Quantifying the Impact of Investments on Distribution System Wind Resilience Using Area Outage Rate Curves

by

Arslan Ahmad

A thesis submitted to the graduate faculty

in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

Major: Electrical Engineering (Electric Power and Energy Systems)

Program of Study Committee:
Ian Dobson, Major Professor
Zhaoyu Wang
Alice Alipour

The student author, whose presentation of the scholarship herein was approved by the program of study committee, is solely responsible for the content of this dissertation/thesis. The Graduate College will ensure this dissertation/thesis is globally accessible and will not permit alterations after a degree is conferred.

Iowa State University

Ames, Iowa

2023

DEDICATION

I want to dedicate this thesis to my father. Without his prayers, support, and encouragement, I would not have been able to complete this work.

TABLE OF CONTENTS

Page

LIST OF TABLES v

LIST OF FIGURES vi

ACKNOWLEDGMENTS ix

ABSTRACT x

CHAPTER 1. OVERVIEW 1
1.1 Introduction 1
1.2 Research Objective 7
1.3 Organization of Thesis 7

CHAPTER 2. LITERATURE REVIEW 9
2.1 Background 9

CHAPTER 3. OUTAGES AND WEATHER DATA 15
3.1 Overview of datasets 15
3.2 Outages Data 15
3.2.1 Dataset Description 15
3.2.2 Outage Cause Codes 15
3.3 Weather Data 18
3.3.1 Dataset Description 18
3.3.2 Data Cleaning 18
3.3.3 Basic Data Statistics 19
3.3.4 Wind Speed Distribution 20

CHAPTER 4. AREA OUTAGE RATE CURVES 26
4.1 Overview 26
4.2 Area Outage Rate Curve 26
4.3 Area Outage Rate Curves and Fragility Curves 27
4.4 Derivation of Area Outage Rate Curves 29
4.5 Software Implementation 32
4.5.1 Spatial Association 32
4.5.2 Temporal Association 35
4.5.3 Joining the datasets 36
4.5.4 Outage Rate Curves 38
4.6 Curve Fitting of the Area Outage Rate Curves 39

4.7 Outage Cause Codes Analysis with Wind Data 40

CHAPTER 5. RESILIENCE INVESTMENTS EVALUATION 45
5.1 Overview 45
5.2 Resilience Metrics 46
5.2.1 AC Algorithm 49
5.3 Relationship between Resilience and Area Outage Rate Curves 49
5.3.1 Area Outage Rate Curve Shifting 50
5.4 Outages Sampling Design 51
5.4.1 Overall Hardening vs. Targeted Hardening 56
5.5 Changes in Events after sampling 56
5.5.1 Super Events 59
5.6 Results 61
5.6.1 Event Size distribution 63
5.6.2 Results with Event Size segregation 64

CHAPTER 6. SUMMARY AND DISCUSSION 68
6.1 Key Strengths and Contributions 68
6.2 Conclusions 70
6.3 Publications 72
6.4 Future Work Avenues 72

BIBLIOGRAPHY 74

APPENDIX A. ZONE-2 RESULTS 80

APPENDIX B. RESILIENCE METRICS DISTRIBUTIONS 86

LIST OF TABLES

Page

Table 3.1 Summary of the Outages dataset 16

Table 3.2 Details of the Weather Stations 19

Table 3.3 Summary of Missing, Duplicate and Suspect values in the Weather dataset . 20

Table 3.4 Summary of the Weather dataset 20

Table 3.5 Parameter estimates and confidence intervals of the Hybrid Weibull distribution of WS-1 23

Table 3.6 Statistics of the Hybrid Weibull distribution of WS-1 24

Table 4.1 Summary of the weather stations' assignment to outages 33

Table 4.2 Goodness of fit properties of fitted Area Outage Rate models 41

Table 5.1 Improvement in Resilience Metrics with 10% reduction in outages of Area-1 61

Table 5.2 Improvement in Resilience Metrics with 20% reduction in outages of Area-1 62

Table 5.3 Improvement in the Mean value of Resilience Metrics of Small Events after 10% reduction in outages of Area-1 64

Table 5.4 Improvement in the Mean value of Resilience Metrics of Medium Events after 10% reduction in outages of Area-1 65

Table 5.5 Improvement in the Mean value of Resilience Metrics of Large Events after 10% reduction in outages of Area-1 65

LIST OF FIGURES

Page

Figure 1.1 Global reported natural disasters by type, from 1970 to 2022 Our world in data (2023) 4

Figure 3.1 Outage Cause Code categories Bar Chart 17

Figure 3.2 Breakdown of the "Tree" and "Weather" Cause Code Categories 17

Figure 3.3 Histograms of WS-1 and WS-2 weather stations' wind speed data 22

Figure 3.4 CDF of the Hybrid Weibull Distribution of WS-1 weather station data . . . 23

Figure 3.5 Comparison of the CDFs of the Hybrid Weibull Distribution and the Two-parameter Weibull Distribution of WS-1 weather station data 24

Figure 3.6 Survival Function (CCDF) of wind speed data of the WS-1 weather station 25

Figure 4.1 Outage Rate function 30

Figure 4.2 Interpolation of wind speed data 31

Figure 4.3 Geographical location of power outages and the associated weather stations in two of the distribution system areas 34

Figure 4.4 Distribution of distance between outages and weather station in Area-1 . . . 35

Figure 4.5 Frequency of time differences between outages and wind speed observations of Area-1 36

Figure 4.6 Time series plot of wind speed and outages data of Area-1 37

Figure 4.7 Individual Outage Rates at each wind speed level in Area-1 38

Figure 4.8 Area Outage Rate Curve of Area-1 39

Figure 4.9 Log-Linear plot of the Area Outage Rate Curve of Area-1 40

Figure 4.10 Estimated Area Outage Rate Curve of Area-1 41

Figure 4.11 Percentage of outages in each cause code category of Area-1 42

Figure 4.12 Variation in the proportion of outage categories with wind speed in Area-1 . 43

Figure 4.13 Area Outage Rate Curves of Area-1 for Selected Outage Cause Categories . 44

Figure 4.14 Area Outage Rate Curves of Area-1 for Selected Outage Cause Categories . 44

Figure 5.1 Resilience Improvement Flow Chart 46

Figure 5.2 Example of a large size Resilience Event 48

Figure 5.3 Comparison of Original and Shifted Area Outage Rate Curve of Area-1 . . . 52

Figure 5.4 Timeline plot of a medium size resilience event 57

Figure 5.5 The probability of change in each resilience metric of events changed after outage sampling 59

Figure 5.6 Empirical Distribution of the Event Sizes 63

Figure 5.7 Comparison of Resilience Metrics with Event Size segregation 66

Figure A.1 Distribution of distance between outages and weather station in Area-2 . . . 80

Figure A.2 Frequency of time differences between outages and wind speed observations of Area-2 81

Figure A.3 Time series plot of wind speed and outages data of Area-2 81

Figure A.4 Individual Outage Rates at each wind speed level in Area-2 82

Figure A.5 Area outage rate curve of Area-2 82

Figure A.6 Log-Linear plot of the Area Outage Rate Curve of Area-2 83

Figure A.7 Estimated Area Outage Rate Curve of Area-1 83

Figure A.8 Percentage of outages in each cause code category of Area-2 84

Figure A.9 Variation in the proportion of outage categories with wind speed in Area-2 . 84

Figure A.10 Area Outage Rate Curves of Area-2 for Selective Outage Cause Categories . 85

Figure A.11 Area Outage Rate Curves of Area-2 for Selective Outage Cause Categories . 85

Figure B.1 Distribution of Mean Event Duration 86

Figure B.2 Distribution of Mean Event Restore Duration 87

Figure B.3 Distribution of Mean Time to First Restore 87

Figure B.4 Distribution of Mean No. Of Customers in Events 88

Figure B.5 Distribution of Mean Customer-Hours 88

Figure B.6 Distribution of Mean Area Under the Performance Curve (Component-Hours) 89

Figure B.7 Variation in Mean Event Size (before and after sampling) with the change in Event Size 90

Figure B.8 Variation in Mean Event Duration (before and after sampling) with the change in Event Size 91

Figure B.9 Variation in Mean Event Restore Duration (before and after sampling) with the change in Event Size 91

Figure B.10 Variation in Mean Time to First Restore (before and after sampling) with the change in Event Size 92

Figure B.11 Variation in Mean No. of Customers Out in an Event (before and after sampling) with the change in Event Size 92

Figure B.12 Variation in Mean Customer-Hours (before and after sampling) with the change in Event Size 93

Figure B.13 Variation in Mean Area Under Performance Curve (before and after sampling) with the change in Event Size 93

Figure B.14 Distribution of Event Size vs. Mean No. of Events that got Split after Sampling 94

ACKNOWLEDGMENTS

I want to take this opportunity to express my thanks and acknowledge the support of those who helped me with various aspects of conducting this research.

First of all, I would like to express my sincerest gratitude to my major professor, Dr. Ian Dobson, for his continuous guidance, encouragement, tremendous support, and patience throughout this research endeavor. This work would not have been possible without his support. I also want to thank my committee members, Dr. Zhaoyu Wang and Dr. Alice Alipour, for consenting to be on the Program of Study Committee and supporting me in this research.

I would like to especially acknowledge the funding and support received from the U.S. Government, the U.S. Department of State, IIE, and USEFP via the Fulbright program. I gratefully acknowledge support in part from NSF grants 2153163 and 2228757.

Moreover, I would like to thank my parents for their prayers and my friends and family for their consistent support during this journey. Finally, I would also like to thank the ECpE department faculty and staff for making the time at Iowa State University a wonderful and memorable experience.

ABSTRACT

Electricity is an essential requirement for modern societies, and its importance continues to grow with time. In order to facilitate economic growth and social harmony, uninterrupted access to electricity is critical. However, power outages are inevitable due to the wide geographical spread of power systems and their exposure to various weather elements. Outages in the distribution system are the most common and can last anywhere from a few hours to several days. Resilience analysis is vital to mitigate the impact of large and infrequent power outages, which are mostly caused by extreme weather events. Improving the resilience of distribution systems is crucial since most of the damage caused by extreme weather events is in the distribution system. Utility companies invest in these systems to make them more resilient. To study the impact of these investments on the resilience metrics of small, medium, and large events, we first create "Area Outage Rate Curves" that describe the impact of wind hazards on the outage rates in different areas of the distribution system. These curves are calculated using observed weather and outage data. We develop these curves by correlating the observed wind speeds at each weather station with the outage rates of the nearest distribution system area. Resilience events and their metrics of size and customer impact are extracted from the outage data. Decisions about improving the distribution system can be made based on these metrics. Investments in overhead lines shift the area outage rate curves, and we calculate the corresponding impacts on the resilience metrics of small, medium, and large-scale events. To test our approach, we utilize outage data from a distribution system and wind speed data from weather stations within that system. The developed techniques and the use of area outage rate curves can be applied to any distribution system with an outage management system and can assist in making informed decisions about improving the system's resilience.

CHAPTER 1. OVERVIEW

1.1 Introduction

The global electricity demand jumped to 24,700 terawatt-hours (TWh) in 2021, after an annual increase of 6% (1538 TWh), the largest ever annual increase in absolute terms IEA (2022a). It is projected to triple by 2050 with a 3-4% projected per-annum growth rate from 2019 to 2050 Insights (2022). According to the World Energy Outlook 2022 report by the International Energy Agency, electricity accounted for about 20% of the world's total final energy consumption in 2021. It is set to grow to 24% by 2040 if countries stay on their present course IEA (2022b). This shows the increasingly high dependence of the world on electricity and the importance of power systems. The electric power system is the backbone of any economy as every sector directly depends on a continuous supply of electricity, be it the health sector, education, industry, or defense. Therefore the continuous and reliable operation of the power system is even more important today than ever. Even a short electric power interruption can have drastic effects. A brief account of some of the recent major electric power failure events and their impacts is given in the following paragraph to signify the extreme importance of the reliable and resilient operation of power systems.

Rotating power outages due to a heatwave in California on 14 and 15 August 2020 left up to 3 million people without electricity for up to 2.5 hours across several days CAISO et al. (2021). Winter storm Uri caused power outages in Texas from 15 February to 17 February 2021, resulting in an average loss of 34,000 MW of generation (nearly half of ERCOT's all-time peak electric load of 69,871 MW), impacting more than 4.5 million people. Some people remained without power for as long as four days. At least 210 people died during the event due to causes mostly connected to the power outages FERC (2021). About 1.1 million customers in the United Kingdom lost electric power for up to 45 minutes when a lightning strike hit the GB transmission system on 9 August 2019. Though short-lived, the power outage disrupted other critical infrastructures, including the

rail network, four hospitals, two water treatment works, two airports, an oil refinery and chemicals manufacturing plant. 60 trains got affected directly, resulting in 23 train evacuations, 371 train cancellations, 220 part cancellations, and 873 train delays Bialek (2020). This shows the high dependence of critical infrastructures on electricity. During October and November 2019, the Pacific Gas and Electric (PG&E) public safety power shutoff (PSPS) events in California caused power shutoffs initially to 2.5 million people and later expanded to over 3 million people. The shutoffs were an attempt to prevent wildfires from being started by the electrical equipment. The mobile communication networks, as well as the landline communication, went down due to these shutoffs, thus amplifying the impacts. Firefighters could not use their electricity-based devices to communicate and draw water from electric pumps Wong et al. (2020). Historically the largest blackout occurred in India in 2012, affecting at least 600 million people across the nation Matthewman and Byrd (2014).

As our daily lives rely heavily on electricity, power failure events immediately disrupt our social behaviors and cause psychological distress. Businesses suffer, commercial activities halt, industrial manufacturing gets hampered, inventory spoils, and operations of other important critical infrastructures like hospitals, water treatment plants, transportation networks, and law enforcement agencies get affected Shuai et al. (2018); Wong et al. (2020); E-ISAC (2016). Moreover, loss of power affects individuals differently depending on their situation and circumstances. The existing literature suggests that power outages have important health consequences ranging from carbon monoxide poisoning, temperature-related illness, gastrointestinal illness, and mortality to all-cause, cardiovascular, respiratory, and renal disease hospitalizations, especially for individuals relying on electricity-dependent medical equipment Casey et al. (2020). Hence, even a very short power outage can be deadly for a person on life support. All of these factors and the "social impact" costs makes estimation of economic losses of power outages difficult. Various studies have been conducted to capture these different dimensions of power outage costs. According to Dale et al. (2018), the damage caused to the transmission and distribution system of selected California areas over 2000-2016 exceeds $700 million. Another detailed study conducted by the Lawrence Berkeley National

Laboratory (LBNL) estimated the total cost of sustained power interruptions in the US to be $44 billion per year (2015-$) LaCommare et al. (2018). Accounting for the uncertainties in estimation, this figure varies between $35-$50 billion with 90% confidence. The weighted average cost per customer for a sustained electricity interruption is estimated to be $8.2, $2009, and $16,259 per interruption for residential, commercial, and industrial customers, respectively.

There are different causes behind outages in power systems, including adverse weather, equipment failures, and cybersecurity attacks. However, among all these reasons, the weather is the dominant factor Preston et al. (2016); DOE-EPSA (2015). Between 2000 and 2021, almost 83% of reported major power outages in the US were caused directly or indirectly by weather Climate Central (2022). The ten most significant power outages in 2018 were weather caused EATON (2018). These weather events include windstorms on East Coast, Hurricane Michael in Florida, Hurricane Florence in Carolinas, a summer storm in Lower Michigan, ice and freezing rain in Ohio, a spring storm in Connecticut, and a blizzard in Kansas City. Similarly, in 2022, mother nature was the cause behind the ten most devastating power outages. One of the catastrophic weather events in 2022 is Hurricane Ian - the deadliest hurricane to strike since 1935. It left 4 million customers without power in Florida, and an additional 1.1 million in North and South Carolina Eaton (2023). Among the total 1542 weather-related outages in 2000-2021, 58% were caused by high winds, rain, and thunderstorms, 22% were caused by winter weather (including snow, ice, and freezing rain), and 15% by tropical storms and hurricanes Climate Central (2022). An alarming situation is that the frequency of these extreme weather events has increased steadily from 1970 to 2002. Figure 1.1 shows the number of global disasters reported annually from 1970 to 2022, segregated according to the disaster type Our world in data (2023). We can see that the numbers are increasing steadily, particularly for floods and extreme weather. This directly impacts the frequency of outages in the power systems. As per Eaton's Blackout and Power Outage Tracker, the average annual number of weather-related power outages increased by almost 78% during 2011-2021, compared to 2000-2010 EATON (2018).

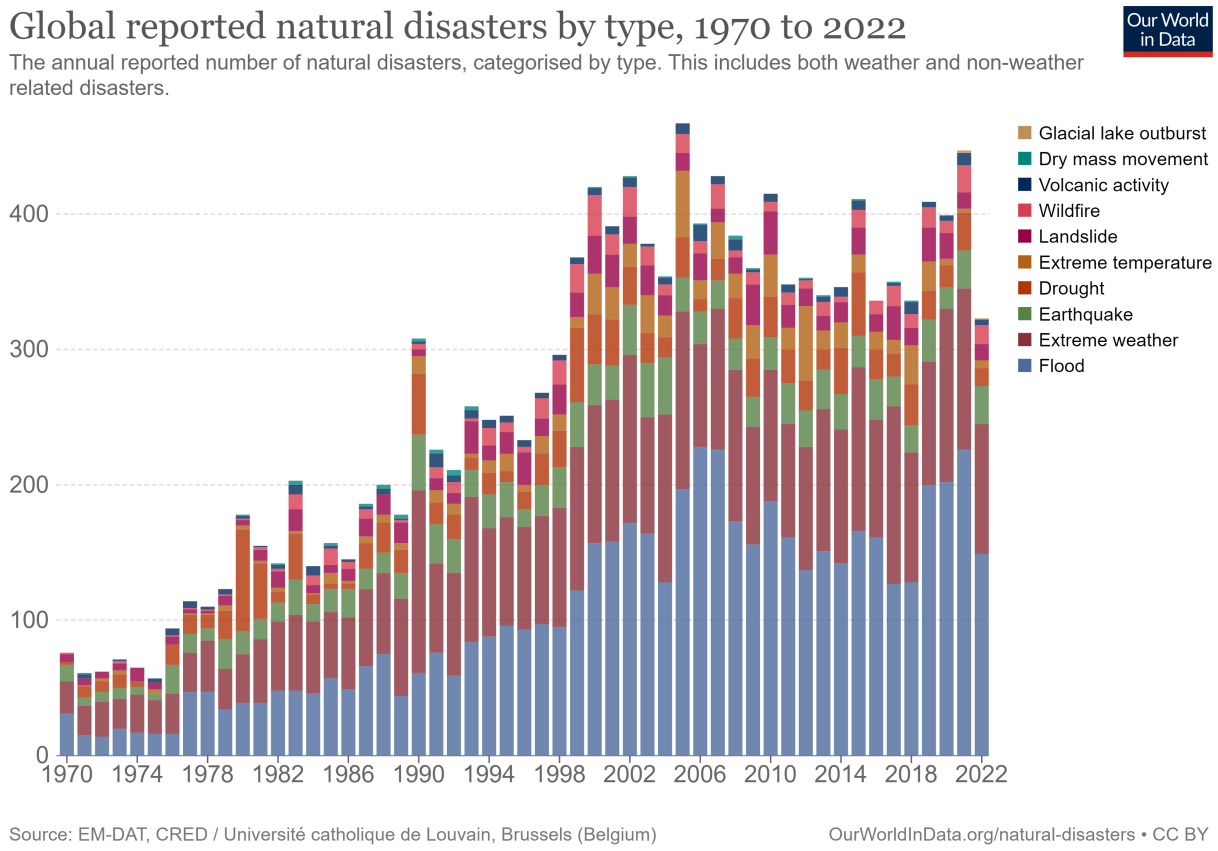

Figure 1.1 Global reported natural disasters by type, from 1970 to 2022 Our world in data (2023)

The discussion mentioned above and the associated references establish the importance of the power system's reliability and resilience. Power system reliability and resilience are two closely related yet different concepts. The incidents mentioned above, though catastrophic, only happen occasionally. Due to their low probability of occurrence, such events are categorized as the high-intensity low probability (HILP) events or high-impact low-frequency (HILF) events SIRTF (2012). While reliability analysis of electric power systems encompasses studying and improving system behavior to more frequent small-scale events, analyzing system response to HILP events falls under resilience analysis. The particular focus on HILP events primarily distinguishes resilience analysis from reliability analysis. The word "resilience" is derived from the Latin word "resilire," which

means "the ability to spring back or rebound" Suri et al. (2014). Multiple definitions of resilience exist in the literature NIAC (2009); NASEM (2017); Finster et al. (2016); PJM Interconnection (2017); Younesi et al. (2022); Gholami et al. (2018); Ouyang et al. (2012); all of these encompass this notion of recovering/rebounding from HILP events. The key elements in all these definitions are summarized by the US Federal Energy Regulatory Commission (FERC) in their definition of resilience: "The ability to withstand and reduce the magnitude and/or duration of disruptive events, which includes the capability to anticipate, absorb, adapt to, and/or rapidly recover from such an event" FERC (2018). Due to the growing importance of the resilience of power systems, it is the prime focus of this research.

The power system is divided into generation, transmission, and distribution systems. Each part has a crucial role in the resilience of the overall power system. However, the transmission and distribution are relatively more exposed and prone to natural hazards than the generation. We discuss the resilience of the distribution system in this thesis because of its importance in resilience and greater impact on consumers. From a consumer perspective, they should be able to get a continuous, uninterrupted electricity supply to their premises even if some transmission lines or distribution components are at fault. As long as they are getting an uninterrupted power supply, they are satisfied. Multiple redundancies are built into the transmission systems to increase their reliability. These are designed in such a way as to have alternate paths of power flow available in case of any failure. So even if there is an outage in the transmission system, it does not always interrupt the supply to the end consumers. But the distribution systems are different. A fault in the distribution system almost always directly impacts the consumer(s). Moreover, distribution system is far more extensive in miles of line than transmission system. Due to this, it has a higher probability of faults than the transmission system. Almost 90% of the customer outages in the US are related to power distribution systems Gholami et al. (2016). Another important aspect is the difference in the restoration times. Faults that are caused due to electrical failures are restored fairly quickly, within hours or days, compared to faults caused by natural hazards with more extensive damage. Since the distribution systems are more vulnerable to natural hazards,

outages in distribution systems take more time to restore. This is evident from the 2004 Florida hurricanes in which repairing the damaged distribution system took weeks to complete DOE (2005). Therefore the performance of the distribution systems is more critical to the restoration process, especially when subjected to wind hazards. Due to all these reasons, distribution system resilience is critically important from a consumer viewpoint.

The resilience of distribution systems is improved by investing in different hardening efforts. For example, distribution poles susceptible to wind speed damages can be replaced by poles with increased resistance to heavy winds. Similarly, tree trimming can be performed to remove trees and their branches from the clearance zones of the distribution lines. Each of the different hardening techniques requires a different amount of investment and impacts the outages in the distribution system in different ways. Therefore, a systematic approach is required to assess the impact of such investments on the distribution system's resilience metrics, so that informed planning decisions can be made. One such approach is introduced in this thesis which quantifies the impact of investments on the distribution system's wind resilience. To study the impact of an hardening investment on the distribution system's resilience, its current resilience state needs to be determined first. Area outage rate curves and resilience metrics are used for this purpose. The area outage rate curves correlate observed wind speeds at weather stations in the distribution area with the outage rate of the area of the distribution system nearest the weather station. The outage data of a USA distribution utility is used to demonstrate the approach, area outage rate curves are developed for two different areas of its power distribution system, and the corresponding resilience metrics are calculated as well. These area outage rate curves show how resilient the distribution system's specific area to a particular hazard is. Since the distribution systems overhead infrastructure is extensive, they are most vulnerable to heavy wind speeds (Gholami et al. (2016)). That is why this paper selects wind speed as the hazard for formulating area outage rate curves. However, the same methodology can also be used to develop the area outage rate curves for other hazards. Distribution system hardening investments shift the area outage rate curves. This shift changes the

resilience metrics, which are then compared with the original (before hardening) resilience metrics to quantify the improvements.

1.2 Research Objective

Use distribution system outage data to quantify the impacts of resilience improvement decisions on the resilience metrics of the power distribution system. This will aid utilities in making informed decisions to manage resilience.

1.3 Organization of Thesis

This thesis has been organized into 6 chapters and two appendices, the contents of which are summarized as:

Chapter 1 introduces the subject, discusses the need for resilience analysis of power distribution systems subjected to natural hazards with references from the literature, and highlights the motivations for the research.

Chapter 2 presents a review of existing literature on approaches and methods used to analyze the resilience of power systems in general and distribution systems in particular. A discussion of model-based analysis techniques and data-based analysis techniques is also given.

Chapter 3 contains the information about the datasets used in this research. Statistical analysis of the outage data and weather data is also given to highlight the various features of the data. Moreover, this chapter also presents the outage cause codes analysis to identify the patterns of outages of different types.

Chapter 4 explains one of the main methods developed in this research, the area outage rate curves. The mathematical treatment of these curves, along with the implementation details, is given in this chapter. Finally, the given methodology is applied to the available datasets, and the resultant area outage rate curves are presented along with their fitted models. The area outage rate curves of one of the areas of the distribution system under study are presented in this chapter, and the area outage rate curves of the second area are given in Appendix A.

Chapter 5 explains the second major contribution of this research, the use of area outage rate curves for resilience investment evaluation. Different resilience metrics and the relevant techniques to compute these from the data are introduced. Then the relationship between area outage rate curves and resilience investments is explained, followed by a detailed explanation and mathematical treatment of the outage sampling. The mechanics of the changes in the resilience metrics due to outage sampling are presented, and finally, the results are given at the end of the chapter.

Chapter 6 summarizes the important findings of this research work, states the conclusions, and suggests future avenues of research in data-driven power distribution system resilience.

Appendix A contains the results of "Area-2" of the distribution system under study. These results are cross-referenced in the body of the thessis but are included separately in this index for better organization of the thesis and to avoid cluttering the main body with figures.

Appendix B contains the distributions of the resilience metrics used in this research and presented in chapter 5. It also contains the plots of the variation in each resilience metric according to the outage event size.

CHAPTER 2. LITERATURE REVIEW

2.1 Background

According to the reports published by NOAA-National Oceanic and Atmospheric Administration (NOAA et al. (2023)), WMO-World Meteorological Organization (Association et al. (2022)), and IPCC-Intergovernmental Panel on Climate Change (Portner et al. (2019)), the intensity and frequency of HILP events are increasing at an alarming pace. Power systems need to be made resilient to such extreme weather events. That is why the resilience assessment of power systems is one of the most important research areas. However, most of the work is model based, where power system simulation models are developed and then analyzed. One particular reason is the lack of data on the power systems that is available to researchers. Nonetheless, the literature review shows that several authors have performed data-driven resilience analyses. Those works are summarized in this section, along with their distinctive approaches and the type of data used.

The literature review on the resilience analysis of power systems can be broadly categorized into the following five categories:

1. Evaluation of system's resilience performance: Research works in this category focus on evaluating the resilience of power systems based on either its current model or based on data. Recommendations and suggestions are then provided on how to improve the performance of the system. It also involves post-event analyses in evaluating the impacts of extreme events on the system and how the damages in those events can be prevented in the future.
2. Real-time resilience assessment: PMUs and micro-PMUs are being installed in the power system at an increased rate. These devices increase observability and provide real-time information about various system parameters. Nonetheless, SCADA and AMI are prevalent in the industry right now. Research in this category develops techniques to utilize this information to assess the current resilience state of the system in real-time.

3. System resilience planning: Most works in this category are model-based. Different modeling approaches are used to plan resilient power systems, and recommendations are provided on expanding or establishing resilient power systems in the future.
4. Resilience Analysis after modification: This category involves research works studying the integration of distributed generation, energy storage, smart switches, microgrids, and other such technologies in the power system and their impacts on resilience. Studies like the optimal placement of distributed generation resources to maintain and improve the system's resilience fall in this category.
5. Resilience Metrics: The topic of power system resilience is not new, but unlike reliability indices, there are still no "standard" metrics to measure the resilience of a power system. Much work has been done on suggesting different types of resilience metrics, and still, more work is being done to devise even more comprehensive resilience metrics and frameworks.

The research works falling in the first category are presented in this literature review because the work done in this thesis also mainly falls under this category.

Fragility curves are typically used to model failure probabilities of individual components subjected to varying hazard intensities. But Dunn et al. (2015) suggested using fragility curves for lifeline infrastructure systems instead of individual components. They developed empirical curves for overhead line components in the distribution system to study their performance under wind storm hazards. They used observed wind gust data supplemented with atmospheric model data. They used faults/outage data from the UK's National Fault and Reporting Scheme (NaFIRS) database. The curves plot the mean number of faults (or consumers involved) against wind speed. However, the outage database used does not have the exact location information of the fault; rather, it has information about the concerned DNO (60000 km^2), Region (15000 km^2), and Area (2000 km^2). So the authors developed the curves separately for the three spatial resolutions. They showed that using wind speed data of a lower spatial resolution (higher distances between the exact fault locations and point of observations of the wind speed) results in a significant underestimate of the failure wind speeds of the components (i.e., components fail at a higher wind speed), as compared

to using a higher resolution wind speed data. It shows the importance of using high-resolution wind speed data for hazard analysis of outages in the power distribution system.

