A PRODUCTION SIMULATION TOOL FOR SYSTEMS WITH INTEGRATED WIND ENERGY RESOURCES

BY

NICOLAS BENOIT PHILIPPE MAISONNEUVE

Diplome d'ingenieur, Supelec, 2007

THESIS

Submitted in partial fulfillment of the requirements
for the degree of Master of Science in Electrical and Computer Engineering
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2009

Urbana, Illinois

Adviser:

Professor George Gross

ABSTRACT

The rapid increase in wind power capacity over the past ten years has benefited tremendously from the technology advancements that brought about significant reductions in their investment costs. In addition, the growing concern about climate change has led several nations to adopt policies that foster the wider use of renewable energy sources in order to reduce CO_2 emissions. In fact, several jurisdictions around the world have specified ambitious targets for the fraction of capacity to be renewable resources, thereby stimulating additional investments in wind. As a result, wind is today the fastest growing source of new capacity for electricity. The high variability in wind speeds poses major difficulties in operating the power system into which wind resources are integrated. To effectively address these difficulties, operators need practical tools to deal with the intermittent nature of wind generation, given the lack of controllability of wind resources. Indeed, these problems become more pronounced as the penetration of wind increases. The nature of the difficulties is exacerbated by the restricted ability to accurately forecast wind speeds as the forecasting period increases. Indeed, wind speed forecasting introduces an additional source of uncertainty that must be considered. Without large-scale storage devices, operators manage the combined uncertainty effects by increasing the levels of the required system operating reserves resulting in increased system production costs. Consequently, there is an acute need for appropriate tools to allow the study of the effects of the integration of wind resources into the grid from a planning perspective. This thesis describes the construction and testing of a probabilistic

production simulation tool with the capability to quantify the variable effects of systems with varying wind penetration.

The initial step in the engine construction is the development of a wind speed model that explicitly represents the variability and the uncertainty in the wind speed. To represent the change in the daily wind patterns, we adopt a regimes-based approach: we make use of statistical clustering algorithms to construct classes of days with similar wind patterns and we probabilistically represent the wind speed for each class or regime. We extend the modeling of a stand-alone wind turbine so as to represent the energy output of a wind farm in a single location and of those in multiple locations. In this way, we are able to represent the aggregate wind power production of a system with wind farms located at disparate sites. The modeling work is for planning purposes and the level of detail is commensurate with the needs for such planning tools.

The incorporation of the wind energy model requires the extension of the widely used probabilistic simulation tool in the studies of systems with integrated wind resources. We recast the load representation so as to make it compatible with the wind power production model and merge the two models with the same level of resolution to develop the model of the load that needs to be served by the controllable units in the system. The resulting methodology is able to simulate systems with integrated wind resources with the explicit representation of various sources of uncertainty.

The principal interest in the use of the extended production simulation tool is to quantify the impacts of the integration of wind resources into the system on the variable effects over longer-term periods. We illustrate the application of the tool using the extensive studies we performed under a wide range of conditions and on systems in

different geographic regions. The studies we discuss provide a realistic assessment of the impacts of the wind energy variability and intermittency on the systems' expected production costs, CO_2 emissions and reliability metrics. In addition, we demonstrate the extent to which the presence of wind farms at different locations is able to help in the effective management of the wind energy intermittency effects. We discuss a set of different sensitivity cases, including a thorough evaluation of the effects of increased wind resource penetration, to obtain insights into the impacts of wind energy penetration.

ACKNOWLEDGMENTS

I would like to thank my adviser, Professor George Gross, for the advice, guidance and support he gave me throughout my studies at the University of Illinois. I also thank him for the time he spent proofreading and improving the content of this thesis.

I also thank Professor Alejandro Dominguez-Garcia with whom I had the chance to discuss some of the issues dealt with in this thesis.

A special thanks to my friends, Mathieu and Christophe with whom I spent a lot of time during the past two years. I enjoyed their continuous presence.

I would also like to thank all the power group students, staff and GGG members for their help and friendship. I am also indebted to my successive officemates, Jonathan, Hector, Jashua and especially Ryan, for the wide range of his topics of conversation and outstanding presence in the office. I enjoyed discussing with all of them.

Finally, nothing would have been possible without the love of my family, who have always supported me during the two years spent in Urbana-Champaign.

TABLE OF CONTENTS

1. INTRODUCTION

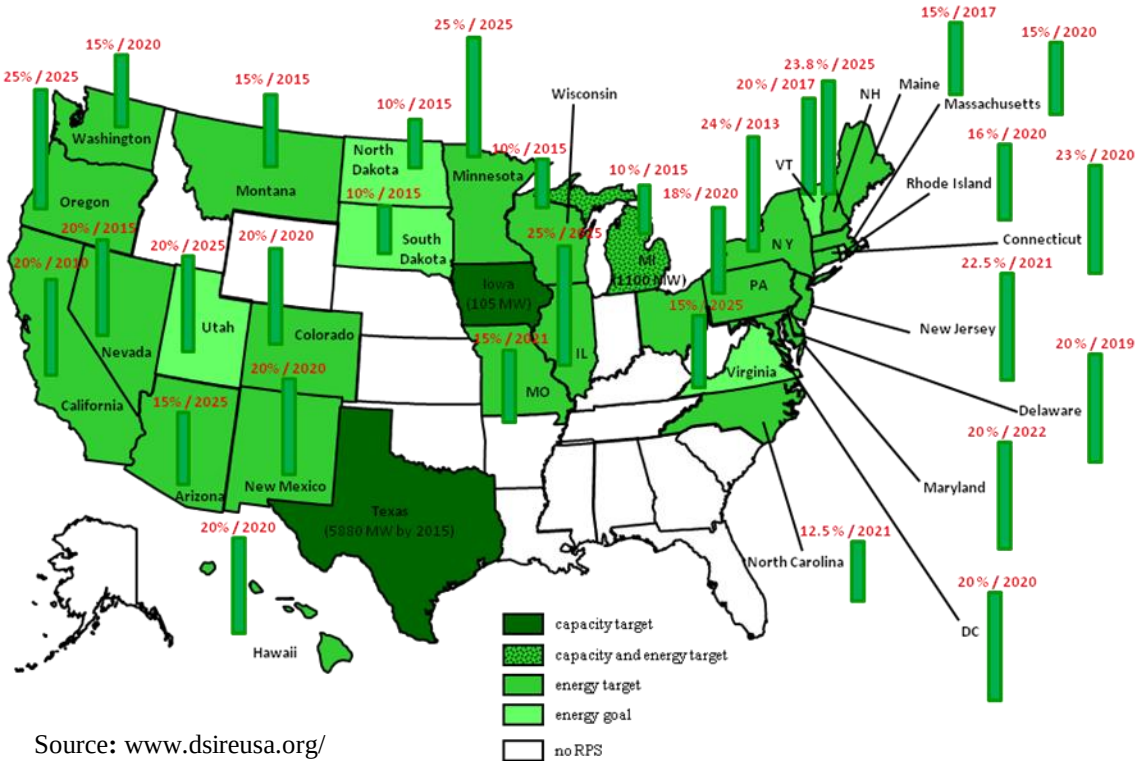
This thesis deals with the development of a production simulation tool for systems with integrated wind resources. In this chapter, we start with establishing the motivation and the need for the work presented in the thesis. In Section 1.2, we review the prior work in this area. Then, we present the scope and the nature of the contributions of the work. We end with outlining the contents of the chapters that follow.

1.1. Background and Motivation

The growing concern about the impacts of global warming has gained momentum in the past few years leading several countries to enact laws to force their citizens and industries to curb greenhouse gas emissions. At the same time, there are concerns about the economic level of available fossil fuel reserves in the world, on the one hand, and the high prices and volatility in the fuel resource markets, on the other hand. Many nations have therefore established an energy independence policy so as to foster the use of alternative resources to generate energy. The key focus is on the exploitation of renewable sources of energy, making use of naturally occurring and quickly replenishable resources. The utilization of such resources paves the way to a sustainable energy path to meet future needs. Consequently, there has been a considerable push for deeper penetrations[1] of renewable energies in the past few years. For instance, in the US, several jurisdictions have set ambitious goals – the so-called Renewable Portfolio Standards – in terms of the percentage of the electric energy that needs to be served by renewable

[1] The energy penetration is defined as the fraction of the demand served by a specified type of resource. The capacity penetration is the ratio between the installed capacity of the considered resource and the peak load of the system.

energies. In Fig. 1.1, we give a visual illustration of the extent and nature of the RPS in the states where it has been adopted. More precise information about such targets is available [1].

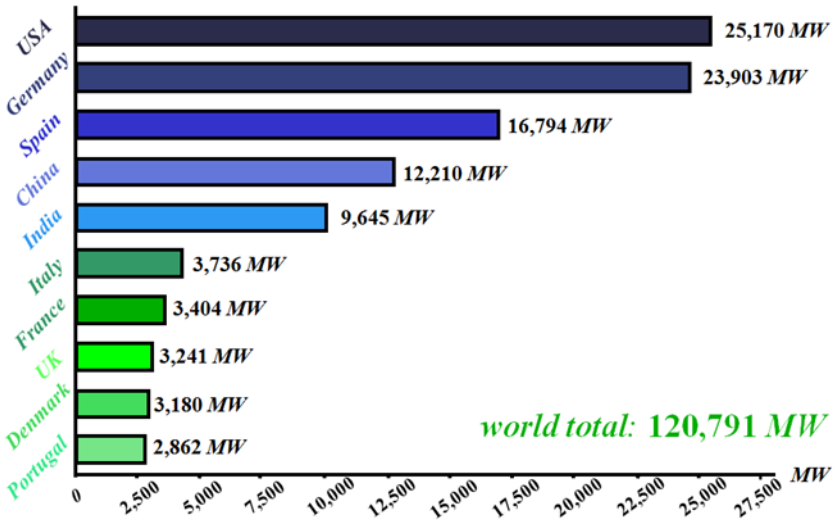


Source: www.dsireusa.org/

Figure 1.1: The renewable portfolio standards map.

Several natural resources may be used to generate electric energy in a renewable fashion: hydro, geothermal, tidal, wind and solar. The use of such energy sources is not only encouraged due to their renewability, but also because of the absence of fuel costs when we harness them. For instance, the energy contained in the wind is harnessed making use of wind turbines and no fuel is needed to produce the electric power. The wind energy has attracted great interest and is the fastest growing energy resource in the world in terms of capacity: the number of installed wind turbines has grown

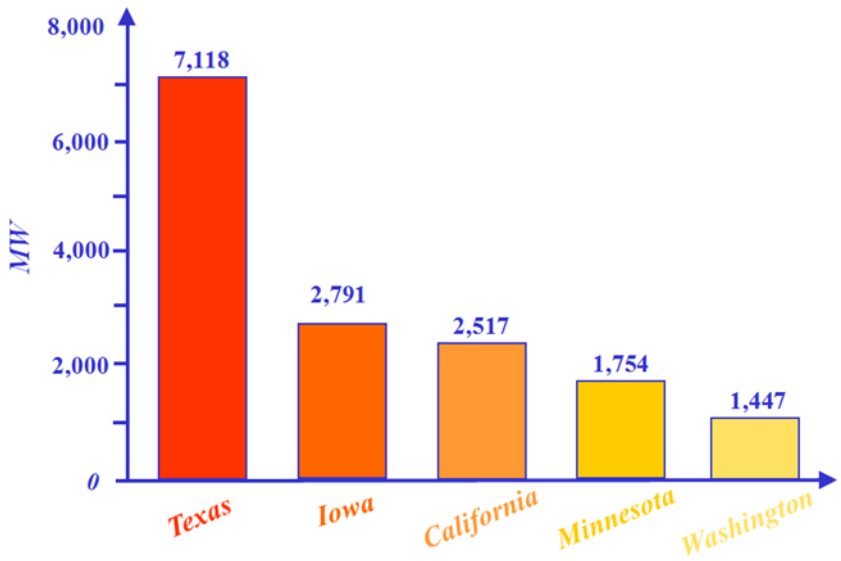
exponentially in the past few years [2] and there is still a huge potential for additional wind power generation devices. In Fig. 1.2, we show the countries with the largest wind capacity: in Europe, countries like Germany, Spain and Denmark have led the effort in the integration of large renewable energy resources into their electric system. In the US, large capacities of wind resource have been installed in California and Texas as shown in Fig. 1.3. The trend toward deeper wind penetrations has encouraged turbine manufacturers to make significant progress in terms of the nameplate capacity of the wind units [2]. The spectacular improvements in the technological capabilities of the wind generation devices are probably due to the large number of turbine manufacturer[2] and the fierce competition between them.



Source**:** GWFC--http://www.renewableenergyworld.com/rea/news/article/2009/02/wind-instalallations-continue-to-break-records-across-the-globe-54658

Figure 1.2: 2008 installed wind capacity – top 10 countries.

[2] The list of the wind turbine manufacturers includes Vestas, Siemens, RePower, General Electric, Suzlon, Gamesa, Acciona, Nordex, and Enercon, as well as smaller firms.



Source: AWEA, http://www.awea.org/pubs/factsheets/Market_Update_4Q08.pdf

Figure 1.3: Installed wind capacity – top 5 states.

However, although the wind energy presents some undeniable advantages, it also has inherent characteristics that complicate the reliable and economic integration of wind resources. For instance, the wind power production is not controlled by the system operator and is driven by climatic conditions. More precisely, the instantaneous wind energy production is directly related to the actual wind speed. As the wind speed fluctuates over time, so does the power produced by the wind farms. An unfortunate consequence of such a characteristic is the wind energy intermittency, i.e., the fact that the wind turbines do not produce energy all the time. Indeed at low wind speeds, the wind does not contain sufficient energy to put in motion the blades of the wind turbines. Also, the wind units are shut down when they are experiencing high wind speeds so as to protect them from being destroyed. As a result, the integration of wind resources into the existing power grid raises some serious issues. Indeed, in a power system, the supply needs to track the demand around the clock. As the energy is a commodity that can be

hardly stored, the power is produced on demand so that the balance constraint is enforced. So, the wind power fluctuations necessitate that the system operators schedule and commit controllable units which have the capability to track the variations of both the wind power production and the load.

Due to its variability, the wind speed is also hardly predictable. The wind power production directly depends on the wind speed and has therefore a lack of predictability, the forecasting error increasing with the forecast horizon. Such a characteristic is unfortunate as it results in additional major complications of the operations of power systems with wind resources. Indeed, a system with integrated wind resources is scheduled making use of the wind power forecast. However, the actual wind power differs from the forecasted one and the system must have the capability to be redispatched in the real-time operations so as to compensate for the mismatch resulting from the wind power forecast error. More precisely, if the wind power has been under-estimated, too much controllable generation is available and some units need to lower their output. In contrast, if the wind power has been over-estimated, there is a lack of supply and the so-called reserves are deployed. So, due to the wind power lack of predictability, the system might face unexpected lack of generation raising reliability issues. Moreover, the adjustment of the schedule to compensate for the wind power forecast error comes at a cost that could have been avoided with better forecasting.

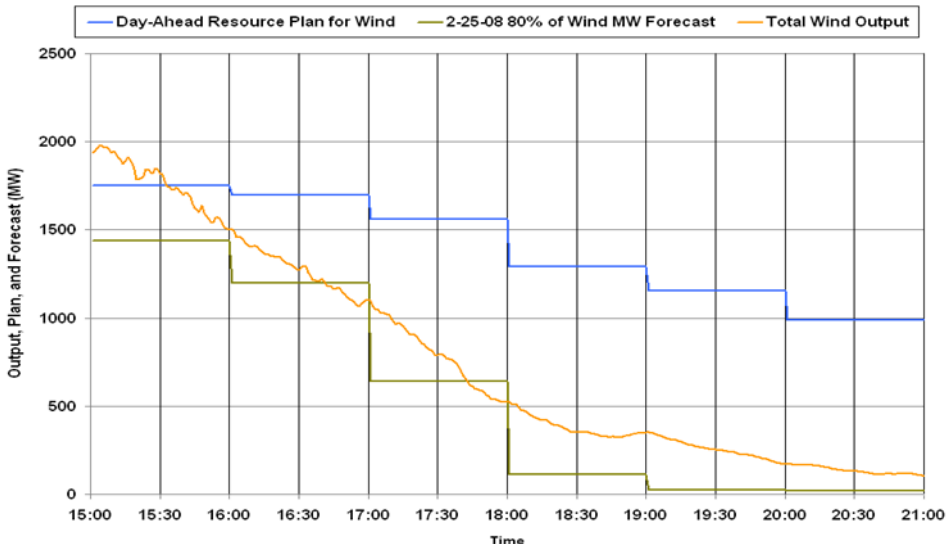
The challenges raised by the integration of wind resources are illustrated with the February 26, 2008, ERCOT[3] event. The state of Texas has adopted a very aggressive policy in terms of the wind resources integration and is therefore at the cutting edge of

[3] The Electric Reliability Council of Texas (ERCOT) operates the electric grid and manages the electric market of 75 % of the Texas land area (85 % of the state load).

this domain. Due to its high wind penetration, the system is facing considerable challenges to ensure its reliability. On February 26, 2008, ERCOT experienced conditions that might have led to the blackout of the entire Texas grid were it not for a combination of fortuitous events. Indeed, the operators had to shed some responsive load as the lack of supply was becoming too pronounced, resulting in a frequency drop and threatening the system stability. Specifically, the conditions were the result of the combination of three separate factors.

- At 15:00, the wind generation started to ramp down with a decrease of 1,500 MW of wind generation over the ensuing three hours;
- There was a rapid load increase of about 2,550 MW over 45 minutes that begun around 18:00; and,
- The system experienced an unexpected drop in conventional generation and a controllable unit also tripped at 17:44 resulting in a 150 MW loss in generation.

The drop in wind generation is noticeable in the plot of the wind output shown by the yellow curve in Fig. 1.4. In addition, in parallel with the load increase, the system suffered from a drop of its reserves making ERCOT enter into an Emergency Electric Curtailment Plan (EECP) step 2 at 18:41. Upon entry into step 2, ERCOT invoked load reduction by reducing the voltage. Also, at 18:49, ERCOT dispatched the LaaRs, i.e., loads acting as a resource. Within the ensuing 10-minute interval, the LaaRs shed 1108 MW of load resulting in a total of 1200 MW of load tripped. At 18:52 the frequency returned to 60 Hz, and at 20:40 the EECP was terminated. The entire sequence of events lasted less than two hours. A detailed discussion of the exact event chronology is given in the report prepared by ERCOT [3].



Source:www.ercot.com/meetings/ros/keydocs/2008/0313/07._ERCOT_OPERATIONS_REPORT_EECP022608_public.doc

Figure 1.4: The real-time wind power output, the wind power forecast and the wind power resource plan of the ERCOT system during the February 26, 2008, event.

While no customer lost power without his agreement, the event threatened the system reliability and its analysis is warranted. Indeed, had ERCOT originally scheduled more capacity to be on-line, the need for emergency response could have been avoided and no financial compensation to the customers who agreed to curtail their load would have been made. Specific recommendations for improvement are given in [4]. A key area is better load forecasting. Indeed, the load forecast heavily relies on historical data and is likely to be good if the actual load pattern does not differ too much from the previous day load shape. However, the load patterns on February 25 and 26 were quite different, impacting the quality of the short-term load forecast and the actual load increased ahead of the predicted time. Another area is the more effective management of wind generation variability, especially in the forecasting field. Actually, the Texas system has considerable experience in handling large drops in wind generation [5]. But, in the February 2008 event, ERCOT did not have an accurate forecast to compensate for the

wind power loss that occured. In Fig. 1.4, we see that the “day-ahead resource plan” did not include the consideration of the magnitude of the drop in wind power generation due to the lack of an accurate forecast. In contrast, the “80% wind forecast” that was available but not considered in the scheduling correctly fits the actual data. Clearly, the management of systems with deep wind power penetration requires the availability and deployment of good forecasting tools. Otherwise, the system must schedule additional reserves at levels commensurate with the uncertainty of the wind power forecast.

The event which occurred in the ERCOT system illustrates the necessity to get a better understanding of the issues that arise when wind resources are integrated into the grid. In this thesis, our aim is to identify the impacts of the wind integration on the system operations so as to develop a planning tool to quantify the variable effects of systems with integrated wind resources. We modify the probabilistic production simulation tool so that it captures the various sources of uncertainty including that of wind generation and reflects the impacts of the wind resource integration on the power system production costs, CO_2 emissions and reliability level. But, before we further discuss the construction of the simulation tool, we first review the literature.

1.2 Survey of the State of the Art

The integration of wind resources into the electric grid has been the subject of much research work in the past few years. We provide a review of the literature related to the issues we are dealing with in the thesis. In particular, we review the work that has been done in terms of wind speed and energy modeling, forecasting issues as well as the impacts of the wind resources integration on the power system operations and planning.

The modeling of the wind speed and the wind energy is a key issue that needs to be addressed in the studies of systems with integrated wind resources as their variability and uncertainty must be appropriately represented. The model's complexity depends on the nature of the study to be conducted and the level of detail of the phenomena we want to capture. In the case where temporal considerations are neglected, the wind speed is typically modeled as a random variable characterized by its distribution functions, the Weibull and Raleigh distributions being widely used [6]. In the case where temporal effects must be represented, more complex approaches need to be considered. Time-series based models [7] are suitable for such applications. If we consider systems where the wind resources have been dispersed over a large geographical area, we also need to take into account the geographical variability of the wind speed [8], [9], [10].

The energy contained in the wind is converted into electric energy by the wind turbines, and their output power production directly depends on the speed of the wind at the location under consideration. So the modeling of the wind turbines requires some attention to obtain a representation of the wind power. A single wind turbine is typically modeled with the so-called power curve [6]. In some cases, the wind power is directly modeled without considering the wind speed. For instance, a Markov-chain approach may be used [11], [12] to model the chronological wind power production of a particular system. Some research has also been done to represent the wind power behavior with a stochastic differential equation. In [13], the author shows that with such a model, the statistical properties of the simulated wind power generation are close to those of the historical wind power. Additional work has also been done to probabilistically characterize the wind power production. The methodology to compute the probability

density function and the cumulative distribution function of a single wind unit is detailed in [14]. Some work on this topic is also proposed in [8].

We next review the relevant work in the forecasting field. The improvement of actual wind speed and power forecasting tools is a critical need for the more effective integration of wind resources and has been identified as a source for potential large savings. Therefore, this area has been subject to a lot of research. The wind speed forecasting techniques are basically divided into two groups [15], [16]: the methods that require physical modeling and the statistical ones. In the physical approach – Numerical Weather Prediction or NWP – the atmosphere behavior is modeled with a set of differential equations that we numerically solve for some grid points. This technique is very demanding in terms of computational resources and is well-suited for forecasting applications with lead-times ranging from 6 to 72 hours. With the statistical approach, we forecast the wind speed from historical values. For instance, time-series and Artificial Neural Network – ANN – approaches have been proposed and are widely described in the literature [17]. The statistical methods are usually simple and require little computational resources. In addition, they outperform the NWP method for forecasting horizons inferior to three hours.

The modeling of the forecasting error is also crucial for conducting studies to determine the impacts of the lack of predictability of the wind resource on the power system operations, reliability and production costs. The wind power forecast is usually modeled as a random variable with a normal distribution [18]. However, the limitations of such a modeling approach have been demonstrated in [19] and the use of a beta

distribution is more appropriate to represent the wind power forecast error random variable [20], [21].

When a wind farm is erected, we are interested in quantifying its economic and reliability impacts on the actual power system. The capacity factor of a wind unit is a widely used indicator to measure its productivity. This metric is defined as the ratio between the actual and the maximum wind energy productions over a given period of time. Typically, wind farms have capacity factors ranging from 25% to 40% [22]. However, such a measure fails to indicate any information on the contribution of the wind farm to the system reliability. For the same capacity factor, two wind farms might not bring the same reliability benefits depending on the wind speed variability characteristics at the two considered locations. For instance, in the case where the wind speed diurnal pattern is positively correlated with the load pattern, the wind farm produces more power during the high-load periods and therefore largely participates in ensuring the system reliability. In contrast, in the case where the wind speed pattern is negatively correlated with the demand, the high wind power production periods occur at night when the load is low and the system reliability is already high. To assess the reliability effects of the integration of wind resources, the notion of capacity credit has been introduced. It measures the amount of controllable capacity that can be replaced by the considered wind farm without modifying the system reliability and is expressed as a percentage of the nameplate capacity. The notion of capacity credit provides us with essential information about the “worth” of a wind farm but its value depends heavily on the reliability metric employed in its evaluation process [23].

Due to the wind power variability and lack of predictability, the modification of the system operations is required to maintain the reliability. Typically, increased reserve levels are dispatched to manage the wind variability associated with wind generation. The quantification of the necessary reserves level is complex and has been the subject of a lot of research. In [24], the authors propose a methodology where they adopt a probabilistic approach to quantify the necessary amount of reserves to mitigate the effects of the uncertainty contained in the load, the wind power production, and the controllable generators' outages. Several comprehensive studies have also studied the impacts of the wind resource integration for several US systems: Minnesota [25] and NYISO [26]. In such studies, the authors perform chronological production simulation where the simulation period is divided into several subperiods and the system operations are simulated step by step. The simulations are repeated several times for different system configurations and wind penetrations so as to get a clear picture of the effects of the wind integration effects. Among other things, the authors emphasize the costs induced by the management of the variability and the intermittency effects attendant with the integration of wind resources.

1.3 Scope and Contributions of the Thesis

While many wind integration studies employ chronological production simulation to study the systems with wind resources, little work has been done on the study of wind resources integration with the use of the probabilistic production simulation framework. More precisely, little emphasis has been put on the adequate modeling of the wind speed and the wind power in the probabilistic simulation. The key requirement is therefore to

develop a wind energy model suitable for probabilistic simulation and that correctly captures the salient wind speed and wind energy characteristics.

In this thesis, we study the nature of the wind speed and the wind energy so as to get a thorough understanding of the natural phenomena that are relevant for our work. Motivated by the insights that we gained, we develop a model for the wind resources which represents the variability and intermittency of the wind energy. The modeling approach we adopt has the capability to take into account the geographic diversity of the wind farm locations so as to correctly model the wind power production resulting from dispersed wind farms.

We use the wind energy model we developed to modify the probabilistic production simulation tool so that it takes into account the variability and intermittency of the wind resources. Additionally, the tool must have the capability to represent the impacts of the change in the reserve level on the system variable effects over periods of longer duration. The usefulness of the probabilistic production simulation tool is demonstrated through many numerical studies. For instance, we make use of the tool to observe the effects of increasing the wind penetration on the system production costs, reliability and CO_2 emissions.

1.4 Outline of the Thesis Contents

This thesis is composed of five additional chapters. In Chapter 2, we start with discussing the salient characteristics of the wind speed and the wind energy. We use historical chronological data to qualitatively analyze their nature, characterize their variability and illustrate the resulting forecasting limitations. Then, we discuss the

complications of the system operations that stem from the integration of wind resources into the grid. We also indicate the potential solutions that must be implemented by the system operators to mitigate the wind power variability effects and facilitate the increase of the wind penetration.

In Chapter 3, we focus on modeling issues. We adopt a wind speed model suitable for longer-term planning analysis and which explicitly represents the wind speed variability and intermittency. The wind model we develop also captures the salient wind speed patterns or regimes at the locations with wind farms. To characterize each regime, we group together the days of wind speed with similar shapes and determine their associated probabilities. The grouping process is performed with the two clustering algorithms we describe and extensively test in this chapter.

In Chapter 4, we describe the probabilistic production simulation framework and justify its use in longer-term planning studies. We next detail the modifications we undertake so that the wind resources characteristics are taken into account. We adopt a load model whose level of detail is commensurate with that of the wind power model. We introduce the notion of "controllable" load which we define as the fraction of the demand that needs to be served by the controllable units. To run a production simulation of a system with wind resources, we therefore need to combine the load and the wind power model to obtain the "controllable" load model which we use as an input of the probabilistic production simulation framework.

In Chapter 5, we illustrate the applications of the modified production simulation tool. We describe the set of studies to be conducted, run the probabilistic production simulations and analyze the results we obtain.

In Chapter 6, we summarize the insights gained from the production simulation results. We also indicate directions for future work so as to extend the capabilities of the production simulation tool and run simulations for even more realistic test systems.

This thesis also has two appendices. Appendix A provides detailed algorithms of the two clustering schemes we use to classify the wind speed data. In Appendix B, we perform a statistical analysis of the wind speed data by evaluating the correlation coefficients between the wind speeds at different locations.

2. DIFFICULTIES IN THE INTEGRATION OF WIND RESOURCES AND THEIR MANAGEMENT

This chapter describes the nature of wind speed and generated energy as well as the impacts of the integration of wind resources into the grid. The wind energy is generated by transforming the kinetic energy of the moving masses of air into electric energy, making use of wind turbines. An attractive feature of the wind energy is that it has no fuel costs. Also, wind units emit no noxious gases into the atmosphere, so wind is generally viewed as a green resource. These two features make very attractive the exploitation of wind resources and have resulted in the construction of wind farms in a large number of locations in the US and the world. Unfortunately, wind energy also has drawbacks that limit the usefulness of the integration of wind resources into the grid. An inherent characteristic of wind speed is its variability, impacting markedly the times and the quantities of wind energy production. In addition, the randomness of wind speed and the lack of controllability of wind energy pose major challenges to the reliable and economic integration of wind generation.

In this chapter, we make use of historical data to qualitatively assess the nature of wind by focusing on the characteristics that impact wind energy. We also identify the principal difficulties in the operations of systems with integrated wind resources. We end the chapter with a description of the actions needed to effectively overcome the problems entailed by the integration of wind resources into the grid.

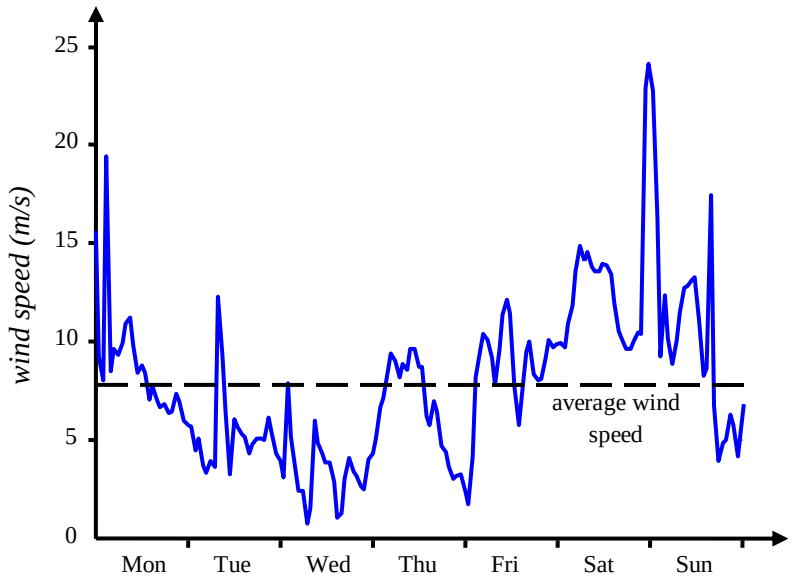
2.1 Characterization of the Wind Speed

Wind speed depends on several physical and climatic factors, including temperature and atmospheric pressure as well as the geographic location. As some of the variables that influence the wind interact with one another and are random, we may view the wind speed as a rather complex phenomenon. Consequently, its modeling is challenging as we need to take into account a large number of variables which influence the speed and direction of the wind. In this section, we provide a qualitative assessment of wind by discussing the key characteristics of wind which impact its usefulness as a resource. We focus on the determination of the physical nature of the wind speed since the output power of a wind turbine is directly related to the speed of the wind at the turbine location.

Without considering all the parameters which influence the wind speed, we study its main variability characteristics. The analysis of wind speed data shows the existence of very short-term – second-to-second and minute-to-minute – variations. In Fig. 2.1 we use data provided by NREL [27] to plot the chronological hourly wind speed in Adair, IA, at a height of 100 m for the week of June 26 – July 02, 2006. We notice that the hour-to-hour wind speed variations are rather substantial.

Also, the fact that wind speed is subject to longer-term variations is well known [28], [29], [30]. Frequency analysis of wind speed data indicates the existence of a daily, a weekly and also a yearly wind speed cycle. More specifically, the wind speed has diurnal variations whose shape depends on many factors including the evolution of the temperature at the study location throughout the day. In some cases, wind speeds tend to be stronger during the daytime [28]. In contrast, wind speed in Ontario is characterized

by peaks during the night due to the temperature drops in that period [30]. As such, no generalization can be drawn about the shape of the diurnal wind speed pattern. Moreover, the wind speed has, typically, longer-term seasonal variations, with the seasonality effect depending on the geographic location [30].



