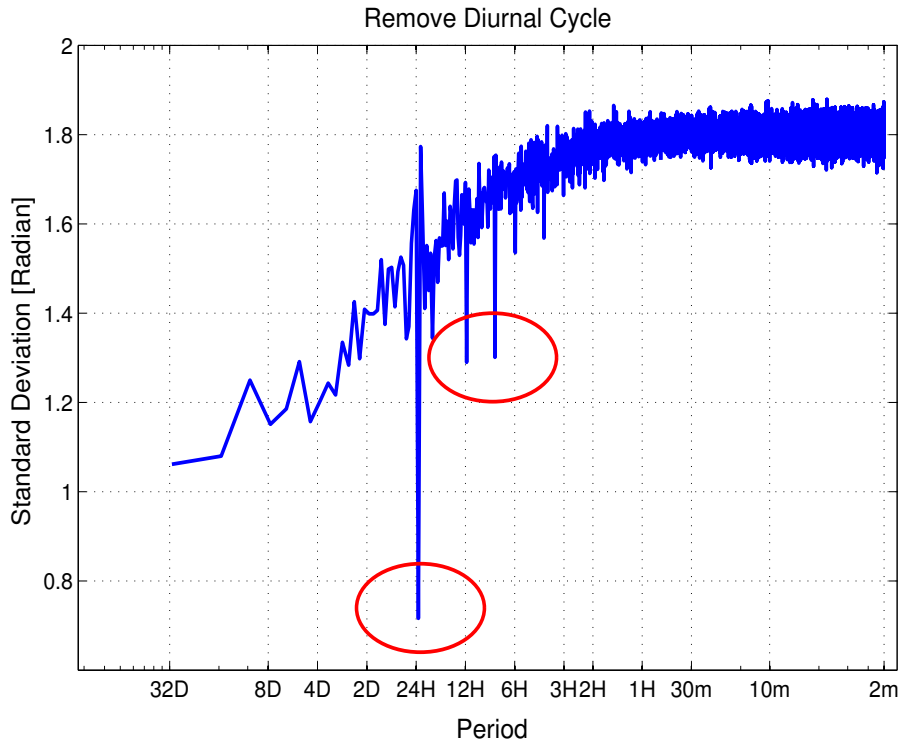
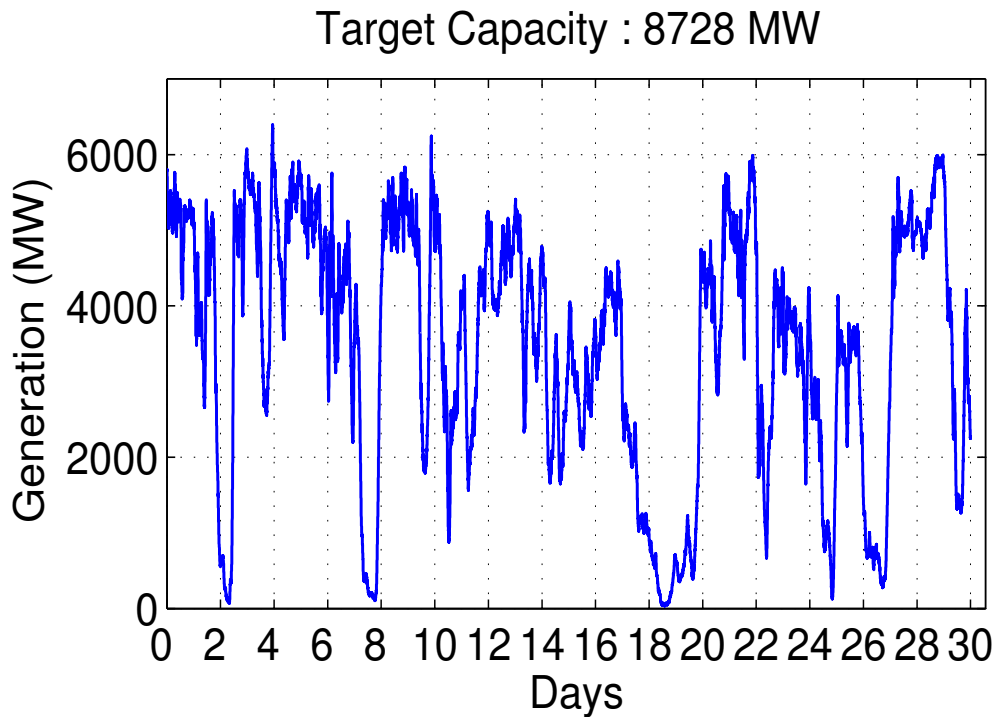


Figure 3.10: The means of the STDs of phase angles per frequency.

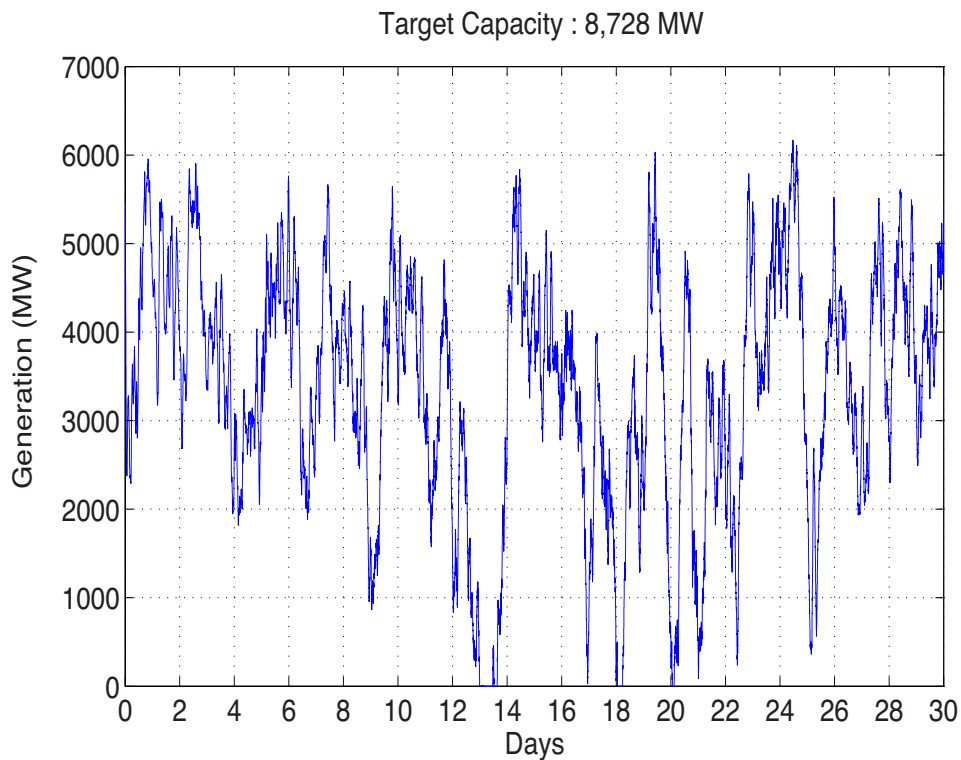
wind power in April 2009. It is synthesized to show that the approach can forecast future statistics and synthesize the waveform satisfying forecasted statistics. It should be noted that the capacity of wind power in 2009 is 8,120 MW capacity. Both scenarios are compared to actual wind power in April 2010. Finally, the third scenario is synthesized with 10,000 MW capacity to show the reliability and stability of the approach. It is assumed that the geographical distribution of wind farms is the same as current wind farms.

The actual wind power scenario sampled in April 2010 is plotted in Figure 3.11. Furthermore, the second scenario is plotted in Figure 3.12. It



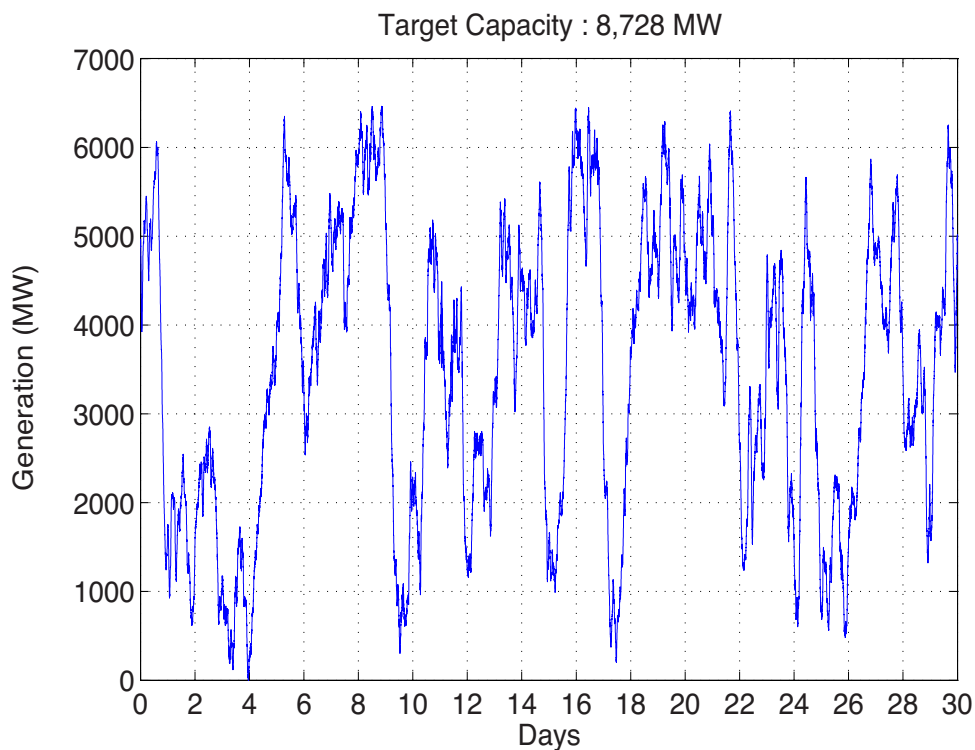
*Figure 3.11: The actual wind power scenario sampled in April 2010.*

is clear that both scenarios are very similar in variability. However, although the wind power fluctuation in Figure 3.11 is concentrated on early in the month, the wind power fluctuation in Figure 3.12 is spread over a month. Furthermore, although the frequency of minimum wind event of the actual wind power is equally distributed, the frequency of minimum wind event of the second scenario is concentrated in the middle of month. The third and fourth scenarios are plotted in Figure 3.13 and Figure 3.14. The third scenario has less fluctuation than the actual wind power, but it has the similar frequency for the minimum wind power occurrence. For the fourth scenario, the maximum



*Figure 3.12: The wind power scenario in 2010 with the installed capacity in 2010 is plotted. This scenario is synthesized based on the wind power in 2009.*

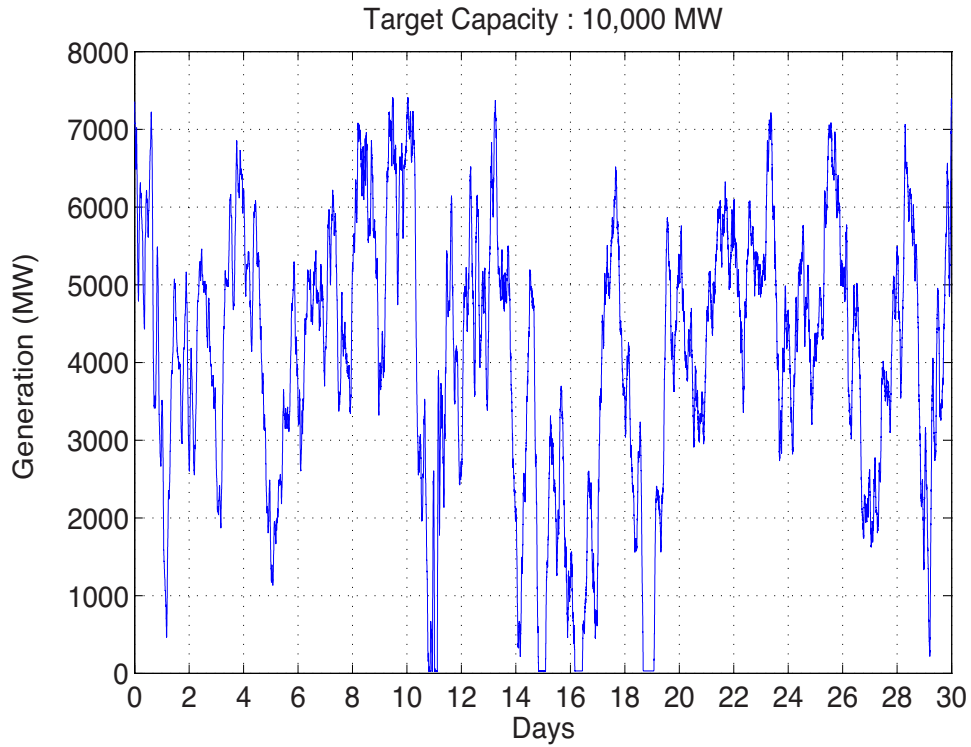
wind power is around 7,500 MW because of the low capacity factor of wind power although the total capacity is 10,000 MW. It is interesting to note that there is still zero wind power in spite of the 10,000 MW installed capacity of wind power. Furthermore, the minimum of actual wind power was 100 MW, and the minimum of scenario is zero, so they have a similar minimum value. Moreover, periodic patterns for two days are strongly observed in Figure 3.11 and 3.14, and patterns correspond to the low standard deviation at the two day period in Figure 3.10. Therefore, scenarios in Figure 3.12, Figure 3.13,



*Figure 3.13: The wind power scenario in 2010 is synthesized based on the wind power in 2010. This scenario is synthesized based on the wind power in 2010.*

and Figure 3.14 are similar in general pattern to the actual wind power in Figure 3.11.

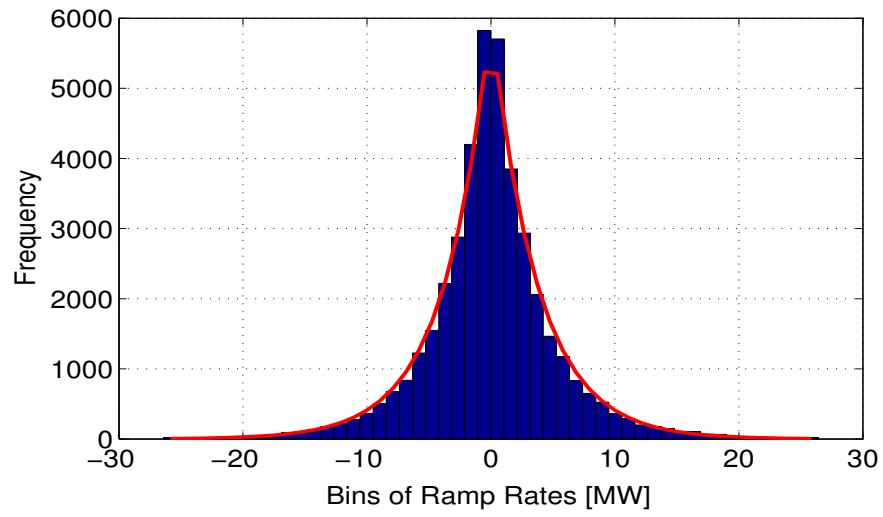
The actual distributions of wind power differences and phase angles are plotted in Figure 3.15. It is clearly shown that the wind power differences follow the Laplace distribution as mentioned above, and the phase angles are uniformly distributed. For the wind power scenario that is synthesized based on the 2010 wind power, the distributions of wind power differences and phase angles are plotted in Figure 3.16. The distribution of the synthesized scenario



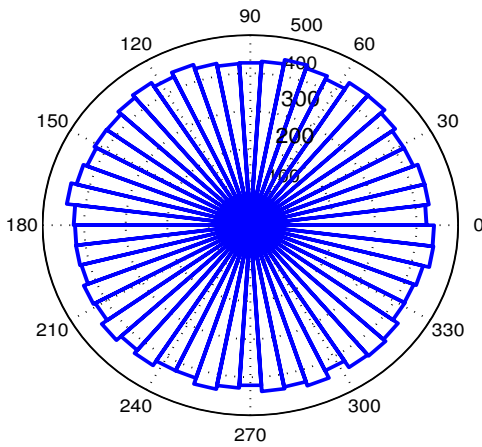
*Figure 3.14: The wind power scenario of 10,000 MW installed capacity in 2030 is plotted. This scenario is synthesized based on the wind power in 2010.*

also follows the Laplace distribution, but has heavier tails than the distribution of the actual measured scenario. Accordingly, the peak value of the distribution of synthesized scenario is also less than that of the distribution of the actual scenario. However, it is clearly shown that the distribution of the synthesized and actual scenarios are very similar. The phase angles of the synthesized scenario are also uniformly distributed, but they are less smooth than the actual phase angles.

