Title
Data-Driven Decision Analysis in Electric Power Systems

Author
Dunn, Laurel N

Publication Date
2020

Peer reviewed|Thesis/dissertation

Data-Driven Decision Analysis in Electric Power Systems

by

Laurel N. Dunn

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Engineering - Civil & Environmental Engineering

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Scott J. Moura, Chair
Professor Alexandra von Meier
Professor Duncan Callaway
Professor Fotini Chow

Fall 2020

Data-Driven Decision Analysis in Electric Power Systems

Abstract

Data-Driven Decision Analysis in Electric Power Systems

by

Laurel N. Dunn

Doctor of Philosophy in Engineering - Civil & Environmental Engineering

University of California, Berkeley

Professor Scott J. Moura, Chair

This dissertation describes tools and techniques for applying data to inform decisions made in electric power systems. We explore how limitations inherent to data sets can contribute to model uncertainty, and formulate quantitative methods for issuing recommendations that are robust to uncertainty. The work provides a basis for justifying advancements in data collection and monitoring to increase confidence in decision-making outcomes.

Data about the past can offer valuable insights about the present and future. Data can improve the fidelity of simulations for characterizing performance, and can shed light on emerging risks. Data offers a basis to validate performance, to inform targeted maintenance, and to predict the outcome of candidate decisions.

Though data can be a powerful tool, our interpretations of data and even the data themselves are subject to error. Data analysts must judge whether the data are sufficient to answer the question at hand, whether the questions themselves are well-posed, or whether more suitable data should be sought after. The current work describes methods to inform these judgements by exploring how they could alter interpretations of risk and performance, contextualizing these different interpretations in terms of recommendations issued to decision-makers.

The work centers around three case studies that apply imperfect data to inform such recommendations. We consider data sets where relevant information is lost, where observational history is sparse, and where spatial coverage is limited. We synthesize data using exploratory data analysis, hierarchical modeling, and Bayesian optimization. These methods internalize what is known about the limitations of underlying data sets to examine uncertainty in decision-making outcomes. Each case study formulates a quantitative decision-support model, and offers a quantitative basis to inform the collection of new information that could help decision-makers to act with greater confidence.

There is but one remedy by which order can be restored to the faculty of thinking; this is, to forget all that we have learned.

Antoine Lavoisier

To my parents, my people, and Layla.

And to Naomi.

Contents

List of Figures

List of Tables

Acknowledgments

First and foremost, I want to thank anyone reading this dissertation. I have have done my best to summarize my current thoughts about what the future of the grid could be, and am pleased to share my ideas with you. There is more work hinted at in this dissertation than I will ever be able to do myself. I hope that reading my work will lead you to discover new and interesting ideas you can explore on your own. If that is the case, I encourage you to get in touch. I would enjoy hearing about the work that you do.

Next, I wish to acknowledge those who have sharpened my thinking and improved my ideas during this time. In particular, I am grateful to Scott Moura and to Sascha von Meier for their guidance, patience, and support. I thank Duncan Callaway and Tina Chow for their valuable feedback on my research and on this dissertation. I also wish to thank Michael Sohn and Ron Hofmann whose contributions will be clear to them upon reading this work.

My colleagues at Berkeley contributed to both the quality of my work and my enjoyment of it. Special thanks goes to Mathilde Badoual and Anna Brockway, who have been outstanding collaborators, role models, and friends. I also wish to acknowledge Ioanna Kavaada, Jaimie Swartz, Isa Ferrall, Michael Berger, Julia Szinai, Colin Shepard, Sangjae Bae, Max Gardner, Joe Eto, Kristina LaCommare, Chris Vermillion, and Patrick Keyantuo. I have learned a great deal through my collaborations with each of you.

Several funding agencies supported this work. The study on outage data was supported by the U.S. Department of Energy, Office of Electricity Delivery and Energy Reliability in accordance with the terms of Lawrence Berkeley National Laboratory Contract no. DE-AC02-05CH11231. The study on component failure was supported by a 2019 Seed Fund Award from CITRIS and the Bantao Institute at the University of California. The study on optimal control in wind energy systems was supported by the National Science Foundation under Award 1437296.

It would not have been possible to complete this degree without the relentless support of my family and my friends.

Chapter 1

Introduction

1.1 Motivation

The electrical grid is one of the pillars of modern society. Construction on the grid began over a century ago, and electric service quickly became a basic necessity. By the early 1900's, state and federal regulations were enacted to keep electricity rates affordable. In 1936, the Rural Electrification Act subsidized construction of grid infrastructure to remote areas that would otherwise have been cost-prohibitive to serve. In the interim decades, grid infrastructure has supported innovations in food systems, heath care, and manufacturing that are central to how we eat, live, and work.

Due to the laws of physics, any device that uses or generates electricity on the grid also influences the flow of electricity. Thus over time, changes in the geography and electrical characteristics of these devices change the electrical dynamics of the system. These changes will eventually (and may already) have implications on infrastructure performance, and there is evidence to suggest that current best-practices will not be able to meet future needs. Design principles and operating policies will need to be revised to address risks associated with ageing infrastructure and climate change, and to ensure that the grid can support new generation and end-use technologies without a decline in performance.

Growth in solar generation, for example, will need to be coupled with technologies that can respond to near-term changes in power output. Load served by solar during the day is re-distributed to other resources at sun set. This re-distribution must happen within the span of a couple hours, which is much faster than most conventional generators can turn on. Those generators that are able to respond quickly enough were not intended for every-day use; they tend to be inefficient, expensive and carbon-intensive to run. Innovations in storage, generation, and grid operations will be instrumental in supporting sustained renewables growth.

Advancements in grid measurement and monitoring are also warranted. Solar, storage, and electronic devices all use direct current, and connect to the grid (which uses alternating current) through power electronic converters. The conversion is not perfect, and converters must be programmed to respond appropriately under different types of fault conditions. Control errors or unforeseen conditions may cause converters to behave in undesirable ways, which can in turn contribute to inefficiencies or equipment damage. But converters actuate on time-scales that are much faster than most grid sensors, which means existing sensors

may detect problematic dynamics unless they escalate to cause more severe issues. Sensor networks will need to become faster, more prevalent, and more versatile to support early detection of emerging issues of this nature.

The environment in which grid infrastructure operates is also changing. Drought, extreme heat, severe storms, flooding, and sea level rise will expose grid components to stress conditions which they may not be designed to handle [14]. Planning for future climate conditions could change decisions about component sizing, placement, and design. Given that many of the components installed today will have operating lifespans on the order of decades, long-term planning decisions will almost certainly be impacted by climate trends. New decision-making frameworks will be needed to issue recommendations that are robust to uncertainty in future climate conditions.

Meanwhile, investments in grid infrastructure have not kept pace with technology and environmental trends. The American Society of Civil Engineers projects a $177 billion shortfall in grid infrastructure investment by 2025 [4]. They give the grid a D+ rating, citing that much of the infrastructure in place today dates back to the 1950's. This investment gap may be due to the fact that planning decisions are largely informed by the needs, technologies, and operating conditions of the past. Even if we consider the grid to be reliable enough to suit current needs, interventions will be needed to maintain comparable performance moving forward.

While some technology trends threaten the reliability of the grid, others pose opportunities. In the last decade, advancements in automation, sensing, communications, and computing have transformed other industries. These technologies lend themselves to adoption in power systems, and are well-suited to address many of the emerging issues in grid planning and operations. To justify investments in innovative solutions will involve navigating a complex regulatory environment that has not historically favored innovation. Quantitative studies showing the benefits to different stakeholder groups could help to strengthen the case for allocating more resources to promising new solutions.

Decision-makers have a rich historical record of data to inform these investments, and to characterize emergent properties of the system that are too complex to model from first principles. However, much of the data on hand was collected with specific use-cases in mind, and will need to be synthesized in new ways to answer new questions. This poses an analytical challenge, as there are no clear guidelines on how best to process the data, formulate models, or benchmark results. There is a need to carefully study how different data sources or analysis methods could lead to different decisions and outcomes.

This dissertation provides a library of tools for exploring the merits and limitations of making data-driven decisions. Each tool uses data to infer statistical properties of the system, and examines different courses of action intended to alter those properties. Each chapter describes a case study showing how advancements in (1) data handling, (2) model formulation, and (3) monitoring could improve our ability to characterize relevant statistics. Each method internalizes known limitations of the data or model formulation to issue recommendations that are robust to uncertainty. We explore how new data sources could reduce uncertainty, thus allowing decision-makers to act with greater confidence. Each case study offers a quantitative basis to inform the discovery and deployment of innovations in data collection and synthesis.

1.2 Contributions and Outline

Chapter 2 reports insights gained through exploratory analysis of power outage data. Data were collected using an automated web scraper to archive real-time outage information posted on the customer-facing website of a large investor owned utility. We report on outage characteristics such as duration, frequency, and size, and show spatial and temporal variability in these properties. Results are compared with insights arrived at from publicly reported reliability metrics, including annual and incident reporting data.

Novel contributions of this chapter are:

- To report the range and variability in outage properties, as evidenced in data with much finer spatial and temporal resolution than other publicly reported information.
- To show that outage characteristics exhibit diurnal and seasonal variability, and that certain customers experience more frequent and longer duration outages than others.
- To compare summary statistics of outage properties against publicly reported reliability metrics, and propose updates in reporting practices that could improve regulatory oversight.
- To summarize inferences drawn from public data for research and policy applications, and discuss how results could change if studies were replicated using more granular outage data.

Chapter 3 describes a new probabilistic model formulation synthesizing data and engineering priors to characterize variability in failure properties of grid components. We use a hierarchical formulation to describe a random failure process, where uncertainty in engineering parameters is captured by representing parameters as if they were random variables. This formulation has a strong physical basis, as it encapsulates random variability in performance arising from manufacturing defects and ageing effects that would be impractical (if not impossible) to monitor directly. The model is applied to estimate the number of repairs needed to meet a targeted risk reduction goal. The application is motivated by the need for quantitative methods for optimizing investments in wildfire risk mitigation.

The novel contributions of this chapter are:

- To formulate a hierarchical model for parameterizing a failure process, while internalizing seemingly random variability among individual components.
- To apply Markov Chain Monte Carlo (MCMC) sampling methods to characterize uncertainty in engineering parameters, given observational data and engineering priors.
- To demonstrate an application using uncertain parameter estimates to inform risk-aware decision-making, where the objective is to minimize the number of repairs needed to achieve a pre-defined system-wide risk tolerance.
- To illustrate differences in decision-making outcomes arrived at using frequentist methods, given observational data that are sparse and potentially biased.

Chapter 4 deals with optimal control of an airborne wind energy system to maximize power production in an operating environment where wind speeds are uncertain. Control methods use Bayesian optimization to balance the trade-off between decisions to maximize power production based on current information (exploitation), against decisions to collect new information that could reduce forecast uncertainty over time (exploration). Performance gains achieved from control are compared against opportunities for innovation in sensing and measurement that could reduce uncertainty in the wind speed forecasts that inform control decisions.

The novel contributions of this chapter are:

- To apply advanced control strategies proposed in the literature to inform optimal altitude control for an airborne wind energy system, given different measurement capabilities.
- To compare performance gains achievable through controls, against gains achievable from innovations in data collection.
- To provide performance benchmarks that could inform techno-economic studies evaluating the possible impact of improving system performance through advancements in controls, against advancements in data collection.
- To show that collecting data that could reduce uncertainty in the environment leads to higher performance gains than are accessible through advancement in controls; thus motivating further study of methods for forecasting vertical wind speed profiles.

1.3 Background and Context

Most decisions in grid operation and design aim to keep the system safe and reliable. These decisions are challenging due to the size and complexity of the grid, and due to the rate at which operating conditions can change. Cost-cutting measures made in long-term planning may introduce challenges in real-time operation. Decisions often come with trade-offs that compromise safety for reliability, security for cost, or vise versa.

The following sections describe the breadth of decisions made in designing and operating the grid. We begin with a description of the tools that inform these decisions, and go on to describe how these tools are applied to make decisions for long-term planning and real-time operations. This discussion is included to familiarize readers with the library of decisions relevant to this field of study. The discussion aims to motivate deeper investigation of how data are used to balance near-term and long-range outcomes across different decision-making processes. Subsequent chapters offer deeper case studies demonstrating how such investigations may be conducted.

Decision Analysis Tools

Decision-makers rely on different types of analytical tools that may offer either prescriptive guidelines or probabilistic recommendations. In some cases, decision-making processes in-

volve modeling pipelines where different types of tools build upon one another. For example, a modeling pipeline could study how engineering or policy decisions – such as expanding fast-ramping generation capacity – would impact power flow. The following paragraphs provide more detail about different types of tools.

Engineering and Design Standards

Standards provide prescriptive guidelines to inform a wide range of decision-making processes. They are written to synthesize conventions, and to bring attention to details that could easily be missed. Standards may specify the configuration (e.g., single- or three-phase), material (e.g., copper or ASCR) and specifications (e.g., thermal limits or wind-speed ratings) of components. Standards may also describe protocols for equipment testing, processing data, communications, measurement, and data formats [32].

Industry-wide standards are set by national, international, or trade organizations (NIST, IEC, IEEE, ASME, etc.). Companies may choose to set more stringent standards, or may choose to opt out of compliance altogether. Industry-wide standards are referred to as "voluntary consensus standards", and are arrived at by reaching consensus among experts [62]. Standards typically lag behind the state of the art, though they are periodically updated to reflect changes as technologies and best-practices evolve.

Power Flow Models

Power flow models build on physical principles of electric circuits to model the flow of electricity through the system. These models use design information such as network configuration and connectivity between generation sources and loads. They are founded upon simple concepts from basic physics, where numerical approximations make it tractable to study large-scale systems. Different approximations are used depending on which characteristics of the system are most relevant to capture. In the context of real-time decision making, computational time is also a limiting factor.

Power flow models form a foundation for various types of decision analysis models. Power flow studies may examine normal steady-state operating conditions, or may study dynamics during faults or contingency events. These models can validate whether operating constraints on the system are being met, and can infer unknown quantities using limited measurement data. Power flow models form the foundation of decisions related to electricity supply and procurement, including long-term and real-time planning.

Scenarios and Simulations

Simulations examine how the grid will perform under specified conditions, or scenarios. Scenarios may examine severe weather events [51] or operating stress [73]. Scenarios may also internalize projections of future loading [64] or climate conditions [70].

Many scenario analysis tools build upon physics-based power flow models [51, 64]. These models, however, are only as accurate as the data or the engineering assumptions underlying them. Studies show that data may be missing [67] or subject to interpretation [66]. In practice, decision-makers tend to use simplified representations of the system that use information readily available to them [15, 7].

Uncertainty and Risk Assessment

Risk assessment tools rely on probabilistic simulations to examine the consequences of different events across a range of scenarios. Risk models often use data-driven [51], engineering-based [58], or hybrid models [28] to generate probabilistic realizations of how the system could perform under different ambient or operating conditions. Risk exposure is typically quantified in terms of downstream consequences – such as dollars, power outages, or lives.

Long-term planning decisions inherently involve making predictions about how the probability of certain risk scenarios is likely to change. These tools traditionally rely on statistical properties computed from historic data [14]. However, infrastructure investments often involve planning horizons on the order of decades. Over those timescales, climate trends are likely to increase risk exposure. Electrification may lead to more severe consequences when risk scenarios do come to fruition.

Risk assessment models play an important role in determining when mitigation costs begin to outweigh the costs of shouldering consequences. These models inevitably involve taking measured risks. Decisions may change depending on local risk exposure and tolerance, resource constraints (e.g., time, personnel, funding), and on the nature of the risks.

Risk mitigation typically involves a portfolio of long-term and real-time solutions. Regulatory filings submitted under the California Public Utility Commission's "wildfire mitigation proceeding" are one example. These documents describe investments in undergrounding, tree trimming, weather monitoring, and other mitigation measures geared at reducing the risk of grid-caused wildfires [15]. Long-term investments (such as undergrounding) reduce the probability of adverse outcomes happening under normal operating conditions, and may make the system more resilient. Real-time monitoring (such as weather stations) allow decision-makers to respond more quickly and effectively when risk scenarios come to fruition.

Long-Term Planning

Long-term planning informs design decisions such as siting, connectivity, and component sizing. Design decisions tend to be highly customized to suit local geography and customer needs.

Network Topology

Grid topology refers to the network and graph structure of power lines connecting generators to loads. In legacy power systems, grid topology was informed by siting constraints on large-scale power generation, which tend to be located outside of load centers. Topological changes may need to be made to accommodate both distributed generation and large-scale renewables.

Studies in the literature cite trade-offs to both more connected and more distributed topologies. Martinez-Anido et al [48] find that more interconnected areas experience more frequent faults, but better overall reliability (i.e., fewer outages). Denholm and Hand [24] find that greater interconnectedness can mitigate frequency deviations in areas with high renewables penetration. Meanwhile islanding capabilities may mitigate the spread of cas-

cading failures [20]. Local microgrids may also improve resilience, protect critical loads, and reduce wildfire risk by allowing grid operators to remove at-risk lines from service.

Capacity Expansion

Capacity expansion decisions rely on power flow models to identify areas of the grid where existing capacity is inadequate to meet future needs. Results inform the siting and sizing of grid components (e.g., generators, transformers, lines, etc.), as well as the "mix" (i.e., the size and number) of power plants on the system. Capacity expansion models build on production cost models which determine the optimal scheduling (or "dispatch") of generators. The optimization objective is to serve power demand while minimizing costs, and without violating operating constraints. These constraints ensure that voltage is within bounds specified in customer service contracts, and that components are not exposed to thermal or mechanical stress beyond what they are designed to withstand. Capacity expansion decisions may change depending on assumptions about load growth, climate trends, and seasonal or diurnal energy consumption [14]. Sustainability targets also play a major role [7].

Grid Hardening and Deferred Maintenance

Grid hardening measures make the grid more robust to failure, while deferred maintenance refers to the practice of putting off work until it becomes necessary to do. Though outages are expensive, perfect reliability would not be cost effective to attain, and some amount of deferred maintenance is inevitable.

By some decision metrics, cost-benefit analysis may prioritize infrastructure investments in urban areas at the expense of remote or rural lines [43]. In areas where the implications of power outages are not as high, this may even mean deferring upgrades until a component fails. This practice contributes to ageing infrastructure [4]. Ageing assets coupled with severe wildfire conditions have contributed to catastrophic utility-caused wildfires [36, 60].

Real-time planning and operations

Long-term planning decisions ensure that it is feasible for the system to operate under most conditions. Real-time decisions involve reacting to circumstances as they evolve, and containing risks that do materialize.

Balancing supply and demand

The primary objective in grid operation is to ensure that supply equals demand at all times. Any imbalance between the two causes the frequency of the grid to change, which can trigger generators to trip offline to avoid damage.

Generators are scheduled to provide a fixed amount of power, or to provide "ancillary services" which involve turning on or off in response to changes in load. Depending on the system need and the capabilities of the generator, response times may be on the order of seconds, minutes, or hours.

Day-ahead scheduling is determined based on a forecast of what power demand and renewable generation will be. Scheduling is coordinated by an independent system operator,

and is optimized to serve demand at least cost without violating system operating constraints. The forecast is then updated on an hourly and sub-hourly basis as confidence in the forecast improves; generators scheduled to provide ancillary services respond to these updates. The remaining uncertainty is managed by making minor adjustments to the rotational speed of spinning generators in response to near-term variability.

Contingency Planning

Grid operators also ensure that the grid can continue to serve system demand in the rare event that a critical piece of infrastructure should fail. These failures are known as "contingencies".

When a contingency occurs, the grid enters an emergency state. Generators on "reserve" are dispatched to increase (or decrease) power supply, and generators on "stand by" may be directed to begin ramping up. Certain operating constraints – such as line transfer limits – may be temporarily relaxed to allow grid operators enough time to restore the system to a stable operating state.

Grid operators plan for a library of deterministic contingency scenarios examining the loss of some critical asset. Each scenario uses power flow models to ensure that operating constraints are not violated. The grid is said to be "N-1 secure" when it is robust to any single-component failure. Some grid operators use more stringent policies that consider the loss of any two components (N-2 security) or the possibility of successive failures (N-1-1 security).

Contingency planning adds to the cost of operating the grid by imposing added constraints. These constraints may prevent grid operators from capitalizing on cheap generation resources [5]. Contingency planning also affects long-term capacity expansion decisions by changing the optimal siting and ramping capabilities of generation fleets.

Protection

Protection systems offer fast-acting controls that are triggered when a fault is detected. A fault is a condition where electricity flows between two objects that are not intended to be electrically connected. Faults can occur between two power lines, internally within a component, or with objects that are foreign to the grid.