Source: ftp://ftp2.nrel.gov/pub/ewits/TimeSeries/LandBased/2006
(File: SITE_1094_OUTPUT_V4_2006.CSV)

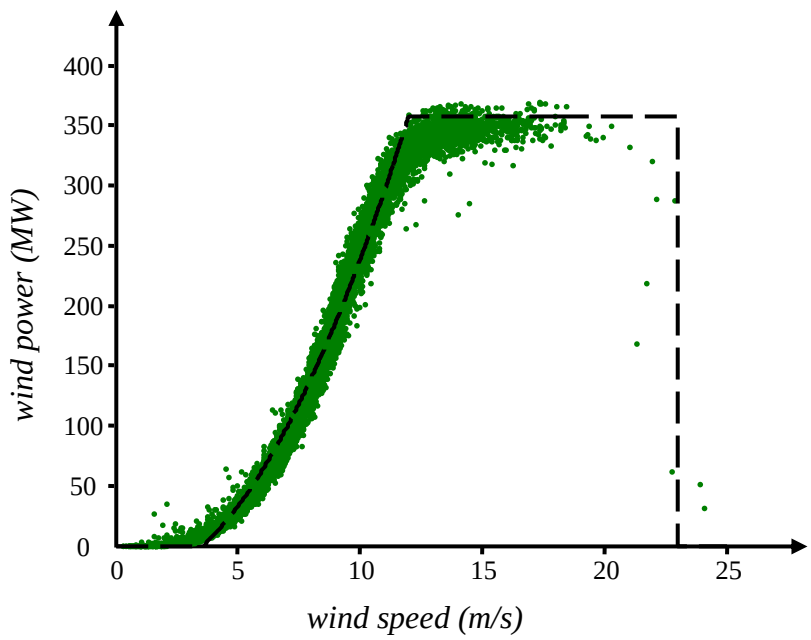
Figure 2.1: The hourly chronological wind speed in Adair, IA, for the period of Monday, June 26, to Sunday, July 02, 2006.

In this section, we reviewed the main characteristics of wind speed variability. Such variability effects influence the wind energy production, which we discuss in the next section.

2.2 The Nature of Wind Energy

The energy obtained at the output of a wind unit depends on the instantaneous wind speed at the specified height of the tower and the location where the turbine is operated. Indeed, the faster the wind is blowing, the more energy it contains and the more

electric power can be produced. Therefore the wind energy presents the same variability characteristics as the wind speed. In addition, the wind energy production also depends on the way the wind turbines are operated. To better understand the nature of wind energy, it is essential to study the so-called power curve which characterizes the wind speed to wind power conversion process. We use the hourly wind speed and power production data collected for the Adair site for the year 2006 to plot the scatter diagram of the 8,760 points as a function of the wind speed in Fig. 2.2. We also plot the averaged power curve, whose shape is a good approximation of the scatter plot formed from all the hours of the year.



Source: ftp://ftp2.nrel.gov/pub/ewits/TimeSeries/LandBased/2006
(File: SITE_1094_OUTPUT_V4_2006.CSV)

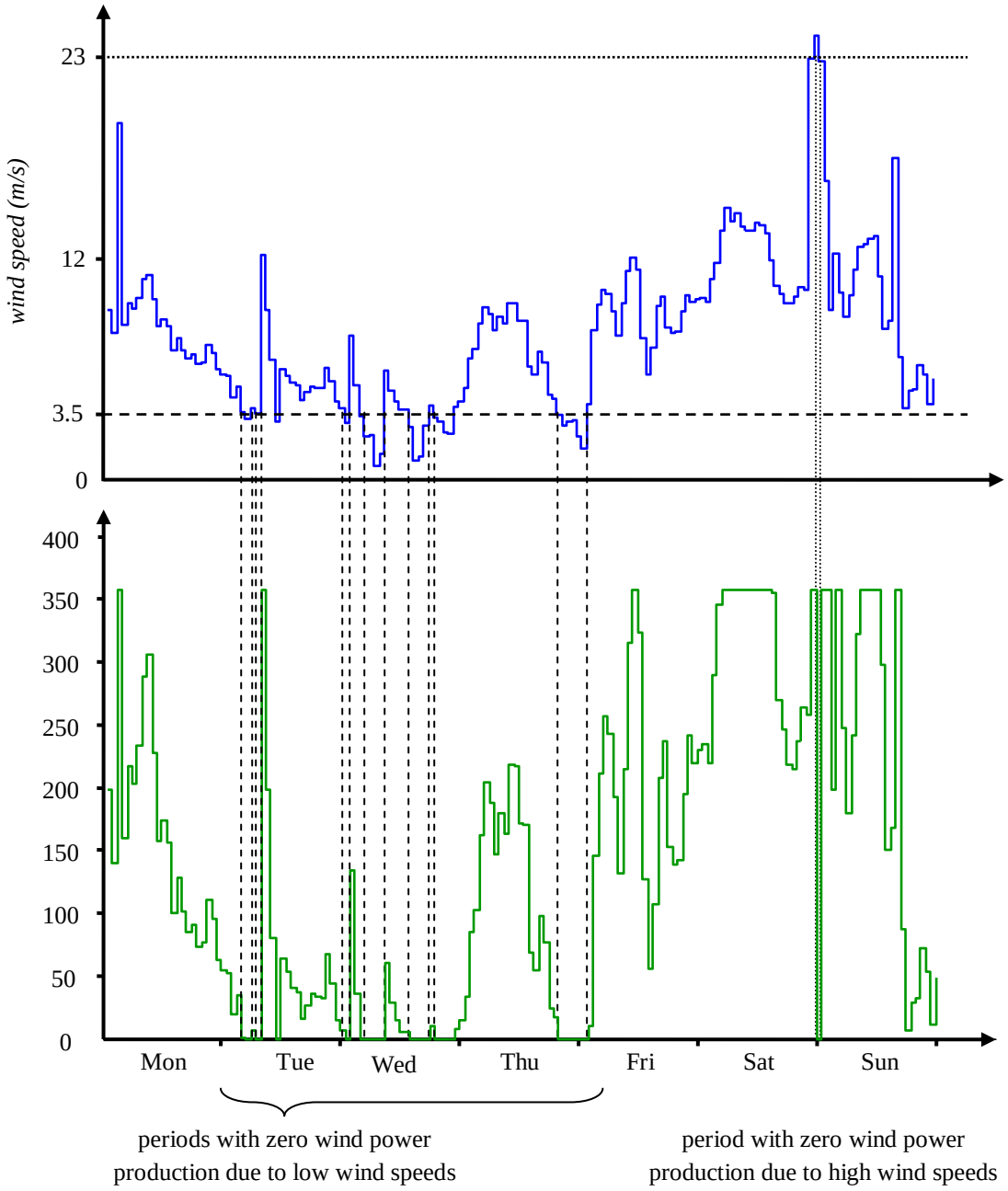
Figure 2.2: Scatter plot showing the 8,760 hourly wind speed and energy production pairs collected at the Adair, IA, location by NREL together with the power curve.

We observe that the wind speed to wind power transformation is non-linear. We also notice that for low wind speeds, the wind farm does not produce energy as the wind is not sufficient to move the blades. Furthermore, for a wind turbine that experiences extremely high wind speeds, the turbine is shut down so as to avoid any damage. This variable output characteristic of wind turbines is a salient characteristic of wind power production and is exhibited by sudden decreases in power output whenever the wind speed is close to the threshold and sudden increases as the wind speed changes but remains within the allowable limits. The fact that the output power of a wind farm may be equal to zero during periods when the wind is blowing too fast or too slow is referred to as the wind energy intermittency effect.

We further illustrate the intermittency effect by computing the hourly wind power production corresponding to the chronological wind speed shown in Fig. 2.1, using the wind farm power curve approximation depicted in Fig. 2.2. Under the assumption that the wind speed and wind power are constant for every hour, we produce the plot in Fig. 2.3. We observe large hour-to-hour wind energy variations and we note that whenever the wind speed is either above 23 or below 3.5 m/s, there is no wind power production.

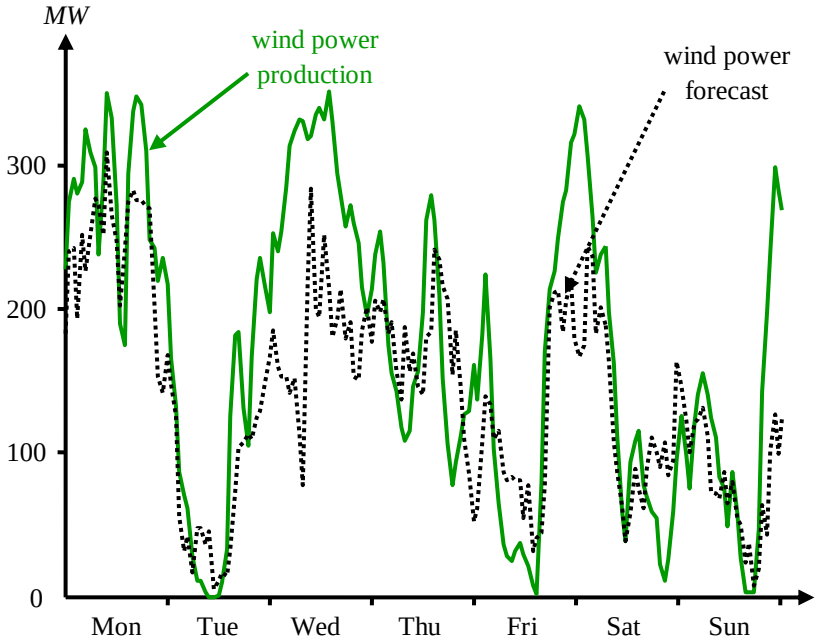
The variability and randomness of the wind energy make its accurate prediction very difficult. This difficulty constitutes a major complication in the operations of systems with integrated wind resources. In Fig. 2.4, we compare the wind power forecasted 4 hours ahead in Adair, IA [27], with the actual wind farm output power.[4] Although the forecast follows the trends in the wind power variations, we observe that the forecasting error is significant and may not be neglected.

[4] The forecasting technique is not mentioned by NREL.



Source: ftp://ftp2.nrel.gov/pub/ewits/TimeSeries/LandBased/2006
(File: SITE_1094_OUTPUT_V4_2006.CSV)

Figure 2.3: The hourly chronological wind speed and power production in Adair, IA, for the period of Monday, June 26, to Sunday, July 02, 2006.

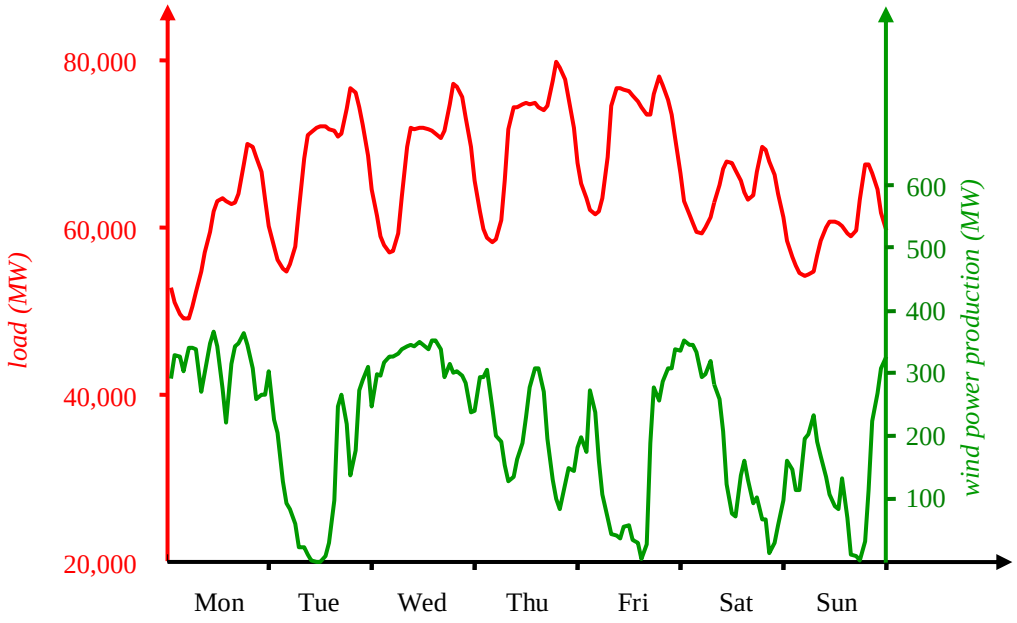


Source: ftp://ftp2.nrel.gov/pub/ewits/TimeSeries/Forecasts/Landbased
(File: SITE_1094_MAIN_4HR_FORECASTS.CSV)

Figure 2.4: The hourly wind power production and four-hours-ahead wind power forecast in Adair, IA, for the period of Monday, January 02, to Sunday, January 08, 2006.

When wind resources are integrated into the grid, the wind energy production is used to serve the total load demand. Since system operators cannot control the outputs of wind resources, the wind energy is used whenever available with the explicit aim of avoiding "spillage" of wind energy – unnecessarily shutting down the wind resource even though there is adequate wind speed to produce energy. The extent to which the wind energy and the load are related is consequently an important consideration in the effective utilization of wind. Clearly, the benefits obtained with the integration of wind resources are higher if the periods with high wind energy production correspond to those with high loads. The assessment of the load characteristics and those of the wind energy production is important, and so is their comparison.

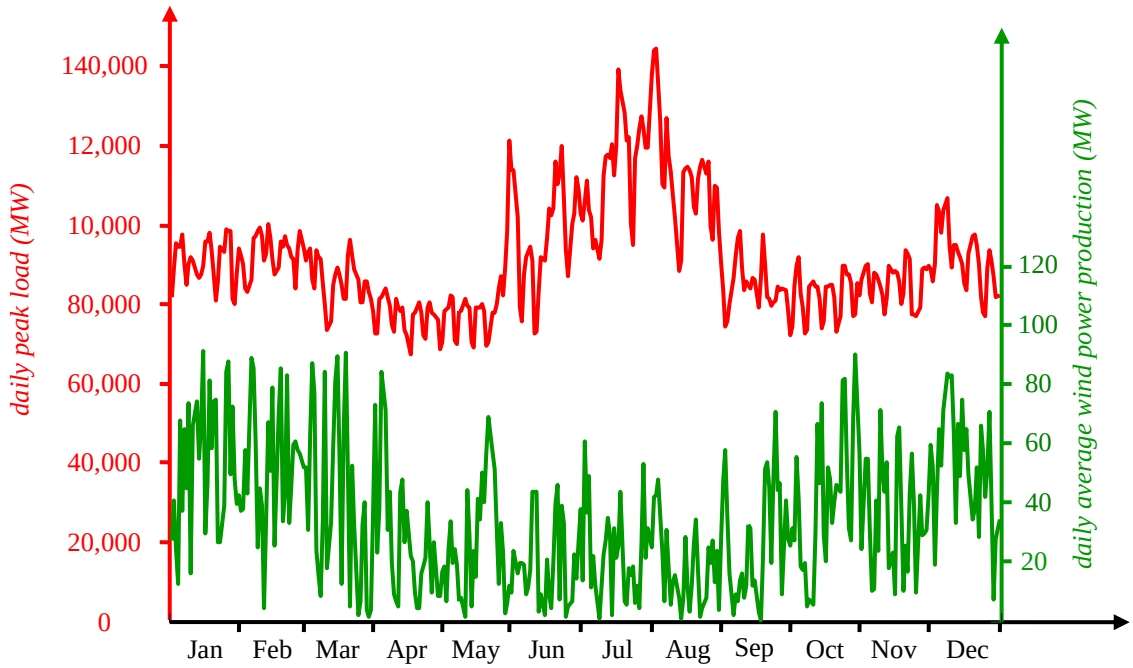
Unlike the wind speed, the load follows a well-defined diurnal pattern with loads higher during the week than on the weekend and reaching peaks during the day and low loads during the night. In Fig. 2.5, we plot a typical chronological hourly load shape of the MISO demand for a winter week. A salient characteristic is the strong similarity between the daily load patterns each day. We add the chronological wind power production plot for Adair, IA, obtained from the NREL data [27] on the same diagram. We note that the periods with high (low) wind power production do not necessarily occur when the high (low) loads do. Such lack of correlation presents a complication in the integration of wind resources into the grid.



Sources:
http://www.midwestmarket.org/publish/Document/4285c4_115415e4f37_-7ec10a48324a?rev=2
ftp://ftp2.nrel.gov/pub/ewits/TimeSeries/LandBased/2006
(File: SITE_1094_OUTPUT_V4_2006.CSV)

Figure 2.5: The hourly chronological load in the MISO system and wind power production in Adair, IA, for the period of January 2-8, 2006.

Another important aspect is the load seasonal characteristics and their alignment or lack thereof with those of the wind resources' outputs. For instance, in the US, most utility systems are summer peaking: the load experiences high values due to the extensive use of AC devices during the summer season. We illustrate this phenomenon in Fig. 2.6, where we provide the plot of the daily peak loads in the PJM system for the year 2006 and the daily average wind power production at Allegheny Ridge, PA, obtained from the NREL data [27]. We observe for the load and wind production a strong disalignment since there is a marked decrease in the wind power production during the summer period when the load reaches its peak. Such misalignment adds to the complexity of wind integration.



Sources:
http://www.pjm.com/.../compliance/historical-load-data/2006-hourly_loads.ashx
ftp://ftp2.nrel.gov/pub/ewits/TimeSeries/LandBased/2006
(File: SITE_4183_OUTPUT_V4_2006.CSV)

Figure 2.6: The daily peak load in the PJM system and the daily average wind power production of a 100 MW wind farm in Allegheny Ridge, PA, for the year 2006.

2.3 Difficulties in the Integration of Wind Resources into the Grid

A power system is planned and operated such that the supply meets the demand at every time instant in the most economic way. The objective of ensuring supply-demand balance around the clock requires that the operators ensure power generated exactly equals the demand 24/7. To meet this objective, the operator adjusts the outputs of the controllable units by sending raise and lower requests to the units as necessary. As there are different time frames in power system operations, the tasks of ensuring the tracking of the demand by the generation must be appropriate for each time frame. In a weekly time frame, the focus is on weekly scheduling for periods of one to several days. The system is scheduled for a specified horizon using a typical time granularity of one hour with the objective to minimize the operating costs. The operators perform a unit-commitment schedule to determine which unit to start up and shut down and at what times. In the economic dispatch time frame, the controllable units are ramped up or down so as to follow the load shape. The units are dispatched to ensure that the production costs are optimal. In the regulation time frame of seconds to a few minutes, the system operators use the so-called Automatic Generation Control (AGC) to have the units track second-by-second the demand to drive the deviations within an acceptable range. As wind units are not controllable or dispatchable, their presence complicates the carrying out of these scheduling, dispatch and control tasks by the operators.

In system planning, all the key considerations and limitations in power system operations need to be taken into account when longer-term studies are conducted for purposes such as the determination of which investments in generation and transmission to undertake or the assessment of the impacts of a new policy. In addition, system

planners must also ensure that there is adequate supply-demand balance for the planning horizon.

When wind resources are integrated into the grid, the variability and randomness of the wind energy are added to those already contained in the load. If only a small fraction of the demand is served by the wind turbines, the additional variability is limited and therefore may be approximated without introducing too many errors. However, as the wind power penetration increases, the additional source of variability and randomness needs to be represented in adequate detail to capture their impacts on system operations. Indeed, as more and more electric power is produced by the wind resources, system operations need to be modified so that the system has the ability to compensate for sudden and unexpected changes in the load and in the wind power production.

With the growing wind power penetration, system operators therefore need effective tools to manage the randomness and variability in the wind energy. We discuss in the next section the methodologies that the operators use.

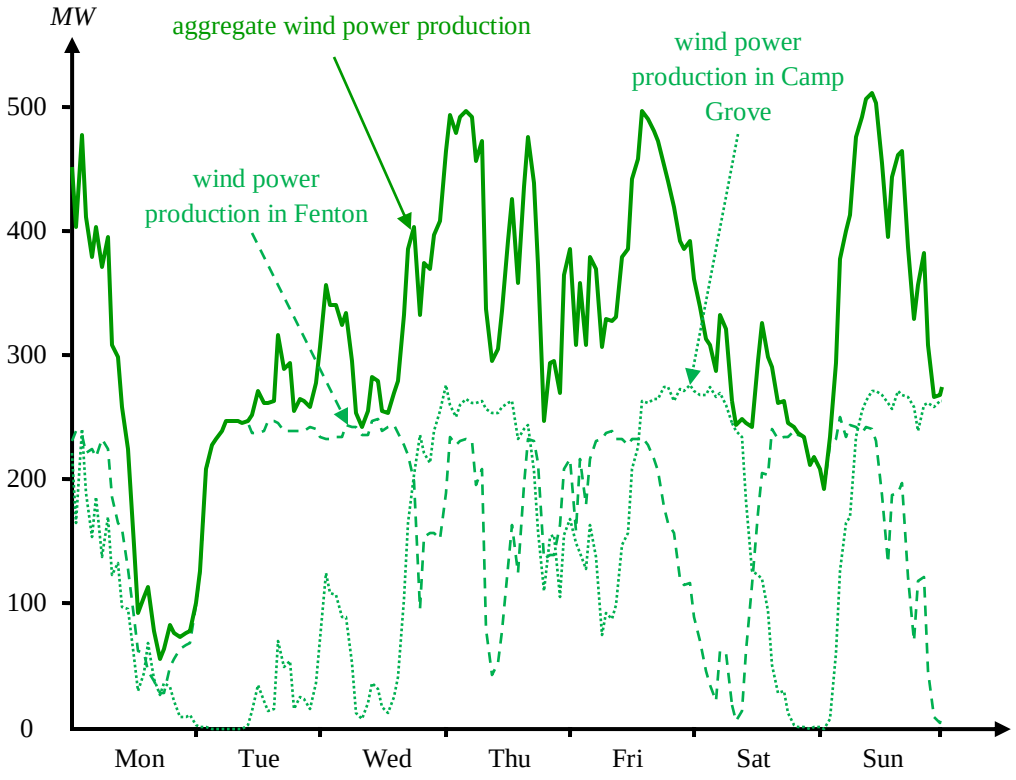
2.4 Management of the Wind Energy Variability and Intermittency

Conceptually, the most effective and reliable way to harness wind energy is to use storage devices [31]. The following is the basic idea. At times when the wind energy production is higher than forecasted, the storage device is charged with the surplus of energy. In contrast, when the system is experiencing a lack of supply – which may be due to insufficient wind energy production – the storage device supplies the demand from its stored energy. In other words, due to the presence of storage, there is no need to keep as close a watch on the load and the wind energy variations; reliance on the storage device

therefore facilitates the integration of variable and intermittent resources into the grid. The techniques for energy storage include batteries, pumped hydro, compressed air, flywheel and hydrogen. The determination of the appropriate technology and size depends on a number of factors including the response characteristics and storage capability of the device. The principal limitation is that large-scale storage capability is not economically available today.

Another technique to effectively manage the variability and intermittency of the wind power production is to disperse geographically the wind farm locations, as opposed to the concentration of the generation sources in a single location [32], [33]. Since the wind does not necessarily blow with the same speed simultaneously at two distinct locations, whenever the wind speed is low at one of the two locations, resulting in a low wind power output, at the other location the higher wind speeds may produce larger output. Indeed, the location diversity uses the notions of portfolio theory to result in a smoother and a less variable energy production from wind resources.

We illustrate the location diversity by considering the two wind farms located at Camp Grove, IL, and Fenton, MN. We arbitrarily select the period of January 9-15, 2006, and plot in Fig. 2.7 the wind power production at the two locations. We add a third plot in Fig. 2.7 of the aggregate wind power production – the sum of the wind power produced at the two sites. We note that the variations of the aggregate wind power production are less pronounced and therefore smoother than those of the outputs at the individual locations. In this way, the smoothened aggregate output levels reduce substantially the intermittency effect.

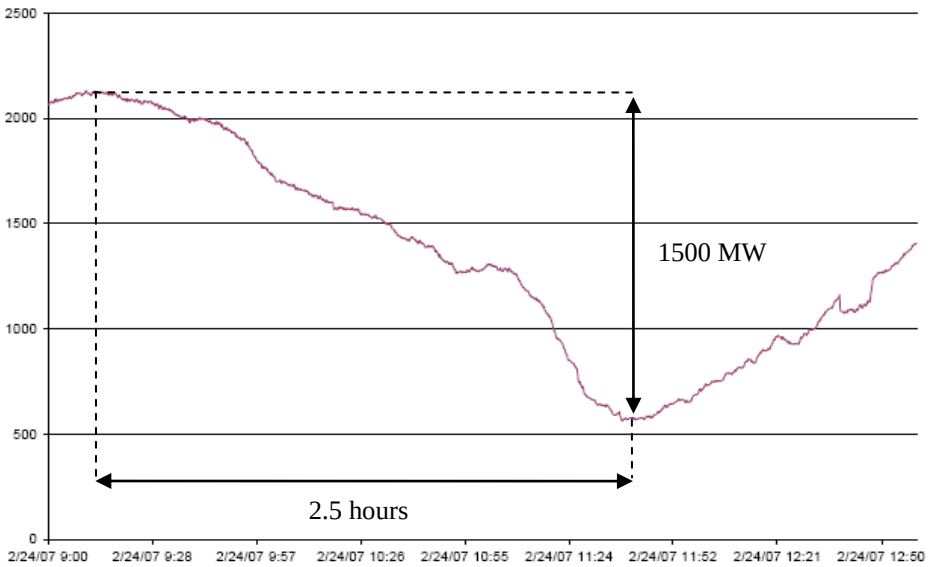


Source: ftp://ftp2.nrel.gov/pub/ewits/TimeSeries/LandBased/2006
(Files: SITE_4022_OUTPUT_V4_2006.CSV and SITE_0251_OUTPUT_V4_2006.CSV)

Figure 2.7: The hourly chronological wind power production in Camp Grove, IL, and Fenton, MN, for the weekly period of Monday, January 9 to Sunday, January 15, 2006.

A very vivid example of the benefits of diversification is the reliability event that occurred in Texas on February, 24, 2007. The Texas installed wind capacity of about 2,900 MW was operating with an output of approximately 2,000 MW. At some point later in the day, the system experienced a sudden wind speed increase leading to the shutdown of some wind turbines. For instance, a 200-MW wind farm had its output reduced by 75% to 50 MW over a short period of 11 minutes. In Fig. 2.8, we show the aggregated power output of all the ERCOT wind farms as a function of time. Although the total decrease in the wind power production amounting to 1,500 MW is high, the ramp rate of the decrease is rather low. While the aggregate wind production dropped by

75% from its peak output level, the reduction took around 13 times longer – approximately 2.5 hours – than the power drop of a single 200-MW wind farm. The conclusion is that the ramp rate of such an event remains acceptable: the diversity of the wind farm locations resulted in a markedly smoother decrease in the wind power production [34].



Source: http://www.isorto.org/atf/cf/%7B5B4E85C6-7EAC-40A0-8DC3-003829518EBD%7D/IRC_Renewables_Report_101607_final.pdf

Figure 2.8: Total wind power production in Texas during the 02/24/2007 event.

In the absence of location diversity and storage, the operators typically compensate for and manage the intermittency and the variability of the wind energy production by increasing the reserve levels. The increased levels are determined to compensate for the increased variability and intermittency effects of wind generation. As there are several reserve services – categorized, typically, by response times – the changed settings must be determined approximately for each reserve service. The

determination of the new values must be done to appropriately compensate for the different variability and random effects in the wind power production. For instance, the units which provide regulation and load following service must be able to follow the simultaneous shorter-term variations of the load and the wind power production. Also, in the determination of the unit commitment, the operators use the wind power forecast to determine the scheduling of the set of controllable units. As there is an unavoidable difference between the forecasted and the actual wind power production, the system must be able to make up the difference at each instant and without interrupting the supply to the loads. To meet this objective, the commitment calls for a higher percentage of reserves in the presence of wind generation. The uncertainty due to wind production is difficult in the setting of the reserve levels so as to avoid the problem of too high or too low reserves. If the reserves are set overly high when the actual wind production is at or above the forecasted levels, there are costs incurred due to the uneconomic use of controllable resources. On the other hand, if the actual wind power production is below the forecasted levels and the reserves scheduled are insufficient, the system may be unable to meet the load resulting in outages and the incurment of outage costs.

So, the difficulty arises in terms of setting the appropriate levels of reserves to ensure reliable operations. The increase in the reserve levels incurs additional unit commitment and dispatch costs. Several studies have been performed to address such an issue and a good summary of the results is given in [35]. The reserve settings presented in the studies are very dependent on characteristics of the power system under consideration and those of its controllable resource.

In this chapter, we have qualitatively described the nature of the wind speed and wind energy variability. We illustrated their variability with concrete examples, making use of historical data. We also explained the impacts of the wind power production variability, lack of controllability and predictability on the system operations and planning. To more effectively integrate wind resources, we indicated the techniques that must be implemented by system operators: the use of storage devices, the geographical diversification of the wind arm locations and the increase in the reserve levels. We make detailed use of the insights we gained in the present chapter in the development of wind speed and energy models that capture the uncertainty and the variability characteristics. We devote the next chapter to these models.

3. WIND SPEED AND RESOURCE MODELING

The analysis of the longer-term impacts of wind resource integration into the grid requires a wind energy model with the capability to represent the wind speed and energy with the appropriate level of detail. The objective of this chapter is to develop such a model to explicitly represent the variability and intermittency characteristics of the wind resources in the probabilistic simulation framework.

This chapter contains four sections. In Section 3.1, we focus on the wind speed modeling with its explicit probabilistic characterization. We introduce the notion of wind speed *regimes* to capture the salient wind patterns in multiple geographical areas. We identify the regimes using historical wind data by grouping together the days with similar wind speed patterns and determine their associated probabilities. We devote Section 3.2 to the testing of the proposed model for different locations and with data sets of varying length. In Section 3.3, we extend the mathematical formulation of the wind turbines' power curve to represent the power output from a wind farm and construct a probabilistic model of its energy output making explicit use of the regime characterization. In the last section, we employ this probabilistic model to determine the wind power production of a power system with wind farms installed at multiple locations.

3.1 Wind Speed Model

The diurnal pattern of power produced by wind farms impacts the scheduling of the controllable resources. Indeed, if the wind energy production is high during the high-load periods, the wind resource can displace one or more peaking – and, typically, costly

– controllable units. On the other hand, if the wind energy production tends to be significant during the low-load periods, the system operators need to lower the output of the base loaded – and not-so-flexible – units. As the shape of the daily wind power production depends directly on the diurnal wind speed pattern, the wind speed model must reflect this characteristic. Since several typical daily patterns may occur throughout the period under consideration, such a phenomenon must be taken into account. To identify the various wind speed regimes, we group the days whose wind speed patterns have similar "shapes" into a class. We identify and characterize all such classes analytically. In our work, we use and compare two well-known clustering methods – the hierarchical clustering and the *k*-means algorithms [36].

We partition a day into H non-overlapping subperiods to analyze the wind speed data. The number of subperiods depends on the level of resolution, the objectives, and the nature of the study. The selected resolution determines the wind speed data sampling and the associated wind output representation. We assume that for each subperiod the wind speed is constant over the subperiod with the value observed at the measurement point. As a result, the subperiod duration is the smallest indecomposable unit of time and no phenomena with shorter durations may be represented in the model we construct.

We first introduce the notation to describe the wind speed data storage. Let D be the number of days in the study period for wind speed data collection, e.g., $D = 365$ for a one-year study, or $D = 90$ for a single season study. We assume D to be sufficiently large to allow a meaningful statistical analysis of the data set. We consider the wind farms to be located at S distinct sites. We define

$$\mathcal{S} \triangleq \{ s : s = 1, 2, \ldots, S \} \tag{3.1}$$

to be the wind farm location index set. Clearly, $|\mathcal{S}| = S$.

Let $v_{s,d}^{(h)}$ denote the wind speed at site s on the day d for the subperiod h, with $s \in \mathcal{S}$, $d = 1, 2, \ldots, D$ and $h = 1, 2, \ldots, H$. We define the wind speed vector at site s on day d by:

$$\underline{v}_{s,d} \triangleq \begin{bmatrix} v_{s,d}^{(1)} \\ v_{s,d}^{(2)} \\ \vdots \\ v_{s,d}^{(H)} \end{bmatrix} \in \mathbb{R}^{H} . \tag{3.2}$$

We also define the super vector of the day d wind speed vectors as:

$$\underline{v}_{d} \triangleq \begin{bmatrix} \underline{v}_{1,d} \\ \underline{v}_{2,d} \\ \vdots \\ \underline{v}_{S,d} \end{bmatrix} \in \mathbb{R}^{(S)(H)} . \tag{3.3}$$

We then construct the matrix $\underline{V} \in \mathbb{R}^{(S)(H)\times D}$ of the wind speeds at the S locations on the D days:

$$\underline{V} \triangleq \left[\underline{v}_1 \vdots \ \underline{v}_2 \vdots \ \cdots \ \vdots \underline{v}_D \right] = \begin{bmatrix} \underline{v}_{1,1} & \underline{v}_{1,2} & \cdots & \underline{v}_{1,d} & \cdots & \cdots & \underline{v}_{1,D} \\ \underline{v}_{2,1} & \underline{v}_{2,2} & \cdots & \underline{v}_{2,d} & \cdots & \cdots & \underline{v}_{2,D} \\ \vdots & \vdots & & \vdots & & & \vdots \\ \vdots & \vdots & & \vdots & & & \vdots \\ \underline{v}_{s,1} & \underline{v}_{s,2} & \cdots & \underline{v}_{s,d} & \cdots & \cdots & \underline{v}_{s,D} \\ \vdots & \vdots & & \vdots & & & \vdots \\ \underline{v}_{S,1} & \underline{v}_{S,2} & \cdots & \underline{v}_{S,d} & \cdots & \cdots & \underline{v}_{S,D} \end{bmatrix}. \tag{3.4}$$

For instance, for the single location case, i.e., $|\mathcal{S}| = 1$, $\underline{V}$ is the $H \times D$ matrix whose entry in position (h, d) is the wind speed for the subperiod h and day d so that:

$$\underline{V} \triangleq \left[\, \underline{v}_1 \vdots \underline{v}_2 \vdots \cdots \vdots \underline{v}_D \,\right] = \begin{bmatrix} v_1^{(1)} & v_2^{(1)} & \cdots & v_d^{(1)} & \cdots & \cdots & v_D^{(1)} \\ v_1^{(2)} & v_2^{(2)} & \cdots & v_d^{(2)} & \cdots & \cdots & v_D^{(2)} \\ \vdots & \vdots & & \vdots & & & \vdots \\ \vdots & \vdots & & \vdots & & & \vdots \\ v_1^{(h)} & v_2^{(h)} & \cdots & v_d^{(h)} & \cdots & \cdots & v_D^{(h)} \\ \vdots & \vdots & & \vdots & & & \vdots \\ v_1^{(H)} & v_2^{(H)} & \cdots & v_d^{(H)} & \cdots & \cdots & v_D^{(H)} \end{bmatrix}, \tag{3.5}$$

where we suppress the notation for the site s since only a single site is under consideration.

