2012

Statistical Analysis of Wind Data and Modeling Regulating Reserves

William Buchanan

University of Vermont

Follow this and additional works at: https://scholarworks.uvm.edu/graddis

Recommended Citation

https://scholarworks.uvm.edu/graddis/33

This Thesis is brought to you for free and open access by the Dissertations and Theses at ScholarWorks @ UVM. It has been accepted for inclusion in Graduate College Dissertations and Theses by an authorized administrator of ScholarWorks @ UVM. For more information, please contact donna.omalley@uvm.edu.
STATISTICAL ANALYSIS OF WIND DATA AND MODELING REGULATING RESERVES

A Thesis Presented
by
William James Buchanan
to
The Faculty of the Graduate College
of
The University of Vermont

In Partial Fulfillment of the Requirements
for the Degree of Master of Science
Specializing in Electrical Engineering

October, 2012
Abstract

The desire to reduce dependence on fossil fuels is resulting in numerous policy incentives for increased renewable energy sources within the power grid. Because wind generation is arguably the most affordable per MWh of the renewable energy sources it is growing nearly as quickly as conventional generation techniques. Due to this significant increase in wind penetration levels, numerous large-scale wind integration studies have been produced to determine the reliability impacts of large-scale wind power. Using data from two large US wind interconnection studies, this thesis provides evidence that mesoscale meteorological models under-predict the variability in wind data particularly on short time scales, indicating that data from mesoscale meteorological models need to be used with caution for some types of analyses. These types of analyses include most notably regulating reserves, which are used to rebalance supply and demand on a second-by-second bias. This thesis will also describe and evaluate a new method for jointly quantifying the amount of spinning and regulating reserves required to meet reliability requirements within a balancing area with significant amounts of wind power using high resolution wind data. The method is based on jointly minimizing dispatch costs and reserve allocations, across two time scales (seconds to minutes, and minutes to hours) to satisfy North American Electric Reliability Corporation (NERC) Area Control Error (ACE) requirements.
Material from Chapter 2 in this thesis has been submitted in the following form:

\textsuperscript{1}These authors have contributed equally to this work
To my loving parents: Jim and Donna Buchanan
Acknowledgements

I would like start by expressing my gratitude to my advisor, Dr. Paul Hines. His enthusiasm, expertise, and patience are responsible for my growth in the field of power engineering. Without his support I would not be where I am today. I would also like to thank the members of my committee, Dr. Jeff Frolik, and Dr. Chris Danforth for the knowledge and assistance towards this research.

I would also like to thank Todd Ryan, Dr. Gabriela Hug, and Dr. Paulina Jaramillo at Carnegie Mellon University for all the assistance that they have contributed towards this research. In addition, I wish to thank Jay Apt and Steve Rose at Carnegie Mellon University, and John Zach from MESCO, Inc. for helpful conversations about this research.

I also wish to thank my research group and all my friends who have shared literature and invaluable assistance for everything from theoretical analysis to programming.

I gratefully acknowledge those involved in making the data for this project publicly available. The datasets were produced by the National Renewable Energy Laboratory, which is operated by the Alliance for Sustainable Energy, LLC for the U.S. Department Of Energy. Use of the data is subject to the user agreement, which is available at http://www.nrel.gov/wind/integrationdatasets/western/disclaimer.html.

This work was supported in part by Carnegie Mellon University’s RenewElec Project, which is funded by the Doris Duke Charitable Foundation, RK Mellon Foundation, National Energy Technology Laboratory, Electric Power Research Institute and the Heinz Endowments.

I would also like to thank Annemarie Gallo for her love and support during my graduate years.

Finally I wish to thank my parents, Jim and Donna Buchanan, for there constant support, both emotional and financially, during both my undergraduate and graduate years.
# Table of Contents

Citations ........................................................................................................... ii  
Dedication ........................................................................................................ iii 
Acknowledgments ............................................................................................ iv  
List of Tables .................................................................................................... vii  
List of Figures .................................................................................................. viii  

Chapter 1  Introduction ................................................................. 1  
  1.1  Motivation ................................................................................. 1  
  1.2  Objectives ................................................................................ 2  
  1.3  Outline ....................................................................................... 4  
  1.4  Contributions ............................................................................. 4  

Chapter 2  Comparing Empirical and Meteorological Wind Speed and Power Data .............................................. 6  
  2.1  Introduction ............................................................................... 6  
  2.2  Data ........................................................................................... 9  
  2.2.1  Meteorological wind speed and power data ....................... 9  
  2.2.2  Wind power plant data .......................................................... 11  
  2.3  Data Analysis Methods ............................................................. 12  
  2.3.1  Power spectral density analysis method .......................... 12  
  2.3.2  Model fitting to find the point at which PSD diverges from the Kolmogorov 14  
  2.3.3  Step-change analysis method .............................................. 16  
  2.4  Results ...................................................................................... 17  
  2.4.1  PSD analysis results for wind speed data ......................... 17  
  2.4.2  PSD analysis results for wind power data ......................... 22  
  2.4.3  Step change analysis results for wind speed data .............. 25  
  2.4.4  Step change analysis results for wind power data .............. 26  
  2.5  Discussion .................................................................................. 32  
  2.6  Conclusion .................................................................................. 34  

Chapter 3  Estimating Regulating Reserve Requirements for High Penetration Wind Scenarios .................................. 35  
  3.1  Introduction .............................................................................. 35  
  3.2  Regulation Modeling Methodology ......................................... 37  
  3.2.1  Optimal dispatch model ...................................................... 39  
  3.2.1.1  5-minute dispatch model ............................................. 41
3.2.1.2 4-second model ......................................................... 42
3.2.2 Linearized dynamic model ............................................. 43
  3.2.2.1 Feedback control system ........................................... 44
  3.2.2.2 LD mathematical model .............................................. 47
  3.2.2.3 Economic dispatch .................................................. 48
3.3 Data .................................................................................. 49
  3.3.1 Feedback control data .................................................. 49
  3.3.2 Load and wind data ...................................................... 50
  3.3.3 Generation capacity and fuel costs ................................... 51
3.4 Results .............................................................................. 54
  3.4.1 Optimal dispatch model results ...................................... 54
    3.4.1.1 SPP Regulation Comparison ...................................... 57
  3.4.2 Linearized dynamic model results ................................... 59
3.5 Discussion ......................................................................... 63
3.6 Conclusions ....................................................................... 64

Chapter 4 Conclusions ............................................................ 65
  4.1 Contributions ................................................................. 65
  4.2 Future Work ................................................................. 67

Bibliography ................................................................. 67

Appendix A ................................................................. 74
  A.1 Spectral Density Analysis .............................................. 74
  A.2 Comparing Welch’s and Bartlett’s Methods for Wind Speed Data .............................................. 75

Appendix B Are Wind Data Gaussian? ........................................ 77
# List of Tables

2.1 General Characteristics of Plant A .................................................. 11
2.2 PSD Parameter Estimates for Wind Speed ......................................... 19
2.3 Linear Best-fit Slopes for Plant A ................................................... 24
2.4 Largest Step-Change with non-zero Probability .................................. 31

3.1 Feedback Control Model Parameters ................................................. 49
3.2 Correlation between OD Model Wind Farms ...................................... 50
3.3 ISO New England Installed Capacities ............................................. 51
3.4 IEEE 39 bus Generator Values ....................................................... 52
3.5 ISO New England Fuel Prices .......................................................... 53
3.6 OD Model Ramp Rates Constraints ................................................ 53
3.7 Regulation Required for OD Model .................................................. 57

A.1 Parameter Estimates for Comparison with Welch’s Method ................. 76

B.1 Statistics of Empirical Wind Data ..................................................... 79
B.2 Probability of Empirical Wind Data being Gaussian ......................... 80
List of Figures

2.1 Single Site Analysis ................................................................. 15
2.2 Wind Speed PSD for Plant A ...................................................... 18
2.3 PSD Analysis of Simulated Wind Speeds ................................... 18
2.4 R-Squared Analysis ................................................................. 20
2.5 Boxplot of Individual Sites ....................................................... 21
2.6 PSD Analysis of Simulated Wind Power ................................... 23
2.7 PSD Analysis of Empirical Wind Power ................................... 24
2.8 Step-Change Analysis of Wind Speed for Farm A ....................... 25
2.9 Step-Change Analysis of Wind Power for Farm B ....................... 27
2.10 10 Minute Step-Change Analysis of Wind Power for Farm A ....... 28
2.11 1 Hour Step-Change Analysis of Wind Power for Farm A .......... 29
2.12 Probability of Wind Power Output over 10 Minutes ................... 30
2.13 Probability of Wind Power Output over an Hour ...................... 31
3.1 Optimal Dispatch Model Flowchart .......................................... 40
3.2 LD Model Design Flowchart ..................................................... 44
3.3 IEEE Case 39 Bus New England System .................................... 45
3.4 Two Area Automatic Generation Control Model ......................... 46
3.5 Optimal Generator Dispatch with 10 Percent Wind Penetration ....... 55
3.6 Optimal Regulation Required with Increasing Wind Penetrations .... 56
3.7 Required Up Regulation using SPP Estimation Method for Increasing Wind Penetrations .............................................. 58
3.8 Required Down Regulation using SPP Estimation Method for Increasing Wind Penetrations .............................................. 58
3.9 LD CPS1 Scores ...................................................................... 60
3.10 LD CPS2 Scores ..................................................................... 61
3.11 LD Regulation Requirements .................................................... 62
A.1 PSD Comparison of Welch’s and Bartlett’s Methods .................... 76
B.1 Gaussian vs Empirical Distribution of Wind Prediction Error ........ 78
B.2 Normal Plot of BPA Prediction Error ...................................... 78
B.3 Gaussian vs Empirical Distribution of Step-Changes of Wind ....... 79
Chapter 1

Introduction

1.1 Motivation

The history and usage of wind energy dates back over 2,000 years to a time when wind-powered machines helped in food production and pumped water. It wasn’t until the the late 1800’s that windmills began to develop into the wind turbines of today first making their appearance in Denmark around 1890 for the purpose of producing electricity. Windmills were once common around the world throughout the 19th century until the rise of fossil fuels in the west lead to their steady decline.

The world’s first megawatt-sized wind turbine was located at Grandpa’s Knob in Castleton Vermont in 1941. After a blade failed after 1100 hours of operation, there would not be another wind turbine of this size deployed for over 40 years. Over the course of the 20th century, and into the 21st, interest in wind power generation has continued to rise as the fear of peak oil and concerns about climate change continues to rise. As a result, power production from wind turbines continues to rise across the world in order to supplement the desire to reduce dependence on fossil fuels.

Currently wind turbines for utility usage tend to be on the size of megawatts with the largest capacity wind turbine rated at 7.48 MW (Thomas, 2008). There are multiple companies attempting to produce turbines with over 10 MW capacities (Richard, 2010). This race to produce larger wind turbines has also lead to the research and development of offshore wind farms. Deployment of anchored fixed-bottom offshore turbines for usage in shallow water began in the early 2000’s. In 2009 the worlds first deep water floating wind turbine was deployed of the coast of Norway (SIEMENS, 2009) and has led to several research initiatives to develop large floating wind turbines.
This increased research and development of large wind turbines is a clear indication of the desire to increase the penetration levels of wind energy in power grids across the world. As a result in the United States, several large-scale wind integration studies have been produced to determine the impacts of increasing penetration levels of wind up to 20 to 30 percent of total demand, utilizing mesoscale meteorological models. With the data from these studies being used to determine the impacts of large scale wind integration, the question must be answered of these models effectiveness in predicting the variability of empirical wind data. In order to accurately determine the best path of action for large-scale integration of wind energy, the model being used to answer important reliability questions must in itself be reliable.

If the models used for these studies were to under or over-predict the variability of empirical wind data, analysis of the data could lead to inaccurate conclusions and thus reliability concerns within the power grid. As a caveat of this regulating reserves, or automatic generation control which is dispatched every few seconds, is of crucial importance when looking to add a power supply such as wind power to the energy grid. Due to the variance at short time scales seen in real wind data, wind power’s effects on regulation must be carefully studied to better understand the impacts that increased wind penetration levels have on the quantity of regulation. If the quantity of regulating reserves scheduled is too small this could lead to reliability concerns and potential failures such as cascading blackouts. As a result, the impacts of increasing wind penetrations on regulating reserves must be studied before large penetrations of wind power is added to the electric grid in order to prevent instabilities.

1.2 Objectives

The work presented in this thesis focuses on two key research goals. The first goal is to determine the effectiveness of mesoscale meteorological models for simulating the production of potential future wind power plants, by comparing both wind speed and power outputs to the statistical properties of real data. The second research goal is to quantitatively determine the effects of regulating reserves on measures of power system reliability as the levels of wind penetration increases using empirical wind power data.

As wind energy continues to grow the effects of adding large amounts of wind power to the grid
must be rigorously studied in order to prevent reliability issues. In order to successfully study these impacts researchers will need large databases of data in order to better understand the consequences of large scale wind integration. Due to the high variability seen within wind speed, data that is used in these studies must be able to accurately represent the statistics seen within empirical data in order to prevent over or under-prediction when calculating reliability problems. This is where data from mesoscale meteorological models, whose horizontal dimensions general range from 2 kilometers up to several hundred kilometers, become extremely important when attempting to answer these questions. The data produced from meteorological models enables researches the potential to no longer rely on statistical assumptions when calculating reliability concerns, in order to better study the effects of renewables on the power grid.

The basic concept behind the electric grid that is used to supply power to homes and business is a basic supply and demand concept. In order to stabilize the power grid and prevent power failures or disturbances, such as blackouts, power being consumed must be the equal to power being produced. In order to accomplish this, independent system operators (ISO’s) use unit commitment with different time scales to ensure that supply and demand are meet at any point and time. Unit commitment breaks down into three distinct categories, day-ahead scheduling, load following and regulation. Day-ahead scheduling is based off from predicted power consumption and historical data to meet the general load pattern of each day. Load following is used to follow the general trending pattern that is seen within each day and regulation is used for balancing for second-to-second variations seen within the load patterns.

In order to account for disturbances that cannot be accounted for by dispatch alone system operators use operating reserves to assist in active power balancing. Operating reserves are defined as any capacity that can be used in active power balancing and can be broken down into regulating reserves, following reserves, and contingency reserves. Regulating reserves are defined as the capacity of generation available for balancing on time scales of seconds to minutes and used in automatic generation control for regulating frequency control. Following reserves is the capacity of generation available for balancing on time scales greater than the economic dispatch of generation and is used for overall load balancing. Finally contingency reserves are defined as the capacity of generation available for infrequent, severe events that are used to correct for non-instantaneous imbalances. In practice contingency reserves tend to be defined as the capacity of the largest generation unit online within a system.
1.3 Outline

The work presented in Chapter 2 will use statistical analysis techniques such as power spectral analysis and step-change analysis on the data produced from these meteorological models. These statistical results will then be compared to the same statistics obtained from empirical wind data obtained from two separate wind farms. The long term goal of this analysis is to determine which types of power system reliability questions data from meteorological models can be used to solve for and what problems will need further data in order to accurately determine.

The work presented in Chapter 3 will use empirical wind data to calculate the percent of regulation needed, as a function of peak demand, given increasing wind penetration levels. This will be achieved by two separate methods. The first method will utilize an optimal dispatch model that neglects transmission and is modeled using linear assumptions for a single balancing area. While the second method represents a dynamic linearized dynamic model that calculates reliability standards based off a two area system. Through comparison of the results obtained from these two models the objective is to successfully determine the minimum regulation needed, as wind penetration levels increase.

1.4 Contributions

Recent large-scale wind integration studies (Corbus, 2011; Hinkle et al., 2010; Kalnay et al., 1996; Lew and Piwko, 2010; Loutan and Hawkins, 2007; Uppala et al., 2005; Walling et al., 2008; Zavadil et al., 2006) have used meteorological models to estimate impacts of large penetrations of wind in systems that currently contain a minimal number of existing wind farms. Using two publicly available datasets from meteorological models (Corbus, 2011; Lew and Piwko, 2010) this work first presents a detailed analysis to determine the implications of the use of this data for estimating power system reliability problems. Through this research the goal is to determine the extent to which wind data from meteorological models correlate to the statistics of empirical wind data.