Protection systems use a suite of devices – fuses, circuit breakers, reclosers, and relays – to detect, locate, and isolate faults. Protective components are triggered in response to high currents, and are calibrated to prevent damage to grid assets. Rapid response also mitigates public safety hazards associated with sparks, arcing, or damaged equipment. Systems are designed for protective elements to trigger as close to the fault as possible to prevent power outages from affecting otherwise healthy areas of the grid.

Though overcurrent fault detection works well in most cases, certain fault conditions are uniquely challenging to detect. High impedance faults, for example, are characterized by high resistance that prevents overcurrent events that would trigger conventional protective gear. Bi-directional power flow due to distributed generation may also make traditional fault detection techniques less reliable.

Preventative Maintenance

The primary objective of preventative maintenance is to repair or replace components before they fail. The alternative is to defer maintenance, which can have safety and reliability consequences. Preventative maintenance can avoid outages, contingencies, and public safety hazards.

Maintenance decisions are often informed by routine inspections. Inspection frequency is prescribed by company or industry standards, and may change depending on findings from previous inspections. Inspections monitor state of health indicators, such as the circumference of poles or conductor hot spots. When issues are detected, at-risk components are replaced or repaired. Inspections may also collect information about trees or other external conditions that threaten grid assets.

Given the vast size of the system, inspection programs can be costly to run. The level of detail for a particular inspection may change depending on what condition a component is in. Even detailed inspections may give limited observability into factors that lead assets to fail. The circumference of a pole, for example, does not tell us how it will perform under high wind speed conditions. However, inspection records have been shown to hold valuable insights in semantic text fields [66]. These data are notoriously difficult to analyze systematically, as domain expertise is often needed to interpret what is written.

Advanced sensor technologies may monitor more direct indicators of failure in real-time, providing more timely information to inform repairs. These sensors may measure tilt on poles, internal temperature of transformers, or electrical waveform signatures that indicate when a component is nearing end-of-life.

Public Safety Power Shutoffs

Where preventative maintenance programs have fallen behind, component failures may pose public safety risks. These risks include the possibility of wildfire ignition.

To eliminate this possibility, grid operators may temporarily de-energize power lines when wildfire conditions are severe. These events (known as "public safety power shutoffs", or PSPS) trade off reliability for wildfire safety. Because long-distance transmission lines often pass through remote and wildfire-prone areas, PSPS events may lead to widespread and long-duration outages that affect customers in urban areas where wildfire conditions are not as severe. Utility wildfire mitigation plans include long-term investments geared at mitigating wildfire risk, as well as investments in microgrids and local generation to mitigate the consequences of PSPS events [15].

Chapter 2

Exploratory Analysis of High-Resolution Outage Data

This chapter documents an exploratory analysis examining the information contents of various data sets collected to track grid reliability and performance. We discuss where the data originate from, and describe filtering and pre-processing that happens to make reporting metrics more digestible to decision-makers. The core contribution of the work is report on the statistical properties of minimally processed data, and to compare the information contents against publicly reported data.

The work synthesizes academic research mining public data sets for statistical, economic, and policy insights. We use exploratory analysis to study how replicating these studies using minimally processed data could lead to different interpretations of system performance. The study finds that aggregation and filtering masks spatial and temporal variability that is important to understand. We propose new reporting metrics that could provide greater insight into these properties, to inform actionable insights for both utilities and regulators. The results are applied to critically evaluate regulatory incentives that may inadvertently contribute to adverse outcomes, such as equity issues in the frequency and duration of outages that different customers experience.

The paper appears in the journal *Energy Policy* and may be cited as follows:

> Laurel N. Dunn et al. "Exploratory analysis of high-resolution power interruption data reveals spatial and temporal heterogeneity in electric grid reliability". In: *Energy Policy* 129 (2019), pp. 206–214

The original funding acknowledgement is included here as a footnote.[1]

2.1 Introduction

With modern monitoring and control equipment, electric power systems deliver high levels of service reliability to grid customers. Monitoring systems provide detailed data about

[1]The work described in this paper was funded by the U.S. Department of Energy, Office of Electricity Delivery and Energy Reliability in accordance with the terms of Lawrence Berkeley National Laboratory Contract no. DE-AC02-05CH11231.

when power interruptions occur and which customers affected. Power interruption data can support empirical analysis exploring a wide range of questions relevant to how we assess, regulate, and invest in grid performance.

Granular outage data are confidential, but are aggregated to compute publicly reported reliability metrics. Information regarding the characteristics of individual power interruptions is lost through aggregation. The implications of losing this information are not well understood. The current work explores whether (or not) the loss of granular information is an impediment to using publicly reported data for policy analysis and research.

Several works in the literature use high-resolution power interruption data to explore different types of questions. Studies use predictive models to compare the effects of different tree trimming policies [34], or to inform preventative maintenance strategies [66]. Vulnerabilities to hurricanes [55, 38] and other weather conditions [39] are also explored. However, these studies rely on utility data that are not readily shared. As such the geographic scope is generally limited to a single service territory. Operational, topological, and climatic differences between service territories raise questions about how generalizable results are to other geographic regions.

Research efforts that are regional or national in scope require datasets that are more geographically comprehensive. In light of the barriers to securing proprietary data from multiple utilities, publicly reported information are better suited to support these efforts. Several studies in the literature aim to capitalize on these datasets, for example to report on trends in grid performance [35, 44, 8], to quantify the implications of power interruptions [41] and to explore the reliability benefits of different undergrounding policies [43]. Further efforts have also been made to develop explanatory models for correlating reliability metrics with local weather patterns [44, 17].

Several of these studies leave questions unanswered, citing low-resolution in the data they use as both a limitation of the analysis they performed and a barrier to future work [35, 44, 45]. This raises questions about whether there are policy implications associated with leaving these research questions unanswered, whether more granular reliability data could advance policy research, and if so whether collection of more granular data is warranted.

Multiple studies examining national grid performance report that the frequency of power interruptions is increasing [35, 44, 8]. This apparent decline in grid performance coincides with an increase in grid infrastructure spending per unit of electricity sold. Between 1990 and 2015, the ratio between infrastructure expenditures (in transmission and distribution systems combined) and electricity sales doubled [31, 30].

Reports that reliability is decreasing suggest that there is some uncertainty about whether or not investments in grid performance have improved the reliability of service to grid customers. Further research is needed to understand what is causing the observed trends, to evaluate whether (or not) there is cause for concern, and to determine what policy interventions (if any) are needed to mitigate a decline in grid performance.

The reported change in grid reliability may not be a cause for concern if grid performance is improving in ways that these aggregate metrics do not capture. For example, investments in grid performance can target measures that benefit those subsets of customers that are most vulnerable to power interruptions. Though such measures may improve grid performance by some measures, these gains are not necessarily evidenced by metrics that describe the characteristics of power interruptions and not their implications.

An active body of research is exploring quantitative methods for evaluating these reliability gains. For example, damage functions estimate costs incurred due to power interruptions by individual grid customers [71] and nationwide [41]. Downtime of critical loads such as hospitals and water treatment facilities have also been considered [51]. Applying these methods to assess the implications of past power interruptions is a critical step towards evaluating whether (or not) the apparent decline in grid performance is cause for concern.

Studies consistently report that the implications of power interruptions are related to when and where the power interruption occurs. For example, Carlsson and Martinsson [16] and Chang et al [18] examine the economic damages due to specific power interruptions in Canada and Sweden. Both studies report that weather exacerbated damages, as the interruptions occurred during severe winter storms and affected customers who relied on electricity to heat their homes. Sullivan et al [71] also show that the damages a given customer incurs vary depending on time of day and time of year. These studies show that accurately characterizing damages requires insight into temporal heterogeneity in grid performance.

Furthermore, the characteristics of power interruptions may change depending on the loads and services that are affected. In an analysis of proprietary power interruption data, Maliszewski and Perrings [47] find that residential customers located near hospitals experience characteristically shorter power interruptions than similar customers located elsewhere. The implication is that aggregate metrics likely do not represent the typical characteristics of power interruptions affecting critical loads. This finding shows that to evaluate grid performance requires insight into geographic heterogeneity in power interruption characteristics.

In the current work, we explore questions related to if and how granular power interruption data could advance data-driven research informing how we understand grid performance. We do so by presenting an exploratory analysis of spatially and temporally resolved power interruption data. We collected these data by archiving real-time information posted on public-facing utility websites about ongoing power interruptions. We mine the data to explore heterogeneity in the characteristics of power interruptions, and report on the patterns we observe. We compare our findings with the information contents of publicly available reliability data, and with findings from the literature reporting on these public data. We discuss whether (or not) past studies have arrived at similar insights from analyzing public data, and whether replicating past studies with more granular data could change results reported in the literature to date. Results showing the two datasets yield similar insights indicate that there is little marginal benefit to collecting additional data. However, in many cases our data reveal new insights that are not evidenced in public datasets. This finding suggests that access to more granular data may be able to resolve questions that remain unanswered in the literature to date (such as if and why grid performance is decreasing), and support new research related to grid performance.

This paper is structured as follows: Section 2.2 describes publicly available information about grid reliability and surveys the literature that draws on these datasets. Section 2.3 describes the dataset we collected to support the current work and methods for calculating conventional reliability metrics from these data. Section 2.4 summarizes our results and discusses the implications of these results in the context of the literature to date. Finally, Section 2.5 summarizes our findings and discusses opportunities to capitalize on granular datasets to inform policy interventions for improving grid performance.

2.2 Publicly available power interruption data

Three main sources of power interruption data are readily available to the public. These include: (1) aggregate metrics of grid reliability, (2) incident reports describing specific reliability events that are of particular interest to regulators, and (3) real-time information about ongoing power interruptions posted on public-facing utility websites.

The first two data sources are published by regulatory agencies who periodically collect information from electric utilities to track grid performance. At the national level, the Energy Information Administration (EIA) collects annual reliability metrics [72], and the Department of Energy's (DOE) Office of Electricity (OE) collects information on specific reliability incidents [57]. Local public utilities commissions often collect data similar to those reported to EIA and DOE, though local reporting requirements may vary. Submission of emergency incident reports to DOE is mandatory, while reporting annual reliability metrics to EIA is voluntary. Reporting requirements may also include sensitive information that are ultimately withheld from publicly available versions of the data.

The EIA collects data on reliability, along with a wide range of information about utility sales and operations via EIA form 861 [72].[2] Utilities submit this form annually. Three reliability metrics are reported: System Average Interruption Frequency Index (SAIFI), System Average Interruption Duration Index (SAIDI) and Customer Average Interruption Duration Index (CAIDI). SAIFI describes the average number of power interruptions customers experience per calendar year. SAIDI describes the average amount of time (in minutes) each customer spends without electricity service per calendar year. CAIDI describes the average duration (in minutes) of individual service interruptions. Standard definitions of these metrics and methods for computing them are detailed in IEEE Standard 1366 [37].

Though EIA only recently began collecting these data at a national scale, many local public utilities commissions have collected similar data for a longer time. Several works in the literature study data collected by local reporting agencies. For example, Larsen et al [44] report on longitudinal SAIDI and SAIFI data to examine long-term trends in grid reliability. LaCommare and Eto [40] link similar data to damage functions reported in Sullivan et al [71] to estimate the scale of damages grid customers incur due to power interruptions nationally each year.

The DOE Office of Electricity (OE) collects more detailed information describing characteristics of specific power interruptions via DOE Form OE-417.[3]. Power interruptions reported to OE are of particular interest due to their size (e.g., more than 50,000 customers affected) or underlying cause (e.g., cyber or physical attack), and are referred to as "emergency incidents".

When an emergency incident occurs, the affected utility (or balancing authority) files an Electric Emergency Incident and Disturbance Report by (via DOE Form OE-417) within 1 or 6 hours of the event, depending on the nature of the incident. Though details about the events leading up to each incident are withheld, much of the reported information is made available to the public including:

[2] Available for download here: https://www.eia.gov/electricity/data/eia861/ (Accessed November, 2018)

[3] Data are available for download here: http://www.oe.netl.doe.gov/oe417.aspx (Accessed November, 2018)

- Affected geographic region
- Start date and time
- End date and time
- Peak megawatts of demand loss
- Number of customers affected
- Alert criteria (i.e., what caused the power interruption to classified as an emergency incident)
- Event type/cause

We refer readers to read the form itself [57] for definitions of these data fields and for a complete list of the criteria that designate an emergency incident.

Finally, many electric utilities post power interruption data on their public-facing websites. These data provide an instantaneous account of the power interruptions currently happening in their service territory. To our knowledge there is no regulatory requirement to provide these data. Though no explanation is given, one viable reason for publicizing the data is to provide real-time updates to grid customers who are experiencing power interruptions. Since there is no regulatory agency overseeing the publication of these data, details such as how frequently information are updated, the minimum size (if any) of the power interruptions posted, and geographic granularity of the information provided are at the discretion of the utility maintaining the website. Websites often state that data are updated every 15 minutes, suggesting that the process is automated.

Utilities that provide the most geographic granularity report power interruptions as point locations described by latitude and longitude coordinates; the data do not indicate the physical significance of these point locations. Most utilities, however, report the number of affected customers aggregated by ZIP code, county, or municipality.

The current work reports on data we collected by archiving these real-time information to compile a historical record of the power interruptions that occurred. In Section 2.3 we discuss the contents of the data and how they were collected. While the bulk of our analysis focuses on this high-resolution data, we also compare information contents of these data with information reported in the EIA and DOE datasets. Through these comparisons, we aim to better understand what information (if any) is lost in computing the aggregate metrics that are publicly reported. The insights we gain allow us to comment on the analysis capabilities that granular data could afford.

2.3 Methods

Data collection

To collect real-time information posted on utility websites, we developed an automated software tool that visits upwards of 30 utility websites on average every 9 minutes and records the current time, the locations of the power interruptions posted, and the number

of customers affected at each location. These data are archived in a PostgreSQL database. These 30 utilities collectively serve about 40% of electricity customers in the United States. We have collected upwards of three years worth of data to date.

Since the current work explores the analysis capabilities of granular data, we focus on a utility that provides high geographic resolution and reports power interruptions by point location. Focusing on a single utility further reduces heterogeneity introduced by operational or climatic differences between utilities.

Without access to utility records, we cannot confirm with certainty that the information we collect provides a precise or complete record of the power interruptions that occurred, though we do show that our observations align well with publicly reported information. In light of these considerations, we want to make it explicit that our aim is not to generalize our results to other utilities or to precisely summarize the interruptions that occurred, but rather to explore what insights granular data can provide that public datasets do not.

We choose a utility in the Western United States that serves nearly 5 million customers across 900 ZIP codes. Of these customers, about 93% are residential, 6% are commercial, and the remaining 1% are split between industrial and agricultural customers. We report on data collected over three years between June 1, 2014 and May 31, 2017. The data include 16 million spatially and temporally distinct observations collected at 120 thousand distinct points in time.

Definition of a power interruption

IEEE Standard 1366 [37] defines a power interruption as the complete loss of electricity service to one or more grid customers. The duration of a power interruption is the time elapsed between when customers lose service and when service is restored.

We form an equivalent definition using the information available to us. We define a power interruption as a series of consecutive observations (or website queries) reporting at least one grid customer without service at a specific point location. By this definition, the number of customers affected by a power interruption can change over subsequent observations, but the location is static.

Figure 2.1 shows time series data illustrating what a power interruption looks like in the data. We show five power interruptions reported on the same day in the service territory we examine for the remainder of this work. Each panel shows the number of customers (y-axis) without service at a specific point location over time (x-axis).

All five power interruptions affect a large number of customers compared with most of the interruptions reported in the data (as we show in Section 2.4), and three of them begin at roughly the same time. Although these observations may lead us to wonder whether there is a causal relationship between them, the data do not explicitly provide information to indicate whether or not this is the case. In the current analysis, we designate these as five distinct power interruptions; we do not attempt to correlate them.

We define the beginning of a power interruption as the first observation reporting that customers are without service at a particular point location. We designate the end of a power interruption as the first observation where no customers remain without service at that location. The duration is simply the time elapsed between these two observations.

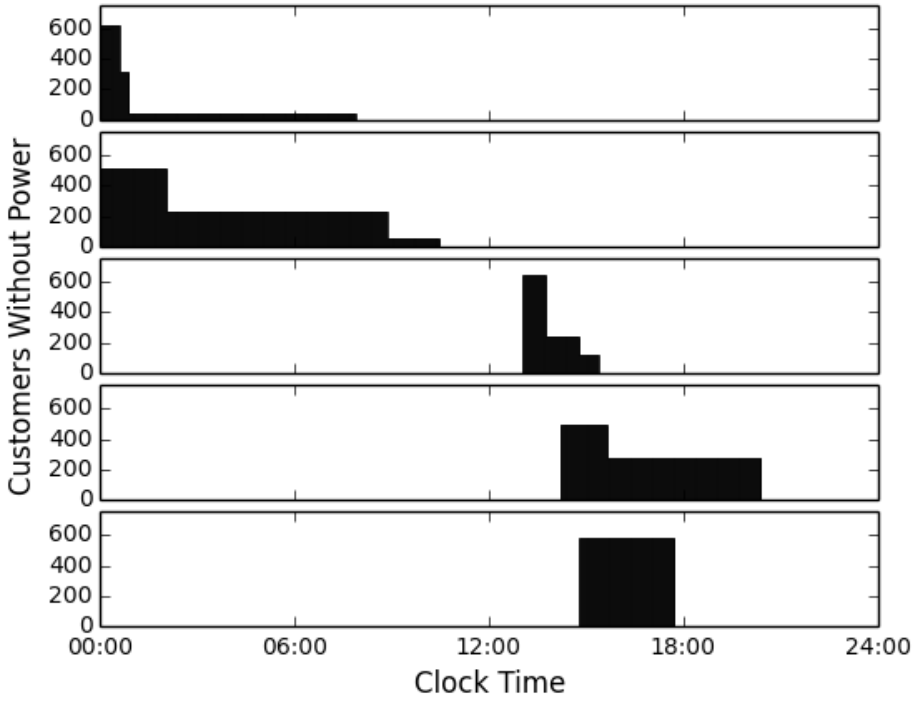

Figure 2.1: Time series of customers reported to be without power at each of five point locations over the course of one day. Locations are all located in the same service territory, and were selected for having large-scale power interruptions that all occurred on the same day.

Intrinsic to this definition is the assumption that customers remain continuously without service between one observation and the next, except when a change in the data indicates that a partial or complete restoration has occurred. Because the data are recorded in discrete sampling intervals, we cannot exclude the possibility that changes occur during the time between consecutive observations, as such changes would not necessarily be evidenced in the data.

By examining how frequently we do observe changes in the data, we can explore how frequently unobserved changes likely occur. Across our entire dataset, the number of customers affected by a particular power interruption changes on average only once every 170 minutes. However, we do observe some change in the overall dataset every time a new sample is recorded (every 9 minutes), for example to indicate when a new power interruption has begun. Because the number of customers at any particular site changes much less frequently than our sampling frequency, the data suggest that unobserved changes in the number of customers affected are rare.

Calculation of SAIDI, SAIFI and CAIDI

From the data described in Section 2.3, we can roughly match the information contents of public reliability datasets described in Section 2.2. We can use these data to compute aggregate reliability metrics including SAIDI, SAIFI and CAIDI [37].

To compute SAIDI, we calculate the total minutes individual customers spend without service during the observed power interruptions and normalize by the number of customers the utility serves, as given by Equation 2.1.

$$SAIDI = \frac{1}{C_{tot}} \sum_{p \in \{P\}} \left[\sum_{i=0}^{N_p} C_{p,i} \times (t_{p,i+1} - t_{p,i}) \right] \tag{2.1}$$

Here C_{tot} describes the total number of customers the utility serves. The set of all power interruptions (p) that begin during a specified time interval is given by P. These time intervals can be continuous (e.g., June 1, 2015) or not (e.g., between 09:00 and 10:00 AM on Weekends). The number of observations recorded between the start and the end of power interruption p is given by N_p. Finally, $t_{p,i}$ and $C_{p,i}$ describe the time stamp and number of customers recorded i observations after the beginning of power interruption p.

Graphically, the summation term for a particular interruption p can also be expressed by the area of the shaded regions in each panel of Figure 2.1.

To compute SAIFI, we calculate the total number of customers that were affected during the observed power interruptions and normalize by C_{tot}, as given by Equation 2.2.

$$SAIFI = \frac{1}{C_{tot}} \sum_{p \in \{P\}} \left[C_{p,0} + \sum_{i=1}^{N_p} \Delta C \right] \tag{2.2}$$

Here $C_{p,0}$ describes the number of customers affected at the beginning of power interruption p, and ΔC describes the magnitude of all *positive* changes in the number of customers affected by each power interruption, given by Equation 2.3.

$$\Delta C = \begin{cases} C_{p,i} - C_{p,i-1}, & \text{if } C_{p,i} > C_{p,i-1} \\ 0, & \text{otherwise} \end{cases} \tag{2.3}$$

Finally, CAIDI is calculated by taking the ratio between SAIDI and SAIFI.