Statistical characteristics of synthesized wind power are enumerated in



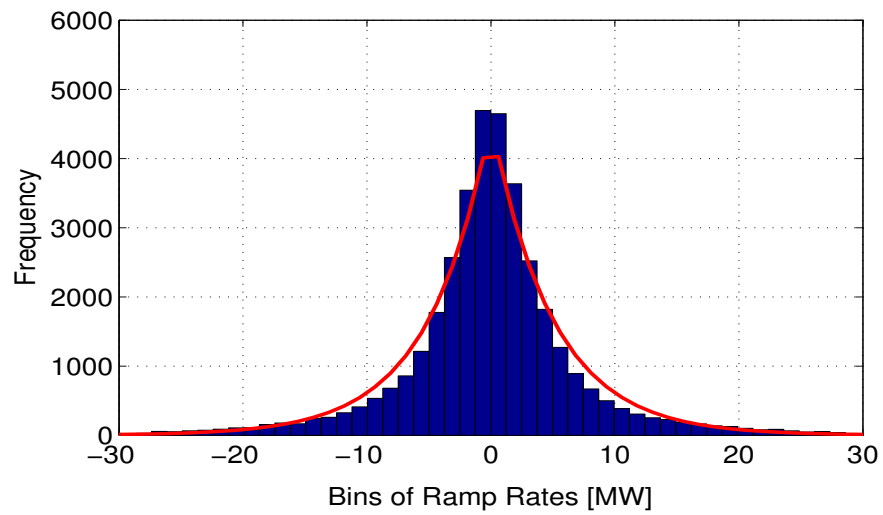
(a)



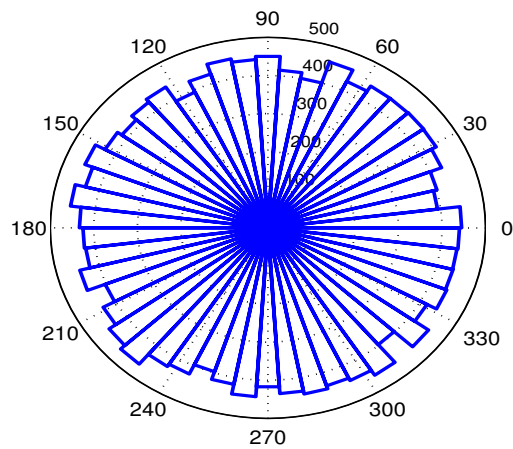
(b)

Figure 3.15: (a) The distribution of actual wind power ramp rates. (b) The distribution of actual phase angles. The phase angles are uniformly distributed.





(a)



(b)

Figure 3.16: (a) The histogram of ramp rates of wind power scenario per minute is plotted. (b) The phase angles of the synthesized PSD of the wind power scenario are plotted. The phase angles are almost uniformly distributed.

Table 3.2. In the second column of Table 3.2, the statistical characteristics of actual wind power in April 2010 are described, and the actual wind power in April 2010 is shown in Figure 3.11. The first scenario in the third column of Table 3.2 is synthesized by actual wind power in April 2009, and the synthesized wind power is plotted in Figure 3.12. Furthermore, the second scenario in the fourth column of Table 3.2 is synthesized using wind power from the month of April for years through 2010, and the synthesized wind power is plotted in Figure 3.13. Statistical characteristics of the third scenario are closer to those of actual wind power than those of the second scenario since the third scenario is trained with wind power from the month of April 2010. It is also observed that properties of the second scenario are similar to actual ones although the statistical characteristics in the second scenario are estimated except for the variability. Since wind power in 2009 was relatively less than the yearly trend of wind power, statistical characteristics might be under-estimated. Therefore, this suggests that multiple years of data should be utilized to synthesize future year.

### **3.9 Ancillary Service Estimation**

The required amounts of CPS1-compliant regulation service is determined based on the synthesized scenarios in this dissertation [44]. The variability that is extracted from the synthesized scenarios at every installed capacity of wind power is used to represent the required amounts of regulation service. Since the wind power variability is defined as the standard deviation of

Table 3.2: Statistical characteristics of actual and synthesized wind power

Type	Actual	Scenario	Scenario	Scenario
Target Month	4/2010	4/2010	4/2010	4/2015
Last Training Year		2009	2010	2010
Capacity [MW]	8,728	8,728	8,728	10,000
Mean [MW]	3,478	3,350	3,493	4,061
STD [MW]	1,666	1,394	1,640	1,743
Max [MW]	6,393	6,166	6,459	7,388
Min [MW]	24.87	0	1.9536	25.729
Mean of Daily Pattern [MW]	3,478	3,355	3,494	4,061
Amplitude of Daily Pattern [MW]	559.82	446.65	509.94	634.56
STD of Ramp Size [MW]	1,816	1,776	1,822	2,076
STD of Ramp Rate [MW/min]	4.001	3.3982	3.23	4.887
Variability [MW/min]	18.986	19.598	18.798	19.4334

power difference, it can be assumed that the variability is the forecast error when the persistent forecasting model is used. Since the regulation service is used to mitigate the fluctuation from the net load variability, and if the load variability is assumed to be fixed, the STD of the future net load can be calculated. Furthermore, the criteria of the required amounts of the regulation service is to satisfy the CPS1 score, which depends on the variance of the system frequency. The simulation was performed under the assumption that the installed wind capacity becomes 18 GW. Then, the required amounts of regulation service will be slightly increased compared to the 2010 installed

wind capacity of 8,728 MW. The estimated amount of regulation service is 632 MW, but the actual amount of regulation service in 2010 is 585 MW. This is a good example that shows how the wind power scenario generation algorithms can be used to determine the required amounts of future ancillary services.

### **3.10 Conclusion**

The algorithm to synthesize the future wind power scenario with respect to the installed capacity of wind power is summarized. As a future work, the unsolved problem that should be handled in future is introduced.

#### **3.10.1 Summary**

Wind power scenarios are synthesized by considering the forecasted future PSD and various statistical information, which includes the variability, STD, maximum wind power, minimum wind power, mean, monthly ramp size, monthly ramp rate, amplitude of the daily cycle, and mean of the daily cycle. The variability of un-normalized wind power is linearly proportional to the square root of capacity. Other statistical characteristics are all linearly proportional to the capacity. Slopes of the PSD follow different trends. The slopes of the first and second segments stay at similar values through each simulation. The slope of the third segment shows an unpredictable trend, so its frequency range is set short, and it is just used as a connector between other slopes. The slope of the fourth segment becomes more negative as the penetration level of wind power increases. The slope of the fourth segment

appears to converge to a constant negative value. Therefore, the frequency range corresponding to the fourth segment is important when wind farms are integrated. Finally, the slopes of the fifth and sixth segments stay at similar values as the total capacity increases. As a result, the monthly future wind power output is rebuilt using the forecasted PSD. Furthermore, as shown in the above section, the future wind power scenarios can be used to estimate the required amounts of future regulation services. Even though this study focuses on a particular case in ERCOT, these results suggest a general framework for analyzing wind power in the frequency domain and generating typical scenarios in the far future by considering future fluctuations.

In conclusion, as the second step to increase the penetration level of wind power, sample paths of total wind power are synthesized In this chapter. Scenarios can be used to simulate the power system so that the power system planner can determine the proper amounts of regulation service, storage size, and incentive signals for EVs. As the third step, improved transmission expansion planning to bring more wind resource to the load center can provide the accurate sum of generation and transmission investment costs by simulating the power system based on the load and wind power scenarios that are synthesized by keeping the stochastic correlation structure between load and wind power. Toward this end, the next chapter discusses the transmission expansion planning based on load and wind power scenarios.

## Chapter 4

# Scenario Generation through GDFM and Transmission Expansion Planning

Optimal transmission expansion plans to bring more wind resources to the load center can be determined by testing the power system using many possible load and wind power scenarios [121]. The optimal plan to expand transmission lines to include additional wind power should secure the system reliability and minimize the total costs of expansion and system operation. In this process, load and wind power scenarios should be represented with their actual correlation coefficients to estimate total costs accurately. For example, positively correlated wind farms sharing the same transmission lines might cause congestion and thus curtailment. Furthermore, negatively correlated load and wind power can reduce the generation adequacy and thus cause higher operation costs [198]. Quantitatively addressing the future variability could affect the plan of the transmission line upgrades when screening sites of new wind farms since additional transmission lines are required to take advantage of the wind diversity over a wide area [33].

## 4.1 Literature Review

Many algorithms have been developed to generate many wind power or load scenarios with a correlation structure. First, negatively correlated total load and wind power were synthesized in order to plan the transmission expansion in [150]. The normalized load and wind power were multiplied by the future total demand and wind power capacity and were graphed in a scatter plot. Samples in the scatter plot were divided into a few partitions so that every partition had the same number of samples. The averages of all samples in each partition were selected as point-wise scenarios, and the probabilities of all partitions were equally set as an inverse of the number of partitions. However, instead of the total load and wind power, it would be advantageous to use higher dimensions of the probability by considering individual wind farms and load zones.

In [149], instead of forecasting individual wind power outputs, the correlation structure among wind farms was incorporated into a special power curve that was designed to match average wind speed to average wind power in California. Average wind speed was modeled as the Autoregressive (AR) process, and then new wind speed scenarios were generated by changing the noise terms in the AR process. Wind speed scenarios were converted to wind power through a special power curve. Wind power scenarios were then used to estimate the reserve requirements for the high penetration level of wind power. In this example, although total wind power in California was synthesized, a better result could be obtained if the correlation between load and wind power

is considered.

Morales et al. [138] synthesized wind speed by considering the cross-correlations of wind speed values at different locations. In [138], wind speed was modeled as the multivariate autoregressive moving average (ARMA) process, and then it was approximated to the univariate ARMA process. Many scenarios were generated using correlated noise terms. These scenarios were used to solve the probabilistic power flow model [137] in which correlated load and wind power scenarios were used to estimate the mean and standard deviations of power flows. However, parameters in the ARMA model were estimated under the assumption that noises are independent. Scenario generation based on an assumption of correlated noise terms could yield better results.

Block scenarios based on historical data were used in [16] to decide optimal locations of wind farms and required transmission lines. Five blocks of different load levels had three different load capacity factors, and each 15-load block had six wind capacity factors, so 90 total load and wind scenarios were used, resulting in a tree graph. Although scenarios consisted of several values, using multiple scenarios that have been generated on each bus might be more practical. Furthermore, instead of the same correlation between wind and load, various correlation coefficients among wind farms and load zones might generate more realistic scenarios.

Recently, power spectral density (PSD) analysis has also been used to synthesize the wind power. Since all time series structures can be converted to transfer functions in the frequency domain, the PSD analysis can also ana-



lyze the temporal correlation in a time series. In [120], under the assumption that wind power consists of discrete sinusoidal waveforms, the temporal correlation in wind power was investigated by analyzing the relationship among amplitudes of sinusoidal waveforms, and this mined relationship was used to synthesize the sample paths of future wind power. However, a single total wind power from wind farms, not multiple sample paths, was synthesized, and phase information was lost in the PSD estimation. On the contrary, the cross-PSDs (CPSD) that are estimated by the Fourier transform of cross-correlation functions can retain phase information, so not only the temporal correlation but also the spatial correlation among wind farms can be maintained, which is beneficial for synthesizing multiple sample paths from wind farms.

As shown above, scenario synthesis has evolved from point values, such as maximum and minimum, to a single time series. Recently, advances in data mining have made it possible to record thousands of wind turbine outputs and load zones. Thus, it might be possible to synthesize huge amounts of time series individually, but the correlation between time series cannot be easily synthesized. Since both wind farms and load are affected by weather, if common factors can be extracted, the computation time and number of parameters can be reduced while keeping the correlation structure among time series. The GDFM analyzes those common factors.

The GDFM has been used widely in forecasting economic indexes [143]. It evolved from the static factor model, which has been used to find latent factors in observation data [164]. In factor models, data is decomposed into the

common component and idiosyncratic component. The common component is assigned by concurrent factors through the factor loading, in which the cross-sectional correlation of data is incorporated. The idiosyncratic component is assumed to arise from measurement errors in individual time series. In order to model not only the cross-sectional correlation but also the temporal correlation, Sargent and Sims [168] proposed the dynamic factor model where time lag and lead structures were implemented in factor loadings.

One unrealistic assumption in both the static and dynamic factor models has been that noise vectors are orthogonal, which makes the covariance matrix of noises diagonal [116]. The effort to relax this condition started from [39] and led to the generalized factor model, in which the covariance matrix of idiosyncratic components is not diagonal. Finally, the GDFM, which combines the dynamic factor model and relaxed condition on noise terms, was proposed in [74]. In short, “Generalized” means that the covariance matrix of idiosyncratic components is not diagonal, and thus idiosyncratic components are weakly correlated noise terms. “Dynamic” means that dynamic factors are loaded with time lag and lead in factor loadings. Since the dimension of dynamic factors is generally less than that of the observation data, a reduction in dimension is expected.

The GDFM has been enhanced to allow it to integrate multivariate time series [75]. It is more parsimonious than the multivariate time series itself because the cross-sectional dimension of factors is generally less than that of observation data. If dynamic factors are designed as a time series

process, dynamic factors can be represented as white noise, which is called the dynamic shock, and thus the common components can be driven by dynamic shocks. For other examples, Bernanke et al. [21] implemented the AR process in factors and observation in order to model observation as time series and to use additional information contained in factors. This is called the Factor-Augmented Vector Autoregressive (FAVAR) because the joint dynamics of observation and factors are implemented simultaneously. Furthermore, the GDFM was approximated in [73] by relaxing the assumption of an infinite dimension of observation data and an infinite order of time lag in factor loadings in [74]. The approximated GDFM is used in this dissertation.

Dimension reduction in the GDFM creates one additional advantage, but at a cost. It was found that the time series structure in the static factor model has an inherent zeroless transfer function that is represented only as the AR process [7]. The AR process is obviously simpler than other time series models. However, if the number of dynamic factors is less than the data dimension, the covariance of dynamic factors becomes singular, so the time series becomes a singular AR process, which could be unstable [55]. However, an insufficient rank on the common component can be increased by adding an idiosyncratic component to the common component.

## 4.2 Additional Motivations and Goals

In this study, scenarios of multiple time-series are synthesized by a systematic process that formulates load and wind power through the GDFM.

The premise of this research is that a few dynamic factors drive the correlated movements of many time series. This process can generate an arbitrary number of scenarios of load and wind power simultaneously by changing the dynamic shocks. The wind power scenarios can be verified by evaluating trajectories of wind power [155], but this study mainly focus on verifying the correlations and statistical information of wind power and load. Furthermore, the importance of representing the correlation structure is verified by solving the economic dispatch and transmission expansion planning with synthesized scenarios. The application of the GDFM to load and wind power is a plausible approach. Since [35], load and wind power have been modeled as the time series model. It was also shown in [117] that factor analysis is a plausible approach to estimate curtailed wind power by using wind power from surrounding wind farms.

### **4.3 Preprocessing and the GDFM**

In this section, load and wind power are normalized, and then the periodic component is extracted. The remaining data is decomposed into the common and idiosyncratic components. Then, dynamic factors and their factor loadings are extracted from the common component. Furthermore, the dynamic factors are modeled as the vector autoregressive (VAR) process of dynamic shocks.

### 4.3.1 Preprocessing Wind Power & Load Data

Wind power data from 96 wind farms and load data from eight areas in ERCOT are used. Data is sampled at every hour in 2013. The eight areas are the Coast, East, Far West, North1, North2, South1, South2, and West. Total wind power capacity is 10,407 MW, and the peak demand is 50,698 MW. Data is normalized by its maximum value. Then, the sum of periodic components is extracted under the assumption that they consist of the harmonics of diurnal periodicity. The sum of periodic components is estimated by averaging data for the same hour every day. In more detail, the each observation of wind or load data  $X \in \mathbb{R}^{24 \cdot 365 \times 104}$  is reshaped into the matrix  $Y \in \mathbb{R}^{24 \times 365 \times 104}$ . The means of every hour are stacked in  $Z \in \mathbb{R}^{24 \times 104}$ , which is averaged over a days. After the periodic component is extracted, residuals are standardized: means are subtracted, and residuals are divided by standard deviations.