The majority of other existing wind integration studies assume that the wind variability can in turn be modeled using normal random variables (Doherty and O’Malley, 2005; Matos and Bessa, 2011; Meibom et al., 2010; Ortega-Vazquez and Kirschen, 2009; Yong et al., 2009). However the statistics of empirical wind deviates substantially for Gaussian models (Apt, 2007). This deviation is noted by some authors (Matos and Bessa, 2011; Ortega-Vazquez and Kirschen, 2009), but
many continue to use Gaussian models to represent the variability of wind production and wind forecast errors. This deviation has also been justified through the usage of the central limit theorem (Ortega-Vazquez and Kirschen, 2009) by referencing geographical diversity. Using high resolution empirical data this research also develops methods to estimate regulation requirements in systems with significant penetrations of wind generation.
Chapter 2

Comparing Empirical and Meteorological Wind Speed and Power Data

2.1 Introduction

The desire to reduce dependence on fossil fuels and mitigate anthropomorphic climate change is resulting in numerous policy incentives for renewable energy. Since wind generation is arguably the most affordable per MWh (EIA, 2011) of the new renewable energy options, wind generation capacity is growing nearly as fast as conventional generation technologies with capacity quadrupling between 2004 and 2008 (EIA, 2010). With this rapid increase in wind generation comes the need to better understand the impact that the variability and intermittency of wind generation has on the reliability of electricity infrastructure. As a result, numerous large-scale wind integration studies produced by government, industry, and academic organizations, have worked to estimate the reliability impacts of large-scale wind power and the feasibility of various levels of wind power production (Corbus, 2011; Hinkle et al., 2010; Kahlay et al., 1996; Lew and Piwko, 2010; Loutan and Hawkins, 2007; Uppala et al., 2005; Walling et al., 2008; Zavadil et al., 2006). However, estimating the impact of large amounts of new wind energy production in systems that currently have only a small number of existing wind farms requires an estimate of the time-varying output from power plants that do
not yet exist. Such estimates typically require large amounts of wind speed data for hypothetical wind generation development locations that is then transformed to produce models of wind power production.

Once generated, large quantities of spatially diverse wind power production data can be used to answer a wide variety of cost and reliability questions using different types of analysis along different time scales. Along monthly to yearly time scales, these data can be used to estimate the seasonal or inter-annual variability in wind farm production. Along multiple-day time horizons, the data can feed unit commitment models, which can estimate the impact of hourly wind farm variability on energy market dispatch costs. Along time horizons of minutes to hours, wind farm data can be used to estimate the need for load following generation resources or generation reserves. Higher resolution data can be used to estimate the requirements for regulation and frequency management as a function of wind power deployment.

However, there are limited quantities of appropriately accurate wind speed data available to those seeking to produce wind integration studies. The US National Weather Service collects large amounts of anemometer data (Elliott et al., 1986; NCDC, 1998). These data, however, are collected at 10 meter elevations, which reduces their utility for studying wind turbines with hub heights of 80 meters and higher. Furthermore wind speed data are typically archived at 10-15 minute intervals, which means that they cannot easily be used to solve problems that require higher sample rate data. Some 50-100 meter wind speed data at 10-15 minute intervals exist, typically in connection with site evaluation studies for new wind power plants, but only for a small number of locations, and these data sets are not generally available for statistical analysis.

Data from existing wind farms clearly provide a more accurate understanding of the statistical properties of wind farms, since they are not subject to the above uncertainties. However, extrapolating from data at one location to produce wind speed or power data for a non-existent wind plant at a distant location will result in errors, since wind patterns are different in different regions and topographies. In addition, the reliability impacts of large-scale wind deployment will depend highly on the correlations between wind output and demand for electricity. In some locations wind is correlated with daily load patterns, whereas in others, wind is anti-correlated.

Because of the difficulty in obtaining empirical wind speed data, almost all large-scale wind integration studies obtain simulated wind speed data from meteorological models and convert these estimates to power using standard turbine power curves. To produce wind speed data, mesoscale
Numerical Weather Prediction (NWP) models are calibrated to approximately reproduce historical meteorological measurements, such as wind vectors, cloud cover and precipitation, using data assimilation methods (see (Costa et al., 2008; Norman A., 1960; Shuman, 1978) for reviews of NWP methods). Such models are termed mesoscale because they are designed to model spatial and temporal patterns between those of synoptic scale models, which model global climate patterns with lower spatial resolution, and microscale models, which focus on spatial resolutions of less than 1 km. Mesoscale models have sufficient spatial resolution to estimate wind speeds at several elevations, and locations with approximately 2-10 km spatial resolution and are capable of archiving data every minute.

Computer-based NWP models have been used for meteorological analysis since the late 1970s, but such models could not produce high time-resolution data until the early 1980’s. Notis et al. (1983) developed a 24-hour ahead prediction model that used an hourly time-step, and produced hourly averaged wind speeds and directions. This method was then improved by Wegley and Formica (1983), who predicted wind speed at three different time steps of 10, 30 and 60 minutes. The first in today’s modern mesoscale NWP models, however, came when Anthes and Warner (1978), who published a time-dependent three-dimensional dynamic prediction model, which was later developed and presented by Anthes et al. (1987) as the Penn State/NCAR Mesoscale Model Version 4 (MM4). MM4 was superseded by the fifth-generation Penn State/NCAR mesoscale model (MM5) in the mid 1990s (Grell et al., 1994). MM5 remained the industry leader in the United States for over a decade (Zhong and Fast, 2003). More recently, the Weather Research and Forecasting (WRF) model (Done et al., 2004) is beginning to supersede MM5 as the leading mesoscale model.

The MM5 and WRF models have been used to produce data for a number of recent large-scale wind integration studies (Hinkle et al., 2010; Loutan and Hawkins, 2007; Walling et al., 2008; Zavadil et al., 2006), which set out to answer important questions about the financial and technological impacts of large-scale wind power deployment. However, the data for these studies are not publicly available. The data for two 2010 wind integration studies are publicly available (Corbus, 2011; Lew and Piwko, 2010).

While accurate wind speed data are a necessary input for most large-scale wind integration studies, wind speed data alone are not sufficient. In order to estimate the power output from potential future wind farms, speed data needs to be translated into estimates of wind farm power production using wind power curves for a particular turbine. This can be done deterministically using the wind
turbine manufacturer’s provided power curve. However doing so may not account for variation in wind speeds across the wind farm; i.e., the power output from a wind farm with \( n \) turbines will not exactly match \( n \) times the output from one wind turbine associated with a particular sensor location. Wind speeds within a wind farm will not be perfectly correlated, particularly at short time scales or in locations with variable terrain. Non-deterministic power curves have been studied previously, for example Potter et al. (2008), however doing so requires additional computation time.

In summary, large-scale wind integration studies require large quantities of wind speed and wind power data. Mesoscale meteorological models are frequently used to generate these data, but because mesoscale models were not originally designed to produce data for wind integration studies, understanding the statistical properties of the wind speed and power data that result would allow analysts to make better decisions about the circumstances under which the wind power plant models that result are representative of actual wind farms. Therefore, the goal of this research is to identify the extent to which wind speed data from meteorological models in general, and the EWITS and WWSIS data in particular, show the statistical characteristics of empirical wind speed and power data.

The remained to this chapter is presented as follows. Section 2.2 discusses the data used in this study, Section 2.3 discusses the statistical methods used, Section 2.4 presents the results, while Sections 2.5 and 2.6 presents our conclusions.

2.2 Data

2.2.1 Meteorological wind speed and power data

The largest, publicly available wind speed/power datasets generated by mesoscale modeling were produced to support two large wind (and solar) integration studies, coordinated by the U.S. National Renewable Energy Laboratory (NREL) in 2010. These two studies, the Eastern Wind Integration and Transmission Study (EWITS, see (Corbus, 2011)) and the Western Wind and Solar Integration Study (WWSIS, see (Lew and Piwko, 2010)), looked at the operational impacts of large-scale integration of wind energy across the U.S.

Each of the EWITS and WWSIS datasets contains three years of wind speed and wind power data for potential wind farm locations in the U.S., both land-based and off-shore. The EWITS dataset consists of 10-minute wind speed and plant output values for 1,326 simulated wind plants,
based on simulated data for 2004-2006. EWITS also provides next-day, six-hour and four-hour wind production forecasts for each simulated wind plant at hub heights of 80 and 100 meters. The MM5 model, with a 2 km spatial resolution, was used to produce data for EWITS. The WWSIS dataset consists of three-year (2004-2006) 10-minute simulated wind speed and plant output values for 32,043 simulated locations at a hub height of 100 meters. WWSIS does not provide forecast data. The WWSIS data come from the WRF model, also with a 2 km grid spacing.

The WWSIS and EWITS datasets use different methods to transform wind speed data into wind farm power output data. Most of the sites in the EWITS dataset represent wind farms between 100 MW and 600 MW in size, whereas each site in the WWSIS dataset represents the power output from ten Vestas V90 3-MW turbines. The EWITS power data was produced using a deterministic power curve that was the composite of multiple manufacture’s power curves (NERL, 2006). WWSIS provides three different output power datasets for each site. The first power output combines the wind speed data with a deterministic power curve. The remaining methods use statistical techniques (termed SCORE and SCORE-lite), which are based on data from several wind farms with anemometers. The SCORE technique, or Statistical Correlation to Output from Record Extension, was designed to correct for some of the statistical smoothing that results from the use of mesoscale NWP models. SCORE-lite was used without dramatically increasing the amount of computation (Potter et al., 2008). A detailed analysis of the power time-series data produced by the SCORE method can be seen in Milligan et al. (2011), where the SCORE and SCORE-lite data are compared to data from an existing wind farm.

Producing data for wind integration studies with the appropriate statistical patterns is clearly a challenging problem. However, it is clear that a prerequisite to doing so is to understand the statistical patterns of the wind speed data that are generally the starting point for wind studies. Therefore, the goal of this research is to identify the extent to which wind speed data from meteorological models in general, and the EWITS and WWSIS data in particular, show the statistical characteristics of empirical wind speed data. Understanding this will allow us to estimate the conditions under which it is appropriate to use simulated wind speed data to support wind integration studies.
2.2.2 Wind power plant data

For comparison purposes, we make use of wind speed and power production data from two wind farms located in the central United States\(^1\). The first (Plant A) is a large farm with approximately 300 MW in capacity. The second (Plant B) is smaller, with a 120 MW capacity.

Plant A consists of three years of wind speed and power data with over 250 wind turbines (approximately 300 MW capacity) and two meteorological towers. The wind speed data are taken at a height of 80 meters and are averaged over ten-minute intervals. The wind power data are also average over ten-minute intervals, and are metered individually for each turbine, as well as for the entire farm.

A number of anomalies in the data for Plant A were identified and removed before analysis. First, any negative wind speeds or non-numeric values (i.e., errors) were eliminated. Secondly, we eliminated data for several periods where the data logger appears to have been stuck. To alleviate this issue, we removed data in locations where the values were identical for three consecutive time points. Finally, for a two-month period, the anemometer appeared to have lost its calibration, and reported wind speeds that were much greater than expected (at one point it read 179 m/s). For this reason, all wind speeds over 40 m/s were eliminated from the data.

**Table 2.1:** General Characteristics of the sets created from Plant A. A map of the plant is available at (Wan et al., 2010).

<table>
<thead>
<tr>
<th>Set Number</th>
<th>Number of Turbines</th>
<th>Area Covered (km(^2))</th>
</tr>
</thead>
<tbody>
<tr>
<td>Plant A, Set 1</td>
<td>1</td>
<td>0.001</td>
</tr>
<tr>
<td>Plant A, Set 2</td>
<td>10</td>
<td>7.7</td>
</tr>
<tr>
<td>Plant A, Set 3</td>
<td>50</td>
<td>51.8</td>
</tr>
<tr>
<td>Plant A, Set 4</td>
<td>100</td>
<td>124.3</td>
</tr>
<tr>
<td>Plant A, Set 5</td>
<td>200</td>
<td>194.25</td>
</tr>
<tr>
<td>Plant A, Set 6</td>
<td>250</td>
<td>285</td>
</tr>
</tbody>
</table>

The data for Plant A was parsed into six sets based on increasing distance to the northwestern meteorological tower. These sets consist of 1, 10, 50, 100, 200, and all turbines, respectively, as described in Table 2.1, and are referred to as Plant A Set 1, . . . , Plant A Set 6. The area of each set was estimated based on the rectangle defined by the extreme latitudes and longitudes of each set.

\(^{1}\)The exact locations cannot be revealed per the terms of non-disclosure agreements
set. The area for Plant A Set 1 was assumed to be 0.25 acres, the area of a single turbine as estimated by NERL (2006). The power for each set was the sum of all turbines in the set, which created six unique time-series of wind power production. Parsing and aggregating the data in this way essentially created empirical data for six wind farms of increasing size and allowed us to study the effect of geographic diversity on the variability of the wind farm power production.

The second empirical dataset came from a smaller wind farm (Plant B) consisting of 80 turbines with a peak production capacity of 120 MW. The dataset includes only power output, sampled at 2-second time intervals, for the entire wind farm. Wind speed data were not available.

### 2.3 Data Analysis Methods

This work employs two methods for quantifying the variability in wind speed and power data. Subsections 2.3.1 and 2.3.2 discuss the power spectral density approach, which allows one to measure the magnitude of variability at different time scales. Subsection 2.3.3 discusses the step-change analysis method, which allows one to see the frequency with which changes of various magnitudes occur.

#### 2.3.1 Power spectral density analysis method

The statistical properties of wind speeds are well documented in the literature. Tuller and Brett (1984) show that the Weibull distribution can be used to model the probability distribution of wind speeds. While the relative frequency of different wind speeds is important for estimating energy production, for wind integration studies that deal with reliability issues, the amount of change in wind power that result from changes in wind speeds will be of greater importance. Power spectrum analysis provides a useful method for measuring the quantity of variability in time-series data at different time scales. The power spectral density (PSD, or power spectrum) for a stream of time-domain data $x(t)$, indicates the amount of variability in $x$ at each frequency for which the PSD was calculated (see Appendix A.1 for details on PSD calculations). Traditional PSD analysis is calculated for either voltage or power signals resulting in units of $V^2 Hz^{-1}$ or $W^2 Hz^{-1}$ depending on the source. The first part of the analysis presented in this work is focused around the PSD of wind speed data, with units of $m/s$, therefore the units resulting from this analysis will be in $(m/s)^2 Hz^{-1}$, whereas the PSD analysis of wind power that is presented in this work has been normalized resulting in units.
of capacity $^2 H z^{-1}$. For wind speed data, it is well known that atmospheric turbulence gives rise to a Kolmogorov distribution in the power spectrum of wind speeds (Kolmogorov, 1962; Oboukhov, 1962). Kolmogorov found that the relative variability (power) of turbulent flows decreases with the $5/3^{rd}$ power of the frequency. The Kolmogorov spectrum of wind data has been discussed in several recent articles (Apt, 2007; Drobinski et al., 2000; Katzenstein et al., 2010; Welter et al., 2009; Zilberman et al., 2008).

While the configuration and size of a wind farm will have some effect on the power spectrum of wind power production from a wind farm, wind speed data from any location should have a power spectral density (PSD), $y(f)$, that fits well to a model of the form in Eq. 2.1, with $\alpha \approx -5/3$, for frequencies ($f$) between $10^{-6}$ and $10^{-2}$ Hz.

$$y(f) = cf^\alpha$$  \hspace{1cm} (2.1)

where $c$ is a scaling constant. More practically, this means that for each doubling in frequency, the amount of signal power (the average of the square of the data) decreases by a factor of $2^{5/3}$. Thus, the amount of signal power in the vicinity of one cycle per hour ($2.8 \times 10^{-4}$ Hz) is $2^{5/3} = 3.17$ times larger than the amount of power in the range of one cycle per half hour ($5.6 \times 10^{-4}$ Hz). As a result of the Kolmogorov spectrum, changes in wind speeds at different time scales have fat-tailed probability density functions (Boettcher et al., 2003).