We start the discussion of the clustering schemes for the wind speed data to classify the wind speed patterns with the description of the steps of the hierarchical clustering algorithm. For the set of D days, we have initially as many classes as we have days, i.e., D classes of wind speed vectors. Each class is simply a singleton set consisting of a day's wind speed vector. The basic idea is to successively reduce the number of classes from D to $k \ll D$ by iteratively combining two classes in an iteration so as to shrink the number of classes by one in each iteration. $\mathcal{R}_j^{(u)}$ denotes the class j in iteration u. For the initial iteration, $u = 0$, $j = 1, 2, \ldots, D$. For iterations $u > 0$, $j = 1, 2, \ldots, D-u$. For $u = D-k$, there are k resulting classes $\mathcal{R}_1^{(D-k)}, \mathcal{R}_2^{(D-k)}, \ldots, \mathcal{R}_k^{(D-k)}$. In principle, some of the k classes may still be a singleton set, but, in practice, each class contains several wind speed vectors.

The reduction of the number of classes requires the use of the "closeness" metric μ. The metric μ that we use for measuring the "closeness" between any two daily wind speed vectors $\underline{v}_d$ and $\underline{v}_{d'}$ is expressed as:

$$\mu\left(\underline{v}_d, \underline{v}_{d'} \right) = \left\| \underline{v}_d - \underline{v}_{d'} \right\|, \tag{3.6}$$

where we use $\| \cdot \|$ to denote the ℓ_2 norm. Note that by definition

$$\mu\left(\underline{v}_d, \underline{v}_{d'}\right) = \mu\left(\underline{v}_{d'}, \underline{v}_d\right), \tag{3.7}$$

and

$$\mu\left(\underline{v}_d, \underline{v}_d\right) = 0 . \tag{3.8}$$

We generalize the metric μ to measure the "closeness" between any two arbitrary classes $\mathscr{X}$ and $\mathscr{Y}$ of daily wind speed vectors by defining:

$$\mu\left(\mathscr{X}, \mathscr{Y} \right) \triangleq \max\left\{ \mu\left(\underline{x}, \underline{y} \right) : \underline{x} \in \mathscr{X}, \ \underline{y} \in \mathscr{Y} \right\}^{5} . \tag{3.9}$$

It follows that the "closeness" between two classes is expressed in terms of the "closeness" between the pair of the two least "close" elements in the two sets: this metric provides the upper bound on the "closeness" of any pair of elements of the two groups based on the measure of the most distant pair consisting of one element from group $\mathscr{X}$ and the other element from group $\mathscr{Y}$.

We start with the D single-element groups, with each element corresponding to $d \in \{1, 2, \ldots, D\}$. In other words, there are D groups and each group consists of the wind speed vector $\underline{v}_d$ for day d:

$$\mathscr{R}_d^{(0)} \triangleq \left\{ \underline{v}_d \right\}, \ d \in \{1, 2, \ldots, D\} . \tag{3.10}$$

The merger criterion of joining two groups with no elements in common is based on the "closeness" of the two groups and so is expressed in terms of the metric μ. We need to

[5] The metric μ is called the complete-link measure in the clustering literature [36].

compute and store the values of μ for each pair of the D vectors so as to determine the "closeness" between any two groups $\mathcal{R}_d^{(0)}$, $d \in \{1, 2, \ldots, D\}$. Let $\underline{Z}^{(0)}$ be the matrix where these values are stored in iteration $u = 0$. $\underline{Z}^{(0)}$ is a $D \times D$ matrix with the entry in position (i, j) corresponding to the "closeness" measure between $\underline{v}_i$ and $\underline{v}_j$. Since μ is a symmetric measure as stated by Equation (3.7), $\underline{Z}^{(u)}$ is a symmetric matrix with all diagonal elements with value 0 due to the property stated in Equation (3.8). Our interest is in the upper-triangular portion only, and the portion of the matrix consisting of the diagonal and the below diagonal elements need not be computed. For $\underline{Z}^{(0)}$, we need therefore $\frac{D.(D-1)}{2}$ values of μ to compute the upper-triangular portion $\underline{\Delta Z}^{(0)}$ of the matrix $\underline{Z}^{(0)}$. We find the smallest element in $\underline{\Delta Z}^{(0)}$ so as to merge the two "closest" classes. Two distinct cases may arise:

- Unique solution: A unique element $z_{x,y}$ in $\underline{\Delta Z}^{(0)}$ exists with the property that

$$z_{x,y} = min\left\{ \underline{\Delta Z}^{(0)} \right\} \leq z_{x',y'} \quad \forall z_{x',y'} \in \underline{\Delta Z}^{(0)} \tag{3.11}$$

  and we merge the two classes $\mathcal{R}_x^{(0)}$ and $\mathcal{R}_y^{(0)}$; without loss of generality, we assume that $x < y$, and then we construct

$$\mathcal{R}_x^{(1)} = \mathcal{R}_x^{(0)} \cup \mathcal{R}_y^{(0)} \tag{3.12}$$

  and update all the indices of the resulting classes so that we end up with the set of classes $\left\{ \mathcal{R}_1^{(1)}, \mathcal{R}_2^{(1)}, \ldots, \mathcal{R}_{D-1}^{(1)} \right\}$:

$$\begin{cases} \mathcal{R}_j^{(1)} = \mathcal{R}_j^{(0)}, & j = 1, 2, \ldots, y-1, \ j \neq x \\ \mathcal{R}_j^{(1)} = \mathcal{R}_{j+1}^{(0)}, & j = y, \ldots, D-1 \end{cases} . \tag{3.13}$$

- Multiple solutions: Two or more pairs of classes have equal values of the μ metric and so are eligible for merging; we arbitrarily choose one of these pairs and merge its two classes: we make the selection of (x, y) unique by adopting a simple rule such as selecting the pair with the smallest index and whose sum of the two indices is also smaller than that of any other pair; we construct the merged subset using Equation (3.12) and upgrade the other classes using Equation (3.13).

We repeat the same procedure on this set of $D-1$ classes and continue with the process until we have k classes. After the last iteration, we determine $\mathcal{R}_r$, $r = 1, 2, \ldots, k$, where:

$$\mathcal{R}_r = \mathcal{R}_r^{(D-k)}, \; r = 1, 2, \ldots, k. \tag{3.14}$$

We may view the hierarchical clustering algorithm in terms of an inverted pyramid structure shown in Fig. 3.1 with the number of classes reduced by one in each successive iteration.

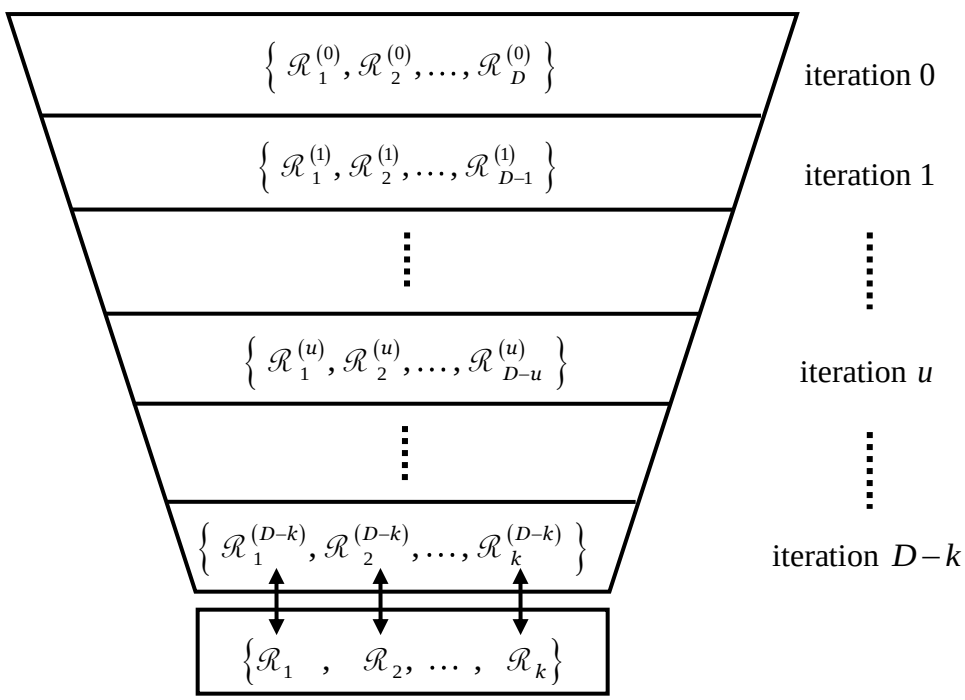


Figure 3.1: Inverted pyramid structure of the hierarchical clustering algorithm.

We now describe the *k*-means method. We start out with k typical daily wind speed vectors, the so-called class centers, and denote the set of the class centers by $\left\{ \underline{a}_j^{(0)} \right\}$, $j = 1, 2, \ldots, k$. Then, to each vector $\underline{a}_j^{(0)}$ is associated a class $\mathcal{R}'^{(0)}_j$ composed of the historical daily wind speed vectors for which $\underline{a}_j^{(0)}$ is the "closest" center with respect to the measure μ:

$$\mathcal{R}'^{(0)}_j = \left\{ \underline{v}_d : \mu\left(\underline{v}_d, \underline{a}_j^{(0)} \right) \leq \mu\left(\underline{v}_d, \underline{a}_{j'}^{(0)} \right), \; j' = 1, 2, \ldots, k, \; j' \neq j \right\}. \quad (3.15)$$

So, we have an initial partitioning of the data: $\left\{ \mathcal{R}'^{(0)}_1, \mathcal{R}'^{(0)}_2, \ldots, \mathcal{R}'^{(0)}_k \right\}$. From this initial partition, the center – or mean daily wind speed vector – of each class is re-computed and we obtain a new set of centers $\left\{ \underline{a}_j^{(1)} \right\}$, $j = 1, 2, \ldots, k$ with:

$$\underline{a}_j^{(1)} = \frac{1}{\left| \mathcal{R}'^{(0)}_j \right|} \sum_{\underline{v}_d \in \mathcal{R}'^{(0)}_j} \underline{v}_d \quad . \quad (3.16)$$

We then repeat the classification process to identify the days associated with the centers $\underline{a}_j^{(1)}$, $j = 1, 2, \ldots, k$, using the "closeness" metric:

$$\mathcal{R}'^{(1)}_j = \left\{ \underline{v}^{(d)} : \mu\left(\underline{v}^{(d)}, \underline{a}_j^{(1)} \right) \leq \mu\left(\underline{v}^{(d)}, \underline{a}_{j'}^{(1)} \right), \; j' = 1, 2, \ldots, k, \; j' \neq j \right\}. \quad (3.17)$$

In each successive iteration $u > 1$, we update the centers $\left\{ \underline{a}_j^{(u)} \right\}$ and the classification of the days in the k groups $\mathcal{R}'^{(u)}_j$, $j = 1, 2, \ldots, k$. The iterations stop at some iteration $\bar{u}$ with:

$$\mathcal{R}'^{(\bar{u})}_j = \mathcal{R}'^{(\bar{u}-1)}_j, \; j = 1, 2, \ldots, k; \quad (3.18)$$

i.e., there is no change in the composition of the k classes of days of wind speed from iteration $\bar{u} - 1$ to iteration $\bar{u}$. Note that whenever a group $\mathcal{R}'^{(u)}_j$ becomes empty for

some u, we need to restart the algorithm with a different set of initial class centers [36]. The outcome of the *k*-means scheme is the k classes $\mathcal{R}'^{(\bar{u})}_j$, $j = 1, 2, \ldots, k$ and we use those to construct the classes $\mathcal{R}'_1, \mathcal{R}'_2, \ldots, \mathcal{R}'_k$. The details of the two algorithms are given in Appendix A.

Once we have identified the k classes composed of the similar days of wind speed data – using either the hierarchical clustering or the *k*-means scheme – we introduce and characterize the notion of wind speed regimes. Without loss of generality, we consider the final classes obtained with the hierarchical clustering algorithm. We may view each class $\mathcal{R}_r$, $r = 1, 2, \ldots, k$, to be the realizations of a set of random variables. We define for site s and subperiod h the random variable (rv) $\underset{\sim}{V}^{(h)}_{s,r}$, and we use the elements of $\mathcal{R}_r$ to determine the density functions of each $\underset{\sim}{V}^{(h)}_{s,r}$, $s = 1, 2, \ldots, S$, $h = 1, 2, \ldots, H$ in the set. Therefore it follows that the mean value $m^{(h)}_{s,r}$ of each r.v. $\underset{\sim}{V}^{(h)}_{s,r}$ may be estimated by:

$$\hat{m}^{(h)}_{s,r} = \frac{1}{\left|\mathcal{R}_r\right|} \sum_{\underline{v}_i \in \mathcal{R}_r} v^{(h)}_{s,i} \quad . \tag{3.19}$$

We then construct the vector

$$\underline{\hat{m}}_{s,r} = \begin{bmatrix} \hat{m}^{(1)}_{s,r} \\ \hat{m}^{(2)}_{s,r} \\ \vdots \\ \hat{m}^{(H)}_{s,r} \end{bmatrix}. \tag{3.20}$$

The vector $\hat{\underline{m}}_{s,r}$ therefore may be viewed as a representation of the average daily wind speed pattern for the class $\mathcal{R}_r$. We can use similar estimators for the higher moments of $\underset{\sim}{V}_{s,r}^{(h)}$ and also to estimate the cdf $\hat{F}_{\underset{\sim}{V}_{s,r}^{(h)}}(\cdot)$ and the pdf $\hat{f}_{\underset{\sim}{V}_{s,r}^{(h)}}(\cdot)$.

The probability α_r we associate with class $\mathcal{R}_r$ is determined from the fact that the class $\mathcal{R}_r$ occurs $|\mathcal{R}_r|$ days out of D days in the data set. We estimate α_r using:

$$\hat{\alpha}_r = \frac{|\mathcal{R}_r|}{D}. \tag{3.21}$$

We refer to the doublet of the estimates of the cdf of the wind speed rv's and the probability of class $\mathcal{R}_r$ by the term regime, which we denote by $\mathcal{R}_r$. We adopt the short-hand notation

$$\mathcal{R}_r \leftrightarrow \left\{ \left\{ \hat{\underline{m}}_{s,r} \right\}_{s=1,2,\ldots,S}, \ \hat{\alpha}_r \right\} \tag{3.22}$$

to denote the regime $\mathcal{R}_r$.

We may note that the wind speed regime model stated in Equations (3.19)-(3.22) uses the partitioning of each day into H subperiods. For concreteness we adopt an hourly resolution and so $H = 24$. So, we have the 24 hourly wind speeds for $h = 1, 2, \ldots, 24$.

In this section, we have described the two clustering schemes to group together the days with similar wind speed patterns. Clearly, they have completely different structures and must yield different classes. In the next section, we test the two clustering algorithms and exhibit their differences.

3.2 Application of Clustering for Regime Determination

In this section, we illustrate the application of the two clustering schemes described in Section 3.1 to identify the wind speed regimes at various US locations using historical wind data. We use parametric variations in the testing work, discuss specific findings and draw conclusions from the results. We provide representative results illustrative of the type of results we obtained in the extensive testing that we performed.

We focus our analysis on the determination of a small number of wind speed regimes so as to capture the average wind speed variability over the longer term. We choose a small number of regimes so as to avoid the typical problems that arise with a higher number of regimes due to the possibility of rare and unusual regimes that may result. Our studies use the wind data for selected sites within the MISO and PJM systems where wind farms are located or are likely to be built in the next few years. The locations of the existing wind farms in the US that are currently in operation are well documented [37]. Figures 3.2 and 3.3 show the sites in the MISO and the PJM system geographic footprints, respectively.

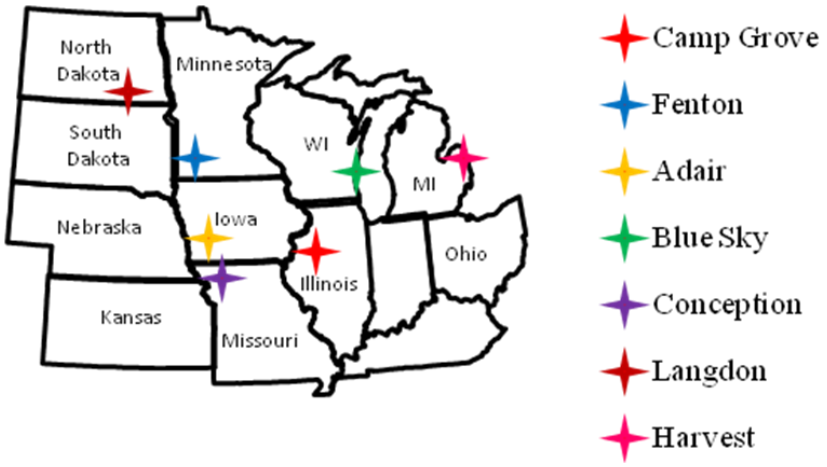


Figure 3.2: The selected sites located within the MISO footprint.

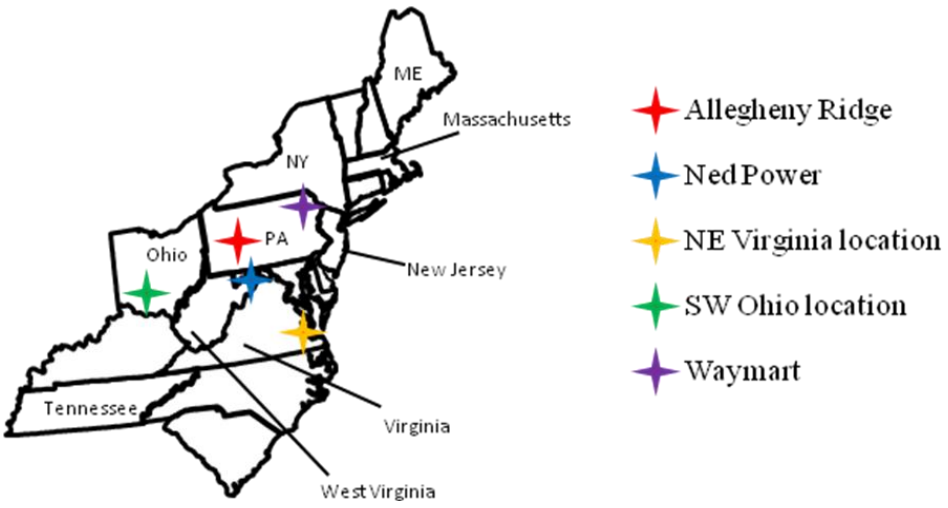


Figure 3.3: The selected sites located within the PJM footprint.

The objective is to make a comparative performance of the two clustering schemes, assess the impacts of the length of the data set and discuss the overall utility of the results. The studies use the wind data collected for the 2004-2006 period by NREL [27].

Example 3.1: We start by considering the wind speed data to determine the wind speed regimes at the location Fenton, MN, using the year 2006 measurements. We apply the two algorithms to identify three regimes. The daily patterns of the hourly averages for the three regimes obtained with the two schemes are shown in Fig. 3.4. The tabulation of the number of days that belong to each regime identified by the two schemes and the corresponding fraction of time are presented in Table 3.1.

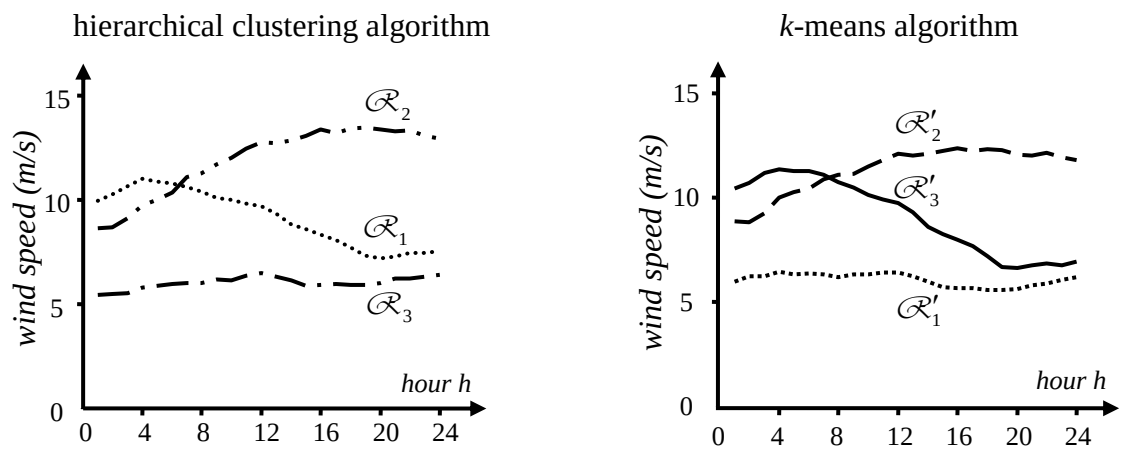


Figure 3.4: The three regimes identified by the two clustering algorithms for the Fenton location, MN, using the 2006 year hourly data.

Table 3.1: Number of occurrences of the three regimes identified by the two clustering algorithms for the Fenton location, MN, using the 2006 year hourly data

hierarchical clustering algorithm

class	occurrence	
	number	fraction
$\mathcal{R}_1$	177	0.48
$\mathcal{R}_2$	58	0.16
$\mathcal{R}_3$	130	0.36

k-means algorithm

class	occurrence	
	number	fraction
$\mathcal{R}'_1$	151	0.41
$\mathcal{R}'_2$	104	0.29
$\mathcal{R}'_3$	110	0.30

We probabilistically characterize the wind speed for regime $\mathcal{R}'_3$ and hour 12 of the day. We collect the wind speed data [27] to construct the set:

$$\mathcal{V}_3^{(12)} = \left\{ v_d^{(12)} : \underline{v}_d \in \mathcal{R}'_3 \right\}, \tag{3.23}$$

where $\underline{v}_d = \left[v_d^{(1)}, \ldots, v_d^{(12)}, \ldots, v_d^{(24)} \right]^T$. We recall that the class $\mathcal{R}'_3$ of wind speed vectors is composed of 110 days of wind speed. So, the set $\mathcal{V}_3^{(12)}$ is constituted of 110 wind speed measurements. We make use of such a set of data

samples to estimate the cdf and pdf of the wind speed rv $\underset{\sim}{V}_3^{(12)}$ and we plot such functions in Fig. 3.5 and Fig. 3.6 respectively. We repeat the same procedure for the three other regimes and the 23 remaining hours of the day.

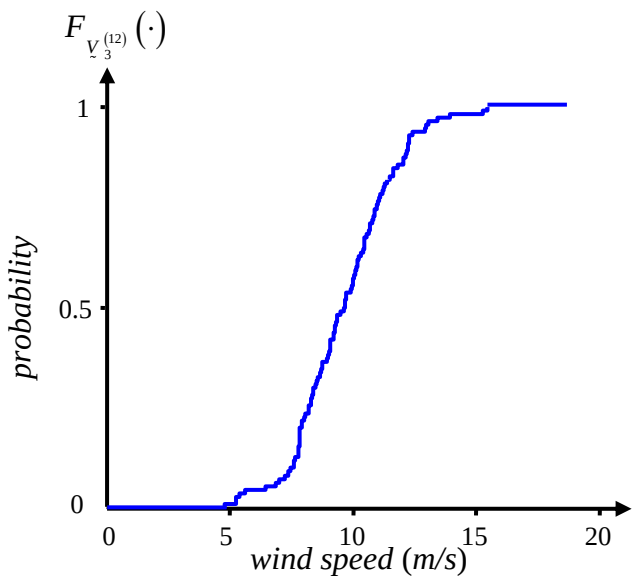


Figure 3.5: The estimated cdf of the random variable $\underset{\sim}{V}_3^{(12)}$ representing the wind speed at Fenton, MN, for regime $\mathscr{R}_3'$ and hour 12.

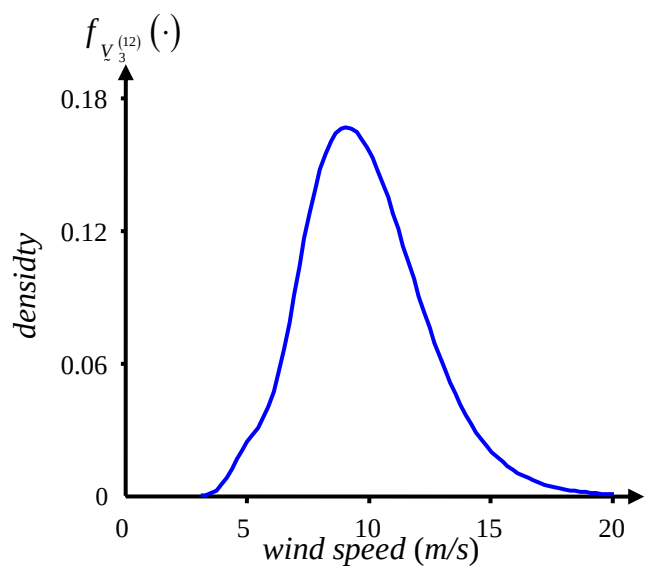


Figure 3.6: The estimated pdf of the random variable $\underset{\sim}{V}_3^{(12)}$ representing the wind speed at Fenton, MN, for regime $\mathscr{R}_3'$ and hour 12.

We compare the different average wind speed patterns obtained by the two algorithms so as to assess the similarity of the two sets of results. We use the "closeness" metric μ to measure the distance between the wind speed patterns of the regimes $\mathcal{R}_i$ and $\mathcal{R}'_j$: $\underline{m}_{s,i}$ and $\underline{m}'_{s,j}$, with $i = 1, 2, 3$ and $j = 1, 2, 3$:

$$\mu\left(\underline{m}_{s,i}, \underline{m}'_{s,j}\right) = \left\| \underline{m}_{s,i} - \underline{m}'_{s,j} \right\|, \; i, j = 1, 2, 3 \quad (3.24)$$

The values of the "closeness" metrics between the two sets of patterns are given in Table 3.2. We remark that $\mathcal{R}'_3$ is the closest to $\mathcal{R}_1$ and the other patterns are more "dissimilar" because $\mathcal{R}'_2$ is less close to $\mathcal{R}_1$ and so is $\mathcal{R}'_1$. Similarly, we observe that $\mathcal{R}'_2$ is the closest to $\mathcal{R}_2$ and $\mathcal{R}'_1$ is the closest to $\mathcal{R}_3$. Clearly, we establish in this case a one-to-one correspondence between the patterns obtained with the two algorithms and therefore conclude that the clustering results obtained under the two schemes are not measurably different.

Table 3.2: Values of the "closeness" metric to determine the similarity between the regimes obtained with the two clustering algorithms for the Fenton location, MN, using the 2006 year hourly data

regime		hierarchical clustering		
		$\mathcal{R}_1$	$\mathcal{R}_2$	$\mathcal{R}_3$
k-means clustering	$\mathcal{R}'_1$	16.01	29.99	1.81
	$\mathcal{R}'_2$	15.38	3.93	26.35
	$\mathcal{R}'_3$	2.22	21.03	17.59

Example 3.2: We perform a similar analysis for the Camp Grove, IL, location making use of the 2006 data set. For this location, we determine four typical wind

speed regimes. In Fig. 3.7, we plot the regimes' patterns identified by the two clustering schemes and we give their frequency of occurrence in Table 3.3.

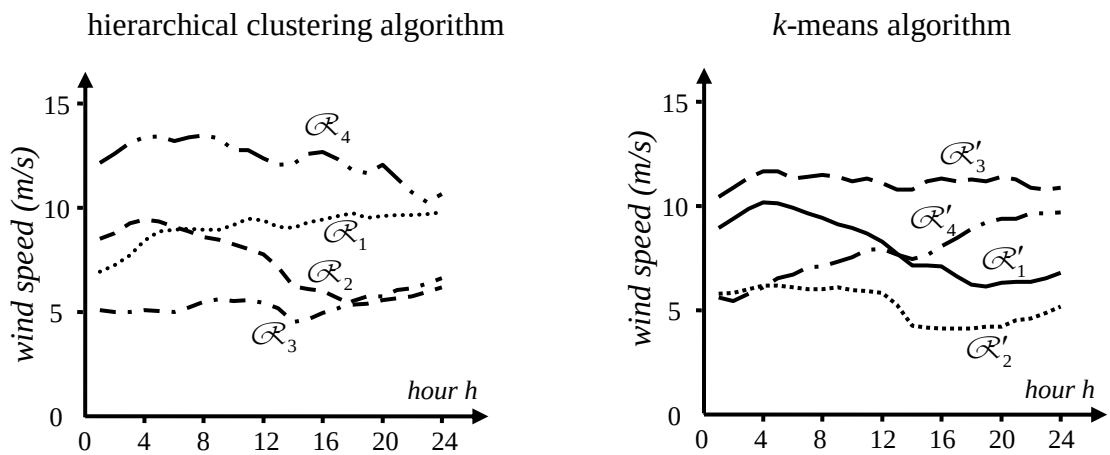


Figure 3.7: The four regimes identified by the two clustering algorithms for the Camp Grove location, IL, using the 2006 year hourly data.