### 4.3.2 Introduction to the GDFM

Suppose that the matrix  $\mathbf{X} \in \mathbb{R}^{N \times T}$  is observed from  $N$  sources for  $T$  hours, and rows of  $\mathbf{X}$  are time series of wind power and load. The column of  $\mathbf{X}$  sampled at time  $t$  is denoted as  $\mathbf{X}_t = \{x_{1t}, x_{2t}, \dots, x_{Nt}\}^T$ , where a doubly indexed sequence is used for the scalar value  $x_{it}$  for  $i = 1, \dots, N$  and for  $t = 1, \dots, T$ . The structural assumption of  $\mathbf{X}$  is that it consists of the common component  $\boldsymbol{\chi} \in \mathbb{R}^{N \times T}$  and the idiosyncratic component  $\boldsymbol{\xi} \in \mathbb{R}^{N \times T}$  as in [74]:

$$\mathbf{X}_t = \boldsymbol{\chi}_t + \boldsymbol{\xi}_t, \quad (4.1)$$

where  $\boldsymbol{\chi}_t$  is the column of  $\boldsymbol{\chi}$  sampled at  $t$ , and  $\boldsymbol{\xi}_t$  is the column of  $\boldsymbol{\xi}$  sampled at  $t$ .

The structural assumption of  $\boldsymbol{\chi}$  is that it has  $Q$  dimensional dynamic factors, so it can be decomposed into the dynamic factor loading  $\mathbf{A}(L) \in \mathbb{R}^{N \times Q \times (2S+1)}$ , which is a set of polynomial matrixes of time leads and lags, and the dynamic factors  $\mathbf{f} \in \mathbb{R}^{Q \times T}$  as:

$$\begin{aligned}\boldsymbol{\chi}_t &= \mathbf{A}(L)\mathbf{f}_t \\ &= \mathbf{A}_{-S}\mathbf{f}_{t-S} + \cdots + \mathbf{A}_0\mathbf{f}_t + \cdots + \mathbf{A}_S\mathbf{f}_{t+S},\end{aligned}\tag{4.2}$$

where  $L$  denotes the forward and backward shifts for  $t = -S, \dots, S$ , and where  $\mathbf{f}_t$  is the column of  $\mathbf{f}$  sampled at  $t$ . The polynomial matrix is a matrix that has a polynomial for each of its entries. The general assumption is that  $Q < N$ , so the common components can be described with fewer dimensions. Since the moving-average structure is encapsulated into  $\mathbf{A}(L)$ ,  $\mathbf{X}_t$  is represented as the moving-averaged  $\mathbf{f}_t$ . Since the starting point of designing the GDFM is  $\mathbf{X}_t$ , it is reasonably assumed that there is a filter  $\mathbf{B}(L) \in \mathbb{R}^{Q \times N \times (2S'+1)}$  that can extract  $\mathbf{f}_t$  from  $\mathbf{X}_t$ :

$$\mathbf{f}_t = \mathbf{B}(L)\mathbf{X}_t.\tag{4.3}$$

It is also assumed that  $\mathbf{f}$  has a time-series structure.  $\mathbf{f}$  is driven by innovation  $\boldsymbol{\epsilon}$  through the AR process with an order  $R$  through parameter matrix  $\mathbf{C}(L) \in \mathbb{R}^{Q \times Q \times (R+1)}$  as:

$$\begin{aligned}\boldsymbol{\epsilon}_t &= \mathbf{C}(L)\mathbf{f}_t \\ &= \mathbf{f}_t - \mathbf{C}_1\mathbf{f}_{t-1} + \mathbf{C}_2\mathbf{f}_{t-2} + \cdots + \mathbf{C}_R\mathbf{f}_{t-R},\end{aligned}\tag{4.4}$$

where  $\boldsymbol{\epsilon}_t$  is the column of  $\boldsymbol{\epsilon}$  sampled at  $t$ . The time series in (4.4) has only time lags because of the causality, although  $\boldsymbol{f}$  can also be modeled with time leads and lags. An additional assumption is that  $\boldsymbol{\epsilon}$  is driven by dynamic shocks  $\boldsymbol{\delta}$  as

$$\boldsymbol{\epsilon}_t = \boldsymbol{Z}\boldsymbol{\delta}_t, \quad (4.5)$$

where  $\boldsymbol{\delta}_t$  is the column of  $\boldsymbol{\delta}$  sampled at  $t$ . The shocks  $\boldsymbol{\delta}$  are assumed to be white noise whose spectral density is constant. The matrix  $\boldsymbol{Z}$  is used to make  $\boldsymbol{\epsilon}$  have the same correlation structure of the actual observation data since  $\boldsymbol{\delta}$  is the uncorrelated white noise. It should be noted that  $\boldsymbol{\epsilon}$  is an estimated value. Then, the correlation structure is extracted from  $\boldsymbol{\epsilon}$  through the Cholesky decomposition, and it is denoted as  $\boldsymbol{Z}$  [138]. The combination of (4.2), (4.4), and (4.5) is called the GDFM.

It should be noted that although the ARMA process is more advanced, the AR process is sufficient to describe the stochastic process in load and wind power. When  $\boldsymbol{X}$  is represented parsimoniously so that  $N > Q$ , the GDFM has a tall transfer function in which the number of output variables is bigger than the number of input variables. When a tall transfer function in the time domain is transformed to a spectral density in the frequency domain, the spectral density becomes zeroless. The zeroless spectral density is equivalent to the AR process, so the AR process can be used to model the dynamic factors [46].

Additional assumptions are required to build the GDFM. It is assumed that  $\boldsymbol{\xi}$  is weakly correlated as mentioned above and that  $\boldsymbol{f}$  is orthogonal to

the idiosyncratic component as:

$$\mathbb{E} [\boldsymbol{\chi}_j \boldsymbol{\xi}_k'] = 0, \quad (4.6)$$

for  $j = 1, \dots, N$  and for  $k = 1, \dots, N$ . It is also assumed that the spectral density of the common component  $\boldsymbol{\mathcal{S}}^X(\omega) \in \mathbb{R}^{M \times N^2}$  is a rational spectral density, which can be represented as a fraction.  $M$  is the number of frequencies. Every rational transfer function can be realized by the time series process by a left or right matrix fraction description [6]. A rational spectral density can have a state space representation in the time domain as:

$$\boldsymbol{\chi}_t = \mathbf{A}(L)[\mathbf{C}(L)]^{-1} \mathbf{Z} \boldsymbol{\delta}_t. \quad (4.7)$$

Then, (4.7) can be decomposed into (4.4) and (4.2), and the rational spectral density can lead to the AR process.

### 4.3.3 Decomposition

The observation data  $\mathbf{X}$  is decomposed into  $\boldsymbol{\chi}$  and  $\boldsymbol{\xi}$  by decomposing the spectral density of the observation data,  $\boldsymbol{\mathcal{S}}^X(\omega_m) \in \mathbb{R}^{M \times N^2}$ , into  $\boldsymbol{\mathcal{S}}^X(\omega_m)$  and the spectral density of the idiosyncratic component,  $\boldsymbol{\mathcal{S}}^\xi(\omega_m) \in \mathbb{R}^{M \times N^2}$ , through the dynamic principal component analysis (DPCA) in [31], where  $\omega_m$  is the discretized frequency for  $m = 1, \dots, M$ , so it is defined as  $\omega_m = 2 * m / (24 * 60)$ .

By following the spectral density definition,  $\boldsymbol{\mathcal{S}}^X(\omega_m)$  can be estimated by the Fourier transform of the covariance matrix of the observation data



$\Phi^X \in \mathbb{R}^{T \times N^2}$ . Therefore, the decomposition of  $\mathcal{S}^X(\omega_m)$  can be processed by decomposing the  $\Phi^X$ , and the appropriateness of the decomposition is supported by the assumption in (4.6). The  $\Phi^X$  is decomposed as:

$$\begin{aligned}
\Phi^X &= E[\mathbf{X}_t \mathbf{X}_{t-k}] \\
&= E[\boldsymbol{\chi}_t \boldsymbol{\chi}_{t-k}] + E[\boldsymbol{\xi}_t \boldsymbol{\xi}_{t-k}] \\
&= \Phi^\chi + \Phi^\xi,
\end{aligned} \tag{4.8}$$

where  $k$  is the time lag,  $\Phi^\chi$  is the covariance matrix of  $\boldsymbol{\chi}$ , and  $\Phi^\xi$  is the covariance matrix of  $\boldsymbol{\xi}$ . The covariance matrix  $\Phi^X$  is estimated by calculating the correlation functions of all pairs of two time series. Since the number of time series is  $N$ , there are  $N^2$  correlation functions including the auto-correlation functions, so the column dimension of  $\Phi^X$  is  $N^2$ . It should be noted that  $\mathcal{S}^X(\omega_m)$  has the CPSD to keep the phase information among time-series, because that  $\Phi^X$  contains the cross-covariance function between different time series and because the Fourier transform of the cross-covariance function is the CPSD.

The assumption in (4.6) and thus the decomposition in (4.8) can be realized by splitting the eigenvalues of  $\mathcal{S}^X(\omega_m)$  through the DPCA in the frequency domain, because eigenvectors corresponding to different eigenvalues are orthogonal to each other. For the given frequency  $\omega_m$ ,  $\mathcal{S}^X(\omega_m) \in \mathbb{R}^{1 \times N^2}$  can be reshaped as  $\mathbf{Y} \in \mathbb{R}^{N \times N}$ . Then, the  $Q$  largest eigenvalues of  $\mathcal{S}^X(\omega_m)$  and their eigenvectors consist of  $\mathcal{S}^X(\omega_m)$ , and the rest of the smaller eigenvalues

and their eigenvectors consist of  $\mathbf{S}^\xi(\omega_m)$  as:

$$\begin{aligned}\mathbf{S}^X(\omega_m) &= \mathbf{V}^X(\omega_m)\mathbf{\Omega}^X(\omega_m)\mathbf{V}^X(\omega_m)^* \\ &\quad + \mathbf{V}^\xi(\omega_m)\mathbf{\Omega}^\xi(\omega_m)\mathbf{V}^X(\omega_m)^* \\ &= \mathbf{S}^X(\omega_m) + \mathbf{S}^\xi(\omega_m),\end{aligned}\tag{4.9}$$

where  $\mathbf{V}^X(\omega_m) \in \mathbb{R}^{N \times Q}$  is a matrix of eigenvectors corresponding to the  $Q$  largest eigenvalues,  $\mathbf{\Omega}^X(\omega_m) \in \mathbb{R}^{Q \times Q}$  is a diagonal matrix of  $Q$  largest eigenvalues,  $\mathbf{V}^\xi(\omega_m)$  is a matrix of the remaining eigenvectors, and  $\mathbf{\Omega}^\xi(\omega_m)$  is a diagonal matrix of the remaining eigenvalues per  $\omega_m$ . It should be noted that  $*$  represents the conjugate transpose but that  $T$  represents the transpose of a matrix. This decomposition process continues for all  $\omega_m$ .

#### 4.3.4 Estimation

The time series  $\mathbf{f}$  and  $\boldsymbol{\chi}$  can be estimated by the linear combination of  $\mathbf{X}$  through (4.2) and (4.3). When  $\boldsymbol{\chi}$  of rank  $Q$  is estimated, without loss of generality, the sum of the auto-covariances of  $\boldsymbol{\xi}$  is minimized.  $\boldsymbol{\xi}_t$  is defined as

$$\boldsymbol{\xi}_t = \mathbf{X}_t - \mathbf{A}(L)\mathbf{B}(L)\mathbf{X}_t.\tag{4.10}$$

Since the sum of the auto-covariances of  $\boldsymbol{\xi}$  is the sum of the diagonal terms of  $\Phi^\xi$ , the problem of minimizing the sum of auto-covariances can be described as

$$\underset{\Phi^\xi}{\text{minimize}} \quad \text{tr}(\Phi^\xi),\tag{4.11}$$

where  $\Phi^\xi$  is defined as

$$\Phi^\xi = \mathbb{E}[(\mathbf{X} - \mathbf{A}(L)\mathbf{B}(L)\mathbf{X})(\mathbf{X} - \mathbf{A}(L)\mathbf{B}(L)\mathbf{X})^T].\tag{4.12}$$

Since the trace is the same as the sum of the eigenvalues, the solution of (4.11) is to select the smallest  $N - Q$  eigenvalues of  $\Phi^X$  when the eigenvalues of  $\Phi^X$  are sorted in ascending order. Therefore, the obvious solution is to select the  $Q$  largest eigenvalues of  $\Phi^X$  for  $\Phi^X$  and the rest of the eigenvalues for  $\Phi^\xi$ . Likewise, in the DPCA, minimizing the largest eigenvalues of  $\mathcal{S}^\xi(\omega_m)$  per  $\omega_m$  will also minimize (4.11) [56].

$\mathbf{A}(L)$  and  $\mathbf{B}(L)$  can be estimated through a similar process as the static principal component analysis (SPCA). In the SPCA,  $\mathbf{A}$  without time lags is given as  $\mathbf{\Lambda}^X \in \mathbb{R}^{N \times Q}$ , which is the  $Q$  largest eigenvectors of  $\Phi^X$ , and  $\mathbf{B}$  without time lags is given as  $(\mathbf{\Lambda}^X)^T \in \mathbb{R}^{Q \times N}$  [183]. Similarly, in the DPCA,  $\mathbf{A}(L)$  is given as  $\mathbf{V}^X(L) \in \mathbb{R}^{N \times Q \times 2S+1}$ , which is the inverse Fourier transform of  $\mathbf{V}^X(\omega_m)$  per  $\omega_m$ , and  $\mathbf{B}(L)$  is given as  $(\mathbf{V}^X(L))^T \in \mathbb{R}^{Q \times N \times 2S+1}$ , which is the inverse Fourier transform of  $\mathbf{V}^X(\omega_m)^*$  per  $\omega_m$  [31]. It should be noted that the time lag is limited to  $S$  because of finite data lengths although the number of discrete frequencies is  $M(M > S)$ . Finally,  $\chi$  can be estimated as:

$$\chi_t = \mathbf{V}^X(L)(\mathbf{V}^X(L))^T \mathbf{X}_t. \quad (4.13)$$

This estimation creates a causality problem when the GDFM is used to forecast the future values because  $L$  represents the time lags. This problem leads to the discussion in [73]. However, ignoring the causality might not cause a problem in this dissertation since the GDFM is used to synthesize load and wind power scenarios.

The next step is to represent  $\chi_t$  as the function of  $\delta_t$ . If (4.4) is com-

bined with (4.2), then  $\boldsymbol{\chi}_t$  can be represented as a function of  $\boldsymbol{\delta}_t$  as (4.7). In order to generate  $\boldsymbol{\chi}$ ,  $\boldsymbol{f}$  is calculated using  $\boldsymbol{C}(L)$  for the given  $\boldsymbol{\delta}$ . Then,  $\boldsymbol{\chi}$  is calculated through (4.2).

Here,  $\boldsymbol{C}(L)$  is estimated by the Yule-Walker equation. Since  $N > Q$ , the variance matrix of  $\boldsymbol{\epsilon}_t$  in the VAR becomes singular, which means that the covariance matrix does not have an inverse matrix. For the singular process, the Yule-Walker equation provides the best stable solution [54]. The stability condition implies that the model can be converted to a moving average form of dynamic shocks because of the Wold theorem. This condition is important in this study because all scenarios are derived from white noise. Besides, the Yule-Walker equation is faster than other methods.

#### 4.4 Load and wind power Scenario Generation

The process to synthesize scenarios is briefly summarized. First,  $\boldsymbol{\chi}$ ,  $\boldsymbol{\xi}$ , and  $\boldsymbol{f}$  are estimated from  $\boldsymbol{X}$ . Second,  $\boldsymbol{f}$  is modeled as the VAR process, and  $\boldsymbol{\epsilon}$  and  $\boldsymbol{Z}$  are estimated simultaneously. Third,  $\boldsymbol{\delta}$  is generated under the assumption that it is normally distributed. Through the (4.13),  $\hat{\boldsymbol{\chi}}$  is built, and  $\hat{\boldsymbol{X}}$  is synthesized by adding  $\hat{\boldsymbol{\xi}}$ , which is synthesized by following the multivariate normal distribution with the covariance matrix of  $\boldsymbol{\xi}$ . Finally,  $\hat{\boldsymbol{X}}$  is scaled up to its original magnitude, and the periodic component is added.

#### 4.4.1 Number of Dynamic Factors

Before synthesizing load and wind power, the penalized information criterion (IC) is used to decide  $Q$ . The ICs, such as the Akaike information criterion (AIC) and Bayesian information criterion (BIC), have been widely used in time-series applications to obtain the optimal number of orders. The criterion to decide  $Q$  is set based on [11]. When the sum of idiosyncratic errors and penalties on the number of variables is minimized,  $Q$  is decided. The IC is given as:

$$IC(P) = \log\left[\frac{1}{NT} \sum_n^N \sum_t^T (\xi_{nt})^2\right] + P \times p(N, T), \quad (4.14)$$

where the penalty function  $p(N, T)$  is given as

$$p(N, T) = k \frac{\log(\min[N, T])}{NT/(N + T)}. \quad (4.15)$$

It is found that 18 is the optimal number of dynamic factors.

