Simulating Long-Term and Short-Term Community and Infrastructure Vulnerability and Response to Natural Hazards

by

Chengwei Zhai

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Industrial and Operations Engineering)
in the University of Michigan
2021

Doctoral Committee:

Professor Seth D. Guikema, Chair
Professor Mark S. Daskin
Professor Ian A. Hiskens
Assistant Professor Allison C. Reilly, University of Maryland

Chengwei Zhai

cwzhai@umich.edu

ORCID iD： 0000-0002-8508-9838

Dedication

To my beloved soul mate Tianhui Maria Ma

and

my wonderful baby Lucas Qianyu Zhai

Acknowledgement

Before the final summary of my PhD studies, please allow me time to pen down my gratitude to some of the most beloved people who have offered great support.

First I would like to thank my advisor, Seth Guikema. It has been my best of luck to be part of Seth's group. I'm thankful for every opportunity to learn from him and to work with him. Seth is not only my research advisor, but also my life mentor and a great friend. He has mentored me to become an independent researcher. He introduced me to both worlds of academia and industry. In the meantime, he has offered so much support to my life and family. No words can express my gratitude to Seth enough.

I met Allison Reilly in the first research class I took upon joining University of Michigan. She's the person who strengthened my determination to become a successful PhD. I have obtained my confidence from her acknowledgements. She has offered countless great advice to my research, writing and other professional skills. I deeply appreciate her guidance and wish for further collaborations in the future.

I would like to thank all my committee members, collaborators, and group mates. Mark Daskin and Ian Hiskens, thank you for being in my committee and providing suggestions to improve my dissertation. Thank you to all of my collaborators in different projects, especially Brooke Anderson and Steven Quiring. I have learned so much from you and really enjoyed working with you. And to my group mates, Thomas Chen, Elnaz Kabir, Tim Williams, Anna

White, Valerie Washington, Julia Coxen, Tom Logan and everyone, it has been a great pleasure to work together and learn from you all.

Finally, I would like to thank my family. To my parents, thank you for loving me, caring for me, and supporting me unconditionally. To Maria, my soul mate and partner of life, thank you for sitting next to me on the flight ten years ago, thank you for loving me, always being with me and making lifetime memories together with me. To our son, Lucas (Cashew), who is 2 year and 3 month old for the time being, you have been such a miracle. I love you.

Table of Contents

List of Figures

`

List of Tables

List of Appendices

`

Abstract

Natural disasters can cause severe damage to infrastructure, the economy, and human lives. A better understanding of the vulnerability of individual households, critical infrastructure, and a community can be vital to in both short-term disaster response and long-term resilience planning. This dissertation develops methods to analyze the long-term vulnerability of communities when facing repeated hurricanes and heat waves in a changing climate, as well as the short-term vulnerability of power systems under different types of disasters. The approaches I develop are a combination of simulation modeling, predictive modeling, and network theory.

The first chapter of this dissertation describes the vulnerability of communities when facing different hazardous events, such as hurricanes, heat waves, and power outages. The second chapter focuses on enhancing understanding of the long-term vulnerability of a community under repeated hurricanes. I discuss how learning, initial beliefs and memory decay effects influence individual decisions and change regional vulnerability under different hurricane climate scenarios. We found how different initial knowledge and the memory effect can result in different community vulnerability under different hurricane climate scenarios. The third chapter develops methods for estimating power system damage and power outages from extreme weather events. I use publicly available data to generate the layout of the distribution system which is not publicly available in most cases. I then use the synthetic distribution layout to simulate damage and power outages from weather events. This model can provide important information to

`

understand regional and individual power outage risks. The fourth chapter studies how the long-term vulnerability of a community may change under repeated heat waves. I build an agent-based model to address the interplay of the vulnerability of community, climate change, individual mitigation, social networks, power outages, and government mitigation. The model shows how each component is triggering the evolution of community heat vulnerability with historical events and what-if scenarios. The fifth chapter summarizes this dissertation and discusses potential limitations and future directions of this study.

Overall, this work develops new methods to study the vulnerability of communities under repeat natural disasters aiming to enhance better short-term response and long-term planning. These models can help decision-makers, policy-makers, and individuals to make better plans when facing unexpected extreme weather events.

Chapter 1: Introduction

Repeated natural disasters cause severe damage to human lives, infrastructure, and the economy. To understand how to mitigate effectively, we must understand the vulnerability of the community from different perspectives including individual, infrastructure, and the government. This work divides vulnerability into long-term and short-term and studies both of them. Long-term vulnerability is a measure of how well or bad a community responds to natural disaster in a dynamic scenario, where the people, climate, land use, etc. can be changing over time. Short-term vulnerability describes the risks to a community of being damaged from a specific event, which can be reflected from mortality, building or infrastructure damage, or loss of life essentials such as electricity power and drinking water. I build simulation models with data analytics approaches to address critical questions that enhance our understanding of both the short-term and long-term vulnerability.

1.1 Hurricanes

Hurricanes cause considerable damage to human lives, properties, infrastructure, and the economy. Hurricane Irma in 2017 caused approximately $42.5 to $65 billion (USD) in total property damage according to some estimates[1]. Hurricane Sandy in 2012 led to 8.2 million customers losing power. Hurricane Maria caused more than 4000 deaths. People and infrastructure are vulnerable to hurricanes, and mitigation needs to be implemented to decrease

[1] https://www.reuters.com/article/us-hurricane-irma-corelogic-idUSKCN1BU28T

`

the vulnerability in the long run[1]. However, to comprehensively understand how to decrease the long-term hurricane vulnerability, further studies into the mitigation efforts made by individual households and the government under a changing climate is crucial.

Household-level mitigation has been studied from multiple perspectives, including what mitigation actions a household can take [2], cost-effectiveness analysis of different measures to increase hurricane resistance [3], and the demographics, wealth, hurricane experiences of the homeowners and the decisions they made [4,5]. Government mitigation efforts, on the other hand, have also been well studied from the measures (incentives, policies) to their effectiveness [6]. However, most of these studies stem from a specific event or from a specific region, which decrease the generalizability of the results from these studies.