2.4 Results and Discussion

Variability in interruption duration

Results

Figure 2.2 shows the probability and cumulative distribution functions (or PDF and CDF), of power interruption duration. We find that durations follow an approximately log-normal distribution with geometric mean and standard deviation equal to 95 and 3 minutes, respectively. While we explored other parametric distributions, we find the log-normal distribution provides the best overall fit to the data.

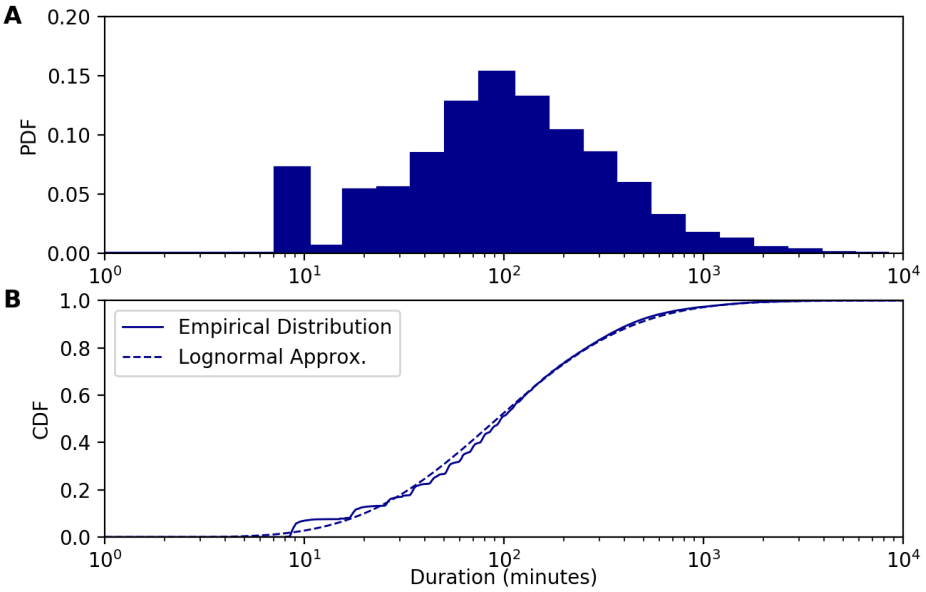

Figure 2.2: (A) Probability and (B) cumulative distributions of power interruption duration (in minutes). 10% of the observed interruptions last fewer than 10 minutes. The dashed line in (B) shows that the distribution is approximately log-normal with a mean of 95 minutes and a standard deviation of 3.

Though short-duration power interruptions (<30 minutes) deviate from the log-normal distribution, this observation may simply be a construct of our dataset. Discrete sampling intervals mean that the duration of power interruptions is not well characterized unless the duration is much longer than our sampling frequency.

Implication 1: The distribution of power interruption duration can be well characterized by only two parameters: one describing the central value of the distribution and another describing the variability among observations.

Implication 2: The range of typical restoration times spans several orders of magnitude.

Discussion

The heterogeneity we observe is not evidenced in public data reporting the mean duration (CAIDI), but not the variance. Implication 1 suggests that the heterogeneity is well approximated by a parametric distribution. Publicly reporting both the mean and the variance would increase transparency into the distribution of restoration times and the relative frequency of long-duration power interruptions.

Furthermore, because the distribution is not symmetric, the mean (CAIDI) is skewed by a relatively small number of power interruptions with durations that are orders of magnitude

higher than average. In our data, about 70% of observations fall below the arithmetic mean and about 30% fall above; these proportions are characteristic of a log-normal distribution. The arithmetic mean is 180 minutes or nearly twice the geometric mean (which describes the center of the distribution). The center of the distribution could be more accurately represented if the median or geometric mean were reported in addition to the arithmetic mean. The geometric mean would better represent the typical duration, without being affected by long-duration interruptions at the tail of the distribution.

These new metrics could support efforts to more accurately compute the damages due to power interruptions, as survey data consistently reports that damages scale with interruption duration (for example, see [71], [18] and [16]). Works in the literature to date typically use CAIDI to represent interruption duration [40, 43]. Our results show that this approximation overestimates damages for the 70% of power interruptions that are shorter than the mean, while underestimating damages (potentially by orders of magnitude) for the 30% of power interruptions that are longer than the mean. It is unclear whether more granular information would increase or decrease overall damage estimates such as those reported in [43] and [40].

Implication 2 underscores the need for further research characterizing damages due to long-duration power interruptions, as damage functions reported in Sullivan et al [71] (and applied in [43, 40]) are only valid for power interruptions shorter than 8 hours. Based on our records, nearly 8% of power interruptions include some customers who remain without service for more than 8 hours. These long-duration interruptions are potentially much more damaging than short interruptions.

Implication 2 also raises questions about whether certain customers are more likely to experience long-duration interruptions than others. The literature suggests that this may be the case, as policies targeting metrics like SAIDI and SAIFI can favor upgrades in high-density areas where service reliability is already high [6]. However, if studies could show that long-duration interruptions contribute a higher share to overall damages, policies targeting damage reductions instead of SAIDI and SAIFI would likely favor measures that benefit customers in underperforming regions of the grid.

However, the lowest performing areas of the grid may be remote areas with few customers per line-mile. The implication is that there may be high costs to improve service for a relatively small number of customers. Where infrastructure upgrades are not cost effective, backup power systems or local generation may provide a cheaper alternative to achieve similar damage reductions. Today, it is customary for individual grid customers to shoulder the costs of backup power systems while ratepayers shoulder the costs of infrastructure upgrades. Our results raise questions about who should pay for backup power, and whether incorporating distributed generation into the planning process could lead to more efficient investment in grid performance, or more equity in how the burden of power interruptions is distributed among different grid customers.

Variability in interruption size

Results

Next we examine the PDF and CDF of the maximum (or peak) number of customers recorded during each individual power interruption (Figure 2.3). Of these power interruptions, 70%

affect only one customer and 80% affect no more than 10 customers. By our definition of a power interruption, we record no single interruption that affected enough customers (50,000 or more) to be classified as an emergency incident based on size alone.

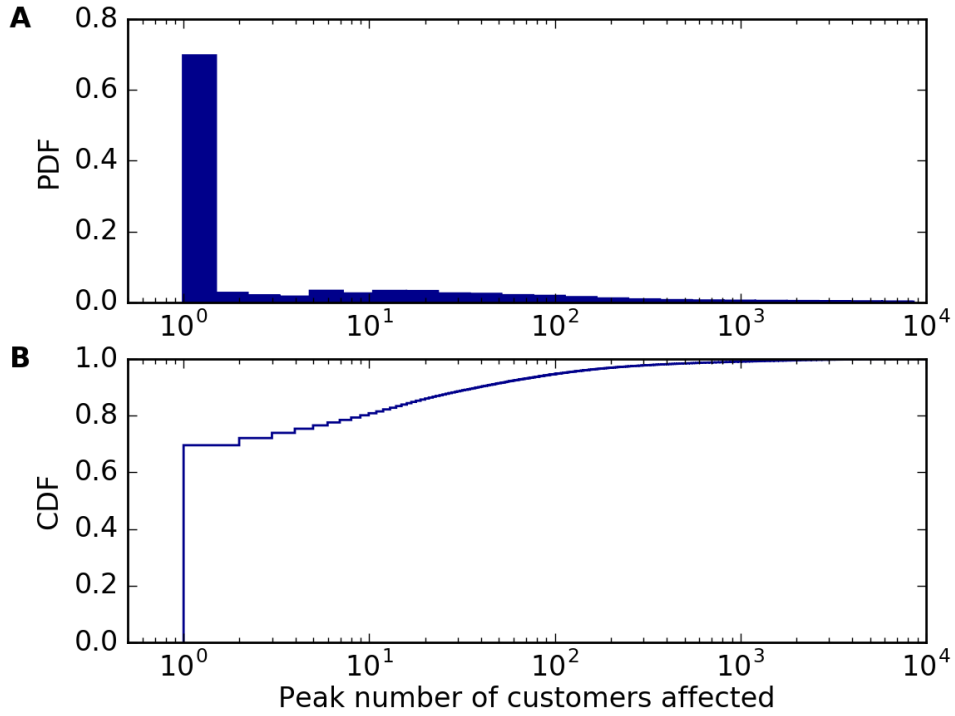

Figure 2.3: (a) Probability and (b) cumulative distribution of the peak (or maximum) number of customers affected by individual power interruptions. 70% of observed power interruptions affect at most one customer; no single interruption affected enough customers (50,000 or more) to be considered an emergency incident.

However, DOE records include several emergency incidents that are spatially and temporally aligned with our dataset. The reason we do not directly observe these large-scale events is that the definition of an emergency incident differs from our own definition of a power interruption. An emergency incident can include power interruptions that occurred at multiple locations so long as they are linked by a common cause (for example a contingency event, or severe weather).

Our dataset does not include enough information to draw causal relationships between interruptions. Instead, we look for possible emergency incidents by examining time intervals where at least 50,000 customers were without power in the service territory as a whole. If these interruptions are causally linked, then we expect them to be included in the emergency incident data.

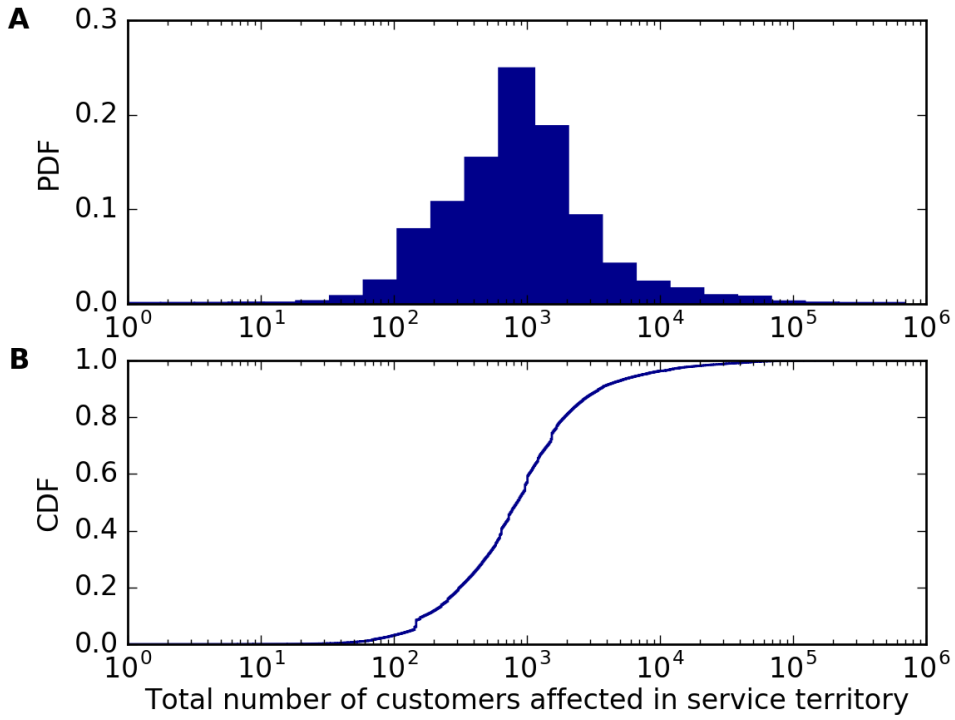

Figure 2.4: (a) Probability and (b) cumulative distribution of the total number of customers without service in the entire service territory at each observation time. 600 observations (0.5% of observations) report that the total number of customers without service exceeds 50,000 customers.

Figure 2.4 shows the PDF and CDF of the total number of customers affected in the service territory across all of the observations we record.

We recorded 600 observations during which time more than 50,000 customers were without power for at least one hour. These observations span 14 continuous time intervals, 11 of which coincide with emergency incidents reported in the DOE dataset. Our dataset does not include enough information to determine what differentiates the remaining 3 incidents from the 11 that were subject to DOE reporting requirements.

Over all three years, these 14 time intervals collectively account for 0.5% of the observations we record, 8% of the power interruptions we observe, 40% of SAIDI and 13% of SAIFI. Ten of these intervals occurred in 2017 (including all three that were not reported in the DOE data). If we exclude observations recorded after January 1 2017, emergency incidents account for only 6% of SAIFI (see Implications 2 and 3 below).

Table 2.1 lists the approximate time of each incident, the peak number of customers affected, and the number of customers reported in the DOE dataset. We note that despite methodological differences in how we calculate the number of customers affected by each

Table 2.1: Time and size of power interruptions affecting a peak number of 50,000 or more grid customers, and corresponding total number of affected customers, as reported to DOE. "Unknown" indicates that an incident was reported to DOE but that the number of customers affected was reported to be unknown; "Not Reported" indicates that an incident was not reported to DOE.

Year	**Season**	**Peak customers recorded**	**Total customers reported**
2014	Winter	120,000	Unknown
2014	Winter	74,000	84,000
2015	Winter	64,000	65,000
2015	Fall	57,000	56,000
2017	Winter	104,000	106,000
2017	Winter	78,000	87,000
2017	Winter	74,000	75,000
2017	Winter	61,000	64,000
2017	Winter	168,000	169,000
2017	Winter	52,000	Not Reported
2017	Winter	76,000	Not Reported
2017	Winter	57,000	Not Reported
2017	Spring	130,000	100,000
2017	Spring	90,000	88,000

incident, we arrive at remarkably similar values as those reported the DOE dataset (see Implications 5 and 6 below).

Implication 3: Year-to-year differences in the size and frequency of large-scale interruptions can have major implications on SAIFI.

Implication 4: Emergency incident data provides only a limited sample of the power interruptions that occurred. Furthermore, the sample is not representative of overall grid performance, as it only includes the most severe interruptions.

Implication 5: While we cannot with certainty confirm the accuracy of the data we collect, we calculate approximately the same number of customers affected as is reported in the emergency incident data. This observation suggests that our data are indeed representative of the interruptions that occurred.

Implication 6: Despite methodological differences in how we estimate the number of customers affected by each incident, we observe less than a 1% overall difference in the number of customers we compute compared with those reported in the DOE dataset. This result indicates that day-to-day power interruptions have a relatively small impact on the overall number of customers affected compared with large-scale disturbances.

Discussion

The bias we observe (Implication 2) raises questions about whether analyzing comprehensive power interruption data could reveal patterns that are characteristically different from the patterns Hines et al [35] observe in emergency incident data. For example, large-scale power

interruption data likely exhibit seasonal patterns more aligned with severe weather events (such as hurricanes) than weather patterns with more nuanced reliability implications. In Section 2.4, we explore the implications of bias by reporting on temporal patterns observed in our dataset and compare our findings with observations reported in Hines et al [35].

The observation of bias also raises questions about how to group power interruptions based on criteria that provide more actionable insights than grouping them simply by size. For example, brief descriptions of vulnerabilities that have led to emergency incidents data include severe weather and cascading failure. However, the same causal factors can also result in power interruptions that affect fewer than 50,000 customers and may not be reported as emergency incidents. Examining how frequently emergency incidents are attributed to a particular cause does not necessarily tell how severe a particular vulnerability really is, or how much the system will benefit from policies for mitigating it.

Novel methods for classifying power interruptions could yield new insights that emergency incident data do not provide. For example, grouping power interruptions based on the environmental stressors that caused them (e.g., high wind speed or lightning) could inform decisions about which vulnerabilities are most urgent to address. Further research is needed to determine the most meaningful ways to classify power interruptions, and to develop statistical methods for doing so.

Temporal patterns in SAIDI, SAIFI and CAIDI

Results

Figure 2.5 shows box-and-whisker plots describing the distribution of SAIDI, SAIFI and CAIDI calculated in hourly increments and grouped by time of day. The data show that both SAIDI and SAIFI typically increase in the morning, peak at about 08:00AM, and decrease over the remainder of the day. The only appreciable change in CAIDI is that variability decreases among power interruptions that begin between 08:00 and 10:00AM. One possible explanation for this trend is that planned maintenance typically occurs during those hours, a hypothesis which is supported by the observation that the number of power interruptions (SAIFI) is an order of magnitude higher in the morning than at night.

Figures 2.6 and 2.7 show box-and-whisker plots of SAIDI, SAIFI and CAIDI by season. We calculate each reliability metric by grouping observations into hourly (Figure 2.6) and daily increments (Figure 2.7).

Figure 2.6 shows no evidence of seasonal heterogeneity in hourly metrics. However, Figure 2.7 shows that the extreme values of daily SAIDI and SAIFI are one to two orders of magnitude higher in winter than in other seasons. The 5 days with the most severe values of SAIDI account for 20% of overall SAIDI; the most severe 30 days account for 50% of overall SAIDI (see Implication 8).

Implication 7: Power interruptions are not homogeneously distributed over the course of the day. Rather, patterns in SAIDI suggest that interruptions are more likely to occur at certain times of day. We see no clear evidence to suggest that interruption duration (CAIDI) follows a similar characteristic diurnal pattern.

Implication 8: In our dataset, seasonal and annual changes in reliability metrics are largely attributable to a small number of days with particularly severe power interruptions.

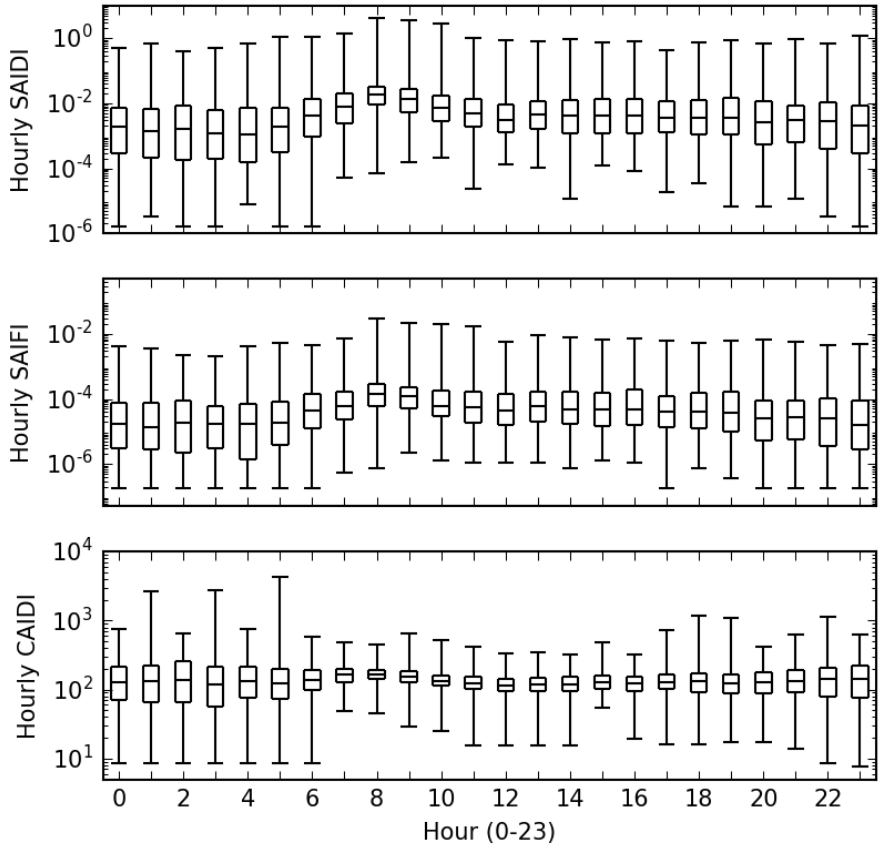

Figure 2.5: Box-and-whisker plot showing the distribution of SAIDI, SAIFI and CAIDI computed in hourly increments and grouped by time of day. Whisker endpoints denote the 5th and 95th percentiles, box endpoints denote the 25th and 75th percentiles, and the horizontal line inside each box denotes the median value.

Implication 9: Daily reliability metrics exhibit seasonal heterogeneity that hourly metrics do not. This suggests seasonality is driven by interruptions that intensify over the course of several hours, rather than interruptions that occur instantaneously.