#### 4.4.2 Scenario Generation

First, the  $\hat{\chi}$  is estimated. For example, the  $\hat{\chi}$  of load in the Coast area is plotted in Figure 4.1. Although the  $\hat{\chi}$  of the load follows the diurnal periodicity of the actual load clearly, it does not explain idiosyncratic movements. Load waveforms follow the weekly trends too. For example, the 41st day in Figure 4.1 corresponds to February, 10, 2013, which was a Sunday. Furthermore, the  $\hat{\chi}$  of wind power is estimated and plotted in Figure 4.2. It can be seen that there was a sudden wind power die off on the 39th day in 2013.

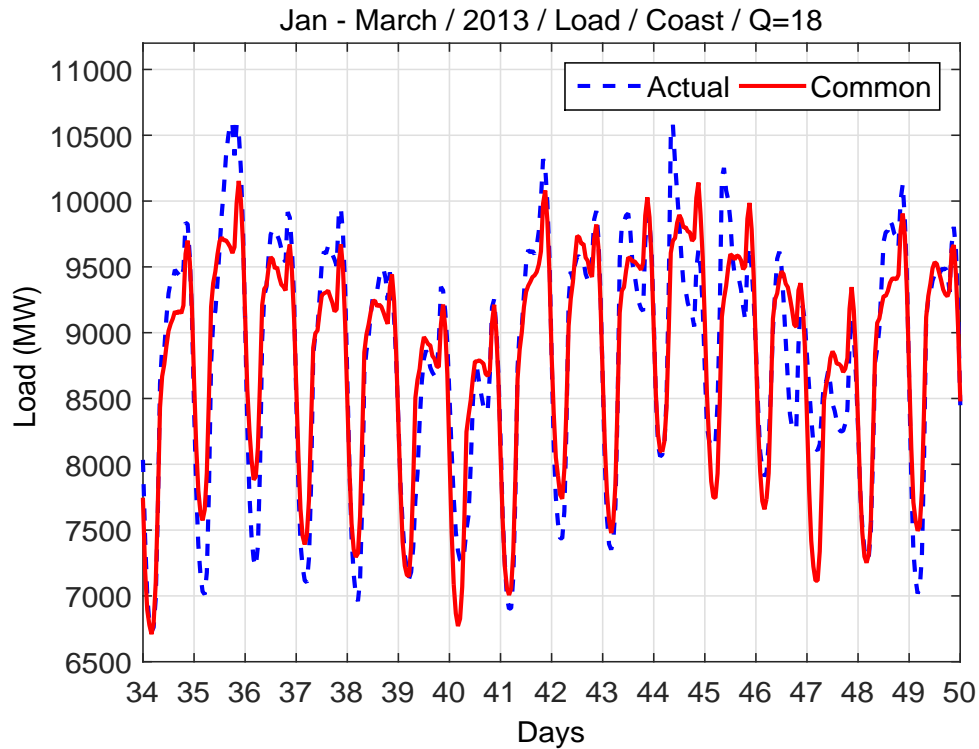
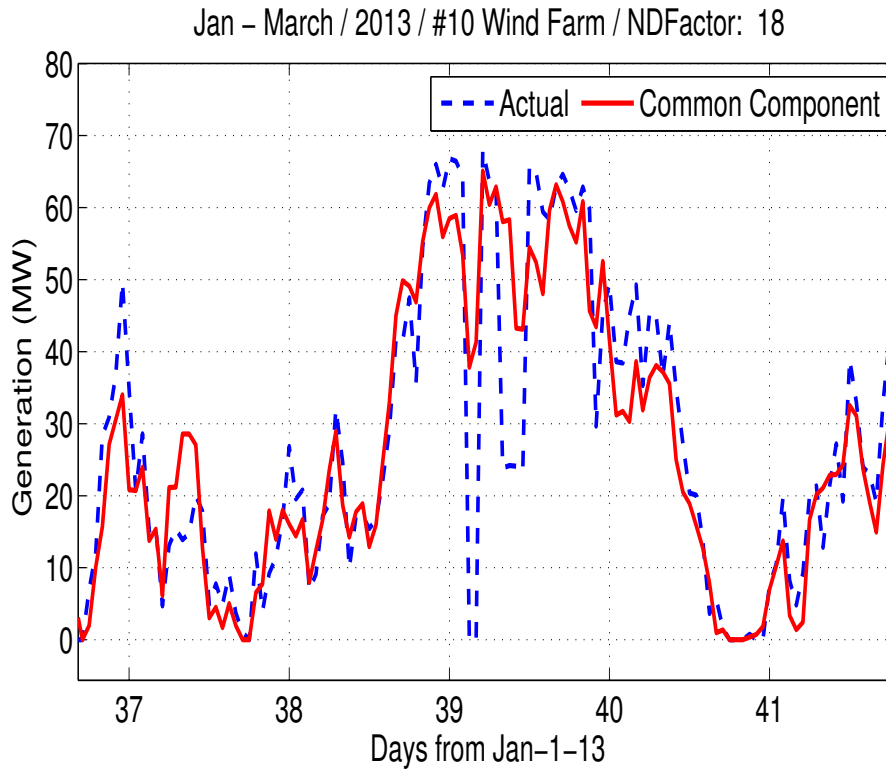


Figure 4.1: Coastal load data and common component. The blue dashed lines represent actual data, and the red lines represent the common component. The actual load and common component of load are plotted.

Considering that the duration was very short and that wind power around the 39th day was very high, it can be assumed that the wind farm was shut down when wind speed went over the cut-out speed. Since it is an idiosyncratic event, the  $\hat{\chi}$  of wind power does not model it.

Second, scenarios are synthesized and plotted in Figure 4.3 and Figure 4.4. The load scenario in the Coast area is plotted in Figure 4.3. It follows the diurnal periodicity clearly, but it also has more severe peaks and bottoms on different days. It is observed that the weekly trend in the actual load, and



*Figure 4.2: The blue dashed lines represent actual data, and the red lines represent the common component. The actual and common components of the 10th wind farm are plotted. Its capacity is 79.3 MW.*

the synthesized load also closely follows it. The load on weekends is relatively lower than the load on weekdays. For example, the 68th day in Figure 4.3 corresponds to March, 9, 2013, Saturday, and the load on that day was relatively lower than the load on other days. The wind power scenario on the 10th wind farm is also synthesized and plotted in Figure 4.4. The wind power scenario has similar ramp rates, maximum values, minimum values, and overall shape to the actual scenario. The scenarios are curtailed at minimum and maximum values in the post processing. Many zero values in both the actual wind power

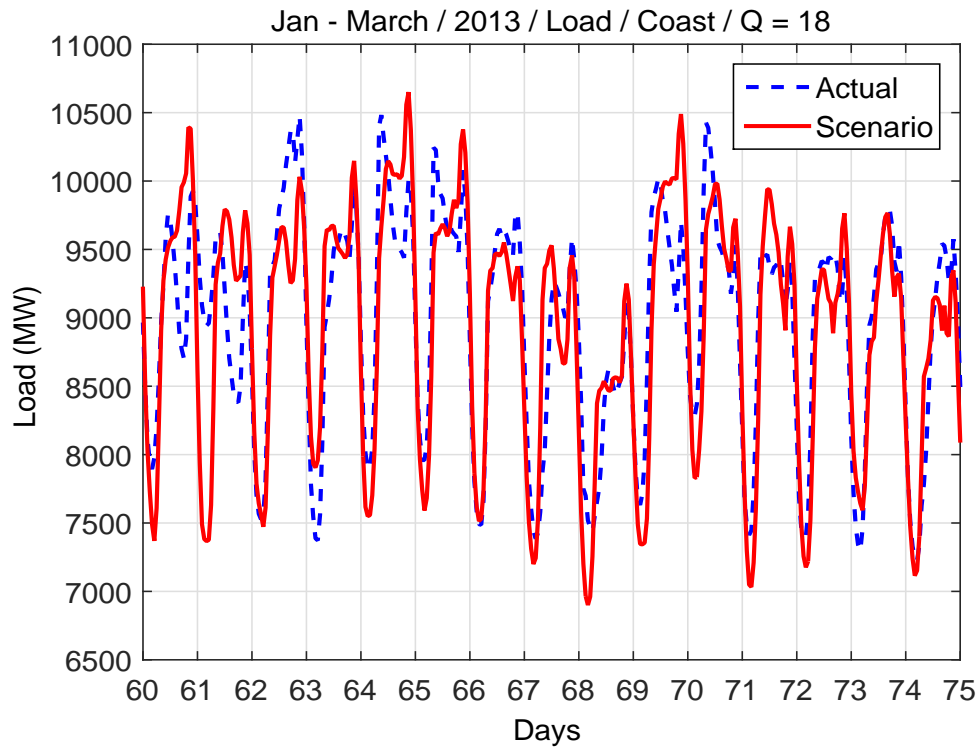
and wind power scenario can be observed.

#### 4.4.3 Statistics & PSD Analysis

Statistics of the scenarios are also compared to actual statistics in Table 4.1. Analyzed statistics are mean, median (MED), standard deviation (STD), coefficient of variation (CV), variability (V), maximum (Max), minimum (Min), 25% value, 75% value, kurtosis (Kur), and skewness (Skw). The variability is defined as the STD of the hour to hour difference between data. Furthermore, the difference of wind power scenarios follows the Laplace distribution just as actual wind power does [189]. Since there are many time series, averaged statistics of scenarios are compared to averaged statistics of actual waveforms, and the normalized root mean square errors (NRMSE)s are also estimated in percentage to measure relative errors. Load and wind power scenarios are separated and averaged because they have different scales. Statistics of the actual waveforms and scenarios are very similar. Although scenarios satisfy most statistics, load scenarios have significant errors in variability and skewness.

In order to verify the periodic component and temporal correlation of an individual scenario, the PSDs of monthly scenarios are measured through the multi taper algorithm and compared to actual PSDs in Figure 4.5 and Figure 4.6. The PSDs of the load scenario and actual load in Figure 4.5 have similar periodic components. Since the periodic component is the sum of the harmonics of diurnal periodicity, the PSD of load scenarios has similar PSD





*Figure 4.3: The blue dashed lines represent actual data, and the red lines represent the final synthesized scenarios. The actual load and load scenario are plotted. It is observed that the load scenario not only follows the diurnal and weekly periodicities but also represents sudden ramps because of weather changes.*

peaks to the PSD of the actual load. Furthermore, the PSD of the wind power scenario is plotted in Figure 4.6. It has a strong amplitude in the 24h period. The PSD of wind power also closely follows the overall shape of the PSD of actual wind power. That is, both PSDs of the wind power scenario and the actual wind power have low-pass filtered waveforms, which are typical AR processes. Through the PSDs, it is also verified that the AR process is a reasonable assumption. In addition, the stability of multivariate time series is

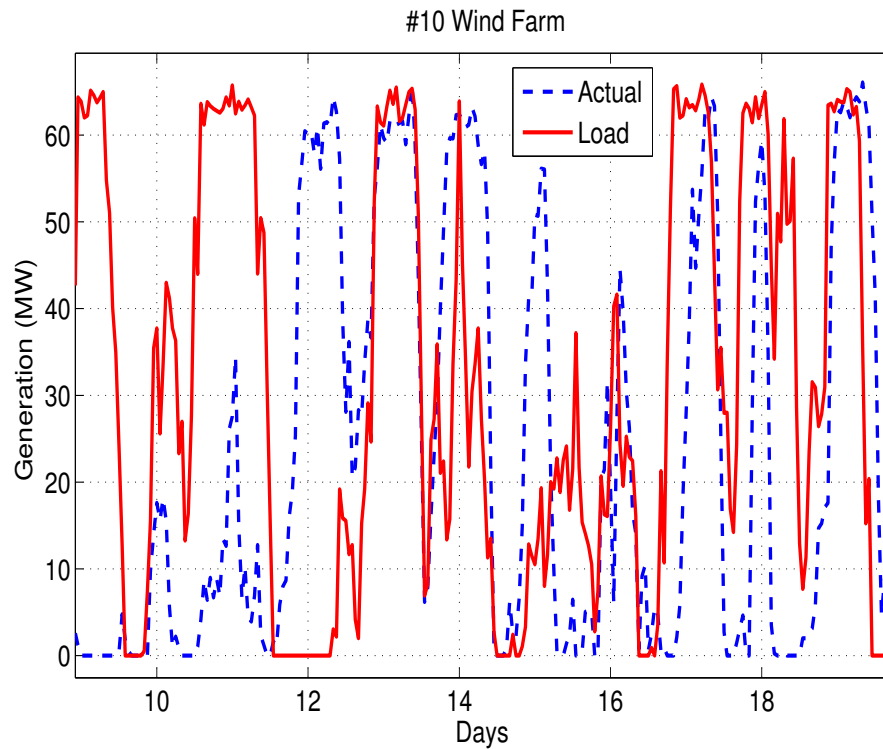


Figure 4.4: The blue dashed lines represent actual data, and the red lines represent the final synthesized scenarios. The actual wind power and wind power scenario of the 10th wind farm are plotted.

checked by testing the eigenvalues of the block companion matrix of  $\mathbf{C}(L)$ . The modulus of eigenvalues should be less than one, and the maximum eigenvalue was 0.971.

#### 4.4.4 Correlation Coefficient

Correlation coefficients of actual waveforms and synthesized scenarios are plotted on shaded surfaces in Figure 4.7 and Figure 4.8 in order to see the relationship among time series. The darkest red in a diagonal represents the

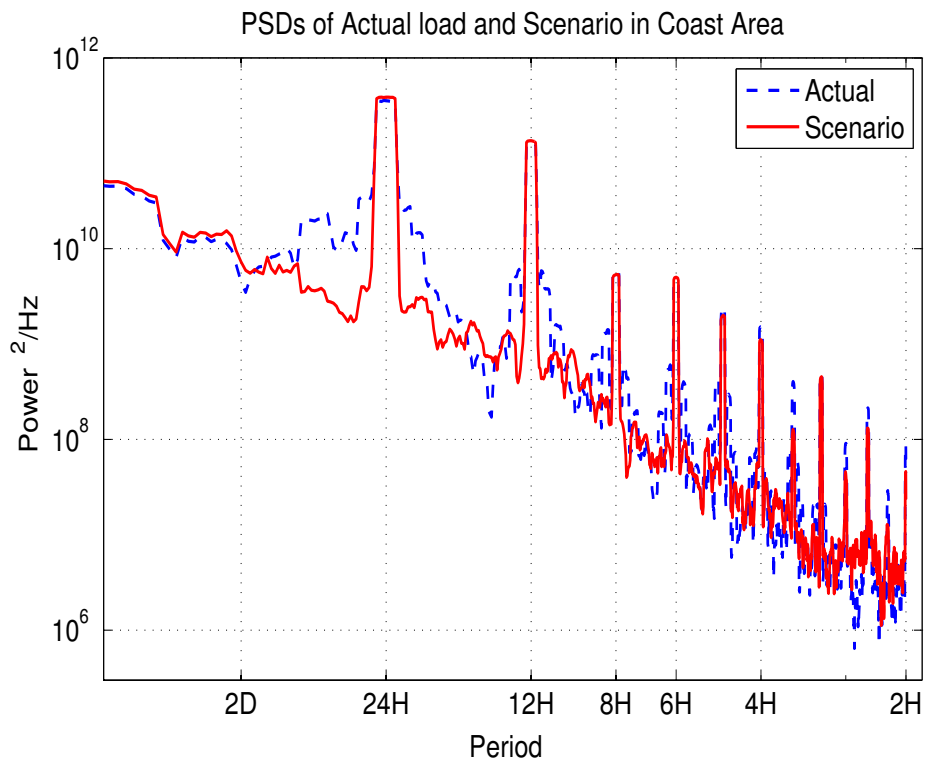


Figure 4.5: The blue dashed lines represent the PSDs of the actual data, and the red lines represent the PSDs of scenarios. The PSDs of actual and synthesized load are plotted. They have a similar strength as the PSDs of the harmonics of the 24th hour period.

most positive correlation, one. The darkest blue represents the most negative correlation. Among 104 time series, the last eight time series are the load. Figure 4.7 shows the correlation coefficients of actual waveforms. It is observed that load and wind power generally have a negative correlation. However, wind farms around the 70th are negatively correlated with other wind farms, but they are slightly positively correlated with the load. The latitude of 70th wind farm is 26N, and the longitude is 97W, so it is located on the Gulf of