Table 3.3: Number of occurrences of the four regimes identified by the two clustering algorithms for the Camp Grove location, IL, using the 2006 year hourly data

hierarchical clustering algorithm

class	occurrence	
	number	fraction
$\mathcal{R}_1$	78	0.21
$\mathcal{R}_2$	166	0.46
$\mathcal{R}_3$	102	0.28
$\mathcal{R}_4$	19	0.05

k-means algorithm

class	occurrence	
	number	fraction
$\mathcal{R}'_1$	111	0.30
$\mathcal{R}'_2$	136	0.37
$\mathcal{R}'_3$	54	0.15
$\mathcal{R}'_4$	64	0.18

We compute the values of the "closeness" metric for the patterns obtained under the two clustering techniques and summarize the results in Table 3.4. For this case, the following correspondences between the wind speed regimes are observed:

$$\mathcal{R}_1 \sim \mathcal{R}'_4, \mathcal{R}_2 \sim \mathcal{R}'_1, \mathcal{R}_3 \sim \mathcal{R}'_2, \mathcal{R}_4 \sim \mathcal{R}'_3. \tag{3.25}$$

Table 3.4: Values of the "closeness" metric to determine the similarity between the regimes obtained with the two clustering algorithms for the Camp Grove location, IL, using the 2006 year hourly data

Regime		hierarchical clustering			
		$\mathcal{R}_1$	$\mathcal{R}_2$	$\mathcal{R}_3$	$\mathcal{R}_4$
k-means clustering	$\mathcal{R}'_1$	10.95	3.69	15.42	21.52
	$\mathcal{R}'_2$	19.84	10.78	5.05	35.05
	$\mathcal{R}'_3$	11.05	20.14	28.40	6.79
	$\mathcal{R}'_4$	7.26	13.30	12.51	24.63

Example 3.3: We next discuss a study with multiple locations. Specifically, we consider the following four sites located in the MISO geographical footprint: Camp Grove, IL, Fenton, ND, Adair, IA and Blue Sky, WI. We use the wind speed data at the four locations for the year 2006 to identify three regimes. We plot the average wind speed patterns for each regime and location in Fig. 3.8 and we provide the frequency of occurrence of each regime in Table 3.5. We also compute the relative distances between each wind speed pattern and list the results in Table 3.6. In this particular case, the similarity is less close and the correspondence between the two sets of regimes is not possible to establish. Indeed, $\mathcal{R}'_1 \sim \mathcal{R}_2$, but so is $\mathcal{R}'_3$ since $\mathcal{R}'_3 \sim \mathcal{R}_1$. Also, $\mathcal{R}'_2 \sim \mathcal{R}_2$ but the "closeness" metric value is worse than for the similarity between $\mathcal{R}'_1$ and $\mathcal{R}_2$. Due to the fact that we no longer have a unique one-to-one correspondence, we

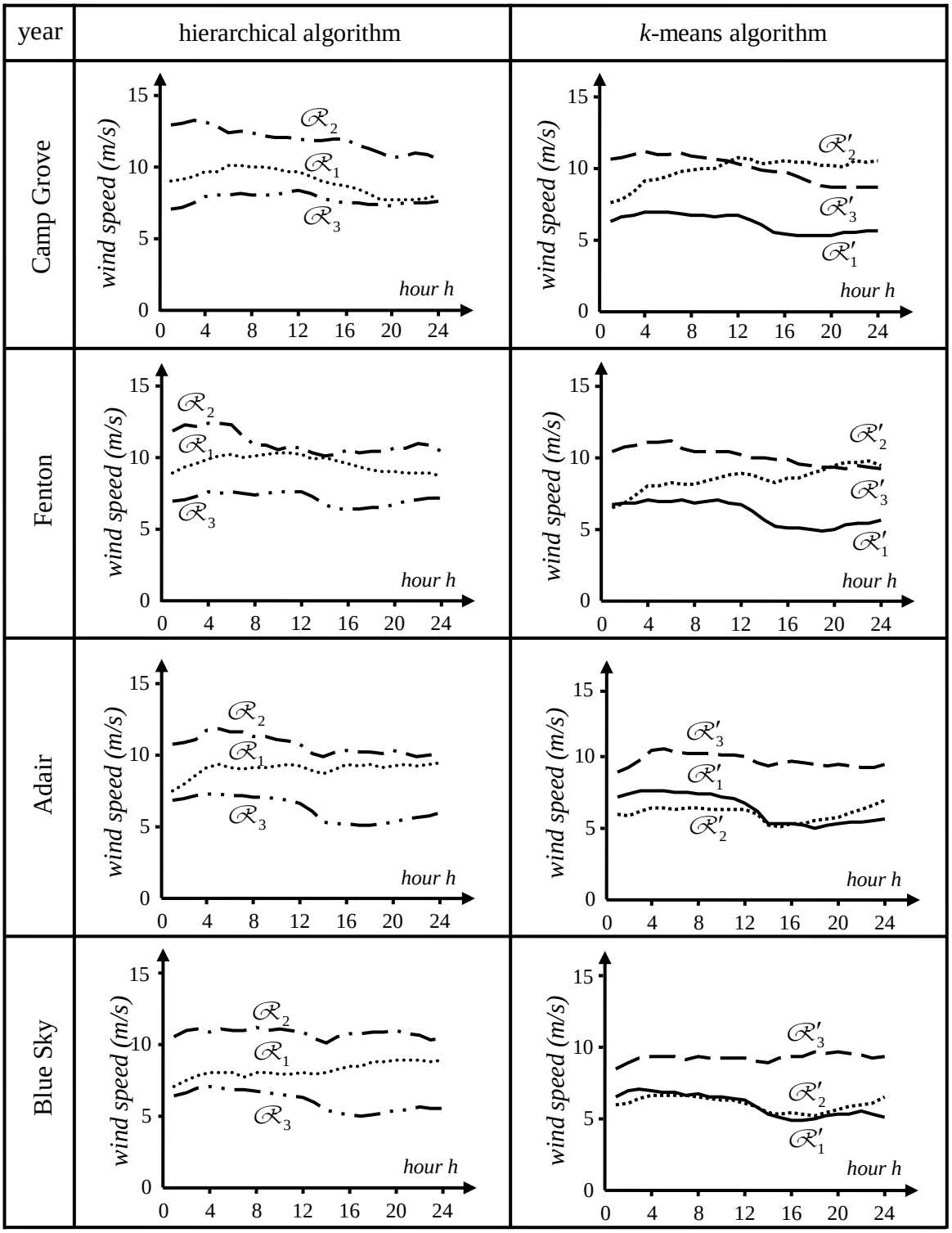


Figure 3.8: The three regimes identified by the two clustering algorithms for the Camp Grove, IL, Fenton, MN, Adair, IA and Blue Sky, WI, locations using the 2006 year hourly data.

can only state that the similarity of the patterns under the two clustering schemes is limited.

Table 3.5: Number of occurrences of the four regimes identified by the two clustering algorithms for the Camp Grove location, IL, using the 2006 year hourly data

hierarchical clustering algorithm

class	occurrence	
	number	fraction
$\mathcal{R}_1$	91	0.25
$\mathcal{R}_2$	233	0.64
$\mathcal{R}_3$	41	0.11

k-means algorithm

class	occurrence	
	number	fraction
$\mathcal{R}'_1$	140	0.38
$\mathcal{R}'_2$	99	0.27
$\mathcal{R}'_3$	126	0.35

Table 3.6: Values of the "closeness" metric to determine the similarity between the regimes obtained with the two clustering algorithms for the Camp Grove location, IL, using the 2006 year hourly data

regime		hierarchical clustering		
		$\mathcal{R}_1$	$\mathcal{R}_2$	$\mathcal{R}_3$
k-means clustering	$\mathcal{R}'_1$	29.24	9.63	49.04
	$\mathcal{R}'_2$	21.82	14.72	37.73
	$\mathcal{R}'_3$	9.16	30.48	14.09

From these three illustrative examples, we make the following findings:

- The wind speed patterns identified by the two clustering algorithms tend to present similarities; however, in some cases, the results differ and one-to-one correspondence cannot be established, in general.

- The sizes of the classes identified by each scheme are sufficiently large to be statistically meaningful; indeed, the smallest class we obtain in the one-year studies is 19 days as indicated in Table 3.3; therefore, care must be taken to ensure that the probability characterization is useful under such conditions. In all the other studies discussed here, the classes have sizes of 41 days or more, which is clearly a size sufficient for determining the probability distribution.
- In general, the *k*-means algorithm tends to produce classes with more uniform sizes than the hierarchical clustering algorithm: such a characteristic is evident in the results summarized in Tables 3.1, 3.3 and 3.5, and, in fact, in each of these three examples, the hierarchical clustering scheme yields a class whose size is larger (smaller) than the size of any of the classes obtained with the *k*-means algorithm.

In the extensive experimentation with the two clustering schemes, we found that a salient characteristic of hierarchical clustering is its sensitivity to outliers.

Example 3.4: We illustrate this characteristic with the data for the Allegheny Ridge location for the year 2006. We identify four classes of similar wind speed days using the two schemes. We present the results in Table 3.7. For the 2006 year data set, we observe that the hierarchical clustering algorithm results in the identification of a singleton class, $\mathcal{R}_1$, which is too small to hold probability information. Such a class is therefore meaningless and no regime can be associated to the class $\mathcal{R}_1$. In contrast, the classes $\mathcal{R}_2$, $\mathcal{R}_3$ and $\mathcal{R}_4$ are sufficiently large that we can associate relatively meaningful regimes to them. In

Fig. 3.9, we plot the average daily wind speed patterns for the four classes we identified. We notice that the day of wind speed constituting the class $\mathcal{R}_1$ has a strange behavior. Therefore, the outlier data must be ignored and the other classes remain unchanged. This characteristic of the hierarchical clustering algorithm for wind speeds is consistent with what has been previously reported in the statistics literature [38].

Table 3.7: Number of occurrences of the four regimes identified by the two clustering algorithms for the Allegheny Ridge location, PA, using the 2006 year hourly data

hierarchical clustering algorithm

class	occurrence	
	number	fraction
$\mathcal{R}_1$	1	0.003
$\mathcal{R}_2$	42	0.115
$\mathcal{R}_3$	19	0.052
$\mathcal{R}_4$	303	0.830

k-means algorithm

class	occurrence	
	number	fraction
$\mathcal{R}'_1$	48	0.13
$\mathcal{R}'_2$	119	0.33
$\mathcal{R}'_3$	70	0.19
$\mathcal{R}'_4$	128	0.35

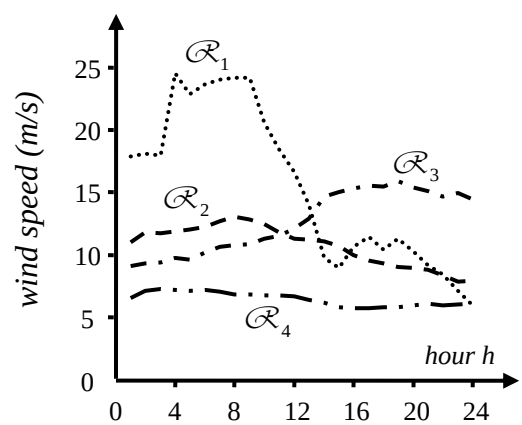
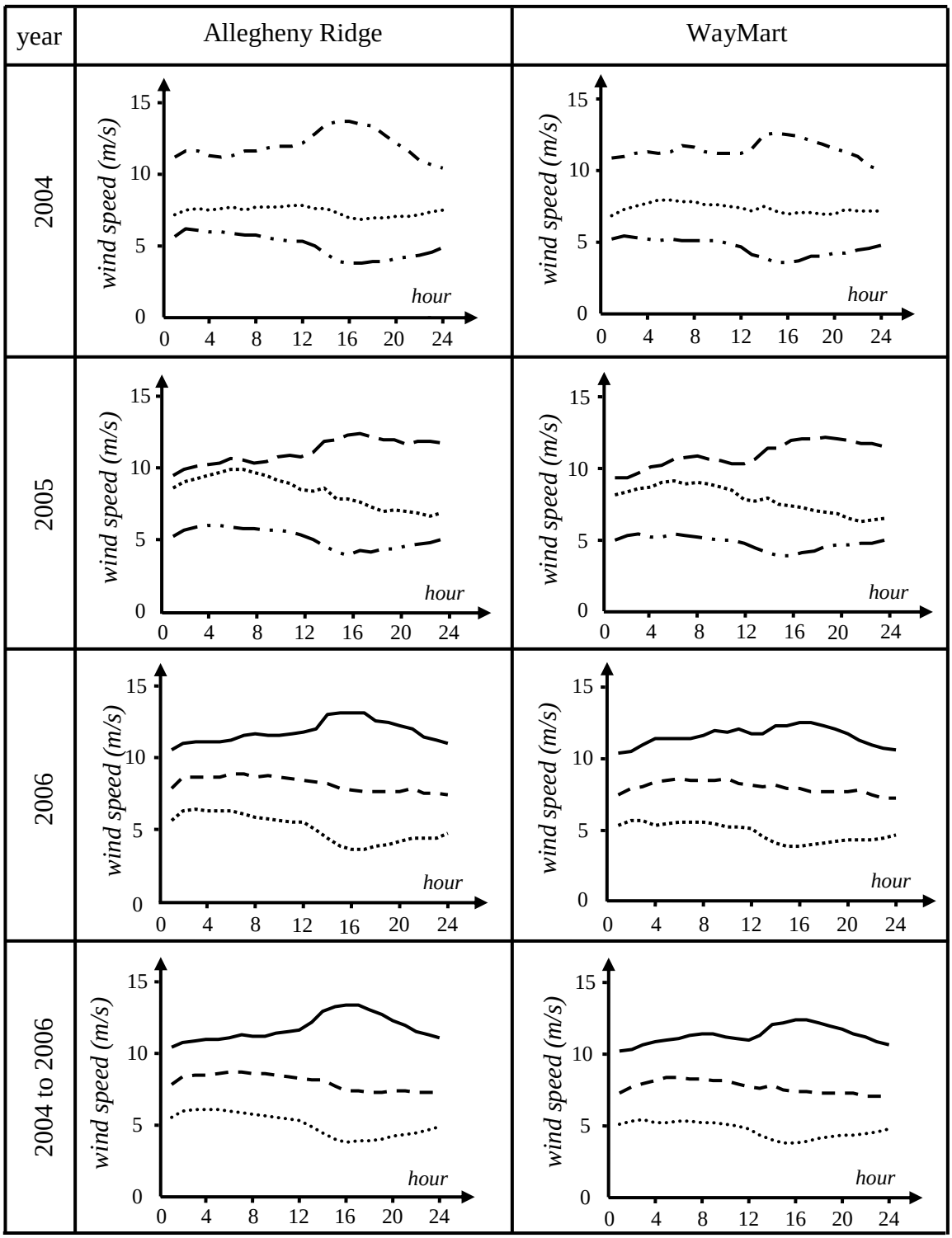


Figure 3.9: The four regimes identified by the hierarchical clustering algorithm for the Allegheny Ridge location, PA, using the 2006 year hourly data.

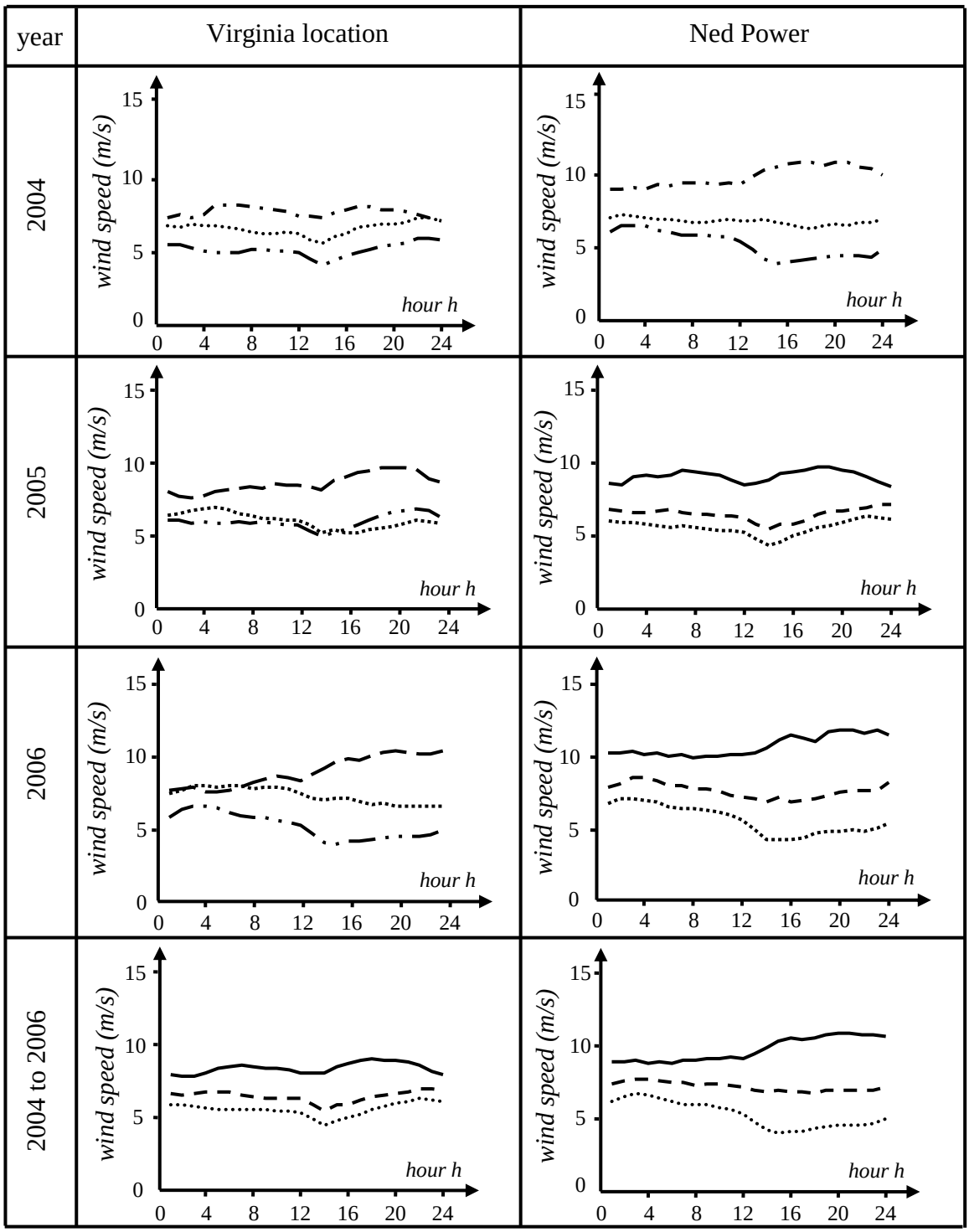
We obtain additional insights from the studies we performed to investigate the persistence of the wind speed patterns from one year to the next so as to assess the impacts of the year-to-year wind speed variations on the wind speed patterns. We are interested in this phenomenon for longer-term studies to determine whether or not the assumption of the persistence of the regimes over the study period is permissible.

Example 3.5: We group the similar days of wind speeds to determine the wind speed regimes for the years 2004, 2005 and 2006 – the years for which we were able to obtain the detailed wind speed data. To provide a representative result for the persistence issue, we consider four different sites located in the PJM geographical footprint – Allegheny Ridge, PA, WayMart, PA, Ned Power, WV and one location in Virginia – shown in Fig. 3.3. We identify three regimes using the *k*-means clustering scheme. We display estimates of the average daily wind speed patterns at the four locations in Fig. 3.9 (a) and Fig. 3.9 (b). The results clearly indicate that the wind speed regimes tend to persist over longer periods. Such a finding need not be true in all the cases as we noted in the extensive testing we performed. We also investigate the effect of the data set length on the clustering results. We identify with the *k*-means classification scheme the three regimes making use of the data set composed of the hourly wind speeds for the three consecutive years 2004-2006. We plot the average daily wind speed patterns of the three classes we obtained in Fig. 3.10 (a) and 3.10 (b). We notice that the results obtained for the 2004-6 period also look similar to those obtained with



(a)

Figure 3.10: The three average daily wind speed patterns identified with the *k*-means scheme for the Allegheny Ridge, WayMart, Ned Power and Virginia locations using different wind speed datasets.



(b)

Figure 3.10: Continued.

considering the years 2004 - 2006 separately. The use of longer data sets reinforced the nature of the results presented here.

We also investigate the seasonality considerations in the identification of wind regimes. In particular, we are interested in the determination of the extent to which the seasonal variability is captured by the identified regimes.

Example 3.6: To provide a representative illustration we consider the Camp Grove location for which we identified four regimes using 2006 year data and presented the results in Fig. 3.10. We analyze the clustering results we obtained so as to observe the regimes seasonality. We partition the year into four quarters, each of three months duration, with Q_1 corresponding to the January to March months, Q_2 to the April to June months, Q_3 to the July to September months and Q_4 to the October to December months. We compute the number of occurrences of each regime within each of the four quarters and plot the respective frequencies of occurrence in Fig. 3.11. We remark that the frequency of occurrence of the wind speed regimes varies throughout the year. For example, the regime $\mathcal{R}'_2$ occurs 28 times in Q_1 while it occurs 48 times – almost twice as frequently – in Q_3. Similarly, the regime $\mathcal{R}'_4$ occurs 20 days in Q_1 and only 10 days in Q_3. With this example, we have emphasized the huge disparity of the frequencies of occurrences of the various regimes depending on the season under consideration. In this way, we have shown that the regimes-based wind speed modeling

approach captures the variability of the wind speed on the seasonal wind speed variability timeframe.

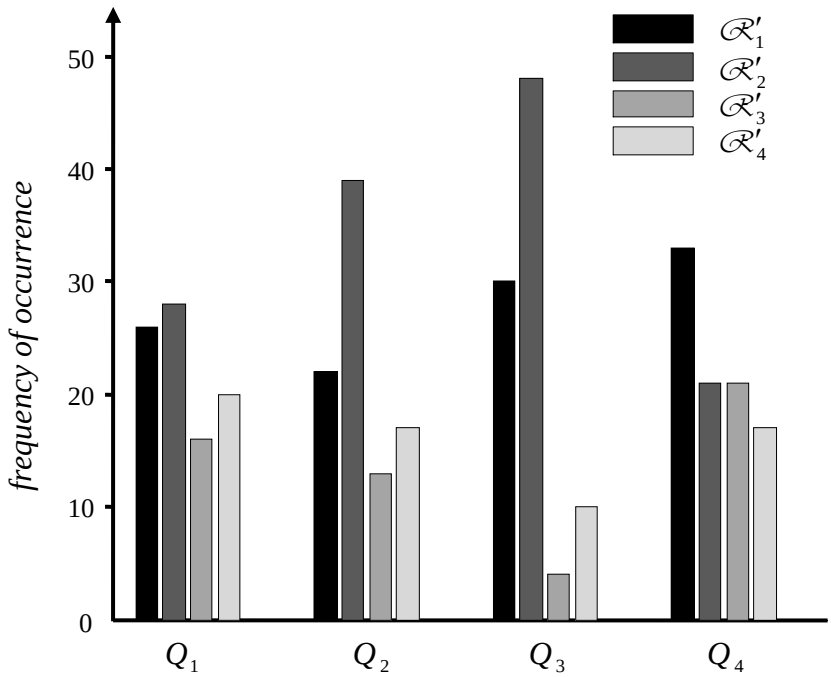


Figure 3.11: Quarterly frequency of occurrence for the four regimes identified with the *k*-means algorithms for the Camp Grove location, IL, using the year 2006 hourly data.

Also note that the regime $\mathcal{R}'_3$ only occurs 4 days in Q_3, which is a relatively small number of occurrences. It is possible to obtain results where some of the classes have no occurrence in a particular quarter. This phenomenon has been observed in [39] where the authors identify the four typical daily wind speed patterns in La Paz, Baja California Sur, Mexico, with the hierarchical clustering algorithm. In our case, we chose to use the *k*-means clustering scheme so as to avoid the problems of having classes with few or no days of wind speed. We perform a similar analysis using a longer data set: we identify four regimes at the Camp Grove location, IL, using the data set composed of the wind speed data for the three consecutive years 2004-2006. We also divide the year into four quarters

and plot the number of occurrences of each regime for each quarter in Fig. 3.12. The use of a longer data set reinforced the nature of the results presented here.

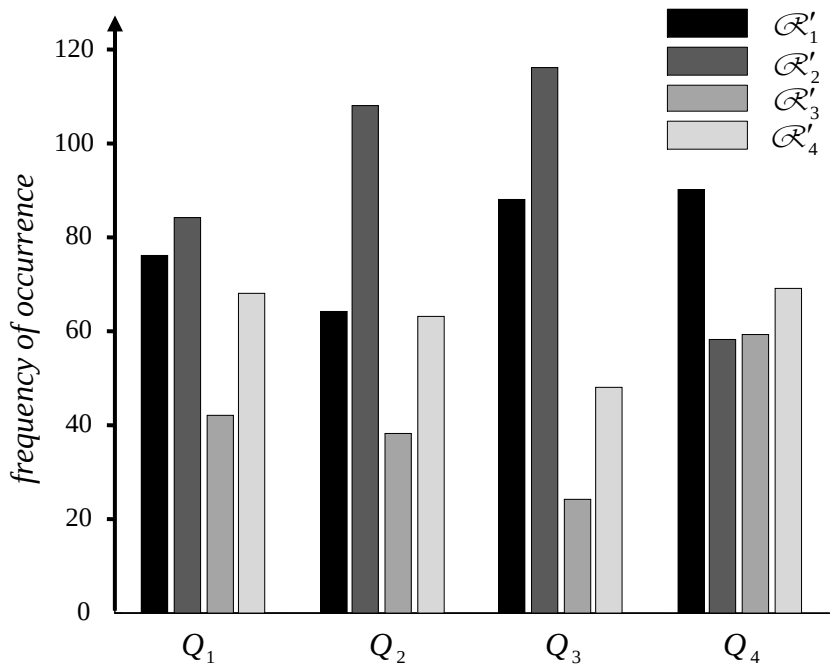


Figure 3.12: Quarterly frequency of occurrence for the four regimes identified with the *k*-means algorithms for the Camp Grove location, IL, using the year 2004-6 hourly data.

We presented this discussion to provide some insights into the types of results that are possible with the two clustering algorithms. We use the regimes for the modeling of the wind power output from a turbine and to construct the model of the wind power output of a wind farm. We discuss these models in the next section.

3.3 Wind Farm Generation Output Model

In this section, we focus on the analytic characterization of the output power of a single wind unit and a wind farm. We analyze the wind speed to wind power conversion process to generate the energy harnessed to supply electricity. We first focus our analysis

on the behavior of a stand-alone wind turbine. The amount of wind power produced by such a turbine depends on the wind speed – and the energy contained in the moving masses of air – as well as the wind turbine technology and its operation. Specifically, the operation of a wind turbine is dependent on the following wind speed levels [6], [40]:

- v_i: the cut-in wind speed: At low wind speeds, the power contained in the wind is insufficient to overcome the various friction forces within the turbine, and so the turbine produces no power as long as the wind speed $v < v_i$; the wind speed must be at least v_i for the turbine to produce power.
- v_r: the rated wind speed: As the wind speed increases, the wind turbine power output also increases when the wind speed is at or above v_i; once the wind speed reaches its rated value v_r, the turbine output power reaches the maximum power the wind unit is designed to produce and this output level is maintained for wind speeds above v_r, by dispatching certain control actions [40].
- v_o: the cut-out wind speed: The highest speed at which the wind turbine generates power since, for speeds exceeding the cut-out wind speed, the operation is shut down so as to protect the turbine from any damage.

The maximum output power of a wind unit is an important characteristic of the wind turbine. We often refer to this value as the rated or the nameplate capacity of the wind turbine and we denote it by p_r. The value of p_r, together with those of v_i, v_r and v_o, completely characterizes the wind turbine. Typical representative values of the characteristics of the equipment in use today are in the ranges indicated in Table 3.8.

Table 3.8: The ranges of the typical values of the actual wind turbines' characteristic parameters

parameter	range	units
v_i	3 - 5	m/s
v_r	11 - 15	m/s
v_o	20 - 30	m/s
p_r	up to 5,000	kW

We next state the mathematical representation of the wind turbine model. We denote by p the wind power produced by a wind turbine driven by wind speed v. The relationship between p and v is given by the nonlinear mapping $g: \mathbb{R}_+ \mapsto \mathbb{R}_+$:

$$p = g(v), \tag{3.26}$$

which defines the so-called power curve using the wind turbine characteristic values. There are different forms of the function $g(\cdot)$. In this thesis we use the following function for the power curve:

$$g(v) = \begin{cases} 0 & 0 \le v < v_i \\ a + b.v^3 & v_i \le v < v_r \\ p_r & v_r \le v < v_o \\ 0 & v \ge v_o \end{cases}. \tag{3.27}$$

The polynomial coefficients a and b are specified so that the power curve is continuous also at v_i and v_r.

We illustrate the nature of our power curve representation using the 1.5 MW rated power turbine whose characteristic wind speeds are given by the vector

$\left[v_i, v_r, v_o \right]^T = \left[3, 12, 20 \right]^T$, with all winds stated in m/s. The power curve of a wind turbine with such characteristics has the following analytical formulation:

$$g(v) = \begin{cases} 0 & 0 \le v < 3 \\ -2,60.10^{-3} + 8,82.10^{-4}.\, v^3 & 3 \le v < 12 \\ 1,5 & 12 \le v < 20 \\ 0 & v \ge 20 \end{cases}. \tag{3.28}$$

We plot the power curve of such a wind turbine in Fig. 3.13.

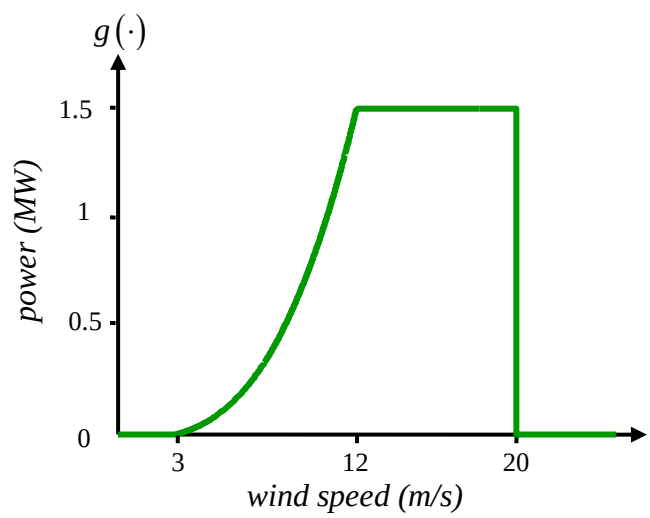


Figure 3.13: The power curve of the 1.5 MW wind turbine characterized by the vector the vector $\left[v_i, v_r, v_o \right]^T = \left[3, 12, 20 \right]^T$.

We next consider the modeling of a wind farm. Typically, a wind farm consists of several wind turbines – from the same or different manufacturers – aggregated at some site s. For modeling purposes, we assume that each turbine is driven by the same wind speed. This assumption is reasonable for actual operations. Even if the wind speeds at two different locations on a wind farm are not identical at each point in time, they are likely to be nearly so on the average over a given period. Additionally, the cannibalization effect among the neighboring wind turbines is assumed to be negligibly

small: in theory, the wind speed experienced by a turbine downwind is smaller than the one experienced by the upwind turbine. The cannibalization effect depends heavily on the actual configuration of the wind farm. Typically, wind farm developers have the necessary expanse to install the wind turbines in such a way to minimize the impacts of interactions among the turbines in the farm. As a result, the losses resulting from the interferences between the various turbines are negligible. We further assume that all turbines are constructed on towers of equal height; such an assumption is reasonable since such uniformity reduces the total construction costs.

We next discuss the steps to construct the power curve $g_s(\cdot)$ for the wind farm at site s. Let $\mathcal{C}_s = \{ c_s : c_s = 1, \ldots, C_s \}$ be the set of the wind turbine technologies implemented and let n_{c_s} be the number of wind turbines of type c_s at site s. A wind turbine of technology c_s produces a wind power p_{c_s} when experiencing wind speed v_s and is therefore specified with the power curve $g_{c_s}(\cdot)$ with:

$$p_{c_s} = g_{c_s}(v_s). \tag{3.29}$$

Under the assumptions above, the total output of the wind farm is the sum of all the outputs of the individual turbines, and so:

$$p_s = g_s(v_s) = \sum_{c_s \in \mathcal{C}_s} n_{c_s} \cdot p_{c_s} = \sum_{c_s \in \mathcal{C}_s} n_{c_s} \cdot g_{c_s}(v_s). \tag{3.30}$$

To illustrate the computation of $g_s(\cdot)$ of a wind farm at location s, we consider a wind farm consisting of wind turbines of three different technologies. The wind farm composition and characteristics of the wind turbines are summarized in Table 3.9. The values of the characteristic points of the wind turbines are taken from [41]. Note that we

only select turbines whose height is around 80 m. We plot the power curves of the three types of turbines in Fig. 3.14 and the power curve of the wind farm is also shown in Fig. 3.15.

Table 3.9: Wind turbine technology data for the wind farm

turbine model	manufacturer	tower height (m)	wind speed (m/s)			power (MW)	number of units
			v_i	v_o	v_r	p_r	n_{c_s}
MM 92	Repower Systems	80	3	24	11.2	2	40
1.5 xle	GE	80	3	20	12	1.5	40
G 80	Gamesa	78	4	25	15	2	40

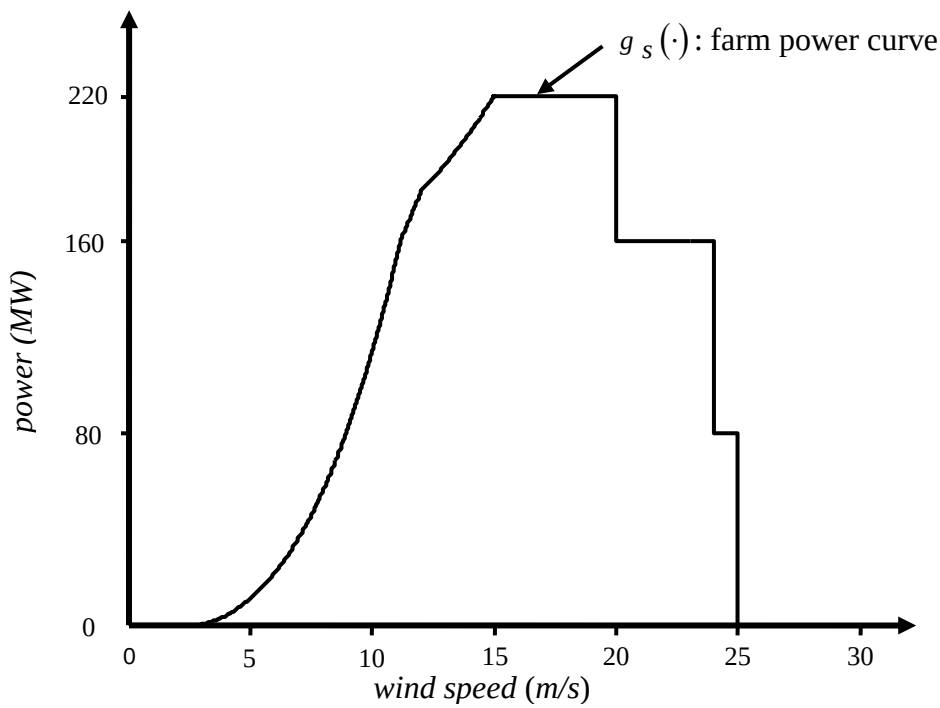


Figure 3.14: Wind farm power curve.