Another challenge comes from climate change, which could be game-changing to all existing policies and preparations. However, the relationship between climate change and hurricane occurrences is not fully understood, which increases the difficulty in predicting, planning, and adapting especially when rare events happen, e.g. hurricane Katerina, 2005 [7]. To fill this gap in long-term hurricane vulnerability studies, we must first address the effect of individual household learning, which changes the mitigation decisions one makes. This in turn changes the regional vulnerability overall. Learning is the most important process that enables the adaptation to an unknown environment [8].

To understand how the effect of learning changes the vulnerability of communities under repeated hurricanes, I built an agent-based model (ABM) that focuses on the influence of individual learning effects on homeowner decisions, which thus affect regional vulnerability to the residential building stock. The resulting model can not only answer our questions about the effect of individual learning, climate change, initial beliefs held by the household, and memory

`

decaying to the change of community vulnerability, but it can also serve as a platform for analyzing how hazard experiences and government policies influence one's understanding of the risk and how this causes communities to change through mitigation and possibly even retreat.

1.2 Heat waves

Heat waves are another type of natural disaster that leads to deaths and in some cases damages infrastructure. Unlike hurricanes, the effects heat waves are felt primarily in specific population groups such as older individuals with health conditions, and those who are socially isolated. There is little research on how long-term heat wave vulnerability changes in a community. Most existing research focuses on a specific event [9,10], or a certain impact . To understand the vulnerability of a community facing heat waves, I must understand the three sources that can change vulnerability over time: individual mitigation behaviors, social networks, and government interventions. For individual households, the most effective mitigation for heat waves is to invest an Air Conditioning unit (A/C). However, this can be prohibitively costly to families with limited budgets. People can also mitigate through their social networks by getting invited to the house of a friend with A/C to avoid heat exposure. However, the effectiveness of mitigation from social networks has not been well-understood. The government, in the meantime, can open cooling centers to host people in need. To understand the impacts of three facets of heat wave mitigation, we need a platform that can systematically model these elements while remaining grounded in reality.

To answer these questions, I build an agent-based model to help understand the long-term vulnerability from heat waves. I model a heat wave environment, where individuals can learn from heat wave events, make mitigation decisions, get help from their social connections, and get help from the government. I use the number of estimated heat-related deaths and the

`

coverage ratio of A/C to quantify the vulnerability of the region. Through this model, I can run different scenarios to test how each component in the process would change the evolution of the region's heat vulnerability.

1.3 Power outages

Most power outages are caused by weather-related events, and power outages cause inconvenience, economic losses, and loss of human lives [(11)]. They can be triggered by both hurricanes and heat waves among many other types of weather events and can be impactful to household mitigation behaviors [(12)]. A good estimation of the severity and spatial distribution of power outages ahead of a forecast event can speed up the response from utility companies and reduce costs when planning personnel and restoration materials. Detailed power outage likelihood forecasts can also help homeowners to better understand and improve their local resilience through considerations of back-up power, power disruption insurance (for a commercial entity), and other measures. Two different routes are taken when studying power outage predictions: statistical-based approaches and fragility-based approaches.

Statistical-based approaches has been used in practice and studied by many [(13–16)]. These approaches rely on historical power outage records with covariates such as hazard information, weather, land use, land topology, and tree cover/species, to build statistical learning models to make predictions for future events [(16)]. In practice, these models require the collection of all the covariates in real time, and the covariates and predictions are restricted to the same spatial scale defined by the dataset, e.g., county level or grid cell level. These limitations pose challenges in using validating models for real-time predictions.

Fragility-based model uses fragility curves for critical power infrastructure assets to simulate the failure of infrastructures and the accessibility of customers to power feeders. The primary

`

limitation of this approach is that the deployment of infrastructure that are susceptible to damages, e.g. power lines, power poles, and the way they are connected from customers to substations are not publicly available due to security concerns.

This work develops a new method that fills the gap of the unknown distribution grid by synthetically generating it with publicly available information. I validate the generated network with an actual distribution grid and use the network to simulate two historical power outage events.

1.4 Bottom-up approaches

Bottom-up and Top-down approaches are two different frameworks widely used in research studies. Experts from different fields might have different definitions of these two approaches. In general, top-down approach models start from an overview of the major question, breaking this down to smaller questions, and explore answers to each of those until gaining an understanding of the system. Typical top-down approaches used in risk analysis are predictive modeling, system dynamics models, and related approaches. Bottom-up approaches, on the other hand, start from modeling each component of the system to create interactions and rules until reaching to an acceptable level of completeness to describe, reflect, or represent the complex system. Typical bottom-up approaches used in risk analysis are agent-based modeling and component-based fragility assessment. There are several benefits of bottom-up approaches. First, bottom-up approaches start from each component of the system. One can easily model the interaction of those components and observe the changes due to those interactions. This is a major benefit of bottom-up approaches in comparison to top-down models, where the latter are often used like a black box, such as neural network model. The second advantage is the bottom-up model is easy

`

to communicate with other researchers or the public, as each component is normally modeled in a simplified way to represent the reality. The third advantage is that bottom-up approaches can be flexible to incorporate new modules. It is a very convenient boundary object in interdisciplinary studies and can be used to answer what-if questions with introducing additional components that may or may not change the output.

In this work, I am modeling the complex system with a bottom-up approach starting from the individual level for both community vulnerability studies and the power outage study. This is a novel contribution to these fields as in the past, the importance of individuals are not well-studied, or individuals are not getting informed as an important responder to the potential risks they might be facing. In community vulnerability studies, each single family or individual is the agent in our agent-based model, and I focus on how they are learning, adapting, and making mitigation decisions can potentially decrease the long-term vulnerability of the community. In the power outage studies, I use a simulation-based method to estimate the risk of each individual building losing power, which provides a metric to support individual homeowners and business owners when making decisions on back up power to protect normal lives and business continuity. I use bottom-up approaches to create models from building level, therefore I can get the flexibility to aggregate the output to whatever resolution I want.