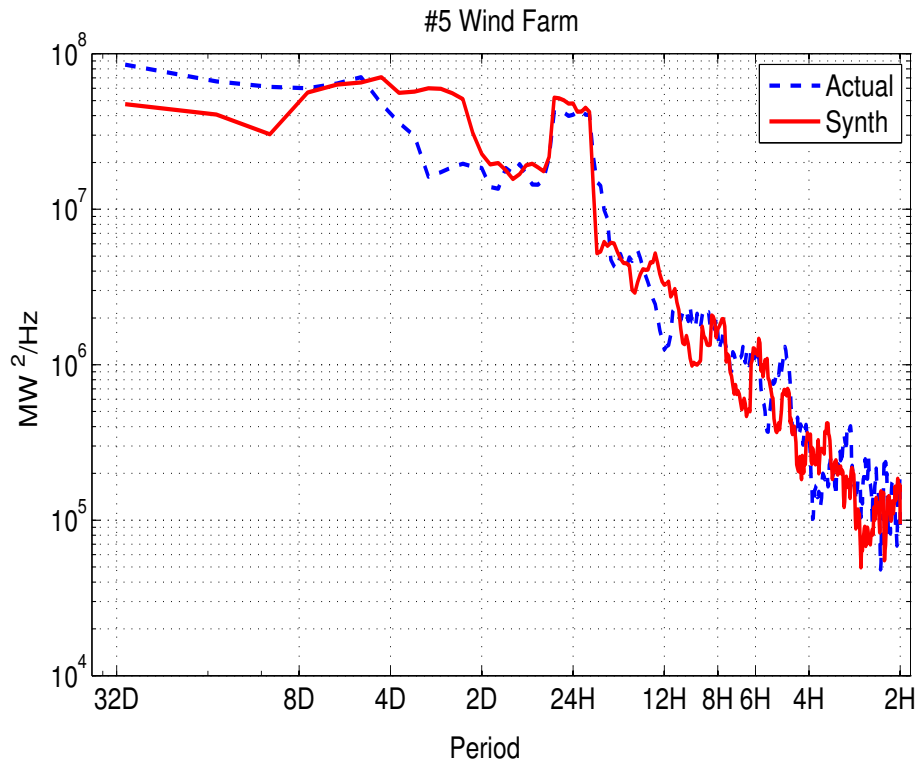


Figure 4.6: The blue dashed lines represent the PSDs of the actual data, and the red lines represent the PSDs of scenarios. The PSDs of actual and synthesized wind power are plotted. They have a similar strength as the PSDs of the 24th hour period and time periods when there are low frequencies.

Mexico. It is shown that offshore wind farms are positively correlated to load. Similarly, Figure 4.8 shows the correlation coefficients of scenarios. The shaded surfaces in Figure 4.7 and Figure 4.8 are similar to each other although some wind power scenarios are more positively correlated to each other than actual wind power outputs are, and some offshore wind farms are more positively correlated to load. On the contrary, although it is not shown here, scenarios that are synthesized from individual time series processes are uncorrelated to

Table 4.1: Evaluation of Statistical Characteristics

STAT	Wind			Load		
Index	Actual	Scen	NRMSE	Actual	Scen	NRMSE
Mean	42.11	42.63	1.721	4.07e3	4.04e3	0.765
Med	34.37	34.42	0.141	4.05e3	4.05e3	0
STD	35.05	35.75	2.318	6.06e2	5.69e2	4.669
CV	0.856	0.861	0.595	0.144	0.138	3.916
V	14.96	13.54	9.425	1.13e2	1.03e2	2.432
Max	109.2	108.9	0.195	6.45e3	6.05e3	5.061
Min	0	0	0	2.82e3	2.63e3	6.709
25%	9.630	10.03	12.34	3.68e3	3.65e3	0.894
75%	72.16	72.44	0.377	4.41e3	4.37e3	0.719
Kur	1.893	1.929	2.473	3.561	3.454	3.047
Skw	0.464	0.464	2.799	0.585	0.478	30.88

each other, so the shaded surfaces are filled with light blue. Therefore, it can be said that the GDFM can synthesize correlated scenarios.

It is also observed that the periodic component and stochastic process have different effects on scenarios. When the load scenario is synthesized, the periodic component has greater effects than the stochastic process. As shown in Figure 4.5, the periodic component of load consists of multiple harmonics of diurnal periodicity. When the wind power scenario is synthesized, the stochastic process has greater effects than the stochastic process has. As shown in Figure 4.6, the low-pass filtered PSD is generated by the AR process, but the

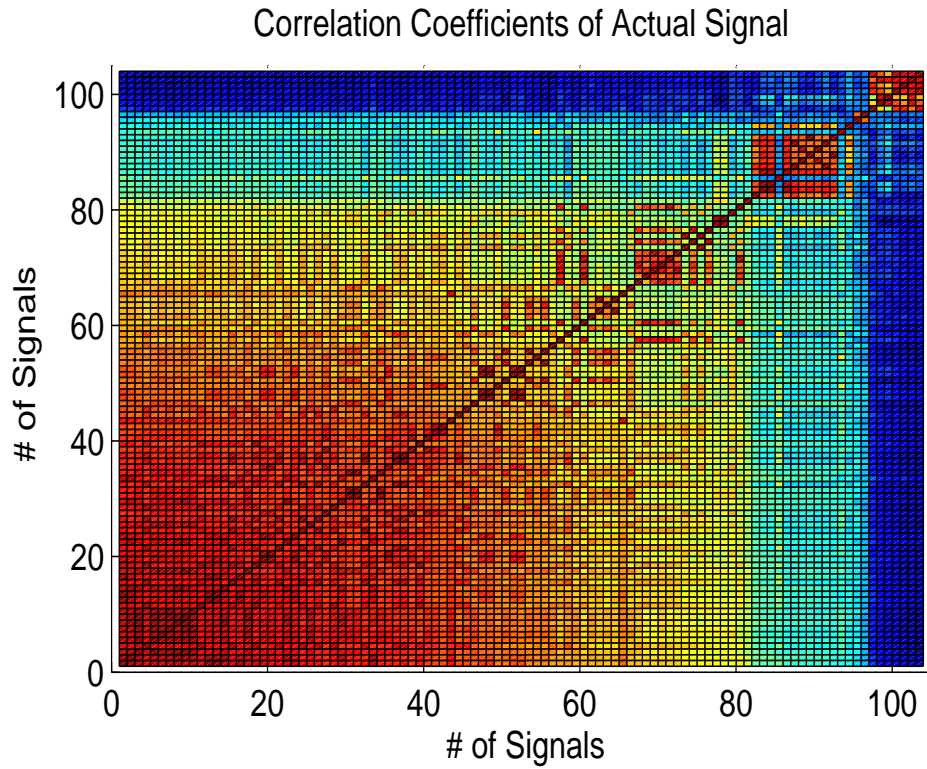


Figure 4.7: The correlation coefficients of actual waveforms are plotted in shaded surfaces. Red shifted colors represent a more positive correlation, and blue shifted colors represent a more negative correlation. The diagonal line represents a perfect correlation.

periodic component generates only a peak value at the 24th hour. For correlation coefficients in Figure 4.7 and Figure 4.8, scenarios that are synthesized using only the stochastic processes create better correlation coefficients than those created using only periodic components. Because periodic components are similar to each other, it is difficult to make different and non-zero correlation coefficients using only periodic components, so the stochastic component is required to have different and non-zero correlation coefficients.

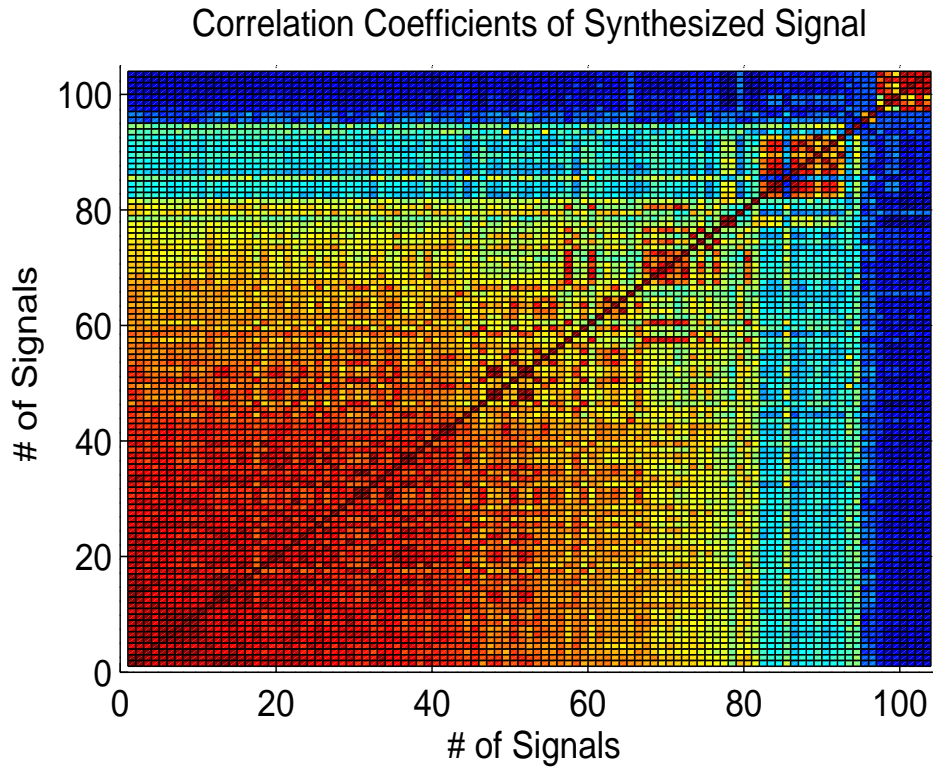


Figure 4.8: The correlation coefficients of synthesized scenarios are plotted in shaded surfaces. Red shifted colors represent a more positive correlation, and blue shifted colors represent a more negative correlation. The diagonal line represents a perfect correlation.

## 4.5 Generation & Transmission Upgrading Costs

In this section, the GDFM scenarios are utilized in a transmission and generation planning problem. The generation and transmission upgrading costs of scenarios of the GDFM are compared to those of uncorrelated, correlated, and actual scenarios. The generation and transmission upgrading costs are calculated by solving the DC optimal power flow (OPF) on the IEEE 300-bus benchmark. The benefits of transmission projects are determined in this

case study by measuring the short-run marginal production costs. It is also observes the effects of the correlation structure in load and wind power to the operation and planning costs.

#### 4.5.1 Simulation Settings

The generation cost is calculated by solving the linear programming, and the upgrading cost is calculated by solving the linear mixed integer programming (LMIP). It is assumed that only existing lines are upgraded, and building new wind farms and transmission lines is not considered. The upgrading cost is levelized as an investment cost for three months. The costs do not include the cost of ancillary services, value of lost load, cost of wind power curtailment, or congestion revenue rights. It should be noted that the LMIP is decomposed by Bender's decomposition because the LMIP has a huge number of variables.

Five different scenarios are generated. First, 104 uncorrelated random signals are generated. Second, uncorrelated total load and wind power outputs are synthesized and equally distributed on buses. The third scenarios are the same as the second scenarios except that total load and wind power outputs are negatively correlated. The correlation coefficient is -0.2. The fourth scenarios are generated by the GDFM, and the fifth scenarios are actual measurements. In order to consider the high penetration level of wind power, wind power is scaled up to double. Furthermore, in order to magnify the transmission expansion, load buses do not have generators. It should be noted that scenarios



are normalized by the actual measurements so that all scenarios have the same amounts of wind and load.

#### 4.5.2 Simulation Results

The linear mixed integer model is as follows.

$$\min_x \sum_{(p,q) \in \mathbf{\Lambda}} L_{(p,q)} x_{(p,q)} + \sum_{t=1}^{720} \pi_t \sum_{g \in \mathbf{G}} G_g(P_g^t) \quad (4.16)$$

subject to

$$P_p^t - \sum_{q \in B(p)} X_{p,q} (\theta_p^t - \theta_q^t) = D_p^t \quad \forall p \in \mathbf{B} \quad (4.17)$$

$$-C_{(p,q)}(x_{(p,q)} + 1) \leq X_{(p,q)}(\theta_p^t - \theta_q^t) \quad \forall (p,q) \in \mathbf{\Lambda} \quad (4.18)$$

$$C_{(p,q)}(x_{(p,q)} + 1) \geq X_{(p,q)}(\theta_p^t - \theta_q^t) \quad \forall (p,q) \in \mathbf{\Lambda} \quad (4.19)$$

$$N_g \leq P_g^t \leq M_g \quad \forall g \in \mathbf{G}, \forall t \in \mathbf{T} \quad (4.20)$$

For line  $(p, q)$ , where  $p$  and  $q$  are buses,  $L_{(p,q)}$  is an upgrading cost,  $C_{(p,q)}$  is the line capacity, and  $X_{p,q}$  is the line admittance. For bus  $p$ ,  $P_p$  is the generation,  $D_p$  is the load, and  $\theta_p$  is the bus angle. For the generator  $g$ ,  $G_g$  is the cost function,  $P_g$  is the generation,  $N_g$  is the generator minimum output, and  $M_g$  is the generator maximum output. Furthermore, the probability  $\pi_t$  is set at one to compare the total generation cost to the upgrading cost.

Simulation results that are averaged over ten runs for 720 hours are in Table 4.2. For the generation cost, the first scenarios have the lowest cost. The GDFM scenarios can be verified since the generation cost of the GDFM

scenarios is similar to that of the actual scenarios. For the second and third scenarios, the total load and wind power are equally distributed on the buses, so sample paths have a perfect correlation. The generation costs of the second and third scenarios are similar and most expensive since all wind power are perfectly correlated. Furthermore, the generation cost of the third scenario is more expensive than the second scenario, since the wind power in the second scenario is negatively correlated with the load in the second scenario.

For transmission upgrading, the sum of upgrading and operation costs is smaller than the generation cost without the transmission upgrading for all scenarios. The GDFM and actual scenarios have similar upgrading and operation costs. These scenarios have the most expensive transmission upgrading cost since actual scenarios have extreme wind power and loads due to extreme weather. By considering that these costs are the optimal solutions of stochastic optimization, for actual scenarios, more money should be invested in the transmission lines more in preparation for extreme cases. In short, the correlation among individual scenarios affects the generation and upgrading costs more than the correlation between the total load and wind power. For more realistic scenarios, wind power is more dynamic and has more peaks, so the peaking characteristic should be reflected on generation and transmission costs.

*Table 4.2: Results of Economic Dispatch and Transmission Expansion*

Scenario Number	Generation Cost	Transmission upgrading cost	
		Upgrading	Operation
Purely randomly generated scenarios	\$3.6157 M	\$ 0.018 M	\$ 3.459 M
Total load and wind power scenarios are uncorrelated	\$3.7784 M	\$ 0.122 M	\$ 3.108 M
Total load and wind power scenarios are negatively correlated	\$3.7974 M	\$ 0.113 M	\$ 3.114 M
GDFM scenarios	\$ 3.7153 M	\$ 0.189 M	\$ 2.150 M
Actual measured scenarios	\$ 3.7180 M	\$ 0.187 M	\$ 2.098 M

## 4.6 Conclusion

In this study, load and wind power scenarios are synthesized through the GDFM, in which observation data is decomposed into the periodic, common, and idiosyncratic components. The common component consists of dynamic factors and dynamic factor loadings, and dynamic factors are derived from the dynamic shocks, which are white noise terms, through the VAR process. Since the dimension of the dynamic factor is less than that of the observation data, the number of variables and parameters is reduced. The study also verifies that scenarios have satisfactory statistical characteristics, PSDs, correlation coefficients, generation, and transmission upgrading costs. Furthermore, the GDFM scenarios have similar simulation results with the actual scenarios, as demonstrated in an generation and transmission upgrading case study.

## Chapter 5

### Conclusion

In this dissertation, the short-term probabilistic and ensemble wind power forecasting algorithm is proposed to overcome the wind power uncertainty. Then, the novel algorithm to generate many wind power scenarios based on the PSD analysis of wind power is also proposed to use scenarios to develop methods for mitigating the wind power variability. Furthermore, it is beneficiary to build the transmission expansion planning based on load and wind power scenarios, which are synthesized through the GDFM by keeping their stochastic correlation structure. In this chapter, these studies are summarized, and additional work that can be done in future work is enumerated.