Dunn et al. use the same wind speed data and faults data again in Dunn et al. (2018) to suggest using the catastrophe (CAT) risk modeling approach to assess the resilience of the distribution system. The CAT risk models are used to assess risk by combining an exposure database with hazard, vulnerability, and economic loss models. The insurance industry typically uses the CAT models to perform risk assessments of geographically dispersed independent assets. However, the authors proposed an extension to this approach for risk assessment of infrastructure systems and applied it to electrical distribution networks as an example. They modeled empirical fragility curves for an average number of faults/1000km in a distribution system during wind storm events (defined by wind speeds $\geq$ 17 m/s). They concluded that the average number of faults, normalized by the line length, increases with an increase in wind speed, exhibiting a power-law relation. Moreover, the spatial distribution of faults (urban vs. rural) does not significantly affect the curves, and most faults occur because of tree cover.

Ouyang et al. (2012) developed a three-stage resilience analysis framework to analyze the resilience of infrastructure systems. The three stages are the "disaster prevention stage", "damage propagation process", and the "recovery process". Resilience-based improvement strategies are discussed for each stage. The framework is tested on the model of the power transmission grid in Harris County, Texas. The authors conduct a comparative analysis of the original model and several hypothetical resilience-improved models. The analysis showed a 0.034% improvement in the time-dependent expected annual resilience (AR) metric. One of the important findings in this paper is that random hazards (a collective name for different hazard types, such as human errors, tree-induced failures, and animal-related events) impact resilience more significantly than hurricane hazards due to their higher frequency and randomness.

Researchers in the Multidisciplinary Center for Earthquake Engineering Research (MCEER) introduced a general framework to assess and quantify the seismic resilience of infrastructure systems Bruneau et al. (2003). Authors in Ouyang and Dueñas-Osorio (2014) applied this framework to

measure the power system's resilience in hurricane-prone areas. Based on the MCEER framework, they proposed a four-dimensional resilience model with four interrelated sub-models: a hazard scenario generation model, a component fragility model, a power system performance model, and a system restoration model. The four dimensions of the resilience model correspond to the technical, organizational, social, and economic dimensions of resilience. The authors used the definition of resilience as the ability of a system to resist, absorb, and recover from hazards. They have addressed each of these three aspects in their work. The authors concluded that "among the technical, organizational and social dimensions, the organizational resilience is the highest with a value of 3.445 in the logarithmic scale while the social resilience is the lowest with a value of 2.620 in the same scale. To improve system resilience and reduce economic loss, strictly obeying the restoration sequences, increasing the restoration resources and their mobilization rate, and enhancing the rate of undergrounding critical distribution lines are all effective strategies in practice."

Bjarnadottir et al. (2013) proposed a probabilistic framework to evaluate the vulnerability of power distribution poles to hurricane risk under the potential impact of a changing climate. They developed fragility curves for wind hazards (3-sec wind gust speed) with and without considering the degradation effects. The authors found that the expected annual probability of failure could increase by up to 50% under an assumed increase in wind speed of 10% over 100 years. Correlation analysis of transmission system faults and weather data by Murray and Bell (2014) suggests that wind gusts are a more suitable measure to explain fault frequency than wind speed. The study shows a logarithmic fit between a 10-meter wind gust and the probability of wind-related faults' occurrence.

The performance of a utility's distribution system in Seattle, Washington, was analyzed during four different wind storms by Reed (2008). The wind storms occurred on 01/1993, 11/1995, 11/1196, and 12/1996. The authors also compared the results of their analysis with those of Davidson et al. (2003) and the French storms of 1999. They defined fragility as the probability of damage conditional on the squared gust wind speed and equated it to the ratio of "damaged feeder length" and "total length of feeders". They also obtained "outage duration curves" to estimate the restora-

tion of the power delivery system. A two-parameter Gamma distribution is used to model the duration of outages. The authors concluded that the gamma distribution with shape parameter correlated with the five-second wind gust speed squared provides a good description of the outage durations during storms. Moreover, the outage durations during wind storms are remarkably consistent, even if the underlying wind phenomena differ. The authors also concluded that the gust speed squared best predicts outage durations and fragility associated with electric utility lifelines. Another conclusion of their analysis is that the fragility is best formulated using the affected length of the feeders rather than the percentage of feeders because the feeder lines have unequal lengths.

In another work, Reed (2009) defined fragility as "the percentage of outages relative to the total number of customers residing per the spatial distribution of a county". They studied the fragility of power systems against wind and rainfall hazards in extreme events (hurricanes). A log-normal model was fitted in the fragility analysis, and they found that fragility curves for wind speed and rainfall hazard exhibit similar trends and, thus, are not independent.

The power transmission, as well as distribution network, is an interconnected "Networked Infrastructure". Reed et al. (2009) outlined a method for evaluating the resilience of such complex network infrastructure for extreme natural hazard events. To illustrate its use, the method is then applied to study the resilience to wind-induced damage to the electric power delivery system in the Gulf region following Hurricane Katrina. The three resilience metrics used by Reed et al. (2009) are fragility, quality, and input-output inoperability.

Davidson et al. (2003) studied the performance of the distribution systems of two of the major electric power companies in the Carolinas (Carolina Power & Light and Duke Power) during five hurricanes: Opal (1995), Fran (1996), Bonnie (1998), Dennis (1999), and Floyd (1999). They used data from five sources: Outage data from CP&L and Duke Power, Hurricane/Weather data from a hurricane wind speed database, a rainfall database (NCDC, 2001), and a land cover database (MRLC, 2001). They concluded from their analysis that most of the damages are caused by trees, and the highest density of outages occurred in areas of high-intensity residential cover. They found the maximum gust wind speed a necessary predictor of disruption in the system. Another finding

was that the shape of the outage restoration curve is weakly related to the overall intensity of the wind and rainfall in a hurricane.

Almost all of the works mentioned above are based on the resilience assessment of a particular resilience event or multiple such events. They used information on the occurrence of extreme events, such as hurricanes, earthquakes, wind storms, etc. But what if such information is not available? How can such events and their timelines be isolated and systematically identified in the power system data? These questions have been answered in Carrington et al. (2020, 2021) by devising methods to automatically extract such resilience events directly from the outage data. Once these events are extracted, any of the analysis techniques discussed in the literature above can be used to evaluate resilience.

The review of literature on infrastructure resilience in general and power system resilience in particular shows that this is a growing area. Even the concept of resilience has been captured and explained differently by different authors (Raoufi et al. (2020); Jasiūnas et al. (2021); Gholami et al. (2018)). Moreover, another difference is in the fragility analysis of the power system, where some works have studied resilience using the fragility of individual components in the power delivery network while others have modeled the fragility of the overall system (Aven (2011)). Yet another distinction, with pros and cons, is using models vs. using real data. Irrespective of all these different approaches, one result is conclusive in all the studies that weather hazards, particularly wind-related weather hazards such as hurricanes and winter storms are the most destructive, and are one of the most important factors in resilience assessment.

CHAPTER 3. OUTAGES AND WEATHER DATA

3.1 Overview of datasets

Two different datasets are used in this research, one with an electric power utility's distribution system outage data and the other containing the same area's weather data. The time span of both datasets is six years. Both datasets are cleaned and pre-processed before conducting the analysis. Details of information present in these datasets, along with the statistical analysis, are given in the following sections.

3.2 Outages Data

3.2.1 Dataset Description

The outages dataset contains records of 32278 individual outages that occurred over six years, with a one-minute resolution, i.e., all outages that occurred within one minute were labeled with the same timestamp. Each outage entry in the dataset corresponds to an outage of a component in the power distribution system and has the location coordinates of that component, the number of customers affected during the outage, the starting and ending time of the outages along with other information. During the dataset cleaning and pre-processing, 149 outage records are removed due to absence of location information in those records. The outage records are then sorted in increasing order of the occurrence time of each outage. Summary statistics of the cleaned data are shown in the Table 3.1

3.2.2 Outage Cause Codes

The outage dataset also has information on the cause of each outage. This information is given in the form of "outage cause codes". There are 64 unique cause codes defined in this dataset. These

Table 3.1 Summary of the Outages dataset (excluding duplicates and missing values)

Statistic	Minimum	Maximum	Mean	Standard Deviation
Outage duration (hours)	0.02	433.0	9.44	17.31
Inter-outage interval (hours)	0.0	1827.9	1.62	15.79
Outage Rate (per hour)	0.0	253.0	0.61	2.83
Customers Affected	0.0	7012.0	53.99	179.49
Customer-Hours	0.0	96294.6	454.63	2347.47

cause codes are grouped into six categories, a summary of which is shown in Figure 3.1. Outages caused by Trees have the maximum percentage, followed by equipment-related faults and weather-related outages. However, when we analyze the cause codes within the "Tree" category, we find that these outages are also primarily caused by heavy winds. A breakdown of the "Tree" and the "Weather" categories is given in Figure 3.2. 92% of the outages in the "Tree" category are caused due to the trees present within and outside the clearance zone around distribution lines. When trees are located within the clearance zone, wind causes the tree branches to touch the distribution lines causing faults in the system. Strong winds also break the tree branches and limbs, which, when falling on the distribution lines, causes faults. Therefore, the wind is also an indirect cause behind the outages within the "Tree" category. Combining these outages with the 47% outages with the "wind/rain" cause code in the "weather" category, we get a total of 44% outages caused directly/indirectly by the wind.

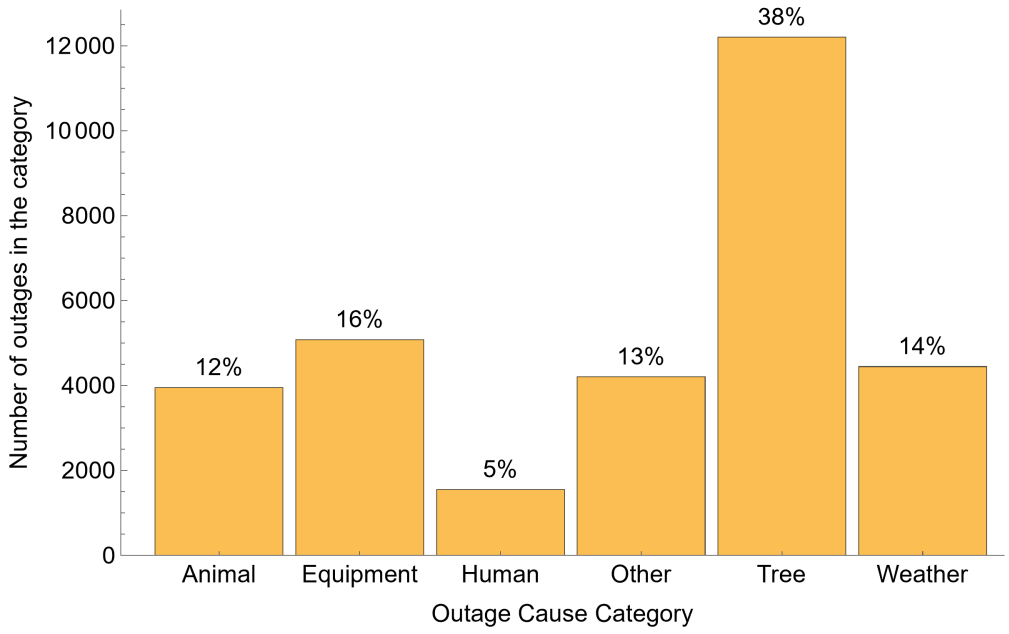

Figure 3.1 Outage Cause Code categories Bar Chart

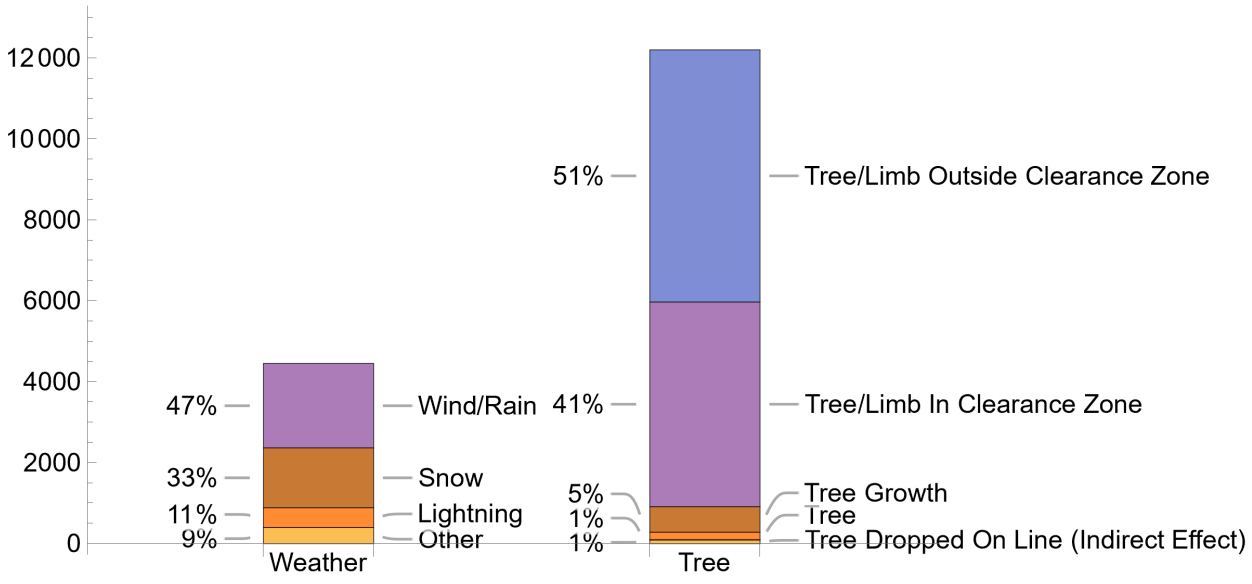

Figure 3.2 Breakdown of the "Tree" and "Weather" Cause Code Categories

3.3 Weather Data

3.3.1 Dataset Description

The weather dataset is sourced from the NCEI (National Centers for Environmental Information) database. It contains data from multiple weather stations in and around the utility area. The data range of each weather station in the dataset is the same six years period as the outage data. The dataset has different weather parameters, including wind speed, wind gust speed, wind direction, precipitation, sky conditions, relative humidity, and monthly summaries; however, only wind speed data is used here. The specifications of wind speed and wind gust observations recorded by each weather station are given in Table 3.2. The wind speed observations were taken hourly, and therefore ideally, there should be 52608 observations (corresponding to 52609 hours in this date range). However, some weather stations have more observations (data points), indicating observations taken at intervals smaller than 1 hour. In the same way, some weather stations have fewer observations, indicating missing data. More details about this are discussed in the next section.

3.3.2 Data Cleaning

As mentioned in the previous section, the weather dataset has inconsistent observation intervals. Therefore, the records with no available wind speed data were removed from the dataset. This resulted in time periods of varying lengths having no wind speed observations. These missing time intervals are summarized in Table 3.3. Similarly, the data of some weather stations has two entries for the same timestamps, corresponding to different wind speeds. The data documentation NCDC (2015) was consulted to find the reason for these duplicate entries. It was found that there are different types of observations corresponding to different data collection methods for each weather station. Most of the duplicate entries had the same wind speed. However, in the case of different wind speeds, the mean of such observations was taken and then rounded to the nearest integer. A summary of such duplicate entries and the missing hours of data is given in Table 3.3.

Some wind speed values are appended with an “s”, indicating that these are suspect values and are not 100% certain. There are 41 such values in the complete wind dataset, summarized in

Table 3.2 Details of Weather Stations

Weather Station Name	Station Elevation	Anemometer Height (above ground surface)	Wind Speed Observations' Specification	Wind Gust Observations' Specification
WS-1	413.1 ft	5 ft	60-min average	maximum 10-sec wind speed during 1 hour
WS-2	166.0 ft	33 ft	2-min average of 5-sec averages	maximum 5-sec wind speed during 1 hour
WS-3	1402.9 ft	33 ft	2-min average of 5-sec averages	maximum 5-sec wind speed during 1 hour
WS-4	279.9 ft	33 ft	2-min average of 5-sec averages	maximum 5-sec wind speed during 1 hour
WS-5	365.2 ft	33 ft	2-min average of 5-sec averages	maximum 5-sec wind speed during 1 hour
WS-6	581.0 ft	33 ft	2-min average of 5-sec averages	maximum 5-sec wind speed during 1 hour

Table 3.3. Almost half (46%) of such suspect observations were recorded when wind gusts were blowing. Therefore most probably higher than the actual wind speed values were recorded in such observations. These values are replaced with a three-hour mean wind speed around the timestamp of the suspect observation.

3.3.3 Basic Data Statistics

The dataset is analyzed after cleaning, and some basic summary statistics are shown in Table 3.4. The weather station WS-1 has no available wind gust speed data; therefore, N/A is shown in the maximum wind gust speed column in Table 3.4. The dataset has weather data from six weather stations; however, only two weather stations, WS-1 and WS-2, are selected to demonstrate the methods developed in this work.

Table 3.3 Summary of Missing, Duplicate and Suspect values in the Weather dataset

Weather Station Name	Total Missing Hours	Maximum Consecutive Missing Hours	Mean Consecutive Missing Hours	Total Duplicate Entries	Total Suspect Values
WS-1	4	1	1	0	0
WS-2	218.8	78	2.3	22	2
WS-3	1471	158	10.8	0	4
WS-4	0	0	0	38	0
WS-5	805.4	240	2.9	5	7
WS-6	5598.7	10	1.2	0	28

Table 3.4 Summary of the Weather dataset (excluding duplicates and missing values)

Weather Station Name	Total Data Points	Minimum time difference between consecutive data points	Maximum Wind Speed	Total Wind Gust Speed Data Points	Maximum Wind Gust Speed
WS-1	52628	60 min	17 mph	0	N/A
WS-2	67725	1 min	36 mph	6174	54 mph
WS-3	152465	15 min	36 mph	30360	53 mph
WS-4	81262	1 min	43 mph	11497	56 mph
WS-5	70809	1 min	71 mph	7056	111 mph
WS-6	49407	1 min	144 mph	7404	78 mph

3.3.4 Wind Speed Distribution

Wind speed is a random quantity, and it can be modeled using the Weibull Distribution or Rayleigh Distribution, a special case of the Weibull distribution. Although other distributions, like Lognormal distribution, Generalized Gamma Distribution, Pearson Type-III Distribution, can also be used to model the distribution of wind speeds, the two-parameter Weibull distribution gives the best fit in most of the cases and is, therefore, the recommended distribution Shi et al. (2021); Bidaoui et al. (2019); Ouarda et al. (2015); Morgan et al. (2011); Zhou et al. (2010); Carta et al. (2009); Jowder (2006); Garcia et al. (1998); Tuller and Brett (1985). Nonetheless, there is no universally

best distribution to model wind speed and each wind speed data needs to be tested with different available options to select the best fit. Since, the two-parameter Weibull distribution is simple and has well-defined forms of mean, variance and moments, it is used here. The two-parameter Weibull distribution is a continuous distribution with a scale parameter $\lambda \in (0, +\infty)$ and a shape parameter $k \in (0, +\infty)$. It is a generalized gamma distribution with both shape parameters equal to k. Its PDF and CDF are given in (3.1) and (3.2), respectively.

$$f(x; \lambda, k) = \begin{cases} \frac{k}{\lambda} \left(\frac{x}{\lambda}\right)^{k-1} e^{(-x/\lambda)^k}, & x \geq 0, \\ 0 & x < 0 \end{cases} \tag{3.1}$$

$$F(x; \lambda, k) = 1 - e^{(-x/\lambda)^k} \tag{3.2}$$

The density function (3.1) has an infinite slope at $x = 0$ as k varies between 0 and 2. Therefore, zero values in the data require some special arrangements. Otherwise, the two-parameter Weibull distribution will not give accurate results Wais (2017); Rosen et al. (1999). If there are not many zero values in the data, then the density can be estimated using only the non-zero values. But if the percentage of zeros is significant, this approach yields misleading results. If there is a known error in the sensing/recording instrument such that it records zero values for very small values, then censoring techniques can be used to handle the zero values in the data. Another approach is to use the three-parameter Weibull distribution. The third parameter is the position parameter which shifts the Weibull distribution on the horizontal axis. It fits well in the low wind speed region but does not accurately model higher wind speed values. Moreover, it also results in negative values, which need to be taken care of separately. A yet another approach Takle and Brown (1978) is to calculate the probability of observing zero wind speed and combine it with the two-parameter Weibull density. Since Weibull is a continuous distribution, the probability of x equaling any constant value is zero. Therefore, the probability of wind speed $V = 0$ is calculated separately and then combined with the density to get the Hybrid Weibull density.

The wind speed data used in this research is obtained from different weather stations and has significant zero wind speed values, indicating the presence of "Calms", as shown in the histograms

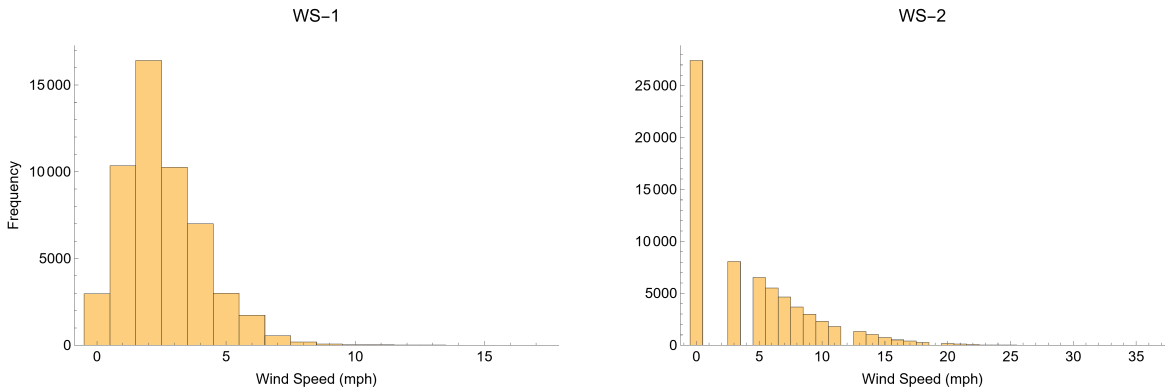

Figure 3.3 Histograms of WS-1 and WS-2 weather stations' wind speed data

in Figure 3.3. However, we can see from the PDF (3.1) of the Weibull distribution that it will have an infinite slope as k varies between 0 and 2. Since it is established that the zeros in the data are not insignificant, we calculated the probability of zeros separately and combined it with the PDF of the two-parameter Weibull distribution to get the following Hybrid Weibull density function:

$$f_H(x;\lambda,k,s) = \begin{cases} s\delta(x) + (1-s)\frac{k}{\lambda}\left(\frac{x}{\lambda}\right)^{k-1} e^{(-x/\lambda)^k}, & x \geq 0, \\ 0 & x < 0 \end{cases} \tag{3.3}$$

The third parameter, s in (3.3), denotes the frequency percentage of "calms" in the data. The $\delta(x)$ is the Dirac Delta function or the Impulse function. The second term of the sum in (3.3) is the density function of the two-parameter Weibull distribution scaled by a factor of $(1-s)$. The CDF of the hybrid Weibull distribution is given in (3.4), and its mean and variance are given in (3.5) and (3.6) respectively.

$$F_H(x;\lambda,k,s) = s + (1-s)(1-e^{(-x/\lambda)^k}), \qquad x \geq 0 \tag{3.4}$$

$$E[X] = \bar{x}_H = \lambda(1-s)\Gamma(1+\frac{1}{k}) \tag{3.5}$$

$$Var[X] = \overline{(x-\bar{x})^2_H} = \lambda^2(1-s)\Gamma(1+\frac{2}{k}) - \bar{x}^2_H \tag{3.6}$$

There are different methods to estimate the scale and shape parameters of the Weibull distribution. For example, the least squares method, power density method, method of moments, and

maximum likelihood estimator Shi et al. (2021); Dorvlo (2002); Conradsen et al. (1984); Stevens and Smulders (1979); Justus et al. (1978). However, the maximum likelihood estimators of both parameters give the best goodness-of-fit results Bidaoui et al. (2019); Jowder (2006); Seguro and Lambert (2000); Conradsen et al. (1984). Therefore, the maximum likelihood estimation method is used to estimate the parameters. The parameter estimates, along with the 95% confidence intervals, of the wind data of WS-1 are given in Table 3.5. The resultant CDF is shown in Figure 3.4.

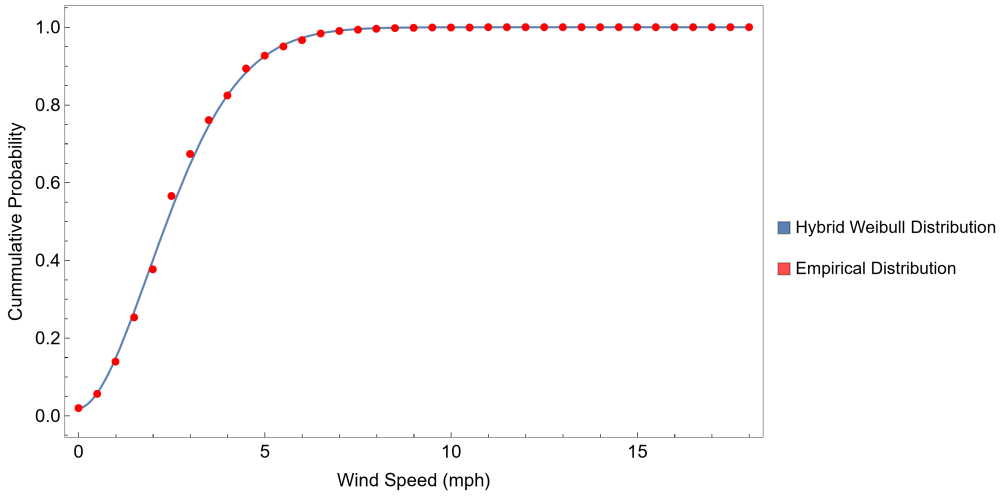

Figure 3.4 CDF of the Hybrid Weibull Distribution of WS-1 weather station data

Table 3.5 Parameter estimates and confidence intervals of the Hybrid Weibull distribution of WS-1

Parameter	Estimated Value	95% Confidence Interval
λ	2.954	(2.939, 2.969)
k	1.803	(1.791, 1.814)
s	0.196	N/A

The empirical CDF of the wind speed data of the WS-1 weather station is also overlaid in Figure 3.4 to show the comparison between the empirical values and the parameterized model. Figure 3.5 shows the comparison between the CDF of the hybrid distribution and the traditional

two-parameter Weibull distribution. The overall shape is almost the same, but the difference lies in the CDF value at zero due to the inclusion of the probability of calms. This addition also changed the mean value of the distribution along with its other statistics. Important statistics of the fitted Weibull model are given in Table 3.6. The same process can be adapted to create models of the wind speed data of other weather stations in the dataset.

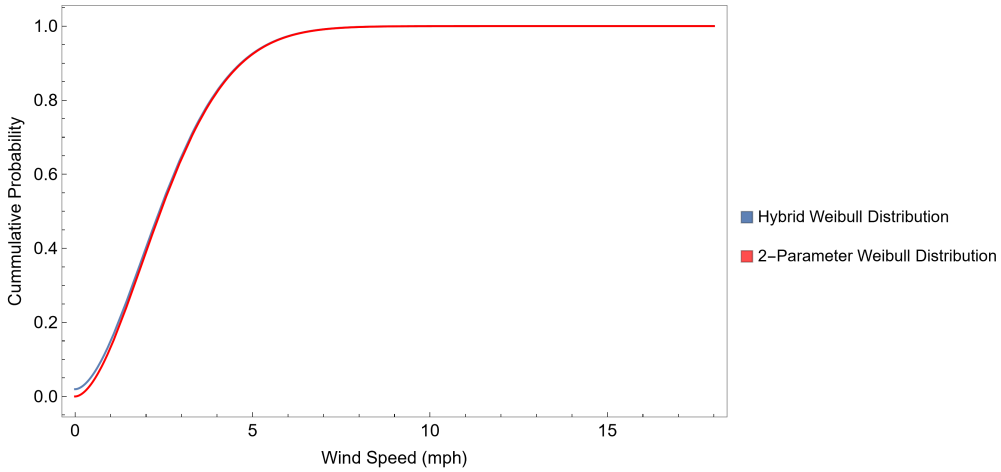

Figure 3.5 Comparison of the CDFs of the Hybrid Weibull Distribution and the Two-parameter Weibull Distribution of WS-1 weather station data

Table 3.6 Statistics of the Hybrid Weibull distribution of WS-1

Statistic	Value
Mean	2.575
Variance	2.363
Standard Deviation	1.537
Interquartile Range	2.090
Skewness*	0.722

* Skewness=$\tilde{\mu}_3 = \frac{\mu_3}{\sigma^3} = \frac{\mu_3}{(\mu_2)^{3/2}}$

After obtaining the appropriate distribution fit for the wind speed data, different useful information can be extracted. For example, the Survival Function or the Complementary Cumulative Distributive Function (CCDF) can be generated to obtain the cumulative probabilities of observing wind speed above a particular threshold. Let V be the random variable representing the wind speed,

then the survival function $S(v)$ is defined by (3.7) and plotted in Figure 3.6. This shows that the probability of observing a wind speed value of more than 7 mph is 0.9%. Since we are conducting an analysis of resilience events, which have a low probability of occurrence, this information about the probability of occurrence of high wind speeds is particularly useful to estimate the chances of observing resilience events around a weather station.

$$S(v) = P[V > v] = 1 - P[V \leq v] = 1 - F_H(v) \tag{3.7}$$

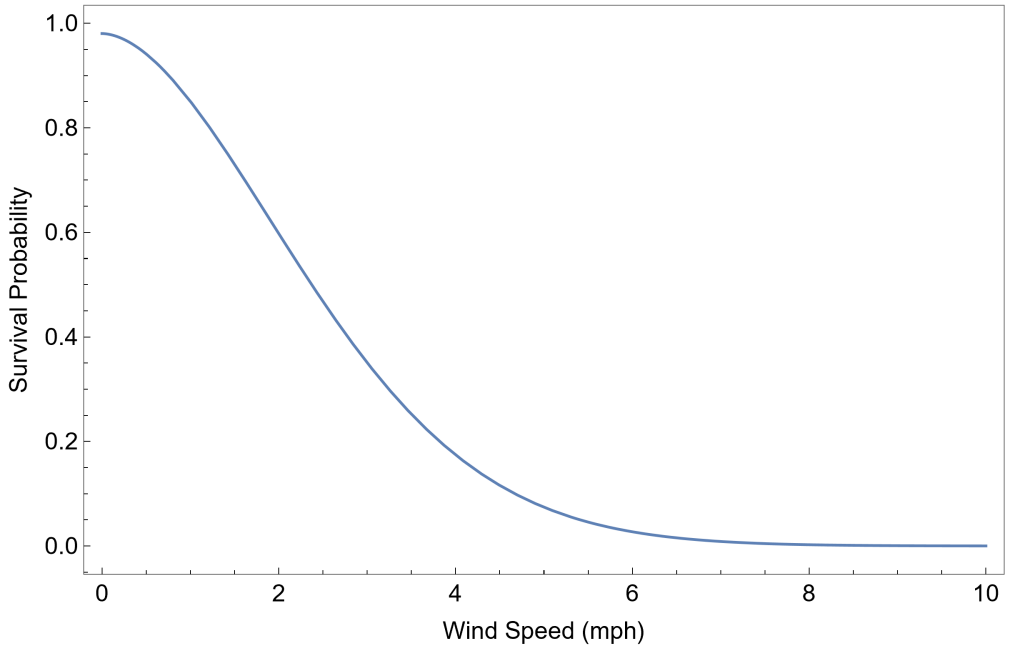

Figure 3.6 Survival Function (CCDF) of wind speed data of the WS-1 weather station

These wind speed modeling results checks that the wind speed measurements in the weather dataset follows a typical distribution known in the literature.