1.5 Overview

This dissertation consists of three independent projects that enhance the understanding of long-term vulnerability of communities under repeated natural disaster events such as hurricanes and heat waves and develop innovative model to estimate short-term vulnerability due to disaster-induced power outages. The main contribution of each project is as follows.

`

Chapter 2: Individual Learning as a Driver of the Evolution of Community Vulnerability under Repeated Hurricane Hazards and Climate Change - I create an agent-based model in nine counties in Maryland that can be used to study the long-term vulnerability of hurricanes for different hurricane climate scenarios. I reveal the impact of learning, initial beliefs, and memory decaying effect on the evolution of community vulnerability. I find that the initial beliefs held by homeowners can have long-term impact on the vulnerability of a community especially when the hurricanes become more frequent and intense. And because of the nature of human memory, post-disaster is the best time to encourage mitigation and decrease the vulnerability.

Chapter 3: Evolution of vulnerability under repeated heat waves - I built an agent-based model that integrates a synthetic heatwave model, dynamically evolving agents, social connections, and a health mortality model to study the impact of repeated heatwaves on the vulnerability of the community. I calibrate our model with available historical data and study multiple scenarios to reveal how individual decisions, social network, and government policies can increase or decrease mitigate heat wave damages. I find that social invitations and cooling centers are very critical mitigation in reducing mortality. Without social invitations, though the A/C penetration ratio becomes the highest, the overall mortality is the highest.

Chapter 4: Power Outage Prediction for Natural Hazards Using Synthetic Power Distribution Systems - I develop an innovative method to generate synthetic distribution grid and how to use these to simulate power outages during hazards. I validate our network with Columbus, OH’s actual distribution system. I identify the model with the best performance for network generation and showed the performance of the model in power outage prediction for the Derecho in 2012 in Ohio and Hurricane Harvey in 2017 in Texas

`

1.6 References

1. Reilly AC, Guikema SD, Zhu L, Igusa T. Evolution of vulnerability of communities facing repeated hazards. Linkov I (ed). PLoS ONE, 2017; 12(9):e0182719.

2. Unanwa CO, McDonald JR. Building wind damage prediction and mitigation using damage bands. Natural Hazards Review, 2000; 1(4):197–203.

3. Orooji F, Friedland CJ. Cost-Benefit Framework to Generate Wind Hazard Mitigation Recommendations for Homeowners. Journal of Architectural Engineering, 2017; 23(4):1–11.

4. Peacock WG. Hurricane Mitigation Status and Factors Influencing Mitigation Status among Florida's Single-Family Homeowners. Natural Hazards Review, 2003; 4(3):149–158.

5. Sadowski NC, Sutter D. Mitigation motivated by past experience: Prior hurricanes and damages. Ocean and Coastal Management, 2008; 51(4):303–313.

6. Burby RJ. Hurricane Katrina and the Paradoxes of Government Disaster Policy: Bringing About Wise Governmental Decisions for Hazardous Areas. The ANNALS of the American Academy of Political and Social Science, 2006; 604(1):171–191.

7. Nafday AM. Strategies for Managing the Consequences of Black Swan Events. Leadership and Management in Engineering, 2009; 9(4):191–197.

8. An L. Modeling human decisions in coupled human and natural systems: Review of agent-based models. Ecological Modelling, 2012; 229(0):25–36.

9. Klinenberg E. Denaturalizing disaster: A social autopsy of the 1995 Chicago heat wave. Theory and Society, 1999; 28(2):239–295.

`

10. De Bono A, Peduzzi P, Kluser S, Giuliani G. Impacts of Summer 2003 Heat Wave in Europe. United Nations Environment Programme, 2004.

11. Nateghi R, Guikema S, Quiring SM. Power Outage Estimation for Tropical Cyclones: Improved Accuracy with Simpler Models. Risk Analysis, 2014; 34(6):1069–1078.

12. Reilly AC, Tonn GL, Zhai C, Guikema SD. Hurricanes and Power System Reliability-The Effects of Individual Decisions and System-Level Hardening. Proceedings of the IEEE, 2017; 105(7):1429–1442.

13. Liu H, Davidson RA, Apanasovich T V. Statistical forecasting of electric power restoration times in hurricanes and ice storms. IEEE Transactions on Power Systems, 2007; 22(4):2270–2279.

14. Shashaani S, Guikema SD, Zhai C, Pino J V., Quiring SM. Multi-Stage Prediction for Zero-Inflated Hurricane Induced Power Outages. IEEE Access, 2018; 6:62432–62449.

15. Kabir E, Guikema SD, Quiring SM. Predicting Thunderstorm-Induced Power Outages to Support Utility Restoration. IEEE Transactions on Power Systems, 2019; 34(6):4370–4381.

16. McRoberts DB, Quiring SM, Guikema SD. Improving Hurricane Power Outage Prediction Models Through the Inclusion of Local Environmental Factors. Risk Analysis, 2018; 38(12):2722–2737.

`

Chapter 2: Learning as a Driver of Changing Community Vulnerability under Repeated Hurricanes and a Changing Climate

2.1 Overview

Hurricanes regularly cause large economic losses and loss of life along the Gulf and Atlantic Coasts of the U.S. The risks from individual hurricanes has been extensively investigated in the literature. However, little is understood about how individual and collective responses to repeated hazards change communities and impact their preparation for future events. Individual mitigation actions may drive how a community's resilience evolves under repeated hazards. In this chapter, we investigate the effect that learning by homeowners has on household mitigation decisions and on how this influences a region's vulnerability to hurricanes over time. To do this, we utilize an Agent-Based Model (ABM) to simulate homeowners' adaptation to repeated hurricanes and how this affects the vulnerability of the regional housing stock. We use a case study to explore how different initial beliefs about the hurricane hazard and how the memory of recent hurricanes could change a community's vulnerability both under current and potential future hurricane scenarios under climate change. In some future hurricane environments, different initial beliefs can result in large differences in the region's long-term vulnerability to hurricanes. We find that getting some homeowners to mitigate soon after a hurricane – when their memory of the event is the strongest – helps to substantially decrease the vulnerability of a community.