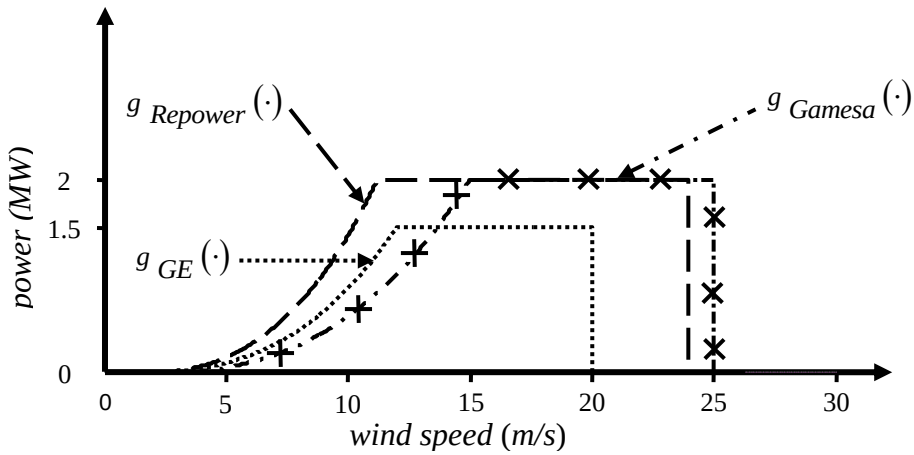


Figure 3.15: Power curves of the three different turbines technology.

Now, we discuss the representation of the variability of the wind power output. Since the wind speed is random, the wind power production – which directly depends on the wind speed – is consequently also random, and we next focus on the development of a probabilistic model for the wind power. We need such a model to explicitly represent the output power uncertainty. We denote by $\underset{\sim}{P}$ be the wind power rv associated with the wind speed rv $\underset{\sim}{V}$ and write:

$$\underset{\sim}{P} = g\left(\underset{\sim}{V}\right). \tag{3.31}$$

The wind speed rv $\underset{\sim}{V}$ is characterized by its cdf denoted by $F_{\underset{\sim}{V}}(\cdot)$ and by its pdf denoted by $f_{\underset{\sim}{V}}(\cdot)$. We analytically express the cdf of the wind power rv $\underset{\sim}{P}$:

$$\begin{aligned} F_{\underset{\sim}{P}}(p) &= Prob\left\{ \underset{\sim}{P} \le p \right\} \\ &= \begin{cases} 0 & p < 0 \\ F_{\underset{\sim}{V}}\left[g^{-1}(p)\right] + \left[1 - F_{\underset{\sim}{V}}(v_o)\right] & 0 \le p < p_r \\ 1 & p \ge p_r \end{cases}, \end{aligned} \tag{3.32}$$

where $g^{-1}(\cdot)$ is the inverse of the restriction of $g(\cdot)$ to the interval (v_i, v_r).[6] In other words:

$$v = g^{-1}(p) \tag{3.33}$$

with

$$g^{-1}: (0, p_r) \mapsto (v_i, v_r). \tag{3.34}$$

By differentiating the cumulative distribution function with respect to p we obtain the pdf of the output power of the wind turbine:

$$f_{\underset{\sim}{P}}(p) = \begin{cases} \left(\left[1 - F_{\underset{\sim}{V}}(v_o)\right] + F_{\underset{\sim}{V}}(v_i)\right)\delta(p) & p = 0 \\ \dfrac{d\left[g^{-1}(p)\right]}{dp} f_{\underset{\sim}{V}}\left[g^{-1}(p)\right] & 0 < p < p_r \\ \left[F_{\underset{\sim}{V}}(v_o) - F_{\underset{\sim}{V}}(v_r)\right]\delta(p - p_r) & p = p_r \end{cases}. \tag{3.35}$$

In Section 3.1, we characterized each wind speed regime $\mathcal{R}_r$ by a set of $(S)(H)$ random variables, one for each site and subperiod of the day. In the case where the wind speed data has an hourly resolution, we have a total of $(S)(24)$ wind speed rv's denoted by $\underset{\sim}{V}_{s,r}^{(h)}$, $s \in \mathcal{S}$ and $h = 1, 2, \ldots, 24$. From the probabilistic characterization of the wind speed rv's at site s, we use the wind farm power curve $g_s(\cdot)$ and the appropriate Equations (3.24) - (3.28) to obtain a characterization of the 24 wind power production rv's for regime $\mathcal{R}_r$ and hour h, which we denote by $\underset{\sim}{P}_{s,r}^{(h)}$:

$$\underset{\sim}{P}_{s,r}^{(h)} = g_s\left(\underset{\sim}{V}_{s,r}^{(h)}\right). \tag{3.36}$$

[6] By construction, the restriction of the power curve $g(\cdot)$ is always invertible on the interval (v_i, v_r).

We illustrate such a process in Fig. 3.16.

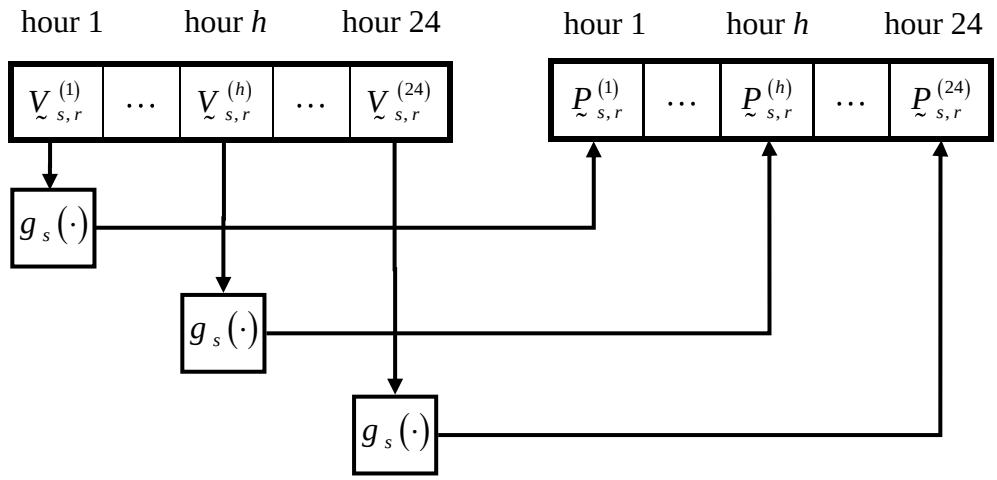


Figure 3.16: The characterization of the 24 wind power rv's for the regime $\mathscr{R}_r$ at site s.

We now use historical wind speed data to illustrate the construction of the cdf and pdf of the wind power produced by a single wind turbine.

Example 3.7: Consider the GE wind turbine whose characteristic parameters are given in Table 3.9. The corresponding power curve is depicted in Fig. 3.13. We plot the restriction of the wind turbine power curve $g(\cdot)$ to the interval $(v_i,\ v_r) = (3,\ 12)$ in Fig, 3.17 (a) as well as its inverse $g^{-1}(\cdot)$ in Fig 3.17 (b). We consider the siting of this turbine in Fenton, MN. In Section 3.2, we identified three wind speed regimes at this location with the *k*-means algorithm and the results were shown in Fig. 3.4 and Table 3.1. We make use of Equation (3.32) and $F_{\underset{\sim}{V}_3^{(12)}}(\cdot)$ – as shown in Fig. 3.5 – to compute the cdf of the output power production of the wind turbine for regime $\mathscr{R}'_3$ and hour 12 and we plot this function in Fig. 3.18. Similarly, we use Equation (3.35) and $f_{\underset{\sim}{V}_3^{(12)}}(\cdot)$ – as shown

in Fig. 3.6 – to estimate the wind power pdf for regime $\mathscr{R}'_3$ and hour 12 and depict it in Fig. 3.19. Note that the impulses at the terminal points are represented with crosses and their values are listed in the associated table; such values correspond to the jumps in the cdf which we observe in Fig. 3.18 for wind power productions of 0 and 1.5 MW.

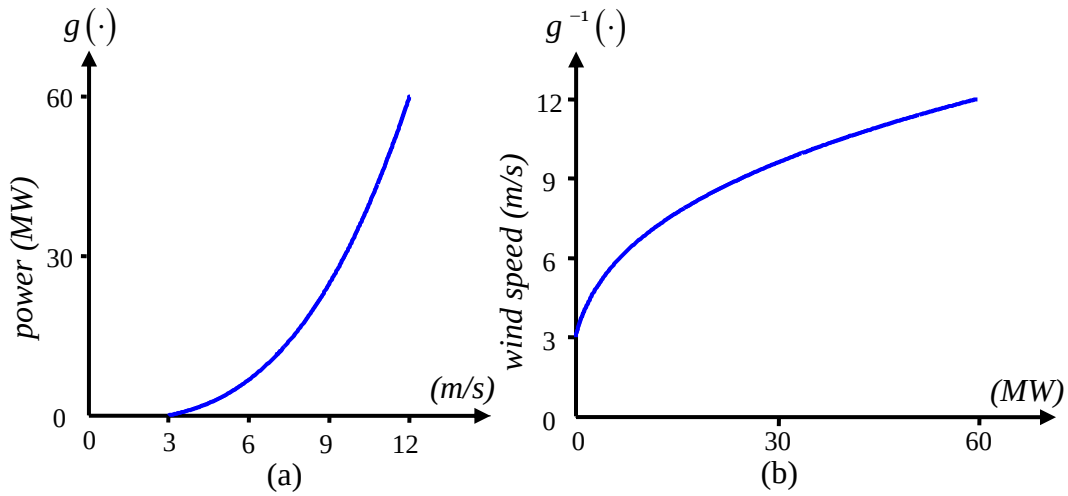


Figure 3.17: The restriction of $g(\cdot)$ on (v_i, v_r) and its inverse $g^{-1}(\cdot)$.

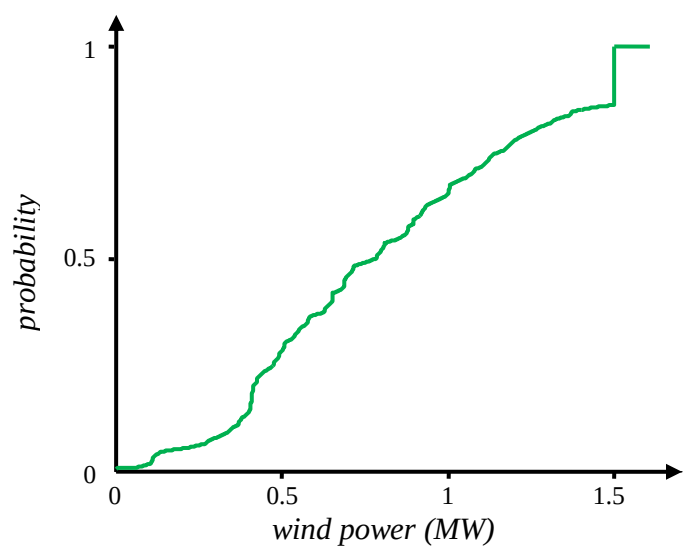


Figure 3.18: The cumulative distribution function of the random variable representing the wind power at Fenton, MN, for regime $\mathscr{R}'_3$ and hour 12.

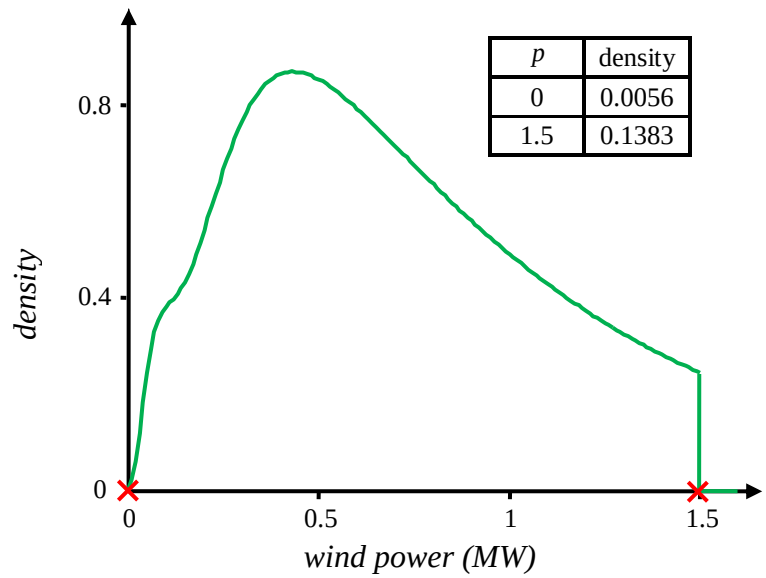


Figure 3.19: The probability density function of the random variable representing the wind power at Fenton, MN, for regime $\mathcal{R}'_3$ and hour 12.

This example illustrates the mechanics of computing the probability functions for the wind power production of a single wind turbine. As we are also interested in estimating the cdf and the pdf of the output power of a wind farm composed of wind units of different technologies, we extend the formulas (3.31) – (3.35) and we use the same historical wind speed data set to illustrate the construction of the cdf and pdf of the wind power produced by a wind farm.

Example 3.8: We consider the wind farm whose power curve is given in Fig. 3.13 and sited at the same location: Fenton, MN; we repeat the same succession of steps so as to estimate the wind farm output power cdf for regime $\mathcal{R}'_3$ and hour 12 – as shown in Fig. 3.20 – and its pdf – plotted in Fig. 3.21. As for the single turbine case, the probability density function also has impulses whose values are given in the associated table. Note that the discontinuities in the wind farm output

power pdf observed in Fig. 3.21 come from the discontinuities in the power curve derivative between the cut-in and rated wind speeds.

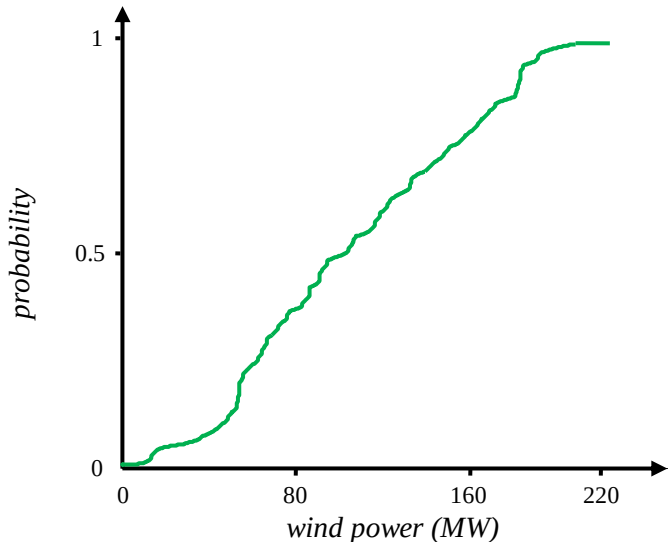
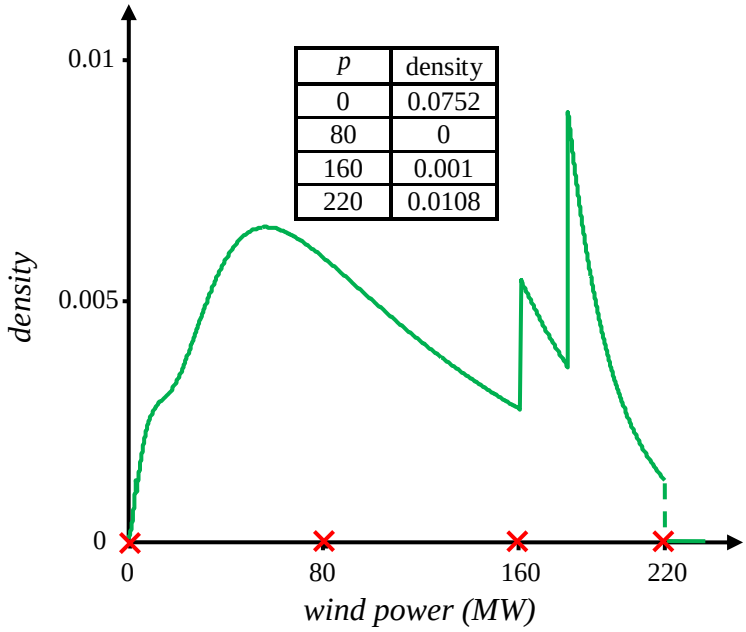


Figure 3.20: The cumulative distribution function of the random variable representing the wind power at Fenton, MN, for regime $\mathscr{R}'_3$ and hour 12.



p	density
0	0.0752
80	0
160	0.001
220	0.0108

Figure 3.21: The probability density function of the random variable representing the wind power at Fenton, MN, for regime $\mathscr{R}'_3$ and hour 12.

In this section, we have detailed the methodology we employ to estimate the cdf and the pdf of the output power of a wind unit / farm located at site s for a single hour h and regime $\mathcal{R}_r$. We need to extend this approach in the case where wind resources are located at multiple locations.

3.4 Wind Power Modeling for Systems with Dispersed Wind Resources

We apply the transformation described in Fig. 3.16 at the S sites making use of the set of the S wind farm power curves: $g_1(\cdot), g_2(\cdot), \ldots, g_S(\cdot)$, so as to obtain, for each regime, the density functions of the $(S)(24)$ wind power random variables $\underset{\sim}{P}_{s,r}^{(h)}$. To illustrate such a process, we may use a matrix representation of the collection of wind speed rv's $\underset{\sim}{V}_{s,r}^{(h)}$. Clearly, the matrix has S rows and 24 columns and each row contains the set of wind speed rv's at a given site s for the 24 hours of the day. We also view the set of wind power rv's in a matrix form whose size is the same as the wind speed rv's matrix. The two matrices are shown in Fig. 3.22.

For every regime and hour, we are interested in the total wind power production resulting from the summation of the wind power production at the S sites. We denote by $\underset{\sim}{P}_{\Sigma,r}^{(h)}$ the aggregate wind power rv corresponding to the wind speed regime $\mathcal{R}_r$ and hour h of the day, and so:

$$\underset{\sim}{P}_{\Sigma,r}^{(h)} = \sum_{s \in \mathcal{S}} \underset{\sim}{P}_{s,r}^{(h)} = \sum_{s \in \mathcal{S}} g_s\left(\underset{\sim}{V}_{s,r}^{(h)}\right). \tag{3.37}$$

For each wind speed regime, we repeat the summation 24 times, once per hour of the day, so as to construct the collection of rv's representing the wind power production. The

summation process of the wind power rv's at the various wind farm locations is illustrated in Fig. 3.23.

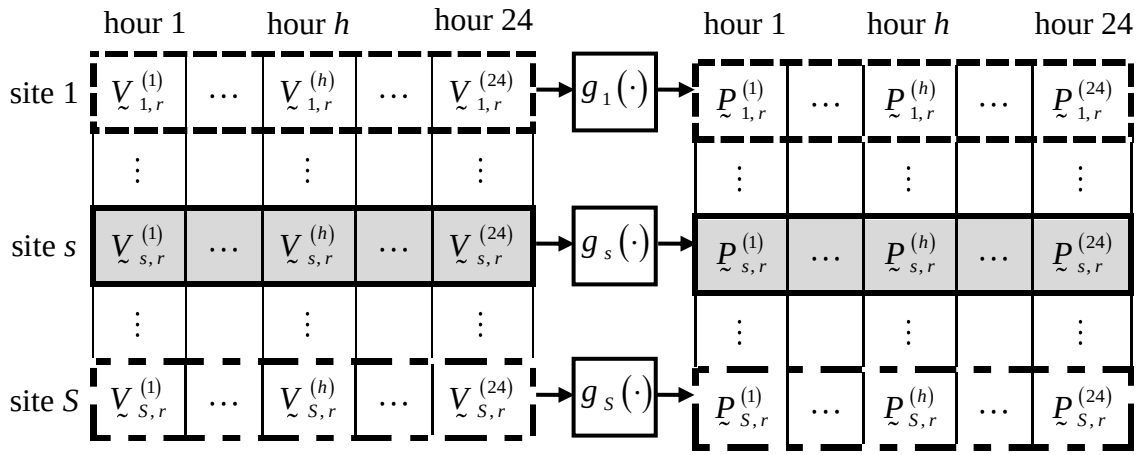


Figure 3.22: The characterization of the $(S)(24)$ wind power rv's for regime $\mathscr{R}_r$ at the S sites.

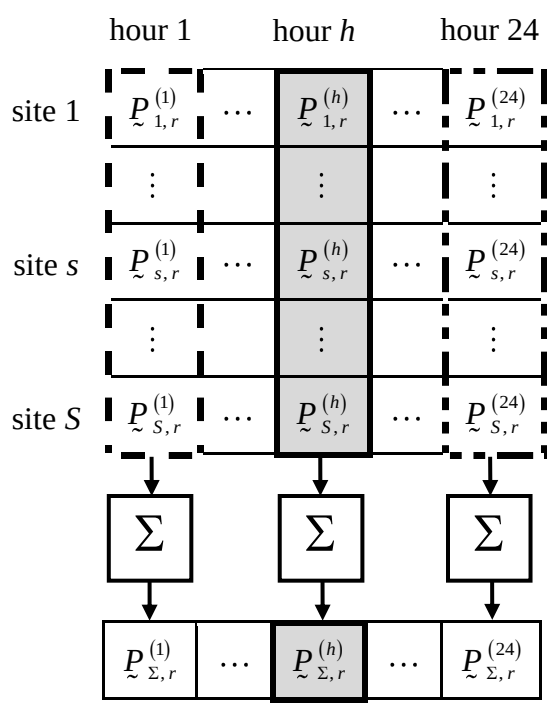


Figure 3.23: The summation of the wind power produced at the S wind farm locations for each hour of the day.

To estimate the cdf and pdf of the 24 hourly wind power rv's $\underset{\sim}{P}_{\Sigma,r}^{(h)}$, we need to make some additional assumptions. With the wind speed regimes-based approach, we have identified the typical days of wind speed which simultaneously occur at the wind farm locations. In this way, we capture the correlation between the daily wind speed patterns at the S sites. However, so far we have not assumed anything about the correlation between the wind speed rv's at two distinct sites $i \in \mathfrak{S}$ and $j \in \mathfrak{S}$, for regime $\mathscr{R}_r$ and hour h of the day: $\underset{\sim}{V}_{i,r}^{(h)}$ and $\underset{\sim}{V}_{j,r}^{(h)}$. To get some insights into the relationship between the wind speeds at two distant locations, we perform a correlation study by using wind speed data at wind farm locations spread over the MISO system footprint for both a single regime and a three-regimes representation of the wind speed. The notion of correlation is discussed in detail in Appendix B. We show that the two random variables $\underset{\sim}{V}_{i,r}^{(h)}$ and $\underset{\sim}{V}_{j,r}^{(h)}$ are weakly correlated in the case where the two locations $i \in \mathfrak{S}$ and $j \in \mathfrak{S}$ are sufficiently distant (> 250 ~ 300 km). If such a condition is met, it is therefore reasonable to assume that the rv's representing the wind speed at two distant locations are independent. As a result, the corresponding wind power rv's, which are functions of independent rv's, are also independent [42]. Consequently, to estimate the pdf and cdf of $\underset{\sim}{P}_{\Sigma,r}^{(h)}$, we iteratively convolve the S pdf's (cdf's) of $\underset{\sim}{P}_{s,r}^{(h)}$.

In our work, we assume that the S sites we consider are widely dispersed so that the correlation among the sites is very small. Consequently, we may assume without loss of generality that the power outputs are independent rv's. In the case where at least two of the S wind farm locations would be too close, we would need to take into account the correlation between the wind power rv's and the use of the convolution formula would

not be valid anymore. We consider such a case to be beyond the scope of the thesis and is left for future work.

In this chapter, we have focused on the modeling of wind speed. We have introduced the notion of wind speed regimes and discussed the schemes used to identify the wind regimes. The aim is to capture the typical daily wind speed patterns at the various locations in a systematic way. We have emphasized the dependence of the clustering results on the available wind speed data set and pointed out the pitfalls that we need to avoid so as to obtain meaningful clustering results. With our model, we capture both the wind speed daily variability – by identifying the typical average daily wind speed patterns – and the within-the-hours uncertainty. Note that we have also shown how the seasonal wind variations are taken into account with the regimes-based approach. Then, we detailed the methodology we employ to estimate the density functions of the wind power rv's for each regime and hour of the day. The modeling is valid in the case where the wind power is produced at a single location and is subject to additional assumptions when we consider multiple wind farms. In the next chapter, we discuss the necessary modifications of the probabilistic production simulation framework so that the wind power model we have proposed can be incorporated. In this way we will be able to run production simulations for systems with wind resources.

4. THE EXTENDED PROBABILISTIC SIMULATION APPROACH

As wind resources produce electricity without the use of fossil fuels, their costs entail no fuel charges and produce virtually zero emissions. However, the wind speed variability and intermittency effects constitute serious limitations to the effective integration of wind resources into the grid. Such effects need to be reflected in the production simulation so as to appropriately evaluate the impacts of the wind resource on the power system production costs and reliability. We extend the widely used probabilistic simulation tool to take into account the impacts of the integration of wind generation resources on system operations. We devote this chapter to review the probabilistic production simulation tool basics and to introduce the required modifications into the probabilistic simulation methodology so as to allow the appropriate representation of the integrated wind resources. We also provide a detailed illustration of the application of the extended simulation approach.

4.1 Review of Production Simulation

The probabilistic production simulation tool is widely used to evaluate the expected energy produced by each unit in the resource mix over a specified period of time. The computation also provides values of the expected total system production costs, the expected emissions, the reliability indices and other figures of merit such as the capacity factor of each unit, i.e., the entire range of variable effects of the system over the specified study period. The probabilistic simulation is a critically important tool for performing longer-term planning studies to investigate issues such as the determination

of the optimal resource mix to serve the forecasted load by evaluating the expected production costs for different resource mixes from which the most economic is selected. Other uses include the evaluation of the economics of various expansion/modification strategies and the investigation of the implications of the introduction of new policies or legislative mandates.

The production simulation study period is specified once the planning horizon of the study is set. For the realistic emulation of the way the system operates, we use a number of shorter simulation periods to cover the entire study period and we perform probabilistic simulation over each simulation period. We define each simulation period in a way so as to capture seasonality effects, as well as changes in the resource mix and resource characteristics, policy and legislative initiatives, investment decisions and the maintenance of the resources. The simulation periods are non-overlapping and they may have unequal durations. For each simulation period, the data are collected and simulated with a resolution commensurate with the level of detail required for the study. The resolution impacts, therefore, the nature of the load and resource representation and entails a specification of the smallest indecomposable unit of time. The load is assumed to be constant over that time unit; phenomena of shorter duration cannot be represented and are ignored in the simulation.

We express the duration of each simulation period in terms of the time unit selected, which we call a subperiod. Let T_t be the number of subperiods that constitute the simulation period t. We define

$$\mathcal{T}_t \triangleq \left\{ 1, \ \ldots \ , \ T_t \right\} \tag{4.1}$$

as the subperiod index set for the simulation period t. For purposes of the simulation, we may view the study period decomposition as shown in Fig. 4.1.

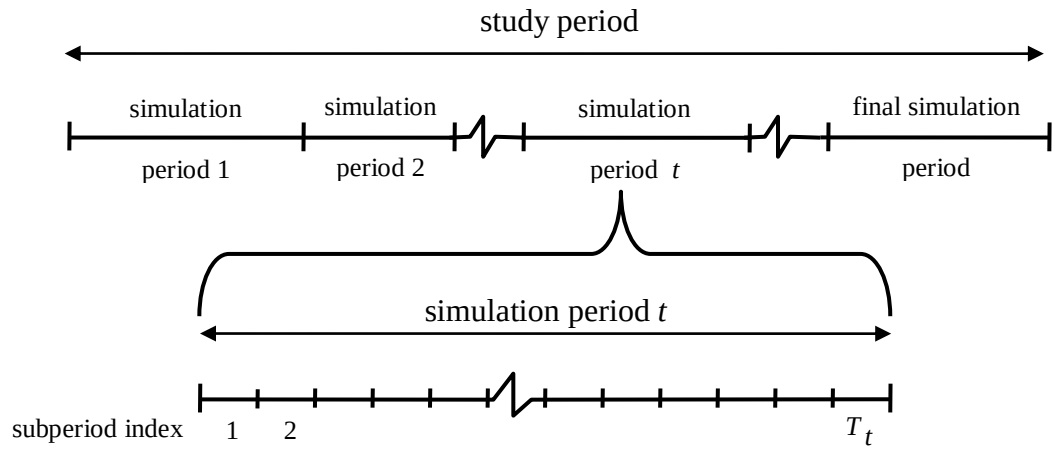


Figure 4.1: Partitioning of the study period into simulation periods in probabilistic simulation.

The chronological load curve is constructed using the values that the load takes on for each subperiod in the simulation period. As an illustration of the chronological load, we consider a weekly simulation period with hourly subperiods. In Fig. 4.2, we plot the chronological hourly load curve of the PJM system for the week of January 16 – 22, 2006. Note that this curve is drawn as a collection of step functions with one step for each of the 168 hours of the week.

We represent the load as an rv $\underset{\sim}{L}$ whose probability distribution we estimate from the chronological load data. We reorder the loads in the chronological curve to construct the so-called load duration curve (ldc) for the simulation period with the order going from the highest to the lowest values of the load. The reordering of the load values removes all

temporal information and all the intertemporal effects are lost. In Fig 4.3, we plot the typical shape of a load duration curve. We easily observe the maximum (minimum) load over the simulation period, which we refer to as the peak (base) load.

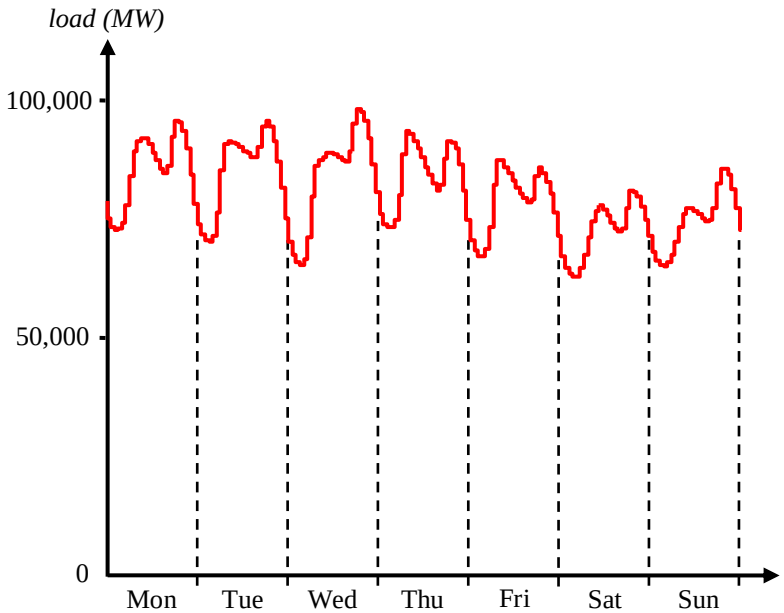


Figure 4.2: The PJM system hourly chronological load for the period of Monday, January 16, to Sunday, January 22, 2006.

We interpret the ldc in the following way. Consider an arbitrary point (ξ, ℓ) on the ldc The load exceeds ℓ for ξ of the T_t subperiods in the simulation period. By normalizing the time, the fraction ξ / T_t may be viewed as an estimate of the probability that the load exceeds ℓ in the simulation period. In Fig 4.4, we plot the ldc of the chronological load shown in Fig. 4.2. From the ldc, we note that the load exceeds 77,893 MW for 96 hours during the week.

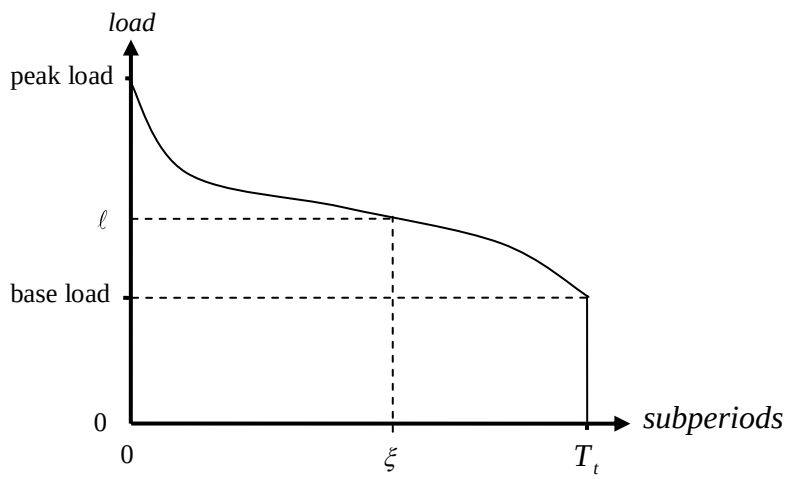


Figure 4.3: Typical shape of a load duration curve.