The remainder of this section describes the method that we use to compare the power spectrum of WWSIS and EWITS wind speed data to the Kolmogorov spectrum. For the Eastern wind data, we analyzed wind speeds for both 80 and 100-meter hub heights. The PSD was computed for each of the 1,326 sites and 2 hub heights in the EWITS data, and 3,200 sites (about 10%) for the Western wind dataset. The WWSIS subset was carefully checked to ensure that no one geographic region was overly represented in the sample. In order to account for geographic diversity we focus our analysis on the average PSD across these sites.

This PSD analysis was replicated for the power output of each dataset. For EWITS we analyzed wind power at 80 and 100-meter hub heights, and for the WWSIS dataset we analyzed the power outputs of the deterministic and SCORE-lite power curves. The PSD for power output was calculated across the 1,326 EWITS sites and the same 3,200 sites used in the WWSIS analysis for wind speeds. As with the previous PSD analysis focusing on wind speeds we have averaged the power output PSD across all sites in order for our results to remain consistent.
Both WWSIS and EWITS include three years of data in separate files, which we combine into a single wind speed time series for each site and height. We use Bartlett’s method for segment averaging (Bartlett, 1948) when calculating the PSD of a site. This method minimizes the effect of noise on the data the PSD without dampening the actual signal variance. To implement Bartlett’s method, we divide our time-series data into and calculate the PSD for 30 equally sized data segments (10 segments per year). Each of these PSD values is subsequently averaged at each frequency to obtain an average power spectrum for each site. If \( y_i(f) \) is the PSD for segment \( i \) at frequency \( f \) the average spectral density \( \overline{y(f)} \) is:

\[
\overline{y(f)} = \frac{1}{30} \sum_{i=1}^{30} y_i(f)^2
\]  

(2.2)

The appendix includes a discussion of different approaches to power spectral analysis, and their impact on the outcomes (Appendix A.2). Using these spectra, we use least squares estimation to find the exponent \( \alpha \) and offset \( c \) from Eq. 2.1 for each site. Figure 2.1 shows the segment averaged PSD for calculated sites given median, highest and lowest exponents for the EWITS and WWSIS datasets, which clearly fall below the Kolmogorov spectrum \( (\alpha < -5/3) \). These three sites were taken from the processed EWITS and WWSIS datasets in order to show that the no wind speed data from a single site within either dataset adhered to the Kolmogorov spectrum.

To obtain an aggregated PSD for each dataset, we averaged the PSD estimates at each frequency, across all sites. As with the individual sites, the PSD exponent is well below that of the Kolmogorov spectrum. However, on further investigation, this steep slope appears to come from the higher frequency portion of the spectrum.

2.3.2 Model fitting to find the point at which PSD diverges from the Kolmogorov

Our initial analysis of the EWITS and WWSIS datasets indicates that there are distinct frequency ranges within the data. That is to say that the PSD of the wind speeds for the mesoscale model predicted wind speeds appears to match the Kolmogorov spectrum for low frequencies and at higher frequencies the spectral densities fall faster than the \( f^{-5/3} \) line. In order to estimate the frequency
Figure 2.1: Power spectral density plots for calculated EWITS and WWSIS individual sites with lowest (A), median (B), and highest (C) power-law slope estimates.
at which this change in slope occurs we fit the data to a model ($\hat{y}(f)$) of the following form:

$$\hat{y}(f) = \begin{cases} c_1 f^{\alpha_1}, & f < f_k \\ c_2 f^{\alpha_2}, & f \geq f_k \end{cases} \quad (2.3)$$

where $\alpha_1$ and $\alpha_2$ are the low and high frequency exponents, $c_1$ and $c_2$ are scaling constants, and $f_k$ ("knee") is the frequency at which the change in slope occurs. To ensure that $\hat{y}(f)$ was not discontinuous at the knee point, we calculate the scaling parameters $c_1$ and $c_2$ to solve:

$$c_1 f_k^{\alpha_1} = c_2 f_k^{\alpha_2} \quad (2.4)$$

$$c_2 = \frac{c_1 f_k^{\alpha_1 - \alpha_2}}{c_1 f_k^{\alpha_2}}$$

We identified ordinary least squares estimates for the slope and scaling parameters for a range of $f_k$ values and chose $f_k$ to be the value that minimized the coefficient of determination ($R^2$) for the model as a whole. This is roughly equivalent to solving for $f_k$ in Eq. 2.5,

$$\min_{c_1, \alpha_1, \alpha_2, f_k} \sum_{i=1}^{N} (\hat{y}(f_i, f_k, c_1, \alpha_1, \alpha_2) - y(f_i))^2 \quad (2.5)$$

where $N$ is the number of frequencies in the PSD, $f_i$ represents an individual frequency within the PSD calculation, and $\hat{y}(\ldots)$ represents the estimated PSD, based on Eqs. (2.3) - (2.4).

### 2.3.3 Step-change analysis method

The PSD method allows one to measure the relative magnitude of variability at different time scales, but does not give one much information about how frequently changes of a certain sizes occur. Knowing how frequently large step changes (such as a 50% reduction in power output in a ten minute period) occur can be useful in estimating the reliability impact of large wind farms in a particular area, as well as estimating the amount of fast-ramping generation needed to balance supply and demand.

To implement the step-change method (see, (Boettcher et al., 2003; Kempton et al., 2010; Sinden, 2007)), we calculated the average wind speed or power for either 10-minute or 1-hour intervals, and then compute the difference between adjacent intervals. For each dataset, we computed both empirical probability density functions (PDF) and complementary cumulative distribution functions (CCDF).

In order to compare the real and synthetic step change data we identified proximate sites for both plants from the EWITS and WWSIS datasets. For Plant A we compared the data to those
from the six closest EWITS and the three closest WWSIS sites, all of which are less than thirty miles from this site. For Plant B we compared the empirical data to data from the three closed EWITS sites, all of which are less than 20 miles away.

The wind speed data for Plant A came from an 80-meter anemometer tower, whereas the EWITS dataset includes data for both 80 and 100-meter heights. WWSIS reports data only for 100-meter hub heights. For the step change analysis we compared the 80-meter EWITS and the 100-meter WWSIS data to the wind farm data.

As previously stated the WWSIS data come from two different methods for translating the wind speeds into power data. The first utilizes a deterministic power curve, based on the manufacturer’s specifications, while the second method is meant to account for the stochastic nature of wind power and is called the SCORE-Lite method. For Plant A we compared both the deterministic time-series and the SCORE-Lite corrected time-series to the real wind farm data.

2.4 Results

2.4.1 PSD analysis results for wind speed data

The single slope least squares fit (Eq. 2.1) of anemometer data from Plant A yielded a power-law slope of $\alpha = -1.68$. As one would expect from previous literature (Apt, 2007; Boettcher et al., 2003; Drobinski et al., 2000; Katzenstein et al., 2010; Welter et al., 2009; Zilberman et al., 2008), the power spectral density of the anemometer data from Plant A follows the Kolmogorov slope over a large frequency range (Figure 2.2). In this case the Kolmogorov frequency range spans from $3 \times 10^{-6}$ to $1 \times 10^{-3}$ hertz.
Figure 2.2: Power spectral density diagram of empirical wind speeds taken from Plant A, along with the fitted power-law slope, which is close to the Kolmogorov spectrum of $f^{-5/3}$, represented by the solid line.

Figure 2.3: Power spectral density results for 80-meter hub height wind speed data from EWITS (right) and 100-meter hub height wind speeds from WWSIS (left). Panels (A) and (B) show the unsegmented PSD results along with the Kolmogorov spectrum (red lines) and the linear best-fit (green lines). Panels (C) and (D) show the segment-averaged PSDs of the EWITS and WWSIS datasets, respectively, divided by the $f^{-5/3}$, meaning that the Kolmogorov spectrum would be a horizontal line with a value of one (in red). Panels (E) and (F) represent the segmented PSD results and the best-fit slopes for frequencies greater and less than the calculated $f_k$ value for the given dataset.
Table 2.2: PSD parameter estimates for the EWITS and WWSIS wind speed spectral density models

<table>
<thead>
<tr>
<th></th>
<th>EWITS, 80 M hub height</th>
<th>EWITS, 100 M hub height</th>
<th>WWSIS, 100 M hub height</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Estimate for the point at which PSD deviates from Kolmogorov spectrum</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$f_k$</td>
<td>$4.70 \times 10^{-5}$ Hz</td>
<td>$4.50 \times 10^{-5}$ Hz</td>
<td>$4.50 \times 10^{-5}$ Hz</td>
</tr>
<tr>
<td><strong>PSD exponent estimate for Eq. 2.1, for averaged PSD with segment averaging</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\alpha(10^{-6} &lt; f)$</td>
<td>-2.2309</td>
<td>-2.2521</td>
<td>-2.2343</td>
</tr>
<tr>
<td>$R^2(10^{-6} &lt; f)$</td>
<td>0.9928</td>
<td>0.9925</td>
<td>0.9936</td>
</tr>
<tr>
<td><strong>PSD exponent estimates for Eq. 2.3, for averaged PSD without segment averaging</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\alpha_1(10^{-6} &lt; f &lt; f_k)$</td>
<td>-1.6094</td>
<td>-1.6067</td>
<td>-1.6548</td>
</tr>
<tr>
<td>$\alpha_2(f \geq f_k)$</td>
<td>-2.3890</td>
<td>-2.4127</td>
<td>-2.3884</td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.9974</td>
<td>0.9973</td>
<td>0.9979</td>
</tr>
<tr>
<td><strong>PSD exponent estimates for Eq. 2.3, for averaged PSD with segment averaging</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\alpha_1(10^{-6} &lt; f &lt; f_k)$</td>
<td>-1.6290</td>
<td>-1.6206</td>
<td>-1.6508</td>
</tr>
<tr>
<td>$\alpha_2(f \geq f_k)$</td>
<td>-2.3693</td>
<td>-2.3907</td>
<td>-2.3623</td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.9981</td>
<td>0.9980</td>
<td>0.9984</td>
</tr>
</tbody>
</table>
In contrast, the PSD of the wind speed data from EWITS and WWSIS does not follow the Kolmogorov spectrum well; this can be seen in Figure 2.2 which shows the average power spectral density results for EWITS and WWSIS (both at 100 m), along with the fit lines from Eqs. (2.1) and (2.3). Panels A and B in Figure 2.3 compare the single slope least squares fit (Eq. 2.1) with the averaged PSD data. A substantial difference between the single log-linear fit and the Kolmogorov spectral density data is clear. Panels C and D show the same PSDs, as in A and B, normalized by $f^{-5/3}$. The Kolmogorov spectrum normalized in such a way would result in a horizontal line with a value of one. This method of visualization helps to show the frequency range where EWITS and WWSIS data match the Kolmogorov spectrum. Panels E and F of Figure 2.3 show the results from fitting the PSDs with separate slopes above and below the knee frequency $f_k$.

In this case, the model fit is improved, as reflected by the decrease in the mean squared error from 0.032 to 0.0097. For frequencies between $10^{-6}$ and $f_k$ the NWP wind speed data follow the Kolmogorov spectrum closely. This is apparent for all three datasets, with both EWITS datasets showing a slope of $\alpha_1 = -1.63$, and with $\alpha_1 = -1.65$ for WWSIS (See Table 2.2). This indicates that mesoscale models at this spatial and temporal resolution approximately reproduce the statistics of wind speed variability for frequencies less than one cycle per 3-6 hours. However, for frequencies

![Figure 2.4: A comparison of the goodness of fit between the two-slope model in Eq. 2.2, and the empirical power spectral density for each dataset, as a function of the cutoff, $f_k$, between the two slopes. The left extreme essentially shows $R^2$ for the single-slope model](image)
above one cycle per 6 hours the spectral density decreases with $f^{-2.36}$ for both EWITS and WWSIS. Table 2.2 gives the results for these fit parameters. In order to validate the models used in this analysis a comparison between the methods used in this work and Welch’s Method can be seen in Appendix A.2.

Both the EWITS and WWSIS begin to diverge from the Kolmogorov spectrum above frequencies of $4.5 \times 10^{-5}$ Hz, or about 1 cycle per 6 hours. Figure 2.4 shows the goodness-of-fit ($R^2$) scores for the model in Eq. 2.3, with different values of $f_k$. When the knee value is set to values greater than $10^{-4}$ Hz (about one cycle per 3 hours) the model quality begins to degrade noticeably. The model also produces poor fits when the knee value is set to values less than $10^{-5}$ Hz (about one cycle per day).

**Figure 2.5:** Box plot depicting all calculated single power-law exponents ($\alpha$ from Eq. 2.1) for each dataset with a 95% confidence interval.

To visualize the diversity among the various sites, Figure 2.5 shows the distributions of single power-law exponent estimates ($\alpha$ from Eq. 2.1) of all individual sites for the three datasets. This
shows that all sites have average slopes that are well below the $f^{-5/3}$, represented by the solid line within the boxes, which is generally observed in real wind speed data, while also illustrating that the average slopes are approximately the same for all three datasets. Perhaps due to the high spatial resolution used in the numerical weather prediction model, the WWSIS data generally came closer to the Kolmogorov spectrum, relative to the EWITS data. It is also important to note that all of the sites have single power-law exponents which are greater significantly than that predicted by the Kolmogorov spectrum.

This reduction in spectral density will have only a small effect on wind speed variability at time scales in the range of 3-6 hours, but will have a more substantial effect at higher frequencies. In order to quantify the magnitude of this effect, we followed the following procedure. Given a spectral density $y$, Eq. 2.6 gives the reduction in spectral density between a knee frequency (e.g., $f_k \approx 6^{-1}\text{hours}^{-1}$) and a higher frequency $f_h$.

$$\frac{y(f_h)}{y(f_k)} = \frac{c_2}{c_2 f_h^\alpha} = \left(\frac{f_h}{f_k}\right)^\alpha$$ (2.6)

To find the loss in spectral density at frequency $f_h$ that results from the exponent $\tilde{\alpha}_2$ being steeper than $\alpha_2$, we can use (2.6) to obtain:

$$\frac{\tilde{y}(f_h)}{y(f_k)} = \left(\frac{f_h}{f_k}\right)^{\alpha_2 - \tilde{\alpha}_2}$$ (2.7)

This assumes that $y(f_k) = \tilde{y}(f_k)$, as is the case for our model. Equation 2.7 gives the relative change in spectral density at $f_h$, due to the steeper slope $\tilde{\alpha}_2$. Thus, Eq. 2.7 indicates that with $\tilde{\alpha}_2 = 2.4$ the power spectral density for 15-minute cycles ($f_h = 1.1 \times 10^{-3} \text{Hz}$) is reduced by a factor of 0.097, or by roughly 90 percent.

### 2.4.2 PSD analysis results for wind power data

As a comparison to the PSD of wind speed we also analyzed the PSD of wind power for the EWITS and WWSIS datasets, as well as the PSD from the two plants and their corresponding geographically proximate EWITS and WWSIS sites.

Figure 2.6 shows the results for the segmented averaged PSD analysis for the power output of the EWITS and WWSIS datasets, along with their corresponding linear best-fit lines. This figure shows that the variability seen within the power output of both EWITS datasets (80-meter and 100-meter hub heights) and the deterministic power curve of the WWSIS dataset are relatively
the same, whereas there is significantly more variability within the SCORE-lite datasets. This difference between the deterministic and SCORE-lite data given that the purpose of SCORE-lite is to add variability at high frequencies in order to compensate for the averaging of the NWP models. In the next step of our analysis we look to compare the PSD of wind power outputs seen in EWITS and WWSIS with empirical wind power outputs.