`

2.2 Introduction

Natural disasters cause considerable property damage and economic loss. In 2017, three major hurricanes, Harvey, Irma, and Maria, made landfall, caused more than $300 billion (USD) in economic loss, damaged more than a million homes, and left misery in their wake(1). Hurricane Irma caused approximately $70 billion (USD) in total economic loss, and half of that was due to damage to residential real estate(2). Residential building vulnerability is a critical component of a region's financial vulnerability when facing disasters. Low penetration rates of flood insurance in most coastal communities often mean that a singular event can erase a family's most valuable asset(3,4). Further, it can take months or even years after a disaster, depending on the availability of financial assistance, contractors, post-disaster economic opportunities, among other factors, before displaced families return to their homes and repair houses to normal, putting the long-term future of the community at risk(5). While there have been studies examining building stock vulnerability to hurricanes, most either, (1) assume that building stock is static(6) or (2) exogenously impose a change to the building stock within a hazard simulation model(7). This ignores the potential endogenous learning that stems from experience with hurricanes that could substantially alter the building stock over time.

Changes in building stock are part of the larger problem of estimating damage from repeated events. Singular hazard events are well-studied, and risk analysis methods are well developed for these events. For example, hundreds of studies have been published on the aftermath of Superstorm Sandy, ranging from evacuation(8), to nursing homes and elderly care(9) , to hardening the New York City and the New Jersey coastline to reduce damage should a similar storm strike again(10). In contrast, relatively little is known about how repeated hazards may

`

induce long-term changes to communities, and more specifically, how hazards change behaviors and policies which in turn influence the built environment. These alterations to the built environment change a community's ability to withstand future events.

There are numerous aspects of communities and their hazards that impact a community's long-term hazard vulnerability. These include decisions made by collective bodies to invest in large-scale protective infrastructure[11,12], land-use changes[12], and climate change[13]. An additional source of change stems from the actions (or lack thereof) by individuals[11]. These decisions result from their knowledge about the risk that hazards pose and personal preferences regarding individual risk tolerances and the opportunity costs of these decisions[14]. Additionally, heuristics and biases can induce decisions that, based on the information at hand, seem counterproductive[12]. Repeated hazards provide experiences from which those affected learn, and thus induce changes to the decision-making process regarding preparation and mitigation. With each new observation (i.e., a new hazardous event), individuals incorporate their new knowledge and update their beliefs about the risks associated with hazards[11]. This influences their actions and changes the trajectory of the region's vulnerability to hazards.

To better understand how regional vulnerability can evolve, we must first better understand the potential contributions of individual-level learning to regional risk. This is particularly critical in communities exposed to repeated hazards as the effects of this learning, and the decisions that follow, compound over time. There is not one way in which individuals understand and respond to hazard risk and different beliefs about risk can lead to different mitigation behaviors. It is important to understand the influence that archetypical behaviors have over time on community vulnerability to identify when interventions are warranted. For example, people who hold strong beliefs that hurricanes are infrequent and inconsequential are harder to sway even when their understanding of the hazard is updated by a new experience[15]. But the impact of this behavior as it compounds over time differs depending on location and climate intensification. The role of memory also influences behavior in a way that has compounding effects. More recent hazard experiences tend to increase an individual's risk perception and decrease their risk tolerance, making individuals more likely to mitigate. When significant time passes between storms, individuals are less likely to be prepared and have a lower likelihood of mitigating in a given year[16]. This is called the memory decay effect[12]. Similarly, availability bias in hazard decision-making occurs when individuals rely more on recent experiences than on their full experience[17].

In this chapter, we evaluate how learning from repeated events may influence how community vulnerability. We allow probabilistic beliefs about the frequency and intensity of hurricanes to evolve over time to evaluate how this may influence community-level risk. We show that even learning about a single aspect of the risk can create a large difference in how community risk – as measured by building stock vulnerability – evolves over time. This study investigates two specific questions: (1) When does learning from the hazard environment and from information campaigns decrease homeowners' vulnerability to repeated hurricanes in a changing climate? (2) How does the effect of memory influence a region's vulnerability in the long term? This work provides a platform for analyzing how hazard experiences influence one's understanding of the risk and how this causes communities to change through mitigation and possibly even retreat.

To accomplish this, we developed an agent-based model (ABM) that focuses on the influence of individual learning on homeowner decisions and then how these decisions affect regional building stock vulnerability. The model specifically evaluates the influence of prior knowledge,

`

the effects of memory, and the confidence that homeowners have in their own beliefs about hazard risk. In the model, homeowners make decisions by choosing from a set of mitigation actions (e.g., installing hurricane shutters or hurricane straps). The approach is place-based, with a case study area based on 357,120 individual homeowners across nine coastal counties in the state of Maryland. However, both the models and insights are generalizable to other areas.

The structure of this chapter is as follows. We first review the literature of relevant studies as the construction of such models is highly interdisciplinary. We start by giving an overview of the literature on (1) vulnerability to and mitigation of hurricanes, (2) the use of ABMs in natural disaster studies, and (3) hurricanes and climate change. In Section 3, we introduce the framework of ABM, which includes a hazard model, a damage model, and models for individual learning and individual mitigation decisions. The case study is presented in Section 4 and details of model evaluation are presented in Section 5. In Section 6, we show results, including how model can be used with different climate scenarios, and discuss the role of learning, the impact of initial perception of hurricane risks, and the impact of memory decay.