#### 5.1 Short-Term Wind Power Forecasting

Wind power fluctuation prevents the participation of wind farm owners in the AS market and the DA market where prices are less volatile and higher than the RT market. In order to promote participation of wind farm owners in the AS and DA markets, wind power and its error distribution are forecasted through novel probabilistic and ensemble forecasting algorithms in the first chapter. The performance of this algorithm is verified through the

data in the wind power competition that was organized by the PES. For the probabilistic forecasting, the error distribution is assumed to follow parametric distributions in the closed form. Training data is clustered in a few clusters, and the error distribution is estimated per cluster in order to reduce the STD of the distribution. Furthermore, error distributions are truncated by the minimum and maximum of the actual values of the point forecast's corresponding cluster.

### 5.1.1 Key Results

The tree-based forecasting machines, such as the RF, GBM, and BAG, show the best performance among the seven forecasting machines although it takes a lot of time to train them. The second best are the kernel-based forecasting machines, such as the NN and SVM. The performance of the GP varies with respect to the data set, and it takes the longest computation time. For the ensemble forecasting, the weight averaging outperforms the simple averaging. Furthermore, fitting the model of the error distribution with the VG distribution outperforms the model of the error distribution with other distributions. In this process, the parametric approach with clustering data in ten clusters shows the best performance. Finally, the discrete STD estimation function with data clustering outperforms the continuous STD estimation function without data clustering.

### 5.1.2 Future Work

Five unsolved issues with this algorithm are enumerated. *a)* The performance of the parametric approach should be compared to the performance of the nonparametric approach. *b)* A single step process to estimate the error distribution in the ensemble forecasting should be developed. Although the parametric approach needs additional steps to estimate the error distribution, forecasting machines can be trained to estimate distributions directly without forecasting point forecasts. *c)* For the GP, a better combination of kernels that are fitted to the wind power forecasting. should be found. *d)* A representative data set that works as the validation data in the cross validation should be selected to reduce the cross validation computation time. The representative data should have all statistical characteristics of actual data. *e)* One-hour-ahead point forecast can be used to forecast the wind power at the given time in order to utilize the temporal correlation structure in wind power outputs since wind power strongly depends on hour ago wind power. If these future work are accomplished, the performance of probabilistic wind power forecasting will increase and participation of wind power owners in the AS and DA markets will be encouraged.

## 5.2 Long-Term Wind Power Scenario Generation

In order to develop smart AS procurement methodologies, the power system should be simulated with many future wind power scenarios. Therefore, in the second Chapter, scenarios of near future wind power are synthesized by

considering the PSD, statistical characteristics, and the future capacity. The PSD of the wind power follows different power laws over different frequency ranges and is approximated by a piecewise function. A scaling exponent of the power law for a particular piece can be approximated by the slope of an affine function fitted to a logarithmic plot of the PSD. Each piece of the function has a different trend as the total capacity increases. Slope trends, the first PSD value, and the last PSD value are regressed to forecast the PSD. Then, future wind power scenarios are synthesized from the forecasted PSD. In this process, phase angles are searched using a genetic algorithm while satisfying forecasted statistical characteristics for the given capacity. This approach is simulated and validated for wind power for seven years in ERCOT and is used to synthesize a future wind power scenario at 10,000 MW capacity.

### **5.2.1 Key Results**

As the penetration level of wind power increases, the absolute magnitude of wind power variability increases, but the relative magnitude of wind power variability with respect to the installed capacity decreases. The dissipation ratio of wind power variability corresponding to the low and high frequency ranges does not change as the installed capacity of wind power increases. On the contrary, the dissipation ratio of wind power variability corresponding to the middle frequency range increases, indicating that the relative amplitudes of wind power variability corresponding to the middle frequency range decrease as the installed capacity of wind power increases. Furthermore,

the PSD slope of the third segment decreases and converges to a certain value. This means that the relative amplitude of wind power variability in the middle frequency range will be fixed at a very high penetration level. Therefore, the absolute amplitude of wind power variability increases as the penetration level of wind power increases. This means that although the wind power fluctuation decreases because of the geographical smoothing effects, certain amounts of wind power variability always remain in the power system, and the absolute magnitude of that wind power variability increases continuously as the installed capacity of wind power increases.

### **5.2.2 Future Work**

Instead of using the GA to search for phase angles, the possibility of modeling phase angles will be investigated by analyzing their stochastic structure. The goal of the future work is to design the stochastic structure of phase angles of the PSD following the power law so that the power differences of the transformed waveform have a certain distribution. The stochastic structure of phase angles of wind power should be able to generate arbitrarily many sample phase angles by changing noise terms as generated phase angles satisfy the distribution of the power difference. Characteristics of phase angles can be inferred from the Laplace distribution of wind power differences. In order for the Laplace distribution to have a heavy tail there must be few big power differences, but many small power differences. A big power difference can be generated by assigning similar phase angles to sinusoidal components of ad-



adjacent frequencies in low frequency ranges so that the sum of the derivatives of sinusoidal components becomes large enough to have a big power difference. Furthermore, big power differences that are generated by the sinusoidal components of adjacent frequency ranges are rare, so the collection of those power differences from all sinusoidal components can have a heavy tail. At the same time, randomly generated phase angles of sinusoidal components in high frequency ranges generate a lot of small power differences. Therefore, a better, systematic method for estimating phase angles can synthesize total wind power scenarios more realistically.

### **5.3 Load and Wind Power Scenario Generation and Transmission Expansion Planning**

The direct way to increase the penetration level of wind power is to build more wind farms and connect them by building more transmission lines. In Chapter 4, therefore, the transmission expansion planning is established based on the load and wind power scenarios that are synthesized through the GDFM and verified through statistics, spectral density analysis, and correlation analysis. The GDFM preserves the correlation structure between load and wind by representing load and wind power as the sum of the periodic component, common component, and idiosyncratic noise component. The periodic component is considered as the sum of the harmonics of the diurnal periodicity and is estimated by averaging data for the same hour every day. The common and idiosyncratic noise components are estimated after extracting the peri-

odic component. The common component consists of the dynamic factors, or is white noise, and the polynomial matrix, which represents the temporal and geographical correlation between load and wind power. Dynamic factors are derived from the dynamic shocks, which are white noise terms, through the VAR process. Factors, factor loadings, and thus the common component are estimated through the DPCA. Furthermore, an infinite number of scenarios can be synthesized by generating different dynamic shocks. The dimension of the GDFM is less than the dimension of observation data, so the scenarios are generated from a parsimonious system.

### **5.3.1 Key Results**

The usefulness of scenarios is verified by calculating the total generation and transmission upgrading costs on the IEEE 300-bus benchmark. The load and wind power scenarios that are synthesized by considering the stochastic correlation structure between them require higher generation and expansion costs than purely randomly generated scenarios require. If the total load and wind power scenarios are equally distributed to multiple buses, they are perfectly correlated to other scenarios although total load and wind power are negatively correlated. In this case, the generation and transmission investment costs are highest. In short, purely randomly generated scenarios or perfectly correlated scenarios on multiple buses can also have similar simulation results to the simulation results of the actual scenarios. However, in order to have more accurate simulation results when scenarios are synthesized, the stochastic

correlation structure should be considered.

### 5.3.2 Future Work

The transmission expansion planning can be expanded to a problem with scenario trees for multiple stages, and it can be solved through the stochastic dual dynamic programming (SDDP) by considering multiple macro scenarios of gas prices, demand level, and installed capacity of wind power. For the given transmission plan from the previous stage, the transmission plan at the current stage is determined by solving the stochastic economic dispatch with multiple micro scenarios of load and wind power for the given macro scenario. It should be noted that micro scenarios of load and wind power are synthesized through the GDFM. One advantage of the SDDP is to decompose the problem into multiple sub-problems according to stages and to change the main objective function by generating cuts using dual variables in the sub stages. Finally, the usefulness of the SDDP can be verified by calculating the total generation and transmission upgrading costs on a simplified electric network in the ERCOT interconnection. Therefore, the proposed transmission expansion planning should be extended to the multistage transmission expansion planning, which is closer to the actual transmission problem than the proposed one, and the multistage planning can be simulated on a simplified electric network in the ERCOT interconnection to build the more realistic transmission expansion planning.

## Bibliography

- [1] 3TIER. Development of regional wind resource and wind plant output datasets. Technical Report SR-550-47676, National Renewable Energy Laboratory, [www.nrel.gov/docs/fy10osti/47676.pdf](http://www.nrel.gov/docs/fy10osti/47676.pdf), 2010.
- [2] Juan Rivier Abbad. Electricity market participation of wind farms: the success story of the spanish pragmatism. *Energy Policy*, 38(7):3174 – 3179, 2010. Large-scale wind power in electricity markets with Regular Papers.
- [3] U.S. Energy Information Administration. Annual energy outlook 2014.
- [4] A. Agresti and B. Finlay. *Statistical Methods for the Social Sciences*. Pearson Practice Hall, New Jersey, 2009.
- [5] Texas Energy Storage Alliance. Storage participation in ERCOT. Technical report, ERCOT, Jan 2010.
- [6] B. D O Anderson and M. Deistler. Generalized linear dynamic factor models - a structure theory. In *Decision and Control, 2008. CDC 2008. 47th IEEE Conference on*, pages 1980–1985, Dec 2008.
- [7] B. D O Anderson and M. Deistler. Properties of zero-free spectral matrices. *Automatic Control, IEEE Transactions on*, 54(10):2365–2375, Oct 2009.

- [8] Jay Apt. The spectrum of power from wind turbines. *Journal of Power Sources*, 169(2):369–374, 2007.
- [9] Peder Bacher, Henrik Madsen, and Henrik Aalborg Nielsen. Online short-term solar power forecasting. *Solar Energy*, 83(10):1772 – 1783, 2009.
- [10] Viorel Badescu. Verification of some very simple clear and cloudy sky models to evaluate global solar irradiance. *Solar Energy*, 61(4):251 – 264, 1998.
- [11] Jushan Bai and Serena Ng. Determining the number of primitive shocks in factor models. *Journal of Business & Economic Statistics*, 25(1):52–60, 2007.
- [12] Jushan Serena Bai, Jushan. Determining the number of factors in approximate factor models. *Econometrica*, 70:1468 – 0262, 2002.
- [13] R. Baldick. The generalized unit commitment problem. *Power Systems, IEEE Transactions on*, 10(1):465–475, Feb 1995.
- [14] Albert-László Barabási and Réka Albert. Emergence of scaling in random networks. *Science*, 286(5439):509–512, 1999.
- [15] L. Baringo and A.J. Conejo. Wind power investment within a market environment. *Applied Energy*, 88(9):3239 – 3247, 2011.

- [16] Luis Baringo and Antonio J Conejo. Transmission and wind power investment. *Power Systems, IEEE Transactions on*, 27(2):885–893, 2012.
- [17] T.P. Barnett, J. Ritchie, J. Foat, and G. Stokes. On the space–time scales of the surface solar radiation field. *Journal of Climate*, 11(1):88, 1998.
- [18] David J Bartholomew, Martin Knott, and Iriini Moustaki. *Latent Variable Models and Factor Analysis: A Unified Approach*, volume 899. Wiley, West Sussex, United Kingdom.
- [19] Jeffrey Basara, Thomas Hamill, David John Gagne, and Amy Mcgovern. Ams 2013-2014 solar energy prediction contest, July 2013. <http://www.kaggle.com/c/ams-2014-solar-energy-prediction-contest/data>.
- [20] C. Bédard, H. Kröger, and A. Destexhe. Does the  $1/f$  frequency scaling of brain signals reflect self-organized critical states? *Phys. Rev. Lett.*, 97:118102, Sep 2006.
- [21] Ben S. Bernanke, Jean Boivin, and Piotr Eliaszc. Measuring the effects of monetary policy: A factor-augmented vector autoregressive (favar) approach. *The Quarterly Journal of Economics*, 120(1):387–422, 2005.
- [22] Dimitris Bertsimas and John N. Tsitsiklis. *Introduction to Linear Optimization*. Optimization and Neural Computation. Athena Scientific, 3rd edition, Feb. 1997.

- [23] R.J. Bessa, M.A. Matos, I.C. Costa, L. Bremermann, I.G. Franchin, R. Pestana, N. Machado, H. Waldl, and C. Wichmann. Reserve setting and steady-state security assessment using wind power uncertainty forecast: A case study. *Sustainable Energy, IEEE Transactions on*, 3(4):827–836, Oct 2012.
- [24] Christopher M. Bishop. *Pattern recognition and machine learning*. Information Science and Statistics. Springer, 2006.
- [25] Arthur Bossavy, Robin Girard, and George Kariniotakis. Forecasting ramps of wind power production with numerical weather prediction ensembles. *Wind Energy*, 16(1):51–63, 2013.
- [26] Audun Botterud, Jianhui Wang, Vladimiro Miranda, and Ricardo J. Bessa. Wind power forecasting in u.s. electricity markets. *The Electricity Journal*, 23(3):71 – 82, 2010.
- [27] F. Bouffard and F.D. Galiana. Stochastic security for operations planning with significant wind power generation. *Power Systems, IEEE Transactions on*, 23(2):306 –316, May 2008.
- [28] T. Boutsika and S. Santoso. Quantifying short-term wind power variability using the conditional range metric. *Sustainable Energy, IEEE Transactions on*, 3(3):369–378, 2012.
- [29] M. Bouzerdoum, A. Mellit, and A. Massi Pavan. A hybrid model (sarima-svm) for short-term power forecasting of a small-scale grid-connected

- photovoltaic plant. *Solar Energy*, 98, Part C(0):226 – 235, 2013.
- [30] H. Breitzkreuz, M. Schroedter-Homscheidt, and T. Holzer-Popp. A case study to prepare for the utilization of aerosol forecasts in solar energy industries. *Solar Energy*, 81(11):1377 – 1385, 2007.
- [31] David R. Brillinger. *Time Series: Data Analysis and Theory*. SIAM, San Francisco, USA, 1981.
- [32] M. Brower. Wind generation and forecasting profiles: A report to GE energy consulting. Technical report, AWS TrueWind, New York, Oct 2007.
- [33] M. Brower. Wind generation assessment: A report to the Electric Reliability Council of Texas (ERCOT). Technical report, AWS TrueWind, New York, Jan 2007.
- [34] Michael Brower. Development of eastern regional wind resource and wind plant output datasets. Technical Report NREL/SR-550-46764, AWS Truewind, March 2008.
- [35] Barbara G. Brown, Richard W. Katz, and Allan H. Murphy. Time series models to simulate and forecast wind speed and wind power. *Journal of Climate and Applied Meteorology*, 23:1184–1195, May 1984.
- [36] Josep Calbo, Josep-Abel Gonzalez, and David Pages. A method for sky-condition classification from ground-based solar radiation measurements. *J. Appl. Meteor.*, 40(12):2193–2199, December 2001.



- [37] D. Cano, J.M. Monget, M. Albuissou, H. Guillard, N. Regas, and L. Wald. A method for the determination of the global solar radiation from meteorological satellite data. *Solar Energy*, 37(1):31 – 39, 1986.
- [38] Jiacong Cao and Xingchun Lin. Study of hourly and daily solar irradiation forecast using diagonal recurrent wavelet neural networks. *Energy Conversion and Management*, 49(6):1396 – 1406, 2008.
- [39] Gary Chamberlain. Funds, factors, and diversification in arbitrage pricing models. *Econometrica*, 51(5):pp. 1305–1323, 1983.
- [40] Gary Chamberlain and Michael Rothschild. Arbitrage, factor structure, and mean-variance analysis on large asset markets. Working Paper 996, National Bureau of Economic Research, 1984 1984.
- [41] Chih-Chung Chang and Chih-Jen Lin. Libsvm: A library for support vector machines. *ACM Trans. Intell. Syst. Technol.*, 2(3):27:1–27:27, May 2011.
- [42] H. Chavez, R. Baldick, and S. Sharma. Regulation adequacy analysis under high wind penetration scenarios in ercot nodal. *Sustainable Energy, IEEE Transactions on*, 3(4):743–750, Oct 2012.
- [43] H. Chavez, R. Baldick, and S. Sharma. Governor rate-constrained opf for primary frequency control adequacy. *Power Systems, IEEE Transactions on*, 29(3):1473–1480, May 2014.