![Figure 2.6: Power spectral density results for 80 and 100-meter hub height wind power data from EWITS (right) and 100-meter hub height wind power data, deterministic and SCORE-lite, from WWSIS (left). Each dataset also is shown with its corresponding linear best-fit line (green lines)](image)

Figure 2.6: Power spectral density results for 80 and 100-meter hub height wind power data from EWITS (right) and 100-meter hub height wind power data, deterministic and SCORE-lite, from WWSIS (left). Each dataset also is shown with its corresponding linear best-fit line (green lines)

Figure 2.7 shows the results for the segmented averaged PSD analysis for the power output of Plant A Set 6 (as well as the linear best-fit line of Plant A Set 1) and Plant B compared to the linear best-fit lines of their corresponding EWITS and WWSIS sites. From this figure it is clear that Plant A Set 6 has a steeper slope than the power output of both EWITS and WWSIS (deterministic and SCORE-lite) power output data, however the difference between Plant A Set 6 and EWITS is minimal compared to the difference with WWSIS. The linear best-fit of Plant A Set 1 illustrates how decreasing the relative plant size increases the variability within the data, which is expected due to the averaging effect of geographical diversity, even within a single wind plant. Table 2.3 shows this decreasing variability by tabulating the linear best-fit slopes of all 6 sets of Plant A. These results
suggest that both the power outputs of EWITS and WWSIS (deterministic power curve) contain more variability than Set 6 of Plant A, but less variability than Set 1. Figure 2.10 further explores these results using step-change analysis and shows that the PSD analysis presented here does not conclusively show the behavior present in these datasets.

![Figure 2.7: Power spectral density results for wind power data from Plant A (top) and wind power data from Plant B (bottom), compared to their corresponding linear best-fit lines from the EWITS and WWSIS datasets](image)

**Table 2.3:** Linear best-fit slopes for the 6 sets of Plant A

<table>
<thead>
<tr>
<th>Set Number</th>
<th>Linear Best-Fit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Plant A, Set 1</td>
<td>-1.8185</td>
</tr>
<tr>
<td>Plant A, Set 2</td>
<td>-2.1088</td>
</tr>
<tr>
<td>Plant A, Set 3</td>
<td>-2.3639</td>
</tr>
<tr>
<td>Plant A, Set 4</td>
<td>-2.4908</td>
</tr>
<tr>
<td>Plant A, Set 5</td>
<td>-2.5652</td>
</tr>
<tr>
<td>Plant A, Set 6</td>
<td>-2.5947</td>
</tr>
</tbody>
</table>
Figure 2.7 also shows that the relative variability of the corresponding EWITS sites to Plant B produce a linear best-fit slope that is slightly steeper than that found in empirical wind power data. This is further explored in Figure 2.9 using step-change analysis and shows that corresponding EWITS sites contain slightly less variability than that contained within the wind power output data of Plant B.

2.4.3 Step change analysis results for wind speed data

In order to better understand the reduced variability indicated by the PSD analysis, this section compares empirical and simulated wind speed data using the step-change analysis method. Figure 2.8 compares the probability of 10-minute and 1-hour step changes of the real wind speeds from Plant A to those in wind speeds from proximate sites in the EWITS and WWSIS data.

![Graph showing CCDF for step-changes in wind speed for data from Plant A and the nearest EWITS and WWSIS plants, for 10-minute (above) and 1-hour (below) averages. Each point shows the probability of an absolute change that is greater than or equal to the value on the X-axis.](image)

**Figure 2.8:** CCDF for step-changes in wind speed for data from Plant A and the nearest EWITS and WWSIS plants, for 10-minute (above) and 1-hour (below) averages. Each point shows the probability of an absolute change that is greater than or equal to the value on the X-axis.
The probabilities of 2-6 m/s changes in wind speeds in the simulated EWITS and WWSIS 10-minute data are substantially (almost an order of magnitude) smaller, compared to the empirical 10-minute data. This results is expected based on the PSD analysis that showed reduced variability in the EWITS and WWSIS datasets at high frequencies. For larger step changes (greater than 7 m/s) the WWSIS and empirical data show similar probabilities to the empirical data, whereas the EWITS data continues to show less variability. As one would expect from the PSD analysis, when the wind speeds are averaged over hourly time intervals, the differences between the simulated and empirical data are less obvious. The hourly step-change probabilities are a factor of 2-5 reduced in the simulated data, relative to the empirical data.

Please note that the 1-hour synthetic step-changes agree very closely with the empirical data. However the 10-minute synthetic is far less variable than the empirical 10-minute data. This is to be expected based on the PSD analysis previously discussed in Subsection 2.3.1. The 1-hour variability corresponds to a frequency of \(2.78 \times 10^{-4}\) Hz, which is very close to the knee frequency found to be approximately \(4.5 \times 10^{-5}\) Hz, meaning that synthetic PSD has not greatly deviated from the actual PSD. However 10-minute variability corresponds to \(1.67 \times 10^{-3}\) Hz, where the PSD of the synthetic wind has greatly deviated from the real PSD.

### 2.4.4 Step change analysis results for wind power data

Figure 2.7 shows that the slopes of the PSD of the simulated wind power output are shallower than that of the empirical data of Plant A, while the slope of the PSD of the simulated wind power output are slightly steeper than the empirical data of Plant B. This suggests that simulated data has less variability than the wind power output data from Plant B and more variability than the output of Plant A.

The second comparison made was of the variability of real wind power to wind power data from mesoscale model predicted wind farms. Figure 2.9 shows the comparison for Wind Plant B and its companion EWITS wind farm.
Figure 2.9: PDF and CCDF comparison of the real wind power output of Plant B and the synthetic wind power obtained from a single EWITS site for step changes of 10 minutes and 1 hour respectively.

From the graphs in Figure 2.9 it is clear that the synthetic data created in the EWITS 10-minute dataset under-predicts the variability for this particular wind farm. From the CCDF in Figure 2.9 one can observe the probability of a 10% or larger step change, over a 10-minute interval is approximately 0.06 for the real wind farm, while the probability of the same change in the EWITS data is less than 0.02, which is almost 4 times less than the probability seen in real wind. At a 50% step change over a 10-minute interval it is observed that there is a probability of 0.0004 for the real wind and 0.0002 for the EWITS data, which is two times less than the probability of the real wind. For one one-hour interval, the wind farm shows a 0.30 probability for a 10% or greater change in wind power, whereas the EWITS data show only about a 0.20 probability. Whereas the probability for a 50% change in one-hour output power is approximately 0.0015 and 0.0030 for EWITS data.

27
and the wind farm data, respectively.

Figure 2.10: PDF and CCDF comparison of the real wind power output of Plant A, the synthetic wind power from its nearest EWITS plant, and the synthetic wind power from its nearest WWSIS plant for both deterministic and SCORE-lite power curves at uniform step changes of 10 minutes.

The same comparison between Plant A and its companion EWITS plant shows far less dramatic results (Figure 2.10 and Figure 2.11). Note that the CCDF’s of Plant A and EWITS agree much more closely than in Figure 2.9, for both the 1-hour and 10-minute data. However the comparison between Plant A and its companion WWSIS plant show a dramatic difference between both its EWITS counterpart and the empirical data observed.

The discrepancy, seen in Plant A and EWITS, between the two comparisons (in Figure 2.9, Figure 2.10, and Figure 2.11) starts to make sense when we look at Plant A Set 1 and how it compares to
the same EWITS farm (Figure 2.10 and Figure 2.11). This result agrees much more with that of
Plant B (Figure 2.9). The key difference between these comparisons is size of the empirical wind
farm: Wind Farm B is 120 MW, Plant A over 300 MW, and Plant A set 1 is 1 MW while Set 3
equals approximately 50 MW.

Figure 2.11: PDF and CCDF comparison of the real wind power output of Plant A, the synthetic
wind power from its nearest EWITS plant, and the synthetic wind power from its nearest WWSIS
plant for both deterministic and SCORE-lite power curves at uniform step changes of 1 hour

Figure 2.10 and Figure 2.11 shows how the CCDF of Plant A changes with the different subsets
we created (see Table 2.1). As more turbines are added the distribution of Plant A converges with
the EWITS wind farm. In the final set, consisting of all the turbines in Plant A, the EWITS farm
data has more variability than the empirical data (Plant A set 6).

However, the discrepancies seen in Figure 2.10 and Figure 2.11 between Plant A and WWSIS
are not correlated to the size of the wind farm, as seen with EWITS. Instead Figure 2.10 shows
that for a 10-minute step-change the deterministic and SCORE-lite WWSIS plant data begins to
converge just above a 33% step-change in capacity, at a value that over predicts the variability seen
in empirical data. Figure 2.10 shows that for a 10-minute step-change at high percent changes both
the deterministic and SCORE-lite data produce wind variability that is greater than the variability seen in all 6 sets of Plant A. While Figure 2.11 shows for a 1-hour step-change the deterministic and SCORE-lite WWSIS variability is approximately the same as the variability seen in Plant A for sets 3 and 6.

In analyzing the wind speed data we have been focused on the affect of the smoothing of the mesoscale model on the synthetic data. The convergence of distributions of power set-changes with increasing number of turbines suggest that geographic diversity, which has a smoothing effect on the real power output of a wind farm, needs to be considered in this comparison.

Figure 2.12 and Figure 2.13 attempt to compare the smoothing effect mesoscale models to the smoothing effect of geographic diversity. On the ordinate these figures plot the likelihood of a given sized step-change in power for the different Plant A sets, the six closest EWITS farms to Plant A, and the three closest WWSIS farms to Plant A; the abscissa is the estimated area of each of these wind farms. Figure 2.12 is comprised of 10-minute data while Figure 2.11 uses 1-hour data.

**Figure 2.12:** Probability of a change in wind power output, over a ten minute period, that is greater than or equal to \{10, 25, 50, or 75\}% of plant capacity, for wind farms of various sizes. The Plant A sets are plotted as solid lines with X’s marking the actual data points, the synthetic data are plotted as markers, and the colors represent different sizes of step-changes in power output of the wind farm, normalized by wind farm capacity.
Figure 2.13: Probability of a change in wind power output, over an hour period, that is greater than or equal to \{10, 25, 50, or 75\}\% of plant capacity, for wind farms of various sizes. The Plant A sets are plotted as solid lines with X’s marking the actual data points, the synthetic data are plotted as markers, and the colors represent different sizes of step-changes in power output of the wind farm, normalized by wind farm capacity.

Table 2.4: The largest step-change in power with a non-zero probability

<table>
<thead>
<tr>
<th></th>
<th>Largest power step change with $p &gt; 0$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>10-minute</td>
</tr>
<tr>
<td>Farm A Set 4</td>
<td>83%</td>
</tr>
<tr>
<td>Farm A Set 5</td>
<td>83%</td>
</tr>
<tr>
<td>Farm A Set 6</td>
<td>82%</td>
</tr>
<tr>
<td>Farm B</td>
<td>83%</td>
</tr>
<tr>
<td>EWITS 1</td>
<td>63%</td>
</tr>
<tr>
<td>EWITS 2</td>
<td>74%</td>
</tr>
<tr>
<td>EWITS 3</td>
<td>76%</td>
</tr>
<tr>
<td>EWITS 4</td>
<td>71%</td>
</tr>
<tr>
<td>EWITS 5</td>
<td>82%</td>
</tr>
<tr>
<td>EWITS 6</td>
<td>82%</td>
</tr>
</tbody>
</table>
A final finding that is worth noting is that the mesoscale predicted power data from EWITS had less weight in the tails of the distributions than would be expected. For example, many of the EWTIS sites that we looked at did not predict step changes greater than 75%. To illustrate this in more detail is Table 2.4 contains the maximum power step-change that has a non-zero probability for each of the plants we analyzed.

2.5 Discussion

The results presented in this chapter indicate that the synthetic wind data that comes from mesoscale meteorological models, such as those used for EWITS and WWSIS, underestimate the variability in wind speed for frequencies above one cycle per 3 hours. The meteorological model data begins to diverge from the Kolmogorov power spectral density at about one cycle per 6 hours, but the divergence is not substantial for frequencies below one cycle per 3 hours. However for higher frequencies this divergence is significant, and appears to result in notable differences between the variability in actual wind farms and that of simulated wind farms from the same location.

This difference is likely to be important to some types of analysis, but not others. Given the large number of measurements that feed mesoscale climate models, the data will likely provide good estimates for the annual energy production from potential wind farms. Also, these results indicate that the mesoscale model data will provide a useful estimate of the day-to-day variability in wind speeds; even the fluctuations at faster time scales, up to one cycle per three hours, remain close to the Kolmogorov spectrum. This means that the mesoscale data can be used for estimating the impact of large-scale wind power production on unit commitment costs and related problems that span hours to days, with some confidence in the statistical outcomes.

However, for problems that require substantial information about the variability in wind farm production over sub-hourly time scales, mesoscale model data need to be treated with some caution. This is particularly true for power grid reliability problems, where understanding the extreme cases is vital. For example, in order to estimate the quantity of load following resources required to satisfy reliability requirements, data with fairly good accuracy at time scales in the range of 5-minutes to hourly is needed. Furthermore, to estimate the impact of large-scale wind integration on regulating reserve requirements wind production data that are accurate over time scales of seconds to minutes is required. An underestimation of wind speed variability along these times will likely lead to
an underestimation of power plant capacities and ramp rates required to maintain grid reliability under high wind penetration scenarios. Similarly, in order to quantify the amount of demand side participation or storage capabilities required to mitigate wind variability, higher fidelity data are likely to be needed. An underestimation of wind variability will lead to an underestimation in required balancing resources independent of how it is provided. The underestimation of variability may also affect analysis of hourly dispatch costs, though to a lesser degree.

A caveat to this conclusion needs to be made. The mapping between single point wind speed measurements and the output from an entire wind farm with many turbines will necessarily result in a filtering of some high-frequency variability in wind speeds. As evidence of this, Milligan et al. (2011) suggest that some methods that map between wind speed data and wind farm power production can result in too much variability. Our analysis of the step change probabilities for one site indicates that this filtering does not eliminate the errors that result from reduced variability in the simulated wind speed data. A filtering effect will also arise from aggregating wind generation at various locations. This aggregation will likely reduce, but will not eliminate the impacts of the decreased variability reported here. Further research is needed to estimate the magnitude of these effects for different wind farm configurations, using high fidelity wind speed and power production data.

This research suggests several directions for future research. This work provides evidence that mesoscale models do a reasonably good job of capturing wind speed dynamics for time scales that are slower than 3-6 hours. A combination of factors result in less accurate results for faster dynamics. Firstly, some of the input data for mesoscale reanalysis models are populated with fewer than 4 samples per day. Because of this, reanalysis cases will naturally interpolate between these points, producing smoothing relative to empirical data. Secondly, the mesoscale (2 km) grid spacing will make it difficult to model atmospheric turbulence. Because turbulence occurs over spatial scales less than 2 km, and across complicated geographic features, mesoscale models will average wind speeds to some extent. Microscale models, also known as Large Eddy Simulation models, are able to capture faster and more fine-grained spatial phenomena, and thus may produce more accurate wind speed data. This suggests that mesoscale models might be useful in combination with microscale models to produce data that have accurate statistics at multiple time scales. The framework of multiscale modeling (e.g., Famiglietti and Wood (1994)) may guide future work in this area.
2.6 Conclusion

It is important to recognize that even though these data sets underestimate the variability of wind speeds, the EWITS and WWSIS studies currently include the best data sets publicly available to researchers. Because public data make detailed and reproducible analysis possible, it is critical that empirical data be made available for research as new wind projects are developed and wind penetration increases. High-resolution (sample frequency of greater than one datum per minute) production, wind speed, and solar radiation data should be made publicly available through the utilities or independent system operators. Such data could be used to improve the understanding of challenges and opportunities for wind power integration.
Chapter 3

Estimating Regulating Reserve Requirements for High Penetration Wind Scenarios

3.1 Introduction

The statistical nature of wind variability will likely have a major impact on the need for balancing ancillary services in systems with significant quantities of wind generation. Power systems with large amounts of wind, or other renewable energy sources, will likely become more prominent as the world looks to move away from fossil fuels and towards renewable fuel supplies. The need for increased integration of renewables in our power grids is based off the knowledge that our current fuel supplies are finite. The BP (2011) Statistical Review estimates under current consumption rates, oil supplies will be used up in 46.2 years, natural gas supplies will be depleted in 58.6 years, and the world’s coal supply will be exhausted in 118 years. Of course these numbers are based on current proven reserves and current world energy usage, which according to the U.S. Energy Information Administration (Conti and Holtberg, 2011), is projected to increase to 770 quadrillion BTU by 2035. As a reference in 2008 the total energy consumption in the world was 505 quadrillion BTU, which means that the projected value in 2035 is a 49% increase from the total energy consumed only three years ago. This increase in the world’s energy consumption, coupled with the remaining of fossil fuels indicates how
important it is to dramatically increase the use of renewable energies in the power grid. As a direct consequence much research needs to be done in order to determine the amount of balancing ancillary services that systems will need, in order to compensate for increased renewable energy penetrations.