CHAPTER 4. AREA OUTAGE RATE CURVES

4.1 Overview

As established in the previous chapter, the occurrence of outages in the power distribution system is very much influenced by heavy winds. The rate of occurrence of outages in a given time interval, i.e., outage rate, therefore, also varies with the wind speed. The effect of wind intensity on the outage rates in the power distribution system is derived from data and modeled using Area Outage Rate Curves, as explained in the following sections.

4.2 Area Outage Rate Curve

The term "Area" in Area Outage Rate Curve refers to the spatial areas of the distribution system associated with each weather station. The weather data is collected from weather stations, and there can be more than one weather station inside the geographical boundaries of a power distribution utility. Therefore, the total area of the utility is divided into small areas such that each "area" has an associated nearest weather station. The power outages in the distribution system are then segregated based on their area, that is according to which weather station they are closet to. This results in multiple areas within the distribution system, each with a weather station and a group of associated outages.

Now we can formally define the area outage rate curve for the wind hazard. "An Area Outage Rate Curve is the functional relationship between the wind speed measured at the area weather station and the mean outage rate of components falling within the distribution system area". That is, the area outage rate curve shows the mean area outage rate as a function of the measured wind speed. Other hazard intensities like precipitation and wind gust speed could also be used instead. The outage rate curves are developed to study the resilience of the power distribution system. A more resilient system would have low outage rates at high hazard intensity levels compared to a

less resilient system which has higher outage rates at high hazard intensity levels. Therefore, these outage rate curves are direct indicators of the resilience state of the distribution system. These can be used to study a particular area's resilience and compare the resilience of multiple inter-system and intra-system areas. The mathematical derivation, implementation, and results of the outage rate curves are discussed in the following sections of this chapter.

4.3 Area Outage Rate Curves and Fragility Curves

The area outage rate curves closely resemble conventional Fragility Curves, yet there are some notable differences between them. Fragility curves are defined as the probability of a component reaching or exceeding a specific damage state under a given stress (hazard intensity) level Mohamed Nazri (2018); Feng et al. (2000); Lallemant et al. (2015). These can also be seen as conditional probability, as reflected in the general equation defining the fragility curve shown in (4.1).

$$Fragility = P[DS|IM = y] \tag{4.1}$$

where,

DS is the damage state

SM is the stress (hazard intensity) measure

y is the specific value of the intensity measure (IM)

The main difference is that conventional fragility curves describe the fragility of a specific component to stress on that component, whereas the area outage rate curves describe the fragility of an entire area of the system with respect to a particular measurement that correlates with stress on the entire area. The area contains many components that differ in type, age, and condition and are subject to different stress values. The historical data encompasses all these effects, so that the area outage rate curve can describe the aggregated fragility of the area for a specific measurement related to the overall stress on the area.

Fragility functions are typically used to model conditional failure probabilities of individual components, like bridges, buildings, and electrical poles Biglari and Formisano (2020); Dunn et al.

(2018). However, while studying the resilience of the power distribution system, we are more interested in the fragility of the overall system instead of its components. Therefore, unlike the traditional component-based fragility analysis approach, we devised an approach to conducting an "area-based fragility analysis" of complete areas in the distribution system. We calculate outage rates of all the components in a given "area" of the distribution system as a function of wind speed observed in a weather station located within that area. The outage rate can be translated to the mean probability of outage of an average component by dividing by the number of components in the area, but this is not necessary for our calculations.

The area outage rate curves are overall similar to the fragility curves introduced by Dunn et al. (Dunn et al. (2018, 2015)), but there are many detailed differences in the processing and implementation. Particularly, the purpose of developing and using these curves differs in both cases. Dunn et al. developed fragility curves to use as damage models in the damage assessment of lifeline infrastructures. In comparison, we developed area outage rate curves to model the historical wind resilience of an area of the distribution system. If we compare the technical details of fragility curves developed by Dunn et al. and the area outage rate curves developed by us, we see that the fragility curves show a relationship between the average number of faults/1000km in a distribution system area subjected to maximum wind gust speed observed in that area at the time of fault occurrence. Since the data used by Dunn et al. does not have the exact fault locations available, so they used the maximum wind gust speed observed in the area where the fault occurred. The area where the fault occurred has three different spatial resolutions; DNO (60000 km^2), Region (15000 km^2), and Area (2000 km^2); Dunn et al. developed separate curves for the three spatial resolutions and then compared the accuracy of results. Moreover, they specified a lower wind gust speed cutoff; wind speeds $\geq$ 17/ms, for identifying wind storm events and then modeled the faults that occurred during those events.

In contrast to the fragility curves developed by Dunn et al., the area outage rate curves developed in this thesis provide a relationship between the mean outage rates (outages/hour) in a distribution system area corresponding to the observed wind speed (not wind gusts!) in that area. We do not

specify any cutoff for the wind speed, and thus we do not exclude any outages. The motivation behind this is to model the overall historical behavior of a particular distribution system area to the average wind speeds observed in that area. The data used in this work have exact locations of outages, so we selected the closest weather stations based on those locations. Therefore we can also estimate the wind stress observed by the faulted component as a function of its distance from the weather station. For example, the distance between the outages in Area-1 and the corresponding weather station is shown in Figure 4.4. We can see from this plot that the maximum distance between an outage and the weather station in Area-1 is almost 70 miles ($\approx$ 113 km). Moreover, we did not use the maximum wind speed observed in the area at the time of fault occurrence; rather, we used the average hourly wind speed observed at a weather station closest to the fault location at the time of fault occurrence. Considering these major technical and implementation differences, we can see that, although conceptually similar, the area outage rate curves are developed differently with a different purpose than Dunn's fragility curves.

4.4 Derivation of Area Outage Rate Curves

The derivation of the outage rate function from data is explained in this section. As mentioned earlier in Section 3.2, the outage data is sampled every minute, giving the total number of outages that occurred during that one-minute time. An outage rate function $R(t)$ is defined to give the outage rate per minute at time t. The function $R(t)$, when drawn against time t, represents a varying amplitude square wave with discrete-valued horizontal and vertical axes, as shown in Figure 4.1. $R(t)$ preserves averages over intervals. Since $R(t)$ is a discrete function, integrals over $R(t)$ correspond to summations. As a result, (4.2) gives the mean outage rate per minute $\overline{R}$ for a total of n minutes.

$$\overline{R} = \frac{1}{n}\int_{1/2}^{n+1/2} R(t)dt = \frac{1}{n}\sum_{i=1}^{n} R(t_i) \tag{4.2}$$

Unlike the outage data, the weather data was originally sampled non-uniformly with one-hour, 15-minutes, and even more than one-hour intervals. Therefore linear interpolation is used to sample the wind speed data at the desired uniform rate of one minute. The linear interpolation function

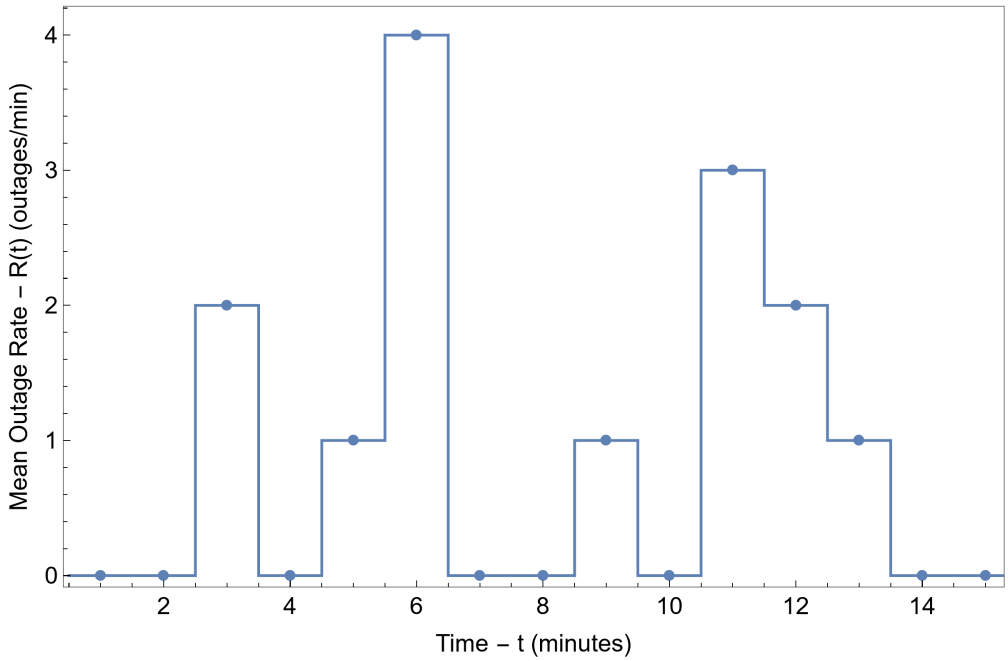

Figure 4.1 Outage Rate function

$V(t)$ is defined in such a way that if the wind speed data values are available at time samples s_1, s_2, s_3, ... , s_k then $V(t)$ gives the wind speed value at any time instant between two consecutive time samples as expressed in (4.3).

$$V(t) = \frac{V(s_k)(s_{k+1} - t) - V(s_{k+1})(s_k - t)}{s_{k+1} - s_k}, \; s_k \leq t \leq s_{k+1} \tag{4.3}$$

Suppose we have the values of $V(t)$ at s_1, s_2, s_3,... as shown in Figure 4.2, then the function $V(t)$ is invertible on a line segment with $s_k \leq t \leq s_{k+1}$ as:

$$t = \frac{[V(s_{k+1}) - v]s_k + [v - V(s_k)]s_{k+1}}{V(s_{k+1}) - V(s_k)} \tag{4.4}$$

$$V^{-1}(v) = \{t \mid V(t) = v\} \tag{4.5}$$

$V^{-1}(v)$ defines a set of time instant values with wind speed value v. Let $W(t) = V(t) - v$ and if $W(s_{k+1})W(s_k) < 0$ then (4.4) can be simplified as follows:

$$t = \frac{W(s_{k+1})s_k - W(s_k)s_{k+1}}{W(s_{k+1}) - W(s_k)} \tag{4.6}$$

Equation (4.6) gives the next time instant t in the set $V^{-1}(v)$ if $W(s_{k+1})W(s_k) < 0$. A union of all such t values gives the complete set of inverse values $V^{-1}(v)$ as:

$$V^{-1}(v) = \bigcup_{k:W(s_{k+1})W(s_k)<0} \left\{ \frac{W(s_{k+1})s_k - W(s_k)s_{k+1}}{W(s_{k+1}) - W(s_k)} \right\} \tag{4.7}$$

While (4.3) gives the interpolated value of wind speed at any time instant t between any two consecutive known data points, (4.7) gives the set of time values corresponding to a particular wind speed v. Figure 4.2 shows an example of $V^{-1}(v) = \{t_1, t_2\}$ when the wind speed value is v at time instants t_1 and t_2. Now that we have defined the mean outage rate per minute and the corresponding wind speed values, we can find the outage rate function. The outage rate function maps the wind speed v to the mean outage rate $\overline{F}(v)$:

$$\overline{F}(v) = \frac{1}{|V^{-1}(v)|} \sum_{t \in V^{-1}(v)} R(t) \tag{4.8}$$

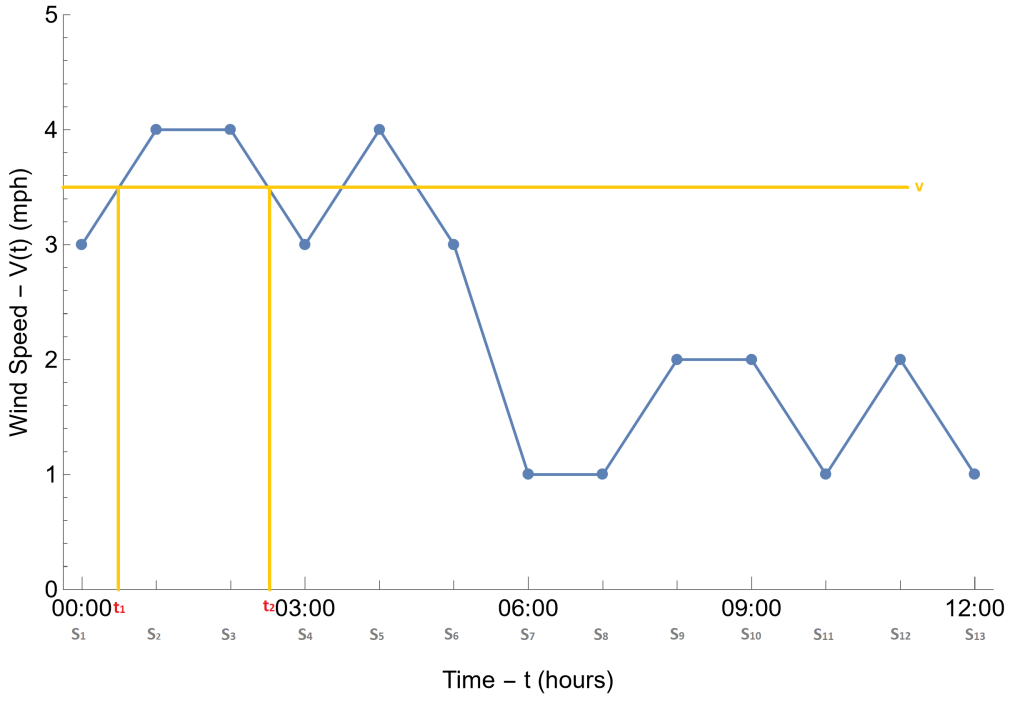

Figure 4.2 Interpolation of wind speed data

4.5 Software Implementation

This section explains the software implementation to calculate the area outage rate curves. In summary, the outages are assigned weather stations using spatial information, thus creating areas. Then the interpolation of wind data is performed, and finally, the area outage rate curves are generated after merging the outage and wind datasets based on temporal data. The fine details of each step are explained in the following sub sections.

4.5.1 Spatial Association

The first task in the implementation is to select the closest available weather station to each outage to have a corresponding wind speed value that is taken closest to the place of outage occurrence. The outage dataset has latitude and longitude values indicating where that outage has occurred. Likewise, the location coordinates of each weather station are available in the weather dataset. The great-circle distance between each outage and each weather station is calculated using the haversine formula (4.9):

$$d = 2r \cdot \sin^{-1}\left(\sqrt{\sin^2\left(\frac{\varphi_2 - \varphi_1}{2}\right) + \cos\varphi_1 \cdot \cos\varphi_2 \cdot \sin^2\left(\frac{\lambda_2 - \lambda_1}{2}\right)}\right) \tag{4.9}$$

where,

r is the radius of earth $\approx$ 3961 miles

φ_1, φ_2 are the latitude coordinates of point 1 and 2 respectively

λ_1, λ_2 are the longitude coordinates of point 1 and 2 respectively

Then the weather station with the minimum distance is assigned to each outage. Details of the number of outages assigned to each weather station and their distance from the weather station are given in Table 4.1. This step divides the total distribution system's area into smaller areas, each with one weather station and its associated outages. Since there are six weather stations, we get six areas within the utility's distribution system. The location of each outage after this spatial association task is shown in Figure 4.3 for two of the areas, along with the locations of the corresponding weather stations (shown with black-filled circles). Only these two areas, Area-1 and

Area-2, are selected as data representatives for further analysis. **Relevant plots and results for Area-1 are included in the main body of the thesis, whereas those for Area-2 are given in Appendix-A for better organization of the contents.**

Table 4.1 Summary of the weather stations' assignment to outages

Weather Station	Assigned outages	Distance to closest outage	Distance to farthest outage	Mean distance between weather station and outages
WS-1	12715	0.1 miles	70.6 miles	28.8 miles
WS-2	7876	0.4 miles	43.4 miles	12.5 miles
WS-3	987	20.5 miles	71.2 miles	32.7 miles
WS-4	3302	27.5 miles	71.3 miles	45.2 miles
WS-5	2706	0.7 miles	44.2 miles	23.6 miles
WS-6	4543	0.1 miles	25.2 miles	10.6 miles

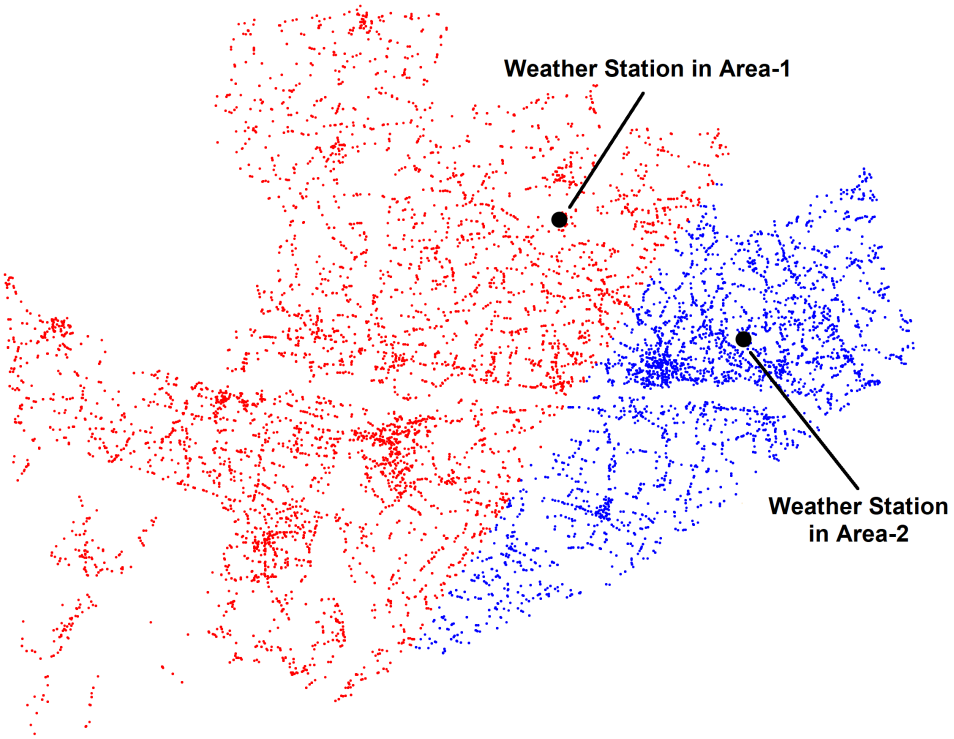

Figure 4.3 Physical location of power outages and the associated weather stations in two of the distribution system areas (Area-1 in red and Area-2 in blue)

The histograms of the distance between each outage and the relevant weather station of Area-1 and Area-2 are shown in Figure 4.4 and Figure A.1, respectively. The bin size in these histograms is 2 miles. We can see from these plots that 50% of outages lie within 16 miles and 41 miles of the weather station in Area-1 and between 7 miles and 15 miles of the weather station in Area-2. Another important information we get from the histogram plot in Figure 4.4 is the three distinct peaks located at 10, 26, and 42 miles. These show that more outages occurred at these distances as compared to the rest of the area around the weather station.

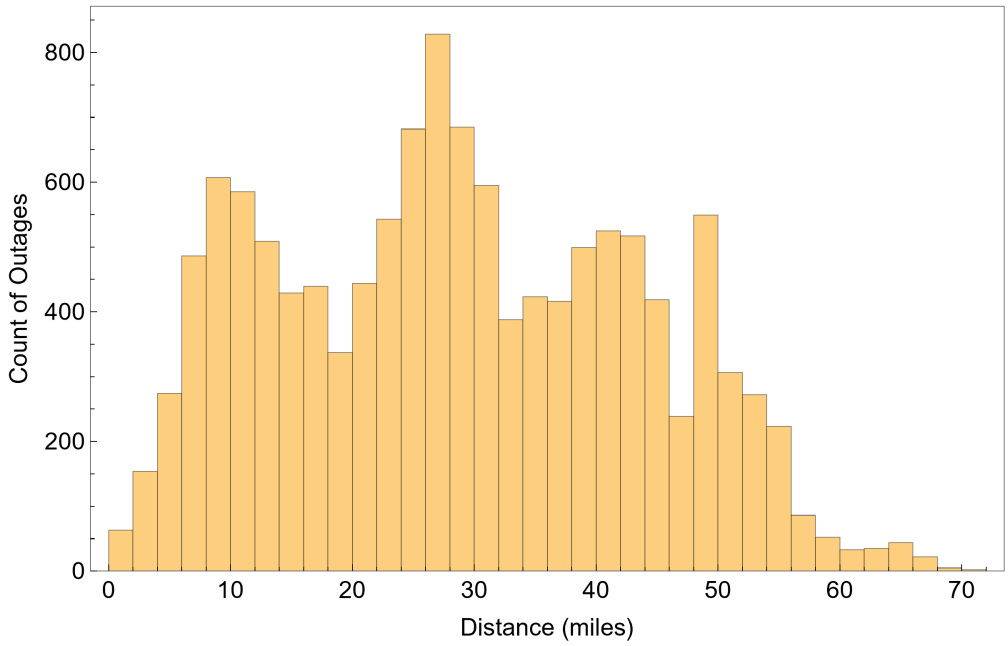

Figure 4.4 Distribution of distance between outages and weather station in Area-1

4.5.2 Temporal Association

The next step is to assign the closest wind speed to each outage according to the time difference between the wind speed observation and the outage occurrence. Each outage has its time of occurrence, and each wind speed has a time of observation, so the time difference between the timestamps of each outage and wind speed value is taken, and the wind speed with the smallest time difference is assigned to each outage. Since the data resolution of the outages dataset is 1 minute, and the data resolution of the wind speed dataset is 60 minutes, there is a time difference in each outage and wind speed observation time. The distribution of these time differences in Area-1 and Area-2 is shown in Figure 4.5 and Figure A.2, respectively.

According to Figure 4.5, the time difference between all outages and the wind speed observations is a maximum of 0.55 hour. However, in Area-2 (Figure A.2), there are 46 outages which have a difference of more than 1 hour in their initiation time and the closest wind speed observation time. Furthermore, there are 18 outages with a time difference of more than three hours. The resampling strategy based on linear interpolation is adopted to resolve this difference, as explained

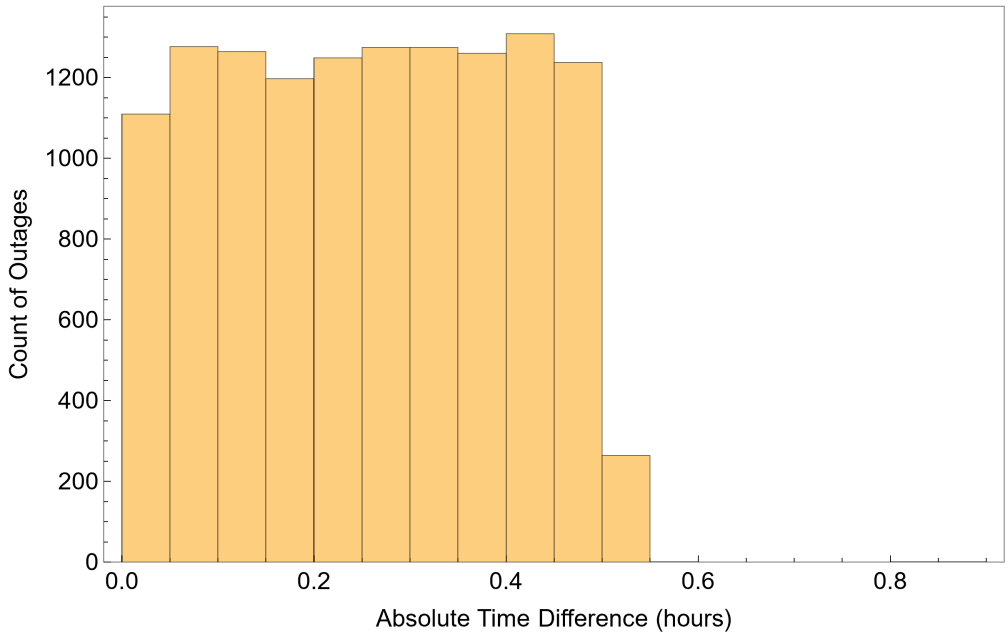

Figure 4.5 Frequency of time differences between outages and wind speed observations of Area-1

in Section 4.4. The resampled data has a resolution of one minute, which directly aligns with the one-minute resolution of the outage data. To ensure accurate results, 18 outages with a time difference of more than 201 minutes are omitted from further analysis. The wind speed values corresponding to these omitted observations are all below 8 mph; therefore, this omission only affects the results a little.

4.5.3 Joining the datasets

The spatial association step divided the system into areas, and the interpolation resulted in the same time resolution of both datasets. The datasets are then joined using the timestamps as the common key between both datasets. A three-column dataset is obtained after the join operation, with each column representing the time (with $\Delta t = 1$ minute), wind speed (at that minute), and outage rate (outages/minute), respectively. This time series wind speed data of Area-1 and Area-2 with overlaid outage rates are shown in Figure 4.6 and Figure A.3, respectively. A 7-day moving average of the wind speed is also plotted on the graphs to visualize the overall wind speed trend

over six years. Although all the calculations are performed with the one-minute resolution, the data in these plots is presented with a one-hour time resolution for ease of understanding.

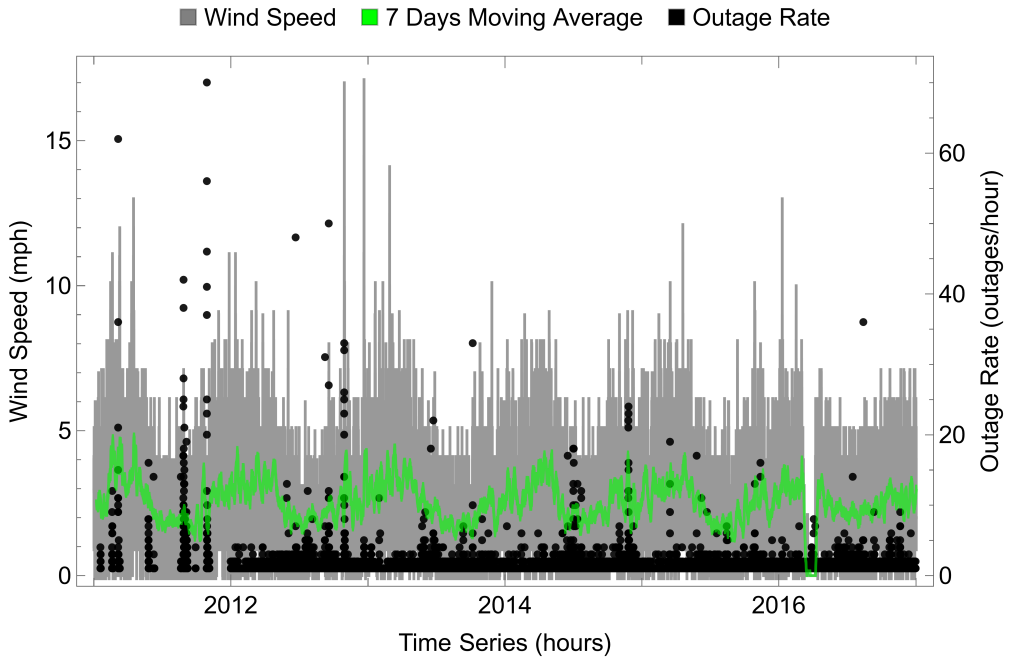

Figure 4.6 Time series plot of wind speed and outages data of Area-1

Figure 4.6 plot shows the wind profile over the years in this particular distribution system area and the corresponding hourly outage rates. Each data point (black colored points) represents the outage rate within an hour. We see vertically stacked outage rates at certain positions in this plot. These outage rates are not stacked on top of each other; rather, these are very close in time, but they appear stacked due to the large time scale. These "clusters" of outage rates are actually resilience events, as discussed more in Section 5.2. When a large number of outages occur in successive hours, they give rise to these resilience events. We can crudely identify such events by specifying an outage rate threshold and then checking the outages with outage rates higher than that threshold. However, these plots don't give us information about how the outage rate varies on average with the wind speed. The area outage rate curves are developed for this purpose.