Discussion

The seasonal and diurnal patterns Hines et al [35] observe in emergency incident data differ from the patterns we report. For example, we record the most power interruptions during winter months, while Hines et al report that the frequency of emergency incidents increases during both summer and winter months. We also observe that most power interruptions begin in the morning, while Hines et al report that most emergency incidents began in the

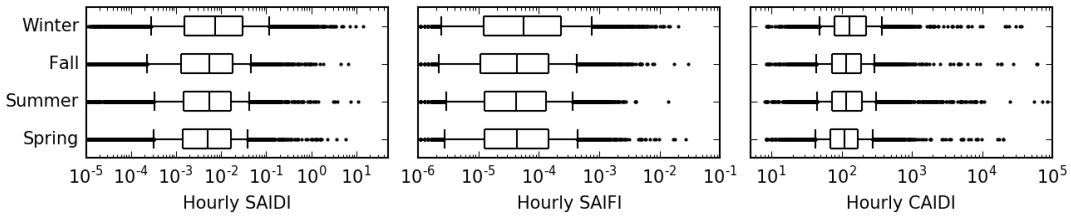

Figure 2.6: Box-and-whisker plot showing the distribution of SAIDI, SAIFI and CAIDI computed in hourly increments and grouped by season. Whisker endpoints denote the 5th and 95th percentiles, box endpoints denote the 25th and 75th percentiles, and the vertical line inside the box denotes the median value, and points denote observations outside the 5th and 95th percentiles for each season.

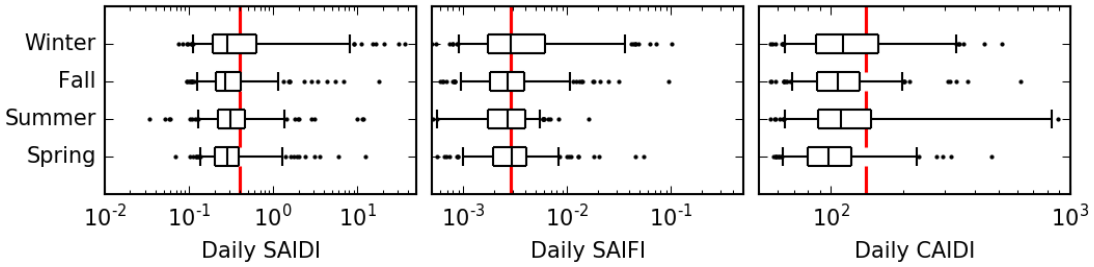

Figure 2.7: Box-and-whisker plot showing the distribution of SAIDI, SAIFI and CAIDI computed in daily increments and grouped by season. Whisker endpoints denote the 5th and 95th percentiles, box endpoints denote the 25th and 75th percentiles, the vertical line inside each box denotes the median value, and points denote observations outside the 5th and 95th percentiles for each season. The red lines denote the average daily values of each metric reported to EIA in 2015.

afternoon.

One explanation for these differences is that our dataset focuses on a single utility service territory, while Hines et al report on emergency incidents observed nationwide. Regional differences in weather patterns could lead to characteristically different seasonal patterns in grid performance. For example, the utility we examine (which is located in the Western United States) may not be exposed to particularly severe weather during the summer. However, utilities in other parts of the country may be vulnerable to lighting storms, tornadoes or hurricanes that occur during the summer.

Another explanation for differences between the two datasets is that the causal factors that lead to emergency incidents follow characteristically different temporal patterns than the factors that lead to small-scale power interruptions. For example, say we could confirm that the diurnal pattern we observe is attributable to scheduled maintenance. Because scheduled

maintenance likely does not lead to emergency incidents, we would not expect emergency incident data to exhibit the same diurnal patterns as we observe in Figure 2.5.

Characterizing temporal patterns can provide insights relevant to (1) grid operations/planning, (2) damage estimation, (3) risk assessment, and (4) identifying the drivers of year-to-year differences in grid performance.

First, Hines et al discuss how patterns observed in large-scale power interruptions could inform policy decisions for improving grid performance. Yet responding efficiently and effectively to large-scale power interruptions is only one component of maintaining a reliable and resilient grid. Low-cost measures may also reduce the duration or frequency of small-scale power interruptions which cumulatively contribute approximately the same amount to annual SAIDI as large-scale power interruptions. Though public datasets do not contain sufficient information to be able to characterize these patterns, doing so could help in identifying opportunities for improving performance.

Second, results in the literature show that damages are related to the time of day and time of year a power interruption occurs [16, 18, 71]. For lack of more granular information, both LaCommare and Eto [40] and Larsen [43] assume that power interruptions are homogeneously distributed with respect to time. Replicating these studies with data describing temporal heterogeneity in SAIDI, SAIFI and CAIDI would lead to more accurate damage estimates than those currently reported.

Third, there is discussion in the literature regarding the growing penetration of electric heating, cooling, and transportation loads, and how these loads will affect electric power systems. One under-explored research area is how growth of high-value loads could change the severity of damages, particularly if patterns in SAIDI and SAIFI coincide with times when these loads are most needed. Case studies of power interruptions in Canada and Sweden (where penetration of electric heating is high) reported loss of electric heating loads to be a leading cause of damages primarily because both interruptions occurred during severe ice storms [16, 18]. Research examining the interplay between grid performance and end use electrification could help in anticipating emerging power system vulnerabilities before high-damage power interruptions occur.

Finally, a deeper understanding of seasonal trends could also advance efforts to understand how and why grid performance is changing over time. Studies reporting that grid performance has been decreasing [44, 8, 35] remain inconclusive as to what is driving this trend. For example, year-to-year changes in grid reliability could be due to weather patterns, or to aging grid infrastructure. Caswell et al [17] discuss methods for normalizing reliability metrics to control for weather in order to isolate changes due to other factors. They point out that to do so, the spatio-temporal granularity of the reliability metrics calculated must be similar to the granularity of the weather phonemona that caused the change in grid performance. Since public metrics are typically reported annually and over large service territory, they are not granular enough to support weather normalization. However, controlling for weather-related differences is a necessary step towards evaluating how much of the variability in performance is due to weather compared with other factors such as operational, maintenance, or investment decisions. Discussion of these methods in the literature to date is largely conceptual, and quantitative research is needed to explore how much aggregation is appropriate.

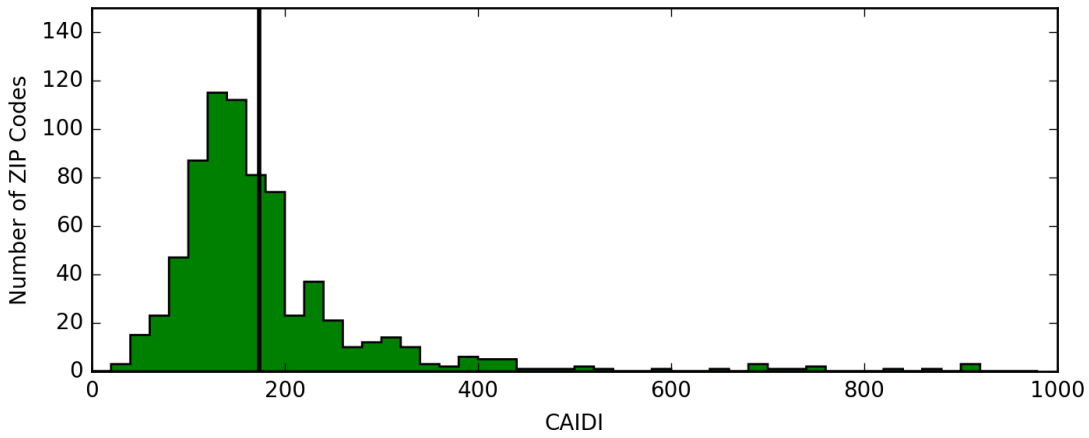

Figure 2.8: Distribution of annual CAIDI for each ZIP code in the service territory. The vertical line denotes overall system CAIDI (170 minutes).

Geographic variability in interruption duration

Results

Finally, we examine geographic variability in CAIDI (or the average restoration time) across different regions of the service territory. We map the reported point locations to ZIP codes and compute CAIDI for interruptions observed in each ZIP code. We note that calculating SAIDI and SAIFI at this geographic resolution is not possible using the current dataset, as C_{tot} is known only for the service territory as a whole. However, C_{tot} is not needed to compute CAIDI, as CAIDI is a ratio between SAIDI and SAIFI (where both quantities are normalized by the same value C_{tot}).

Figure 2.8 shows the distribution of CAIDI for 700 ZIP codes in the service territory, excluding ZIP codes with fewer than 50 power interruptions. Across all 700 ZIP codes, we calculate that CAIDI is 170 minutes (denoted by the vertical line). Within individual ZIP codes, CAIDI ranges from 10 to 1,000 minutes.

Implication 10: Customers in different parts of the service territory experience characteristically different restoration times.

Implication 11: Because they are averaged over large geographic regions, system-wide metrics such as CAIDI (and perhaps SAIDI and SAIFI as well) miss a great deal of variability in the typical characteristics of power interruptions that customers in different subsets of the service territory experience.

Discussion

Characterizing geographic heterogeneity in CAIDI could lead to potentially large changes in damage estimates reported in the literature. Damage functions reported in Sullivan et al [71] are sensitive to the duration of a power interruption, as well as the mix of customers affected. For lack of more granular data, works in the literature to date assume that SAIFI

and CAIDI are homogeneous throughout the service territory. In other words, these studies assumed that in a given year, all customer types experience the same number and duration of power interruptions. If it were shown (for example) that industrial customers experience characteristically shorter restoration times than residential customers, damage estimates would decrease because commercial and industrial customers make up the largest share of these damages [40].

Reporting reliability metrics for different subsets of customers (e.g., customer type, ZIP code, rate class, or service level) would also unlock opportunities to explore a wide range of policy questions that to our knowledge remain unanswered. For example, reliability metrics reported by customer type would allow us to more accurately assess damages. Furthermore, reliability metrics reported by ZIP code could provide insight into whether the costs of power interruptions are being equitably distributed among customers, or whether policy intervention is needed to redistribute these costs. Detailed consideration of these (and other) use cases would be needed to determine how much granularity is warranted in publicly reported data, or whether granular data should be made available to researchers on a case-by-case basis.

Characterizing heterogeneity in power interruption characteristics could also inform data-driven methods for optimizing investment in performance upgrades. For example, Larsen [43] describes a cost-benefit framework for evaluating if and where undergrounding distribution infrastructure is cost effective. The author finds that undergrounding is typically (though not always) cost effective in urban but not in rural areas. However, that result is sensitive to the assumption that restoration times are homogeneous throughout the service territory, and that they are the same for all customers.

To illustrate how these assumptions could affect damage estimates, consider hypothetical results showing (1) that grid customers in urban areas experience fewer and shorter power interruptions than customers in rural areas, and (2) that there are more commercial and industrial customers located in urban areas than in rural areas. After updating damage functions to account for heterogeneity in interruption duration (result 1), the benefits of undergrounding urban power lines would likely decrease because urban customers typically experience restoration times that are shorter than CAIDI. On the other hand, updating damage functions to account for heterogeneity in the customer mix (result 2) would increase the benefits of undergrounding in certain areas, as commercial and industrial customers incur much higher damages than residential customers [71]. Understanding how the characteristics of power interruptions differ for different groups of customers is critical to informing undergrounding policies.

This example illustrates how spatially resolved power interruption data can help in developing more accurate models for evaluating the costs and benefits of a particular policy for investing in grid performance. Such policies can support decisions to invest in system upgrades if and where higher levels of service reliability are warranted.

2.5 Conclusion and Policy Implications

The main contribution of this work is to show that power interruption characteristics exhibit substantial geographic and temporal heterogeneity that are not evidenced in publicly

available datasets.

The heterogeneity we observe sheds light on a number of important policy questions. Detailed analysis exploring these questions will require high-resolution data, as aggregate metrics provide limited insight into the characteristics of the power interruptions that individual grid customers experience. We discuss a wide range of analysis questions that could leverage high-resolution data to advance how we evaluate, regulate, and invest in grid performance.

We contextualize our results by examining how granular data could change results reported in the literature to date, and show that limited access to high-resolution data has been a barrier to research progress.

As energy sources, grid infrastructure, and electricity end uses evolve, the causes and implications of power interruptions are also likely to change. To ensure that power systems are able to support electrification of new end uses such as transportation or heating without adverse consequences, granular research is needed to examine the drivers of regional and year-to-year differences in performance. Research is also needed to assess adequacy and equity of grid performance today, and to confirm that both the burden of underperformance and the costs of improving performance are equitably distributed among ratepayers. Examining these questions could help us to better understand and quantify grid performance, and ensure that policymakers are leveraging state-of-the-art data and methods to make the most informed decisions available to them.

Although high-resolution data suitable for exploring these questions do exist, these data are proprietary and are not readily made available to researchers conducting policy analysis. However, the consequences associated with leaving unanswered questions related to current and future grid performance could be severe. To ensure that grid performance is adequate to support the needs of a society that relies heavily on electricity, it is our recommendation that high-resolution power interruption data be made more readily available to support data-driven reliability research.

Chapter 3

Bayesian Hierarchical Methods for Quantifying Failure Risk

This chapter gives a Bayesian hierarchical formulation of conventional engineering models examining the failure probability of components under stress. The formulation provides a basis for calling into question the adequacy of the data and the validity of the model. The work explores the limits of making decisions informed by sparse data, and provides a basis for internalizing expert judgement in the form of Bayesian priors. The methods are applied to optimize the number of repairs needed to achieve a specified risk tolerance for the system overall. Results compared against more conventional methods where model uncertainty is not considered.

This chapter was originally presented at the Power Systems Computation Conference (PSCC) and appears in the journal *Electric Power Systems Research*. It may be cited as follows:

> Laurel N. Dunn et al. "Bayesian hierarchical methods for modeling electrical grid component failures". In: *Electric Power Systems Research* (July 2020). URL: %5Curl%7Bhttps://arxiv.org/abs/2001.07597%7D

The original funding acknowledgement is included here as a footnote.[1]

3.1 Introduction and Motivation

Millions of individual components make up electric power systems; each of these has a finite operating life time. Grid components are subjected to physical force, chemical processes, and operational stress that can give rise to failure. Grid hardening and preventative maintenance can mitigate the risk of unexpected failures, but risk reduction measures may also negatively impact cost and even performance. High-fidelity models for representing failure processes can provide insights valuable to a number of resource allocation decisions, including system upgrades [52], component stockpiling [21], and disaster response [54].

[1]This research was supported by a 2019 Seed Fund Award from CITRIS and the Bantao Institute at the University of California.

The current work presents a framework for probabilistically modeling the risk that components will fail when exposed to stress. This work provides a basis for leveraging grid data to detect and quantify unmitigated vulnerabilities in operational power systems, and to inform upgrade decisions. We explore challenges that arise when using data-driven methods to characterize failure properties of real-world systems. These challenges include heterogeneous performance characteristics of individual components (e.g., due to differences in state-of-health), and sparse observational data of past failure events. The contributions of this work are twofold:

First, we propose a Bayesian hierarchical framework for capitalizing on data and domain expertise to model component-level and system-wide failure properties. Our formulation captures the stochastic and heterogeneous nature of failure processes by assuming model parameters are uncertain and random. Accounting for this randomness better captures the range of outcomes that could occur due to inherent uncertainties. This feature makes our model particularly well-suited to examining the possibility of low-probability risk scenarios that may not be evidenced in past data. Applications could include scenario analysis of evolving risks due to climate change, including wildfires and severe weather events.

Second, we formulate a mathematical model for optimizing component upgrades to achieve specified risk thresholds. This upgrade policy accounts for uncertainty in the failure parameters of the system, and provides insight into the performance characteristics of existing components. We describe how decision-makers can apply these insights to identify the subset of components with the highest probability of failure, thus minimizing the number of upgrades needed to achieve risk reduction goals.

This paper is structured as follows. In Section 3.2 we contextualize our work within the existing literature. Section 3.3 motivates the use of Bayesian hierarchical models in power systems, and describes our model formulation. Section 3.4 describes methods for parameter estimation, model selection, and model evaluation–and formulates the optimal upgrade policy. Section 3.5 compares the parameters, statistical properties, and decision-making implications of our model with alternative models for systems exhibiting a range of failure properties. Section 3.6 presents concluding remarks and outlines opportunities for future work.

3.2 Literature Review

A number of studies in the literature examine failure properties of regional power systems during extreme events. Some of these studies use prescriptive failure models to inform probabilistic models for assessing risk and weighing upgrade policies in simulation [52, 58, 64]. These studies typically do not call into question the validity of the failure models they build upon. Other studies mine failure data from past events to characterize vulnerabilities in operational power networks, for example to cascading failures [19], hurricanes [63] and earthquakes [59]. However, these studies generally do not consider uncertainty from erroneous or sparse data.

A separate body of literature focuses on estimating the probability that individual components will fail when exposed to certain ambient or operating conditions. It is conventional to use parametric failure models (or "fragility curves") to describe these failure probabilities.

These curves describe the probability that a particular component will fail under stress. For example, a fragility curve could describe the probability that a wood pole will fail under extreme wind speeds. A survey of models and methods used to fit fragility curves is given in [42]. A number of studies report models characterizing failures that occurred under severe conditions such as hurricanes [63, 50] and earthquakes [59].

The literature also suggests, however, that failure properties are not the same for all grid components. For example, [23] and [66] mine grid data to characterize differences in failure probabilities based on what is known about components that failed in the past. However, grid data are often noisy and sparse, and the degradation mechanisms that give rise to failure may be costly (if not impossible) to monitor [22].

These studies underscore two critical research needs. First, there is a need to better understand failure properties of grid components in light of the fact that grid data can be sparse, noisy, costly, and inaccurate (see [66]). Second, there is a need to examine if or how uncertainty in component-level failure models impact system-level operation in low-probability yet potentially high-impact events. Our contribution is to provide a quantitative basis for characterizing this uncertainty using what is known about the system from first principles, design-oriented, and empirical models.

3.3 Model Formulation

Here, we formulate a Bayesian hierarchical model to characterize failure properties of grid components in a regional power system. We begin by providing philosophical and mathematical context on Bayesian hierarchical models, and discuss why they are uniquely well-suited to the application at hand. We then describe the functional form of the models used to characterize failure probabilities for individual grid components, and draw a mathematical link between component-level failure probabilities and system-wide failures.

Bayesian Hierarchical Model

Hierarchical models provide a basis for characterizing systems where the relationship between the input and output variables is probabilistic and uncertain. The model is structured as a multi-level hierarchy where the lowest level describes some probabilistic process whose outcome is conditioned on the outcomes at other levels in the hierarchy, in light of the uncertainty therein. The various levels of the model may describe uncertainty in the outcome, in the data, in the system of equations relating the inputs to the outputs. These uncertainties can be modeled collectively or individually to inform our understanding of different aspects of the system.

To formalize this mathematically, let us consider a system where the output of the system y is related to some input x by a given model $M_\theta(x)$ with parameters θ, such that

$$y = M_\theta(x) \tag{3.1}$$

If the input x is a realization of some random variable X, then the output Y is also a random variable which is conditionally dependent on X. The probability of observing a particular

set of outcomes $P(x, y)$ can be written

$$P(x, y) = P(y|\theta, x)\ P(x) \tag{3.2}$$

where the outcome y depends on the realization of x that was observed, and on the parameters θ that are given. This formulation can be extended to account for additional sources of uncertainty, for example in how the system is parameterized.

Hierarchical models capitalize on this structure of conditional probabilities to characterize the likelihood that a particular parameterization is correct, in light of the observations x and y available to us. Inherent to this approach is the notion that parameter estimates can only be as definitive as the data that are available to compute them. When the data are sparse, or when the mapping of X onto Y is imprecise (e.g., due to measurement noise, or unobservable system dynamics), then parameter estimates will be uncertain.

To demonstrate this mathematically, we return to (3.2) above. Suppose we wish to estimate the probability $P(\theta|x, y)$, or the distribution of unknown parameters θ, given an observation of x and y. The Bayesian formulation is as follows:

$$P(\theta|x, y) = \frac{P(\theta)\ P(x, y|\theta)}{P(x, y)} \tag{3.3}$$

Here, $P(\theta)$ is the prior distribution of the parametes θ, and $P(\theta|x, y)$ is the posterior distribution given the data (x and y). A well-informed specification of the prior distribution $P(\theta)$ allows Bayesian models to internalize expert judgement, while also being informed by empirical data. Such specifications may be advantageous in cases where empirical data are lacking.

The term $P(x, y|\theta)$ is referred to as the "likelihood function". It describes the probability of observing the data (x, y) given a particular estimate of the parameters. Finally, the denominator $P(x, y)$ is the likelihood of observing the data independent of the parameters. Using the law of total probability, we can rewrite the denominator as

$$P(x, y) = \int_{\theta \in \Theta} P(x, y|\theta) p(\theta) d\theta \tag{3.4}$$

Where $p(\theta)$ is the probability density function of the random variable θ. The quantity $P(x, y)$ is referred to as the "normalizing constant" and generally cannot be computed analytically since the density function $p(\theta)$ is unknown. Instead, we use Markov Chain Monte Carlo methods (described in Section 3.4) to compute the posterior numerically.