2.3 Background

2.3.1 Hurricane vulnerability and mitigation

Systems-level research on hurricane risk typically focuses on quantifying regional damage, the benefits and cost-effectiveness of mitigation, and how federal and local policies may induce homeowners to mitigate their risks of being damaged to reduce region-wide vulnerability. We begin with a discussion on regional damage assessment. Many studies rely on FEMA’s HAZUS

`

framework to simulate hazards and hazard impact, including estimates of economic losses from infrastructure and residential housing damage[7,18,19]. HAZUS uses a whole-entity approach to modeling; each building class is assigned a hazard fragility curve (derived from the literature, reconnaissance, and experimental testing), which is based on a conditional probability of damage state given a hazard intensity. For example, for a hurricane, the intensity measure is typically gust wind speed. This can be too simplistic, especially when considering an individual house in a specific location. However, on a regional-scale, it provides an approximation of the expected building damage given a hazard. HAZUS has additional known shortcomings, including out-of-date building stock, crude hurricane and flooding scenarios – mostly accredited to accommodating the computational limits of most users – and deterministic hazard scenario modeling. Despite these limitations, it offers both an easy-to-use platform for researchers and practitioners to test hypotheses and its ubiquity has allowed for comparison among studies. Component-based approaches, on the other hand, are more granular. They focus on the vulnerability of individual building components, and the interactions among them, providing more accurate estimates of singular building damage[20–22]. This approach requires significantly more knowledge of the building stock and computational resources. Our work relies on the whole-entity approach given its ubiquity and less inputs requirements. We do not use HAZUS directly, but we do use the building fragility curves provide by HAZUS. This allows us to overcome problems surrounding out-of-date building stock and crude hazard scenarios while still leveraging the HAZUS fragility functions.

Mitigation reduces the likelihood that a building experiences damage due to a hurricane. This chapter focuses on household-level mitigation, such as adding wind-straps or hurricane shutters. Not all mitigation is warranted based on its cost and the likelihood of damage[23]. For example,

`

Pinelli et al. (2009) studied the cost effectiveness of various mitigation strategies using a Monte Carlo simulation for different regions in Florida and created a map indicating the benefit/cost effectiveness for different combinations of housing mitigation decisions(24). Rose et al. contributes to this discussion on the cost effectiveness of mitigation using empirical data on hazard mitigation grants(25). They find a 4.7-fold return-on-investment for wind mitigation from FEMA mitigation grants – grants given to communities particularly susceptible to high windspeeds. These studies, and others like it, assume a static building stock and do not include the compounding benefits from mitigation from repeated hazard events.

Additional considerations for understanding how regional building stock vulnerability may change include homeowners' proclivity towards mitigation and federal, state, and local policies that influence homeowner mitigation behavior. Federal policies are also commonly evaluated with benefit-cost analysis or survey-based approaches(26), though these too are often based on a static snapshot of the situation.

Little of this past work has explicitly investigated the effects of repeated hurricanes – and their effects on mitigation behavior - on communities with the exception of Jain et al., Reilly et al.(7,11,27). Jain et al. propose a method to consider temporal changes in building inventory when estimating changes in expected losses from hurricanes over time(7). This simulation model considers how building vulnerability responds to changes in building codes. Reilly et al. model that homeowners can increase their houses resistance to wind following a set of different rules defined following rules defined by probabilistic distributions. However, these studies do not explicitly investigate the potential role of learning as a determinant of long-term building stock vulnerability. Other studies that do consider the role of sequential hazards focus more on the degree to which learning takes place, and not the impact of learning(28). Siegrist et al. explained

`

the importance past flooding experiences and how they might change people's risk perception and motivate mitigation behavior. Their survey results suggested that negative experiences increase the likelihood of household mitigation behaviors but not necessarily when people also doubt the effectiveness of mitigation or perceive the cost as too high. Better understanding this potential role of learning is critical in understanding repeated hazards and their impacts on communities over time.

2.3.2 Agent-based models and their use in natural disaster studies

An ABM is a bottom-up *in silico* modeling approach that simulates how heterogeneous agents (e.g., homeowners) interact with and learn from other agents and the environment and how, with this knowledge, they make decisions[(29,30)]. Their decisions then change the environment – and in our case, their homes. ABMs have been broadly used in tackling problems involving humans and their decisions. Applications of ABMs span demography, the social sciences, economics, public health and environmental science among other fields[(31,32)].

The primary benefit of ABMs is that they allow researchers to create a large collection of scenarios (e.g., different policy alternatives or learning behaviors) to conduct sensitivity analysis which facilitates answering various questions under different circumstances[(14)]. This allow researchers to quantify the impact of various interventions (e.g., mitigation policies) or exogenous changes to the hazard environment (e.g., via climate change). Compared to top-down modeling approaches, which include statistical and optimization models, another benefit of ABMs are their ability to model the compounding effects of individual actions and learning[(27)]. The results from ABM simulations are challenging, and often impossible, to validate due to the complex nature of human behavior and the use of scenarios (or policies) that have not been

`

observed in reality yet[14]. This is overcome by using behaviors and environments that have been observed by researchers or practitioners. For example, an ABM can be calibrated to past observed patterns, and behavioral models can be based on real-world surveys or observations of behavior. ABMs provide an advantage in that the scenarios that are examined need not have occurred, unlike e.g., statistical models which rely upon prior data. ABM output is often high-dimensional and usually requires sophisticated analytical approaches to interpret. One approach that can help in comparing different scenarios is using common random numbers. This reduces the variance of the differences between scenarios by fixing the randomness of the input, thus making the underlying differences from scenarios more obvious.

ABMs have been widely used to model natural disasters and their impacts. ABMs are useful boundary objects able integrate domain knowledge from multiple disciplines, making them increasingly popular [33]. However, most ABM work in the hazards realm centers around short-term actions after hazards occur to evaluate how agents might evacuate, take shelter, or access primary care services, and what infrastructure is needed to accommodate emergent behaviors[34–36]. An ABM could be used to model the long-term effects of natural disasters[11,27,37]. For example, Reilly et al. (2018) quantified how hurricane-induced power outages that induce particular behaviors, may, in turn, influence a region's power system reliability in the long-run.