- [44] Hector Chavez, Duehee Lee, and Ross Baldick. CPS1-compliant regulation using a psd analysis of wind expansion in a single balancing authority. *Sustainable Energy, IEEE Transactions on*, 2015.
- [45] Changsong Chen, Shanxu Duan, Tao Cai, and Bangyin Liu. Online 24-h solar power forecasting based on weather type classification using artificial neural network. *Solar Energy*, 85(11):2856 – 2870, 2011.
- [46] Weitian Chen, Brian D.O. Anderson, Manfred Deistler, and Alexander Filler. Solutions of Yule-Walker equations for singular AR processes. *Journal of Time Series Analysis*, 32:531–538, 2011.
- [47] Weitian Chen, Brian D.O. Anderson, Manfred Deistler, and Alexander Filler. Properties of blocked linear systems. *Automatica*, 48(10):2520 – 2525, 2012.
- [48] A. Clauset, C. Shalizi, and M. Newman. Power-law distributions in empirical data. *SIAM Review*, 51(4):661–703, 2009.
- [49] D Corbus, J King, T Mousseau, R Zavadil, and B Heath. Eastern wind integration and transmission study. Technical Report SR-5500-47078, National Renewable Energy Laboratory, Feb 2010.
- [50] Katie Coughlin and Joseph H. Eto. Analysis of wind power and load data at multiple time scales. Technical Report DE-AC02-05CH11231, Ernest Orlando Lawrence Berkeley National Laboratory, 2010.

- [51] Nicholas Cutler, Merlinde Kay, Kieran Jacka, and Torben Skov Nielsen. Detecting, categorizing and forecasting large ramps in wind farm power output using meteorological observations and wppt. *Wind Energy*, 10(5):453–470, 2007.
- [52] Robert J. Davy, Milton J. Woods, Christopher J. Russell, and Peter A. Coppin. Statistical downscaling of wind variability from meteorological fields. *Boundary-Layer Meteorology*, 135(1):161–175, 2010.
- [53] Lew Debra and R. Piwko. Western wind and solar integration study. Technical report, National Renewable Energy Laboratories, 2010.
- [54] M Deistler, A Filler, and B Funovits. AR systems and AR processes: The singular case. *Communications in Information and Systems*, 11(3):225, 2011.
- [55] Manfred Deistler, Brian D.O. Anderson, Alexander Filler, Ch. Zinner, and Weitan Chen. Generalized linear dynamic factor models: An approach via singular autoregressions. *European Journal of Control*, 16(3):211 – 224, 2010.
- [56] Manfred Deistler and Christiane Zinner. Modelling high-dimensional time series by generalized linear dynamic factor models: An introductory survey. *Communications in Information & Systems*, 7(2):153–166, 2007.
- [57] Stephen M. Kogon Dimitris G. Manolakis, Vinay K. Ingle. *Statistical and Adaptive Signal Processing*. McGraw-Hill, 2000.

- [58] Potomac Economics. 2013 state of the market report for the ERCOT wholesale electricity markets. Technical report, Potomac Economics, LTD, Sept 2014.
- [59] Jimmy S.G. Ehnberg and Math H.J. Bollen. Simulation of global solar radiation based on cloud observations. *Solar Energy*, 78(2):157 – 162, 2005.
- [60] Robert Engle and Mark Watson. A one-factor multivariate time series model of metropolitan wage rates. *Journal of the American Statistical Association*, 76(376):774–781, 1981.
- [61] ERCOT. ERCOT quick facts. Technical report, Electric Reliability Council of Texas, 2007. [http://www.ercot.com/content/news/presentations/2015/ERCOT\\_Quick\\_Facts\\_11215.pdf](http://www.ercot.com/content/news/presentations/2015/ERCOT_Quick_Facts_11215.pdf).
- [62] ERCOT. ERCOT methodologies for determining ancillary service requirements. Technical report, ERCOT, 2011.
- [63] ERCOT. ERCOT concept paper. Technical report, Electric Reliability Council of Texas, Nov. 2013.
- [64] ERCOT. ERCOT system planning monthly status report. Technical report, ERCOT, April 2014.
- [65] ERCOT. ERCOT quick facts. Technical report, Electric Reliability Council of Texas, Jan 2015.

- [66] B. Ernst. Analysis of Wind Power Ancillary Services Characteristics with German 250-MW Wind Data. Technical Report NREL/TP-500-26969, National Renewable Energy Laboratory, Golden, Colorado, Nov. 1999.
- [67] M. J. Bertin et al. *Pisot and Salem Numbers*. user Verlag, Berlin, 1992.
- [68] Lucas Eustaquio and Gomes da Silva. A blending approach to solar energy prediction. Technical report, 12th Conference on Artificial and Computational Intelligence and its Applications to the Environmental Sciences, 2014. <https://ams.confex.com/ams/94Annual/webprogram/Session35375.html>.
- [69] Alexander Filler. *Generalized Dynamic Factor Models: Structure Theory and Estimation for Single Frequency and Mixed Frequency Data*. PhD thesis, Vienna University of Technology, 2010.
- [70] Ulrich Focken, Matthias Lange, Kai Monnich, Hans-Peter Waldl, Hans Georg Beyer, and Armin Luig. Short-term prediction of the aggregated power output of wind farm—a statistical analysis of the reduction of the prediction error by spatial smoothing effects. *Journal of Wind Engineering and Industrial Aerodynamics*, 90(3):231–246, 2002.
- [71] Mario Forni and Luca Gambetti. The dynamic effects of monetary policy: A structural factor model approach. *Journal of Monetary Economics*, 57(2):203 – 216, 2010.

- [72] Mario Forni, Domenico Giannone, Marco Lippi, and Lucrezia Reichlin. Opening the black box: Structural factor models with large cross sections. *Econometric Theory*, 25:1319–1347, 10 2009.
- [73] Mario Forni, Marc Hallin, Marco Lippi, and Lucrezia Reichlin. The generalized dynamic factor model: One-sided estimation and forecasting. *Journal of the American Statistical Association*, 100(471):830–840, 2005.
- [74] Mario Forni, Marc Hallin, Lippi Marco, and Lucrezia Reichlin. The generalized dynamic-factor model: Identification and estimation. *Review of Economics and Statistics*, 82(4):540–554, Nov. 2000.
- [75] Mario Forni, Marco Lippi, and Lucrezia Reichlin. Opening the black box: Structural factor models versus structural vars. *Centre for Economic Policy Research*, 4133, 2003.
- [76] Jerome H. Friedman. Stochastic gradient boosting. *Computational Statistics & Data Analysis*, 38(4):367 – 378, 2002. Nonlinear Methods and Data Mining.
- [77] John Geweke. *Latent VARIables in Socio-Economic Models*, chapter The dynamic factor analysis of economic time series models. University of Wisconsin, 1976.
- [78] G. Giebel, R. Brownsword, and G. Kariniotakis. The state-of-the-art in short-term prediction of wind power: A literature overview. Technical

report, Advanced Tools for the Management of Electricity Grids with Large-Scale Wind Generation, Roskilde, Denmark, January 2011.

- [79] William Glassley, Jan Kleissl, C. P. Van Dam, Henry Shiu, Junhui Huang, Gerry Braun, and Ronnie Holland. California renewable energy forecasting, resource data and mapping. Reports 500-99-013, California Institute for Energy and Environment, 2010.
- [80] T.N. Goh and K.J. Tan. Stochastic modeling and forecasting of solar radiation data. *Solar Energy*, 19(6):755 – 757, 1977.
- [81] S. Granville, M.V.F. Pereira, G.B. Dantzig, B. Avi-Itzhak, M. Avriel, A. Monticelli, and L.M.V.G. Pinto. Mathematical decomposition techniques for power system expansion planning: Volume 2, analysis of the linearized power flow model using the bender decomposition technique: Final report. Technical Report EPRI-EL-5299-Vol.2, Electric Power Research Institute (EPRI), Feb. 1988.
- [82] D.M. Gruenbacher and D.R. Hummels. A simple algorithm for generating discrete prolate spheroidal sequences. *Signal Processing, IEEE Transactions on*, 42(11):3276 –3278, Nov. 1994.
- [83] RS & H. CREZ progress Report No. 17. Technical report, Public Utility Commission of Texas, Dec 2014.
- [84] D.A. Halamay, T.K.A. Brekken, A. Simmons, and S. McArthur. Reserve requirement impacts of large-scale integration of wind, solar, and ocean

- wave power generation. *Sustainable Energy, IEEE Transactions on*, 2(3):321–328, July 2011.
- [85] Marc Hallin and Roman Liska. Determining the number of factors in the general dynamic factor model. *Journal of the American Statistical Association*, 102(478):603–617, 2007.
- [86] Sekyung Han, Soohee Han, and K. Sezaki. Development of an optimal vehicle-to-grid aggregator for frequency regulation. *Smart Grid, IEEE Transactions on*, 1(1):65–72, June 2010.
- [87] Poul Sorensen Hannele Holttinen, Ritva Hirnonen. *Wind Power in Power Systems*. Wiley, Hoboken, NJ, Feb 2009.
- [88] Trevor Hastie, Robert Tibshirani, and Jerome Friedman. *The Elements of Statistical Learning*. Springer, 2nd edition, 2009.
- [89] Monson H. Hayes. *Statistical Digital Signal Processing and Modeling*. John Wiley & Sons, 1996.
- [90] Simon Haykin. *Neural Networks and Learning Machines*. Pearson International Edition, 2009.
- [91] H.S. Hippert, C.E. Pedreira, and R.C. Souza. Neural networks for short-term load forecasting: a review and evaluation. *Power Systems, IEEE Transactions on*, 16(1):44–55, Feb 2001.



- [92] Fatih O. Hocaoglu, Omer N. Gerek, and Mehmet Kurban. Hourly solar radiation forecasting using optimal coefficient 2-d linear filters and feed-forward neural networks. *Solar Energy*, 82(8):714–726, 8 2008.
- [93] Nynke Hofstra and Mark New. Spatial variability in correlation decay distance and influence on angular-distance weighting interpolation of daily precipitation over europe. *International Journal of Climatology*, 29(12):1872–1880, 2009.
- [94] <http://www.dsireusa.org>. Database of state incentives for renewables and efficiency.
- [95] Jin Hur. *Spatial Prediction of Wind Farm OOutput for Grid Integration using the Augmented Kriging-based Model*. PhD thesis, The University of Texas at Austin, 2012.
- [96] Anil K Jain and Richard C Dubes. *Algorithms for clustering data*. Prentice-Hall, Inc., 1988.
- [97] Gareth James, Daniela Witten, Trevor Hastie, and Robert Tibshirani. *An Introduction to Statistical Learning with AApplication in R*. Springer, 2013.
- [98] Jing-Jong Jang and Yu-Lin Lee. A study of along wind speed power spectrum for taiwan area. *Journal of Marine Science and Technology*, 6(1):71–77, 1998.

- [99] Jooyoung Jeon and James W. Taylor. Using conditional kernel density estimation for wind power density forecasting. *Journal of the American Statistical Association*, 107(497):66–79, 2012.
- [100] Wu Ji and Keong Chan Chee. Prediction of hourly solar radiation using a novel hybrid model of ARMA and TDNN. *Solar Energy*, 85(5):808 – 817, 2011.
- [101] P.D. Jones, T. J. Osborn, and K. R. Briffa. Estimating sampling errors in large-scale temperature averages. *Journal of Climate*, 10:2548–2568, Oct 1997.
- [102] Henry. F. Kaiser. The varimax criterion for analytic rotation in factor analysis. *Psychometrika*, 23(3):187–200, 1958.
- [103] Rajesh Karki, Po Hu, and Roy Billinton. A simplified wind power generation model for reliability evaluation. *Energy conversion, IEEE Transactions on*, 21(2):533–540, 2006.
- [104] Warren Katzenstein, Emily Fertig, and Jay Apt. The variability of interconnected wind plants. *Energy Policy*, 38(8):4400 – 4410, 2010.
- [105] Y Kemmoku, S Orita, S Nakagawa, and T Sakakibara. Daily insolation forecasting using a multi-stage neural network. *Solar Energy*, 66(3):193 – 199, 1999.
- [106] A. Khosravi, S. Nahavandi, D. Creighton, and A.F. Atiya. Lower upper bound estimation method for construction of neural network-based pre-

- diction intervals. *Neural Networks, IEEE Transactions on*, 22(3):337–346, March 2011.
- [107] Andrey Nikolaevich Kolmogorov. Dissipation of energy in locally isotropic turbulence. In *Dokl. Akad. Nauk SSSR*, volume 32, pages 16–18, 1941.
- [108] Samuel Kotz, Tomasz Kozubowski, and Krzysztof Podgorski. *The Laplace distribution and generalizations: a revisit with applications to communications, economics, engineering, and finance*. Number 183. Springer Science & Business Media, 2001.
- [109] Birk Kraas, Marion Schroedter-Homscheidt, and Reinhard Madlener. Economic merits of a state-of-the-art concentrating solar power forecasting system for participation in the spanish electricity market. *Solar Energy*, 93(0):244 – 255, 2013.
- [110] Aksomitis Kris. Short-term wind integration: Discussion paper. Technical report, Alberta Electric System Operator, Canada, May 2010.
- [111] Max Kuhn and Kjell Johnson. *Applied Predictive Modeling*. Springer, 2013.
- [112] M. H. Kutner, C. J. Nachtsheim, and J. Neter. *Applied Linear Regression Models*. McGraw-Hill Irwin, New York, fourth edition, 2004.
- [113] Leslie Lamport. *TEX: A document preparation system*. Addison-Wesley, 2nd edition, 1994.

- [114] Eric Lantz, Daniel Steinberg, Michael Mendelsohn, Owen Zinaman, Ted James, Gian Porro, Maureen Hand, Trieu Mai, Jeffrey Logan, Jenny Heeter, et al. Implications of a ptc extension on us wind deployment. Technical report, National Renewable Energy Laboratory (NREL), Golden, CO., 2014.
- [115] M. Lave, J. Kleissl, and J.S. Stein. A wavelet-based variability model (wvm) for solar pv power plants. *Sustainable Energy, IEEE Transactions on*, 4(2):501–509, April 2013.
- [116] D.N. Lawley and A. E. Maxwell. *Factor Analysis as a Statistical Method*. London Butterworths, second edition, 1971.
- [117] Duehee Lee and R. Baldick. Synthesis of sample paths of wind power through factor analysis amp; cluster analysis. In *North American Power Symposium (NAPS), 2013*, pages 1–6, Sept 2013.
- [118] Duehee Lee and R. Baldick. Future wind power scenario synthesis through power spectral density analysis. *Smart Grid, IEEE Transactions on*, 5(1):490–500, Jan 2014.
- [119] Duehee Lee and R. Baldick. Short-term wind power ensemble prediction based on gaussian processes and neural networks. *Smart Grid, IEEE Transactions on*, 5(1):501–510, Jan 2014.
- [120] Duehee Lee and Ross Baldick. Analyzing the variability of wind power output through the power spectral density. In *Power and Energy Society*

*General Meeting, 2012 IEEE*, pages 1–8. (Accepted), July 2012.

- [121] Wenyuan Li. *Probabilistic Transmission System Planning*. IEEE Press, 2011.
- [122] Jiaqi Liang, S. Grijalva, and R.G. Harley. Increased wind revenue and system security by trading wind power in energy and regulation reserve markets. *Sustainable Energy, IEEE Transactions on*, 2(3):340–347, July 2011.
- [123] N. R. Lomb. Least-squares frequency analysis of unequally spaced data. *Astrophysics and Space Science*, 39:447–462, 1976. 10.1007/BF00648343.
- [124] E. Lorenz, J. Hurka, D. Heinemann, and H.G. Beyer. Irradiance forecasting for the power prediction of grid-connected photovoltaic systems. *Selected Topics in Applied Earth Observations and Remote Sensing, IEEE Journal of*, 2(1):2–10, March 2009.
- [125] H. Louie. Evaluation of probabilistic models of wind plant power output characteristics. In *Probabilistic Methods Applied to Power Systems, 2010 IEEE 11th International Conference on*, pages 442–447, June 2010.
- [126] Gilles Louppe and Peter Prettenhofer. Forecasting daily solar energy production using robust regression techniques. Technical report, 12th Conference on Artificial and Computational Intelligence and its Applications to the Environmental Sciences, Feb 2014.