Component Fragility Curves

Fragility curves are commonly used in reliability engineering to describe the probability that a system will fail when exposed to different magnitudes of stress. These curves provide a quantitative basis for evaluating how vulnerable a system is to different modes of stress, and to stress conditions that are rarely (if ever) observed in practice.

For example, consider an analysis examining the number of wood poles likely to be destroyed during a high wind speed event. For any given pole, the failure probability is related to the severity of the winds, to structural loading on the pole, and of the structural integrity

of the pole. One approach for examining failure risk is using finite element analysis–or high-dimensional models. These models, however, rely on simplifying assumptions and presume that our understanding of the physics underlying failure mechanisms is complete. Fragility curves offer reduced order models characterizing a simplification of the failure process, such that the parameters can be empirically derived.

We use a generalized linear model with a logistic link function relating the stress input conditions X to failure occurrence, as is consistent with common formulations expressed in [42] and [53]. We refer readers to these texts for a more detailed discussion of fragility analysis.

We define $g_j(x)$ to be the probability that a particular component j will fail given stress conditions x observed over some time interval, written mathematically as

$$\begin{aligned} g_j(x) &= P(\text{“component fails”}|x) \\ &= \frac{1}{1+\exp\left[-\sum_{i=1}^{p}\beta_{j,i}(x_i-\alpha_{j,i})\right]} \end{aligned} \tag{3.5}$$

Here x_i is a vector of time series observations for a particular stress condition. Elsewhere in the text, we refer to X—a p-dimensional matrix of stress conditions measurement data. For each stress condition x_i, the threshold $\alpha_{j,i}$ describes the ability of component j to withstand mild to moderate stress conditions. The coefficient $\beta_{j,i}$ describes how quickly the failure probability increases as stress conditions approach (and exceed) the threshold $\alpha_{j,i}$. These thresholds and slopes parameterize the failure mechanics of the system, providing a probabilistic relationship between the stress conditions X and failures. We refer to the parameters collectively as θ. Parameters may differ among components j based on engineering specifications and equipment condition. Here, we assume that the conditions x are homogeneous across all components.

System-Wide Failure Properties

Here, we extend the component-level failure model in (3.5) to model system-wide failure. Power systems are designed and operated to be resilient to the loss of one or two components. However, losing additional components can cause the system to become inoperable, even when only a small fraction of components are affected. Since the ambient conditions that give rise to failures (e.g., storms, heat waves, etc.) typically occur over a time-span of hours to days, the probability that any one specific component will be lost to the system is small (component lifespans are typically on the order of decades).

At the component level, failure occurrence is a binary random variable. Assuming that all the components have the same failure probability $g(x)$, the total number of failures in the grid can be modeled as a Bernoulli process with failure probability $g(x)$ (see Section 3.3). It follows that from a systems perspective, the overall number of failures Y can be modeled as a sequence of Bernoulli trials. The number of trials is simply equal to the number of components in the system N, and Y follows a Binomial distribution where $Y \sim Binomial(N, g(x))$.

The Poisson limit theorem tells us that if N is large, a Binomial random variable can be approximated by a Poisson process with rate $\lambda = N \times g(x)$. The probability of observing a

particular number of failures y during the interval of time considered is

$$P(Y = y) = e^{-N \times g(x)} \frac{(N \times g(x))^y}{y!} \tag{3.6}$$

To put this in words: the system-wide failure process Y can be approximated by a Poisson process, where the failure rate on the system can be generalized from the failure probabilities of individual components, as specified in (3.5).

Having formalized the failure probability distributions, we revisit our description of the model $M_\theta(x)$. We define M_θ to be a function mapping ambient conditions $x \in \mathbb{R}^p$ onto a random variable Y describing the number of failures in the system. The variable Y is an inhomogenous Poisson process; the observed failures y are a realization of Y. Mathematically, $M_\theta(x)$ is defined as:

$$M_\theta : x \mapsto Y \tag{3.7}$$

$$Y \sim \text{Poisson}\left(\frac{N}{1 + \exp\left[\sum_{i=1}^{p} -\beta_i(x_i - \alpha_i)\right]}\right) \tag{3.8}$$

3.4 Analysis Methods

Figure 3.1 provides a visual overview of the framework we use to fit and evaluate our Bayesian hierarchical model. The following paragraphs provide details about (A.) methods for estimating parameters given a candidate model $M_\theta^i(x)$, (B.) choosing the best model $M_\theta^*(x)$ from a library of options $\mathcal{M}_\theta(x)$, and (C.) benchmarking the performance of our model against others.

Parameter Estimation

We use Markov Chain Monte Carlo (MCMC) methods to compute the posterior distribution of the parameters θ.

MCMC encompasses a category of algorithms that approximate an unknown probability distribution by sampling candidate values from some proposed distribution, and then updates the proposed distribution at each iteration. Updates are made according to some rule based on the likelihood function. In our case, we examine the likelihood of the data given a candidate set of parameters θ^*, given by $P(x, y|\theta^*)$.

Different MCMC algorithms use different rules for updating the sampling distribution of θ^* at each iteration. However, these rules are all designed such that the sampling distribution asymptotically converges to a target distribution. Convergence is achieved either by updating the sampling distribution at each iteration, or using rejection sampling. We use the common Metropolis-Hastings algorithm [49] which relies on rejection sampling. We refer readers to [68] and [65] for details on implementation.

At each iteration k, updated parameters θ^* are proposed according to the following rule:

$$\theta^* = \theta^k + z \tag{3.9}$$

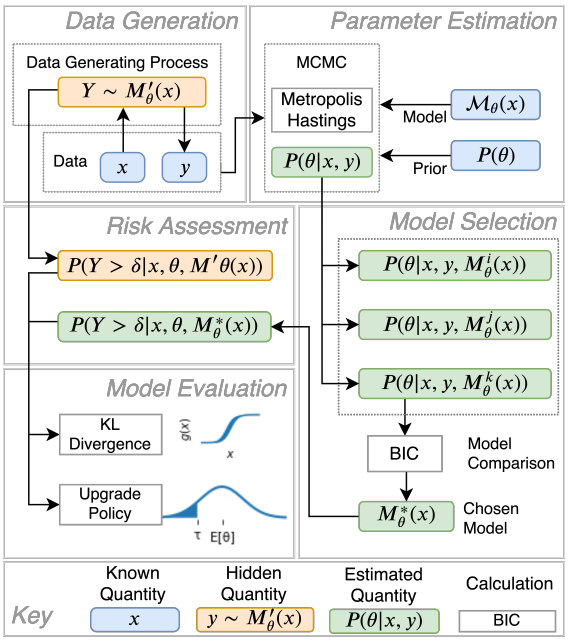

Figure 3.1: Block diagram illustrating analysis procedure.

where z is a random perturbation sampled from a multivariate normal distribution.

As the chain advances, the parameters θ^{k+1} are set equal to θ^* if certain acceptance criteria are satisfied. If the criteria are not met, the new parameters are rejected and the current parameters are kept (i.e., θ^{k+1} is set equal to θ^k). The acceptance rule is given as follows:

$$P(\text{accept}) = \min\left\{1, \frac{P(\theta^*|y)}{P(\theta^k|y)}\right\} \tag{3.10}$$

In words: if the likelihood of the new parameters $P(\theta^*|y)$ exceeds the likelihood of the current parameters $P(\theta^k|y)$, the update is accepted. If the proposal decreases the likelihood function, then the update is accepted with some probability equal to the ratio between the two likelihoods.

We use empirical methods to generate the prior and to determine the correlation structure of z: the so-called "jumping distribution". We do so by computing maximum likelihood parameter estimates $\hat{\theta}$ for bootstrapped sub-samples of the data. We then adjust the covariance of z by some scalar value to achieve an acceptance rate of 0.25 to ensure that the Markov chain exhibits desirable convergence properties [65].

Model Selection

We use the Bayesian Information Criterion (BIC) to weigh different parametric models $\mathcal{M}_\theta(x)$ based on parsimony and fit. Mathematically, the BIC is given by:

$$BIC_{M^i_\theta(x)} = -2\log\left(\ell_{M^i_\theta}\right) + K\log(n) \tag{3.11}$$

Here, $\ell_{M_\theta^i}$ describes the likelihood that a particular model $M_\theta^i(x)$ describes the data, K is the number of features in the model, and n is the number of observations in the data. The likelihood is calculated by taking the total probability of the model across all candidate parameterizations θ' with the probability density function $p(\theta')$, given by

$$\ell_{M_\theta} = \int_{\theta' \in \Theta} P(x, y | M_\theta, \theta') p(\theta') d\theta' \tag{3.12}$$

The BIC describes the information contents of one model compared with another. Thus, we choose the model that minimizes the BIC, and compare the information contents of different models based on the differences between the BIC for one model compared with the model selected. We refer readers to [61] for a comparison of BIC and other metrics for model selection.

Evaluating Model Performance

Here, we describe metrics for evaluating how accurately the model learned from data recovers the true model.

In real-world systems, the true model would be unknown. In simulation, however, we have specified the true model in order to generate the data y. We capitalize on the fact that the failure model is known to evaluate the performance of our Bayesian hierarchical model. We evaluate the performance of our model by examining how accurately it represents the statistical properties of the system (measured using the KL divergence), and in terms of its ability to recommend suitable upgrade policies.

Kullback-Leibler (KL) Divergence

KL divergence is an information theoretic metric describing the statistical distance between two probability distributions P and Q (with the probability density functions p and q respectively), given by:

$$D_{KL}(P||Q) = \int_{y \in \mathcal{Y}} p(y) \log \left(\frac{p(y)}{q(y)} \right) dy \tag{3.13}$$

Here P is the distribution of the probability of observing y failures given the true distribution of θ

$$P(Y) = \int_x \int_\theta P(Y|x, \theta) p(x) p(\theta) \, d\theta \, dx \tag{3.14}$$

while Q is the distribution of the probability of observing y failures given the distribution of $\hat{\theta}$ approximated using MCMC.

Optimal Upgrade Policy

Next, we examine upgrade policies that a decision-maker would arrive at using the proposed model, compared with the true model. The objective of the upgrade policy is to identify the

minimum number of components that must be upgraded to achieve a target distribution of acceptable failure rates. The current work considers replacements that altogether eliminate risk. Extensions on this work could consider reparable components, and upgrade measures that alter (but do not eliminate) the risk profile of the system.

We define the target distribution $\tilde{Y}$ such that the probability that the number of failures exceeds some specified damage threshold δ must be below a specified value ϵ:

$$P\left(\tilde{Y} > \delta\right) \leq \epsilon \tag{3.15}$$

$$P\left(\tilde{Y} > \delta\right) = 1 - P\left(\tilde{Y} \leq \delta\right) = 1 - \sum_{\tilde{y}=0}^{\delta} P(\tilde{Y} = \tilde{y}) \tag{3.16}$$

where $P(\tilde{Y} = \tilde{y})$ is given by (3.6). For example, the policy could be to upgrade the system such that damages exceeding $> 1\%$ of components occur once every 100 years.

The optimal policy finds the minimum number of components M that must be upgraded to achieve the condition (3.15). Upgrade decisions modify the parameters θ_i of individual components, thus changing the distribution of parameters in the overall system $p(\tilde{\theta})$.

In Section 3.3, we introduce the notion that uncertainty in the parameters of the system stems from variability in the parameters of individual components. The implication is that the distribution $p(\theta)$ provides meaningful insight into the failure probabilities of individual components, and that the optimal upgrade policy would target the subset of components that are at highest risk of failure (given x). Though the failure parameters of individual components may not be known, exogenous information about the age or state-of-health of individual components are often available. We assume that decision makers can capitalize on these data to assess the relative performance of one component verses another.

For simplicity, we consider failure probabilities that decrease monotonically with increasing θ. For example, in the case study the probability that a pole fails will decrease as the pole's wind speed rating increases. The parameter θ represents wind speed rating. The optimal upgrade policy for such a measure is depicted graphically in Fig. 3.2. Thus the target distribution $p(\tilde{\theta})$ is simply a truncation of the original distribution $p(\theta)$, where upgrades are targeted to replace components with relatively lower failure thresholds.

The optimization objective is to identify the truncation threshold τ. Mathematically, we can compute τ by numerically solving the following:

$$1 - \sum_{\tilde{y}=0}^{\delta} \int_{x \in \mathcal{X}} \int_{\theta=\tau}^{\infty} P(\tilde{y}|x, \theta)\, p(x)\, p(\theta)\, d\theta\, dx \leq \epsilon \tag{3.17}$$

As presented here, this formulation assumes that the failure probability of upgraded components is zero.

We assume that the parameters of upgraded components are modified such that the failure probability becomes negligible. Once τ is known, the optimal number of components to upgrade can be computed from the integral:

$$M = \left\lceil N \int_0^{\tau} p(\theta)\, d\theta \right\rceil \tag{3.18}$$

where $\lceil \cdot \rceil$ rounds up to the nearest integer.

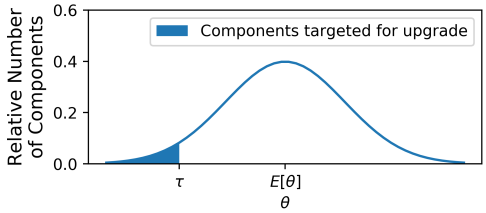

Figure 3.2: Illustrative diagram showing failure parameters for components targeted for upgrades. The number of components is given by (3.17). The threshold τ is obtained by solving (3.15)

.

3.5 Results and Discussion

Here, we present results from a series of tests designed to evaluate how accurately the Bayesian Hierarchical model (BHM) detailed in Sections 3.3 and 3.4 approximates failure processes representative of real power systems.

We train a model using synthetic failure data generated in simulation, where the parameters of the failure model are specified. In the following paragraphs, we outline our procedure for generating the data, and describe the characteristics of the failure processes we examine. We report results of the parameter estimation, model selection, and model evaluation procedures described in Section 3.4. Where relevant, we compare results against the true model used to generate the data, and with a model trained using traditional maximum likelihood estimation (MLE), as described in [42].

Data Generation

We generate synthetic failure data by defining a failure model that relates ambient conditions x to the random variable of the number of failures Y. Data are generated from a model where x is a vector of wind speed measurements. For x we use 10-years of hourly weather data recorded at a site in Northern California to represent stress conditions on the system. The data generation as well as the code used in this paper is open sourced and available at https://github.com/lndunn/gridfix [26].

Given the weather conditions at each time step and the specified failure model $M'_\theta(x)$, we compute the failure probability $g(x)$ using (3.7). From these component-level failure probabilities we generate a realization y of the number of failures in the system overall. The random variable Y follows the Poisson distribution specified in (3.6). The current analysis focuses on learning the underlying parametric model given a single realization of failures over the specified time interval. We note that results are sensitive to the realization that is chosen, and future work could examine a wider range of simulations. That said, the approach taken here is consistent with practical applications, where a historical record provides only one of many realizations that could occur.

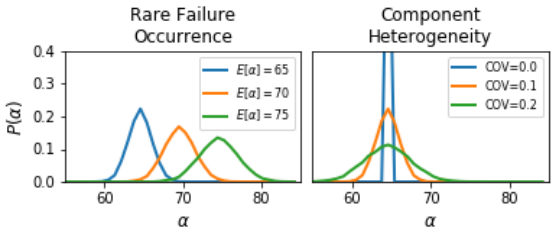

Figure 3.3: Probability density of threshold parameter α for different failure model specifications. Here, θ describes the wind speed (in m/s) at which the failure probability $g(x)$ is 0.5.

Experimental Design

We test our method given failure processes that exhibit a range of statistical properties. First, we examine cases where components are increasingly robust to failure. Because the probability $g(x)$ of failure decreases as ambient conditions x become increasingly severe, robust performance can be achieved by increasing the parameter α in (3.5). Note that failure models are challenging to learn for robust components because as $P(x > \alpha)$ decreases, the probability of observing failure events becomes small. We examine cases where α is normally distributed and is centered around three wind speed thresholds (65, 70, and 75 m/s), as depicted in Fig. 3.3.

Next, we examine cases where failure properties of different components in the system exhibit varying degrees of heterogeneity. We represent heterogeneity by increasing the coefficient of variation (COV) in failure parameters, or the ratio between the standard deviation and the mean. A higher COV indicates more variability in the performance characteristics of individual components.

We execute parameter estimation, model selection, and model evaluation using the methods discussed in Section 3.4 for each of the models specified. We evaluate how well our model compares with the true model. We also compare against a model of the same functional form as the selected model $M^*_\theta(x)$ that uses maximum likelihood estimation to estimate the parameters of the model (rather than using MCMC to estimate the posterior). That is, $M^*_\theta(x)$ estimates scalar values for the parameters as opposed to probability distributions.

Discussion

Here, we summarize results for the failure conditions specified above. We use the BIC to select the functional form of $M^i_\theta(x)$ from candidate models where failures are correlated with wind speed only. In all considered cases, we find that minimizing the BIC results in selecting the correct model.

Once a model is selected, we examine estimates of the threshold parameter α describing the wind speed x at which $g(x)$ equals 0.5 (see (3.5)). The following paragraphs examine the magnitude and distribution of the parameter estimates fitted to failure data generated for

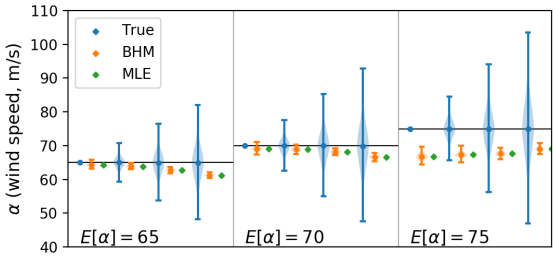

Figure 3.4: True values and BHM & MLE estimates of α. The horizontal lines denote the mean of the true distribution $E[\alpha]$. From left to right, the COV for parameters within each panel are 0, 0.1, 0.2, and 0.3.

	COV=0	COV=0.1	COV=0.2	COV=0.3
$E[\alpha] = 65$	0.201	0.201	0.203	0.206
$E[\alpha] = 70$	0.201	0.201	0.199	0.197
$E[\alpha] = 75$	0.239	0.232	0.220	0.199

Table 3.1: Estimated wind speed coefficient β for BHM and MLE models. In the true model, β is always 0.2.

each of the specified failure models. We also report the KL divergence and upgrade policy, as defined in Section 3.4.

Parameter Estimates

Figure 3.4 shows the range and distribution of α for the true model, and for the BHM and MLE approximations. Both the BHM and MLE models produce mean values within $< 5\%$ of the true value when the true mean $E[\alpha]$ is 65 and 70 m/s, and within $< 10\%$ when the true mean is 75 m/s. However, the mean BHM and MLE estimates are systematically lower than the true mean. This difference increases in cases where the variance is high.

This result suggests that parameter estimates are systematically baised. The reason for this bias is that the probability of observing damaging wind speed conditions is higher when the threshold α is low. It follows that the incidence of failures is relatively higher among components with small α than among components with high α. The implication is that parameter estimates are more heavily influenced by a relatively small subset of components that are more prone to fail than others.

The estimated coefficients β for each case are listed in Table I. In the true model, the coefficients are taken to be constant and are equal to 0.2. BHM estimates report very narrow uncertainty bounds on slope parameter estimates. We note that in cases when the wind speed threshold is high, β tends to decrease as the COV increases. This result has implications on upgrade decisions, as discussed in detail below.

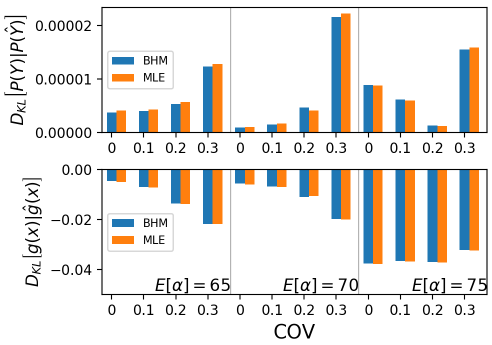

Figure 3.5: Statistical distance between the true distribution $P(Y)$ and estimated distribution $P(\hat{Y})$ of system failures (top), and between the true failure probability $g(x)$ and estimated failure probability $\hat{g}(x)$ of individual components.

KL Divergence

Fig. 3.5 shows the statistical distance between the true and estimated probabilities $P(Y)$ (top) and $g(x)$ (bottom), as given by (3.13). The true distribution is omitted, as the distance is by definition zero. The KL divergence is computed by taking the integral across all parameter values $\theta \in \Theta$ (as indicated in (3.14)), and can be thought of as a metric of error in the underlying probability distributions.