2.3.3 Modeling Individual Learning in ABMs

People’s behaviors are driven, in part, by their beliefs[38] and preferences[39]. Beliefs can change over time when individuals gain additional information – perhaps through an experience

`

or through information shared others – and then update their understanding of the process. Because learning affects how individuals make decisions over time, learning can be an essential component in an ABM framework for studying long-term implications of hazard vulnerability[40]. In ABMs, agents are considered autonomous and can interact with one another, meaning they can learn from both the environment and from other agents.

Learning can be viewed as a process in which beliefs are updated. In reality, this process is complex and is the subject of numerous psychological studies[41]. In an ABM, however, learning is modeled in a logical structure compatible with computer coding. There are several ways to model this process in an ABMs. The primary methods include information modeling, Bayesian updating, reinforcement learning, and coevolutionary algorithms. We briefly review each below.

A common approach – information modeling or risk modeling – uses new information about disasters and risks perceived by the agents as the knowledge learned by the agents. For example, Du et al. explored evacuation processes during flooding events using an ABM[42]. In this study, agents learn of a flood from a news broadcast, social media, and neighbor observations, which, when combined, triggers decisions by agents. In other work, Tonn et al. quantifies how different risk perception factors (i.e., flood experiences) and coping perception factors (i.e., mitigation behaviors) combine to estimate perceived risk factors. This value will then be compared with a risk tolerance threshold to determine when an agent will consider taking action [37].

Bayesian learning or Bayesian updating is another approach for modeling learning[38]. This method assumes an agent's beliefs about future events, such as the likelihood of a hurricane, follow specified probability distributions. These distributions are updated as new information becomes available using Bayes theorem. For example, Reilly et al. (2017) modelled agents as having categorically distributed beliefs over the likelihood of zero, one, or more power outages

`

in a given year[11]. This is then combined with Dirichlet conjugate priors which are updated in a Bayesian manner every year given the number of power outages that occurred.

Another popular learning algorithm used in ABMs is reinforcement learning[39]. In reinforcement learning, agents receive rewards from their actions, and they make decisions based on a new environment and the consequences caused by their prior actions. Krause et al. model the process of power suppliers submitting their bids to the electricity market using reinforcement learning[43]. The agents are trying to maximize their payoffs and after each bid, they observe their gains or losses to update their behavioral policies or expected reward functions. As a result, they discover that after many rounds of bidding, the stable decisions each agent made to achieve their optimal output is the same as the existence of a unique Nash equilibrium or multiple equilibria in the system.

For each learning method, it is possible to incorporate learning and knowledge biases that individuals exhibit. For example, bandwagon effect or herd behavior is a cognitive bias in humans that people follow what others are doing instead of using their own information or making independent decisions, which is commonly observed and can be detrimental to disaster evacuations[44]. In natural hazards research, researchers have observed that the memory of events tends to fade over time (i.e., memory effect), which makes individuals less likely to take preparative actions[11,12], and this behavior is critical to capture in the model process. Memory fading can be a result of time or of new people moving into the area. In ABMs, the memory effect is often modelled as either a decay parameter on the awareness of the hazard[12] or as a window that allows only those events in the window to be 'remembered' by the agents[11].

`

In this work, we use a Bayesian updating model with a memory effect. This captures agent learning even without intervening actions (as in reinforcement learning) and does so in a way that is consistent with existing knowledge, particularly of the importance of the memory effect.

2.3.4 Modeling Decisions in an ABM

The decision-making component of an ABM is a critical element for modeling emergent behavior. There are many different types of decision rules that can be used within ABMs. For ABMs in the hazards field specifically, we divide behavioral decision models into four groups: 'if-then', descriptive, empirical, and prescriptive.

The first approach is "if-then" models which take the format of "if this happens, then one will do that with some likelihood." This approach is more ad-hoc and often hard to be validated, though it is widely used in the ABM literature. The probabilities can be populated based on observation (i.e., empirical) or based on subject matter experts, but that is typically not the point. The objective is to find the marginal influence of different types of decisions on model outcomes, and these types of models require extensive sensitivity analysis(11,37,42).

Descriptive decision theory attempts to explain the actual behaviors of decision-makers. This is often different from their utility optimizing actions. Prospect theory is one example of a descriptive framework. Here, individuals evaluate outcomes based on possible gains and losses rather than expected utility(45). Another well-known collection of descriptive decision theory method is bounded rationality. Models for bounded rationality assume individuals would be rational in their decision-making process, except that they are limited by the information they

`

have, their cognitive limitations, and/or the amount of time they have to make the decision[46]. Various heuristics and biases have also been studied to help explain observed actions. An examples of using these methods in the ABM literature is a model of water scarcity management for repeated droughts.[47]

Empirical models apply observed decision behavior (e.g., observed mitigation rates). The approach is reasonable when building an ABM for a particular region using observed data from that same region provided that the behavior is unlikely to change in the future. Then with a well-established dataset, statistical approaches can be applied to model agent's behaviors. However, individuals from different regions have been observed making different choices under similar hazards[48] – likely because of different background understanding of the hazard – meaning empirical data are at times less relevant when applied to different regions. These methods are also less useful for situations in which agents must make decisions in environments they have not seen before.

Prescriptive decision theory models focus on identifying the "best" decisions by using expected utility maximization. This model assumes that a rational decision-maker lists the actions or alternatives available to him or her, identifies the possible outcomes associated with each action along with the likelihood of occurrence of these outcomes, and finally quantifies the desirability of each outcome using a utility function. The alternative with the highest expected utility is selected. While attractive for improving individual decision-making, utility theory is not necessarily a descriptive approach that captures the actual decision process people use[49].

In this chapter we use a utility-based model with subjective information for agent decision-making. Given that the decision that the agents are making are both expensive and are associated

`

with long-term changes to their homes, the agents are likely to consider costs and benefits in depth, analyzing their options.