The research presented here will focus on increasing penetrations of wind due to its quantity cost compared to other renewable sources, and the rate at which wind power generation is growing. As a result of this rapid growth in wind generation, numerous large-scale wind integration studies have been produced to estimate the reliability impacts of large-scale wind power and the feasibility of different levels of wind power penetration (Corbus, 2011; Hinkle et al., 2010; Lew and Piwko, 2010; Loutan and Hawkins, 2007; Tsuchida et al., 2010). These studies are crucial in helping researchers to better understand the reliability impacts of high wind penetration scenarios on the grid as wind penetration levels are expected to increase dramatically in the years to come. According the World Wind Energy Report (WW EA, 2011) wind power projections are as high as 600 GW globally in 2015 and 1.5 TW in 2020. This report along with the Renewables 2011 Global Status Report (Sawin et al., 2011) report that in 2010 the world’s total wind capacity was rated at 198 GW with 40.2 GW located within the United States. As a comparison in 1996 the total wind power capacity installed in the world was equal to 6.1 GW. In 2010 alone, 39 GW of new wind capacity was installed globally, with 5 GW in the U.S. This is three time the 11.5 GW of wind that was installed globally just five years ago.

The majority of existing wind integration studies implicitly or explicitly assume that the variability and/or prediction error of wind power production can be modeled as a Gaussian random process (Jaramillo and Hines, 2010). Doherty and O’Malley (2005) propose a model for estimating reserve requirements, and treat forecast errors for both electricity demand and wind as normal random variables. Similar Gaussian assumptions are present in (Matos and Bessa, 2011; Meibom et al., 2010; Ortega-Vazquez and Kirschen, 2009; Yong et al., 2009). However real wind data deviate substantially from Gaussian models (Apt, 2007). Some note this deviation (Matos and Bessa, 2011; Ortega-Vazquez and Kirschen, 2009), but continue to use normal random variables to model wind production or forecast errors. Ortega-Vazquez and Kirschen (2009) justify their usage of a normal assumption by referencing geographic diversity and the central limit theorem.

In order to evaluate these assumptions, the goal of this research is to develop and evaluate methods for estimating reserve requirements in systems with significant amounts of wind generation, based on the measured statistical properties of wind, rather than relying on Gaussian models. Matos
and Bessa (2011) also evaluate new methods for estimating primary (regulating) and secondary (load-following or spinning) reserve requirements, but continue to use the standard deviation as the primary measure of variability, which implicitly assumes Gaussian behavior. A recent industry study (Tsuchida et al., 2010) presents a method for estimating regulating reserves using 95/5% confidence intervals rather than the second-order moment, but do not extensively evaluate the new approach.

Our estimation method aims to estimate the amount of regulating reserves required to satisfy NERC’s BAL-001 standards (NERC, 2012). The standard BAL-001 is known as real power balancing control performance and is used to control frequency and power limits within each balancing authority. The beginning of Section 3.2 shows the mathematical formulation and implantation of the BAL-001 standard.

The remaining to this Chapter is presented as follows. Section 3.2 discusses the methodology developed in our work, Section 3.3 discusses the data used, Section 3.4 presents the results obtained from our models, and Sections 3.5 and 3.6 presents our conclusions.

### 3.2 Regulation Modeling Methodology

The goal of this research project is to determine the quantity of regulation needed within a system with increasing levels of wind penetration in order to keep the system in compliance with NERC reliability standards. In order to accomplish this goal this work is focused on adhering to the same main standard that is practiced in industry, which is to insure that our system is in compliance with the NERC standard BAL-001. The purpose of BAL-001 is to maintain interconnection steady-state frequency within predetermined limits by controlling and balancing the real power demand within the system and the real-time supply produced by the generators. This standard is sequentially broken apart into separate requirements. The two that are the focus of this chapter are known as Control Performance Standard 1 (CPS1) and Control Performance Standard 2 (CPS2).

The first requirement that this work addresses is CPS1, which is shown below in Eq. 3.1,

\[
\text{CPS1} = (2 - \text{CF}) \times 100%
\]

where CPS1 must be greater than or equal to 100% for an average of all clock minute averages over a rolling 12-month basis. In Eq. 3.1 the compliance factor, or CF, is the ratio of all one-minute
compliance parameters over the rolling 12-month basis as shown in Eq. 3.2 below,

\[ CF = \frac{AVG_{Period}[\left(\frac{\text{ACE}_{i}}{10B_i}\right) \ast \Delta F_i]}{\epsilon_1^2} \]  

(3.2)

where the average of the clock minute averages over the rolling 12-month basis must be less than a specified limit known as \( \epsilon_1^2 \). The constant \( \epsilon_1 \) is derived from the targeted frequency bound, which is separately calculated for each interconnection, and is the targeted RMS value of the clock minute averages of the frequency error over a given year. In the above equation the formal definition of ACE is shown in Eq. 3.3,

\[ ACE = (NIA - NIS) - 10B(F_A - F_S) - I_{ME} \]  

(3.3)

where \( NIA \) and \( NIS \) represent the algebraic sum of the actual and scheduled flows on all tie lines, \( B \) represents the frequency bias setting (MW/0.1Hz) for the balancing authority, \( F_A \) and \( F_S \) represent the actual and scheduled frequencies and \( I_{ME} \) is the meter error correction factor which is normally very small or equal to zero.

The second requirement of BAL-001 that our work must satisfy is known as CPS2 and is shown below in Eq. 3.4.

\[ CPS2 = [1 - \frac{V_{\text{month}}}{(TP_{\text{month}} - UP_{\text{month}})}] \ast 100\% \]  

(3.4)

In Eq. 3.4 the variable \( TP_{\text{month}} \) represents the total periods within the month, \( UP_{\text{month}} \) represents the number of unavailable periods in each month, and \( V_{\text{month}} \) refers to the number of violations per month where a violation occurs when Eq. 3.5 fails to be satisfied.

\[ \frac{1}{10} \int_{t_i}^{t_i+10} ACE(t_i) \leq L_{10} \]  

(3.5)

Where \( L_{10} \) is the targeted root-mean square (RMS) value of the ten-minute averages of the frequency error over a given year and is defined by Eq. 3.6.

\[ L_{10} = 1.65\epsilon_{10} \sqrt{(-10B_i)(-10B_s)} \]  

(3.6)

Equation 3.5 states that a violation occurs when the 10-minute average of ACE is greater than a specified limit of \( L_{10} \). In order to satisfy the CPS2 requirement each balancing authority must operate so that its average ACE in at least 90% of its clock-ten-minute periods during the calendar month must not violate Eq. 3.5.
The analysis outlined in this work utilizes two distinct models for calculating the amount of regulation needed within a system in order to adhere to the NERC standards. Subsection 3.2.1 describes the methodology of an optimal dispatch model, while Subsection 3.2.2 discusses the methodology for a linearized dynamic model.

### 3.2.1 Optimal dispatch model

The objective of the proposed optimization method, for our optimal dispatch (OD) model, is to determine the minimum quantity of regulation required to keep ACE (Eq. 3.7) in compliance with NERC standards CPS1 and CPS2, given increasing quantities of wind generation within a balancing area.

\[
ACE(t) = E(t) + P_R(t) \tag{3.7}
\]

\[
E(t) = \sum_{i=1}^{n_g} P_{Gi}(t) + W(t) - P_D(t) \tag{3.8}
\]

In Eq. 3.7 the variable \( E(t), \) which is defined in Eq. 3.8, represents the total net power generation for each 4-second time period, where \( P_{Gi}(t) \) represents the power of generation for each of the generators, \( W(t) \) represents the total wind power production and \( P_D(t) \) represents the total demand seen within the system at time \( t. \) While the variable \( P_R(t) \) represents the regulation power produced during each 4-second time period. The design of our proposed method is outlined in Figure 3.1 where it is shown that our model involves solving a two step dispatch problem, the first being at a 5-minute time scale and the second being at a 4-second time scale. Our 5-minute dispatch model takes in plant data, 5-minute averages of wind speed and 5-minute load data in order to calculate the quantity of power produced by each generator for each 5-minute time period. The next step in our model takes in the net power observed in the system, as a function of the power generated by the generators, the wind within the system and the total demand in order to determine the amount of regulation needed at each time step in order to keep ACE in compliance with CPS1 and CPS2 and is represented by our 4-second model. The 4-second model will then determine the required quantity of regulating reserves needed for each hour in order to meet the NERC standards.
However, prior to determining the amount of regulation that is needed at each 4-second time period we must determine the dispatch schedule of the generators in the balancing authority, which is (in our model) adjusted every 5-minutes. Our model is designed to optimally dispatch generation.
in order to minimize the total cost. In this model we will be assuming that the regulation costs are constant for both up and down regulation, meaning that minimizing the cost of the regulation will not be needed during the computation of our 4-second model. The methodology of the 5-minute dispatch model is discussed in Subsection 3.2.1.1 while the methodology for the 4-second dispatch model is contained in Subsection 3.2.1.2.

3.2.1.1 5-minute dispatch model

The objective for the 5-minute dispatch model (Eq. 3.9) is to minimize the total cost of generation for each of the 5-minute time steps for of all the generators. Soft constraints for surplus and deficit power imbalances were incorporated into this objective statement in order to ensure a close balance between dispatched generation and demand.

\[
\min_{\sum_{k=1}^{n} \sum_{i=1}^{n_g} C_{G_i}(t_k)} \sum_{k=1}^{n} \sum_{i=1}^{n_g} C_{G_i} P_{G_i}(t_k) + C_S S_+(t_k) - C_S S_-(t_k)
\]

\[\text{s.t. } RR_D \Delta t \leq P_{G_i}(t_k) - P_{G_i}(t_{k-1}) \leq RR_U \Delta t, \forall i, k\]

\[\sum_{i=1}^{n_g} P_{G_i}(t_k) + W(t_k) - P_D(t_k) + S_+(t_k) + S_-(t_k) = 0, \forall k\]

\[P_{G_i} \leq P_{G_i}(t_k) \leq \bar{P}_{G_i}, \forall i, k\]

\[S_+(t_k) > 0, \forall k\]

\[S_-(t_k) < 0, \forall k\]

Equation 3.9 $C_{G_i}$ represents the cost of the generation ($$/MWh) of each of the power plants in the system and $P_{G_i}(t_k)$ refers to the total power produced by each of the generator for each 5-minute time period. The size of the surplus and deficit soft constraints of the objective statement, represented by $S_+(t_k)$ and $S_-(t_k)$ respectively, are constrained by $C_S$, which represents the cost of the imbalance and is set to a value much greater than the actual generator cost to ensure that a close balance between supply and demand is found.

The objective statement for this model is subject to the constraints in Eqs. (3.10)-(3.14). Equation 3.10 constraints the ramping of the generators by requiring that the difference between the current generation and the generation at the next time step is not more than the 5-minute ramp rate capacity of each generator. Where $RR_D$ and $RR_U$ represent each generators ramp rates (MW/hr) and $\Delta t$ represents the 5-minute step change. As shown in Eq. 3.11 this optimization method operates with the constraint that the total generation in the system must equal the total demand in the
system at each of the time intervals. Due to the variability of the wind in the system, this constraint would be occasionally impossible to match without the soft constraints shown in Eqs. 3.13 and 3.14, which refer to the surplus and deficit power imbalance. The last constraint (Eq. 3.12) states that the generation of each generator must lie between the generator’s maximum and minimum output values at all times.

This formulation results in an optimal generation dispatch, disregarding the losses in transmission, in order to match the demand in the system given increasing wind penetration levels up to 25 percent of the demand. The resulting generation totals from each of the generators will then be interpolated down to a 4-second time step, which will then be used in the second part of our model (Subsection 3.2.1.2) to determine the amount of regulation needed at each 4-second time interval in order to keep the system in compliance with NERC standards.

### 3.2.1.2 4-second model

The objective of the 4-second model (Eq. 3.15), is to minimize the quantity of regulation needed at each 4-second time step. The total regulation, \( R \), is determined on a hourly basis and is the total amount of regulation needed to satisfy the NERC reliability requirements. This regulation variable applies to both up and down regulation, as shown in Eq. 3.16, which states that up and down regulation are treated as the same value for each hour. For a system in which up and down regulation prices differ, this would have to be taken into account, however for our purposes this constraint holds true.

\[
\begin{align*}
\min_{R, P_R} & \quad R \\
\text{s.t.} & \quad -R \leq P_R(t_k) \leq R, \forall k \\
& \quad -RK_R \leq P_R(t_{k+1}) - P_R(t_k) \leq RK_R, \forall \kappa \\
& \quad -\epsilon_1 \leq \frac{1}{15} \sum_{k=\kappa}^{\kappa+15} \frac{E(t_k) + P_R(t_k)}{-10B} \leq \epsilon_1, \forall \kappa \in \{1, 15, \ldots\} \\
& \quad -L_{10} \leq \frac{1}{150} \sum_{k=\kappa}^{\kappa+150} E(t_k) + P_R(t_k) \leq L_{10}, \forall \kappa \in \{1, 76, 151, \ldots\}
\end{align*}
\]  

(3.15) \hspace{2cm} (3.16) \hspace{2cm} (3.17) \hspace{2cm} (3.18) \hspace{2cm} (3.19)

This objective is subject to the constraints shown in Eqs. (3.16)-(3.19). Equation 3.16, ensures that the regulation at each 4-second time interval can be any value, positive or negative, between the total regulation purchased \( (P_R(t)) \) for that hour. The ramp rate constraint (Eq. 3.17) ensures that the regulation at the following 4-second time step is within limits. This amount in our model is set
as the fastest ramping limit at which the ISO will pay its generators to ramp, which is represented by
the constant $K_R$ and is the same for both up and down regulation. This model essentially assumes
that all regulation is provided by a single generator, resulting in a uniform regulation price and as
a result the total ramp rate of the system for each 4-second period is defined as the total ramp
rate of all the generators over a given period. This assumption is not valid for systems with highly
constrained systems, but allows one to gain initial intuition about the affect of wind on regulation.

The last two constraints in this model enforce the NERC standards CPS1 and CPS2. The
constraint for CPS1 (Eq. 3.18) requires that the average ACE for each 1-minute time period,
divided by the balancing areas frequency bias (B) must be between the positive and negative $\epsilon_1$.
This linear approximation for CPS1 is valid because our model is only dependent on the internal
frequency and isn’t influenced by external frequencies. The final constraint in the 4-second model
(Eq. 3.19) enforces CPS2, and states that the 10-minute averages of ACE in the system must be
between the positive and negative value of $L_{10}$. The numerical values for the constants used to
evaluate the NERC constraints seen in Eqs. 3.18 and 3.19 were obtained from (NERC, 2010) for
New England in 2010. This document is updated yearly, but its values tend to be constant over
time.

This 4-second model estimates in the total regulation required for each hour, as well as the
power output from plants providing regulation at each 4-second time step. By running these two
optimization processes one is able to observe the effect of increasing wind penetration levels on the
regulation needed in a system to satisfy NERC reliability requirements.

3.2.2 Linearized dynamic model

The goal of this work is to calculate the amount of regulating reserves needed to balance the quantity
of demand with the quantity of the generation, given increasing levels of wind penetration. In this
section we use a linearized dynamic (LD) power system, coupled with optimal economic dispatch
to accurately determine the quantity of regulation needed in order to keep the power system within
the NERC standards. Figure 3.2 depicts the overall design of our LD model and shows that the first
step of this process is to optimally dispatch generation units and solve for the steady-steady solution
to the problem formulation. After the steady-state solution has been determined our model iterates
though a day’s worth of data solving the mathematical model and re-dispatching the generation
units to meet the demand within the system for each 5-minute time period until the completion of
the process. The remainder of this section describes the feedback control system used in our analysis, as well as the mathematical models for both the economic 5-minute dispatch and the linear-time invariant model.