4.5.4 Outage Rate Curves

The interpolated wind speed values in the joined dataset are rounded to the nearest integer. The data is then grouped based on the integer wind speed values to get individual outage rates at each wind speed. The individual outage rates for Area-1 are shown in Figure 4.7 whereas the individual outage rates for Area-2 are shown in Figure A.4 in appendix-A. The bubble size in Figures 4.7 and A.4 is proportional to the occurrence of that specific outage rate value. It is shown to visualize the overall distribution of outage rates for each wind speed. We can see the trends in the individual outage rates as the wind speed changes. There are many small outage rate values in the lower wind speed levels, which result in a smaller mean outage rate for lower wind speeds. Another information conveyed in Figures 4.7 and A.4 is that the data gets sparser as the wind speed increases. There are two factors behind this phenomenon, the low probability of occurrence of higher wind speeds and the low probability of higher outage rates.

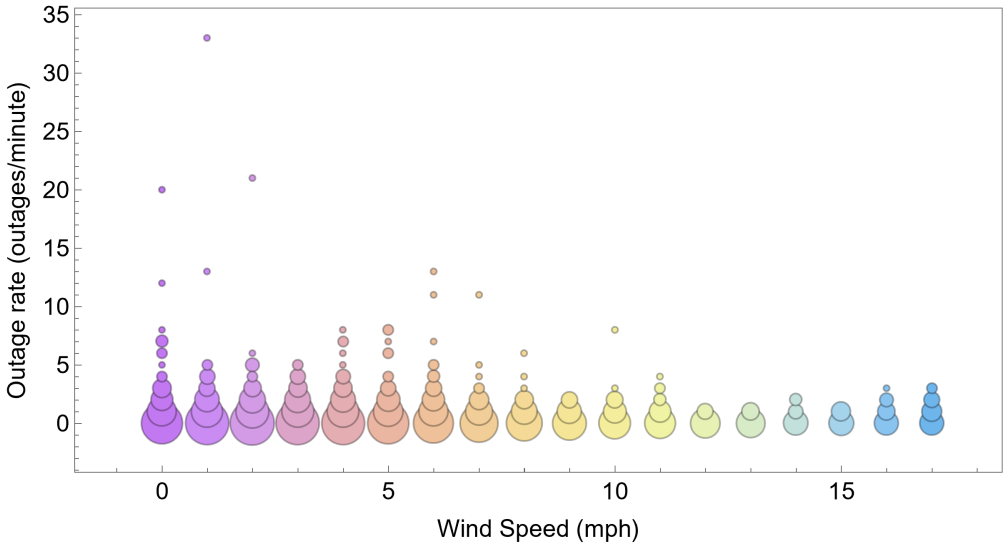

Figure 4.7 Individual Outage Rates at each wind speed level in Area-1

The outage rate curve for Area-1, i.e., the area outage rate curve, is calculated by taking the mean of all individual outage rate values at each wind speed level. These curves are shown in Figure 4.8 for Area-1 and in Figure A.5 for Area-2.

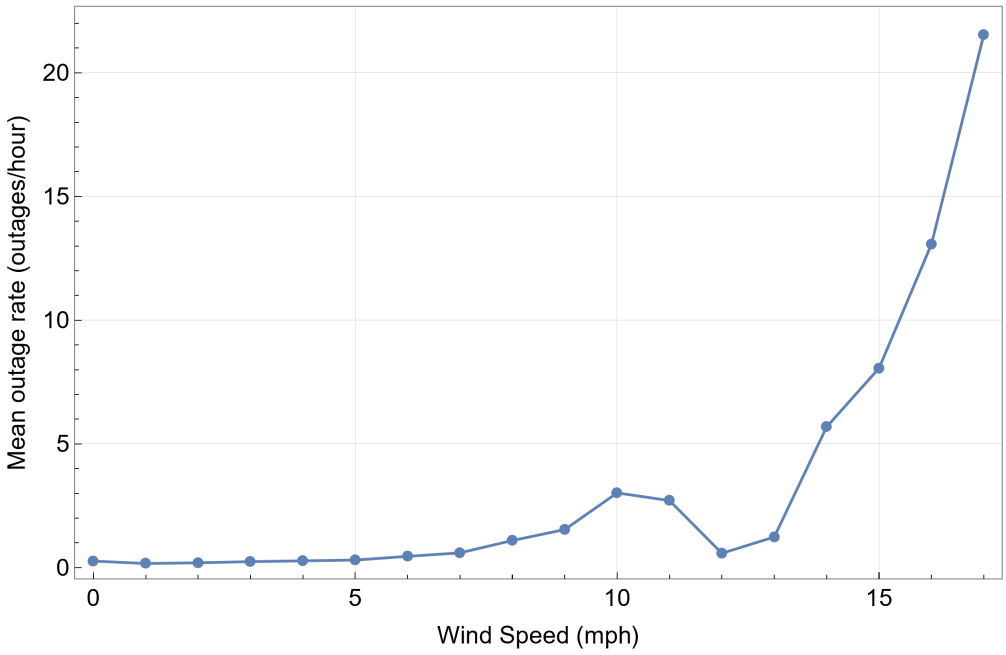

Figure 4.8 Area Outage Rate Curve of Area-1

4.6 Curve Fitting of the Area Outage Rate Curves

The area outage rate curves obtained in the previous section are specific for each area of a certain distribution system. But, comparing both the presented curves (Figures 4.8 and A.5) shows certain similarities in the overall form (In fact, this has been tested with all six areas of the distribution system under study, and the overall shape of the curves is similar to the ones shown in Figures 4.8 and A.5). The data points of these area outage rate curves are plotted on a log-linear scale to find the most suitable model for these curves. The result is shown in Figure 4.9 for Area-1 and Figure A.6 for Area-2.

We can see straight lines forming roughly in both of these plots that suggest an exponential model is the most suitable fit for the area outage rate curve specifying the mean outage rate $\overline{F}(v)$ as a function of the wind speed v:

$$\overline{F}(v) = b_0 e^{q_1 v} \tag{4.10}$$

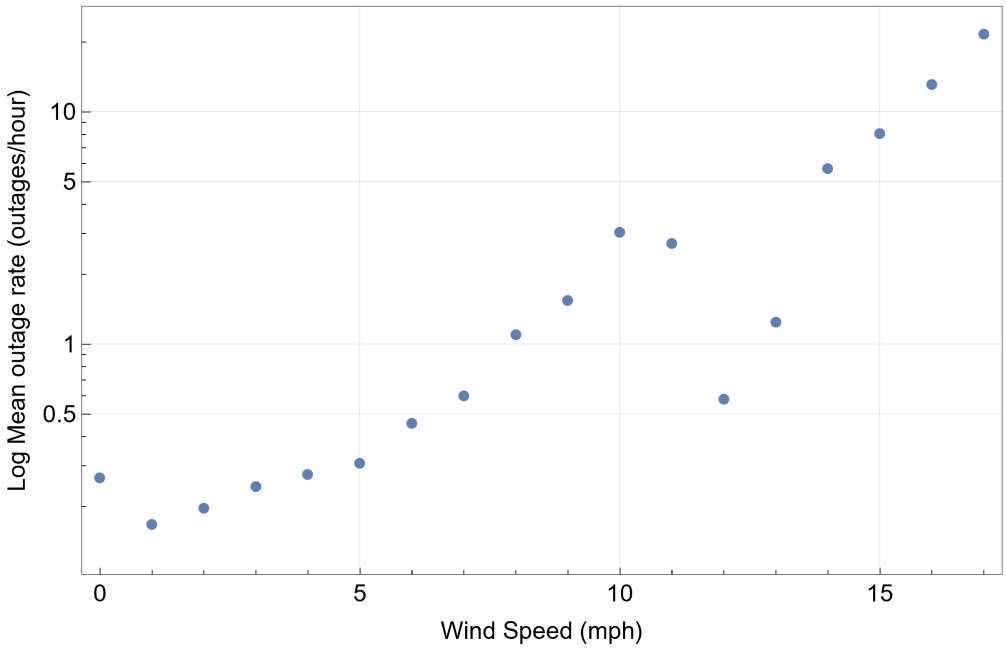

Figure 4.9 Log-Linear plot of the Area Outage Rate Curve of Area-1

The exponential model (4.10) is fitted using the Levenberg-Marquardt method (LM algorithm, also known as the damped least-squares method) with a 95% confidence level for parameters and predictions. After fitting the exponential model, the parameters obtained are $b_0 = 0.006$, and $q_1 = 0.481$. The fitted model is given in (4.11) and plotted in Figure 4.10.

$$\overline{F}(v) = 0.006e^{0.481v} \tag{4.11}$$

Similarly, the fitted exponential model for Area-2 is given in (4.12) and plotted in Figure A.7.

$$\overline{F}(v) = 1.44 \times 10^{-7}e^{0.6v} \tag{4.12}$$

The goodness of fit parameters used to select the best-fit model are given in Table 4.2 for both Area-1, and Area-2 fitted models.

4.7 Outage Cause Codes Analysis with Wind Data

As mentioned in Section 3.2.2, the outage data also has outage cause code information. After the spatiotemporal association of the weather data and the outage data, we can study the outages'

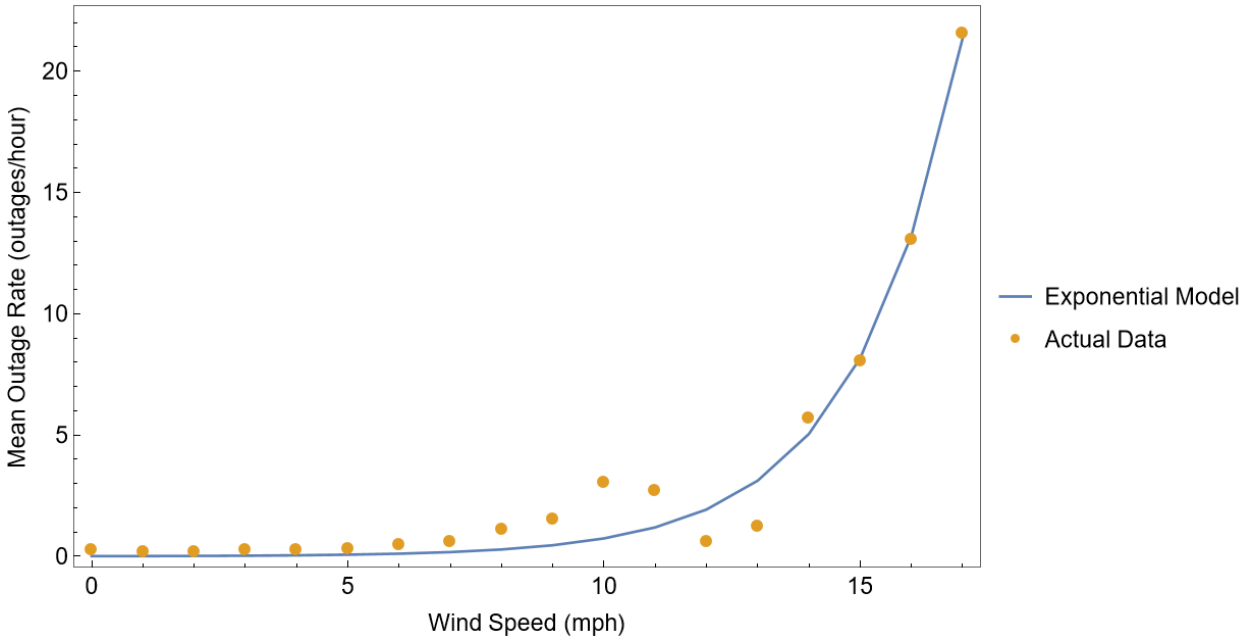

Figure 4.10 Estimated Area Outage Rate Curve of Area-1

behavior based on weather variations. The number and percentage of outages in each cause category of Area-1 and Area-2 are shown in Figure 4.11 and A.8 respectively. These bar charts resemble the bar chart shown in Figure 3.1, which means the proportion of outages falling in each category is almost the same as the overall outages in all areas. Figure 4.12 and A.9 shows how the percentage of outages in each cause category changes as the wind speed changes in Area-1 and Area-2. It is evident from these plots that the maximum number of outages at each wind speed level fall in the "weather" and "tree" categories, and as the wind speed increases, the outages of other categories

Table 4.2 Goodness of fit properties of fitted Area Outage Rate models

Property	Area-1	Area-2
Adjusted R^2	0.98	0.86
R^2	0.98	0.87
AIC	54.82	151.00
BIC	57.49	155.49

decrease, and the outages with weather and tree-related cause codes dominate. This shows that weather is the main cause of outages.

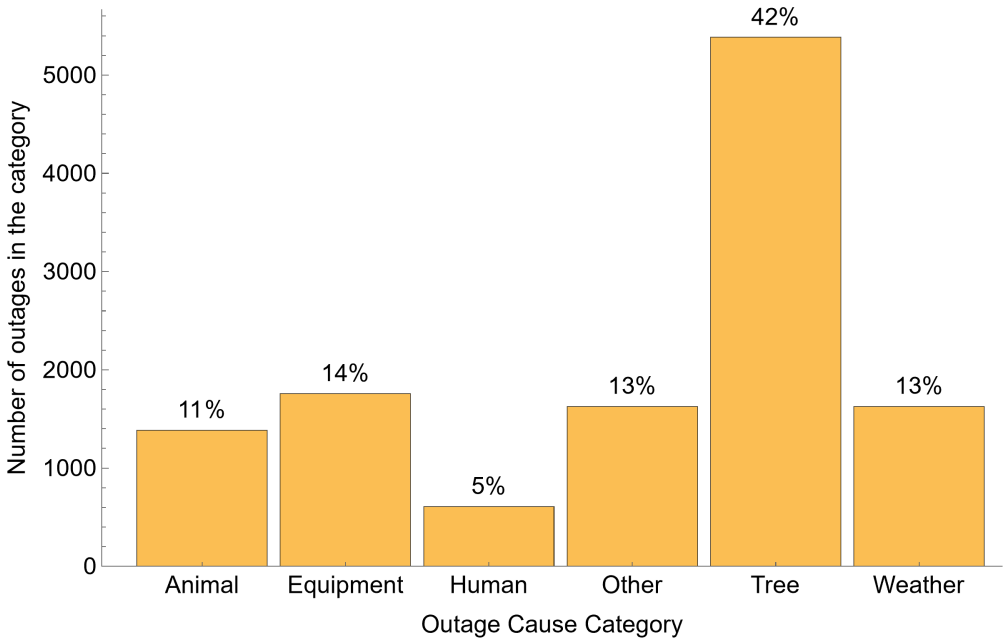

Figure 4.11 Percentage of outages in each cause code category of Area-1

The area outage rate curves in the previous section are formulated using all outages in each area without differentiating the outage cause categories. We can also develop these curves for outages of a particular cause category. These selective outage curves give us three pieces of information. First, these show us how accurate are the assigned outage cause codes. Secondly, these show the variation of outage rates in a particular cause category. Thirdly, we can check from these curves which cause categories are influenced by heavy winds and which are not. Figures 4.13 and 4.14 show these outage rate curves with the selected outage cause categories in Area-1. Figure 4.13 shows that the outage rate increases the most with the increase in the wind speed for the "Weather" cause code category, followed by the "Tree" and "Equipment" cause code categories. This shows that wind is indeed the dominant cause of outages in these categories, and the assigned cause codes are correct. Therefore, even if the cause code information is unavailable in the data, it can be safely assumed that as the outage rate increases exponentially, the weather, particularly

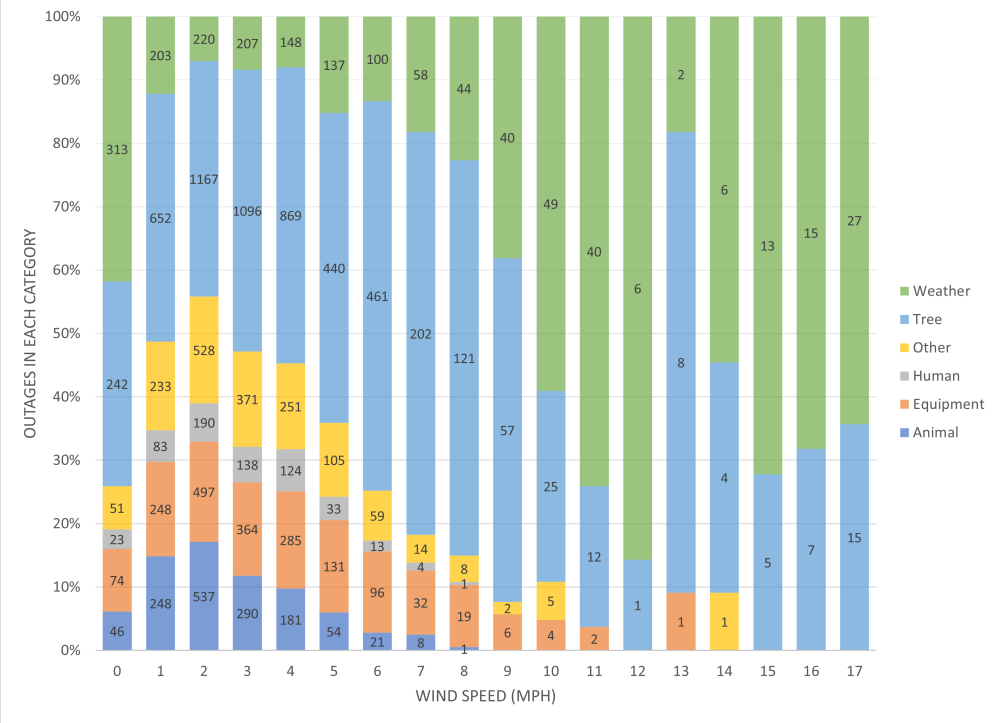

Figure 4.12 Variation in the proportion of outage categories with wind speed in Area-1

wind, becomes the dominant outage cause code. These plots (Figures 4.13 and 4.14) also show that the tree-related and equipment-related faults in the distribution system increase when wind speed increases. So if a system needs to be made resilient to heavy winds, then necessary measures must be taken to minimize fault in these two categories. On the other hand, the response of the "Human", "Animal", and "Other" cause categories shown in Figure 4.14 reflects that these three categories are least affected by heavy winds. Therefore, less priority can be given to addressing the causes behind outages in these categories when the goal is the improve system's resilience to heavy wind speeds, and more priority should be given to the cause categories shown in Figure 4.13. The plots with selective outage cause categories of Area-2, shown in Figure A.10 and A.11, show similar trends and thus verify these arguments.

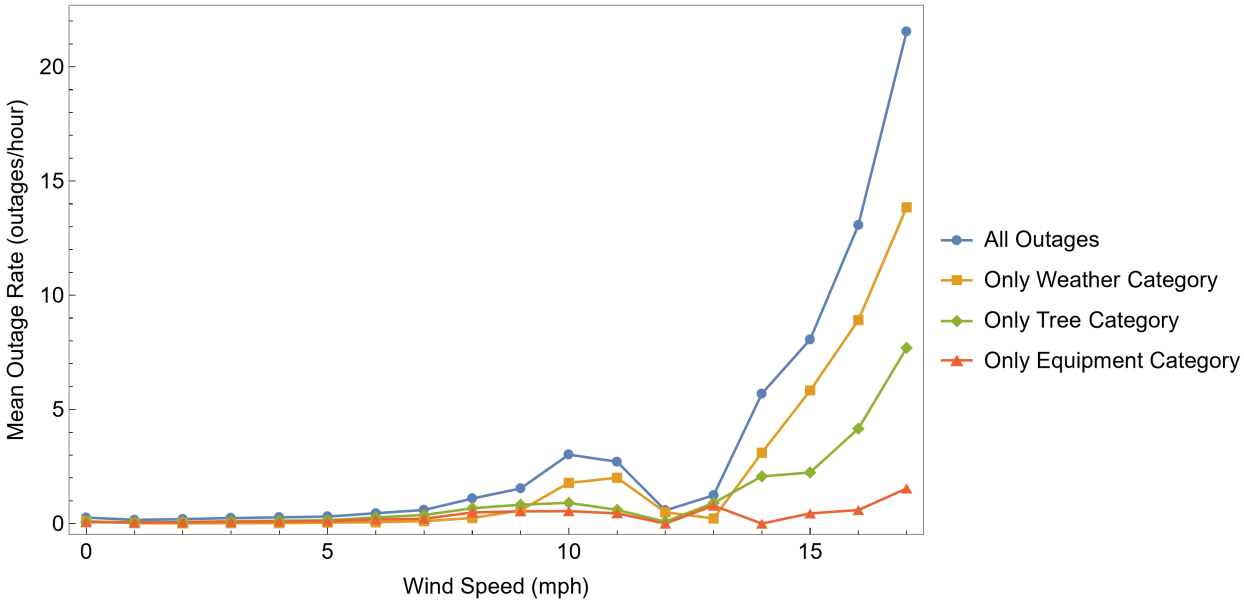

Figure 4.13 Area Outage Rate Curves of Area-1 for Selected Outage Cause Categories

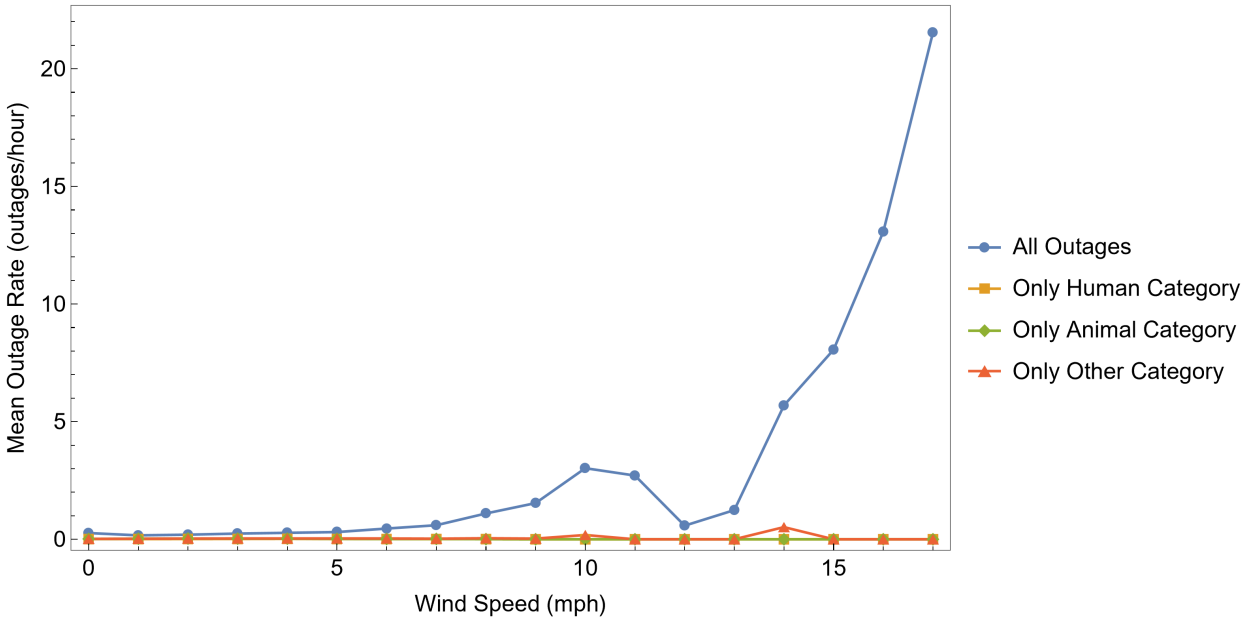

Figure 4.14 Area Outage Rate Curves of Area-1 for Selected Outage Cause Categories

CHAPTER 5. RESILIENCE INVESTMENTS EVALUATION

5.1 Overview

As discussed in the introduction, one of the goals of this work is to quantify the impact of resilience-enhancement investments made in distribution systems. This chapter elaborates on how the area outage rate curves help evaluate the benefits of such investments. When an investment is made in the distribution system to improve its resilience, it improves the resilience metrics. For example, these resilience metrics include the number of outages, restoration time, and duration of outages. More details about the resilience metrics are in Section 5.2. So, the resilience investment improves the resilience metrics, but which specific resilience metric will be improved? And by what percentage? Moreover, different kinds of investments can be made in the distribution system to improve its resilience. For example, the poles in the distribution system can be replaced with new poles with increased resistance to heavy winds, the overhead distribution lines can be buried underground, or more maintenance crews can be hired to quickly restore outages. Not all investments have an equal effect on the resilience metrics; some will target particular ones, whereas others will target all resilience metrics generally. So, how can the impact of a specific resilience-improvement investment on the resilience metrics be determined? All these questions can be answered with the help of area outage rate curves.

One important aspect to note here is that we are studying the historical outage data of the distribution system. So the analysis is retrospective, i.e., our goal is to see how much the resilience metrics would have improved if a certain investment was made in the past. The outage data used in this study is from six years past, so the area outage rate curves represent the system's resilience to wind hazards during those past six years. If a resilience-improvement investment had been made at the beginning of those past six years, then we address the question: what would have been its

impact on the resilience metrics of the system? We have used area outage rate curves and outage sampling techniques (discussed in Section 5.4) to address this question.

Figure 5.1 illustrates the use of area outage rate curves in resilience improvement. We assume that the utility wants to improve the resilience of a certain distribution system area to wind hazard by x mph. The area outage rate curves of that particular area are shifted to the right by x mph. This gives new outage rates at each wind speed level. The outage data is sampled according to these new outage rates, and then the resilience metrics are recomputed to get the percentage improvement in each resilience metric. Each stage of this process is explained in detail in the following sections.

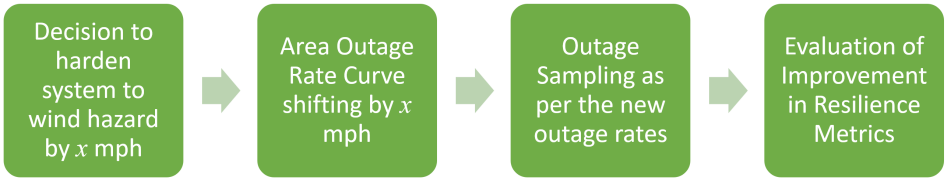

Figure 5.1 Resilience Improvement Flow Chart

5.2 Resilience Metrics

The power system resilience is its ability to maintain its performance and quickly recover from a degraded state after an extreme event. The extreme event is often called High-Intensity Low Probability (HILP) or High-Intensity Low Frequency (HILF) event Liu et al. (2022); Gholami et al. (2018). Resilience Metrics, also known as Resilience Indices or Resilience Indicators Raoufi et al. (2020), are used to measure a power system's resilience. These can also be used to compare the resilience of different power systems with each other. These resilience metrics differ from the more common Reliability Indices; SAIDI, SAIFI, CAIDI, CAIFI, and ASAI; in the sense that reliability indices measure the system's steady state performance averaged over the years. The reliability indices are dominated by the ability of the power system to withstand "regular short-term disturbances". Whereas the resilience metrics measure the performance of a power system

when exposed to "extreme events". This difference in including extreme events primarily separates the resilience metrics from the reliability indices.

There is a multitude of resilience metrics and their calculation methodologies proposed, a review of which is given in Raoufi et al. (2020); Liu et al. (2022). A few examples of such resilience metrics are the Unsupplied Load, Generation Capacity Not Connected, Loss of Load Expectation (LOLE), Storm Average Interruption Duration Index (STAIDI), Number of Disconnected Customers, and Loss of Load Probability (LOLP). However, this study uses an event-based approach to calculating resilience metrics. We follow the specific approach devised by Carrington et al. (2021) that is summarized in the following paragraph. One peculiarity about this approach is that while other resilience assessment approaches found in the literature use pre-defined events, Carrington et al. (2021) used automatic data processing to extract events from the data and then calculate their metrics.

Outages in a power system occur and then get restored by the restoration crew. This is a common and routine phenomenon. But, when there is an extreme event (tornado, hurricane, wildfires, flood, etc.), these outages occur at an increased rate. Due to this, the outages start to accumulate. These accumulated outages are called outage events. The number of outages in each event represents the size of the event. Large-size events are considered resilience events since these are rare events resulting from abnormal circumstances in the power system. An example of such a resilience event is given in Figure 5.2.