2.3.5 Climate change and hurricanes[50]

Climate change is likely to alter the pattern of hurricane occurrence in the next century[51]. While there is uncertainty in how climate change will affect hurricane frequency and intensity, it is likely that hurricanes will become less frequent though more intense in the Atlantic basin[52]. On the other hand, climate change will also cause sea level rise, which will inundate coastal regions and result in flooding of areas that might have never seen storm surge in the past. In addition, greenhouse gas-induced warming may lead to gradually increasing risk of the occurrence of highly destructive category-5 storms[53]. We use a scenario-driven approach, similar to Staid et al., to better understand how individual behavior interactions with a changing hurricane climate to influence community vulnerability over time[13].

2.4 Methods

2.4.1 Overview of Model structure

Our approach builds on the agent-based model in Reilly et al. by integrating computational learning and decision-making models and synthetic stochastic hurricane models[27]. The addition of synthetic hurricane, as opposed to the use of historical hurricanes, allows for us to test the implications of different possible future climate scenarios. Our agents are individual homeowners, modeled as one homeowner per land parcel and we focus on the impact of

`

hurricane wind hazard. The model is initialized by assigning each agent attributes of their house (i.e., their house's construction type, a.k.a., building class, and the house's ability to resist wind forcing, a.k.a., resistance level) and a wealth (approximated by their house's improved value). The agents make mitigation decisions on an annual basis. Decisions are driven by an agent's perception of the risks from hurricanes. We simulate how the region could evolve over 100-year periods in various hurricane scenarios. We choose 100-year as it is a sufficiently long period to allow us to see the results of individual behavior in a setting in which events happen relatively infrequently. The goal is to find the marginal contribution of decision-making related to mitigation on the vulnerability of the regional building stock. It is not to predict the building stock or its quality in 100-years. We compare different scenarios and analyze the level of vulnerability of the community and discuss the importance of learning and initial knowledge of the agents to their decisions. Figure 2-1 provides an overview of the model used in this study. We first give an overview of this structure, and the subsections below provide more detail on the model's components.

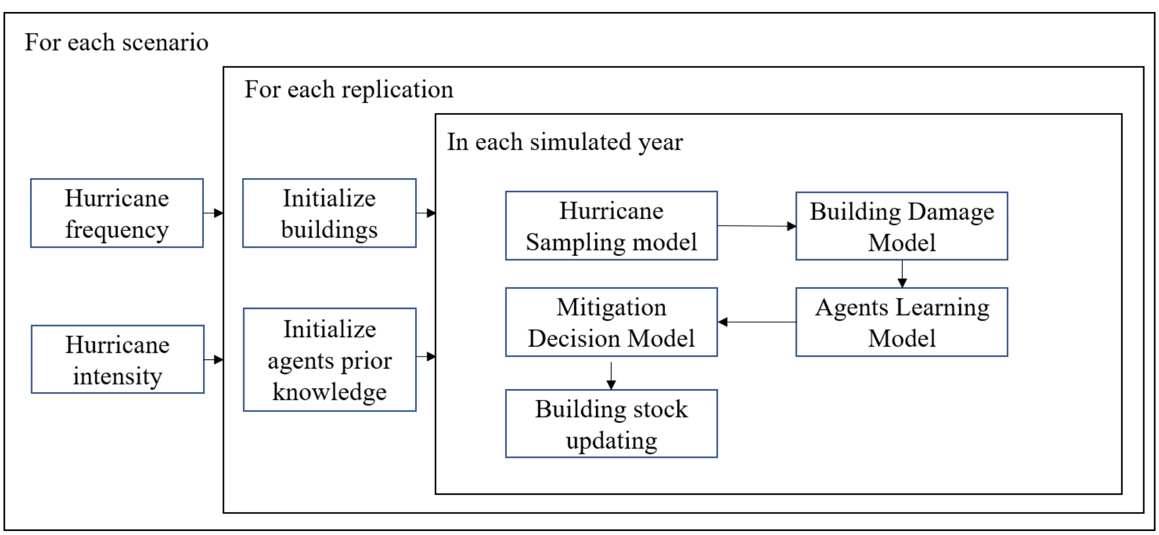


Figure 2-1 Overview of the computational flow of the ABM.

`

We create different hurricane climate scenarios by controlling the frequency and intensity of synthetic hurricanes used in our analysis. Then we run a number of replications of the entire 100-year history using a simulation model that can be divided in to four components: the hurricane sampling model, the building damage model, the learning model, and the mitigation decision model. Many of these components have multiple steps to compute that are described in later subsections. The general process is that in a given time step (i.e., year), we sample the number and intensity of hurricanes to occur (zero hurricanes is a possibility in a given year) from distributions initially parameterized to reflect the conditions of the study area. The hurricane(s) that most closely matches the intensities are selected from a large library of synthetic but possible storms. This process is described in Section 3.2. Each house is probabilistically assigned damage (i.e., a damage level) based on the intensity of the hurricane, where it made landfall, and the construction type and resistance level of each home (which together, have their vulnerabilities represented by fragility curves). This process is described in Section 3.3. Agents then learn from this experience (which could include no hurricanes or no damage even if there is a hurricane) and then make mitigation decisions. This process is described in Section 3.4. This entire process repeated for 100-years under different climate scenarios. The model purposefully excludes numerous confounding factors, including insurance, disaster policy, and relocation, to isolate the impacts of learning from disasters for various climate scenarios. Future work could explore these effects.