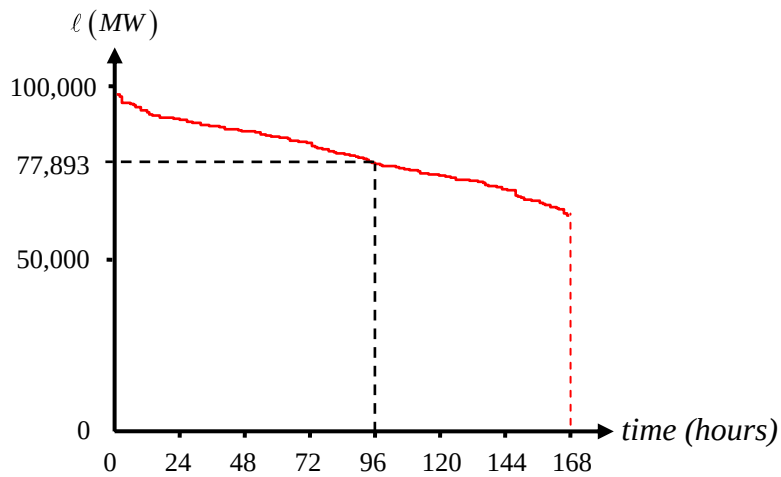


Figure 4.4: The load duration curve corresponding to the chronological load in Fig. 4.2.

We use the probability interpretation to construct the distribution function of $\underset{\sim}{L}$. Indeed, the normalization of the time axis results in the loads being over the $[0,1]$ interval. We flip and invert the ldc to obtain the so-called inverted ldc, which we denote by $\mathcal{L}(\cdot)$. The curve $\mathcal{L}(\cdot)$ is derived from the ldc by normalizing the horizontal axis of the ldc, inverting the curve and rotating through a 90° clockwise angle. Clearly, for a

specified value ℓ of the load, we interpret $\mathcal{L}(\ell)$ to provide the value of the probability that $\underset{\sim}{L} > \ell$, i.e.,

$$\mathcal{L}(\ell) = \text{Prob}\{\underset{\sim}{L} > \ell\}. \tag{4.2}$$

Indeed, $\mathcal{L}(\cdot)$ is simply the complement of the cumulative distribution function (cdf) of the load rv $\underset{\sim}{L}$. Therefore $F_{\underset{\sim}{L}}(\cdot)$, the cdf of $\underset{\sim}{L}$, is given by:

$$F_{\underset{\sim}{L}}(\ell) = \text{Prob}\{\underset{\sim}{L} \le \ell\} = 1 - \mathcal{L}(\ell). \tag{4.3}$$

We illustrate the $\mathcal{L}(\cdot)$ for the ldc in Fig. 4.5 (a) and the cdf of $\underset{\sim}{L}$ in Fig. 4.5 (b).

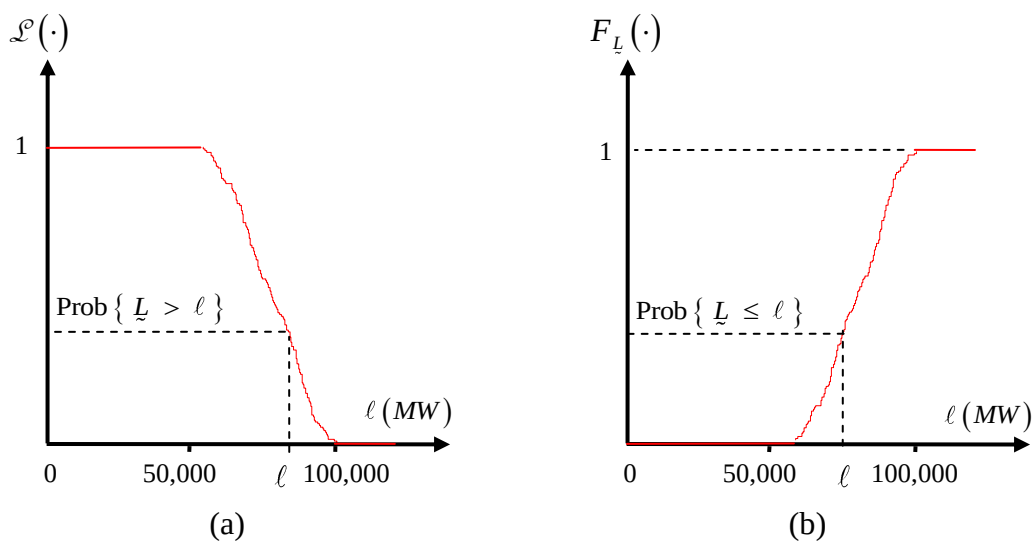


Figure 4.5: The $\mathcal{L}(\cdot)$ and the cdf of the load rv representing the PJM chronological load in Fig. 4.2.

We next consider the modeling of the set of units that are in the resource mix used to meet the load. For the time being, we assume that each unit in the resource mix is controllable/dispatchable. Let I be the number of controllable units of the system we consider. We define

$$\mathcal{I} \triangleq \{i : i = 1, \ldots, I\} \tag{4.4}$$

as the index set of the units in the resource mix. Each controllable unit has its output level set by the system operator but the output is a function of the availability of the unit. We model the availability of each unit by a discrete random variable to represent the multi-state capacities with which a unit may be dispatched. Consider a controllable unit with capacity c_i; we denote by $\underset{\sim}{A}_i$ the availability rv of the unit. The cdf of $\underset{\sim}{A}_i$ is determined from historical operating data of the unit. For the simulation, we use the outage capacity rv $\underset{\sim}{Z}_i$ of a unit where:

$$\underset{\sim}{Z}_i \triangleq c_i - \underset{\sim}{A}_i. \tag{4.5}$$

We introduce the following assumptions in the probabilistic production simulation:

- The availability rv of each unit is independent of any other unit and also independent of the load rv.
- The parameters characterizing each unit – including the availability cdf – remain unchanged for each subperiod of the simulation period.

The probabilistic simulation emulates the energy production of each unit in the resource mix used to meet the load. A unit that is not on planned maintenance may be scheduled for generation in a given simulation period. Once scheduled, each unit is loaded so as to meet the load in the most economic manner. We represent the actual resource scheduling of the controllable units by a priority list. The list is constructed to reflect the actual power system operations. Typically, the blocks are loaded from the cheapest to the most expensive using the unit economics. The unit economics may be expressed in terms of the input-output curve of a unit where the input is the cost of generation as a function of the unit output level in MW or in terms of the prices the blocks of energy are offered by each unit. In this report, we use the marginal cost (mc)

information to describe the unit economics. The blocks with the smallest mc are loaded ahead of those with higher mc to reflect the objective of efficient production costs.

The simulation makes use of the notion of equivalent load. We define the equivalent load to be the rv obtained by summing the load and the outage capacity rv's of the loaded units. Consider the loading of units $1, 2, \ldots, i-1$ to meet the load. We compute the equivalent load $\underset{\sim}{L}_{i-1}$ iteratively and derive from it

$$\underset{\sim}{L}_i = \underset{\sim}{L}_{i-1} + \underset{\sim}{Z}_i, \quad i = 1, 2, \ldots, I \tag{4.6}$$

with

$$\underset{\sim}{L}_0 = \underset{\sim}{L}. \tag{4.7}$$

In (4.6), $\underset{\sim}{L}_{i-1}$ is the equivalent load after the first $i-1$ units have been loaded and $\underset{\sim}{Z}_i$ is the outage capacity rv of the unit i. We may view $\underset{\sim}{L}_i$ as the load that has to be served by the units which will be loaded after unit i is loaded. The computation is carried out in terms of the inverted ldc $\mathcal{L}(\cdot)$ and we successively compute $\mathcal{L}_i(\cdot)$ as the complement of the cdf of the equivalent load rv $\underset{\sim}{L}_i$. The independence assumption allows the use of convolution for this purpose. We make use of $\mathcal{L}_{i-1}(\cdot)$ to compute the expected energy - ε_i - generated by the unit i over the simulation period. The capacity factor CF_i of unit i is defined as the ratio between the expected energy produced ε_i and the maximum energy production over the simulation period:

$$CF_i = \frac{\varepsilon_i}{\varepsilon_i^{\max}} = \frac{\varepsilon_i}{c_i \, . \, T_t}. \tag{4.8}$$

The production simulation also provides as a byproduct the values of the system reliability metrics of interest. We use the notion of equivalent load rv to determine the

loss of load probability (LOLP) and the expected unserved energy $\mathscr{U}$ of the system for the simulation period. The $\mathscr{L}_I(\cdot)$ of the equivalent load rv after all the units are loaded provides the complement of the cdf of the load that remains unserved. Therefore, we interpret the LOLP as the probability of the load remaining unserved after all units are loaded and so:

$$\text{LOLP} = \mathscr{L}_I(C_I) \tag{4.9}$$

where

$$C_I = \sum_{k=1}^{I} c_k \ . \tag{4.10}$$

Similarly, we use $\mathscr{L}_I(\cdot)$ to compute:

$$\mathscr{U} = T_t \int_{C_I}^{\infty} \mathscr{L}_I(x)\, dx \, . \tag{4.11}$$

In actual production costing the units are loaded in blocks to reflect the economics of generation. The details are available in the literature where the modeling of limited energy plants and storage devices such as pumped-storage units are also given. We refer the reader to [43], [44], [45] for additional details.

4.2 Reexamination of the Load Representation in Production Simulation with Time-Dependent Resources

The principal challenge of extending production simulation to systems with integrated resources is to mesh the probabilistic framework of production simulation with that for representing the variability and intermittency of wind resources. We describe in this section the approach that we adopt in the work. In Chapter 3, we developed a regime-

based wind power output model that probabilistically characterizes the wind power on a diurnal period basis. The model provides the wind power production rv for each regime for each subperiod τ of the T subperiods representing the daily period. These models represent the production over the period of simulation for which they are valid. The incorporation of the diurnal regime-based wind output models into the production simulation framework requires several "synchronization" steps. We explain the "synchronization" steps by adopting an hourly resolution for the discussion. This simplifies both the notation and the understanding of the steps. The scheme, however, is sufficiently general to allow the adoption of any other level of resolution. Under an hourly resolution, each wind power regime is characterized by a collection of 24 wind power rv's, one for each hour of the day. To incorporate these models in the production simulation framework, we need a commensurate resolution in the production simulation models.

We consider the load representation in the production simulation and use the hourly resolution for each subperiod. This is the resolution of the chronological load shown in Fig. 4.2 where the weekly load is depicted as having 168 values, one for each hour of the week. The hourly random wind output needs to be directly related to the hourly random load that must be met. However, the load rv $\underset{\sim}{L}$ incorporates no temporal information. To mesh the two models, we start with the sample space of the load rv consisting of the T_t hourly values of load in the simulation period t. We classify the sample values by days and obtain a subset for each day. In this way, we decompose the sample space into as many subsets J as there are days in the simulation period with each subset consisting of that day's 24 hourly load values. We may view the sample space as a

matrix with J rows and 24 columns. We define the rv whose state space corresponds to column h, $h = 1, 2, \ldots, 24$, as $\underset{\sim}{L}^{(h)}$. Conceptually, $\underset{\sim}{L}^{(h)}$ is the rv whose cdf $F_{\underset{\sim}{L}^{(h)}}(\cdot)$ is the cdf of the rv $\underset{\sim}{L}$ conditionally on the hour h, which we denote by $F_{\underset{\sim}{L}|h}(\cdot)$. The distribution of $\underset{\sim}{L}^{(h)}$ is estimated from the J samples in column h. Let $\mathcal{J}_t^{(h)}$, $h = 1, 2, \ldots, 24$, be the subset of $\mathcal{J}_t$ of the indices of the hour h of the J days constituting the simulation period. Then the cdf of $\underset{\sim}{L}^{(h)}$ is estimated from the ldc constructed for the J load samples. The use of conditional probability allows us to restate the probability of the event that $\underset{\sim}{L} \le \ell$ in terms of that of the rv's $\underset{\sim}{L}^{(h)}$. Thus:

$$
\begin{aligned}
F_{\underset{\sim}{L}}(\ell) &= Prob\{\, \underset{\sim}{L} \le \ell \,\} \\
&= Prob\{\, \underset{\sim}{L} \le \ell \;\; \textit{in every hour h in the simulation period} \,\} \\
&= \sum_{h=1}^{24} Prob\{\, \underset{\sim}{L} \le \ell \;\mid\; \textit{hour } h \,\} \; Prob\{\, \textit{hour } h \,\} \qquad (4.12) \\
&= \sum_{h=1}^{24} Prob\left\{\, \underset{\sim}{L}^{(h)} \le \ell \,\right\} \; Prob\{\, \textit{hour } h \,\} \\
&= \sum_{h=1}^{24} F_{\underset{\sim}{L}|h}(\ell) \; \frac{1}{24} \qquad ,
\end{aligned}
$$

where we made use of the non-overlapping subsets that constitute the sample space and the fact that each such subset has the uniform probability of occurrence of 1/24. Note that the computation of the final expression requires the addition of at most 24 terms. Indeed, the event $\left\{\, \underset{\sim}{L}^{(h)} \le \ell \,\right\}$ might be an empty set for certain values of ℓ, e.g., always true for any value of ℓ below the base load. By definition the probability of an empty set is zero and so, for these values of ℓ, those terms do not contribute to $Prob\{\, \underset{\sim}{L} \le \ell \,\}$. In Fig.

4.6, we depict the classification and its representation for obtaining compatibility with the wind modeling.

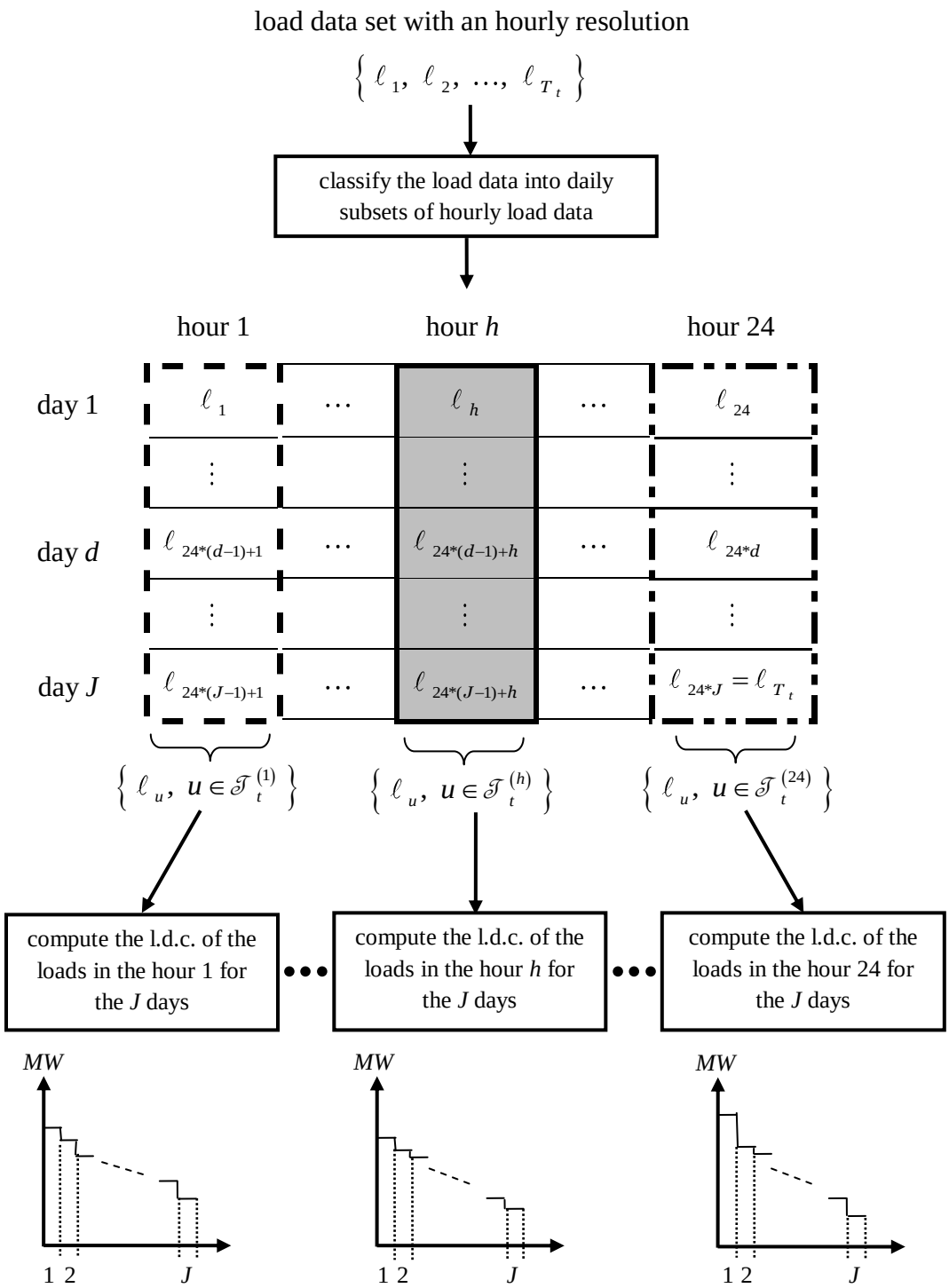


Figure 4.6: The process for the load model refinement.

We can extend the conditioning characterization also for the equivalent loads $\underset{\sim}{L}_i$, $i = 1, 2, \ldots, I$: we replace $\underset{\sim}{L}$ by $\underset{\sim}{L}_i$ in Equation (4.12) and $\underset{\sim}{L}^{(h)}$ by $\underset{\sim}{L}_i^{(h)}$. Such an operation is valid since the unit characteristics are the same for the entire simulation period and therefore for every hour. Therefore, we have:

$$
\begin{aligned}
F_{\underset{\sim}{L}_i}(\ell) &= Prob\left\{ \underset{\sim}{L}_i \le \ell \right\} \\
&= \sum_{h=1}^{24} Prob\left\{ \underset{\sim}{L}_i \le \ell \mid hour\ h \right\} Prob\left\{ hour\ h \right\} \\
&= \sum_{h=1}^{24} Prob\left\{ \underset{\sim}{L}_i^{(h)} \le \ell \right\} Prob\left\{ hour\ h \right\} \\
&= \sum_{h=1}^{24} F_{\underset{\sim}{L}_i \mid h}(\ell) \frac{1}{24} \quad .
\end{aligned}
\tag{4.13}
$$

In this way, we can recast all the probabilistic simulation results obtained with the initial model for the load, making use of the load model with a finer resolution.

In this section, we have introduced a refined load model and shown the resulting modifications of such a transformation on the production simulation framework. So far the resource mix of the system we have considered is only constituted of controllable units. In the next section, we describe the combination of the refined load model with the wind power model described in Chapter 3 so that we can employ the production simulation framework for systems with wind resources.

4.3. Extension of Production Simulation with Time-Dependent Resources

Whenever a wind turbine produces electric power at a level which depends on the actual meteorological conditions, we make use of its output power to supply the demand. Therefore, at any time, the controllable units need to serve the load which has not been

served by the wind resources. The modified load which needs to be served by the controllable units is referred to as the "controllable" load which we model with a random variable denoted by $\underset{\sim}{C}$. We can directly make use of the hourly decomposition to compute the "controllable" load rv for each hour, which we denote by $\underset{\sim}{C}^{(h)}$, and we have:

$$\underset{\sim}{C}^{(h)} \triangleq \underset{\sim}{L}^{(h)} - \underset{\sim}{P}_{\Sigma}^{(h)}, \tag{4.14}$$

where $\underset{\sim}{P}_{\Sigma}^{(h)}$ represents the total wind power production rv for the system under study and for hour h, $h = 1, 2, \ldots, 24$. If we consider multiple wind speed regimes, we denote by $\underset{\sim}{P}_{\Sigma,r}^{(h)}$ the wind power production rv for hour h and wind speed regime $\mathcal{R}_r$. In this case, we also denote the "controllable" load rv for hour h and wind speed regime $\mathcal{R}_r$ by $\underset{\sim}{C}_r^{(h)}$, and so:

$$\underset{\sim}{C}_r^{(h)} \triangleq \underset{\sim}{L}^{(h)} - \underset{\sim}{P}_{\Sigma,r}^{(h)}. \tag{4.15}$$

To run a probabilistic production simulation with non-controllable wind generation, we need to probabilistically characterize the "controllable" load rv – denoted by $\underset{\sim}{C}_r$ – for the simulation period and for each wind speed regime $\mathcal{R}_r$ by estimating its cdf or duration curve. Under the assumption that $\underset{\sim}{L}^{(h)}$ and $\underset{\sim}{P}_{\Sigma,r}^{(h)}$ are independent, we compute the cdf of the rv $\underset{\sim}{C}_r^{(h)}$ by convolving the cdf of the rv $\underset{\sim}{L}^{(h)}$ with the pdf of $\underset{\sim}{P}_{\Sigma,r}^{(h)}$. Using an analogous reasoning to the derivation of (4.12), we derive the relationship between the cdf of $\underset{\sim}{C}_r$ and the 24 cdf's of the hourly "controllable" loads:

$$
\begin{aligned}
F_{\underset{\sim}{C}_r}(c) &= Prob\left\{ \underset{\sim}{C}_r \le c \right\} \\
&= Prob\left\{ \underset{\sim}{C}_r \le c \quad in\ every\ hour\ h\ in\ the\ simulation\ period \right\} \\
&= \sum_{h=1}^{24} Prob\left\{ \underset{\sim}{C}_r \le c \ \mid hour\ h \right\} . \frac{1}{24} \qquad (4.16) \\
&= \sum_{h=1}^{24} Prob\left\{ \underset{\sim}{C}_r^{(h)} \le c \right\} . \frac{1}{24} \\
&= \sum_{h=1}^{24} \frac{1}{24} F_{\underset{\sim}{C}_r | h}(c) \quad .
\end{aligned}
$$

So, the cdf of the “controllable” load is also the weighted average of the cdf’s of the 24 hourly “controllable” loads. We illustrate the procedure with the diagram in Fig. 4.7. Note that we need to repeat such a procedure for the k wind speed regimes that have been determined, and for each regime we make use of the “controllable” load cdf as well as the controllable units characteristics to run a production simulation. In this way, we evaluate the figures of merit of the various controllable units conditionally on the wind speed regime. The expected value of the production simulation results over the entire simulation period is the weighted average of their expected value conditionally on the regime.

We next illustrate the application of the analysis developed in this section. We consider the weekly chronological PJM load shape for the year 2006 – available on the PJM web site – and construct the scaled version of the load shape. The scaling factor is set so as to have a 2850 MW peak load – the IEEE reliability test system peak load [46]. As the PJM peak load for the year 2006 is equal to 144,640 MW, the scaling factor is therefore equal to 2,850/144,640. The simulation period is one week: Monday, January

16, to Sunday, January 22, 2006. We compute and plot in Fig. 4.8 the empirical cdf's of three of the 24 rv's representing the hourly loads.[7]

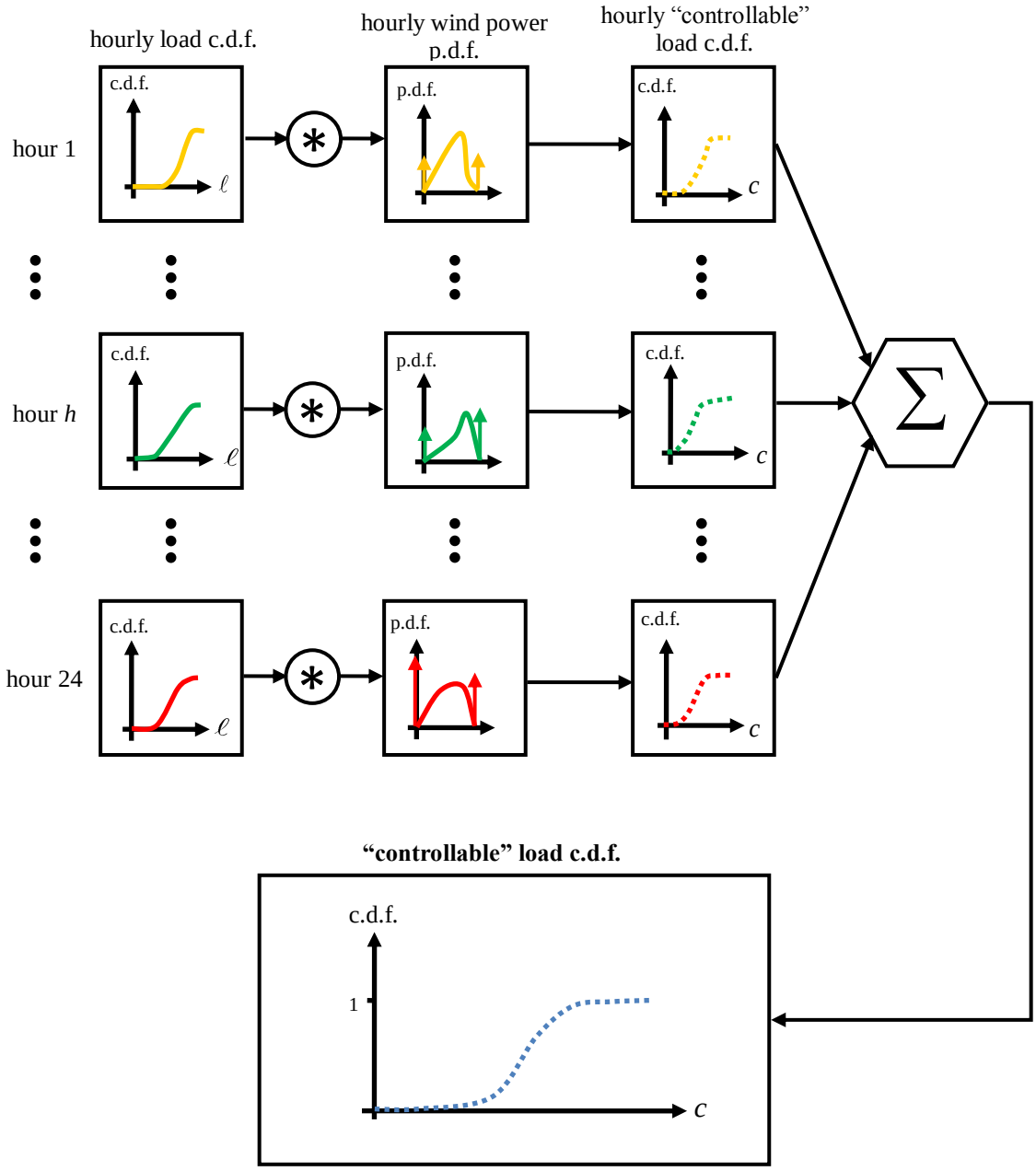


Figure 4.7: Procedure to construct the "controllable" load cdf.

[7] In actual computation, we replace the staircase function by piecewise linear segments.

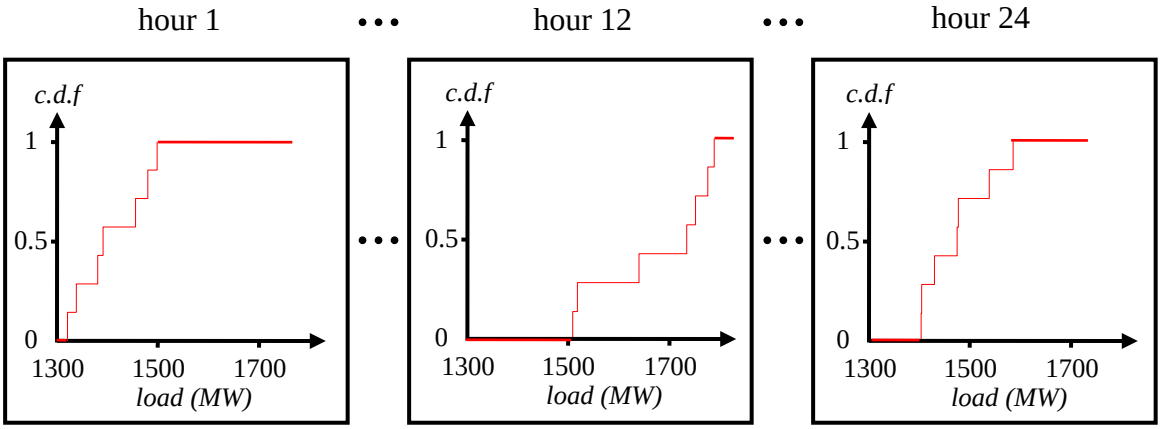


Figure 4.8: The computation of the 24 hours load cdf's.

For computational purposes, we do not consider the stairs empirical cumulative distribution function for the load, but rather its linear interpolation between the vertices of the stairs. Such an action is valid especially if we have a large number of load data points: in this case, the difference between the two curves is negligible.

We consider the integration of a wind farm located at Allegheny Ridge, PA, on the PJM system footprint. It is composed of 200 GE wind turbines whose characteristics are given in Table 3.9. We identify three typical regimes at this location, as shown in Fig. 4.9.

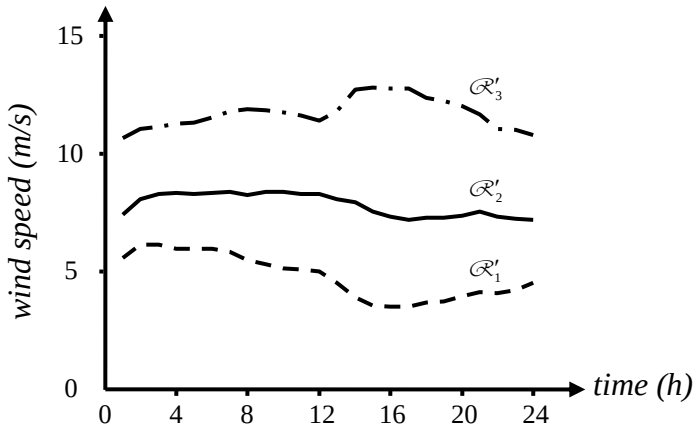


Figure 4.9: The three regimes identified by *k*-means clustering algorithms for the Allegheny Ridge location, PA, using the 2006 year hourly data.

We use Equation (3.35) to estimate the pdf's of the hourly wind power production rv's for the regime $\mathscr{R}'_2$ and the hours 1, 12 and 24 of the day: $\underset{\sim}{P}_2^{(h)}$, $h = 1, 12, 24$, where we suppress the Σ notation since only one site is under consideration. The three density functions are depicted in Fig. 4.10.

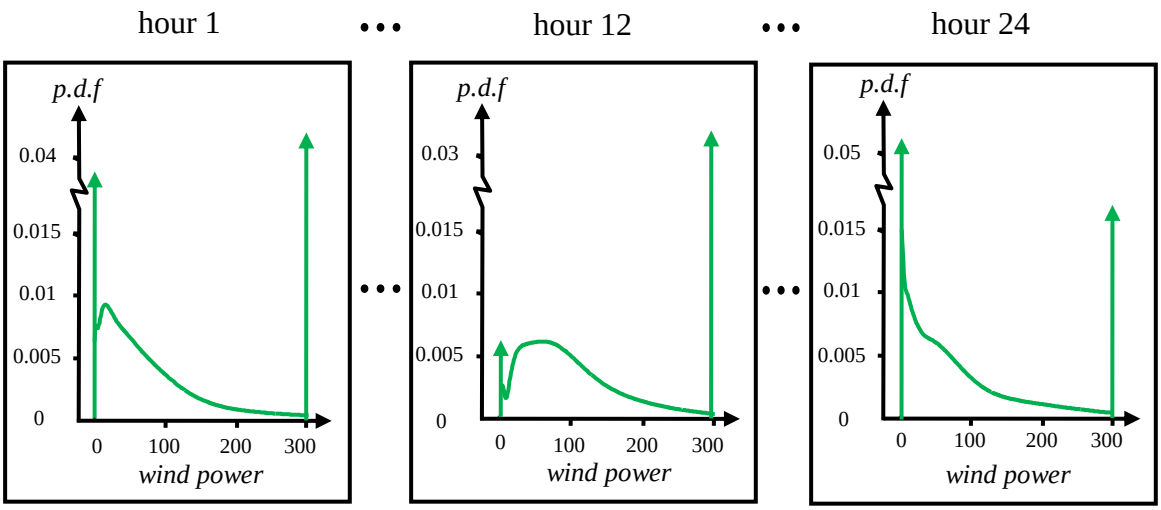


Figure 4.10: The computation of the 24 hourly wind power production pdf's.

We convolve the previously obtained hourly load cdf's and wind power pdf's to compute the cdf's of the 24 hourly "controllable" loads: $\underset{\sim}{C}_2^{(h)}$, $h = 1, 2, \ldots, 24$. In Fig. 4.11, we plot such functions for the hours 1, 12 and 24 of the day. After we have computed the cdf's of the 24 hourly "controllable" load rv's $\underset{\sim}{C}_2^{(h)}$, we combine them so as to get the cdf of the "controllable" load rv $\underset{\sim}{C}_2$. It is the weighted average of the 24 hourly "controllable" load cumulative distribution functions. Such a curve is depicted in Fig. 4.12 where we also plot the empirical cdf of the load for the simulation period. Note

that although the “controllable” load cdf looks like a continuous function, it is piecewise-continuous.