Performance curves that track in time the negative of the cumulative number of unrestored outages or customers are often used to track the progress in time of real or simulated resilience events (Carrington et al. (2021)) as shown in Figure 5.2. In real distribution outage data, Carrington et al. (2020) extracts the events by noting the event start time at which outages start, and the performance curve drops below zero, and then the event end time when all the outages of the event are restored, and the performance curve returns to zero. In each event, the performance curves can always be decomposed into an outage process, that is, the cumulative number of outages or customers affected, and a restore process, that is, the cumulative number of outages or customers restored

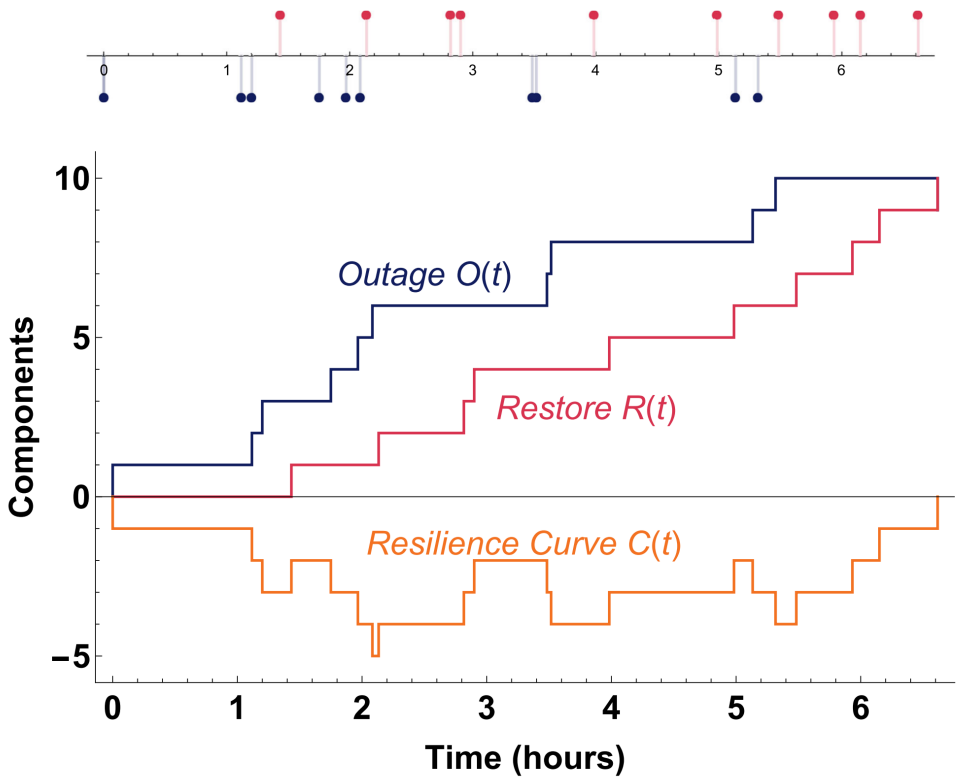

Figure 5.2 Example of a large size Resilience Event (Carrington et al. (2021))

Carrington et al. (2020). Outage and restore processes are shown in Figure 5.2. Then outage, restore, and overall process metrics for each event follow from the dimensions of the outage, restore, and performance curves, respectively. This extraction of events and obtaining the performance curves, outage and restore processes, and their associated metrics are applied to the practical processing of real utility data used in this thesis. The outage data in the distribution system is divided into events using an automatic event processing algorithm, the "AC algorithm", as used by Carrington et al. (2020). This "AC algorithm" is explained in the following section. Resilience metrics are then calculated based on these events. The following is the list of resilience metrics calculated and used in the analysis in this work:

1. Event Size (number of outages in an event)
2. Event Duration

3. Event Restore Duration
4. Time to First Restore of Event
5. Customer-Hours of Event
6. Area under the Performance Curve of Event (also called Component-Hours of Event)

5.2.1 AC Algorithm

There are different ways to extract events from the outage data. One such method is introduced here; the "AC algorithm". "AC" stands for accumulation, thus "AC algorithm" abbreviates "Accumulation algorithm". A set of outages, sorted by the outage occurrence time, is given as input to this algorithm. It then iterates over all the outages, starting from the first one. It accumulates all those outages which overlap in time, i.e., the first outage is not restored yet, and another outage occurs, that outage is accumulated with the first one, and if yet another outage occurs before that second outage is restored, it gets accumulated as well. All the accumulated outages constitute an event, and this process continues until the last outage in the event is restored, and no more outage start before that.

5.3 Relationship between Resilience and Area Outage Rate Curves

The Area Outage Rate Curves for the wind hazard of an area in the distribution system shows the outage performance with respect to the wind speeds observed at a specific location in or near that area. Since resilience analysis is associated with extreme events, we want to know how the distribution system behaves at high wind speeds. We use the area outage rate curve to obtain this information. The area outage rate curve not only shows the range of wind speed observed in an specific location in the distribution system but also how the system reacted to each wind speed level during the observation period. A lower mean outage rates at lower wind speeds tell us that the distribution system in that area has been resilient to those wind speed levels. Similarly, the elevated mean outage rates at higher wind speed levels indicate a less resilient response. For example, let us look at Figure 4.10. We see that the wind speeds observed at WS-1 in Area-1 go up to 18 mph,

and the distribution system in this area is resilient to wind speeds (measured at weather station WS-1) up to 9 mph because the mean outage rate remains less than one outage per hour. However, the mean outage rates started to increase exponentially as the wind speed increased from 9 mph to 18 mph. Now, if this exponential increase started at 14 mph for the same distribution system, the outage rate curve would have shifted to the right. Similarly, if the exponential increase had started at 6 mph, the outage rate curve would have shifted to the left. Therefore, in the context of assessing the system's resilience to wind hazard, we can say that if a given distribution system is more resilient to high wind speeds, then it would have an outage rate curve shifted towards the right, whereas if the same system is less resilient to the same wind hazard, then it would have an outage rate curve shifted towards the left. We use this important aspect of the area outage rate curves to model the impact of hardening investments on the distribution system. To see the impact of hardening on the wind hazard, we shift the area outage rate curve of the wind hazard to the right, resample the outages accordingly and then evaluate the improvement in the metrics.

5.3.1 Area Outage Rate Curve Shifting

We refer to Figure 5.1 and start from the left side, where we have a wind speed value x. We want to quantify the impact on the resilience metrics if the distribution system is hardened to withstand x mph more wind speed. An example of this is the replacement of the existing distribution poles in a certain area with poles having higher resistance to wind loads, say resistance to 10 mph higher wind speeds as compared to the existing poles. The new poles have a certain cost that, along with the replacement cost of the existing poles, will be the total investment cost to improve the distribution system's resilience to heavy wind loads. This overall investment causes a decrease in the outage rates at each wind speed level. To study the impact and benefit of this investment, a right shift of x mph in the area outage rate curve is introduced, and the impacts are calculated. For example, if such a hardening investment had been made on the distribution system of Area-1, then the area outage rate curve of Area-1 given in Figure 4.8 would have been shifted to the right by x mph. This right shift gives us new mean outage rate values at each wind speed level. We then

do random sampling of outages (details in Section 5.4) to get a subset of outages corresponding to the shifted area outage rate curve and recompute the resilience metrics with that outage subset. Those resilience metrics are then compared with the original resilience metrics corresponding to the original outage rate curve (Figure 4.8) to measure the impact of the hardening investments.

Since the area outage rate curves follow an exponential model (Figure 4.10), as discussed in Section 4.6, we can easily calculate the change in the outage rates corresponding to an x mph shift in the area outage rate curves. After a right shift of x mph, the exponential model given in (4.10) becomes (5.1):

$$\overline{F}_{new}(v) = \overline{F}(v-x) = b_0 e^{q_1(v-x)} = \overline{F}(v)e^{-q_1 x} \tag{5.1}$$

Dividing the shifted model (5.1) by the original model (4.10) gives the multiplier of change $y(v)$ at wind speed value v:

$$y(v) = \frac{\overline{F}_{new}(v)}{\overline{F}(v)} = e^{-q_1 x} \tag{5.2}$$

Equation 5.2 shows that a right shift of x mph results in a uniform $y(v) = e^{-q_1 x}$ decrease in the mean outage rate at each wind speed v. If initially, we had an outage rate $\overline{F}(v')$ corresponding to k number of outages at wind speed v', then after the right shift, we get $\overline{F}_{new}(v')$ and a corresponding k_{new} number of outages for same wind speed level v', such that $\overline{F}_{new}(v') < \overline{F}(v')$ and $k_{new} < k$. The decreased number of outages k_{new} is realized by sampling from the original set of k outages (details in Section 5.4). Resilience metrics are then computed with this new subset of outages and compared with the pre-investment system's resilience metrics to get percentage improvements in each resilience metric. The shift in the exponential model of the area outage rate curve of Area-1 is shown in Figure 5.3.

5.4 Outages Sampling Design

This section describes the sampling design to select a reduced subset of outages from all the outages and then calculate the mean of the improved metric. The area outage rate curve gives the mean outage rate $\overline{F}(v)$ for each wind speed value v according to (4.2). A shift in the area outage

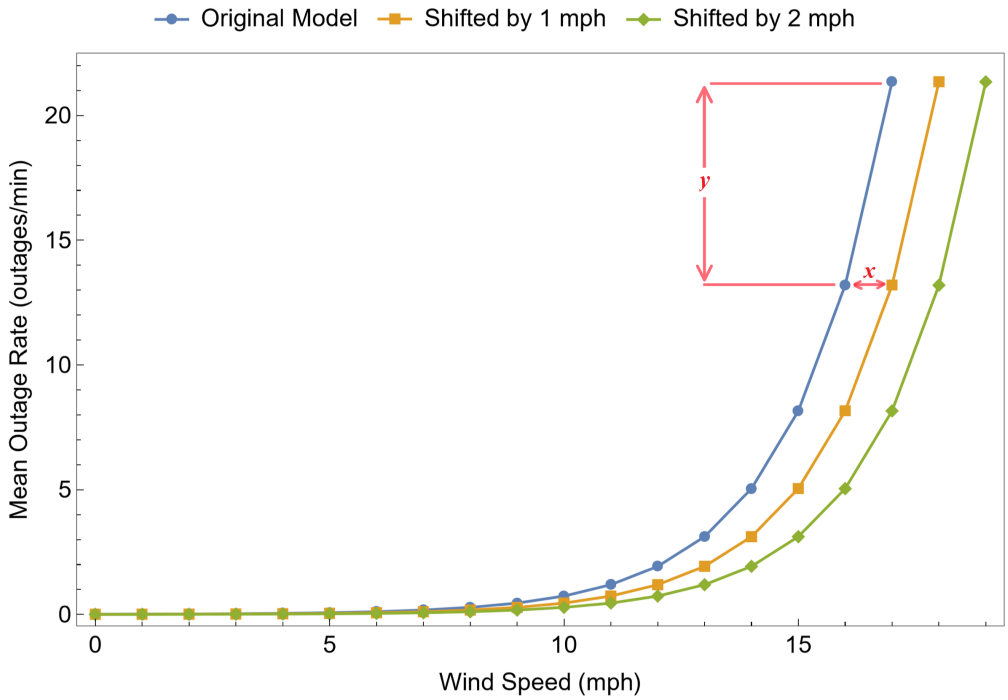

Figure 5.3 Comparison of Original and Shifted Area Outage Rate Curve of Area-1 after 1 mph and 2 mph hardening to wind hazard

rate curve gives new values for the mean outage rate $\overline{F}_{new}(v)$ for each wind speed value v. For a particular wind speed value v, while $\overline{F}(v)$ was calculated using an outages set $\{o_1, ..., o_k\}$ of size k, the $\overline{F}_{new}(v)$ is calculated using a new subset of outages $\{o_1, ..., o_{k_{new}}\}$ having size k_{new} (5.4) such that $k_{new} \leq k$ when a right-shift is introduced in the curve.

$$\frac{\overline{F}_{new}(v)}{\overline{F}(v)} = \frac{k_{new}}{k} \tag{5.3}$$

$$k_{new} = k\frac{\overline{F}_{new}(v)}{\overline{F}(v)} \tag{5.4}$$

In the case of exponential outage rate curves discussed in Section 4.6, (5.4) simplifies so that $\frac{\overline{F}_{new}(v)}{\overline{F}(v)} = e^{-qx}$ as detailed in (5.2). To get this new subset of k_{new} outages is from k outages we use simple random sampling without replacement. It is then used to calculate a function M of the outage data, where M is the value of a particular resilience metric. Therefore, each resilience metric (given in Section 5.2) has its own specific function M. If we call the outages subset S_i, then there are a total of $N = {}^{k}C_{k_{new}}$ possible subsets $\{S_1, ..., S_N\}$, each with size k_{new}. Each

subset results in a different value of the resilience metric M. This gives us a finite population $\{x_1, ..., x_N\}$ of size N of the resilience metric M with a finite variance σ^2 and mean μ. We want to estimate the mean of this finite population, i.e., the expected value of the resilience metric M. Since $N = {}^{k}C_{k_{new}}$, N has its maximum value when $k_{new} = k/2$. For smaller values of N, the mean value of the resilience metric can be computed directly by finding all S_i (the complete population), computing the value of the resilience metric for each S_i and then taking the mean of all values. However, the value of N increases exponentially, so for larger values of N, finding all the S_i subsets is computationally expensive. Therefore, the mean value of the resilience metric is estimated using the following approach.

We draw a simple random sample $\mathbf{X} = \{X_1, ..., X_n\}$ of size m from the population of a resilience metric M such that each of the ${}^{N}C_m$ possible samples have same probability of selection, i.e., $({}^{N}C_m)^{-1}$. It should be noted here that the concept of having an equal probability of section is typically associated with an infinite population, whereas in this case, the population of the resilience metric is not an infinite population. However, since the population size is very large, it can be assumed to have an infinite population, as removing one value from such a large population will negligibly affect the probability of selection of the other value. Therefore, assuming an infinite population, the sample mean and sample variance is defined by (5.5) and (5.6), respectively.

$$\bar{X} = \frac{1}{m}\sum_{i=1}^{m} X_i \tag{5.5}$$

$$S^2 = Var(X) = \frac{1}{m}\sum_{i=1}^{m}(X_i - \bar{X})^2 \tag{5.6}$$

$\bar{X}$ is a random variable itself because its value varies from one sample to another. As its estimated value (5.7) equals the population mean (the parameter we want to estimate), $\bar{X}$ is an unbiased estimator for the population's mean.

$$E(\bar{X}) = \mu_{\bar{X}} = \mu \tag{5.7}$$

Now we evaluate how many samples m are needed for acceptably close estimates. According to the central limit theorem, the sampling distribution of the sample mean $\bar{X}$ approaches the normal

distribution with a good approximation if the sample size $m \geq 30$ and $N > m/0.05$:

$$\bar{X} \sim N(\mu, \frac{\sigma^2}{m}) \tag{5.8}$$

$$\Rightarrow Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{m}} \sim N(0,1) \tag{5.9}$$

Therefore, the (1-α)100% confidence interval around $\bar{X}$ will be:

$$\left(\bar{x} - z_{\frac{\alpha}{2}} \cdot \sigma_{\bar{X}}\right) < \mu < \left(\bar{x} + z_{\frac{\alpha}{2}} \cdot \sigma_{\bar{X}}\right) \tag{5.10}$$

$$\left(\bar{x} - z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{m}}\right) < \mu < \left(\bar{x} + z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{m}}\right) \tag{5.11}$$

This means that the (1-α)100% confidence interval will include the true value of the population mean μ with probability 1-α, i.e., if $\alpha = 0.05$, the probability is about 0.95 that the 95% confidence interval will include the true population mean μ. $z_{\frac{\alpha}{2}}$ in (5.11) is the critical value of the standard normal distribution, which is just another word for 1-$\alpha/2$ quantile of the standard normal distribution. Its values can be found by (5.13)

$$\mathbf{P}(Z \leq z_{\frac{\alpha}{2}}) = \Phi(z_{\frac{\alpha}{2}}) = 1 - \frac{\alpha}{2} \tag{5.12}$$

$$z_{\frac{\alpha}{2}} = \Phi^{-1}(z_{\frac{\alpha}{2}}) \tag{5.13}$$

However, since we don't know the value of the population variance σ, we use the sample variance S^2 as defined in (5.6). Due to this substitution, we can't use Z defined in (5.9), instead we use T which follows a Student's t-distribution with $m-1$ degrees of freedom:

$$T = \frac{\bar{X} - \mu}{S/\sqrt{m}} \sim t_{m-1} \tag{5.14}$$

Now, the (1-α)100% confidence interval around $\bar{X}$ is:

$$\left(\bar{x} - t_{\frac{\alpha}{2},m-1} \cdot \frac{s}{\sqrt{m}}\right) < \mu < \left(\bar{x} + t_{\frac{\alpha}{2},m-1} \cdot \frac{s}{\sqrt{m}}\right) \tag{5.15}$$

Where $t_{\frac{\alpha}{2},m-1}$ is the critical value of the student-t distribution with $m-1$ degrees of freedom. The error in the estimation of the population mean μ is the absolute difference between the estimated value and the actual value:

$$E = |\bar{x} - \mu| \tag{5.16}$$

Suppose we specify the desired significance level α and an acceptable upper limit d of the estimation error E, such that $\mathbf{P}(E > d) < \alpha$. In that case, the minimum sample size can be approximated using the following equation:

$$d \leq t_{\frac{\alpha}{2},m-1} \cdot \frac{s}{\sqrt{m}} \tag{5.17}$$

$$\Rightarrow m \geq \left(t_{\frac{\alpha}{2},m-1} \cdot \frac{s}{d}\right)^2 \tag{5.18}$$

However, (5.18) has m on both sides of the equation, as one of the parameters of the t-distribution is the degrees of freedom, which is $m-1$. Since m is the quantity we try to find, we cannot find the value of $t_{\frac{\alpha}{2},m-1}$ as it requires the values of m. Therefore, we need to numerically solve (5.18) to find a solution. Another way to approach this problem is to use the z value, as given in (5.11):

$$\Rightarrow m \geq \left(z \cdot \frac{\sigma}{d}\right)^2 \tag{5.19}$$

But in that case, we need to specify the population variance σ, which is unknown in this case. Although the value of population variance can be estimated using a pilot study in the case of survey-based analysis, we cannot do that in the present case. Therefore two alternate approaches can be adopted to get an estimate for m:

1. We specify significance level α and then solve (5.18) iteratively by increasing the sample size m until the desired values of estimation error d is achieved.

2. Instead of using (5.18), we can use Standard Error SE to estimate the sample size m. We have already discussed that the distribution of the sample mean approaches Normal Distribution as the sample size increases. The standard deviation of the distribution of the sample mean is called the Standard Error. An estimator of this standard error is $\hat{SE} = s/\sqrt{m}$. The smaller the standard error, the narrower the distribution of the sample mean. Therefore, an iterative approach can be used again to calculate the standard error by increasing the sample size m until a desired value of standard error is achieved.

The first approach mentioned above is used to calculate the mean value of all resilience metrics with a 99% confidence level with the minimum number of samples m to minimize the computational resources.

5.4.1 Overall Hardening vs. Targeted Hardening

The approach explained above involves reducing all types of outages with equal probability. This helps us understand the effects of "Overall Hardening" of the distribution system. However, this same approach can be used to quantify the effects of "Targeted Hardening", where only specific outages are removed, and others are left as it is. An example of such targeted hardening is installing some squirrel deterrent systems in certain areas of the distribution system to keep the outages caused by the squirrels at a minimum. Yet another example of targeted hardening is trimming trees in a certain area with high wind-related outages. To see the impact of such targeted hardening on the system's resilience and reliability metrics, the outages with specific cause codes are removed only instead of all the outages with equal probability. The outage dataset used in this study has "Cause Codes" assigned to each outage identifying the cause behind the outage occurrence (details in subsection 3.2.2). Therefore, to simulate the effects of "targeted hardening", the outages of specific cause codes are isolated first, and then the methodology explained in this chapter is used to randomly remove/sample outages, and then the relevant metrics are calculated with the new outages subset.

5.5 Changes in Events after sampling

We have used an event-based resilience assessment approach, as discussed in Section 5.2. When outages are removed using sampling, the events get modified as well. The different ways in which an event can change after sampling are described in this section.

A timeline plot of a typical event of size 9 is shown in Figure 5.4. Each horizontal line represents an outage within the event, with bubbles at the start and end representing each outage's start and end time. The outages are labeled for easy referencing. There are 3706 resilience events in the

complete outage data. When a certain number of outages are removed randomly from the data, any outage can be removed from any event. The removal of an outage from an event can change the event in different ways depending on the position of that outage in the event. Correspondingly, the resilience metrics of that event can increase, decrease, or remain the same. Therefore, it is necessary to understand how an event can evolve due to outage removal.

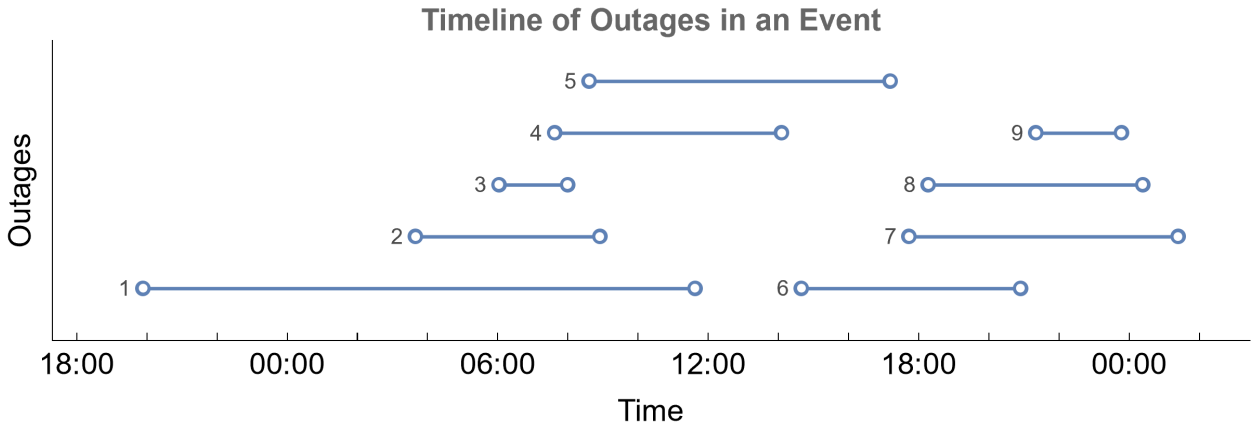

Figure 5.4 Timeline plot of a medium size resilience event

If we look at the timeline plot in Figure 5.4, we see that there are certain outages which, if removed, will change the event metrics significantly, whereas there are other outages that will not have a considerable impact on event's resilience metrics when removed from the event. For example, if Outage 1 is removed, it will decrease the event duration, number of customers, customer hours, and area of the performance curve. Similarly, if Outage 7 is removed, it will also decrease the event duration as it is the last outage of the event. However, if Outage 2 is removed, it will not affect the event duration as it is neither the first outage nor the last outage in the event. Outage 3 is the outage that is restored first, and removing it will increase the time to the first restore of the event because the next outage is Outage 2, which is restored later than Outage 3. This shows that removing the outages does not necessarily mean that all the event metrics will also decrease. In fact, the "Time to First Restore" metric is the only metric that can increase as a result of the removal of an outage from an event. And it, too, only increases if the removed outage

is that outage that is restored first in the event. Other than that, it will either decrease or stay the same. Two metrics that will always decrease, if an outage is removed, are the "Event Size" and "Component-Hours" (Area under the performance curve) because these are directly dependent on the number of outages in an event, and if an outage is removed from an event, then its size and area will decrease. Similarly, the "Number of Customers" and the "Customer-Hours" of an event will always decrease unless the removed outage has zero customers affected.

Apart from the change in event metrics, there are certain outages that, if removed, break the event into two events. For example, if we refer to Figure 5.4, we see that after the restoration of Outage 4 and the start of Outage 6, there is only one outage in the event, i.e., Outage 5. If this Outage 5 is removed, it will break the event into two, each with its own resilience metrics. Similarly, Outage 6 is the only outage between Outage 5 and Outage 7; if removed, the event will get divided into two events. More details about this are given in Section 5.5.1.

Each of the different event changes discussed above affects the resilience metrics of an event differently. To find the probability of a change in each resilience metric after sampling, we repeated the outage sampling process 2000 times to generate 2000 unique samples of 11444 outages out of the original 12715 outages of Area-1. Since the outage sampling is a random process, some events get changed, and some remain unchanged after sampling because no outage belonging to those events gets removed in the sampling process. The probability of occurrence of a specific change in each resilience metric of events that got changed after outage sampling is calculated and shown in Figure 5.5. We can see from Figure 5.5 that the "Time to first restore" remains the same more than 50% of the time, increasing more times than decreasing. Moreover, the event duration metric remains unchanged 48% of the time and decreases 52% of the time. The net decrease in the event duration metric is smaller than the other metrics. Similarly, event size and the component-hours of an event always decrease whenever an outage is removed from an event.

This section discussed the different possibilities of change in an event due to outage sampling and the impacts of those changes on each resilience metric of an event. However, how can these

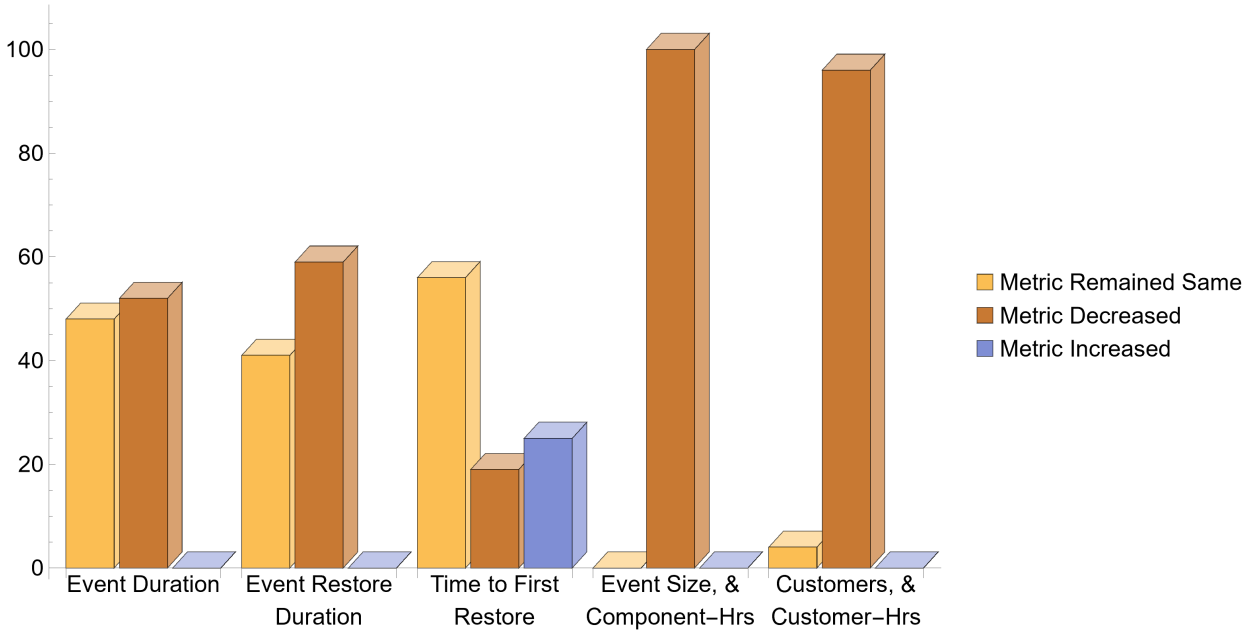

Figure 5.5 The probability of change in each resilience metric of events changed after outage sampling

event changes be tracked after outage sampling? The concept of "Super Events" is introduced to answer this question, which is explained in the next section.

5.5.1 Super Events

As discussed in the previous section, a resilience event in the outage data can change in different ways when outages are removed. In summary, there are four types of changes an event can undergo:

1. Deletion of small-size events, thus decreasing the overall total number of events.
2. Creation of new events, due to larger events getting split into smaller events, thus decreasing the number of larger size events and increasing the number of smaller size events.
3. Shrinkage of events due to removing one or more outages from them.
4. Events remaining intact, as no outage belonging to those events got removed in the random outage sampling process.

The original outage data of Area-1 has 3706 events, extracted from 12715 total outages, using the "AC algorithm" (explained in Section 5.2.1). The outage sampling is performed afterward, and the 3706 events change to 3706 "Super Events". A "Super Event" is the original event after it has undergone one of the four changes in the list above. Therefore a super event can be an empty set (if all the outages in it got deleted due to sampling), it can be a set of two or more events (due to event split), it can be a set containing "reduced" version of the original event (if some of the outages got deleted during sampling), or it can be a set containing the exact same event as it was before sampling. A super event thus encompasses all the changes that occurred to an event due to sampling. The super events approach allows efficient tracking of changes in original events and their metrics. Since each super event corresponds to the same event before sampling, event tracking before and after sampling becomes easy, and all the changes in an event can be easily identified. Moreover, Some of the events get deleted after the sampling process. If we compare the mean resilience metrics of all events before and after sampling, we will get inaccurate results as we will not adjust for the deleted events. So using the super events approach, the deleted events can also be tracked and added to the results to get an accurate metrics comparison.

In order to form "super events" after sampling is complete, the "AC algorithm" is run on each of the individual 3706 events, and it creates a "super event" encompassing all the different event changes. To understand how the resilience metrics are computed for a super event, suppose event A is split into events A1, A2, and A3. The super event A will have A1, A2, and A3 events in it. Metrics of A1, A2, and A3 are calculated separately and then summed together to get a single value for each metric of the super event. An exception is the "Time to First Restore". The mean of the "Time to First Restore" metric of events A1, A2, and A3 is calculated instead of the sum to get a more representative value. Similarly, the metrics for an empty super event are all zero. Suppose the event A gets deleted after sampling. In that case, the super event A still exists in the sampled outage data, with zero outages in it and thus zero assigned to all its resilience metrics. This way, the number of events in the original data and the number of super events in sampled data remain the same.