Results show that there is little difference between the BHM and MLE models compared with the true model. When the mean threshold is 65 or 70 m/s, the divergence increases as heterogeneity among parameters increases. This result does not hold when the threshold is 75 m/s, though it is unclear to what extent this result is driven by the particular realization of failures y that occurred.

At the component level, the divergence is consistently negative–indicating that both the BHM and MLE models under-represent component-level failure probabilities. This result is consistent with the previous observation that both models under-estimate the threshold parameter α. When the failure threshold is high, the statistical error in component-level failure models is high, though error in the distribution of failures in the system overall may be low. Future work will explore how sensitive these results are to the particular realization of failures y that occurred.

Upgrade Policy

Finally, we compare differences in upgrade decisions using the true model, and using BHM and MLE parameter estimates. We use the upgrade policy detailed in Section 3.4, where the risk tolerance is defined by:

$$P(Y > 0.1N) \leq 0.05 \tag{3.19}$$

In words, our objective is to limit the probability that $> 10\%$ of components will be damaged (i.e., $\delta = 0.1N$); we aim to limit the probability to below 5% (i.e., $\epsilon = 0.05$). From (3.17),

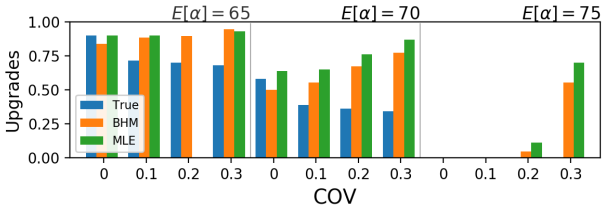

Figure 3.6: Percentage of components recommended for upgrade.

it is clear that upgrades are related not only to the failure model, but also to the ambient conditions x. Here, we take Y to be the total number of failures in a year, and take $P(x)$ to be a log-normal distribution fitted to 10 years of historic data.

Figure 3.6 summarizes the number of upgrades necessary to achieve this risk target for each of the cases we consider. We also show the threshold τ that defines the subset of components targeted for upgrades. Systems where $P(\alpha)$ is centered around 65 or 70 m/s do not meet the specified risk target given the wind speed conditions that are present, and the model recommends substantial upgrades to achieve the specified risk threshold. On the other hand, when $P(\alpha)$ is centered around 75 m/s, the system is overbuilt–and no components are targeted for upgrades.

In general, the BHM formulation outperforms the MLE model in recommending upgrades that are less aggressive and are more closely aligned with the true model. The reason is that by representing variability in the failure parameters of the components, the BHM can more accurately represent the fact that the benefits to the system are higher when upgrade policies target components with the highest failure probability.

Errors in upgrade policies informed by the BHM and MLE models tend to favor more aggressive upgrades than are necessary. This result is consistent with our previous finding that these models systematically underestimate the mean threshold α. What is less intuitive is that the difference between the true and recommended upgrade policies increases as the variance in α grows. To understand this result, we return to Table I. Here, we can see that although β is held constant across all systems in the true model, our estimates of β tend to decrease as the variability in α grows. The consequence of underestimating β is that the estimated failure probability $g(x)$ is larger than the true value when $x < \alpha$. In other words, underestimating β causes us to overstate the risk of failures under mild ambient conditions, resulting in unnecessary upgrades. Nevertheless, this end result is consistent with a mentality of hedging against uncertainty – that is, we more aggressively upgrade components due to the presence of uncertainty.

3.6 Concluding Remarks

This paper outlines a Bayesian hierarchical modeling framework designed to capitalize on grid data of historic failures to characterize failure probabilities, and inform upgrade decisions. We train our model on synthetic failure data generated from specified failure models, and

benchmark our results against both the true model and against a model fitted using the maximum likelihood estimation.

When we assume that components exhibit heterogeneous failure properties, we find that BHM tends to systematically under-represent variability relative to the true model. Furthermore, we find that bias in the failures recorded in the data leads to systematic bias in the parameters of both Bayesian and maximum likelihood approximations.

This finding suggests a critical flaw in current best-practices for evaluating risk in electric power systems. Where risk assessment has historically relied primarily on engineering insights, next-generation models are beginning to leverage data collected in the field. However, our results show that data-driven models must account for sparsity and bias that can arise in rare event data. One opportunity to do so would be to explore sampling techniques–such as stratified MCMC sampling or importance sampling–which can adjust for known bias. The Metropolis Hastings algorithm does not account for this. Another opportunity is to utilize the so-called "value of information" to more precisely determine the component failure properties that do (or do not) pose a risk, or could lead to more informed priors.

Chapter 4

Value of Real-Time Information in Optimal Control

This chapter examines optimal control under uncertainty. The work builds on methods from Bayesian optimization to inform altitude adjustments of an airborne wind energy turbine. Here, we present a risk-seeking formulation where control decisions favor altitudes with high uncertainty, due to the possibility of high reward. More risk adverse variations could be developed to avoid the possibility of adverse outcomes. The model is applied to examine how more timely and comprehensive data could improve performance or mitigate risks. The methods are applied to inform design and control decisions in an airborne wind energy system. Specifically, we compare the effectiveness of candidate sensing and measurement systems that could provide more timely and more accurate information to increase confidence in the outcome of candidate control decisions.

This chapter was originally presented at the American Controls Conference (ACC) and may be cited as follows:

> Laurel N. Dunn, Christopher Vermillion, and Scott J. Moura Tina K. Chow. "On wind speed observability for altitude control of airborne wind energy systems". In: *American Controls Conference* (July 2019)

The original funding acknowledgement is included here as a footnote.[1]

4.1 Introduction and Motivation

Airborne wind energy (AWE) turbines are an emerging wind generation technology. AWE systems differ from conventional turbines in that they are attached to the ground by an adjustable tether, rather than by a fixed tower. This tethered design reduces capital costs and makes it possible to achieve a higher capacity factor. We refer readers to [25] for an in-depth description of airborne wind energy. Information about modern AWE systems can be found at [1, 2, 3].

[1]This work was supported by the National Science Foundation under Award 1437296. The authors would like to thank Cristina Archer at the University of Delaware for providing the wind speed data used in this study.

Three characteristics of AWE systems contribute to a high capacity factor compared with conventional turbines. First, AWE systems can reach higher altitude winds, which tend to be stronger and more consistent than surface-level winds [9]. Second, AWE systems can use crosswind flight patterns to increase the apparent wind speed [46], leading to more power output than would be achieved with stationary flight [33, 56]. Finally, by adjusting the operating altitude in real-time, AWE systems can vertically track favorable wind speeds in a spatially and temporally varying wind field [10]. The current work focuses on understanding how sensor placement impacts the performance of an AWE system with various altitude control schemes.

Currently, research on AWE altitude control systems focus on control schemes that find the optimal operating altitude [11] and are robust to uncertainty [12]. For example, [12] borrow control objectives from Bayesian optimization to balance exploration and exploitation. More specifically, these objectives balance the trade off between capitalizing on known wind resources (exploitation) and collecting observations at altitudes where wind speed estimates are uncertain (exploration), given that wind speed profiles are only partially observable and that altitude adjustments are costly to make. Over time, control decisions favoring exploration can reduce uncertainty in wind speed models, though doing so may come at a cost to near-term performance. We refer readers to [13] for information on Bayesian optimization and a more detailed discussion of the trade-off between exploration and exploitation.

In the context of AWE system altitude control, this trade off is directly related to uncertainty in wind speed estimates that inform control decisions. A survey of state-of-the-art algorithms for forecasting wind power production is given in [69]. Many of the algorithms cited require large training data sets. Training such algorithms may not be possible for AWE systems which typically rely on sparse data streams collected online. Observations are sparse because wind speeds are recorded by a single sensor that tracks (vertically) with the operating altitude of the turbine. A simpler persistence model is also shown to perform quite well, and is recommended as a benchmark for evaluating the performance of more complex wind forecasting algorithms.

Aside from using complex forecasting algorithms, uncertainty in wind speed can be reduced by increasing the spatial coverage of wind speed sensors. For example, if wind speeds were recorded continuously at all altitudes, a simple forecasting model (e.g. the persistence model) may be able to outperform a system that uses a more sophisticated forecasting model trained on single sensor data. This possibility motivates a comparative analysis of different sensor configurations and forecasting methods for AWE altitude control, an issue currently unexplored in the literature.

The main contribution of the current work is to develop a framework for evaluating performance gains achievable using different wind speed sensor configurations. We demonstrate this framework with a case study of a particular wind field. We use the control objectives proposed in [12] to determine the optimal altitude trajectory. Questions related to the performance gains from coupling sensor data with different statistical forecasts are reserved for future research.

This paper is organized as follows. Section 4.2 provides methodological details. Section 4.2 outlines the sensor configurations. Section 4.2 details the forecasting methods. Section 4.2 formulates the altitude control objectives. Section 4.2 describes the comparative analysis benchmarks and metrics. Section 4.3 provides the results and discussion. Finally, Section

4.4 summarizes the paper's main conclusions.

4.2 Methods

In this work we simulate the altitude trajectory of a buoyant airborne turbine (pictured in Fig. 4.1) in a spatially and temporally varying wind field. The simulation relies on wind speed data recorded by a 915-MHz wind profiler between July 1, 2014 and August 31, 2014 at Cape Henlopen State Park in Lewes, DE [9]. Wind speed data are measured every 50 meters in 30 minute intervals. We use the same spatial and temporal discretization in simulation.

We examine single-sensor, multiple-sensor and remote sensor configurations. We track the wind speed measurements that would be recorded online given a particular sensor configuration and the altitude trajectory followed up to a particular point in time. Observations are used to train a persistence model that generates a probabilistic wind speed forecast. We use three different control objectives to determine the optimal altitude trajectory given the current wind speed forecast.

Altitude trajectories are generated for a range of scenarios, each of which uses:

1. One of three control objectives
2. One of three sensor configurations
3. A persistence forecast

The current work examines differences in power production with each sensor configuration. We build on the control objectives presented in [12] to do so, and leave it to future work to explore the performance implications of using more sophisticated wind forecasting methods.

Sensor configurations

Single-sensor configuration

AWE systems are typically designed with a single anemometer to measure the wind speed at the current operating altitude. The vertical position of these measurements changes when altitude adjustments are made.

Multiple sensor configuration

A novel sensor design would be to record wind speed measurements along the length of the tether. The system is still considered partially observable as wind speeds are not measured above the hub height. However, flying the turbine at the highest altitude within operating bounds would provide a complete wind speed profile.

The focus of the current work is to assess the performance implications rather than to explore sensor technologies themselves. However, we have identified several technologies that could collect these measurements. One option is to affix anemometers in regular intervals along the tether, which provides reliable wind speed measurements at altitudes below the AWE system but requires a winch system to house the anemometers upon spool-in. Another

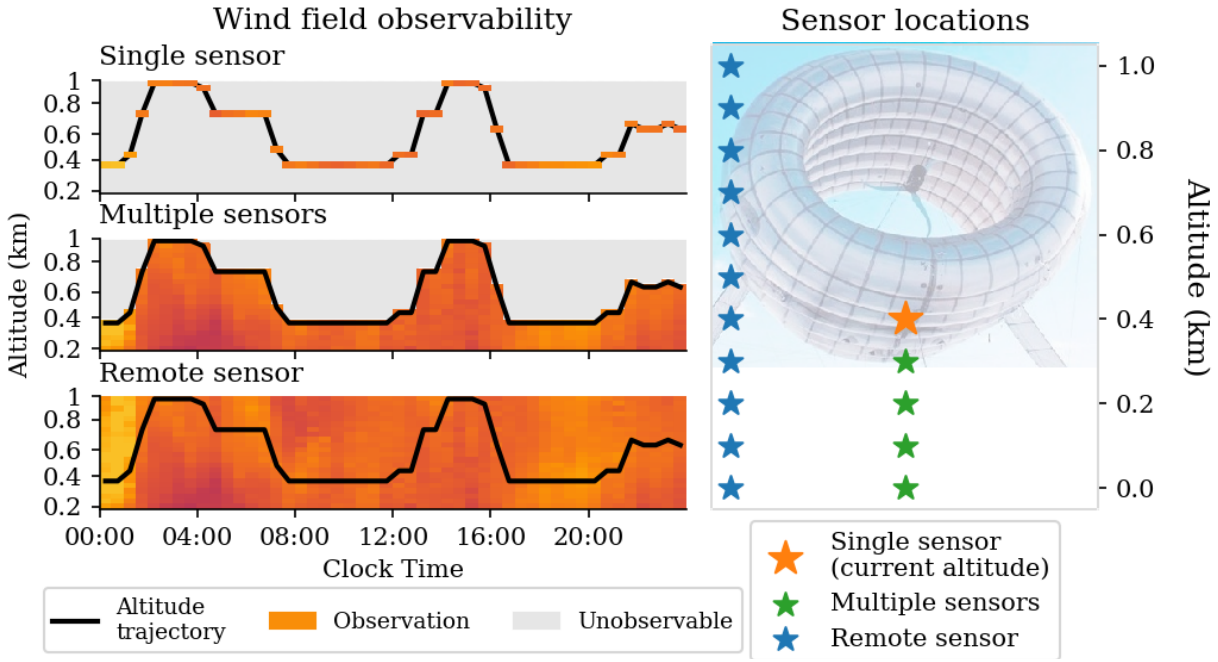

Figure 4.1: Schematic diagram showing wind field measurements recorded by each of three sensor configurations. Right panel shows measurement locations for single, multiple, and remote sensor configurations. Left panel shows the altitude (y-axis) and magnitude (color scale) of the measurements that would be recorded by each sensor configuration given some trajectory (black line) of altitudes with respect to time (x-axis).

alternative is to attach telltales to the tether and use image processing to estimate wind speeds from the angle of the telltale relative to the tether. Finally, one could potentially compute local wind velocities based on the catenary geometry of the tether, though to do so would require a detailed characterization of the tether's structural and aerodynamic properties. Further research is needed to assess the technical and economic feasibility of specific solutions for collecting these measurements.

Remote sensor configuration

The third configuration relies on a remote sensor to record wind speed measurements in discrete intervals across a wide range of altitudes simultaneously. For example, the wind data we use to inform our simulation was collected using a vertical profiler that measures wind speeds in 50 meter increments. With remote sensing wind speed profiles are fully observable at each time step. Because the same measurements are recorded regardless of the current position of the turbine, this configuration decouples data acquisition from altitude control. The implication is that exploring the wind field for data acquisition is not necessary, and that control decisions can focus instead on exploitation of known wind resources.

Persistence Forecast

We use a persistence model adapted from [69] to generate a probabilistic wind speed forecast. We extend the model to extrapolate wind speeds into the spatial domain, and to characterize uncertainty. The persistence model is based on the premise that wind speeds change very little (or not at all) from one time step to the next. The spatial extension of this model presumes that there is little (or no) change vertically either. Based on this premise, the forecast mean μ at altitude h and s time steps into the future is given by:

$$\mu_{h,t+s} = X_t(h') \tag{4.1}$$

where $X_t(h)$ describes the wind speed observation recorded at time t and altitude h. We define h' to be the measurement altitude in X_t that is either equal to the forecast altitude h, or is nearest to it. For example, with multiple sensors, h' would be equal to h for altitudes below the current operating altitude, and would equal the current altitude (i.e., the highest observable altitude) otherwise. The fundamental assumption is that the measurement $X_t(h')$ recorded at (h', t) persists across all unobserved times and altitudes.

We characterize uncertainty by estimating how erroneous the assumption of persistence could be based on past observations. We define $\mathbf{\Delta}$ to be a matrix composed of column vectors $\Delta X/\Delta h$ and $\Delta X/\Delta t$ describing the finite differences between measurements in X in space and time. Based on an exploratory analysis of the data, we characterize $\mathbf{\Delta}$ as a joint Gaussian distribution with zero mean.

Next, we define the vector $\mathbf{d}$ describing the distance in space $(h - h')$ and in time (s) between the current observation $(X_t(h'))$ and the forecast. We can then characterize forecast uncertainty (σ^2) as:

$$\sigma^2_{h,t+s} = \Sigma\left(\mathbf{d}\mathbf{\Delta}^T\right) = \mathbf{d}\Sigma(\mathbf{\Delta})\mathbf{d}^T \tag{4.2}$$

Here $\Sigma(\mathbf{\Delta})$, for example, is the covariance of $\mathbf{\Delta}$.

Where forecast uncertainty is high, we truncate the distribution between 0 and 17 m/s to ensure that high uncertainty does not lead to unrealistic predictions. We find that 98% of all observations in the data are within these bounds. We assume that the underlying distribution of $\mathbf{\Delta}$ is stationary, though characterizing non-stationarity in wind speed dynamics presents an interesting opportunity for future work.

Control Methods

We describe altitude control using the following simple integrator dynamics:

$$h(t+1) = h(t) + u(t) \tag{4.3}$$

where $h(t)$ is the altitude at discrete time index t, and $u(t)$ is the controlled altitude adjustment.

The control objective at a given time t is to select the trajectory of optimal future operating altitudes $\{h^*(t+1), h^*(t+2), \cdots\}$ and controlled altitude adjustments in $\{u^*(t), u^*(t+1), \cdots\}$ that maximize some objective function, given the wind speed forecast V. We express

Table 4.1: List of model parameters and values.

Variable	**Value**
h_{min}	0.15 km
h_{max}	1.0 km
r_{max}	0.01 km/min
v_r	12 m/s
Δt	30 min
T	90 min
k_1	0.0579 kW s^3/m^3
k_2	0.09 kW s^2/m^2
k_3	1.08 kW $s^2/m^2 \cdot$ km

this objective J mathematically as:

$$\max_{h(t),u(t)} J = \sum_{s=t}^{t+T} g(h(s), u(s), V_{\forall h,s}) \tag{4.4}$$

Here $V_{\forall h,s}$ to refers collectively to the random wind speed forecasts for all candidate altitudes h at time s, and T is the planning horizon.

The problem is constrained such that operating altitudes are bounded to within $h_{\min}$ and $h_{\max}$, and the rate of change in altitude is below $r_{\max}$.

$$\begin{aligned} h_{\min} \leq h(t) \leq h_{\max} \\ |u(t)| \leq r_{max} \end{aligned} \tag{4.5}$$

Table I provides numerical values for operating constraints.

This formulation uses model predictive control to optimize the control and state trajectories over the upcoming T time-steps given the current state $h(t)$, dynamical model (4.3) and wind speed forecasts for all altitudes. Only the first control action $u^*(t)$ is physically implemented, and the process is repeated using the measured state in the next time step.

We use a planning horizon of 90 minutes (or three 30-minute time steps). Given the values listed in Table I, the planning horizon is sufficient for the turbine to travel between any two operating altitudes. We use dynamic programming to solve for the optimal trajectory at each time step.

We examine three formulations of $g(h, u, V)$ that aim to maximize power production. These formulations differ in how they account for uncertainty in power production (which stems from uncertainty in wind speed forecasts). Although the objective function differs for each formulation, the long-term goal is always to maximize power production.

Equation (4.6) lists the system of equations used to calculate power production, as described in [11]. Here power production $p(u(t), v)$ is a function of the altitude adjustment during some time interval $u(t)$ and the true wind speed v.

$$\begin{aligned} p(u,v) &= p_1 - p_2 - p_3 \\ p_1 &= k_1 \cdot \min\{v_r, v\}^3 \\ p_2 &= k_2 v^2 \\ p_3 &= k_3 v^2 \cdot |u| \end{aligned} \tag{4.6}$$

In words, the total power production $p(u, v)$ is the difference between the amount of energy the turbine generates (p_1) and the amount of energy required to maintain (p_2) and to adjust (p_3) the operating altitude. The rated wind speed of the turbine is given by v_r, and k_1, k_2 and k_3 describe lumped parameters representing the mechanical and aerodynamic properties of the system. Numeric values for these constants are provided in Table I.

We highlight that p_1 is maximized when v is equal to v_r. However, p_2 and p_3 continue to increase at wind speeds greater than v_r. The implication is that $p(u, v)$ increases as v approaches v_r, but decreases if v increases beyond v_r.

The wind speed v can represent either a measurement reported in the data, or some realization of the wind speed forecast. In order words, if $V_{h,t}$ is a random variable describing the wind forecast at altitude h and time t, then we can use the function $p(\cdot)$ to derive a probabilistic forecast of power production, denoted by $P_{h,t}$. Though $P_{h,t}$ is not explicitly related to h or t, it is implicitly related by the fact that the wind forecast changes with respect to both quantities.

In the following sections we describe three candidate formulations of $g(h, u, V)$. Though the specific objective functions are different, the aim of all three formulations is to maximize overall power production. The objective functions are borrowed from Bayesian optimization, and their application to real-time control of AWE systems is motivated in [12].