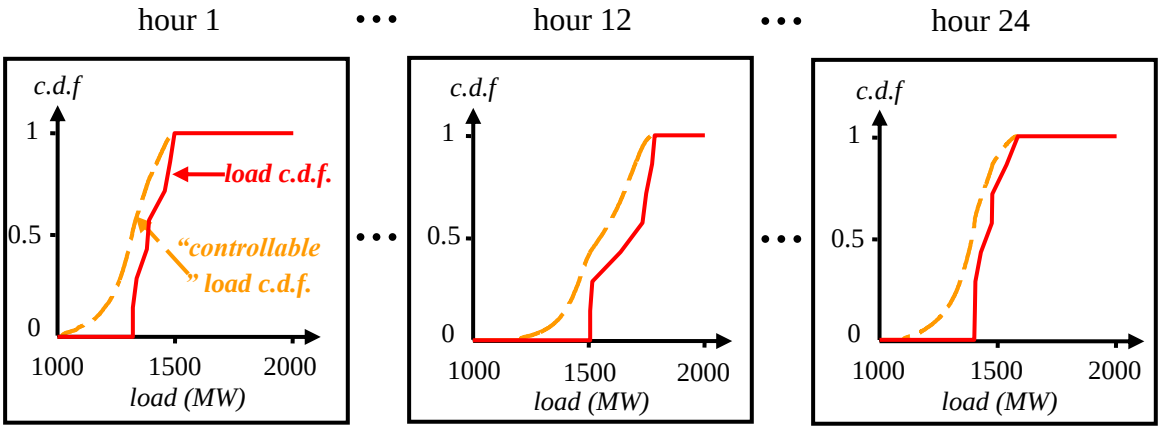


Figure 4.11: The computation of the 24 hourly “controllable” load cdf’s.

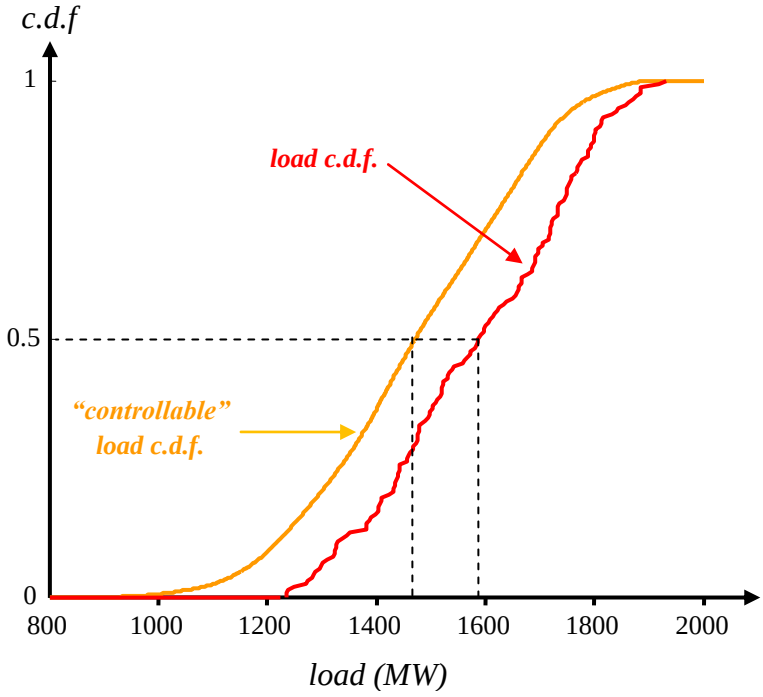


Figure 4.12: The computation of the “controllable” load cdf.

In this chapter, we have described the probabilistic production simulation framework we use for systems whose resource mix is only constituted of controllable units. When wind resources are integrated into the grid, we want to be able to assess the expected value of the production costs of the system as well as its reliability level and its expected CO_2 emissions. In order to do so, we modified the probabilistic production simulation tool by refining the load model so that it is compatible with the regimes-based wind power probabilistic representation developed in Chapter 3. We have also illustrated the methodology with an example to concretely show the succession of steps that we need to undertake when we want to run a production simulation for a system with wind resources. The simulation studies we present in Chapter 5 make use of the tool.

5. SIMULATION STUDIES OF SYSTEMS WITH INTEGRATED WIND RESOURCES

In this chapter, we describe the application studies using the production simulation methodology extended to systems with wind resources. We have carried out an extensive set of simulations to demonstrate the effective capabilities of the tool to capture effects that are observed with the integration of wind resources and to evaluate their impacts over a wide range of conditions. We provide and discuss representative results on two test systems that we investigated in these studies; the systems are synthetic and are constructed to represent different US regions.

Our initial focus is to quantify the effects of the integration of wind resources on the system production costs, CO_2 emissions and reliability metrics. We also investigate the impacts of the wind energy variability on the production simulation results and demonstrate the benefits obtained with the wind farm location diversification to mitigate the intermittency effects. Over the longer-term planning horizon, another major issue to address regarding the integration of wind resources is the impacts of the wind penetration increase. In this chapter, we perform sensitivity studies by varying the size of the wind resources. We analyze the simulation results and identify the most significant effects of the wind penetration increase.

5.1 The Test Systems

We present the simulation results on two test systems to explicitly represent the impacts of location on wind generation. One of the systems, which we refer to as the

Midwestern test system, is located within the MISO region. The other is sited within the PJM area and is called the *Eastern test system*. Additionally, the two test systems have similar sizes based on the fact that they have an identical set of controllable resources and they serve a load with the same peak load (2,850 MW). For both systems, we are also going to consider the integration of wind resources at one or more locations in the area of the respective Independent System Operator. The key differences between the two test systems are summarized in Table 5.1.

Table 5.1: The characteristics of the two test systems

test system	region	wind farm location(s)
Eastern	PJM	Allegheny Ridge, PA
Midwestern	MISO	Camp Grove, IL Fenton, MN Langdon, ND Harvest, MI

Since the Midwestern (Eastern) system is located in the MISO (PJM) region, we synthesize its load to be a scaled-down version of the load shape in the MISO (PJM) area. The MISO (PJM) peak load for the year 2006 is equal to 112,368 MW (144,640 MW), and the scaling factor to construct the load for the Midwestern (Eastern) system is therefore equal to 2,850/112,368 (2,850/144,640).

We now describe the controllable resource mix we constructed for the two test systems. We considered the set of controllable units described in the IEEE reliability test system [46]. It provides us with the forced outage rate, number of blocks and blocks capacities of the various units. The units' economics are specified in terms of the heat

rate at minimum capacity and the incremental heat rate for each block.[8] The CO_2 emission rates of the units are also provided. For computational purposes, we use a slightly modified version of the IEEE reliability test system[9] composed of 29 units of 8 different types. We summarize the characteristics of the units constituting the controllable resource mix of our test systems in Table 5.2.

To characterize the wind resources, we employ the wind speed data set provided by NREL [27] and collect the chronological hourly wind speeds at the locations listed in Table 5.1 for the single-year 2006 period. We consider wind farms sited at such locations and composed of wind turbines – GE 1.5 xle, Repower Systems MM92 and Gamesa G80 – whose characteristics are given in Table 3.9.

Due to the randomness of the wind power production, system operators need to increase the reserve levels so that the system has the ability to track the wind energy variations and compensate for the mismatch between supply and demand resulting from the wind energy forecast error. Such a modification of the system operations leads to higher production costs since more units need to be scheduled and the use of more flexible – and more expensive – units is required. In our work, we represent the reserve levels increase by modifying the loading order of the controllable units: we load the first block of some flexible, and costly, units as base-loaded blocks, the capacity of the remaining blocks being counted as a contribution to reserves. The number of units whose first block is base-loaded is determined so that the summation of the individual units'

[8] Such rates are expressed in Btu/kWh. Throughout the chapter, we use a $6/MMBtu rate to convert the production costs from Btus to dollars.
[9] We turn the multiple-block peaking units into single block units. Also the capacities of some of the units' blocks are slightly modified.

Table 5.2: Characteristics of the test system controllable resources

unit size	number of units	forced outage rate	scheduled maintenance (weeks/year)	CO_2 emissions rate (lbs/MMBtu)	minimum capacity (MW)	heat rate at the minimum capacity (Btu/kWh)	output		incremental heat rate (Btu/kWh)
							% of maximum output	MW	
12	5	0.02	2	170	2	16,017	100	12	10,996
20	4	0.10	2	160	16	15,063	100	20	12,377
80	6	0.02	3	210	16	17,107	20	16	9,548
							50	40	9,966
							80	64	11,576
							100	80	13,311
100	6	0.04	3	170	25	12,999	25	25	8,089
							50	50	8,708
							80	80	9,420
							100	100	9,877
155	4	0.04	4	210	62	11,244	40	62	8,265
							60	93	8,541
							80	124	8,900
							100	155	9,381
200	3	0.05	4	170	70	10,750	35	70	8,348
							60	120	8,833
							80	160	9,225
							100	200	9,620
350	1	0.08	5	210	140	10,200	40	140	8,402
							70	245	8,896
							80	280	9,244
							100	350	9,768
400	2	0.12	6	/	100	12,751	25	100	8,848
							50	200	8,965
							80	320	9,210
							100	400	9,438

contribution to reserves exceeds the target reserve level. Based on the reserve levels proposed in the literature [24], [47], those we adopt in our work depend on the size of the wind resource and are listed in Table 5.3. We may note that, as we have removed the notion of time in probabilistic production simulation, the reserve levels we adopt are uniform over the entire simulation period. Consequently, this approach yields rather conservative results.

Table 5.3: Reserves increase to manage the variability and the intermittency of the wind power production

wind farm nameplate capacity (MW)	additional reserves (MW)
300	60
600	120
900	180

In our work, we also take into account the effects of the scheduled maintenance of the controllable units. We constructed a maintenance schedule enforcing the requirements listed in Table 5.2. The schedule is realistic since most units are shut down for maintenance during the periods with lower load levels: the spring and the fall seasons.

In this section, we constructed the two test systems we used to run production simulations for systems with integrated wind resources. In the remainder of this chapter, we present and analyze the results obtained from such production simulation studies. Note that for purposes of illustration, we report and discuss figures for a single year of simulation: 2006. The extension of the results to longer periods is a straightforward exercise.

5.2 The Base Case for the Two Test Systems

In this section, we define a base case for the two test systems where we assume that the generation mix is only constituted of controllable resources. The objective is to compute the variable effects of the two test systems without wind generation. We will use them as benchmarks to which we will compare the values of the variable effects of the systems with integrated wind resources.

We start with the base case for the Eastern test system. We arbitrarily select a one-week simulation period corresponding to the week of January 2-8, 2006. We computed the LOLP, the expected unserved energy, the expected production costs and CO_2 emissions making use of the production simulation methodology discussed in Section 4.1. We repeated the same process for the 52 weeks constituting the year 2006 so as to assess the variable effects of the system over a study period of one year. We summarize the production simulation results for the Eastern test system base case and for both study periods in Table 5.4. We performed an identical study for the Midwestern test system and we list the results in Table 5.5.

Table 5.4: Base case production simulation results for the Eastern test system for the year 2006 and for the weekly period of Monday, January 2, to Sunday, January 8, 2006

metric	simulation period	
	single week	year 2006
LOLP	$7.78\ 10^{-6}$	$1.26\ 10^{-4}$
expected unserved energy (MWh)	$2.25\ 10^{-2}$	18.6
expected production costs ($)	$1.64\ 10^{7}$	$8.42\ 10^{8}$
expected CO_2 emissions (lbs)	$3.93\ 10^{8}$	$2.09\ 10^{10}$

Table 5.5: Base case production simulation results for the Midwestern test system for the year 2006 and for the weekly period of Monday, January 2, to Sunday, January 8, 2006

metric	simulation period	
	single week	year 2006
LOLP	$2.73\ 10^{-5}$	$1.66\ 10^{-4}$
expected unserved energy (MWh)	$7.84\ 10^{-2}$	24.4
expected production costs ($)	$1.71\ 10^{7}$	$8.94\ 10^{8}$
expected CO_2 emissions (lbs)	$4.02\ 10^{8}$	$2.17\ 10^{10}$

In the next sections, we consider the expansion of the two test systems resulting from the integration of wind resources, while the load and the controllable resource mix remain unchanged.

5.3 Quantifying the Impacts of Wind Integration

In this section, we study the impacts of the integration of wind resources on the system variable effects. We start with quantitatively demonstrating the economic and reliability benefits obtained with the integration of wind resources into the Eastern test system: we consider the siting of a wind farm constituted of 200 GE 1.5xle turbines located in Allegheny Ridge, PA. In Table 3.9 we indicated that such wind units have a 1.5 MW rated power. The wind farm has therefore a 300 MW nameplate capacity corresponding to a 10.5% wind penetration. In this particular case, we defined the penetration as the ratio between the nameplate capacity of the wind resource and the system peak load. We use this definition for the notion of wind penetration throughout the chapter.

We used a single-regime wind speed representation to compute the "controllable" load duration curves for each of the 52 weeks of the year 2006 with the methodology discussed in Section 4.3. In Fig. 5.1, we depict the load and the "controllable" load duration curves for the week of January 2-8, 2006.

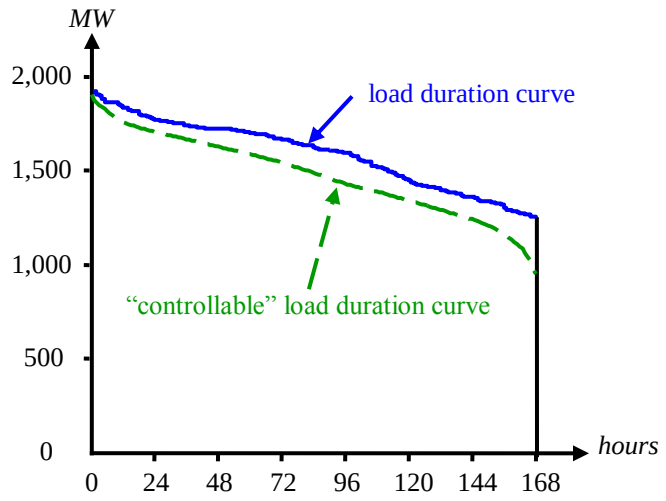


Figure 5.1: The load and the "controllable" load duration curves for the Eastern test system and for the week of Monday, January 2, to Sunday, January 8, 2006.

By definition, the "controllable" load duration curve gives a measure of the number of hours for which the "controllable" load exceeds a certain load level during the simulation period. Throughout this chapter, we make extensive use of "controllable" load duration curves since we use them as inputs of the extended production simulation tool. It is therefore necessary to get a better understanding of such curves and determine how they are related to the load duration curve. As we have abstracted time out to construct both the load and the "controllable" load duration curves, the comparison of the two curves does not give any information on the time when the wind power production curtails the load and on the curtailments durations. Actually, the only information which may be obtained is the expected wind power production over the simulation period. In

Fig.5.2, we plot the duration curve of the power generated by a 300 MW wind farm located in Allegheny Ridge, for a one-week period. The area below the curve represents the expected wind energy production over a period of one week.[10] By construction, this area must be equal to the area between the two duration curves which we illustrate in Fig. 5.3.

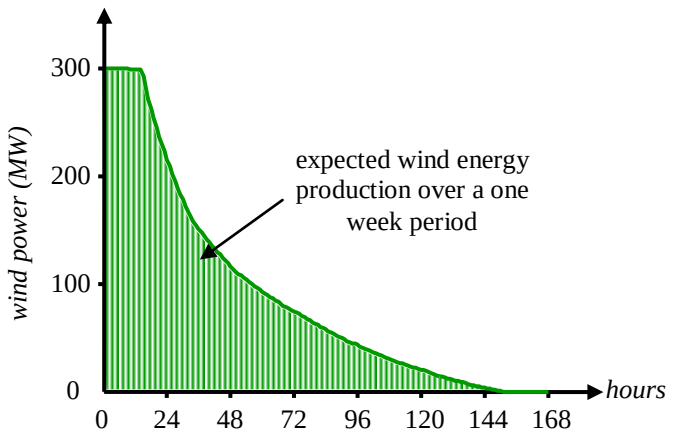


Figure 5.2: The duration curve of the wind power production of the 300 MW wind farm located in Allegheny Ridge, PA, for a one-week period.

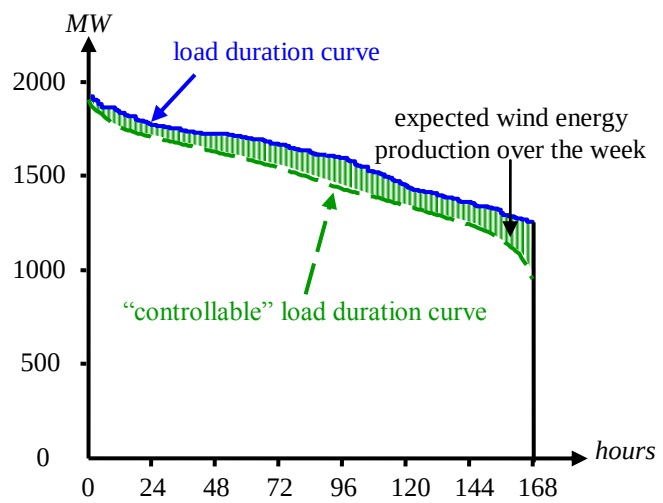


Figure 5.3: The expected wind energy production of the 300 MW wind farm located in Allegheny Ridge, PA.

[10] In this particular case, the expected wind energy production over a period of one week is equal to 15,295 MWh.

We used the "controllable" load duration curve plotted in Fig. 5.3 to run a production simulation for the Eastern test system and for the week of January 2-8, 2008. We summarize the production simulation results in Table 5.6 and we compare them with those obtained for the base case.

Table 5.6: Production simulation results for the Eastern test system and for the weekly period of Monday, January 2, to Sunday, January 8, 2006

metric	wind farm nameplate capacity (MW)	
	0	300
LOLP	$7.78\ 10^{-6}$	$3.46\ 10^{-6}$
expected unserved energy (MWh)	$2.25\ 10^{-2}$	$1.00\ 10^{-2}$
expected production costs ($)	$1.64\ 10^{7}$	$1.56\ 10^{7}$
expected CO_2 emissions (lbs)	$3.93\ 10^{8}$	$3.88\ 10^{8}$

Clearly, the integration of the 300 MW wind farm significantly increases the system reliability since both the LOLP and the expected unserved energy have been approximately divided by two. In addition, the more expensive and polluting resources have been displaced by the wind generation, resulting in a reduction of the system production costs and emissions.

It is also crucial to investigate the influence of the size of the wind resource on the variable effects of a system with integrated wind resources. We consider the Eastern test system with a wind farm located in Allegheny Ridge, PA, and whose nameplate capacity is equal to either 300 MW, 600 MW or 900 MW. We computed the "controllable" load duration curves for the systems with the three different wind resources and for the 52 weeks constituting the year 2006 to run production simulations. We summarize the

results of such a parametric study in Table 5.7. In Fig. 5.4, we observe the dependence of the LOLP (a), expected unserved energy (b), expected production costs (c) and expected CO_2 emissions (d) with respect to the increase of the wind power penetration.

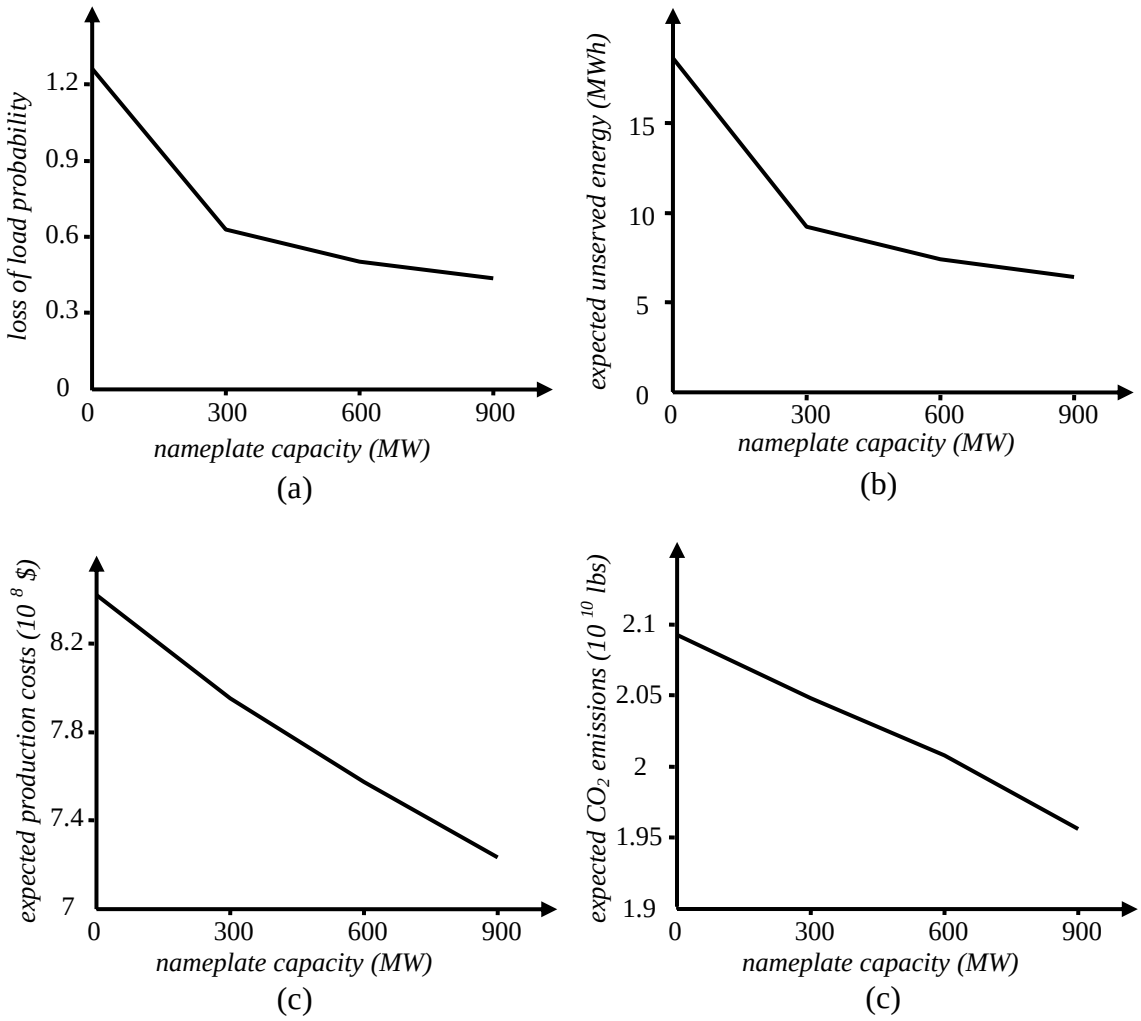


Figure 5.4: The dependence of the LOLP (a), expected unserved energy (b), expected production costs (c) and expected CO_2 emissions (d) for the Eastern test system with respect to the increase in the wind penetration.

Table 5.7: Production simulation results for the Eastern test system and for the year 2006 simulation period

metric	nameplate capacity (MW)			
	0	300	600	900
LOLP	$1.26\ 10^{-4}$	$6.26\ 10^{-5}$	$5.01\ 10^{-5}$	$4.35\ 10^{-5}$
expected unserved energy (MWh)	18.6	9.23	7.38	6.41
expected production costs ($)	$8.42\ 10^{8}$	$7.95\ 10^{8}$	$7.57\ 10^{8}$	$7.23\ 10^{8}$
expected CO_2 emissions (lbs)	$2.09\ 10^{10}$	$2.05\ 10^{10}$	$2.01\ 10^{10}$	$1.96\ 10^{10}$

In general, we notice a decrease in the system expected production costs and expected emissions as well as the increase in the reliability. To quantitatively assess the benefits obtained with the integration of wind resources, we compare the variable effects computed for the three different wind penetrations with those obtained for the base case. We list such figures of merit in Table 5.8.

Table 5.8: Figures of merit of the wind farm integrated into the Eastern test system for three wind penetrations

figure of merit	nameplate capacity (MW)		
	300	600	900
wind penetration	10.5	21	31.5
wind energy production (% of total demand)	5.8	11.6	17.4
LOLP reduction (%)	50.3	60.2	65.5
expected unserved energy reduction (%)	50.4	60.3	65.5
production costs reduction (%)	5.6	10.1	14.1
CO_2 emission reduction (%)	1.9	3.8	6.2

We remark that the expected production costs and CO_2 emissions decrease in a linear fashion with respect to the increase in the wind penetration increase. In contrast, there is a strong inflection in the reliability increase as the penetration rises. Such an effect may be explained by the intermittency of the wind resource: no matter the penetration, there is always a fraction of the simulation period when there is no wind power production. During such periods, the wind resources do not contribute to the system reliability and the effects of the integration of additional wind resources are limited.

We performed an identical study for the Midwestern test system. We summarize the production simulation results in Tables 5.9 and 5.10. The results look similar to those obtained with considering the Eastern test system. Note that such results depend on the location where the wind farm is located. Such an issue is discussed in a further example.

Table 5.9: The production simulation results for the Midwestern test system and for the year 2006 simulation period

metric	nameplate capacity (MW)			
	0	300	600	900
LOLP	$1.66\ 10^{-4}$	$7.92\ 10^{-5}$	$6.06\ 10^{-5}$	$5.07\ 10^{-5}$
expected unserved energy (MWh)	24.4	11.7	8.93	7.46
expected production costs ($)	$8.94\ 10^{8}$	$8.46\ 10^{8}$	$8.08\ 10^{8}$	$7.71\ 10^{8}$
expected CO_2 emissions (lbs)	$2.17\ 10^{10}$	$2.13\ 10^{10}$	$2.09\ 10^{10}$	$2.04\ 10^{10}$

Table 5.10: Figures of merit of the wind farm integrated into the Midwestern test system for three wind penetrations

figure of merit	nameplate capacity (MW)		
	300	600	900
wind penetration	10.5	21	31.5
wind energy production (% of total demand)	5.6	11.2	16.8
LOLP reduction (%)	52.3	63.5	69.5
expected unserved energy reduction (%)	52.0	63.4	69.4
production costs reduction (%)	5.4	9.6	13.8
CO_2 emission reduction (%)	1.8	3.7	6.0

In our work, we have represented the increase in the reserve levels by changing the loading order of the controllable units. Such a modification results in higher production costs since the units are no longer loaded from the cheapest to the most expensive. So we need to assess the extent to which the additional reserves impact the system production costs. We used the "controllable" load duration curves computed in the previous example to run production simulations for the systems without increased reserve levels and compare the results with those obtained in the previous examples where the change in the reserve levels was taken into account. In this way, we quantified the production costs raise incurred by the increased reserves levels so as to mitigate the impacts of the wind power variability and intermittency. We summarize the results we obtain in Table 5.11 and observe that, even though the increase in the production costs remains limited, it is not negligible and must be taken into account, especially for high wind penetrations.

Table 5.11: The production costs for the Eastern and Midwestern test systems with and without reserves for different levels of wind penetration

wind farm nameplate capacity (MW)	test system					
	Midwestern			Eastern		
	reserves		increase (%)	Reserves		increase (%)
	no	yes		no	yes	
300	$8.40\ 10^8$	$8.46\ 10^8$	0.7	$7.89\ 10^8$	$7.95\ 10^8$	0.8
600	$7.95\ 10^8$	$8.08\ 10^8$	1.6	$7.45\ 10^8$	$7.57\ 10^8$	1.6
900	$7.51\ 10^8$	$7.71\ 10^8$	2.7	$7.04\ 10^8$	$7.23\ 10^8$	2.7

In the previous studies, we adopted a wind speed model with a single regime. In the next example, we consider the Eastern test system with a 300 MW wind farm located in Allegheny Ridge. To probabilistically represent the wind speed, we used the *k*-means scheme to identify three wind speed regimes whose average daily patterns are plotted in Fig. 4.9. For each of the 52 weeks of the year 2006, we determined the distribution functions of the "controllable" load conditionally on the three regimes. In Fig. 5.5, we plot the load duration curve corresponding to the week of January 2-8, 2006, along with the conditional "controllable" load duration curves, the conditioning being on $\mathscr{R}'_1$ (a), $\mathscr{R}'_2$ (b) and $\mathscr{R}'_3$ (c).

With the regimes-based wind speed representation, we implicitly take into account the seasonality of the wind power production, as shown in Section 3.2. To illustrate this salient feature of the regimes approach, we partition the year 2006 into four quarters – denoted by Q_i, $i = 1, 2, 3, 4$ – of 13 weeks each. We used the set of

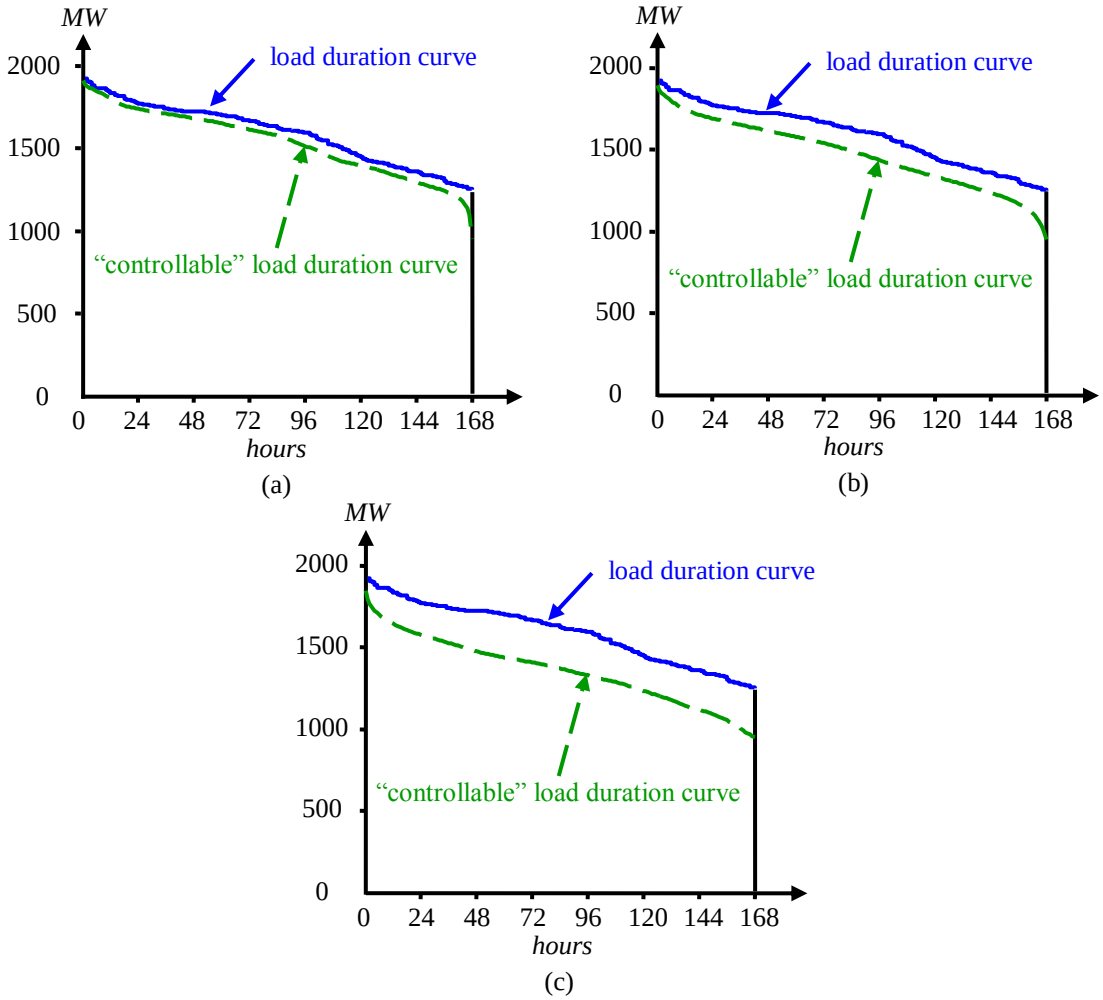


Figure 5.5: The duration curves of the "controllable" load of the Eastern test system conditionally on regime $\mathscr{R}'_1$ (a), regime $\mathscr{R}'_2$ (b) and regime $\mathscr{R}'_3$ (c) for the weekly period of Monday, January 2, to Sunday, January 8, 2006.

"controllable" load duration curves to compute the variable effects of the system for each quarter and conditionally on the wind speed regime. The weighted average of the results we obtained provides us with the expected value of the variable effects for each quarter. Note that the weighted average is computed making use of the frequency of occurrence of each regime within the quarter under consideration. Such coefficients have been identified from the clustering results and we list them in Table 5.12. We also present the

variable effects of the Eastern test system for the first quarter and three regimes as well as the weighted average in Table 5.13.

Table 5.12: Frequency of occurrence of the three wind speed regimes identified at the Allegheny Ridge location, for the four quarters of the year 2006

quarter	regime		
	$\mathcal{R}'_1$	$\mathcal{R}'_2$	$\mathcal{R}'_3$
Q_1	0.20	0.44	0.36
Q_2	0.56	0.42	0.02
Q_3	0.54	0.45	0.01
Q_4	0.21	0.55	0.24

Table 5.13: Variable effects of the Eastern test system for the three wind speed regimes and for the quarter Q_1 of the year 2006

metric	regime (probability)			weighted average
	$\mathcal{R}'_1$ (0.20)	$\mathcal{R}'_2$ (0.44)	$\mathcal{R}'_3$ (0.36)	
LOLP	$1.07\ 10^{-5}$	$5.77\ 10^{-6}$	$1.77\ 10^{-6}$	$1.3\ 10^{-6}$
expected unserved energy (MWh)	0.40	0.22	0.067	0.20
expected production costs ($)	$2.07\ 10^{8}$	$1.99\ 10^{8}$	$1.86\ 10^{8}$	$1.95\ 10^{8}$
expected CO_2 emissions (lbs)	$4.95\ 10^{9}$	$4.86\ 10^{9}$	$4.70\ 10^{9}$	$4.82\ 10^{9}$

The analysis of the results presented in Table 5.13 shows the large disparity of the production simulation results we obtain depending on the wind regime under consideration. For instance, the LOLP computed when considering the regime $\mathcal{R}'_1$ is around 6 times higher than that obtained when we considered $\mathcal{R}'_3$. Similarly, the

expected production costs and CO_2 emissions are respectively 11% and 5% higher, so the shape of the daily wind speed pattern highly impacts the system economics and reliability. In a similar way, we computed the variable effects of the Eastern test system for the three other quarters of the year 2006: Q_2, Q_3 and Q_4. We used the estimated probability of occurrence of the three regimes within each quarter – listed in Table 5.12 – to compute the variable effects of the system in each of the four quarters of the year. We show the results in Table 5.14.