5.6 Results

Resilience metrics are first calculated using all outages in Area-1. These metrics are tabulated in Table 5.1. 3706 events are formed using all the outages in Area-1. Each of the resilience metrics shown in Table 5.1 is first calculated for each event, and then the mean of resilience metrics of all events is taken to get the mean value of each resilience metric. A right shift is then applied to the Area Outage Rate Curve of Area-1 (as discussed in Section 5.3.1), and reduced outage rates are obtained. For ease of understanding, the right shift is introduced such that the reduced outage rates are 10% lower than the original outage rates at each wind speed. There are 12715 outages in total in Area-1, and random sampling (as discussed in Section 5.4) 10% of outages at each wind speed resulted in a subset of a total of 11444 outages. We repeated this random sampling process m times to obtain m different subsets of 11444 outages each. The value of m is calculated using (5.18) by specifying the estimation error $|d|$ <0.01 and confidence level (1-α)=99%. Mean Resilience Metrics are calculated for each of the m subsets of outages, and then the Mean of all the Mean Resilience Metrics is taken, shown in the 3rd column of Table 5.1. The distribution of the each of the Mean Resilience Metrics is given in Figures B.1 - B.6 in Appendix B. Table 5.2 shows the results of a higher value of right shift resulting in 20% overall reduction in the outages at each wind speed.

Table 5.1 Improvement in Resilience Metrics with 10% reduction in overall outages of Area-1

Resilience Metric (Mean of all events)	Base Case Value	After Hardening	Percentage Change
Event Size (outages in an event)	3.43	3.09	-10.0%
Event Duration (hours)	4.29	4.05	-5.7%
Event Restore Duration (hours)	2.06	1.89	-8.5%
Time to First Restore (hours)	2.23	2.12	-4.9%
No. of Customers Affected	186.62	168.00	-10.0%
Customer-Hours	1573.63	1414.93	-10.1%
Area Under Performance Curve	32.40	29.15	-10.0%

Table 5.2 Improvement in Resilience Metrics with 20% reduction in overall outages of Area-1

Resilience Metric (Mean of all events)	Base Case Value	After Hardening	Percentage Change
Event Size (outages in an event)	3.43	2.75	-20.0%
Event Duration (hours)	4.29	3.79	-11.8%
Event Restore Duration (hours)	2.06	1.71	-17.2%
Time to First Restore (hours)	2.23	2.00	-10.2%
No. of Customers Affected	186.62	149.18	-20.0%
Customer-Hours	1573.63	1258.2	-20.0%
Area Under Performance Curve	32.40	25.93	-20.0%

We notice in Table 5.1 and Table 5.2 that all the metrics decrease with the decrease in the number of outages, or other words, all the resilience metrics show improvement in system resilience in response to the shift in the outage rate curve. However, while the Event Size, Number of Customers, Customer-Hours, and Component-Hours show an approximately same percentage decrease as the percentage decrease in the number of outages (i.e., 10% against 10% and 20% against 20%), the Event Duration, Event Restore Duration and the Time to First Restore metrics show comparatively less percentage change. To understand the reason behind this phenomenon, we refer to Figure 5.5. The time to first restore shows the least percentage decrease in Table 5.1 and Table 5.2 because it is the only metric that can increase after sampling, and it decreases more often than it increases, as shown in Figure 5.5 so the net decrease is less as compared to the other metrics. Moreover, the event duration metric shows the second least percentage decrease among all metrics because it remains unchanged 48% of the time and decreases 52% of the time, as per the results in Figure 5.5. For the same reason, the event restore duration shows only a 5.7% decrease for a 10% overall decrease in the number of outages.

5.6.1 Event Size distribution

Event Size is the number of outages in an event or a super event. The size of an event changes due to outage sampling; it either decreases or remains the same. The PDF of the empirical distribution event sizes in the original outage data of Area-1 is shown in Figure 5.6 on a log-log scale.

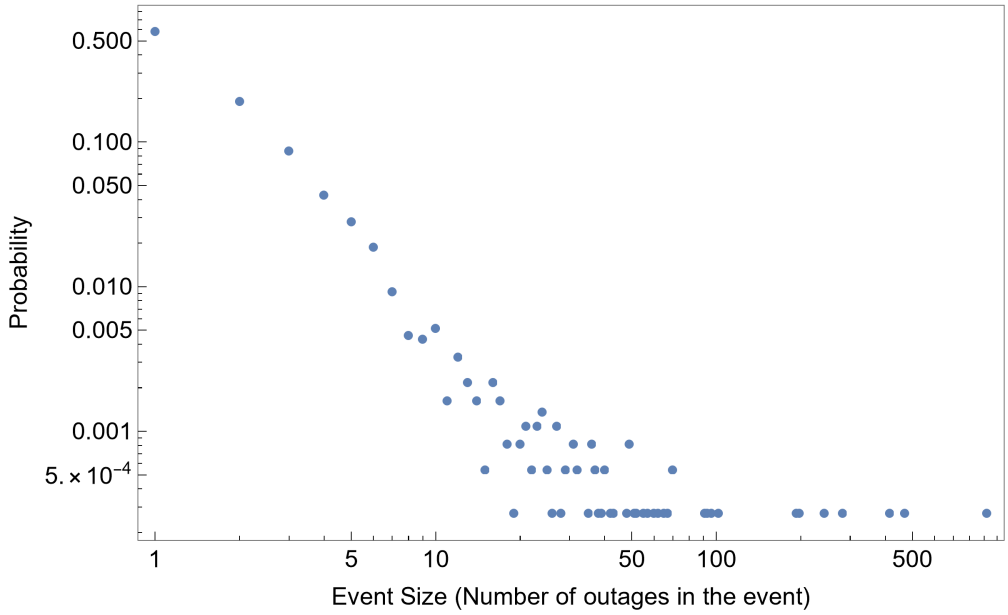

Figure 5.6 Empirical Distribution of Event Sizes of the events in Area-1 outages data (Log-Log scale)

We see two straight lines forming with negative slopes, one from event size 1 to 8 and the other from size 8 onwards. This hints at the power law behavior of the event sizes with a right-side heavy tail. In particular, this shows a significantly larger number of small events compared to large ones. Small events, particularly those with only one or two outages, are more likely to get deleted during outage sampling. Because they are large in number, they are more likely to get randomly selected during the outage sampling process, and second, they have only one or two outages which, if removed, deletes the event. Due to the abundance of small-size events, they dominate the mean metrics. Therefore, event size segregation is introduced below to calculate the resilience metrics of events of different sizes separately.

5.6.2 Results with Event Size segregation

To avoid skewed conclusions by lumping the resilience metrics of all event sizes together, the events are segregated into three groups:

1. Small: events with 1 outage
2. Medium: events with 2 to 15 outages
3. Large: events with more than 15 outages

Resilience metrics of each event size group are calculated separately before and after sampling and compared to quantify the improvements. The results are presented in Tables 5.3, 5.4, and 5.5 for small, medium and large event size groups, respectively. The results from these three tables are summarized as a bar chart in Figure 5.6.

Table 5.3 Improvement in the Mean value of Resilience Metrics of Small Events after 10% reduction in outages of Area-1

Resilience Metric	Base Case Value	After Hardening	Percentage Change
Event Size (outages in an event)	1	0.9	-10.0%
Event Duration (hours)	2.05	1.84	-10.0%
Event Restore Duration (hours)	0	0	0%
Time to First Restore (hours)	2.05	1.84	-10.0%
No. of Customers Affected	45.00	40.48	-10.0%
Customer-Hours	69.58	62.60	-10.0%
Area Under Performance Curve	2.05	1.84	-10.0%

A comparison of results in Table 5.1 and Tables 5.3, 5.4, 5.5 immediately shows that resilience metrics, before and after sampling, are very different in the segregated groups. This difference clearly shows the importance of the segregation of events. Starting from the results of small size events in Table 5.3, we see that the improvement is uniformly 10% in all of the resilience metrics for a 10% decrease in the outages. It is because the small-size events are all those events with only one outage. So after outage sampling, an event will exist or get deleted. There is no third option for the event change. Since there is only one outage in each event, the Event Restore Duration

Table 5.4 Improvement in the Mean value of Resilience Metrics of Medium Events after 10% reduction in outages of Area-1

Resilience Metric	Base Case Value	After Hardening	Percentage Change
Event Size (outages in an event)	3.44	3.10	-10.0%
Event Duration (hours)	5.56	5.29	-5.1%
Event Restore Duration (hours)	3.09	2.71	-12.3%
Time to First Restore (hours)	2.48	2.50	0.8%
No. of Customers Affected	159.75	143.71	-10.1%
Customer-Hours	350.46	315.33	-10.0%
Area Under Performance Curve	10.26	9.23	-10.0%

Table 5.5 Improvement in the Mean value of Resilience Metrics of Large Events after 10% reduction in outages of Area-1

Resilience Metric	Base Case Value	After Hardening	Percentage Change
Event Size (outages in an event)	62.48	56.21	-10.0%
Event Duration (hours)	37.67	37.14	-1.3%
Event Restore Duration (hours)	35.11	34.08	-2.9%
Time to First Restore (hours)	2.56	2.62	3.0%
No. of Customers Affected	4084.2	3679.4	-10.0%
Customer-Hours	58699.5	52775.1	-10.1%
Area Under Performance Curve	1142.45	1028.05	-10.0%

metric always has zero value, as it is the time difference between the start of the first restore and the completion of the last restore in the event.

There are 2142 small-size events, 1476 medium-size, and 88 large-size events. When the results (Table 5.1) are calculated without event segregation, the small and medium events, being large in number, significantly influence the overall mean values of metrics. This effect is more noticeable if we compare Table 5.4 with Table 5.1. For example, the base case values of the mean customer-hours and mean area under the performance curve in Table 5.1 are more than three times higher than the corresponding values shown in Table 5.4. We observe a pattern of decrease in the event

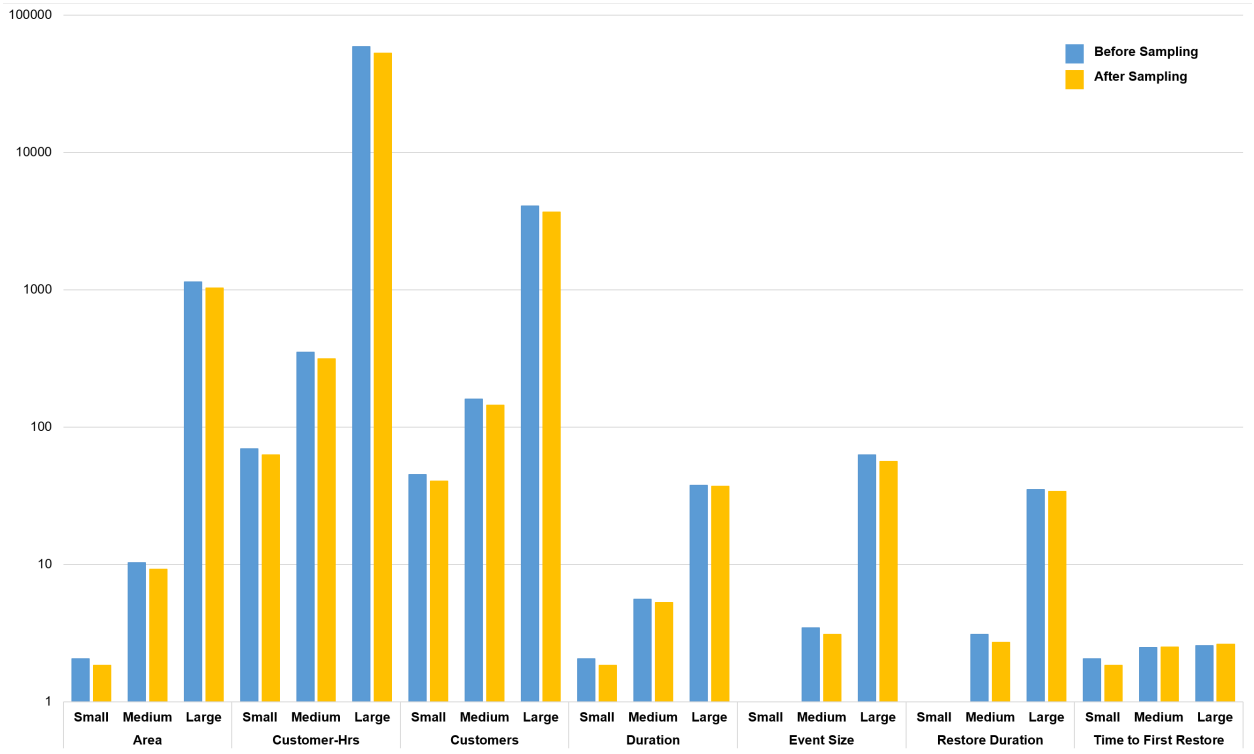

Figure 5.7 Comparison of Resilience Metrics with Event Size segregation (Vertical Log Axis)

resilience metrics of Table 5.4 similar to that observed in Table 5.1; i.e., all the resilience metrics decrease by approximately 10%, except the event duration, restore duration, and time to first restore metrics. The event restore duration decreases even more than 10% because the size of the events in the medium-size category starts from 2 outages per event up to 15 outages per event. Events in this event size range have the highest probability of getting split into two or more events for two reasons: large in number (1476 events) and small enough to easily get split. We already discussed in Section 5.5 that the event restore duration decreases significantly when an event gets split. This is why the event restore duration decreases more than 10% in Table 5.4. On the other hand, the "Time to First Restore" metric increased by almost 1%. This happens due to a large number of the medium size events. According to Figure 5.5, the time to first restore metric increases more times than it decreases when the metrics are calculated together (without event segregation).

However, this effect is more prominent in medium and large-size events, as can be seen in the segregated results in Table 5.3, 5.4, and 5.5. The time to first restore of changes by -10%, 0.8%, and 3% respectively for small, medium and large size events. The variation in each resilience metric according to the change in the event size is shown in Figures B.7 - B.14 in Appendix B.

Results of the large-size events are given in Table 5.5, which are particularly important because we are modeling resilience that deals with large events. We see some very large numbers in the base case results, like the average duration of a large-size event is 37.67 hours. The average duration reduces by 1.3% after sampling. So the event duration is not affected much by a reduction in the number of outages, at least not as much as the event size, customers, customer-hours, and component-hours. The "Time to First Restore" increases the most in this category, i.e., by 3%. This category has the largest events, ranging up to 922 outages in a single event. But the number of events is the smallest in this category, i.e., 88. When percentage outage sampling is performed, multiple outages are removed from each event due to the large number of outages. Therefore, multiple outages restored earlier in the event cue get removed as well, leading to an increase in the time to first restore of the event.

In conclusion, we have found that the "Event Size", "Number of Customers", "Customer Hours", and "Area under the Performance Curve" metrics decrease with approximately the same percentage by which the outages have been sampled. This decrease occurs with the same percentage, irrespective of the event size. Moreover, this has an almost linear relationship with the percentage reduction of outages, as seen in Tables 5.1 and 5.2.

CHAPTER 6. SUMMARY AND DISCUSSION

The resilience of distribution systems is important as most of the damage caused by extreme events is in the distribution systems. Utilities make investments to improve the resilience of the distribution networks. We study the impact of resilience-improving investments on the resilience metrics of small, medium, and large events in the distribution system. First, we model the impact of wind hazards on the mean outage rates in different areas of a distribution system in the form of "Area Outage Rate Curves", calculated from the observed outage and weather data. Then we use these curves to quantify the impacts of investments on resilience metrics. The area outage rate curves correlate observed wind speeds at weather stations in the distribution area with the mean outage rates of the area of the distribution system nearest the weather station. Resilience events and their metrics of size and customer impact are extracted from the outage data. Distribution system improvement decisions and investments in overhead lines shift the area outage rate curves. Shifting of curves provides reduced outage rates at each wind speed value. We use random outage sampling to select outages from the outage data to match the reduced outage rates. We then calculate the resilience metrics of small, medium, and large-scale events with the sampled outages to quantify the corresponding impact of investments. We test the developed techniques with outage data from a distribution system and wind speed data from weather stations within that distribution system.

6.1 Key Strengths and Contributions

The contributions and key strengths of this work are summarized as follows:

- "Area outage rate curves" are calculated to show the variation in the mean outage rates of a certain distribution system area in response to the wind speed observed in a nearby

weather station. The curves thus quantify the historical wind resilience of different areas in a distribution system with respect to an easily measured quantity.

- The impact of hardening investments on the resilience metrics of small, medium, and large-size events in the distribution system is quantified using area outage rate curves. Wind hazard hardening by x mph shift area outage rate curves and reduce outage rates. Reduced outage rates are implemented by sampling from the historical data. Resilience metrics of small, medium, and large events are calculated from the sampled data and compared with the original resilience metrics to assess the impact of x mph wind hardening investments. This information is expected to help make decisions about the types and benefits of investments and maintenance activities to get the maximum improvement in the system's performance. Area outage rate curves have strong overall similarities to the fragility curves pioneered by Dunn (Dunn et al. (2018, 2015)) using British outage and wind data but are intended to be used as a direct description of area wind resilience with respect to a specific wind measurement rather than a fragility model for a larger region to be used as a damage model for infrastructure damage assessment. The concept of Dunn's fragility curves is similar to the area outage rate curves, but many technical and implementation details differ.

- We have used complete outage data of the system and presented the results separately for small, medium, and large size events. The segregated results of small, medium, and large events show which particular metric and group of events are affected the most by a certain investment. Thus, decisions and investments can be prioritized accordingly. In particular, the impact characterization of large, infrequent events enables an evaluation of resilience beyond reliability considerations.

- A data-driven approach is used to calculate the distribution system's resilience in contrast to model-based approaches with modeling assumptions. The absence of models and fewer assumptions greatly improves the realism of the results. The historical outage data of the distribution system captures in it all the operational complexities, system configuration changes,

component behaviors, and other factors. Therefore, the "Area outage rate curves" developed using this data encompass all kinds of system characteristics and variations that are difficult and computationally expensive to capture in model-based analysis, thus providing a robust quantification of historical wind resilience.

- The outage data and weather data required to use this approach are easily available to utilities, and the process is fast and does not require many computational resources.

- The area outage rate curves provide information regarding which outage cause codes or cause code categories are more sensitive to a certain hazard. For example, the dependence of various cause categories on wind hazard is shown in Figure 4.13 and 4.14. This process can be adapted to identify outages caused by a particular hazard (for example, wind in this case) even if this cause code information is unavailable in the data.

- The concept of "Super Events" is introduced and implemented to track event changes after random sampling of outages from the outage data. It provides a simple yet effective way to keep track of all the changes in an event, such as event split and event deletion.

6.2 Conclusions

Electricity is a basic need of every society, and its dependence is increasing steadily. Therefore, an uninterrupted supply of electricity is a must for economic growth as well as social harmony. However, due to the wide geographical spread of the power systems and their exposure to weather elements, electricity interruptions are inevitable. These interruptions can occur in electric power generation, transmission, and distribution systems. However, the frequency of their occurrence is the highest in distribution systems. There are different causes behind outages in power distribution systems, including adverse weather, equipment failure, human errors, natural elements, and accidents. Out of all these causes, trees, where present, are the major cause of outages. For the distribution system studied in this work, we find that 38% of the total outages are caused by trees present within and outside the distribution lines' clearance zone. The wind is the most significant

contributor to outages in the distribution system among all weather elements. Out of 14% weather-caused outages, 47% are found to be caused by wind and rain. But the tree-caused outages are also indirectly caused by wind because trees and tree branches break down due to heavy winds and cause outages in the system. Therefore, combining those outages, we find that 44% of the total outages are caused directly or indirectly by the wind.

Power interruptions can be small and large durations, ranging from several hours to even days. However, larger interruptions are less common as compared to smaller ones. Resilience analysis is performed to study and mitigate the large and less frequent occurring power outages. We introduce area outage rate curves to study the historical resilience of the power distribution systems. The area outage rate curves tell us that the average outage rates in the distribution system remain very low up to a certain wind speed level (that varies from one area of the distribution system to the other), and after that, these outage rates start to rise exponentially. That is, an exponential relationship exists between wind speeds and the average outage rates in the distribution system.

Distribution system improvement decisions and investments in overhead lines shift the area outage rate curves. Outage rates at each wind speed level reduce corresponding to this shift. Outages at each wind speed level are sampled according to these reduced outage rates, and thus a smaller set of outages is obtained. This subset is used to calculate the corresponding impacts on the resilience metrics of small, medium, and large-scale events. Events with one outage are categorized as small events, 2 to 15 outages as medium-size events, and events with more than 15 outages as large-size events. The results show that investment to improve the distribution system's resilience affects events' resilience metrics differently according to the event size. For a 10% reduction in outages, small-size events show a uniform 10% reduction in all resilience metrics; event size, event duration, event restore duration, time to first restore, number of customers, customer hours, and area under performance curve. For medium and large-size events, all resilience metrics reduce by 10%, corresponding to a 10% reduction in outages, except the event duration, event restore duration, and time to first restore metrics. The event duration decrease by 5% -1% for medium and large events, whereas the event restore duration decrease by 12% and 3%, respectively. This shows

that small, medium, and large-size events need to be studied separately for resilience investment purposes.

6.3 Publications

The research work presented in this thesis is being rewritten for submission to a peer-reviewed journal.

6.4 Future Work Avenues

Following are some of the future work that can be explored:

- We have quantified the improvements in the resilience metrics of events corresponding to investments. This work can be extended to conduct a data-driven risk analysis to show how a hardening investment changes the risk of small, medium, and large events.
- The area outage rate curves are generated using wind data only; however, other weather parameters like temperature and precipitation are also available in the weather. These can be used to model the system behavior to high temperatures and rains.
- The devised resampling of reduced outages methodology can also be used to evaluate the impact on reliability indices. For example, it can be used to answer questions like how much the SAIDI would have been improved if squirrel guards were placed in a certain area of the distribution system.
- Future work can use elaborate weather data using Climate Reanalysis Models and combine it with wind data from weather stations to calculate more detailed area outage rate curves.
- In our work, we assumed that for a certain investment, there would be a certain percent hardening of the system components to the wind hazard, and then we used that percentage to conduct the analysis. However, there is a research gap in the detailed translation of investments to the actual hardening of the system.

- We have shown the results with overall hardening; however, the technique can also be used to see the impacts of targeted hardening using outage cause codes. For example, if we were to invest in the installation of squirrel guards, then we would only remove the outages caused by squirrels instead of removing outages from all the cause categories.
- This technique can be extended to do a comparative analysis between the impacts of hardening and faster restoration in the distribution system.

BIBLIOGRAPHY

Association, W. M. et al. (2022). State of the global climate 2021.

Aven, T. (2011). On some recent definitions and analysis frameworks for risk, vulnerability, and resilience. *Risk Analysis*, 31(4):515–522.

Bialek, J. (2020). What does the GB power outage on 9 August 2019 tell us about the current state of decarbonised power systems? *Energy Policy*, 146:111821.

Bidaoui, H., Abbassi, I. E., Bouardi, A. E., and Darcherif, A. (2019). Wind speed data analysis using Weibull and Rayleigh distribution functions, case study: Five cities Northern Morocco. *Procedia Manufacturing*, 32:786–793. 12th International Conference Interdisciplinarity in Engineering, INTER-ENG 2018, 4–5 October 2018, Tirgu Mures, Romania.

Biglari, M. and Formisano, A. (2020). Damage probability matrices and empirical fragility curves from damage data on masonry buildings after Sarpol-e-zahab and Bam earthquakes of Iran. *Frontiers in Built Environment*, 6.

Bjarnadottir, S., Li, Y., and Stewart, M. G. (2013). Hurricane risk assessment of power distribution poles considering impacts of a changing climate. *Journal of Infrastructure Systems*, 19(1):12–24.

Bruneau, M., Chang, S. E., Eguchi, R. T., Lee, G. C., O'Rourke, T. D., Reinhorn, A. M., Shinozuka, M., Tierney, K., Wallace, W. A., and Von Winterfeldt, D. (2003). A framework to quantitatively assess and enhance the seismic resilience of communities. *Earthquake spectra*, 19(4):733–752.

CAISO, CPUC, and CEC (2021). Final root cause analysis, mid-august 2020 extreme heat wave. Technical report, California Independent System Operator (CAISO).

Carrington, N. K., Dobson, I., and Wang, Z. (2021). Extracting resilience metrics from distribution utility data using outage and restore process statistics. *IEEE Transactions on Power Systems*, 36(6):5814–5823.

Carrington, N. K., Ma, S., Dobson, I., and Wang, Z. (2020). Extracting resilience statistics from utility data in distribution grids. In *2020 IEEE Power & Energy Society General Meeting (PESGM)*, pages 1–5.

Carta, J., Ramírez, P., and Velázquez, S. (2009). A review of wind speed probability distributions used in wind energy analysis: Case studies in the Canary Islands. *Renewable and Sustainable Energy Reviews*, 13(5):933–955.

Casey, J. A., Fukurai, M., Hernández, D., Balsari, S., and Kiang, M. V. (2020). Power outages and community health: a narrative review. *Current environmental health reports*, 7:371–383.

Climate Central (2022). Surging power outages and climate change. Technical report, Climate Central. September 12, 2022.

Conradsen, K., Nielsen, L., and Prahm, L. (1984). Review of Weibull statistics for estimation of wind speed distributions. *Journal of Applied Meteorology and Climatology*, 23(8):1173–1183.

Dale, L., Carnall, M., Wei, M., Fitts, G., and McDonald, S. L. (2018). Assessing the impact of wildfires on the California electricity grid. *California Energy Commission. CCCA4-CEC-2018-002*, 1.

Davidson, R. A., Liu, H., Sarpong, I. K., Sparks, P., and Rosowsky, D. V. (2003). Electric power distribution system performance in Carolina hurricanes. *Natural Hazards Review*, 4(1):36–45.

DOE (2005). Florida state's energy emergency response to the 2004 hurricanes. online. Retrieved from: "https://www.oe.netl.doe.gov/docs/fl2004energy.pdf" on 02.28.2023. A report published by the United State Department of Energy, Office of Electricity Delivery and Energy Reliability.

DOE-EPSA (2015). Climate change and the U.S. energy sector: Regional vulnerabilities and resilience solutions. Technical report, U.S. Department of Energy's Office of Energy Policy and Systems Analysis (DOE-EPSA). October, 2015.

Dorvlo, A. S. (2002). Estimating wind speed distribution. *Energy Conversion and Management*, 43(17):2311–2318.

Dunn, S., Wilkinson, S., Alderson, D., Fowler, H., and Galasso, C. (2018). Fragility curves for assessing the resilience of electricity networks constructed from an extensive fault database. *Natural Hazards Review*, 19(1):04017019.

Dunn, S., Wilkinson, S., Galasso, C., Manning, L., and Alderson, D. (2015). Development of empirical fragility curves for electrical supply systems subjected to wind hazard. In *International Conference on Applications of Statistics and Probability in Civil Engineering (ICASP) (12th - 2015)*.

E-ISAC (2016). Analysis of the cyber attack on the Ukrainian power grid. Technical report, E-ISAC: Electricity Information Sharing and Analysis Center. March 18, 2016.

EATON (2018). Blackout tracker, United States annual report 2018. online. accessed: 02.27.2023.

Eaton (2023). Blackout and power outage tracker. online. accessed: 02.27.2023.

Feng, M. Q., Shinozuka, M., Kim, H.-K., and Kim, S.-H. (2000). Statistical analysis of fragility curves. *Journal of Engineering Mechanics*.

FERC (2018). Order terminating rulemaking proceeding, initiating new proceeding, and establishing additional procedures - issued January 8, 2018. online. Docket Nos. RM18-1-000 and AD18-7-000 (accessed: 02.27.2023).

FERC (2021). The February 2021 cold weather outages in Texas and the South Central United States — FERC, NERC and regional entity staff report. Technical report, Federal Energy Regulatory Commission (FERC).

Finster, M., Phillips, J., and Wallace, K. (2016). Front-line resilience perspectives: The electric grid. Technical report, Global Security Sciences Division, Argonne National Laboratory, United States.

Garcia, A., Torres, J., Prieto, E., and de Francisco, A. (1998). Fitting wind speed distributions: a case study. *Solar Energy*, 62(2):139–144.

Gholami, A., Aminifar, F., and Shahidehpour, M. (2016). Front lines against the darkness: Enhancing the resilience of the electricity grid through microgrid facilities. *IEEE Electrification Magazine*, 4(1):18–24.

Gholami, A., Shekari, T., Amirioun, M. H., Aminifar, F., Amini, M. H., and Sargolzaei, A. (2018). Toward a consensus on the definition and taxonomy of power system resilience. *IEEE Access*, 6:32035–32053.

IEA (2022a). Electricity market report - January 2022. Technical report, International Energy Agency.

IEA (2022b). World energy outlook 2022. Technical report, International Energy Agency.

Insights, M. E. (2022). Global energy perspective 2022. online. Retrieved from: "https://www.mckinsey.com/industries/oil-and-gas/our-insights/global-energy-perspective-2022" on 03.05.2023. A report developed by Energy Insights in collaboration with McKinsey Sustainability and the Global Energy and Materials and Advanced Industries practices. Published in April-2022.

Jasiūnas, J., Lund, P. D., and Mikkola, J. (2021). Energy system resilience – a review. *Renewable and Sustainable Energy Reviews*, 150:111476.