3.2.2.1 Feedback control system

For the model developed in this section we have designed a two-area feedback control system taking into account generator dynamics, generation limits, and power flow in order to determine the quantity of regulation needed as wind penetrations increase. Unlike the the OD model this LD model will account for power flow between areas, with the assumption that there is no scheduled power flow between the areas for any time period. This method avoids using linear assumptions.

In order to calculate the regulation needed within a system with increasing penetrations of wind energy this work utilizes the IEEE 39 bus system (Pai, 1989), seen in Figure 3.3, as our benchmark test case with a few alterations. The original IEEE 39 bus system is comprised of 10 generators, however for the purposes of this study four additional generators were added to each area for the purposes of regulation dispatch, seen in Figure 3.3 as $G_{11}$, $G_{12}$, $G_{13}$ and $G_{14}$. For the analysis presented in this work two of these systems have been connected together in parallel to produce a single interconnection containing two separate balancing areas. In this design these two areas are connected by two tie lines located at buses 23 and 29 in the first area, which are then connected to

Figure 3.2: LD Model Design Flowchart
buses 2 and 9 in the second area. Producing this two areas system allows for the calculation and control of the quantity of ACE within each of these areas by controlling the amount of power being transmitted through the tie lines. For the purposes of this study each area is comprised of 39 buses, 14 generators, 19 loads and 46 branches. The generator parameters used in this study are shown in Subsections 3.3.1 and 3.3.3, while all other parameters were taken from Pai (1989).

Figure 3.3: IEEE Case 39 Bus New England System

For this analysis wind farms are located on buses that have neither a generator nor a load currently located on them. This approach was chosen in order to both simplify the model to an extent and due to current practices when building wind farms. Due to their size and appearance it is currently infeasible to build a large wind farm near a residential consumer population due to zoning restrictions and fierce opposition by some residents, so taking this into account we have decided to island the wind farms on buses with no direct load. We used four wind farms in each area placed on different buses to calculate the amount of regulation needed to solve the system starting with zero percent wind penetration and increasing that value up to 25 percent.

The two-area feedback control system that was implemented for this analysis is shown in Figure
3.4, which was derived from the models seen in (Bergen and Vittal, 2000; Shoureshi et al., 2000), where the generator blocks are defined by the following first order equations.

\[
G_M(s) = \frac{1}{T_g s} \quad (3.20)
\]

\[
G_p(s) = \frac{1}{M + D} \quad (3.21)
\]

Equation 3.20 represents the control block for the governor for each of the machines, where \( T_g \) is the generators time constant for each of the machines. Whereas Eq. 3.21 represents the swing equation of the machine rotor for each of the generators within the system. In this model we have neglected power flow limits on the transmission lines, therefore in this model \( M \) represents the rotating inertia of each machine, and \( D \) is the mechanical resistance of the machines, with the values for \( T_g, M, \) and \( D \) shown in Table 3.1.

![Two Area Feedback Control Model](image)

**Figure 3.4:** Two Area Feedback Control Model

In the feedback model shown in Figure 3.4, \( B \) represents the frequency bias setting for each of the areas, \( R \) represents the droop control within the system, the DCPF Grid is defined as direct current power flow grid in order to incorporate power flow into the model, and lastly \( K \) represents a positive scaling factor to determine \( \Delta P_c \) from the changing ACE value. In this feedback model
the total power that can be produced by any generator is rail limited by each generator's minimum and maximum output values in each area. Generator output is also rate limited by the mechanical constraints from each of the generators. We also have limited the total quantity of regulation that can be provided to the system as a percentage of peak demand.

### 3.2.2.2 LD mathematical model

The linear-time invariant model used in this research simultaneously solves an algebraic power flow coupled with nonlinear differential equations to solve for the state variables in the system. The differential equations for the mathematical model for the feedback control loop are shown below in the Eqs. (3.22)-(3.25),

\[
\frac{d\delta}{dt} = \omega_0 \Delta \omega \tag{3.22}
\]

\[
\frac{d\omega}{dt} = \frac{1}{M} P_m - \frac{1}{M} P_g - \frac{D}{M} \Delta \omega \tag{3.23}
\]

\[
\frac{dP_m}{dt} = \frac{1}{T_g} P_{ref} + \frac{1}{T_g} \Delta P_c - \frac{1}{T_g} P_m - \frac{1}{RT_g} \Delta \omega \tag{3.24}
\]

\[
\frac{dP_c}{dt} = -KACE \tag{3.25}
\]

where ACE can be calculated by Eq. 3.26, since we assume that there is no scheduled interchange between the two areas.

\[
ACE = P_{tie} + B \Delta \omega \tag{3.26}
\]

In this model ACE is the sum of the power transmitted through the tie lines and the mean averaged weighted frequency deviation multiplied by the areas frequency bias.

The algebraic equation for the power network model is shown in Eq. 3.27,

\[
P = \mathbf{B} \theta - P_G + P_D \tag{3.27}
\]

where the product of the nodal admittance matrix (\(B\), note the difference between the variable for frequency bias) and the voltage angles of the bus (\(\theta\)) represents the DC power flow within the area. The total power injected into each node is the difference between the total power generated (\(P_G\)) and the sum of the power flow and demand (\(P_D\)) in the system. With the system designed to drive the net power within the system (\(P\)) to zero within each area.
3.2.2.3 Economic dispatch

In order to accomplish this objective the following economic dispatch model has been developed, as shown in Eq. 3.28, where the objective is to minimize the total cost of the system with respect to power produced by each of the generators, in terms of dispatch and regulation. This objective statement is the same as the one used for the OD model in Subsection 3.2.1.1 with the addition of a regulation constraint \((C_{R_i} R_i)\).

\[
\begin{align*}
\text{min} & \quad \sum_{i=1}^{n_g} C_{G_i} P_{G_i} + C_{R_i} R_i + C_S S_+ - C_S S_- \\
\text{s.t.} & \quad R_{R_D} \Delta t \leq P_{G_i}(t_k) - P_{G_i}(t_{k-1}) \leq R_{R_U} \Delta t, \forall i, k \quad (3.29) \\
& \quad \sum_{i=1}^{n_g} P_{G_i}(t_k) + W(t_k) - P_D(t_k) + S_+(t_k) + S_-(t_k) = 0, \forall k \quad (3.30) \\
& \quad P_{G_i} \leq P_{G_i}(t_k) \leq P_{GiT}, \forall i, k \quad (3.31) \\
& \quad S_+(t_k) > 0, \forall k \quad (3.32) \\
& \quad S_-(t_k) < 0, \forall k \quad (3.33) \\
& \quad P_{G_i} + R_i \leq P_{G_i}(t_k) \leq P_{GiT} - R_i, \forall i, k \quad (3.34) \\
& \quad \sum_{i=1}^{n_g} R_i = R_T, \forall i \quad (3.35) \\
& \quad 0 \leq R_i \leq R_{R_{reg}}, \forall i \quad (3.36)
\end{align*}
\]

The constraint Eqs. (3.29)-(3.33) used in this economic dispatch are the same as those used in Subsection 3.2.1.1 for the 5-minute dispatch formulation of the OD model. In addition to these constraints this model uses regulation constraints in Eqs. (3.34)-(3.36). Equation 3.34 states that the generation of each of the generators must lie between the generator’s minimum and maximum values, plus or minus respectively, the quantity of regulating reserves that the generator is providing at all times. Equation 3.35 constrains the total regulation within the system and states that the sum of the regulation provided by each of the generators must be equal to the total regulation provided by the system. While Eq. 3.36 governs the amount of regulation that can be provided by each of the generators and states that the regulation provided by each generator \((R_i)\) must lie between zero and its total regulation ramp rate \((R_{R_{reg}})\) over the five minute period in respect to each generator, which for this model is the same quantity as the generator’s ramp rates.

This economic dispatch allows for an optimal approach of dispatching generators as well as reserves in a single model for each 5-minute period. These results are then fed into the LD model,
as $P_{ref}$ and the regulation limits of the generators, which give dynamic results and is repeated for each 5-minute time period.

### 3.3 Data

#### 3.3.1 Feedback control data

Table 3.1 contains all of the parameters show in Figure 3.4, as well as Eqs. 3.20 and 3.21. The frequency bias ($B$) for each area was set as 1% of the peak demand of the system using the guidelines outlined by the NERC Resources Subcommittee (NERC, 2011). Unlike frequency bias the values for $K$ and $R$ are not defined by the NERC reliability standards, so these values were taken from Shoureshi et al. (2000). The values for rotating inertia ($M$) where taken from Pai (1989), for generators $G_1 - G_{10}$, and the values for the mechanical resistance of the machines ($D$) was set to half that of its rotating inertia. The generator time constants ($T_g$) were set to uniform values of 1, for generators $G_1 - G_{10}$, with some variation based off of the size of the generator. The values for the generators $G_{11} - G_{14}$ were based off their size relative to the original generation units, and fact that these are meant for fast ramping regulation services.

**Table 3.1:** Parameters used in Two Area Feedback Control Model

<table>
<thead>
<tr>
<th>Area Parameters</th>
<th>Area 1</th>
<th>Area 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>$B$</td>
<td>522 MW/Hz</td>
<td>522 MW/Hz</td>
</tr>
<tr>
<td>$K$</td>
<td>0.015 sec</td>
<td>0.015 sec</td>
</tr>
</tbody>
</table>

| Machine Parameters for Area 1 (Same for Area 2) in p.u. values |
|------------------|------------------|------------------|------------------|------------------|------------------|------------------|------------------|------------------|------------------|------------------|------------------|
|                  | $G_1$ | $G_2$ | $G_3$ | $G_4$ | $G_5$ | $G_6$ | $G_7$ | $G_8$ | $G_9$ | $G_{10}$ | $G_{11}$ | $G_{12}$ | $G_{13}$ | $G_{14}$ |
| $M$              | 1000  | 60.6  | 71.6  | 57.2  | 52.0  | 69.6  | 52.8  | 48.6  | 69.0  | 84.0  | 30.0  | 25.0  | 30.0  | 25.0  |
| $D$              | 500   | 30.3  | 35.8  | 28.6  | 26.0  | 34.8  | 26.4  | 24.3  | 34.5  | 42.0  | 15.0  | 12.5  | 15.0  | 12.5  |
| $T_g$            | 1.2   | 1     | 1     | 1     | 1     | 1     | 1     | 1.1   | 0.8   | 0.6   | 0.5   | 0.6   | 0.5   | 0.5   |
| $R$              | 0.05  | 0.05  | 0.05  | 0.05  | 0.05  | 0.05  | 0.05  | 0.05  | 0.05  | 0.05  | 0.05  | 0.05  | 0.05  | 0.05  |
3.3.2 Load and wind data

For both models used in this work, ISO New England and the IEEE 39 bus system, there does not exist substantial wind plants installed, which could be included in our model. In the case of the ISO New England system there are wind farms located within the actual balancing area, however the output data from those farms are not publicly available. The IEEE 39 bus system does not include wind farms, as is the case with most test cases. Therefore in order to introduce wind farms into the model, the data that was used in Chapter 2 from Plant B was used, which contains 60 days of 2-second resolution wind data from a wind farm in central United States.

For the OD model discussed in Subsection 3.2.1 the 60 days of wind data were broken into 10-day segments allowing for each of the segments to be used as a single wind farm, with power from wind being treated as a negative load. These segments were then used to create six different wind profiles, which were then scaled to wind penetration levels of up to 25 percent in order to observe the effects on regulation, given the optimal dispatch of generation within the system. The six wind profiles were created by taking the first wind farm and combining sequential wind farms in numerical order, as seen in Table 3.2, to create the six different wind profiles. Table 3.2 lists the correlation factors between the individual 10-day segments, and shows that there is little correlation between the segments. The one exception to this is the correlation between the segments that represent Farm 4 and Farm 5, which is due to the curtailment of the wind data within the dataset. However, due to the method used in the creation of the wind farms the law of large numbers will produce a wind dataset that is largely uncorrelated. However, if a randomized approach were used to create the different wind profiles this may not be the case given the data used in this study.

<table>
<thead>
<tr>
<th>Correlation between the six wind farms used in the OD model</th>
</tr>
</thead>
<tbody>
<tr>
<td>Farm 1</td>
</tr>
<tr>
<td>Farm 1</td>
</tr>
<tr>
<td>Farm 2</td>
</tr>
<tr>
<td>Farm 3</td>
</tr>
<tr>
<td>Farm 4</td>
</tr>
<tr>
<td>Farm 5</td>
</tr>
<tr>
<td>Farm 6</td>
</tr>
</tbody>
</table>

For the LD model discussed in Subsection 3.2.2 the 4 wind farms in each of the balancing areas
are treated independently, meaning that their outputs are not connected to the outputs of all other wind farms. To accomplish this task, a randomly selected day of wind power data was taken from the 60 day dataset for each of the wind farms within the two area system to account for the total power being provided by each wind farm over the course of each day. These randomly selected days were then scaled in order to reach the desired wind penetration level and processed through interpolation for the given period for usage in the LD model, with the model treating wind as a negative load. Table 3.2 also shows that much like the results for the OD model we can assume that the wind data for each random day will be highly uncorrelated based on the assumption that not all of the data will come from the 20 days worth of data that comprises Farm 4 and Farm 5.

The load profile of the net demand was based off from historical 5-minute load data from ISO New England (ISONE, 2012). For the OD model discussed in Subsection 3.2.1, the load data were already of the appropriate size, due to the usage of the ISO New England system as our power system. For the LD model the chosen load profile was first normalized, at which point a mean reverting random walk was applied to the profile to produce diversity in the data. From here the given load profile was scaled to an appropriate size for each load within the balancing area to ensure that the total demand seen within the system was less than the generation capacity. The total wind profile, for both models, was determined based off the percentage of energy demand within the system.

3.3.3 Generation capacity and fuel costs

The OD model discussed in Subsection 3.2.1 uses the ISO New England balancing authority as the test bench. The ISO New England model includes 86 generators, which are spread out over all of New England. Table 3.3 shows the current installed generator capacities in the ISO New England system. Almost half of the installed capacity in ISO New England is natural gas, while the second highest capacity of installed plants are oil-based, which are the most expensive plants to run.

<table>
<thead>
<tr>
<th>Plant type</th>
<th>Biomass</th>
<th>Coal</th>
<th>Nuclear</th>
<th>Natural Gas</th>
<th>Oil</th>
</tr>
</thead>
<tbody>
<tr>
<td>Capacity (MW)</td>
<td>531.1</td>
<td>4316.7</td>
<td>4074.9</td>
<td>15316.2</td>
<td>6349.7</td>
</tr>
</tbody>
</table>

The LD model uses the IEEE 39 bus system with a few alterations. The first alteration to this test case was the addition of four additional generating units in order to match regulation requirements.
to maintain grid reliability. Table 3.4 shows the generation capacities, along with the assigned plant type and corresponding fuel costs for the altered IEEE 39 bus system. Due to the objective of our LD model (Eq. 3.28), which simultaneously dispatches both generation and reserve resources the total amount of regulation that can be provided is limited by the size of the demand. Since the amount of regulation dispatched is a percentage of the peak demand, the maximum regulation that can be provided is equal to the difference between the peak demand and sum of total generation possible. In order to increase the amount of generation available for regulation four additional generators (short cycle gas and oil plants) at a base of 100 MW in size were added to the system.