Though the formulations we use are conceptually the same as the control objectives presented in [12], we have adapted them in two important ways. First, we optimize over a finite planning horizon extending T time steps into the future. Second, we use a probabilistic wind speed forecast to compute a probability distribution of power production. Since power production is a nonlinear function of wind speed, the power production is not Gaussian, and generally does not follow a parametric distribution. As shown below, our approach handles non-parametric distributions directly, and does not require parametric approximations.

Maximize Expected Energy

The first control strategy chooses the altitude trajectory that maximizes expected power production across time steps within the planning horizon. This can be viewed as an exploitative control approach, in the sense that no reward is explicitly provided to explore portions of the state-space where uncertainty is high. Instead, the goal is to directly maximize expected power production.

In continuous form, the expected power at time t is given by (4.7), where $f_{V_{h,t}}(v)$ is the probability density function (PDF) of wind speed forecast random variable $V_{h,t}$.

$$\mathbb{E}[P_{h,t}|u, V_{h,t}] = \int_0^\infty f_{V_{h,t}}(v) p(u, v) dv \tag{4.7}$$

To accommodate non-parametric wind speed forecasts with no closed form solution to (4.7), we use the discrete approximation given by

$$\begin{aligned} \mathbb{E}[P_{h,t}|u, V_{h,t}] &\approx \frac{1}{n}\sum_{q=1}^{n} p(u, v_{q/n}) \\ v_{q/n} &= Q_{V_{h,t}}(q/n) \\ Q_{V_{h,t}}(q/n) &:= \Pr(V_{h,t} \leq v) = q/n \end{aligned} \tag{4.8}$$

Here $Q_Y(q/n)$ is the inverse cumulative density function (CDF) of some random variable Y evaluated at quantile q for a specified number of quantile bins n. For example, $Q_Y(q/n)$ evaluates to the q^{th} quartile of Y when n is 4, or to the q^{th} percentile when n is 100. We set n equal to 100.

Maximize Upper Confidence Bound (UCB)

The second control strategy chooses the altitude trajectory that maximizes performance under an optimistic realization of the forecast. This is also known as quantile optimization. For example, one might maximize the 90^{th} percentile of $P_{h,t}$. Equation (4.9) expresses this mathematically for an arbitrary probability threshold $\alpha > 0.5$ and usually near 1.

$$g(h, u, V) = Q_{P_{h,t}|u,V_{h,t}}(\alpha|u, v) \tag{4.9}$$

where $Q_{P_{h,t}|u,V_{h,t}}$ describes the inverse conditional CDF of the random variable $P_{h,t}$ conditioned on $u = u$ and $V_{h,t} = v$.

This approach rewards trajectories that explore altitudes where uncertainty may be large yet potentially yields high power production. Tuning the upper confidence bound α adjusts the degree to which exploration is rewarded. The drawback is that when an overly optimistic control objective is used (i.e., if α is close to 1), there is a high risk that the observed power production will be much lower than the value used to inform a control decision.

Algorithms favoring exploration may under-perform relative to purely exploitative methods, except insofar as they reward acquisition of new data that reduces long-run uncertainty in the wind speed forecasts. As uncertainty bounds become narrow, the difference in power production between control objectives favoring exploration and exploitation also decreases.

Maximize Probability of Improvement

The last control strategy selects the altitude trajectory with the highest probability of improving performance relative to maintaining the current altitude. We calculate the probability of improvement by taking the log probability that power production for a particular trajectory will exceed power production if the altitude were to remain fixed at current altitude h. Mathematically, this is expressed in (4.10).

$$g(h, u, V) = \log \Pr\left(P_{h+u,t} > p(0, v)\right) \tag{4.10}$$

where $P_{h+u,t} = p(u, V_{h,t})$ is a random variable describing power generation at altitude $h + u$, and $p(0, v)$ is the current power output (i.e. no altitude adjustment $u = 0$ and assuming wind speed v stays constant).

Performance Metrics and Baseline Scenarios

In the current section we define metrics and benchmarks to evaluate the performance of each sensor configuration and control strategy. We calculate these metrics and benchmarks by simulating the altitude trajectory over the course of three months.

Performance metrics

The most fundamental metric we use to evaluate performance is average power production (in kW). Though power production could be compared against the nameplate capacity (in this case 100 kW), this is not a practical target as it is not physically possible to achieve that level of performance. A more practical target would need to account for the physics of the simulation environment, including variations in wind speed and the energy required to adjust and maintain altitude. The omniscient baseline (described below) provides just such a target. In addition to reporting power production, we express performance as a ratio (between 0 and 1) of the energy harvested in a particular scenario and the energy harvested in the omniscient baseline. We refer to this quantity as the "actualized power ratio" (as in Figs. 4.3 and 4.4).

Performance baselines

Omniscient baseline

The omniscient baseline is obtained by simulating the altitude trajectory the AWE would follow if perfect information were available to inform control decisions. The result provides an upper limit on power production, given the characteristics of the wind field and the specified operating constraints.

Fixed altitude baseline

We also compare our results against baselines where the altitude is fixed for the entire of the simulation at the altitudes that would achieve the highest (h_{best}) and lowest power production (h_{worst}). Though these provide a useful basis for comparison, we note that it is not possible to know h_{best} or h_{worst} without omniscient information about the wind speeds at each altitude.

Unlike the omniscient baseline, fixed altitude trajectories do not bound system performance. Instead, they provide a benchmark against a naïve, but possibly effective, control strategy. A real-time control scheme can under-perform relative to a fixed altitude trajectory if control decisions are made based on sufficiently erroneous wind speed forecasts, or if the power production observed at the new operating altitude does not compensate for the energy expended in making the altitude adjustment.

4.3 Results and Discussion

Here we summarize the performance of an AWE system evaluated in simulation for nine scenarios. What differentiates each scenario is the specific combination of sensor configuration

and control scheme used to inform altitude adjustments. A persistence forecast is trained on observational data collected in simulation, given the sensor configuration and altitude trajectory up to that point. We examine a range of values for α in the upper confidence bound (UCB) control and report results for the value that achieves the highest performance in each sensor configuration. We compare the performance in each scenario against the omniscient and fixed altitude baselines.

Fig. 4.2 shows the altitude trajectory over one week in August for the omniscient baseline and four control schemes in the multiple-sensor scenarios. Commenting on the similarities and differences between trajectories highlights the merits (and pitfalls) of using a particular control scheme with each sensor configuration.

Observation 1: Though altitude trajectories follow very different patterns at times when the wind speed is low (e.g., July 11-13), they all follow a relatively fixed course when wind speeds are high (e.g., July 8-9). The reason for this is that p_1 is constant for wind speeds in excess of v_r, while p_2 and p_3 continue to increase. The incentive to explore is only in place if the potential increase in power production exceeds the cost of making altitude adjustments. When the wind speeds are near or in excess of v_r, a fixed altitude is favored because exploration comes at a relatively high cost without the possibility of increasing power production.

Observation 2: Maximizing the probability of improvement leads to a fixed altitude trajectory in both the single- and multiple-sensor cases. The reason for this is that wind speed forecasts are centered around the current observation. In other words, the forecast estimates that exploring some unobserved altitude is equally likely to reduce performance as to improve performance. Thus the probability of improvement is only 50%, and control decisions favor maintaining a constant altitude to avoid power loss from making altitude adjustments.

Though the objective to maximize expected energy is also centered about the current observation, control decisions in that case are informed not only by the probability but also the magnitude of improvement potential. Since the magnitude of improvement scales with v^3, the distribution of p is skewed to the right and there is some incentive to explore.

Observation 3: In the multi-sensor case, when the objective is to maximize the upper confidence bound (UCB), trajectories tend towards higher altitudes rather than lower altitudes. This happens because the uncertainty is greater at unobserved altitudes above the current hub height than at altitudes where current measurements are available. This high uncertainty creates a strong incentive to explore higher altitudes. However, once the highest altitude is reached, the system becomes completely observable and uncertainty is equal at all altitudes, so exploitation is favored.

At this point the trajectory will tend downwards if the best wind resource is below the uppermost altitude. Decreasing the hub height also makes the system only partially observable, reinstating the reward to explore higher altitudes. This process repeats, causing the oscillations observed in the lowermost panel in Fig. 4.2. These oscillations come at a high energy cost and do not necessarily lead to gains in overall performance. We examine how energy is allocated when α is set to 0.7, and compare it against energy allocation when the optimal value (0.54) is used. Although the higher incentive to explore leads to a 4% increase in power production (p_1), the system expends twice as much energy on altitude adjustments (p_3). The additional energy cost leads to a 3% reduction in overall performance. Fig. 4.3

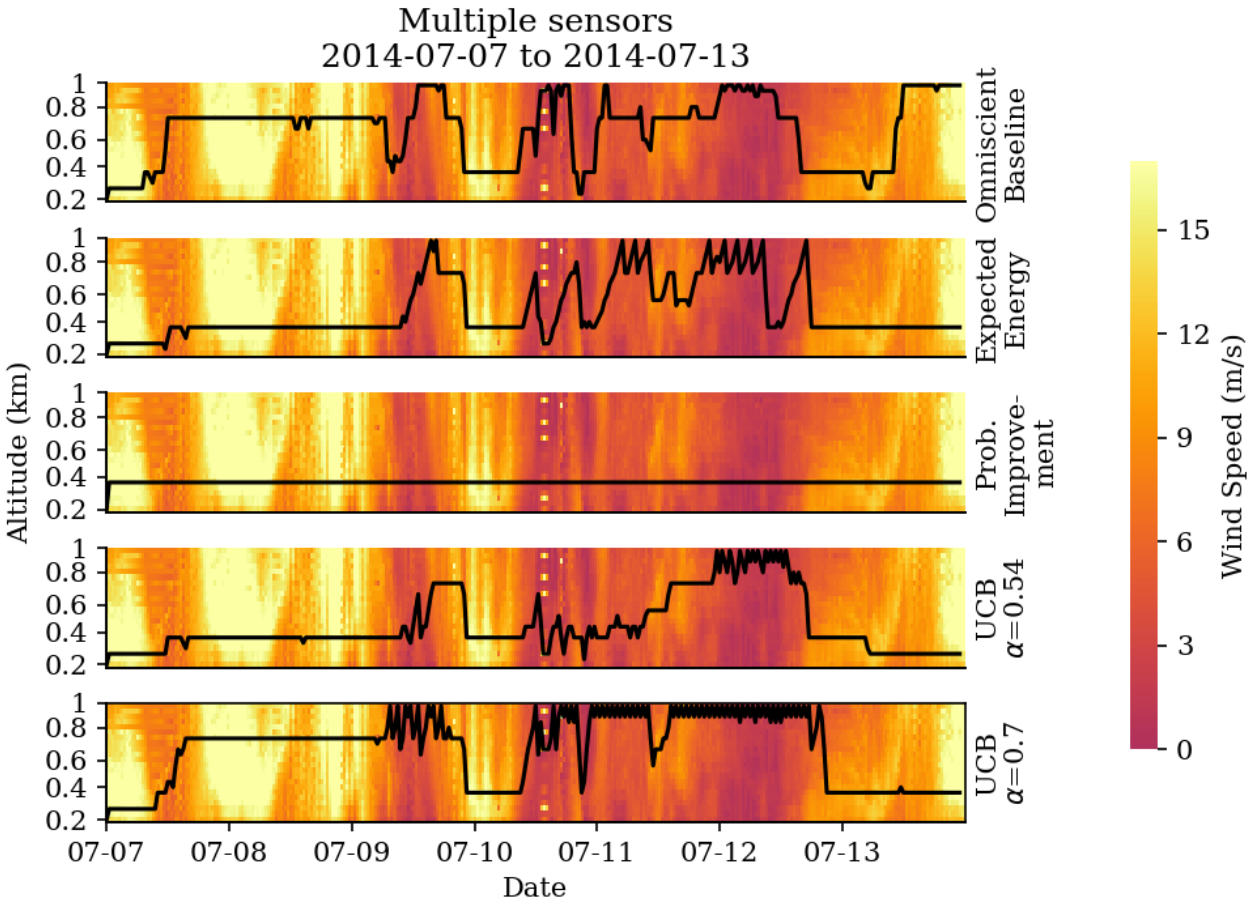

Figure 4.2: Altitude trajectory of AWE turbine (black line) over one week (July 7-13, 2014) for control scenarios indicated on the right. We include UCB control scenarios where $\alpha = 0.54$ (the optimal), and $\alpha = 0.7$ (included for illustrative purposes). The color scale indicates the wind speed at each altitude (y-axis) with respect to time (x-axis).

shows that performance tends to decrease as α increases.

Fig. 4.4 summarizes overall power production across the three-month simulation in all nine scenarios, and compares them against baselines presented in Section 4.2.

Our results show that the remote sensor improves performance by 11% compared with deploying the system with multiple sensors, and by 15% compared with deploying a system with only a single sensor. Fig. 4.3 shows that although performance declines as α increases, the remote sensor consistently outperforms the other two sensor configurations. In all scenarios the multiple and single-sensor cases perform about on par with the optimal altitude (h_{best}) for stationary control. In practice, however, it is not possible to know h_{best} in advance.

These results are based on the highest performing control strategy for each sensor configuration. However, the optimal value for α is not known *a priori*, and likely depends on many factors such as the spacing of sensors and the characteristics of the wind field. Fig. 4.3 shows that the consequences of choosing a sub-optimal control strategy are particularly severe in the multiple sensor configuration.

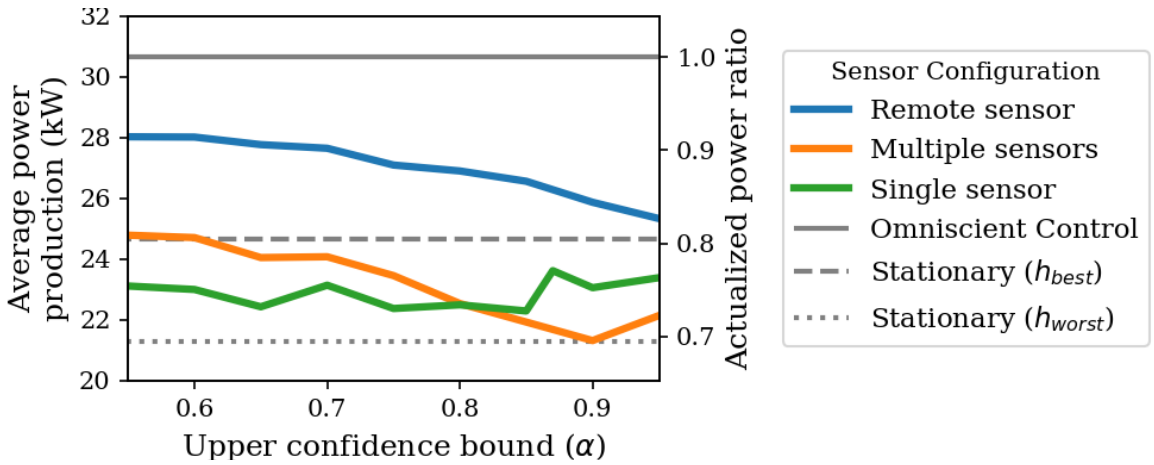

Figure 4.3: Average power production (y-axis) for upper confidence bound control scenarios, as a function of confidence level α (x-axis).

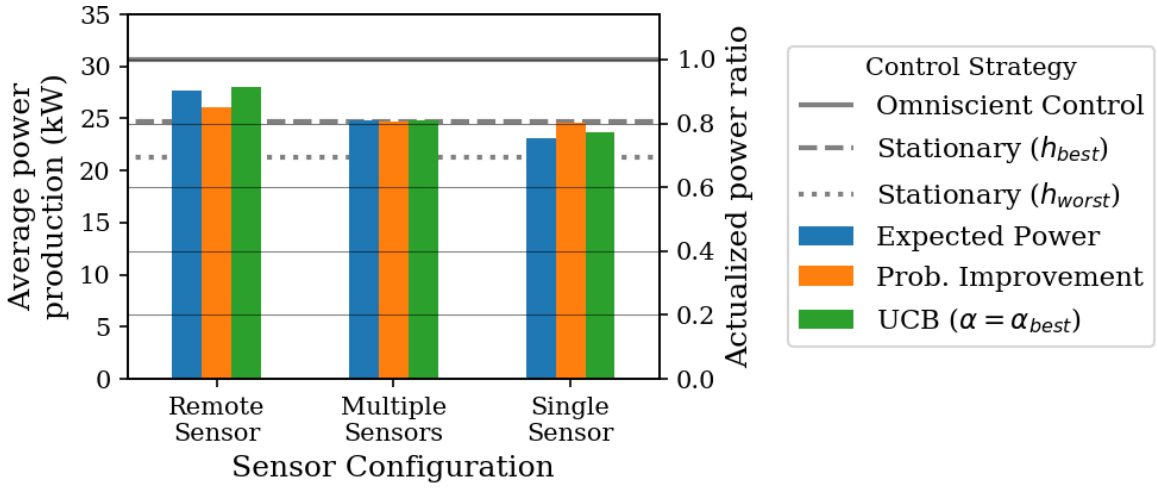

Figure 4.4: Average power production for each control scheme & sensor configuration. Horizontal lines denote omniscient baseline (solid line), and the best (dashed line) and worst (dotted line) fixed altitude trajectories. Performance is measured in terms of average power production (left y-axis) and in terms of the "actualized power ratio" between power production and the omniscient baseline (right y-axis).

Finally, Fig. 4.3 shows that heavily rewarding exploration may slightly improve performance in the single-sensor configuration, but actually decreases performance in the multi-sensor configuration. This decline in performance is due to the altitude oscillations discussed in Observation 3.

4.4 Conclusions and Future Work

In this work we report on differences in AWE performance achieved in simulation using various different control strategies and wind speed sensor configurations. The key difference between the control strategies is how (or if) uncertainty in the forecast is incorporated into the control objective.

We demonstrate that an AWE system with remote sensing equipment can achieve a high level of performance using the most recent measurement to inform control decisions. As the amount of information available to characterize wind speed profiles decreases, forecast uncertainty increases and performance declines.

Both results are related to the quality of the wind speed forecast, raising the question: Can a high-fidelity statistical model improve performance and/or close the gap in performance between different sensor configurations? Our work underscores the need for further research exploring statistical methods for characterizing vertical wind speed profiles.

Chapter 5

Conclusions

5.1 Thesis Review

This thesis explores the merits and limitations of re-purposing data to inform design, operational, and regulatory decisions. Three case studies couple analytical methods with domain expertise to inform data synthesis and interpretation. Each case study highlights instances where data can be misinterpreted to lead decision-makers astray. The core contribution of the work is to describe data-mining techniques that shed light on these limitations, and to offer a quantitative basis for internalizing uncertainty into decision-making processes. In each case study, we explore opportunities for advancements in data collection or monitoring to increase confidence in decision-making outcomes.

The first chapter presents a case study of outage data. The study explores results of an exploratory analysis of high-resolution outage data which we collected. We report summary statistics of minimally processed outage data, and compare them against the information contents of public reporting data. Public data sets report reliability metrics that are heavily aggregated, resulting in substantial information loss. We discuss how public data are used in the literature, and show that analysis of more granular data could lead to different conclusions than are reported to date. We propose new reporting metrics that capture more of the relevant variability, and suggest new models for regulation that could improve equity across customer segments.

The second chapter presents a case study where failure data are applied to inform upgrade decisions targeted at mitigating failure risk. We formulate a Bayesian hierarchical model to probabilistically approximate an unknown failure process. The model is fit to synthetic failure data generated under known stress conditions. We examine failure processes where data are sparse, or where failure properties are heterogeneous across components. Results show that sparse observational data may cause purely data-driven methods to systematically under-estimate failure risks. The study provides strong motivation for data-mining techniques that internalize subject-matter expertise, for example as Bayesian priors.

The third chapter examines optimal control of an airborne wind energy turbine. Conventionally, these systems are designed with a single wind speed sensor co-located with the turbine, providing measurements that track (in real time) with the operating altitude. The implication is that control decisions are informed by uncertain estimates of the wind speed at

candidate operating altitudes. We propose innovations in system design that could improve wind speed observability across a range of operating altitudes, thus increasing confidence in the outcome of control decisions. Design innovations are compared against control strategies that internalize uncertainty to balance the trade-off between exploration and exploitation. Results show that advancements in measurement capabilities lead to greater performance gains than advancements in controls.

5.2 Future Work

Identify socio-economic trends in grid service quality

Chapter 2 shows that different customer segments experience characteristically different standards of reliability. These differences likely arise from the fact that remote infrastructure is more costly and time-consuming to maintain. We also describe how current upgrade policies may favor measures that benefit urban areas that serve more customers. Future work should explore the equity implications of such policies, and what types of investments or policies could help to correct the issue.