Jowder, F. A. L. (2006). Weibull and Rayleigh distribution functions of wind speeds in Kingdom of Bahrain. *Wind Engineering*, 30(5):439–445.

Justus, C., Hargraves, W., Mikhail, A., and Graber, D. (1978). Methods for estimating wind speed frequency distributions. *Journal of Applied Meteorology (1962-1982)*, pages 350–353.

LaCommare, K. H., Eto, J. H., Dunn, L. N., and Sohn, M. D. (2018). Improving the estimated cost of sustained power interruptions to electricity customers. *Energy*, 153:1038–1047.

Lallemant, D., Kiremidjian, A., and Burton, H. (2015). Statistical procedures for developing earthquake damage fragility curves. *Earthquake Engineering & Structural Dynamics*, 44(9):1373–1389.

Liu, W., Yao, Y., and Jain, R. (2022). Quantitative power system resilience metrics and evaluation approach. In *2022 IEEE Power & Energy Society General Meeting (PESGM)*, pages 1–5. IEEE.

Matthewman, S. and Byrd, H. (2014). Blackouts: a sociology of electrical power failure. *Social Space*, pages 1–25.

Mohamed Nazri, F. (2018). *Fragility Curves*, pages 3–30. Springer Singapore, Singapore.

Morgan, E. C., Lackner, M., Vogel, R. M., and Baise, L. G. (2011). Probability distributions for offshore wind speeds. *Energy Conversion and Management*, 52(1):15–26.

Murray, K. and Bell, K. R. (2014). Wind-related faults on the GB transmission network. In *2014 International Conference on Probabilistic Methods Applied to Power Systems (PMAPS)*, pages 1–6. IEEE.

NASEM (2017). *Enhancing the Resilience of the Nation's Electricity System.* The National Academies Press, Washington, DC. National Academies of Sciences, Engineering, and Medicine.

NCDC (2015). Federal climate complex data documentation for integrated surface data. NCDC: National Climatic Data Center, URL:https://www.ncei.noaa.gov.

NIAC (2009). Critical infrastructure resilience: Final report and recommendations. Technical report, National Infrastructure Advisory Council (NIAC). September 8, 2009.

NOAA et al. (2023). U.S. climate extremes index (cei) - January 2023.

Ouarda, T., Charron, C., Shin, J.-Y., Marpu, P., Al-Mandoos, A., Al-Tamimi, M., Ghedira, H., and Al Hosary, T. (2015). Probability distributions of wind speed in the UAE. *Energy Conversion and Management*, 93:414–434.

Our world in data (2023). Global reported natural disasters by type, 1970 to 2022. online. Published online at OurWorldInData.org. Retrieved from: "https://ourworldindata.org/grapher/natural-disasters-by-type" on 02.28.2023. Data published by: EM-DAT, CRED / UCLouvain, Brussels, Belgium.

Ouyang, M. and Dueñas-Osorio, L. (2014). Multi-dimensional hurricane resilience assessment of electric power systems. *Structural Safety*, 48:15–24.

Ouyang, M., Dueñas-Osorio, L., and Min, X. (2012). A three-stage resilience analysis framework for urban infrastructure systems. *Structural Safety*, 36-37:23–31.

PJM Interconnection (2017). PJM's evolving resource mix and system reliability - March 30, 2017. *March*, 30:2017.

Portner, H.-O., Roberts, D., Masson-Delmotte, V., Zhai, P., Tignor, M., Poloczanska, E., Mintenbeck, K., Alegría, A., Nicolai, M., Okem, A., Petzold, J., Rama, B., and (eds.), N. W. (2019). IPCC special report on the ocean and cryosphere in a changing climate. Technical report, IPCC (Intergovernmental Panel on Climate Change).

Preston, B. L., Backhaus, S. N., Ewers, M., Phillips, J. A., Silva-Monroy, C., Dagle, J., Tarditi, A., Looney, J., and King, T. (2016). Resilience of the US electricity system: A multi-hazard perspective. *US Department of Energy Office of Policy. Washington, DC.*

Raoufi, H., Vahidinasab, V., and Mehran, K. (2020). Power systems resilience metrics: A comprehensive review of challenges and outlook. *Sustainability*, 12(22).

Reed, D. A. (2008). Electric utility distribution analysis for extreme winds. *Journal of Wind Engineering and Industrial Aerodynamics*, 96(1):123–140.

Reed, D. A. (2009). *Multi-hazard analysis of electric power delivery systems*, pages 1–7. American Society of Civil Engineers. Technical Council on Lifeline Earthquake Engineering Conference (TCLEE) 2009, DOI: 10.1061/41050(357)30, URL: https://ascelibrary.org/doi/abs/10.1061/41050%28357%2930.

Reed, D. A., Kapur, K. C., and Christie, R. D. (2009). Methodology for assessing the resilience of networked infrastructure. *IEEE Systems Journal*, 3(2):174–180.

Rosen, K., Van Buskirk, R., and Garbesi, K. (1999). Wind energy potential of coastal eritrea: An analysis of sparse wind data. *Solar Energy*, 66(3):201–213.

Seguro, J. and Lambert, T. (2000). Modern estimation of the parameters of the Weibull wind speed distribution for wind energy analysis. *Journal of Wind Engineering and Industrial Aerodynamics*, 85(1):75–84.

Shi, H., Dong, Z., Xiao, N., and Huang, Q. (2021). Wind speed distributions used in wind energy assessment: A review. *Frontiers in Energy Research*, 9.

Shuai, M., Chengzhi, W., Shiwen, Y., Hao, G., Jufang, Y., and Hui, H. (2018). Review on economic loss assessment of power outages. *Procedia Computer Science*, 130:1158–1163. The 9th International Conference on Ambient Systems, Networks and Technologies (ANT 2018) / The 8th International Conference on Sustainable Energy Information Technology (SEIT-2018) / Affiliated Workshops.

SIRTF (2012). Severe impact resilience: Considerations and recommendations. *NERC, May*, 4.

Stevens, M. and Smulders, P. (1979). The estimation of the parameters of the Weibull wind speed distribution for wind energy utilization purposes. *Wind Engineering*, 3(2):132–145.

Suri, N., Cabri, G., et al. (2014). *Adaptive, dynamic, and resilient systems (1st Edition)*. Auerbach Publications.

Takle, E. S. and Brown, J. M. (1978). Note on the use of Weibull statistics to characterize wind-speed data. *Journal of Applied Meteorology and Climatology*, 17(4):556 – 559.

Tuller, S. E. and Brett, A. C. (1985). The goodness of fit of the Weibull and Rayleigh distributions to the distributions of observed wind speeds in a topographically diverse area. *Journal of Climatology*, 5(1):79–94.

Wais, P. (2017). Two and three-parameter Weibull distribution in available wind power analysis. *Renewable energy*, 103:15–29.

Wong, S. D., Broader, J. C., and Shaheen, S. A. (2020). Review of California wildfire evacuations from 2017 to 2019. *UC Office of the President: University of California Institute of Transportation Studies*.

Younesi, A., Shayeghi, H., Wang, Z., Siano, P., Mehrizi-Sani, A., and Safari, A. (2022). Trends in modern power systems resilience: State-of-the-art review. *Renewable and Sustainable Energy Reviews*, 162:112397.

Zhou, J., Erdem, E., Li, G., and Shi, J. (2010). Comprehensive evaluation of wind speed distribution models: A case study for North Dakota sites. *Energy Conversion and Management*, 51(7):1449–1458.

APPENDIX A. ZONE-2 RESULTS

The results (figures and tables) of Area-1 are presented in the main body of this thesis, whereas the results of Area-2 are given in this Appendix to avoid too many pictures in the main text.

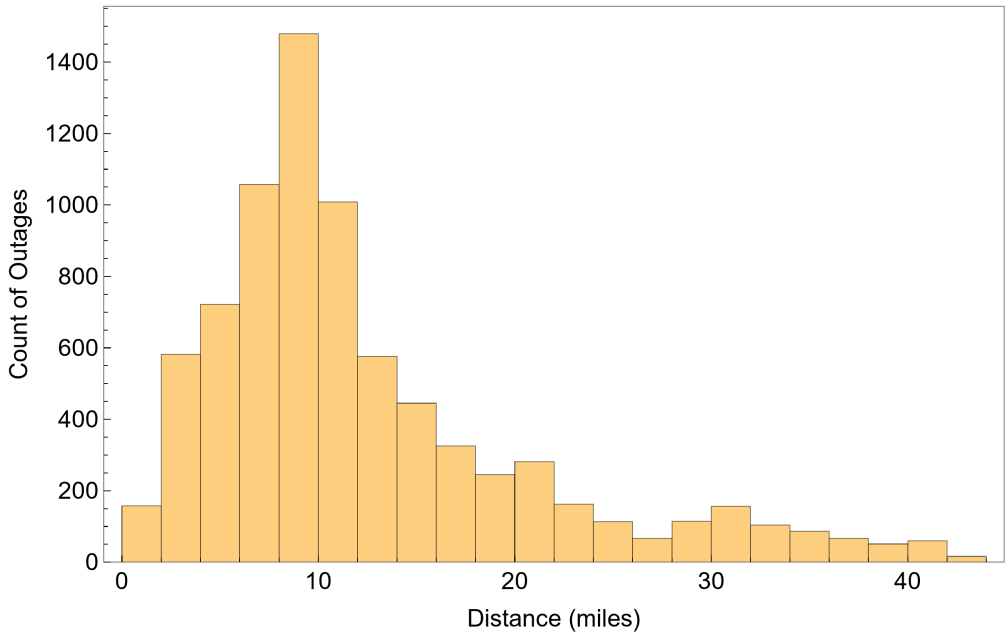

Figure A.1 Distribution of distance between outages and weather station in Area-2

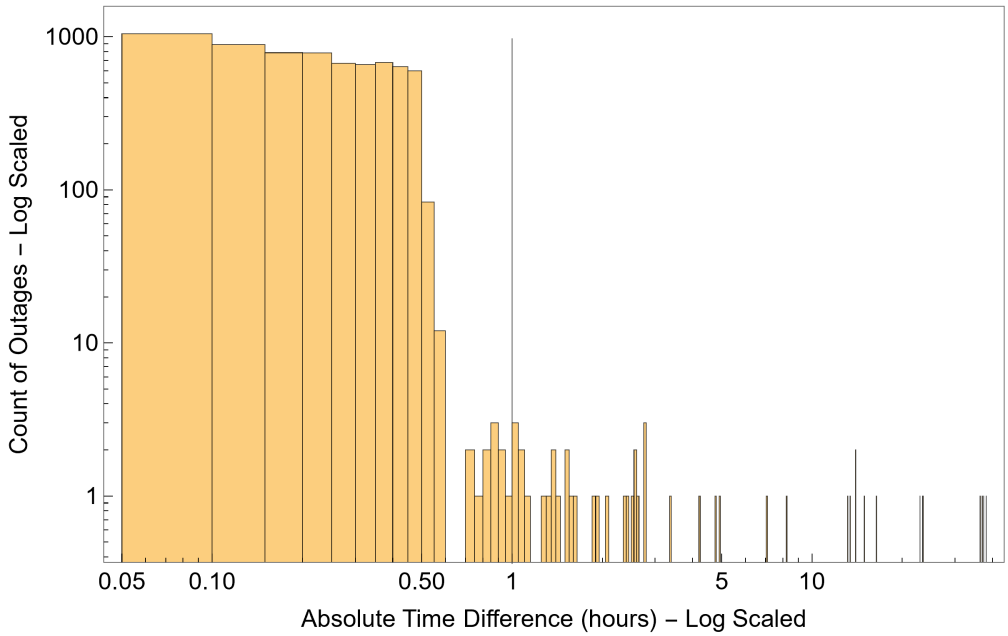

Figure A.2 Frequency of time differences between outages and wind speed observations of Area-2 (log-log scale)

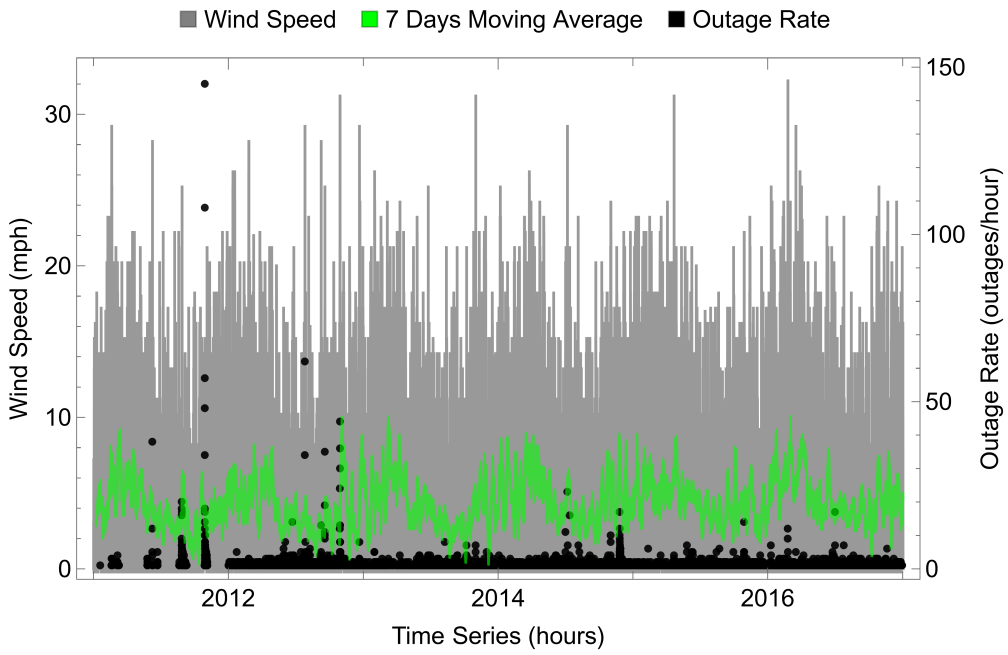

Figure A.3 Time series plot of wind speed and outages data of Area-2

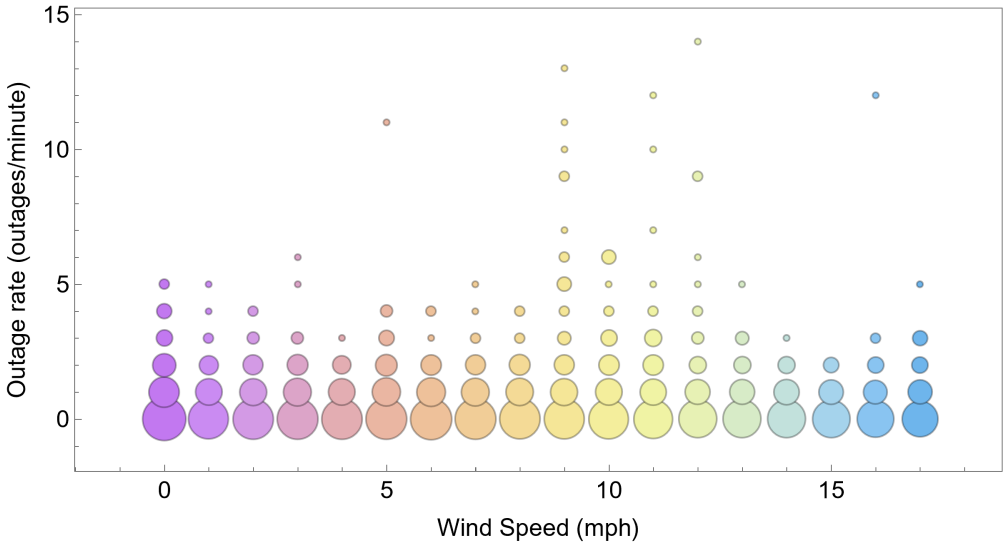

Figure A.4 Individual Outage Rates at each wind speed level in Area-2

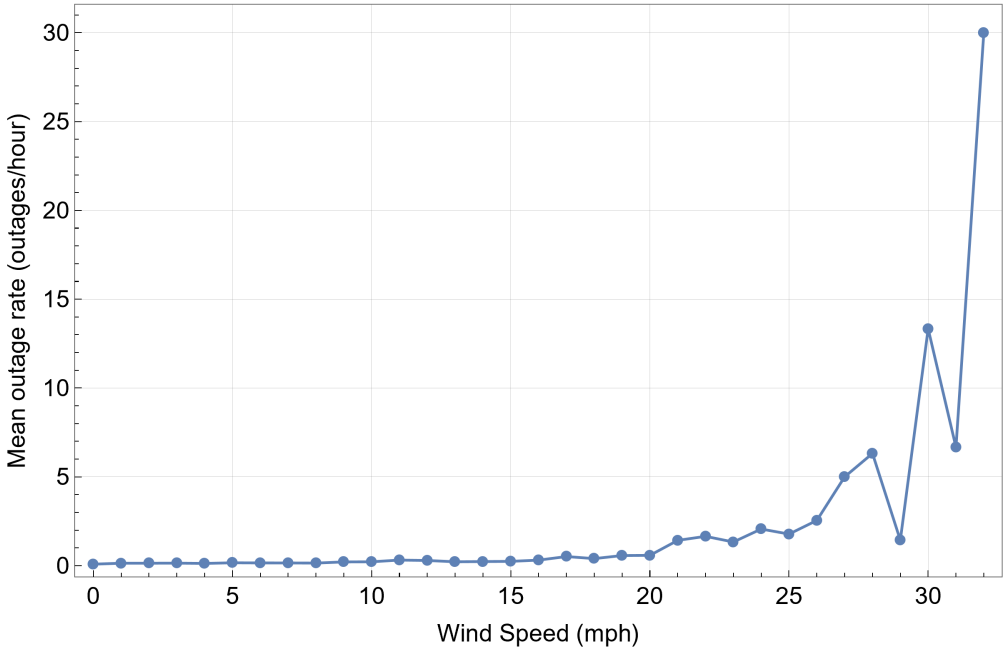

Figure A.5 Area outage rate curve of Area-2

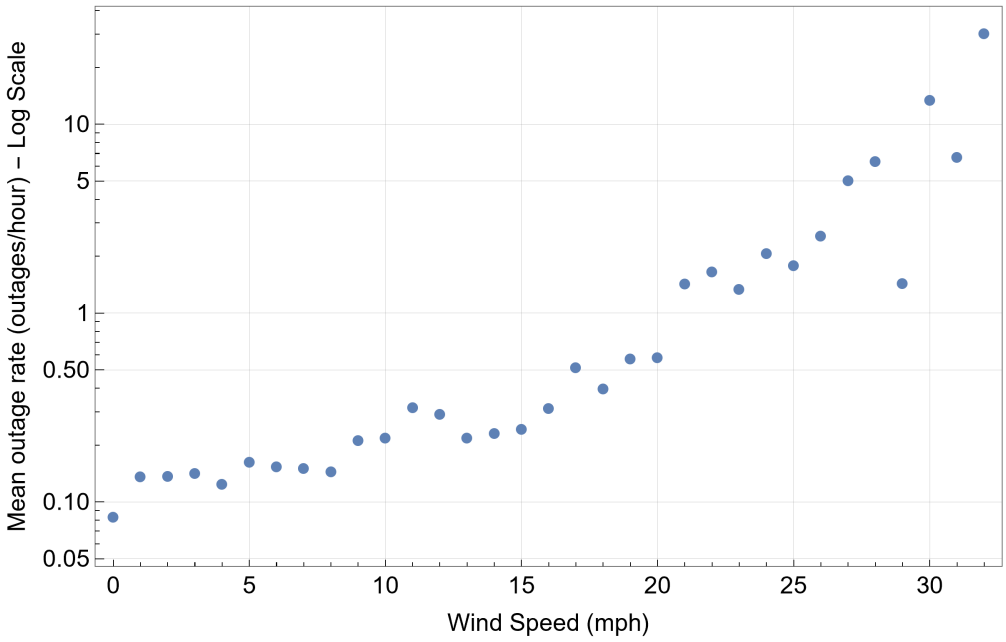

Figure A.6 Log-Linear plot of the Area Outage Rate Curve of Area-2

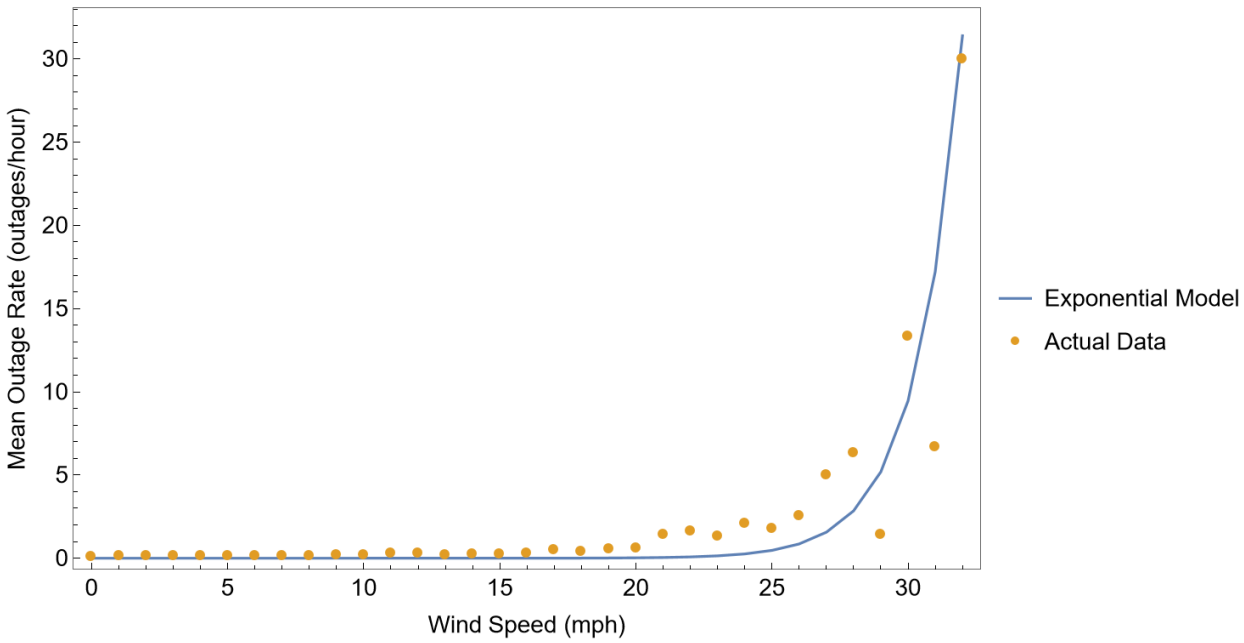

Figure A.7 Estimated Area Outage Rate Curve of Area-1

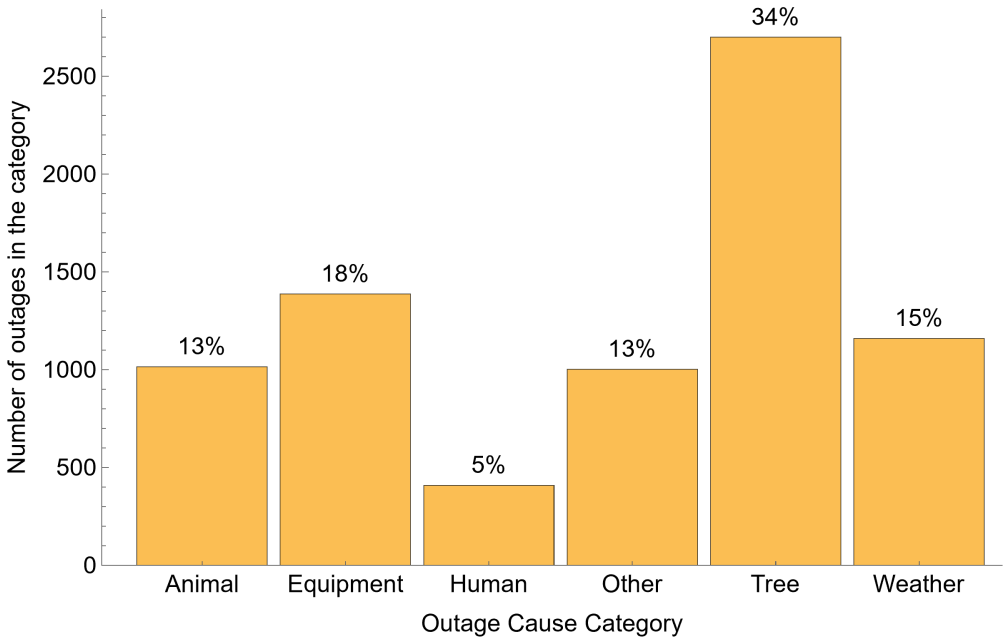

Figure A.8 Percentage of outages in each cause code category of Area-2

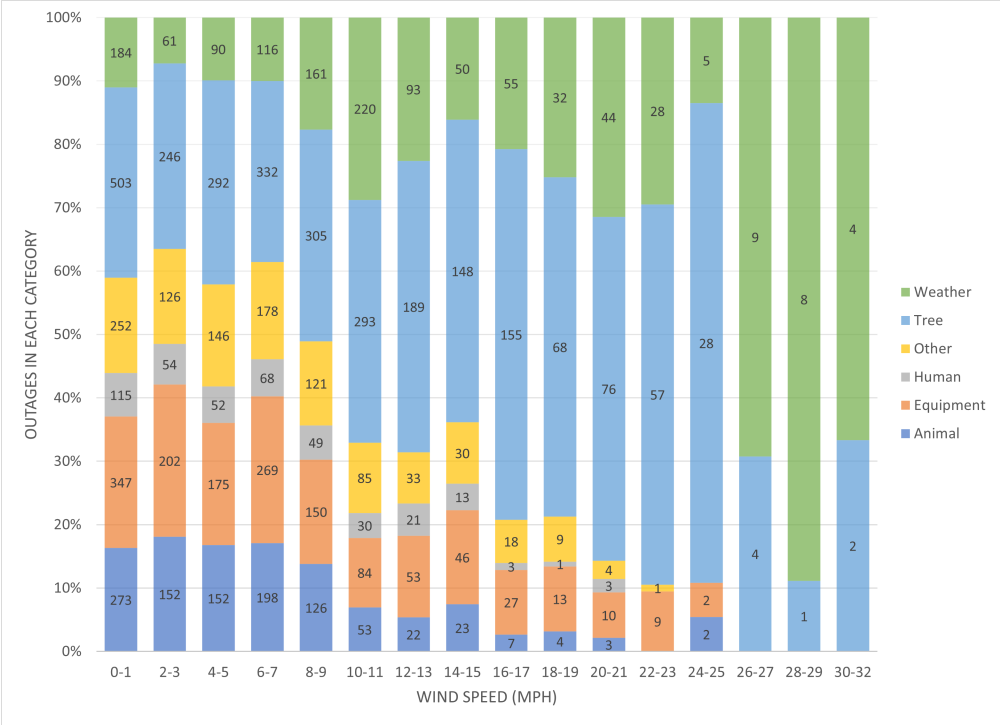

Figure A.9 Variation in the proportion of outage categories with wind speed in Area-2

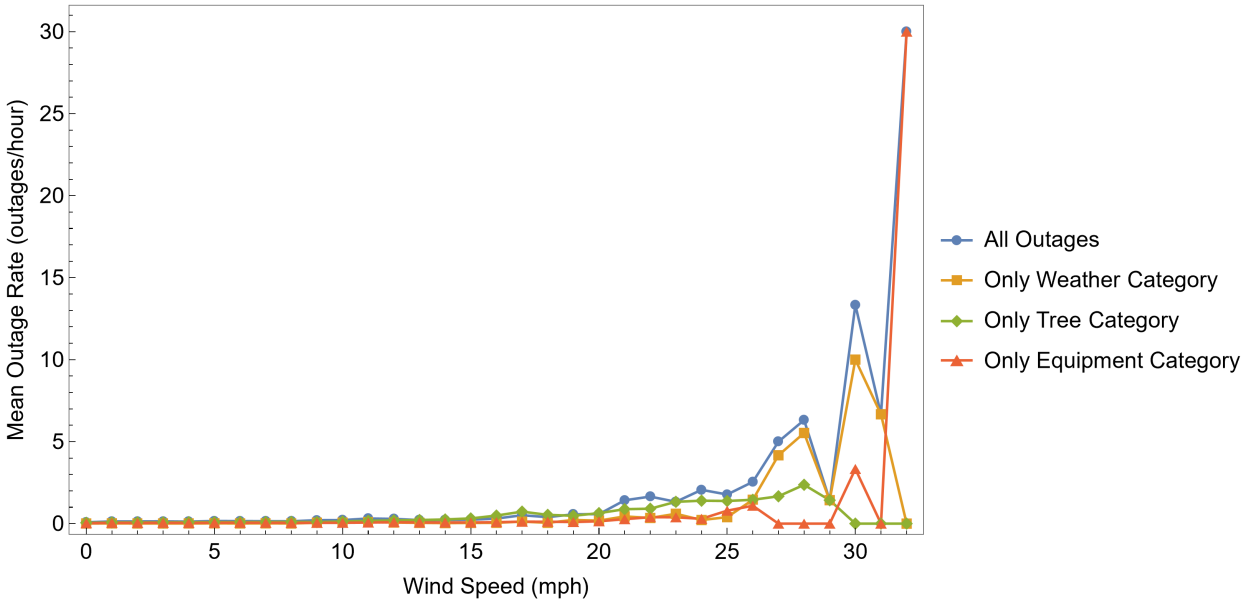

Figure A.10 Area Outage Rate Curves of Area-2 for Selective Outage Cause Categories

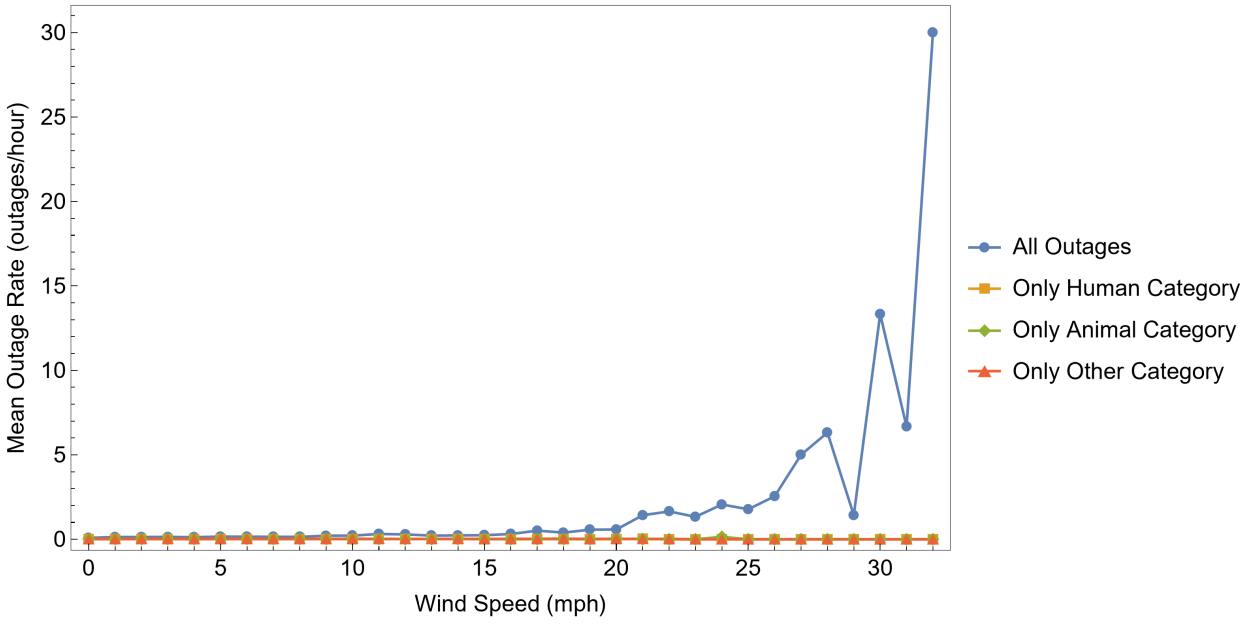

Figure A.11 Area Outage Rate Curves of Area-2 for Selective Outage Cause Categories

APPENDIX B. RESILIENCE METRICS DISTRIBUTIONS

To get robust results for each resilience metric within 99% confidence intervals, the outage sampling process is performed m times, the mean resilience metrics are calculated for each of the m samples, and then the mean of those are taken and reported in Section 5.6. The distributions of each of the mean resilience metrics of all samples are given below to illustrate the variability in each.