<table>
<thead>
<tr>
<th>Generator</th>
<th>Max Capacity (MW)</th>
<th>Plant Type</th>
<th>Fuel Cost ($/MWh)</th>
<th>Ramp Rate (MW/min)</th>
</tr>
</thead>
<tbody>
<tr>
<td>G1</td>
<td>1100</td>
<td>Coal</td>
<td>16.212</td>
<td>5.50</td>
</tr>
<tr>
<td>G2</td>
<td>1145.55</td>
<td>Nuclear</td>
<td>1.706</td>
<td>1.91</td>
</tr>
<tr>
<td>G3</td>
<td>750</td>
<td>Coal</td>
<td>16.519</td>
<td>3.75</td>
</tr>
<tr>
<td>G4</td>
<td>732</td>
<td>CC Gas</td>
<td>25.598</td>
<td>12.20</td>
</tr>
<tr>
<td>G5</td>
<td>608</td>
<td>CC Gas</td>
<td>27.987</td>
<td>10.13</td>
</tr>
<tr>
<td>G6</td>
<td>750</td>
<td>Coal</td>
<td>16.894</td>
<td>3.75</td>
</tr>
<tr>
<td>G7</td>
<td>660</td>
<td>Coal</td>
<td>16.041</td>
<td>3.30</td>
</tr>
<tr>
<td>G8</td>
<td>640</td>
<td>CC Gas</td>
<td>23.038</td>
<td>10.67</td>
</tr>
<tr>
<td>G9</td>
<td>930</td>
<td>CC Gas</td>
<td>23.891</td>
<td>15.50</td>
</tr>
<tr>
<td>G10</td>
<td>350</td>
<td>CC Gas</td>
<td>30.205</td>
<td>5.83</td>
</tr>
<tr>
<td>G11</td>
<td>100</td>
<td>SC Gas</td>
<td>34.813</td>
<td>5.00</td>
</tr>
<tr>
<td>G12</td>
<td>100</td>
<td>Oil</td>
<td>43.004</td>
<td>5.00</td>
</tr>
<tr>
<td>G13</td>
<td>100</td>
<td>SC Gas</td>
<td>37.884</td>
<td>5.00</td>
</tr>
<tr>
<td>G14</td>
<td>100</td>
<td>Oil</td>
<td>52.560</td>
<td>5.00</td>
</tr>
</tbody>
</table>

The size of these four plants are altered according to the total amount of regulation that is needed in order to satisfy reliability standards. This is accomplished by treating the additional generation units as $N \times 100$ MW units, where $N$ is scaling factor of the number of 100 MW sized plants that would be needed to match regulation requirements. For the analysis presented in this work we have set regulation to 15% of the peak demand. This requires a $N$ equal to 3, which correlates to generators $G_{11}, G_{12}, G_{13},$ and $G_{14}$ each being represented by three distinct plants, each with a
generation capacity of 100 MW. Though more expensive at high regulation levels this approach provides fast ramping plants, which are useful when dealing with the variability seen in wind data.

Table 3.5: Fuel Prices used in OD Model, based off real price data from ISO New England in 2010

<table>
<thead>
<tr>
<th>Fuel Type</th>
<th>Fuel Prices ($/MWh)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bituminous Coal</td>
<td>17.539</td>
</tr>
<tr>
<td>Diesel Oil</td>
<td>57.069</td>
</tr>
<tr>
<td>Jet Fuel</td>
<td>58.686</td>
</tr>
<tr>
<td>Kerosene</td>
<td>58.686</td>
</tr>
<tr>
<td>Natural Gas</td>
<td>17.570</td>
</tr>
<tr>
<td>Nuclear</td>
<td>1.672</td>
</tr>
<tr>
<td>Residual Oil</td>
<td>39.052</td>
</tr>
<tr>
<td>Sub-bituminous Coal</td>
<td>4.662</td>
</tr>
<tr>
<td>Biomass</td>
<td>12.594</td>
</tr>
</tbody>
</table>

Table 3.6: Ramp rates constraints of generators used in OD Model

<table>
<thead>
<tr>
<th>Fuel Type</th>
<th>Ramp Rate (% of Max Generation)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nuclear</td>
<td>10% per hour</td>
</tr>
<tr>
<td>Coal</td>
<td>30% per hour</td>
</tr>
<tr>
<td>CC Gas</td>
<td>100% per hour</td>
</tr>
<tr>
<td>SC Gas</td>
<td>100% per 20 minutes</td>
</tr>
<tr>
<td>Oil</td>
<td>100% per 20 minutes</td>
</tr>
</tbody>
</table>

Due to the optimization nature of our models the cost of producing energy from each type of plant is extremely important. Table 3.4 shows the fuel prices used the LD model, whereas Table 3.5 details the pricing used in the OD model. Due to the objective statements (Eqs. 3.9 and 3.28) of our models the cheapest generation units will be dispatched first, with more expensive plants coming online as demand increases. For these models the cost of wind generation has been set to zero. There are operation and management costs associated with wind, but for this work we have assumed them to be negligible.

Regardless of the sizes of the generators used in the models the key aspect of each model that
must be taken into account is the quantity that each generator can ramp, up or down, during each dispatch period. For both models generators were dispatched at 5-minute intervals, which is slightly faster than current practice in most balancing areas. Most balancing areas re-dispatch generation every 10-minutes to 1-hour, however dispatching on a 5-minute interval will yield more control over the power balancing within the system allowing for a more focused analysis on regulating reserves. Tables 3.4 and 3.6 show the ramp-rate assumptions of the generation capacities used in this work. From these tables it is clear that nuclear plants are used as a base-load due to ramp rate limits, while coal and combined cycle (CC) gas will be used in load following and providing regulation, along with the short cycle (SC) gas and oil plants.

3.4 Results

The results from this study aim to determine the quantity of regulating reserves that will be required in a system with up to 25 percent wind penetration levels. The results from the two models developed in this work will be discussed independently and compared to validate the results from each.

3.4.1 Optimal dispatch model results

Figure 3.5 shows the generator dispatch results using a six wind farm design and 10% wind penetration from solving for the objective statement shown in Eq. 3.9 for the OD model. The generation totals, as expected by the fuel cost data in Table 3.5, show that the nuclear plants remain at a constant power output over the course of the study, while the coal and natural gas plants ramp up and down over the course of each day.
The results from Figure 3.5 show the overall behavior of the dispatch of the generators for each 5-minute time period, but it shows no relative information on the regulating reserve requirements within the system. Figure 3.6 represent the results yielded from the 4-second optimization model (Eq 3.15). This figure depicts the optimal quantity of regulating reserves necessary, as a percentage of peak demand, in order to solve the linear approximations of the NERC standards seen in Eqs. 3.18 and 3.19, for six different wind profiles.

Figure 3.5: Optimal generator dispatch resulting from the OD Model for the ISO New England system with 10% wind penetration
The results shown in Figure 3.6 provide substantial insight into the impact of large-scale wind, perhaps most notably is the impact of geographical diversity on regulating reserves. The results in this figure clearly depict that increased geographical diversity drastically reduces the quantity of regulation needed to satisfy the NERC standards. In contrast, if there is little-to-no diversity, as is the case when only one wind farm exists in the system, the regulation requirements increase dramatically with wind penetration. This pressing issue is tabulated in Table 3.7, which shows that with a wind penetration equal to 25% for a single wind farm, the amount of regulation needed is greater than the quantity of the wind. However, Table 3.7 shows that the addition of geographical diversity within the system through the addition of more wind farms indicates a steady decrease in the regulation that is needed for the same wind penetration. The regulation requirement decreases to approximately 4% of the peak demand, for a system with six wind farms.
Table 3.7: Regulation required using OD model for 25% wind penetration

<table>
<thead>
<tr>
<th>Number of Wind Farms</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regulation Required (% Peak of Demand)</td>
<td>26.69</td>
<td>16.9</td>
<td>10.01</td>
<td>6.48</td>
<td>4.95</td>
<td>4.004</td>
</tr>
</tbody>
</table>

It is clear from these results that geographic diversity is a major contributing factor to the amount of regulation that would be needed as wind penetrations increase. This is due to the fact that as diversity is added to the system, smoothing across the data takes effect. Given the results obtained through this OD model a system will require significant levels of geographical diversity, at least 4 wind farms in this model, in order to maintain a reasonable quantity of regulating reserves. However, before making any conclusions about the results obtained from this optimization model it must be compared to the more detailed LD model.

3.4.1.1 SPP Regulation Comparison

The results obtained from the OD model were then compared to the regulation requirements proposed by Tsuchida et al. (2010) shown in Eqs. 3.37 and 3.38,

\[
R_{up} = \sqrt{(0.01l_{\text{peak}} + L_{10})^2 + a\Delta W_{95}^2 - L_{10}} \tag{3.37}
\]

\[
R_{down} = \sqrt{(0.01l_{\text{peak}} + L_{10})^2 + a\Delta W_{5}^2 - L_{10}} \tag{3.38}
\]

where \(R_{up}\) and \(R_{down}\) represent the up and down quantity of regulation needed, \(l_{\text{peak}}\) represent the peak of the demand, \(a\) is assumed to be equal to 2, and \(\Delta W_5\) and \(\Delta W_{95}\) are the 95\(^{th}\) and 5\(^{th}\) percentiles of the wind power increments.
Figure 3.7: Required up regulation using SPP estimation method (Eq. 3.37) for increasing wind penetrations up to 25%.

Figure 3.8: Required down regulation using SPP estimation method (Eq. 3.38) for increasing wind penetrations up to 25%.
Figures 3.7 and 3.8 show the corresponding quantity of up and down regulation needed based off the SPP estimation method. Similar to the results obtained with our OD model increased geographical diversity lowers the total quantity of regulating reserves, however the values calculated by the SPP estimation method are significantly lower than those calculated by our OD model. This estimation method yields results with a maximum regulation percentage of less than 3.5%, compared to the over 26% regulation requirement yielded by our OD model for a single wind farm at 25% wind penetration. Thus the estimation method proposed by Tsuchida et al. (2010) states that the wind used in this study should have little to no effect on regulating reserves, even at high wind penetration levels. This estimation method also states that geographical diversity has minimal impact on the quantity of regulating reserves required within the system, as the minimum quantity of regulation needed is less than 2% away from the maximum value required. The results from this estimation method will be compared to our LD model in order to determine its effectiveness at predicting the quantity of regulating reserves required as wind penetration levels increase within a system.

3.4.2 Linearized dynamic model results

In order to validate the results obtained from the OD model, the results are compared to a dynamic model that incorporates power flow between balancing areas. This is where the development of the linearized dynamic regulation model from Subsection 3.2.2 became essential. The results obtained from this model allow for the validation of results from the optimization model covered previously. Unlike the previous model the results from the LD model cannot be approximated using linear assumptions, therefore the model uses power imbalances and frequency to calculate the CPS1 and CPS2 scores in Eqs. (3.1)-(3.5).

In practice CPS1 and CPS2 are calculated using information spanning 1-year of data for CPS1 and monthly data for CPS2, and there is no historical data for the IEEE 39 bus system. This also means that there is not a predefined value for $\epsilon_1$ or $L_{10}$ seen in Eqs. 3.2 and 3.5. In order to determine values for $\epsilon_1$ and $L_{10}$ we used standards from the NERC resources subcommittee (NERC, 2011), which provides instructions to determine these values. Due to the size of the balancing areas used this study, $L_{10}$ was set at 50 MW, while $\epsilon_1$ was determined by calculating the value that gave CPS1 equal to 160% for a base case load with zero wind penetration and 1% regulation. Due to the lack of historical data for this model, the experiment was set so that instead of averaging CPS1 and CPS2 scores over the standard time frame, these values were calculated on a rolling average over
the time frame of the model simulation. Though this process is less than ideal, it is the only way to account for the changes seen by the newly introduced wind farms without averaging out the effects.

The model was then run over the course of a single day for given regulation and wind penetration levels and at the completion of each day's simulations the CPS1 and CPS2 scores were calculated. This process was completed for wind penetration levels ranging from 0 to 25% penetration in 1% increments. Each wind penetration level was processed for regulation values ranging from 1 to 15% of the peak demand, also with 1% increments. In order to ensure accuracy across the simulation, the same random set of wind days was used for each wind penetration level. This process was then repeated for 20 different random sets of wind days in order to observe the effects of different wind patterns.

Because regulation increases flexibility, CPS1 and CPS2 scores should increase as the regulation within the system increases. Figure 3.9 illustrates this result for 5, 10, 15 and 20% wind penetration levels for a single random set of wind days, where the two curving lines represent the CPS1 scores...
Figure 3.10: Resulting CPS2 scores from the LD model for single random wind set. The solid line with △’s and the solid line with X’s represent the CPS2 scores of the two balancing areas in respect to the minimum CPS2 score allowed, represented by the solid line at 90%.

for the two balancing areas in respect to the black line, which is the minimum CPS1 score to satisfy NERC requirements. From this graph it is clear that the maximum value for CPS1 is 200% while there is no defined minimum CPS1 score, as the compliance factor seen in Eq. 3.2, can be much greater than 2 depending on the values of ACE and the interconnections frequency resulting in a large negative CPS1 score. As expected, the values for CPS1 increases with respect to regulation percentage for all wind penetration levels shown.

Figure 3.10 shows that similar results were obtained for CPS2, however, with this model even when CPS1 was in violation CPS2 tended to still be in compliance with the 90% bound limit. The results for CPS2 in Figure 3.10 show that the lowest CPS2 score produced by the model is approximately equal to 66%, which corresponds to a CPS1 score of roughly -2500% showing that for this model failure of CPS1 will in most cases not produce a failure in CPS2. This result is due
to the fact that the model is re-dispatched every 5-minutes instead of every 10-minutes to 1-hour. Because the system is re-dispatched within the 10-minute averaging period used when calculating the number of violations, the model is in a way setting the constraint on CPS2 to hold true for every 10-minute time period. Due to the dispatch procedure used in this work if CPS2 was to be outside of the permissible bounds it would be a clear sign of a major power flow imbalance within the system.

Figure 3.11 shows the resulting regulation percentages from the LD model for all of the random wind day sets. Due to the size of the load profile used and the generation capacity of the 39 bus system, regulation percentage exceeding 15% could not be accurately calculated, thus any value exceeding a regulation percentage of 15% was not taken into account when calculating the mean regulation across the 20 wind sets. This is the reason that the mean average of regulation does not

Figure 3.11: Quantity of regulation needed, as a percentage of peak demand resulting from the LD model. The dotted line represents the range between the 10th and 90th percentiles, while the X’s represent the minimum and maximum extremes produced by the model. The solid line with * markers represents the linear interpolation between the mean regulation of all wind sets for wind penetration levels ranging from 0 to 25%.
increase with each increase in the wind penetration level.

Despite the fact that the regulation required at 25% wind penetration is approximately equal to 7.7%, the minimum and maximum values required are 3% and greater than 15% respectively. Despite not knowing the exact maximum regulation required, Figure 3.11 does show that at one point for one of the random wind scenarios the system requires more regulation than wind. This instance occurs for a single wind set at a wind penetration level of 13%, which yields a regulation percentage of 15 percent. If it is assumed that this quantity of regulation were to double at 25% wind the system could easily require more than 30% regulation, which is infeasible with todays energy markets.

### 3.5 Discussion

The fascinating aspect of the LD model results is how well they correlate to those found from the OD model. The results obtained from the LD model show a range of regulation values from approximately 3 percent to values with potential upwards of 30 percent at 25 percent wind, while the result of the optimal dispatch model yields a span from approximately 4 percent to 27 percent regulation. In both cases the models produced scenarios that required more regulation than wind, and cases with regulation percentages less than 5 percent. Even more interesting, is the fact that the mean regulation obtained from the LD model at 25 percent wind of 7.7 percent per area is extremely close to the optimal regulation for a system containing four wind farms from the first model, which yielded a regulation of 6.5 percent. The results also show that for low wind penetration levels (below 5 percent) wind has little to no effect on the quantity of regulating reserves needed within a system. This knowledge is particularly useful for utilities looking to add small wind farms, as it shows that through proper dispatch of generation there should be little to no effect on the overall system, both in generation cost and reserves required.

Despite the fact that the first model developed in this work neglected transmission flow and power flow between external balancing areas the linear approximations yielded very similar results seen from a dynamic linearized dynamic model that incorporated power flow and frequency. The correlations seen from the results of these models show that for a system with no wind forecast errors can potentially handle large penetrations of wind power without causing grid failures due to large power or frequency imbalances with less than 10 percent regulation. The results from this study
also indicate that the opposite outcome is possible as wind penetrations are increased. Both models show cases where regulation requirements will actually exceed the percentage of wind within the system. If this were the case it would not be possible to replace fossil fuels with wind power after about 15 percent. If the regulation requirements are higher than the wind penetration levels, there is virtually no cost or environment benefit to producing large quantities of wind, as the regulating plants would have to produce more energy than if there was no wind in the system at all.