Table 5.14: Variable effects of the Eastern test system for the four quarters of the year 2006

metric	quarter			
	Q_1	Q_2	Q_3	Q_4
LOLP	$1.3\ 10^{-6}$	$1.74\ 10^{-5}$	$2.63\ 10^{-4}$	$3.81\ 10^{-6}$
expected unserved energy (MWh)	0.20	0.64	9.70	0.14
expected production costs ($)	$1.95\ 10^{8}$	$1.90\ 10^{8}$	$2.22\ 10^{8}$	$1.87\ 10^{8}$
expected CO_2 emissions (lbs)	$4.82\ 10^{9}$	$5.19\ 10^{9}$	$5.64\ 10^{9}$	$4.87\ 10^{9}$

We observe huge differences between the values of the metrics computed for the four quarters of the year 2006. For instance, the LOLP is approximately 200 times higher during Q_3 than during Q_1, so the vast majority of the loss of load events of the test system must occur during Q_3. Such a result is coherent since the third quarter corresponds to the summer months with higher loads. The analysis of the expected unserved energy metric leads to an identical conclusion. We also remark that the expected production costs are highly seasonal since the expected production costs are around 16% lower in Q_4 than in Q_3.

We now investigate the impacts of the geographical variability of the wind speed on the variable effects of the Midwestern test system with integrated wind resources. We performed production simulations with considering the same 300 MW wind farm located at one of the four sites listed in Table 5.1. The analysis of the results – detailed in Table 5.15 – indicates that the wind resource location greatly influences the system variable effects: clearly, the "value" of a wind farm depends on its location. For instance, the 300 MW wind farm located in Harvest, MI, has a 26.6% capacity factor while that of the same wind resource located in Fenton, MN, is 61% higher. Also, the loss of load probabilities of the two systems with wind resources at the two locations significantly differ: it is 23% higher for the test system with wind resources located in Fenton, MN. Similar remarks may be made concerning the other variable effects. The extended production tool may therefore be used to determine the optimal location for a wind resource so that it brings more reliability/economic benefits to the actual system.

To manage the variability associated with the wind power production, system operators typically disperse the wind farms at various locations within the system region. In this way, we take advantage of the wind speed geographical variability, resulting in a wind power production with smoother variations and a less pronounced intermittency effect. As a result, the diversification of the wind farm locations increases the system reliability. In the present example, we quantify the reliability benefits obtained with the dispersion of wind farms in two particular cases:

- Case A: We consider four 75 MW wind farms – composed of 50 GE turbines each – located at the four Midwestern sites.

- Case B: We consider four 66 MW wind farms – composed of 44 GE turbines each – located at the four Midwestern sites. The expected wind energy production over the year 2006 is the same as that of the 300 MW located in Camp Grove.

In Table 5.16, we compare the loss of load probability and the expected unserved energy computed for these two cases with those computed when considering a single 300 MW wind farm in Camp Grove, IL. We remark that in both cases, the reliability metrics have been significantly reduced by the dispersion of the wind farm locations.

Table 5.15: The impact of the wind speed geographical diversity on the variable effects of the Midwestern test system

metric	wind farm location			
	Camp Grove, IL	Langdon, ND	Fenton, MN	Harvest, MI
capacity factor	31.2	41.1	42.8	26.6
wind energy production (% of total demand)	5.6	7.4	7.7	4.8
LOLP	$7.92\ 10^{-5}$	$6.55\ 10^{-5}$	$6.26\ 10^{-5}$	$8.12\ 10^{-5}$
expected unserved energy (MWh)	11.7	9.63	9.22	11.95
expected production costs ($)	$8.46\ 10^{8}$	$8.32\ 10^{8}$	$8.29\ 10^{8}$	$8.52\ 10^{8}$
expected CO_2 emissions (lbs)	$2.13\ 10^{10}$	$2.10\ 10^{10}$	$2.10\ 10^{10}$	$2.13\ 10^{10}$

Table 5.16: The impact of the diversification of the wind farm locations on the Midwestern test system reliability

metric	case	
	A	B
LOLP	$6.04\ 10^{-5}$	$6.57\ 10^{-5}$
expected unserved energy (MWh)	8.88	9.67

In this section, we have detailed the steps of the production simulation procedure and shown how to compute the variable effects of systems with integrated wind resources. We have illustrated the impacts of the wind speed temporal and geographical variability and intermittency on the variable effects of systems with wind farms.

5.4 The Effective Load Carrying Capability of a Wind Farm

Another application of the extended probabilistic production simulation tool is the evaluation of the effective load carrying capability – ELCC – of a wind resource. This notion has been widely described in the literature [43], [48]. It is defined as the load increment that a system can support after a change in its resource mix while maintaining the same reliability level. The ELCC is therefore a key metric we want to assess as it gives a capacity value for the wind resource and provides us with an indication of the reliability benefits resulting from such a modification in the resource mix. To illustrate the notion of effective load carrying capability, we computed the LOLP of the Eastern test system without wind resources and for various peak load levels. We also evaluated the LOLP of the system after a 300 MW wind farm had been integrated. We plot the two LOLP curves in Fig. 5.6 and observe that the 300 MW wind farm has a 103 MW ELCC.

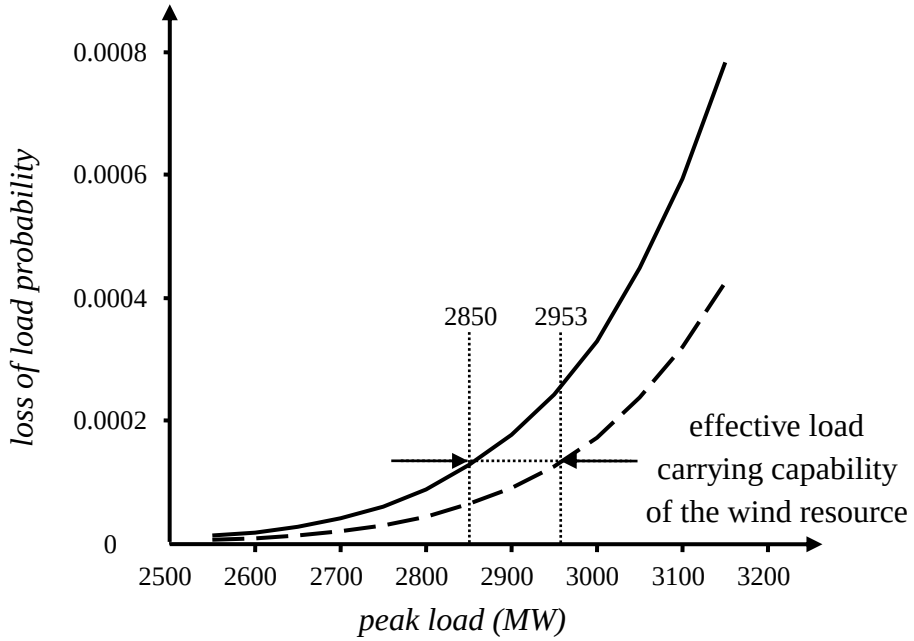


Figure 5.6: The effective load carrying capability of the 300 MW wind farm located in Allegheny Ridge, PA.

We now discuss a new simulation where we study the dependence of the effective load carrying capability of a wind resource with respect to the wind penetration. We used the methodology previously described to compute the ELCC of wind farms with different sizes. We express the results as a fraction of the wind farm nameplate capacity and plot them with respect to the wind penetration in Fig. 5.7. We notice that the effective load carrying capability of the wind resource decreases with respect to the wind penetration. In other words, the integration of a wind farm into a system with a low wind penetration is more beneficial – in terms of the reliability – than the integration of the same farm into a system which already has large wind resources.

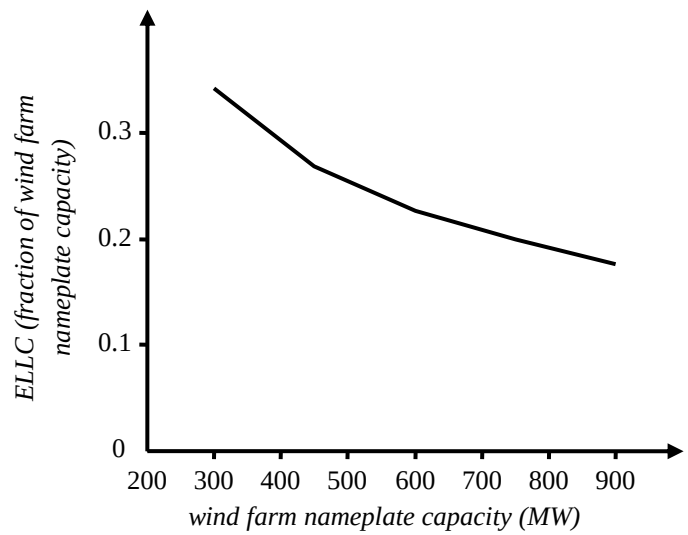


Figure 5.7: The effective load carrying capability of a wind farm with different sizes and expressed as a fraction of the wind farm nameplate capacity.

5.5 The Impacts of the Choice of Wind Technology

Wind energy production depends not only on the wind speeds at the wind farm location, but also on the technologies of the wind units constituting the wind resource. As a result, the integration of wind farms with the same nameplate capacity but different characteristics might impact the system differently. In this section, we make use of the wind farm model developed in Section 3.3 to assess the impacts of the wind farm constitution on the variable effects of the Eastern test system. We consider three wind farms with a 300 MW nameplate capacity and whose characteristics are given in Table 5.17. We ran a production simulation for the Eastern system with the three wind resources located in Allegheny Ridge, PA. The production simulation results are summarized in Table 5.18. The noticeable differences between the results obtained for each of the three wind farm configurations demonstrate the necessity to employ the

proposed wind farm model to get accurate production simulation results when considering wind farms constituted of wind turbines of different technologies.

Table 5.17: Constitution of the three different wind farms

turbine model	wind farm constitution		
	A	B	C
GE	400	200	200
Gamesa	0	150	75
Repower	0	0	75

Table 5.18: Production simulation results for the Eastern test system with integrated wind resources in Allegheny Ridge, PA, for the three wind farm constitutions

metric	wind farm constitution		
	A	B	C
LOLP	$5.01\ 10^{-5}$	$5.57\ 10^{-5}$	$5.14\ 10^{-5}$
expected unserved energy (MWh)	7.38	8.20	7.57
expected production costs ($)	$7.57\ 10^{8}$	$7.75\ 10^{8}$	$7.62\ 10^{8}$
expected CO_2 emissions (lbs)	$2.01\ 10^{10}$	$2.04\ 10^{10}$	$2.02\ 10^{10}$

In this chapter, we have focused our attention on the impacts of the wind speed and energy temporal variability on the system. We have shown to what extent the integration of wind resources increases the reliability and reduces the production costs and emissions of a power system. We have also quantified the increase of the system production costs that result from the increase in the system reserve levels to manage the variability and intermittency of the wind resource.

6. CONCLUSION

In this chapter, we summarize the work presented in this thesis and identify directions for future work.

6.1 Summary

This thesis presents the development and testing of a planning tool to quantify the variable effects of systems with integrated wind resources over longer-term periods. Throughout the thesis, we detail the successive steps that are necessary to extend the widely used probabilistic production simulation tool. We also present some of the results we obtained from extensive testing so as to demonstrate the capabilities of the tool and show the usefulness of our work.

We started with identifying the nature of the wind speed and illustrated its variability, randomness and lack of predictability making use of historical wind data. We pointed out that the wind power production obtained at the output of a wind unit directly depends on the wind speed at the turbine location and has therefore the same salient characteristics. On top of that, due to the way the wind turbines are operated, the wind power production is also intermittent. We explained how such characteristics complicate the operations of power systems with wind resources. We thoroughly described these impacts and listed the techniques that have been identified and implemented to effectively manage the wind power uncertainty and intermittency.

We used the insights we gained to develop a wind speed model that appropriately represents the salient characteristics of the wind speed. As the shape of the diurnal wind

power production pattern impacts the system scheduling outcomes, the modeling approach must reflect the variability of the daily wind power production patterns. The identification of wind speed regimes and the construction of the corresponding regime-based generation output model effectively capture such effects. We described the construction of the wind speed regimes making use of two widely used statistical clustering techniques: the hierarchical and the *k*-means clustering schemes. We pointed out their differences and provided some directions for their effective use.

We extended the probabilistic production simulation tool for systems with integrated wind resources. The principal challenge here was to mesh the production simulation framework with the representation of the wind power. In order to do so, we reexamined the load model so that its level of detail is adequate with that of the wind power model. We described in detail all the necessary steps that we undertook to modify the tool and we illustrated them with a concrete example.

The extensive simulations we conducted were useful to get a better understanding of the impacts of the integration of wind resources on the system economics, emissions and reliability. We demonstrated the implications of the wind temporal and geographic variability on the system variable effects. We showed the extent to which the variability and intermittency of the wind power production may be managed. Indeed, we quantified the reliability benefits obtained with the dispersion of the wind resources. We also investigated the economic impacts of the necessary increase of the reserve levels. We performed the simulations for several configurations of the wind resources and analyzed the sensitivity of the results with respect to the wind penetration.

6.2、 Future Research

The extension of the probabilistic production simulation tool constitutes a major improvement to study systems with wind resources. However, the variability and intermittency effects are also typically managed with the integration of storage devices into the grid. Over the long-term planning horizon, the big issue to address regarding storage is to determine optimal sizes for the storage devices. The determination of the optimal sizes can be done with a production simulation tool that can appropriately take into account the wind variability impacts on production costs. Today's probabilistic simulation tools have the capability to study storage technology to represent pumped and compressed air energy storage devices and certain exchange contracts. We will ensure that the modifications introduced into the probabilistic production costing tool will not perturb such capability. In this way, we will quantify the benefits of using storage devices to harness wind resources more effectively in a planning context.

APPENDIX A: CLUSTERING ALGORITHMS

We provide a formal statement of the two clustering algorithms – the hierarchical and the *k*-means clustering schemes – we use in grouping the similar days of wind speed patterns. We use the notation of Section 3.1 throughout this appendix.

A.1 The Hierarchical Clustering Algorithm

% initialization phase

- set the iteration counter to zero: $u = 0$
- build the initial classes:

  $\mathcal{R}_d^{(u)} = \left\{ \underline{v}^{(d)} \right\},\ d \in \{ 1,\ 2,\ \ldots, D \}$.

- construct the upper triangular part of the "closeness matrix" which we call: $\underline{\Delta Z}^{(u)}$.

while (number of classes $\neq k$)

- construct the set

  $\mathcal{Z} = \left\{ (x,\ y) : \left(\underline{\Delta Z}^{(u)} \right)_{x,\, y} \leq \left(\underline{\Delta Z}^{(u)} \right)_{i,\, j},\ i = 1, 2, \ldots, D-u,\ \ j > i \right\}$

- if $\mathcal{Z}$ is a singleton, set:

  $(\bar{x},\ \bar{y}) = (x,\ y)$;

  else pick the element $(\bar{x},\ \bar{y})$ with $\bar{x} + \bar{y}$ strictly less than $x' + y'$ of any other element $(x',\ y') \in \mathcal{Z}$

- increment the counter: $u = u + 1$
- merge the classes with indices $\bar{x}$ and $\bar{y}$:

  $\mathcal{R}_{min\{\,\bar{x},\ \bar{y}\,\}}^{(u)} = \mathcal{R}_{\bar{x}}^{(u-1)} \bigcup \mathcal{R}_{\bar{y}}^{(u-1)}$

- update the indices of all other classes:

  $\forall w \ni max\{\,\bar{x},\ \bar{y}\,\} \leq w \leq D-(u-1)$ set $\mathcal{R}_w^{(u)} = \mathcal{R}_{w+1}^{(u-1)}$

  $\forall w \ni w < max\{\,\bar{x},\ \bar{y}\,\},\ w \neq min\{\,\bar{x},\ \bar{y}\,\}$ set $\mathcal{R}_w^{(u)} = \mathcal{R}_w^{(u-1)}$

- compute the new "closeness" matrix: $\underline{\Delta Z}^{(u)}$

end

A.2 The *k*-means Algorithm

% initialization phase

- set the iteration counter to zero: $u = 0$
- pick a set of initial class "centers": $\left\{ \underline{a}_j^{(u)} \right\}, \; j = 1, \ldots, \; N$
- construct the corresponding classes: $\left\{ \mathcal{R}'^{(u)}_j \right\}, \; j = 1, \ldots, \; N$ with:

$$\mathcal{R}'^{(u)}_j = \left\{ \underline{v}^{(d)} : \forall l \neq j, \; \mu\left(\underline{v}^{(d)}, \; \underline{a}_j^{(u)} \right) \leq \mu\left(\underline{v}^{(d)}, \; \underline{a}_l^{(u)} \right) \right\}$$

while ($\forall \; j = 1, \; 2, \ldots, k, \; \mathcal{R}'^{(u)}_j \neq \mathcal{R}'^{(u-1)}_j \; or \; u = 0$)

- increment: $u = u + 1$
- compute the new classes "centers":

$$\underline{a}_j^{(u)} = \frac{1}{\left| \mathcal{R}'^{(u-1)}_j \right|} \sum_{\underline{v}^{(d)} \in \mathcal{R}'^{(u-1)}_j} \underline{v}^{(d)}$$

- construct the corresponding classes $\mathcal{R}_j^{(u)}$ such that:

$$\mathcal{R}'^{(u)}_j = \left\{ \underline{v}^{(d)} : \forall l \neq k, \; \mu\left(\underline{v}^{(d)}, \; \underline{a}_k^{(u)} \right) \leq \mu\left(\underline{v}^{(d)}, \; \underline{a}_l^{(u)} \right) \right\}$$

end

APPENDIX B: GEOGRAPHICALLY DISPERSED WIND SPEED CORRELATION

In this appendix, we study the correlation between the wind speeds at two distinct locations within a power system. In particular, we observe the evolution of such correlations with respect to the distance between the two locations under consideration. Throughout this appendix, we use the same notation as in Chapter 3.

The correlation coefficient ρ between the rv's $\underset{\sim}{V}_{s,r}^{(h)}$ and $\underset{\sim}{V}_{s',r}^{(h)}$, $s, s' \in \mathfrak{S}$ is computed in the following way:

$$\rho\left(\underset{\sim}{V}_{s,r}^{(h)}, \underset{\sim}{V}_{s',r}^{(h)}\right) = \frac{cov\left(\underset{\sim}{V}_{s,r}^{(h)}, \underset{\sim}{V}_{s',r}^{(h)}\right)}{\sqrt{var\left(\underset{\sim}{V}_{s,r}^{(h)}\right).var\left(\underset{\sim}{V}_{s',r}^{(h)}\right)}}.$$

By definition, $-1 \leq \rho \leq 1$. In the above equation, the terms $cov\left(\underset{\sim}{V}_{s,r}^{(h)}, \underset{\sim}{V}_{s',r}^{(h)}\right)$, $Var\left(\underset{\sim}{V}_{s,r}^{(h)}\right)$ and $Var\left(\underset{\sim}{V}_{s',r}^{(h)}\right)$ are estimated from the wind speed data.

We consider the seven sites spread over the MISO system footprint and mapped in Fig. 3.2. The distances between any two of the seven locations are listed in Table B.1. We identify three wind speed regimes making use of the *k*-means clustering scheme. We collect the wind speed data corresponding to the site s, wind speed regime r and hour h of the day so as to construct the data sets $\mathcal{V}_{s,r}^{(h)} = \left\{ v_{s,d}^{(h)} : \underline{v}_d \in \mathcal{R}'_r \right\}$ constituted of the realizations of the rv's $\underset{\sim}{V}_{s,r}^{(h)}$. We make use of the two data sets $\mathcal{V}_{s,r}^{(h)}$ and $\mathcal{V}_{s',r}^{(h)}$ to compute the correlation coefficient ρ between the rv's $\underset{\sim}{V}_{s,r}^{(h)}$ and $\underset{\sim}{V}_{s',r}^{(h)}$, $s, s' \in \mathfrak{S}$. Note that for each pair of sites, we compute a total of $(3)(24)$ correlation coefficients.

Table B.1: The distances between the wind farm sites located within the MISO footprint

distance (km)	Camp Grove	Fenton	Adair	Blue Sky	Concep-tion	Langdon	Harvest
Camp Grove		350	260	200	270	565	390
Fenton			170	360	260	210	620
Adair				370	90	370	610
Blue Sky					410	520	250
Conception						530	640
Langdon							770
Harvest							

We list the values of the correlation coefficients in Table B.2 corresponding to a few pairs of sites and for the three regimes. Due to the large size of the results, we do not provide the correlation coefficients for the 24 hours of the day but rather give the minimum, the maximum and the mean ones. There is no exact interpretation of the values of the correlation coefficient and we interpret them in the following way: the correlation is considered to be low if $|\rho| \leq 0.3$. It is considered to be moderate if $0.3 \leq |\rho| \leq 0.7$ and high in the remaining cases. From the results listed in Table B.2, we conclude that for two sufficiently distant locations (> 250-300 km), the correlation coefficients are either low or in the lower part of the moderate range. In contrast, for two close locations (< 200 km), the correlation is either moderate or even high.

Table B.2: The correlation coefficients of the wind speed data

site s site s' (distance - km)	hourly correlation	regime 1	regime 2	regime 3
Langdon Harvest (770)	min	-0.12	-0.05	-0.04
	mean	0.05	0.07	0.07
	max	0.30	0.21	0.18
Langdon Conception (530)	min	-0.35	-0.18	-0.07
	mean	-0.13	-0.07	0.04
	max	0.26	0.15	0.14
Camp Grove Harvest (390)	min	-0.14	0.07	-0.03
	mean	0.03	0.21	0.08
	max	0.22	0.36	0.24
Camp Grove Fenton (350)	min	-0.16	-0.20	-0.06
	mean	-0.02	-0.02	0.14
	max	0.14	0.17	0.37
Camp Grove Conception (270)	min	0.18	0.03	0.30
	mean	0.32	0.22	0.39
	max	0.49	0.38	0.52
Harvest Blue Sky (250)	min	-0.05	0.23	0.04
	mean	0.14	0.37	0.30
	max	0.38	0.52	0.50
Camp grove Blue Sky (200)	min	0.26	0.18	0.37
	mean	0.44	0.30	0.51
	max	0.64	0.40	0.65
Adair Conception (90)	min	0.58	0.50	0.64
	mean	0.73	0.61	0.77
	max	0.88	0.71	0.87

REFERENCES

[1] DSIRE – Database of State Incentives for Renewables & Efficiency. [Online]. Available: www.dsireusa.org/

[2] Spanish Wind Energy Association, "Wind power 2008," 2008. [Online]. Available: http://www.aeeolica.es/userfiles/file/aee-publica/AEE_anuario_ingles 08.pdf

[3] "ERCOT operations report on the EECP event of February 26, 2008," 2008. [Online]. Available: www.ercot.com/content/meetings/ros/keydocs/2008/0313/ 07._ERCOT_OPERATIONS _REPORT_EECP022608_public.doc

[4] E. Ela and B. Kirby, "ERCOT Event on February 26, 2008: Lessons learned," NREL, CO, Tech. Rep. NREL/TP-500-43373, July 2008. [Online]. Available: http://www.nrel.gov/wind/systemsintegration/pdfs/2008/ela_ercot_event.pdf

[5] "ERCOT operations report on the EECP event of february 24, 2007," 2007. [Online]. Available: www.ercot.com/content/meetings/ros/keydocs/2007/0315/ 07._ERCOT_OPERATIONS_REPORT_EECP022407_final.doc

[6] G. Masters, *Renewable and Efficient Electric Power Systems*. New York, NY: John Wiley & Sons, 2007.

[7] B. Brown, R. Katz, and A. H. Murphy, "Time series models to simulate and forecast wind speed and wind power," *Journal of Climate and Applied Meteorology*, vol. 23, no. 8, pp. 1184-1195, August 1984.

[8] J. Carlin and J. Haslett, "The probability distribution of wind power from a dispersed array of wind turbine generators," *Journal of Applied Meteorology*, vol. 21, no. 3, pp. 303-313, March 1982.

[9] M. S. Miranda and R. W. Dunn, "Spatially correlated wind speed modeling for generation adequacy studies in the UK," in *IEEE Power Engineering Society General Meeting, 2007*, 24-28 June 2007, pp. 1-6.

[10] M. Gibescu, B. C. Ummels, and W. L. Kling, "statistical wind speed interpolation for simulating aggregated wind energy production under system studies," in *9th International Conference on Probabilistic Methods Applied to Power Systems*, KTH, Stockholm, Sweden, June 11-15, 2006, vol.2, pp. 1101-1107.

[11] G. Papaefthymiou and B. Klöckl, "MCMC for wind power simulation," *IEEE Transactions on Energy Conversion*, vol. 23, no. 1, pp. 234-240, March 2008.

[12] J. Mur-Amada and A. Bayold-Rujula, "Wind power variability model Part I – Foundations," in *9th International Conference, Electrical Power Quality and Utilization*, Barcelona, Spain, October 2007, pp. 485-490.

[13] M. Doquet, "Use of a stochastic process to sample wind power curves in planning studies," in *2007 IEEE Lausanne Power Tech*, 1-5 July 2007, pp. 663-670.

[14] S. H. Karaki, R. B. Chedid, and R. Ramadan, "Probabilistic performance assessment of wind energy conversion systems," *IEEE Transactions on Energy Conversion*, vol. 14, no. 2, pp. 217-224, June 1999.

[15] G. Giebel, R. Brownsword, and G. Kariniotakis, "The state-of-the-art in short-term prediction of wind power – A literature overview," Anemos, Report D1.1, 2003. [Online]. Available: http://anemos.cma.fr/download/ ANEMOS_D1.1_ StateOfThe Art_v1.1.pdf

[16] B. Ernst, B. Oakleaf, M. L. Ahlstrom, M. Lange, C. Moehrlen, B. Lange, U. Focken, and K. Rohrig, "Predicting the wind," *IEEE Power and Energy Magazine*, vol. 5, no. 6, pp. 78-89, Nov.-Dec. 2007.

[17] Y. Wu and J. Hong, "A literature review of wind forecasting technology in the world," in *2007 IEEE Lausanne Power Tech*, 1-5 July 2007, pp. 504-509.

[18] Y. Makarov, D. Hawkins, E. Leuze, and J. Vidov, "California ISO wind generation forecasting service design and experience," 2002. [Online]. Available: http://www.caiso.com/docs/2002/06/12/2002061211030528205.pdf

[19] M. Lange, "On the uncertainty of wind power predictions – Analysis of the forecast accuracy and statistical distribution of errors," *Journal of Solar Energy Engineering*, vol. 127, pp. 177-184, 2005.

[20] A. Fabbri, T. Gómez San Román, J. Rivier Abbad, and V. Méndez Quezada, "Assessment of the cost associated with wind generation prediction errors in a liberalized electricity market," *IEEE Transactions on Power Systems*, vol. 20, no. 3, pp. 1440-1446, August 2005.

[21] H. Bludszweit, J. A. Dominguez-Navarro, and A. Lombart, "Statistical analysis of wind power forecast error," *IEEE Transactions on Power Systems*, vol. 23, no. 3, pp. 983-991, August 2008.

[22] American Wind Energy Association, "Wind energy basics," 2009. [Online]. Available: www.awea.org

[23] M. Amelin, "Comparison of capacity credit calculation methods for conventional power plants and wind power," *IEEE Transactions on Power Systems*, vol. 24, no. 2, pp. 685-691, May 2009.

[24] R. Doherty and M. O'Malley, "A new approach to quantify reserve demand in systems with significant installed wind capacity," *IEEE Transactions on Power Systems*, vol. 20, no. 2, pp. 587-595, May 2005.

[25] D. Brooks, E. Lo, R. Zavadil, S. Santoso, and J. Smith, "Characterizing the impacts of significant wind generation facilities on bulk power system operations planning," Xcel Energy – North Case Study, Final Report, Prepared for Utility Wind Interest Group, May 2003. [Online]. Available: http://www.uwig.org/ UWIGOpImpacts Final7-15-03.pdf

[26] R. Piwko, X. Bai, K. Clark, G. Jordan, N. Miller, and J. Zimberlin, "The effects of integrating wind power on transmission system planning, reliability, and operations," March 4, 2005. [Online]. Available: http://www.nyserda.org/public ations/wind_integration_report.pdf

[27] Eastern Wind Integration and Transmission Study – NREL data. [Online]. Available: http://wind.nrel.gov/public/EWITS/

[28] S. Kennedy and P. Rogers, " A probabilistic model for simulating long-term wind-power output," *Wind Engineering*, vol. 27, no. 3, pp. 167-181, 1 May 2003.

[29] Danish Wind Industry Association website. [Online]. Available: http://www.windpower.org/en/tour/wres/variab.htm

[30] Canadian Wind Energy Association, "An analysis of the impacts of large-scale wind generation on the Ontario electricity system," April 2005. [Online]. Available: http://www.uwig.org/IESO_Study_final_document1.pdf

[31] J. McDowall, "Integrating energy storage with wind power in weak electricity grids," *Journal of Power Sources*, vol. 162, no. 2, pp. 959-964, 22 November 2006.

[32] G. Czisch and B. Ernst," High wind power penetration by the systematic use of smoothing effects within huge catchment areas shown in a European example," 2001. [Online]. Available: http://www.iset.uni-kassel.de/abt/w3-w/projekte/awea_ 2001_czisch_ernst.pdf

[33] M. Milligan, K. Porter, B. Parsons and J. Caldwell, "Wind energy and power system operations: A survey of current research and regulatory actions," *The Electricity Journal*, vol. 15, no. 2, pp. 56-67, March 2002.

[34] ISO / RTO Council, "Increasing renewable resources: How ISOs and RTOs are helping meet this public policy objective," October 16, 2007. [Online]. Available: http://www.isorto.org/atf/cf/%7B5B4E85C6-7EAC-40A0-8DC3-003829518EBD%7D/IRC_Renewables_Report_101607_final.pdf

[35] J. C. Smith, M. R. Milligan, E. A. DeMeo, and B. Parsons, "Utility wind integration and operating impact state of the art," *IEEE Transactions on Power Systems*, vol. 22, no. 3, pp. 900-908, August 2007.

[36] G. Gan, C. Ma, and J. Wu, *Data Clustering: Theory, Algorithms, and Applications*. Philadelphia, PA: SIAM, 2007.

[37] US Wind Energy Projects – AWEA website. [Online]. Available: http://www.awea.org/projects/

[38] Q. Nguyen, "Machine learning: Algorithms and applications," 2008. [Online]. Available: http://www.inf.unibz.it/~qnhatnguyen/mlearning/slides/Clustering.pdf

[39] V. M. Gomez-Munoz and M. A. Porta-Gandara, "Local wind patterns for modeling renewable energy systems by means of cluster analysis techniques," *Renewable Energy*, vol. 25, no.2, pp. 171-182, 2002.

[40] J. Twidell and T. Weir, *Renewable Energy Resources*, 2nd ed, New York, NY: Taylor & Francis, 2006.

[41] Wind Energy Market – German Wind Energy Association (BWE) website. [Online]. Available: http://www.wind-energy-market.com/

[42] K. Imai, "Expectations and functions of random variables," March 10, 2006. [Online]. Available: http://imai.princeton.edu/teaching/files/Expectation.pdf

[43] G. Gross, "Electricity Resource Planning", class notes for ECE 588, Department of Electrical and Computer Engineering, University of Illinois at Urbana-Champaign, Fall 2008.

[44] A. Wood and B. Wollenberg, *Power Generation, Operation and Control.* New York, NY: John Wiley & Sons, 1996.

[45] R. Sullivan, *Power System Planning*. New York, NY: McGraw-Hill, 1977.

[46] C. Grigg et al., "The IEEE reliability test system-1996. A report prepared by the reliability test system task force of the application of probability methods subcommittee," *IEEE Transactions on Power Systems*, vol. 14, no. 3, pp. 1010-1020, August 1999.

[47] R. Zavadil et al., "Final Report – 2006 Minnesota wind integration study," November 2006. [Online]. Available: http://www.uwig.org/windrpt_vol%201.pdf

[48] L. L. Garver, "Effective load carrying capability of generating units," *IEEE Transactions on Power Apparatus and Systems*, vol. PAS-85, no. 8, pp. 910-919, August 1966.