Bottom-up economic study of household outage costs

There is a large body of literature using survey data to study the economic impacts of outages. These studies are founded on the premise that customers are able to accurately self-report the costs they incur during an outage. While commercial and industrial customers may know these figures, the costs are not as easy to quantify for residential customers. Costs change depending on the characteristics of the outage (i.e., time of day, time of year, duration, etc.), and on the affected end uses. The proposed study would conduct a bottom-up study to estimate the value of services (e.g., internet, heating, refrigeration, etc.) and the costs incurred when these services are interrupted. The model could be applied to study how changes in customer end uses could change the burden of outages. The results will provide valuable insights to policy-makers and utilities evaluating the need for rate increases to support grid modernization. Results could also help decision-makers to better understand the costs of PSPS events that sacrifice reliability to mitigate other risks.

Online calibration of failure parameters due to aging effects

Engineering models for characterizing failure risk under different stress conditions are typically trained on experimental test data collected in a laboratory setting. The results describe ageing effects of well-defined stress modes for a “typical” component. In practice, manufacturing variability and secondary stress modes (e.g., chemical stress, mechanical stress, animal contact, etc.) cause real-world failure processes to deviate from laboratory conditions. This work would extend the methods proposed in Chapter 3 to study how failure parameters change over time. Results could be applied to improve situational awareness of failure risk, or to generate empirical estimates of end-of-life based on the failures that have been observed.

Techno-economic study of next-generation grid sensors for equipment health monitoring

A wide range of sensor technologies are entering the market to provide real-time situational awareness about the health of distribution assets. Technologies include RF sensors, waveform sensors, and pole-mounted sensors, among others. The sensitivity of these sensors and the quantities that they measure lead to differences in siting and coverage necessary to detect and/or locate issues. A detailed techno-economic study is warranted to examine the capabilities and costs of emerging sensor technologies. The results could inform targeted innovations in hardware or data analysis, and could compare the cost-effectiveness sensor-based solutions against conventional reactive maintenance practices.

Sensor fusion for characterizing local variability in high-altitude winds

Wind turbine controllers operate at time-scales on the order of seconds to maximize performance in the face of near-term changes wind speed and direction. Applications that have traditionally motivated high-altitude wind speed measurement do not require the same granularity that is needed for studying wind turbine controls. Remote sensing data are typically reported in discrete sampling intervals that are too coarse (both spatially and temporally) for high-fidelity simulations examining with power production. This study could would focus on synthesizing remote sensing data with higher-resolution measurements (e.g., from anemometers) recorded at different altitudes. Data sources could include surface sensors, tower sensors, or data collected from weather balloons.

Bibliography

[1] Website. Accessed: Sept 24, 2018. URL: http://www.altaeros.com/.

[2] Website. Accessed: Sept 24, 2018. URL: https://www.skysails.info/.

[3] Website. Accessed: Sept 24, 2018. URL: http://www.ampyxpower.com/.

[4] *2017 Infrastructure Report Card*. 2017.

[5] *2018 Vermont Long-Range Transmission Plan*. June 2018.

[6] Julio Romero Aguero et al. "A Reliability Improvement Roadmap Based on a Predictive Model and Extrapolation Technique". In: *Power Systems Conference and Exposition*. IEEE, 2009, pp. 1–8. DOI: 10.1109/PSCE.2009.4840220.

[7] Peter Alstone et al. *2015 California Demand Response Potential Study: Charting California's Demand Response Future*. Interim Report. California Public Utilities Commission, Apr. 2016. URL: %5Curl%7Bhttp://www.cpuc.ca.gov/General.aspx?id=10622%7D (visited on 05/03/2016).

[8] Massoud Amin and Phillip F. Schewe. "Preventing blackouts: Building a smarter power grid". In: *Scientific American* (2008).

[9] Cristina L. Archer. *Wind profiler at Cape Henlopen*. Website. Accessed: Sept 23, 2018. URL: https://www.ceoe.udel.edu/our-people/profiles/carcher/fsmw.

[10] Cristina L. Archer, Luca Delle Monache, and Daran L. Rife. "Airborne wind energy: Optimal locations and variability". In: *Renewable Energy* 64 (2014), pp. 180–186. ISSN: 0960-1481. DOI: https://doi.org/10.1016/j.renene.2013.10.044. URL: http://www.sciencedirect.com/science/article/pii/S0960148113005752.

[11] A. Bafandeh and C. Vermillion. "Real-time altitude optimization of airborne wind energy systems using Lyapunov-based switched extremum seeking control". In: *2016 American Control Conference (ACC)*. July 2016.

[12] Ali Baheri et al. "Real-time control using Bayesian optimization: A case study in airborne wind energy systems". In: *Control Engineering Practice* 69 (2017), pp. 131–140. ISSN: 0967-0661. DOI: https://doi.org/10.1016/j.conengprac.2017.09.007. URL: http://www.sciencedirect.com/science/article/pii/S0967066117302101.

[13] Eric Brochu, Vlad M. Cora, and Nando de Freitas. "A Tutorial on Bayesian Optimization of Expensive Cost Functions, with Application to Active User Modeling and Hierarchical Reinforcement Learning". In: *CoRR* abs/1012.2599 (2010). arXiv: 1012.2599. URL: http://arxiv.org/abs/1012.2599.

[14] Anna Brockway and Laurel N. Dunn. "Weathering Adaptation: Grid infrastructure planning in a changing climate." In: *Submitted to Climate Risk Management* (Nov. 2019). URL: %5Curl%7Bhttps://arxiv.org/abs/1912.02920%7D.

[15] California Public Utilities Commission. "Unified Resource Adequacy and Integrated Resource Plan Inputs and Assumptions – Guidance for Production Cost Modeling and Network Reliability Studies". In: *California Public Utilities Commission* March (2019), pp. 1–81. URL: http://www.cpuc.ca.gov/uploadedFiles/CPUCWebsite/Content/UtilitiesIndustries/Energy/EnergyPrograms/ElectPowerProcurementGeneration/irp/2018/Unified%7B%5C_%7DRAIRP%7B%5C_%7DIA%7B%5C_%7DFinal%7B%5C_%7D20190329.pdf.

[16] Fredrik Carlsson and Peter Martinsson. "Does it matter when a power outage occurs? — A choice experiment study on the willingness to pay to avoid power outages". In: *Energy Economics* 30.3 (2008), pp. 1232–1245. ISSN: 0140-9883. DOI: https://doi.org/10.1016/j.eneco.2007.04.001. URL: http://www.sciencedirect.com/science/article/pii/S0140988307000618.

[17] Heidemarie C. Caswell et al. "Weather Normalization of Reliability Indices". In: *IEEE Transactions on Power Delivery* 26.2 (Apr. 2011).

[18] Stephanie E. Chang et al. "Infrastructure failure interdependencies in extreme events: power outage consequences in the 1998 Ice Storm". In: *Natural Hazards* 41.2 (2007), pp. 337–358. ISSN: 1573-0840. DOI: 10.1007/s11069-006-9039-4. URL: http://dx.doi.org/10.1007/s11069-006-9039-4.

[19] L. A. Clarfeld et al. "Assessing Risk from Cascading Blackouts Given Correlated Component Failures". In: *2018 Power Systems Computation Conference (PSCC)*. June 2018, pp. 1–7. DOI: 10.23919/PSCC.2018.8442655.

[20] Harrison K. Clark et al. "The Application of Segmentation and Grid Shock Absorber Concept for Reliable Power Grids". In: *12th International Middle-East Power System Conference*. IEEE, 2008, pp. 34–38. DOI: 10.1109/MEPCON.2008.4562303.

[21] Carleton Coffrin, Pascal Van Hentenryck, and Russell Bent. "Strategic Stockpiling of Power System Supplies for Disaster Recovery". In: *Power Engineering Society General Meeting* 836 (2011). ISSN: 978-1-4577-1001-8. DOI: 10.1109/PES.2011.6039414.

[22] IEEE Transformer Standards Committee. "IEEE Guide for Application for Monitoring Equipment to Liquid-Immersed Transformers and Components". In: *IEEE Std C57.143-2012* (Dec. 2012), pp. 1–83. DOI: 10.1109/IEEESTD.2012.6387561.

[23] Yousef Mohammadi Darestani and Abdollah Shafieezadeh. "Multi-dimensional wind fragility functions for wood utility poles". In: *Engineering Structures* 183 (2019), pp. 937–948. ISSN: 0141-0296. DOI: https://doi.org/10.1016/j.engstruct.2019.01.048. URL: http://www.sciencedirect.com/science/article/pii/S0141029618327238.

[24] Paul Denholm and Maureen Hand. "Grid flexibility and storage required to achieve very high penetration of variable renewable electricity". In: *Energy Policy* 39.3 (2011), pp. 1817–1830. DOI: 10.1016/j.enpol.2011.01.019.

[25] Moritz Diehl. "Airborne Wind Energy: Basic Concepts and Physical Foundations". In: *Airborne Wind Energy*. Ed. by Uwe Ahrens, Moritz Diehl, and Roland Schmehl. Berlin, Heidelberg: Springer Berlin Heidelberg, 2013, pp. 3–22. ISBN: 978-3-642-39965-7. DOI: 10.1007/978-3-642-39965-7_1. URL: https://doi.org/10.1007/978-3-642-39965-7%5C_1.

[26] Laurel N. Dunn and Mathilde D. Badoual. *GridFix GitHub Repository*. Apr. 2020. URL: https://github.com/lndunn/gridfix.

[27] Laurel N. Dunn, Christopher Vermillion, and Scott J. Moura Tina K. Chow. "On wind speed observability for altitude control of airborne wind energy systems". In: *American Controls Conference* (July 2019).

[28] Laurel N. Dunn et al. "Bayesian hierarchical methods for modeling electrical grid component failures". In: *Electric Power Systems Research* (July 2020). URL: %5Curl%7Bhttps://arxiv.org/abs/2001.07597%7D.

[29] Laurel N. Dunn et al. "Exploratory analysis of high-resolution power interruption data reveals spatial and temporal heterogeneity in electric grid reliability". In: *Energy Policy* 129 (2019), pp. 206–214.

[30] *Electric Power Annual 2001*. Tech. rep. U.S. Energy Information Administration, Mar. 2003.

[31] *Electric Power Annual 2012*. Tech. rep. U.S. Energy Information Administration, Dec. 2013.

[32] Institute of Electrical and Electronics Engineers. *IEEE Guide for the Statistical Analysis of Thermal Life Test Data*.

[33] L. Fagiano et al. "Automatic Crosswind Flight of Tethered Wings for Airborne Wind Energy: Modeling, Control Design, and Experimental Results". In: *IEEE Transactions on Control Systems Technology* 22.4 (July 2014), pp. 1433–1447.

[34] S. D. Guikema, R. A. Davidson, and Haibin Liu. "Statistical models of the effects of tree trimming on power system outages". In: *IEEE Transactions on Power Delivery* 21.3 (July 2006), pp. 1549–1557. ISSN: 0885-8977. DOI: 10.1109/TPWRD.2005.860238.

[35] Paul Hines, Jay Apt, and Sarosh Talukdar. "Large blackouts in North America: Historical trends and policy implications". In: *Energy Policy* 37.12 (2009), pp. 5249–5259. ISSN: 0301-4215. DOI: http://dx.doi.org/10.1016/j.enpol.2009.07.049.

[36] "How PG&E Ignored Fire Risks in Favor of Profits". In: *New York Times* (Mar. 2019).

[37] *IEEE Guide for Electric Power Distribution Reliability Indices: IEEE Standard 1366-2012*. Tech. rep. IEEE Power & Energy Society, May 2012.

[38] Chuanyi Ji et al. "Large-scale data analysis of power grid resilience across multiple US service regions". In: *Nature Energy* 16052 (2016). DOI: 10.1038/NENERGY.2016.52.

[39] P. Kankanala, S. Das, and A. Pahwa. "AdaBoost$^+$: An Ensemble Learning Approach for Estimating Weather-Related Outages in Distribution Systems". In: *IEEE Transactions on Power Systems* 29.1 (Jan. 2014), pp. 359–367. ISSN: 0885-8950. DOI: 10.1109/TPWRS.2013.2281137.

[40] Kristina Hamachi LaCommare and Joseph H. Eto. *Cost of Power Interruptions to Electricity Consumers in the United States U.S.* Technical Report LBNL-58164. LBNL, Feb. 2006.

[41] Kristina Hamachi LaCommare and Joseph H. Eto. *Understanding the cost of power interruptions to U.S. electricity customers*. Technical Report LBNL-55718. LBNL, Sept. 2004.

[42] David Lallemant, Anne Kiremidjian, and Henry Burton. "Statistical procedures for developing earthquake damage fragility curves". In: *Earthquake Engineering & Structural Dynamics* 44.9 (July 2015), pp. 1373–1389. ISSN: 1096-9845. DOI: 10.1002/eqe.2522. URL: https://doi.org/10.1002/eqe.2522.

[43] Peter H. Larsen. "A method to estimate the costs and benefits of undergrounding electricity transmission and distribution lines". In: *Energy Economics* 60 (2016), pp. 47–61. ISSN: 0140-9883. DOI: http://dx.doi.org/10.1016/j.eneco.2016.09.011.

[44] Peter H. Larsen et al. *Assessing changes in the reliability of the U.S. electric power system*. Technical Report LBNL-188741. LBNL, Aug. 2015.

[45] Peter H. Larsen et al. *Projecting Future Costs to U.S. Electric Utility Customers from Power Interruptions*. Tech. rep. LBNL-1007027. LBNL, Jan. 2017.

[46] Miles L. Loyd. "Crosswind Kite Power". In: *J. Energy* 2.80-4075 (3 1980).

[47] Paul J Maliszewski and Charles Perrings. "Factors in the resilience of electrical power distribution infrastructures". In: *Applied Geography* 32.2 (2012), pp. 668–679. DOI: 10.1016/j.apgeog.2011.08.001.

[48] C. Branucci Martinez-Anido et al. "European power grid reliability indicators, what do they really tell?" In: *Electric Power Systems Research* 90 (2012), pp. 79–84. DOI: 10.1016/j.epsr.2012.04.007.

[49] Nicholas Metropolis et al. "Equation of State Calculations by Fast Computing Machines". In: *The Journal of Chemical Physics* 21.6 (1953), pp. 1087–1092. DOI: 10.1063/1.1699114. URL: https://doi.org/10.1063/1.1699114.

[50] Kirsty Murray and Keith Bell. "Wind related faults on the GB transmission network". In: *2014 International Conference on Probabilistic Methods Applied to Power Systems, PMAPS 2014 - Conference Proceedings* (Nov. 2014). DOI: 10.1109/PMAPS.2014.6960641.

[51] H. Nagarajan et al. "Optimal Resilient transmission Grid Design". In: *2016 Power Systems Computation Conference (PSCC)*. June 2016, pp. 1–7. DOI: 10.1109/PSCC.2016.7540988.

[52] Harsha Nagarajan et al. "Optimal Resilient transmission Grid Design". In: *19th Power Systems Computation Conference, PSCC 2016* (2016). DOI: 10.1109/PSCC.2016.7540988.

[53] Franklin R. Nash. *Reliability Assessments: Concepts, Models and Case Studies*. CRC Press Taylor and Francis Group, 2016.

[54] R. Nateghi. "Multi-Dimensional Infrastructure Resilience Modeling: An Application to Hurricane-Prone Electric Power Distribution Systems". In: *IEEE Access* 6 (2018), pp. 13478–13489. ISSN: 2169-3536. DOI: 10.1109/ACCESS.2018.2792680.

[55] Roshanak Nateghi, Seth D. Guikema, and Steven M Quiring. "Forecasting hurricane-induced power outage durations". In: *Natural Hazards* 74.3 (2014), pp. 1795–1811. DOI: 10.1007/s11069-014-1270-9.

[56] P. Nikpoorparizi, N. Deodhar, and C. Vermillion. "Modeling, Control Design, and Combined Plant/Controller Optimization for an Energy-Harvesting Tethered Wing". In: *IEEE Transactions on Control Systems Technology* 26.4 (July 2018), pp. 1157–1169.

[57] Office of Cybersecurity, Energy Security, and Emergency Response. *Electric Disturbance Events (OE-417)*. Tech. rep. Department of Energy, Jan. 2012. URL: https://www.oe.netl.doe.gov/oe417.aspx.

[58] M. Panteli et al. "Power System Resilience to Extreme Weather: Fragility Modeling, Probabilistic Impact Assessment, and Adaptation Measures". In: *IEEE Transactions on Power Systems* 32.5 (Aug. 2017), pp. 3747–3757. ISSN: 0885-8950. DOI: 10.1109/TPWRS.2016.2641463.

[59] Jaewook Park, Nobuoto Nojima, and Dorothy A. Reed. "Nisqually Earthquake Electric Utility Analysis". In: *Earthquake Spectra* 22.2 (2006), pp. 491–509. URL: %5Curl%7Bhttp://grouper.ieee.org/groups/td/dist/sd/doc/2005-03-Review-of-Undergrounding.pdf%7D.

[60] "PG&E Knew for Years Its Lines Could Spark Wildfires, and Didn't Fix Them". In: *The Wall Street Journal* (July 2019).

[61] David Posada and Thomas R. Buckley. "Model Selection and Model Averaging in Phylogenetics: Advantages of Akaike Information Criterion and Bayesian Approaches Over Likelihood Ratio Tests". In: *Systematic Biology* 53.5 (Oct. 2004), pp. 793–808. ISSN: 1063-5157. DOI: 10.1080/10635150490522304. eprint: http://oup.prod.sis.lan/sysbio/article-pdf/53/5/793/24197659/53-5-793.pdf. URL: https://doi.org/10.1080/10635150490522304.

[62] Prem K. Rachakonda, Balasubramanian Muralikrishnan, and Daniel S. Sawyer. "Roles and Responsibilities of NIST in the Development of Documentary Standards". In: *Proceedings of the CMSC*. National Institute of Standards and Technology. June 2018.

[63] D.A. Reed et al. "Multi-hazard system-level logit fragility functions". In: *Engineering Structures* 122 (2016). ISSN: 0141-0296. DOI: https://doi.org/10.1016/j.engstruct.2016.05.006.

[64] Hui Ren, Ian Dobson, and Benjamin A Carreras. "Long-term effect of the n-1 criterion on cascading line outages in an evolving power transmission grid". In: *IEEE Transactions on Power Systems* 23.3 (2008), pp. 1217–1225. DOI: 10.1109/TPWRS.2008.926417.

[65] Gareth O. Roberts and Jeffrey S. Rosenthal. "Optimal Scaling for Various Metropolis-Hastings Algorithms". In: *Statistical Science* 16.4 (2001), pp. 351–367. ISSN: 08834237. URL: http://www.jstor.org/stable/3182776.

[66] Cynthia Rudin et al. "A process for predicting manhole events in Manhattan". In: *Machine Learning* 80 (2010), pp. 1–31. DOI: 10.1007/s10994-009-5166-y.

[67] Abdollah Shafieezadeh et al. "Fragility Assessment of Wood Poles in Power Distribution Networks against Extreme Wind Hazards". In: *IEEE Transactions on Power Delivery* 29.1 (2014), pp. 131–139. URL: https://ieeexplore.ieee.org/document/6616005.

[68] A. F. M. Smith and G. O. Roberts. "Bayesian Computation Via the Gibbs Sampler and Related Markov Chain Monte Carlo Methods". In: *Journal of the Royal Statistical Society. Series B (Methodological)* 55.1 (1993), pp. 3–23. ISSN: 00359246. URL: http://www.jstor.org/stable/2346063.

[69] S. S. Soman et al. "A review of wind power and wind speed forecasting methods with different time horizons". In: *North American Power Symposium 2010*. Sept. 2010, pp. 1–8. DOI: 10.1109/NAPS.2010.5619586.

[70] Andrea Staid et al. "Simulation of tropical cyclone impacts to the U.S power system under climate change scenarios". In: *Climatic Change* 127 (2014), pp. 535–546. DOI: 10.1007/s10584-014-1272-3.

[71] Micahel J. Sullivan, Josh Schellenberg, and Marshall Blundell. *Updated value of service reliability estimates for electric utility customers in the United States*. Technical Report LBNL-6941E. LBNL, Jan. 2015.

[72] U.S. Energy Information Administration. *Electric power sales, revenue, and energy efficiency Form EIA-861 detailed data files*. Oct. 2016.

[73] Vaiman et al. "Risk Assessment of Cascading Outages: Methodologies and Challenges". In: *IEEE Transactions on Power Systems* 27.2 (2012), pp. 631–641.