This makes the variability seen in wind data extremely important. Through better control of wind power outputs, accurate forecasting, energy storage and curtailment, utilities can mitigate the impacts of the variability within wind data. This work also shows that increasing geographic variability can have an enormous effect on the quantity of regulating reserves by significantly lowering the amount of regulation that must be purchased to keep a system within NERC operating standards. If approached correctly, utilities potentially increase wind levels to reach 20 percent wind scenarios without compromising grid reliability, avoiding potential catastrophic events such as a massive cascading black out similar to the Northeast blackout of 2003. Further work is needed in order to determine the cost of added transmission for increased wind production as well as determining the appropriate energy storage levels needed to increase wind power to an appropriate level for usage in the energy market across all time periods for high wind penetration levels.

3.6 Conclusions

This chapter has studied the effects of increasing wind penetration levels on the quantity of regulating reserves. The results in this chapter show that with no forecast error or wind power output control, there is a steady increase in the quantity of regulating reserves required in order to keep a given system within NERC operating standards. Acknowledging that high wind penetration levels are indeed feasible given the correct factors. A system with significant geographical diversity will require significantly less regulation than a system with less diversity for the same wind penetration level. As worldwide wind penetration levels increase, diversity of the wind power within systems will be key, leading to added focus on forecasting and further research and development of more economical energy storage technologies. Through further research of methods to more accurately control the power produced by wind farms, the implementation of wind penetration levels of 20 percent or more may be feasible.
Chapter 4

Conclusions

This thesis presents results from two research projects related to large-scale wind integration. The work in Chapter 2 studied data from mesoscale meteorological models for the purpose of wind integration studies and analysis. Chapter 3 presented an analysis of regulating reserves with increasing wind penetration levels using high resolution wind speed data.

4.1 Contributions

Results from Chapter 2 indicate that synthetic wind data from meteorological models, such as those used to produce EWITS and WWSIS, underestimate the variability of wind speed for frequencies above one cycle per 3-hours. Analysis of these models show that these studies begin to diverge from the Kolmogorov spectrum at about one cycle per 6-hours, with minimal substantial difference for frequencies below one cycle per 3-hours. It is at the higher frequencies that this divergence becomes significant, and results in notable differences between the variability in actual wind farms and that of simulated wind farms from the same relative geographical area.

As a result of this divergence from the expected spectrum data from studies such as these will be appropriate for some types of power analysis, but will not be applicable for others. Data from studies such as these can likely provide effective estimates for annual energy production of a potential wind farm, as well as useful estimates of day-to-day variable of wind speeds. This results in models that are applicable in estimating the impact of large-scale wind power production on unit commitment costs and related problems that span hours to days, with some confidence in the statistical outcomes.
However, for problems that require substantial information about the variability in wind farm production over sub-hourly time scales, mesoscale model data need to be used with caution. This is particularly true for power grid reliability problems such as estimating the quantity of load following resources required to satisfy reliability requirements, which requires accuracy at time scales ranging from 5-minutes to hourly. These data sets will also be inadequate for estimating the impacts on regulating reserves requirements, which requires accuracy over time scales of seconds to minutes. The underestimation of wind speed variability observed in these models will likely lead to underestimation in power and ramp rates required to maintain grid reliability with high wind penetrations.

Despite the inadequate variability in data produced from mesoscale meteorological models, it must be recognized that the EWITS and WWSIS studies currently include the best data sets publicly available to researchers. Due to the fact that public data makes it possible for detailed and reproducible analysis it is essential that high resolution empirical data be made available for research as wind development and penetration levels increase. When such data are released, it becomes possible to improve the understanding of challenges and opportunities for wind power integration.

The regulating reserve analysis from Chapter 3 used two different models to arrive at similar conclusions. Both models produced wind scenarios that would require more regulation than the wind energy being supplied to the system, while also producing results that show only a minimal increase from current industry standards for regulating reserves for substantial penetrations of wind energy. With only four wind farms within a balancing area at 25 percent wind penetration, the models indicate that the required regulation increase will remain less than 10 percent of the peak demand. The models also indicate that, for wind penetration levels of less than 5 percent, there is wind integration has a relatively small effect on the quantity of regulation needed within the system. These results are particularly important to utilities looking to add small wind farms to their system, as few or no alteration, may be needed to incorporate the additional resources.

The results for increasing regulation show that high-quality wind forecasts or simple control of generator dispatch system can potentially handle large quantities of wind penetration while still maintaining reliability standards using significantly less regulation than the actual percent of wind penetration. However, the models also displayed that there is a potential for scenarios in which more regulation is required than the penetration of wind, which would actually negate the environmental benefits of wind power and require more generation from fossil fuel plants than if there was no wind in the system. It is this conclusion that makes understanding and dealing with the variability seen
in wind data extremely important. If wind power outputs are controlled in a systematic way to smooth the effects of wind variability, utilities may be able to mitigate the impacts of highly volatile wind. Incorporation of large geographical diversity coupled with good forecasting and added storage could in-turn help to meet the goal of “20% wind by 2030” without compromising grid reliability.

4.2 Future Work

At the conclusion of this work there are several future research topics that will need further study to facilitate increases in wind penetration levels. This work indicates that mesoscale models do a reasonably good job at capturing wind dynamics for times scales slower than 3-6 hours, however, a combination of factors result in less accurate results for faster dynamics. This includes the fact that input data for mesoscale models may contain as few as 4 samples per day resulting in smoothing relative to empirical data. Also the size of mesoscale (2 km) grid spacing makes it difficult to accurately model atmospheric turbulence, leading to the suggestion that it might be more useful to couple the results from a mesoscale model with that of microscale models in order to produce accurate statistics over multiple time scales. This work has also shown that as wind penetration levels increase there will be a need to better understand the predicted output power of wind farms, which is directly correlated to the findings in this work on mesoscale models. More research is needed to better understand the control of wind farm power outputs, possibility through increased usage of energy storage, as wind penetration levels increase.
Bibliography


BPA (2010). Wind generation and total load in the bpa balancing authority.


Richard, M. G. (2010, December). We could have 10 mw wind turbines by 2011 and 15 mw turbines by 2020. treehugger.


Appendix A

A.1 Spectral Density Analysis

The goal of spectral density analysis is identify the relative magnitudes of frequency components in a time domain signal. To estimate the PSD the discrete Fourier transform of the time series measurements is computed, as shown in Eq. A.1.

\[ X_k = \sum_{n=0}^{N-1} x_n e^{-2\pi i \frac{k}{N} n}, \quad k = 0, 1, \ldots, N - 1 \]  

(A.1)

For a time-domain signal \( x(t) \), the periodogram estimate of the signals power at the frequency domain point \( k \) is given by Eq. A.2,

\[
\begin{align*}
P_0(t) &= x_0^2(t) \\
P_{f_k}(t) &= x_k^2(t) + x_{N-k}^2(t), \quad k = 1, 2, \ldots, \frac{N}{2} - 1 \\
P_{f_{\text{max}}}(t) &= x_{N/2}^2(t)
\end{align*}
\]  

(A.2)

where the frequency and the frequency domain point \( k \) is related by Eq. A.3,

\[ f_k = 2f_{\text{max}} \frac{k}{N}, \quad k = 0, 1, \cdots, \frac{N}{2} \]  

(A.3)

which represents the amount of power that would be emitted if \( x(t) \) were a voltage applied to a 1 \( \Omega \) resistor. The spectral density, \( y(f) \), over some frequency range \([f_1, f_2] \) of \( x(t) \) gives the amount of signal power within that frequency range. This can be used to estimate the variance of a signal within a given frequency range:

\[ \sigma^2[f_1, f_2] = 2 \int_{f_1}^{f_2} y(f) df \]  

(A.4)
Techniques for estimating the spectral density of a discrete time signal $x(t)$ can be broken down into two distinct groups: parametric and non-parametric methods. Parametric methods assume that the stochastic process being sampled is comprised of a certain structure, which can be used to define some parameters within the model, such as the auto-regressive or moving average (ARMA) model. Non-parametric methods on the other hand estimate the power spectrum without assuming that the sample is comprised of any particular structure. The classic non-parametric method is known as the periodogram, which is computed by applying a discrete Fourier transform to the sample data, however due to the spectral bias and the fact that the variance does not reduce at a given frequency as the number of samples increase this method does not result in a good spectral estimate. The technique that was utilized in this paper (Bartlett, 1948) reduces the variance and bias by taking the original $N$ point data segment and splitting it into $K$ data segments of length $M$, computing the Fast Fourier Transform (which gives complex numbers for each frequency), squaring the absolute value of each frequency component, and dividing by $M$. The results from the segment analysis are then averaged together to produce a single periodogram. Another classic approach to this problem is know as Welch’s method (Welch, 1967), which uses a modified version of Bartlett’s method in which the segments are split into overlapping segments and windowed in order to minimize information losses.

### A.2 Comparing Welch’s and Bartlett’s Methods for Wind Speed Data

In order to determine the appropriateness of using Bartlett’s method, as was done in this paper we compared our results to those obtained using Welch’s method. For this analysis we decided to compare the results from each method using a single site from the EWITS dataset using the data for a turbine with a hub height at 80 meters. The resulting PSD graphs are shown below in Figure A.1. When the segmented method is compared to Welch’s method it is clear that both methods result in harmonics at the same frequencies as well as similar magnitudes of the data for each frequency. As would be expected Welch’s method includes a larger number of frequency estimates, relative to Bartlett’s method with 30 segments. Table A.1 shows the resulting power-law slope estimates for each method. From this table it is clear that the results from this analysis are similar to those found in Table 2.2 with the exception that the data in the low frequency set have a slightly shallower slope,
relative to the Kolmogorov spectrum. This difference is most like due to the lack of spatial averaging that is observed between multiple wind farms.

Figure A.1: Comparison of the results obtained from the methods used in the paper for unsegmented and segment analysis compared to the classic Welch approach for a single site in the EWITS dataset at a hub height of 80 meters

Given that we do not observe major differences in the resulting slope estimates between Welch’s method and Bartlett’s method, we conclude that Bartlett’s method is sufficient for the purposes of this paper. If one were to need more refined detail about the variability of wind speed or power data at specific frequencies, one would need to consider Welch’s method, as it does provide more detailed information about the signal variance at specific frequencies.

Table A.1: Linear Regression slopes for comparing unsegmented, segmented and Welch results for a single site in the EWITS dataset at a hub height of 80 meters

<table>
<thead>
<tr>
<th></th>
<th>Linear Regression Slope for Unsegmented Analysis</th>
<th>Linear Regression Slope for Segmented Analysis</th>
<th>Linear Regression Slope for Welch Method</th>
</tr>
</thead>
<tbody>
<tr>
<td>$f \leq f_k$</td>
<td>-1.4525</td>
<td>-1.4062</td>
<td>-1.3857</td>
</tr>
<tr>
<td>$f &gt; f_k$</td>
<td>-2.4260</td>
<td>-2.4366</td>
<td>-2.4359</td>
</tr>
<tr>
<td>$f &gt; 10^{-6}$</td>
<td>-2.2701</td>
<td>-2.2627</td>
<td>-2.2612</td>
</tr>
</tbody>
</table>
Appendix B

Are Wind Data Gaussian?

As discussed in the literature review of Chapter 3, most papers in the area of research pertaining to regulation and reserve requirements use Gaussian statistical assumptions in determining requirements needed as wind penetration levels increase. Therefore, this work seeks to determine the implications of this assumption for calculating reserve requirements in systems with increasing wind penetrations. In order to determine if current techniques used to predict the amount of wind in a system should affect regulation requirements, we compare both prediction error and step-change analysis data of empirical wind data to their corresponding Gaussian distributions. For our prediction error analysis we used the 2010 wind data set from the Bonneville Power Administration balancing authority (BPA, 2010) to determine the total forecast prediction error for the BPA balancing area averaged over 10-minute intervals. While for our analysis of step-changes we compared 10-minute average step-change distributions of the BPA wind data from 2010, as well as the 10-minute averaged step-change analysis of Plant A Set 6 and Plant B from Chapter 2.

In order to determining the correlation of the empirical results for both prediction error and step-change analysis a normal assumption Figure B.1 and Figure B.3 were used to depict the difference between the empirical and corresponding Gaussian PDF’s. Table B.1 shows the statistics of the empirical data used to create the Gaussian PDF’s used in Figure B.1 and Figure B.3. Our analysis also included the results obtained from a normal probability comparison plot of BPA’s prediction error from 2010, shown in Figure B.2, in order graphically determine if the prediction error data could have come from a normal distribution.
Figure B.1: Empirical wind prediction error (solid line) versus the Gaussian distribution (slashed line) of the wind prediction error using the mean and standard deviation of the BPA balancing authorities wind load data for 2010.

Figure B.2: Normal plot of BPA prediction error data for 2010, where the dotted line represents a Gaussian distribution and the empirical distribution is represented by the + symbols.
Figure B.3: Step-Change analysis of empirical wind data (solid lines) versus their corresponding Gaussian distributions (slashed lines) using BPA’s wind load data from 2010, as well as Plant A Set 6 and Plant B from Chapter 2.

Table B.1: Statistics of the empirical wind data used in this study to create the corresponding Gaussian distributions

<table>
<thead>
<tr>
<th></th>
<th>BPA Prediction Error</th>
<th>BPA Step-Change</th>
<th>Plant A Step-Change</th>
<th>Plant B Step-Change</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \mu )</td>
<td>-0.0031</td>
<td>1.8863 \times 10^{-6}</td>
<td>-4.0371 \times 10^{-5}</td>
<td>-2.5797 \times 10^{-5}</td>
</tr>
<tr>
<td>( \sigma )</td>
<td>0.0589</td>
<td>0.014</td>
<td>0.0381</td>
<td>0.0509</td>
</tr>
</tbody>
</table>

From the results shown in Figures B.1, B.2 and B.3 it is clear that the distributions for both prediction error and step-change analysis fails to match the expected results from a corresponding Gaussian distribution. Most notably is that the Gaussian distribution curve fails to account for the values at the tails of the empirical distribution. Table B.1 shows the mean and standard deviation from each dataset used to create the corresponding Gaussian distribution. It is also known for a Gaussian distribution 99.6% of the data will be within plus or minus three standard deviations of
the mean. This knowledge coupled with the results shown in Figure B.1 and Figure B.3 show that this is not the case for the PDF for the empirical data, and can more clearly be seen in Figure B.2, which shows that the actual probability of the wind prediction error does not lie on the slashed linear line, representing the Gaussian assumption.

Coupled with these graphical results the mathematical probability was solved for to determine the odds that the wind prediction error could have come from a normal distribution using the Kolmogorov–Smirnov and Lilliefors tests. The Kolmogorov-Smirnov test works by comparing two distribution functions to each other and determining if they are equal, while the Lilliefors test is used to test if the data came from a normally distributed population. The difference with the Lilliefors test is it does not specify which normal distribution is being used, in other words it does not specify the expected value and variance of the data in question. Table B.2 shows the mathematical results obtained comparing the empirical wind data to a corresponding normal assumption. From these results it is clear that the empirical data does not correlate to a normal distribution.

**Table B.2:** Probability that the statistics of wind data used in this study matches that produced by a Gaussian distribution. Note that the smallest tabulated value for the Lilliefors test has a probability equal to 0.001

<table>
<thead>
<tr>
<th></th>
<th>BPA Prediction Error</th>
<th>BPA Step-Change</th>
<th>Plant A Step-Change</th>
<th>Plant B Step-Change</th>
</tr>
</thead>
<tbody>
<tr>
<td>$p_{\text{Kolmogorov-Smirnov}}$</td>
<td>$1 \times 10^{-3}$</td>
<td>$1 \times 10^{-3}$</td>
<td>$1 \times 10^{-3}$</td>
<td>$1 \times 10^{-3}$</td>
</tr>
<tr>
<td>$p_{\text{Lilliefors}}$</td>
<td>$5.35 \times 10^{-5}$</td>
<td>$1.36 \times 10^{-27}$</td>
<td>$6.205 \times 10^{-15}$</td>
<td>$5.712 \times 10^{-16}$</td>
</tr>
</tbody>
</table>

From this analysis it is clear that the statistics of wind data can not be accurately approximated with a normal distribution. The Gaussian assumption fails to model the fat tails of the probability density function of the data. This analysis supports the work presented in Chapter 3 by acknowledging the fact that Gaussian assumptions for do not accurately represent real wind data and as a result could lead to under-prediction when using models that incorporate these assumptions.