- [127] Liancong Luo, David Hamilton, and Boping Han. Estimation of total cloud cover from solar radiation observations at lake rotorua, new zealand. *Solar Energy*, 84(3):501 – 506, 2010.
- [128] Jordan Macknick, Jordan Macknick, Robin Newmark, Garvin Heath, Kc Hallett, Jordan Macknick, Robin Newmark, Garvin Heath, and Kc” Hallett. A review of operational water consumption and withdrawal factors for electricity generating technologies. national renewable energy laboratory. Technical Report NREL/TP-6A20-50900, National Renewable Energy Laboratories, 2011.
- [129] Y. V. Makarov, P. Nyeng, B. Yang, J. Ma, J. G. DeSteese, D. J. Hammerstrom, S. Lu, V. V. Viswanathan, and C.H. Miller. Wide-area energy storage and management system to balance inermittent resources in the Bonneville administration and California ISO control areas. Technical report, Pacific Northwest Naional Laboratory, June 2008.
- [130] Ricardo Marquez and Carlos F.M. Coimbra. Forecasting of global and direct solar irradiance using stochastic learning methods, ground experiments and the {NWS} database. *Solar Energy*, 85(5):746 – 756, 2011.
- [131] Patrick Mathiesen, Craig Collier, and Jan Kleissl. A high-resolution, cloud-assimilating numerical weather prediction model for solar irradiance forecasting. *Solar Energy*, 92(0):47 – 61, 2013.
- [132] P. Meibom, R. Barth, B. Hasche, H. Brand, C. Weber, and M. O’Malley. Stochastic optimization model to study the operational impacts of high

- wind penetrations in ireland. *Power Systems, IEEE Transactions on*, 26(3):1367 –1379, Aug. 2011.
- [133] A. Mellit, M. Benganem, and S.A. Kalogirou. An adaptive wavelet-network model for forecasting daily total solar-radiation. *Applied Energy*, 83(7):705 – 722, 2006.
- [134] Adel Mellit and Alessandro Massi Pavan. A 24-h forecast of solar irradiance using artificial neural network: Application for performance prediction of a grid-connected PV plant at trieste, italy. *Solar Energy*, 84(5):807 – 821, 2010.
- [135] Martin Fodslette Moller. A scaled conjugate gradient algorithm for fast supervised learning. *Neural Networks*, 6(4):525–533, 1993.
- [136] C. Monteiro, R. Bessa, V. Miranda, A. Botterud, J. Wang, and G. Conzelmann. Wind power forecasting: State-of-the-art 2009. Technical report, Argonne Natinal Laboratory: Decision and Information Sciences Division, Chicago, Illinois, November 2009.
- [137] J.M. Morales, L. Baringo, A.J. Conejo, and R. Mínguez. Probabilistic power flow with correlated wind sources. *IET Generation, Transmission & Distribution*, 4(5):641 – 651, May 2010.
- [138] J.M. Morales, R. Minguez, and A.J. Conejo. A methodology to generate statistically dependent wind speed scenarios. *Applied Energy*, 87(3):843 – 855, 2010.

- [139] E. Muljadi, V. Gevorgian, M. Singh, and S. Santoso. Understanding inertial and frequency response of wind power plants. In *Power Electronics and Machines in Wind Applications (PEMWA), 2012 IEEE*, pages 1–8, July 2012.
- [140] J. Mur-Amada and A.A. Bayocl-Rujula. Characterization of spectral density of wind farm power output. In *Electrical Power Quality and Utilisation, 2007. EPQU 2007. 9th International Conference on*, pages 1–6, 2007.
- [141] Toshiya Nanahara, Masahiro Asari, Takamitsu Sato, Koji Yamaguchi, Masaaki Shibata, and Tsutomu Maejima. Smoothing effects of distributed wind turbines. part 1. coherence and smoothing effects at a wind farm. *Wind Energy*, 7(2):61–74, 2004.
- [142] Henrik Aalborg Nielsen, Henrik Madsen, and Torben Skov Nielsen. Using quantile regression to extend an existing wind power forecasting system with probabilistic forecasts. *Wind Energy*, 9(1-2):95–108, 2006.
- [143] Christophe Van Nieuwenhuyze. A generalised dynamic factor model for the Belgian economy-useful business cycle indicators and GDP growth forecasts. *National Bank of Belgium Working Paper*, (80), 2006.
- [144] Per Norgaard and Hannele Holttinen. A multi-turbine power curve approach. In *Nordic Wind Power Conference*, volume 1, pages 1–2, 2004.



- [145] M.A. Ortega-Vazquez and D.S. Kirschen. Estimating the spinning reserve requirements in systems with significant wind power generation penetration. *Power Systems, IEEE Transactions on*, 24(1):114–124, Feb 2009.
- [146] T. J. Osborn and M. Hulme. Development of a relationship between station and grid-box rainday frequencies for climate model evaluation. *Journal of Climate*, 10(8):1885–1908, 2011/09/24 1997.
- [147] Christophe Paoli, Cyril Voyant, Marc Muselli, and Marie-Laure Nivet. Forecasting of preprocessed daily solar radiation time series using neural networks. *Solar Energy*, 84(12):2146 – 2160, 2010.
- [148] G. Papaefthymiou and B. Klockl. Mcmc for wind power simulation. *Energy Conversion, IEEE Transactions on*, 23(1):234–240, March 2008.
- [149] A. Papavasiliou, S.S. Oren, and R.P. O’Neill. Reserve requirements for wind power integration: A scenario-based stochastic programming framework. *Power Systems, IEEE Transactions on*, 26(4):2197–2206, Nov 2011.
- [150] Heejung Park and R. Baldick. Transmission planning under uncertainties of wind and load: Sequential approximation approach. *Power Systems, IEEE Transactions on*, 28(3):2395–2402, Aug 2013.
- [151] O. Perpignan and E. Lorenzo. Analysis and synthesis of the variability of irradiance and {PV} power time series with the wavelet transform.

- Solar Energy*, 85(1):188 – 197, 2011.
- [152] Mansoureh Peydayesh and Ross Baldick. The effects of very fast response to frequency fluctuation. In *31th United States Association of Energy Economics Conference*, 2012.
- [153] Haurant Pierrick, Menezo Christophe, Gaillard Leon, and Dupeyrat Patrick. Dynamic numerical model of a high efficiency pv-t collector integrated into a domestic hot water system. *Solar Energy*, 111(0):68 – 81, 2015.
- [154] P. Pinson, C. Chevallier, and G.N. Kariniotakis. Trading wind generation from short-term probabilistic forecasts of wind power. *Power Systems, IEEE Transactions on*, 22(3):1148–1156, Aug 2007.
- [155] P. Pinson and R. Girard. Evaluating the quality of scenarios of short-term wind power generation. *Applied Energy*, 96(0):12 – 20, 2012. Smart Grids.
- [156] P. Pinson and G. Kariniotakis. Conditional prediction intervals of wind power generation. *Power Systems, IEEE Transactions on*, 25(4):1845–1856, Nov 2010.
- [157] P. Pinson, H.Aa. Nielsen, H. Madsen, and G. Kariniotakis. Skill forecasting from ensemble predictions of wind power. *Applied Energy*, 86(7):1326 – 1334, 2009.

- [158] Pierre Pinson. *Estimation of the Uncertainty in Wind Power Forecasting*. PhD thesis, Ecole des Mines de Paris, 2006.
- [159] Pierre Pinson and Henrik Madse. Ensemble-based probabilistic forecasting at horns rev. *Wind Energy*, 12:1099 – 1824, 2009.
- [160] ERCOT System Planning. Panhandle renewable energy zone (prez) study report. Technical report, Electric Reliability Council of Texas, April 2014.
- [161] C. E. Rasmussen and C. K. I. Williams. *Gaussian Processes for Machine Learning*. MIT Press, 2006.
- [162] Carl Edward Rasmussen. Gaussian processes in machine learning. In Olivier Bousquet, Ulrike von Luxburg, and Gunnar Rätsch, editors, *Advanced Lectures on Machine Learning*, volume 3176 of *Lecture Notes in Computer Science*, pages 63–71. Springer Berlin Heidelberg, 2004.
- [163] Gordon Reikard. Predicting solar radiation at high resolutions: A comparison of time series forecasts. *Solar Energy*, 83(3):342 – 349, 2009.
- [164] Alvin C. Rencher and William F. Christensen. *Method of Multivariate Analysis*. Wiley Series in Probability and Statistics, third edition, 2012.
- [165] Greg Ridgeway. Generalized boosted model: A guide to the gbm package, May 2012.

- [166] J. Salmeron, K. Wood, and R. Baldick. Analysis of electric grid security under terrorist threat. *Power Systems, IEEE Transactions on*, 19(2):905–912, 2004.
- [167] E. F. Sanchez-Ubeda and L. Wehenkel. The hinges model: A one-dimensional countinuous piecewise polynomial model. In *Information Processing and Management of Uncertainty in Knowledge-based Systems*, pages 878–885, July 1998.
- [168] Thomas J Sargent and Christopher A Sims. Business cycle modeling without pretending to have too much a priori economic theory. *New Methods in Business Cycle Research*, 1:145–168, 1977.
- [169] Eugene Seneta. Fitting the variance-gamma model to financial data. *J. Appl. Probab.*, 41A:177–187, 02 2004.
- [170] A. Sfetsos and A.H. Coonick. Univariate and multivariate forecasting of hourly solar radiation with artificial intelligence techniques. *Solar Energy*, 68(2):169 – 178, 2000.
- [171] N. Sharma, P. Sharma, D. Irwin, and P. Shenoy. Predicting solar generation from weather forecasts using machine learning. In *Smart Grid Communications (SmartGridComm), 2011 IEEE International Conference on*, pages 528–533, Oct 2011.
- [172] Sandip R Sharma. ERCOT pilot project for fast responding regulation service (frs). In *IEEE T & D Energy Storage Applications and Business*

*Models*, 2014.

- [173] M. Sherman. *Spatial Statistics and Spatio-Temporal Data*. Wiley Series in Probability and Statistics, 2011.
- [174] Jie Shi, Wei-Jen Lee, Yongqian Liu, Yongping Yang, and Peng Wang. Forecasting power output of photovoltaic systems based on weather classification and support vector machines. *Industry Applications, IEEE Transactions on*, 48(3):1064–1069, May 2012.
- [175] Robert H. Shumway and David S. Stoffer. *Time Series Analysis and Its Applications*. Springer Texts in Statistics. Springer, third edition, 2011.
- [176] G. Sideratos and N.D. Hatziargyriou. Probabilistic wind power forecasting using radial basis function neural networks. *Power Systems, IEEE Transactions on*, 27(4):1788–1796, Nov 2012.
- [177] Ramteen Sioshansi. Increasing the value of wind with energy storage. *Energy Journal*, 32(2):1–29, 2011.
- [178] Ramteen Sioshansi and David Hurlbut. Market protocols in ERCOT and their effect on wind generation. *Energy Policy*, 38(7):3192 – 3197, 2010. Large-scale wind power in electricity markets with Regular Papers.
- [179] E. Sortomme and M.A. El-Sharkawi. Optimal scheduling of vehicle-to-grid energy and ancillary services. *Smart Grid, IEEE Transactions on*, 3(1):351–359, March 2012.

- [180] C. Spearman. "general intelligence," objectively determined and measured. *The American Journal of Psychology*, 15(2):pp. 201–292, 1904.
- [181] James H Stock and Mark W Watson. Forecasting using principal components from a large number of predictors. *Journal of the American Statistical Association*, 97(460):1167–1179, 2002.
- [182] James H Stock and Mark W Watson. Macroeconomic forecasting using diffusion indexes. *Journal of Business & Economic Statistics*, 20(2):147–162, 2002.
- [183] James H Stock and Mark W Watson. Dynamic factor models. *Oxford Handbook of Economic Forecasting*. Oxford University Press, USA, pages 35–59, 2011.
- [184] Steven Stoft. *Power system economics*, volume 8. THE OXFORD INSTITUTE FOR ENERGY STUDIES, 2002.
- [185] J.O. Street., R. J. Carroll, and D. Ruppert. A note on computing robust regression estimates via iteratively reweighted least squares. *Ann. Statist.*, 42:152–154, 1988.
- [186] S. Subbayya, J.G. Jetcheva, and Wei-Peng Chen. Model selection criteria for short-term microgrid-scale electricity load forecasts. In *Innovative Smart Grid Technologies (ISGT), 2013 IEEE PES*, pages 1–6, Feb 2013.

- [187] System Planning Department. Pjm manual 21: Rules and procedures for determination of generating capability. Technical report, PJM, May 2014.
- [188] Nobuo Tanaka. Harnessing variable renewables: A guide to the balancing challenge. Technical report, International Energy Agency, Paris, France, March 2011.
- [189] S. Tewari, C.J. Geyer, and N. Mohan. A statistical model for wind power forecast error and its application to the estimation of penalties in liberalized markets. *Power Systems, IEEE Transactions on*, 26(4):2031–2039, Nov 2011.
- [190] David J. Thomson. Quadratic-inverse spectrum estimates: Applications to palaeoclimatology. *Philosophical Transactions: Physical Sciences and Engineering*, 332(1627):pp. 539–597, Sept. 1990.
- [191] D.J. Thomson. Spectrum estimation and harmonic analysis. *Proceedings of the IEEE*, 70(9):1055 – 1096, Sept. 1982.
- [192] A. Tuohy, P. Meibom, E. Denny, and M. O’Malley. Unit commitment for systems with significant wind penetration. *Power Systems, IEEE Transactions on*, 24(2):592 –601, May 2009.
- [193] T.S. Viana, R. R  ther, F.R. Martins, and E.B. Pereira. Assessing the potential of concentrating solar photovoltaic generation in brazil with

- satellite-derived direct normal irradiation. *Solar Energy*, 85(3):486 – 495, 2011.
- [194] Claire Louise Vincent, Gregor Giebel, Pierre Pinson, and Henrik Madsen. Resolving nonstationary spectral information in wind speed time series using the hilbert-huang transform. *Journal of Applied Meteorology and Climatology*, 49(2):253–267, 2010.
- [195] R. A. Walling. Analysis of Wind Generation Impact on ERCOT Ancillary Services Requirements. Technical report, GE Energy, Schenectady, New York, March 2008.
- [196] Can Wan, Zhao Xu, P. Pinson, Zhao Yang Dong, and Kit Po Wong. Optimal prediction intervals of wind power generation. *Power Systems, IEEE Transactions on*, 29(3):1166–1174, May 2014.
- [197] Jianhui Wang, M. Shahidepour, and Zuyi Li. Security-constrained unit commitment with volatile wind power generation. *Power Systems, IEEE Transactions on*, 23(3):1319–1327, Aug 2008.
- [198] W. Wangdee and R. Billinton. Considering load-carrying capability and wind speed correlation of WECS in generation adequacy assessment. *Energy Conversion, IEEE Transactions on*, 21(3):734–741, Sept 2006.
- [199] WindFarmer. <http://www.gl-garradhassan.com/en/ghwindfarmer.php>, May 2012.



- [200] Michael Wittmann, Markus Eck, Robert Pitz-Paal, and Hans Müller-Steinhagen. Methodology for optimized operation strategies of solar thermal power plants with integrated heat storage. *Solar Energy*, 85(4):653 – 659, 2011. SolarPACES 2009.
- [201] Dan Woodfin and Walter Reid. Response to request for comments issued on august 13, 2014. Technical Report 42647, Electric Reliability Council of Texas, 2014.
- [202] Dazhi Yang, Panida Jirutitijaroen, and Wilfred M. Walsh. Hourly solar irradiance time series forecasting using cloud cover index. *Solar Energy*, 86(12):3531 – 3543, 2012. Solar Resources.
- [203] D.L. Yao, S.S. Choi, K.J. Tseng, and T. T. Lie. A statistical approach to the design of a dispatchable wind power-battery energy storage system. *Energy Conversion, IEEE Transactions on*, 24(4):916–925, 2009.
- [204] A. Yona, T. Senjyu, A.Y. Saber, T. Funabashi, H. Sekine, and Chul-Hwan Kim. Application of neural network to 24-hour-ahead generating power forecasting for pv system. In *Power and Energy Society General Meeting - Conversion and Delivery of Electrical Energy in the 21st Century, 2008 IEEE*, pages 1–6, July 2008.

## Vita

Duehee Lee was born in Daegu, Republic of Korea on November 29, 1981, the son of Kicho Lee and Sunhee Whang. He received the Bachelor of Science degree in Electronic and Electrical Engineering from the Pohang University of Science and Technology (POSTECH) and was commissioned an Officer in the Republic of Korea Navy (ROK NAVY) in 2004. He entered active duty in July, 2004, and was stationed in Jinhae, Kyungsangnamdo. He separated from the ROK NAVY in 2007 and applied to the University of Texas at Austin for enrollment in their Electrical and Computer Engineering. He was accepted and started graduate studies in September, 2007. He continued pursuing his Ph.D. from the Department of Electrical and Computer Engineering at The University of Texas at Austin, Austin, TX under supervision of Professor Ross Baldick and graduated in May 2015. His primary research interests are the power system, wind power forecasting, wind power analysis, and electricity market.

Permanent address: 10800 Lakeline Blvd, #13102  
Austin, Texas 78717

This dissertation was typeset with L<sup>A</sup>T<sub>E</sub>X<sup>†</sup> by the author.

---

<sup>†</sup>L<sup>A</sup>T<sub>E</sub>X is a document preparation system developed by Leslie Lamport as a special version of Donald Knuth's T<sub>E</sub>X Program.