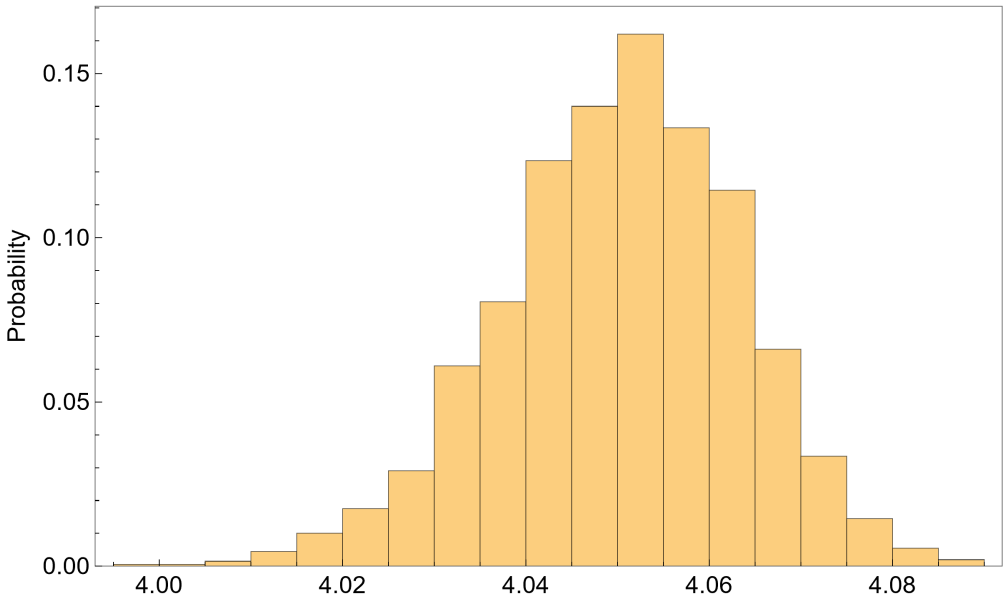

Figure B.1 Distribution of Mean Event Duration

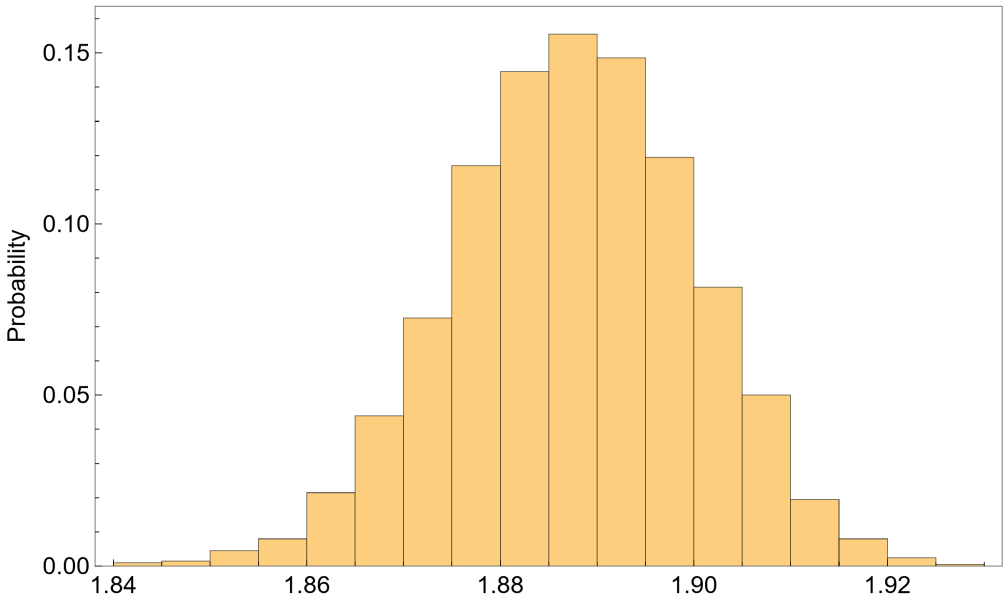

Figure B.2 Distribution of Mean Event Restore Duration

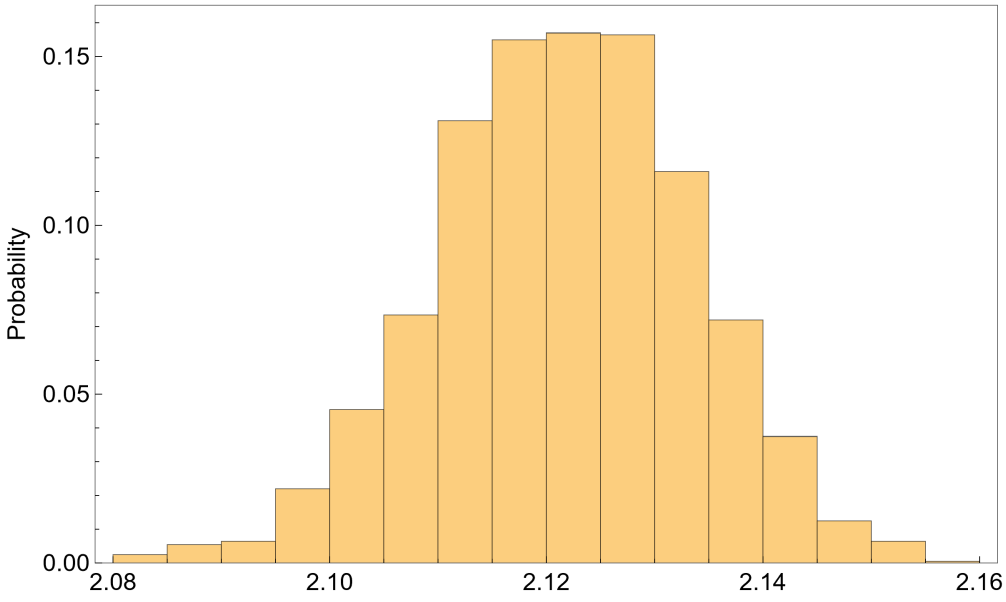

Figure B.3 Distribution of Mean Time to First Restore

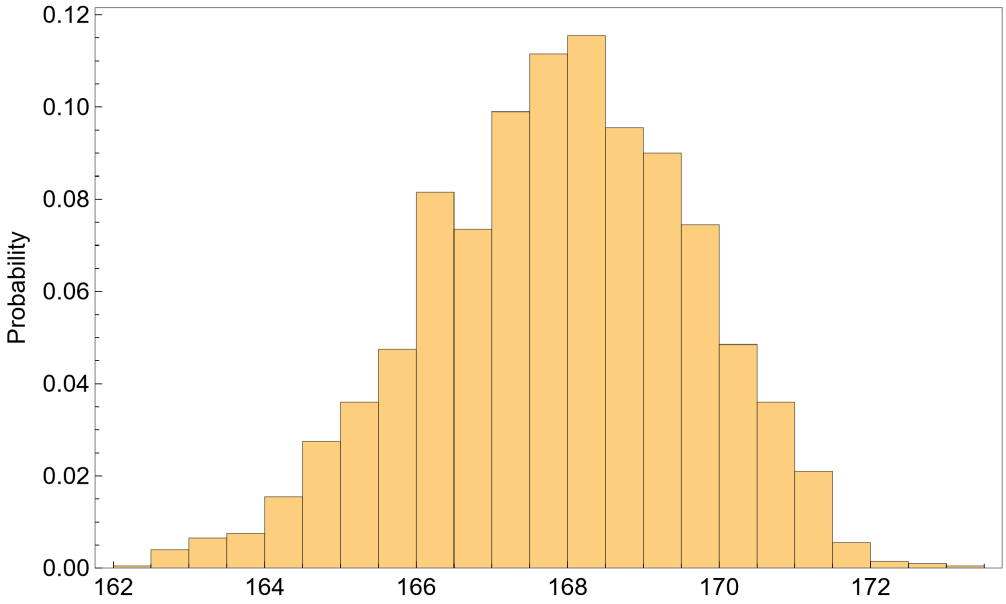

Figure B.4 Distribution of Mean No. Of Customers in Events

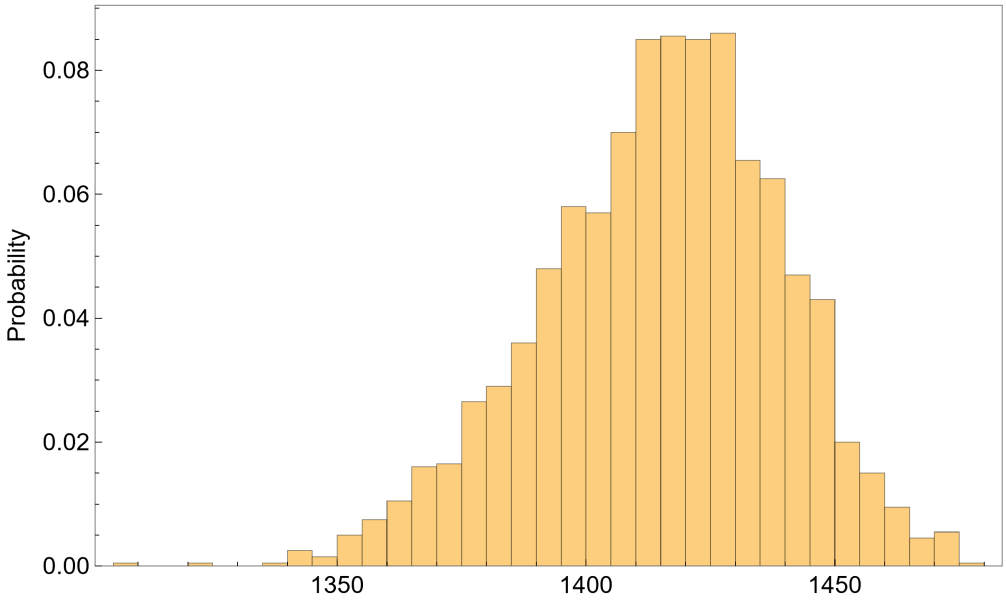

Figure B.5 Distribution of Mean Customer-Hours

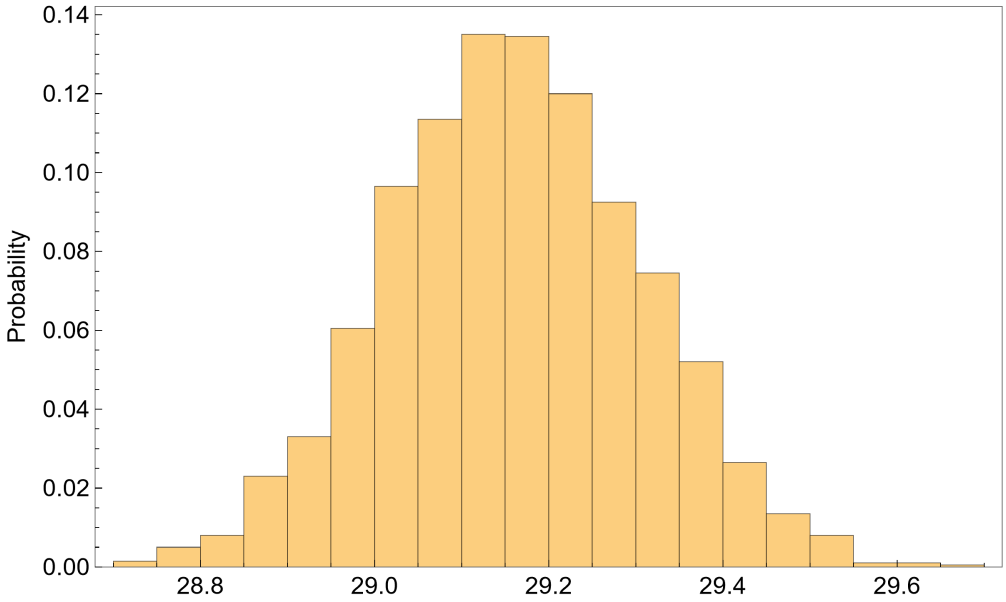

Figure B.6 Distribution of Mean Area Under the Performance Curve (Component-Hours)

Variation of Resilience Metrics with Event Size

Each resilience metric depends on event size (number of outages in an event). Moreover, the percentage change in each resilience metric also varies with the change in event size. Therefore, when 10% outages are removed from the outage data, it does not necessarily cause a 10% decrease in each resilience metric. To illustrate this effect, the random sampling is performed 2000 times to generate 2000 unique outage subsets with 10% of the total outages. Resilience metrics are then calculated, and the mean of all resilience metrics of 2000 samples are calculated for events of a given size. The original resilience metrics (with all the outage data of Area-1) and the mean resilience metrics after sampling are plotted in the following figures with event size on the horizontal axis.

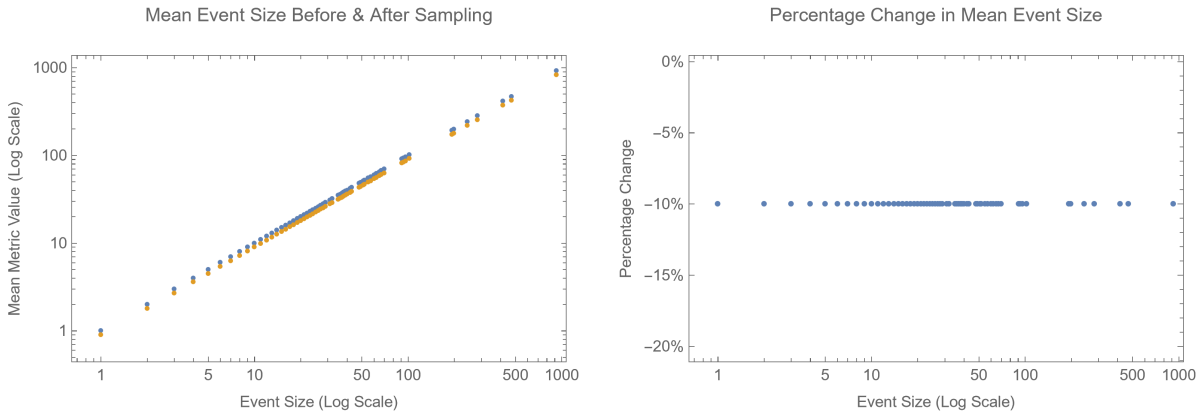

Figure B.7 Variation in Mean Event Size (before and after sampling) with the change in Event Size

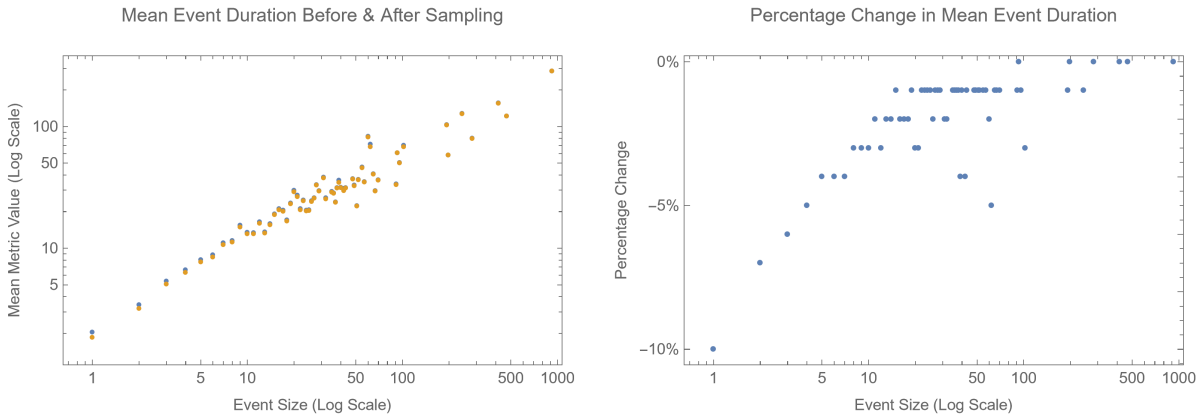

Figure B.8 Variation in Mean Event Duration (before and after sampling) with the change in Event Size

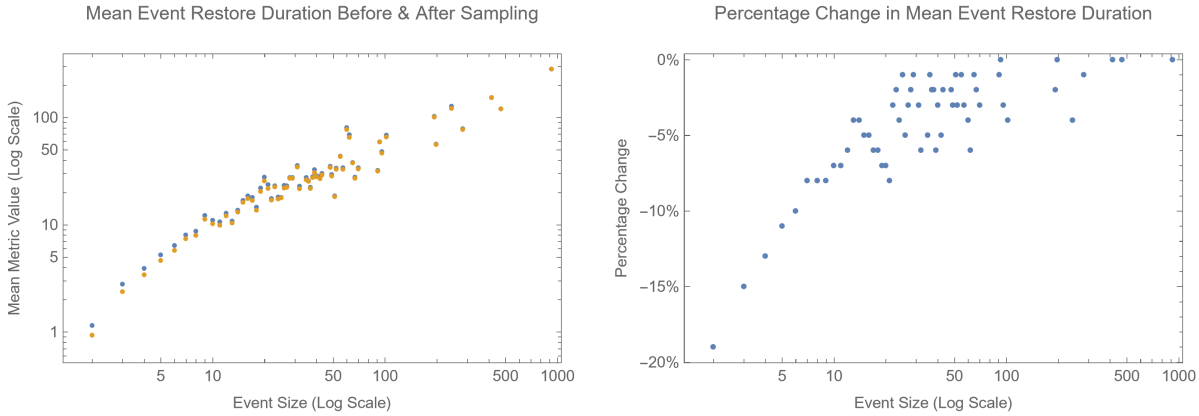

Figure B.9 Variation in Mean Event Restore Duration (before and after sampling) with the change in Event Size

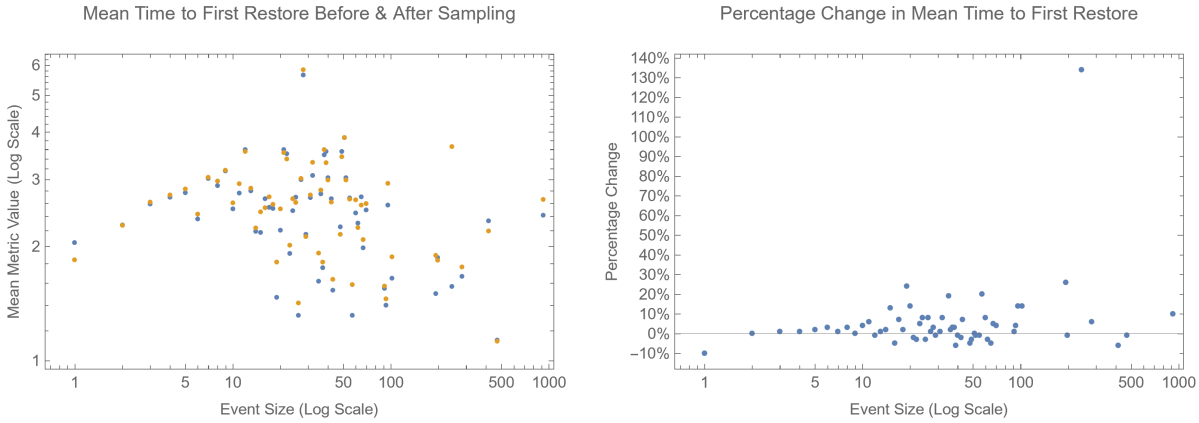

Figure B.10 Variation in Mean Time to First Restore (before and after sampling) with the change in Event Size

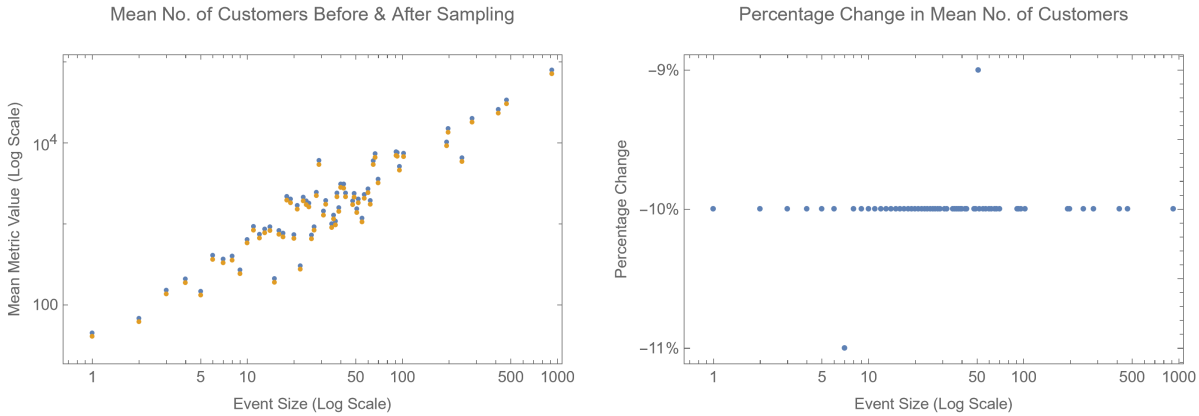

Figure B.11 Variation in Mean No. of Customers Out in an Event (before and after sampling) with the change in Event Size

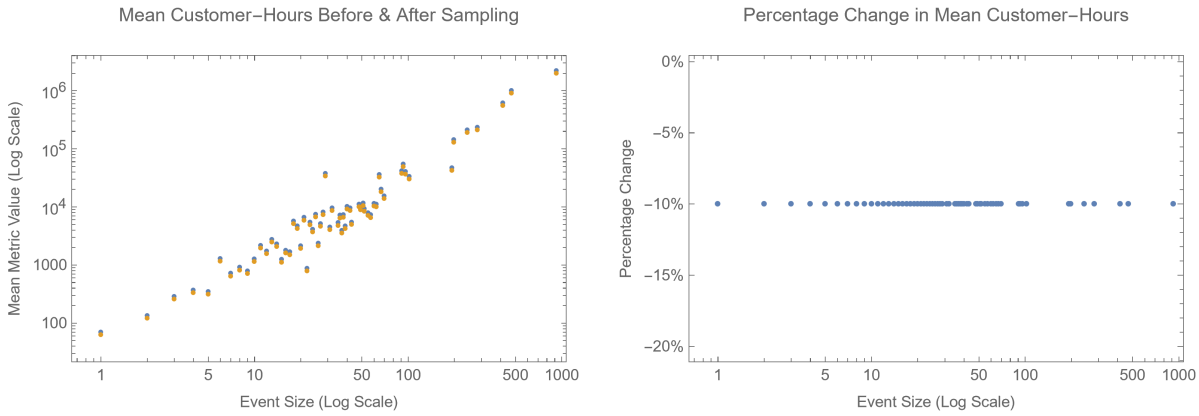

Figure B.12 Variation in Mean Customer-Hours (before and after sampling) with the change in Event Size

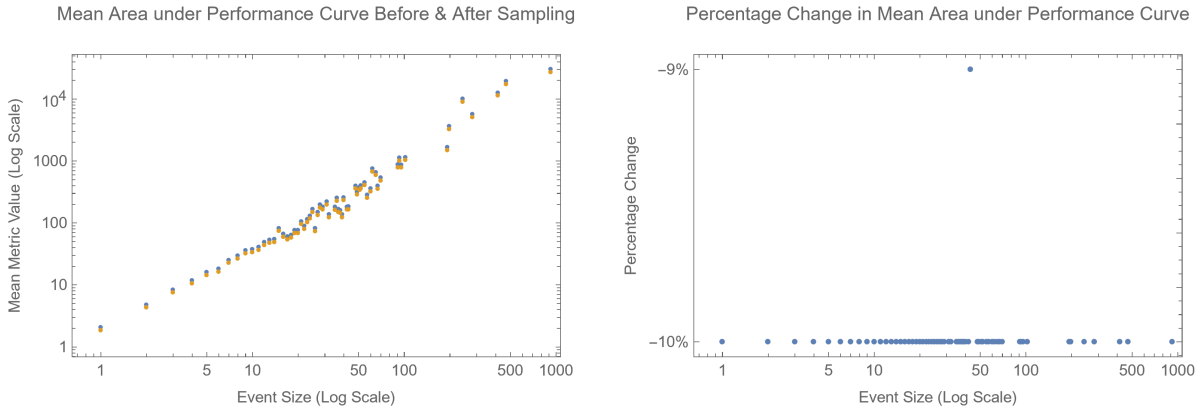

Figure B.13 Variation in Mean Area Under Performance Curve (before and after sampling) with the change in Event Size

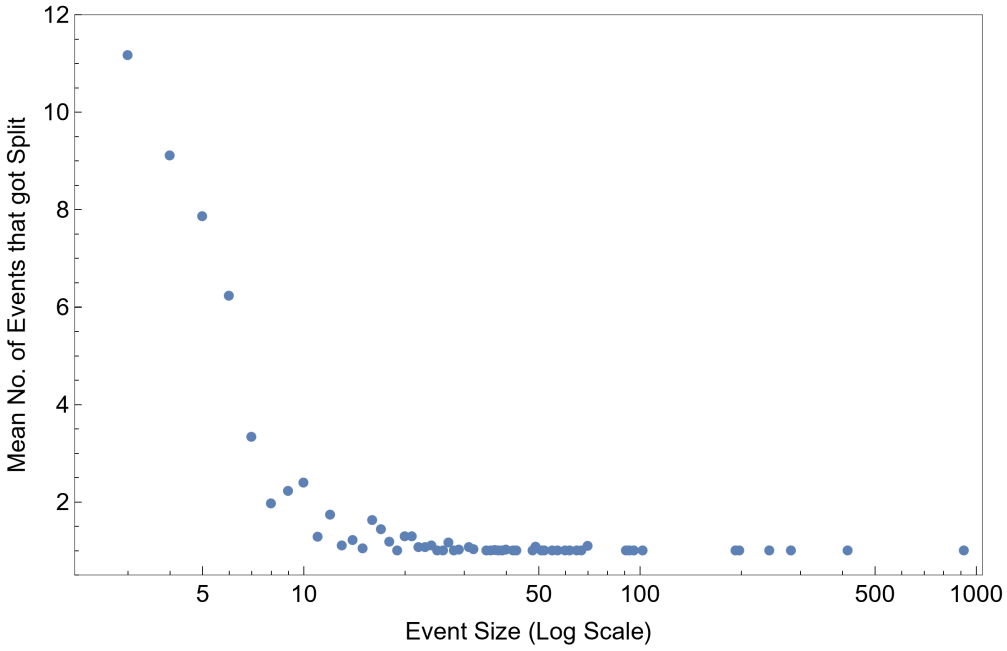

Figure B.14 Distribution of Event Size vs. Mean No. of Events that got Split after Sampling