The number of replications needed for stochastic convergence is determined by replicating the entire 100-year time horizon 10,000 times. We use the coefficient of variation shown in equation (1) to calculate the output from the ABM. We then use equation (2) to find the minimum number of replications, n_{min}, needed to achieve desired level of convergence, E,

\`

between the coefficient of variation from *n* replications and *m* replications, where *m* is a large number. c_V^n is the coefficient of variation from *n* replications.

$$c_V = \frac{\sigma}{\mu} \qquad (1)$$

$$n_{\min} = argmax_n |c_V^n - c_V^m| < E, \forall m > n \qquad (2)$$

2.4.2 Hurricane sampling model

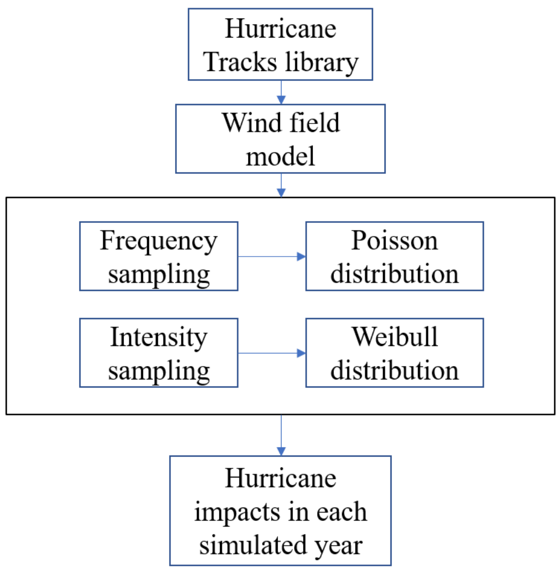


Figure 2-2 Hurricane Sampling model

An overview of the hurricane sampling model of the ABM is shown in Figure 2-2. A synthetic library of possible hurricanes to affect the region is initially created and is used for all simulations (see Appendix A). The library consists of 36,399 plausible storm tracks that could affect the study area, ranging from tropical storms to Category-5 hurricanes. With these tracks, we apply hurricane wind field model to compute the 3-second 10-meter peak wind gust at the

`

centroid of each parcel [54]. The library itself contains no information on the likelihood of any of these synthetic hurricanes over other synthetic hurricanes. Next, baseline hurricane frequency and intensity distributions are established for the region. The region's historic yearly hurricane frequency and maximum hurricane windspeed are fitted to Poisson and Weibull distributions, respectively, to create baseline distributions. The fitted parameters are later perturbed for different climate change scenarios (e.g., the rate parameter for climate scenario with 10% more hurricane is multiplied by 1.1). Once the simulation has begun, in each simulated year, we sample from the intensity and frequency distributions to identify how many hurricanes affect the region that year (0 or more) and how intense each hurricane is. The hurricane in the synthetic library that is closest in intensity to the sampled intensity is selected. A detailed description of this model is provided in Appendix A.

2.4.3 Building damage model

For each hurricane, the 3-second 10-meter peak wind gust (which is typically used in the design code for wind resistance in structural engineering) at the centroid of each parcel is computed[54]. If more than one hurricane occurs in a year, the hurricane that is most intense is used. That is, damage from subsequent storms does not compound. Then, each house is assigned a damage state by combining building's fragility curve with the downscaled windspeed at the parcel's location.

HAZUS fragility curves are used to model each house's fragility. Each parcel is assigned to a building class based in its characteristics (e.g., number of stories, construction material). Each building class has distinct possible resistance levels, depending on the possible ways mitigation can be conducted (e.g., wind straps). As a house is mitigated, its resistance level improves. Each

`

resistance level has five fragility curves, one for each of the possible damage states. These are: Damage State 0, (DS0, no damage), Damage State 1 (DS 1, minor damage or 10% or less of the building sustained damage), Damage State 2 (DS2, moderate damage or about 25% of the building sustained damage), Damage State 3 (DS3, severe damage or about 50% of the building sustained damage), and Damage State 4 (DS4, completely destroyed)[(55)]. We assume every house will return to its original state at each simulated year. Fragility curves produce a probability that a house of a specific type and resistance level will be in one of five damage states conditioned on the wind speed.

The damage model randomly samples a number between 0 and 1 and a damage state is selected in proportion to the likelihood of being in a specific damage state given a wind speed. The losses are quantified by multiplying the improved value of the parcel by the fraction of sustained damage. Given the simplicity of these fragility curves, criticism exists[(27)]. Our study is not dependent on this model and alternate fragility curves can easily be incorporated as long as they give similar output[(27)].

After a hurricane, we assume that any agents which experienced damages rebuilds. Additionally, some or all agents may decide to mitigate. A wood-framed home, which constitutes the vast majority of the houses in our study area, could be upgraded by installing roof-wall straps, application of secondary water resistance, installing storm shutters, strengthening the roof deck attachment, or by changing the shape of the roof. Part of this decision depends on the cost of mitigation. We estimated the cost for individual homeowners to make house upgrading decisions using the RS Means cost pricing database[(56)].

`

2.4.4 Agent Learning and Mitigation Decision Models

Mitigation decisions are essential components of how a community's vulnerability evolves. These decisions are partially controlled by an individual's personal beliefs about the risk a hazard poses(57). By observing a hazard and the damage it causes, or the lack thereof, individuals may update their beliefs about the risks posed by the hazard, and potentially change future mitigation decisions.

After each hurricane season in a simulated year, agents learn from their experience and may take mitigatory action. For instance, they may believe that the likelihood of a hurricane in any given year is now greater especially if they recently experienced a hurricane. These beliefs are combined with damage probabilities for each of the mitigation strategies and are used as inputs for a decision model on whether to mitigate.

The ABM is initialized by assigning each agent partial information or "knowledge" about the hazardscape. Specifically, each agent holds an initial belief about the frequency of each category of hurricane (including no hurricane). This is updated annually given their experiences. In each year of the ABM run, regardless if hurricane or damage occurs, agents "learn" from their experience by updating their beliefs about their risk and then decide whether to act. Based on this knowledge, agents may choose to mitigate, which, in turn, may change the vulnerability of their homes to damage in future hurricanes.

2.4.4.1 Agent Learning Model

We model agent learning using a Bayesian updating framework. Learning in this model focuses on how likely agents believe hurricane force winds of varying magnitudes are to occur

`

on their parcel in a given year. For example, an agent could believe that there is an 80% likelihood of no hurricane occurring in a given year, a 19% likelihood that their parcel experiences tropical storm force winds, and a 1% likelihood that a Category 1 or stronger storm would occur. If this agent were to experience Category 2 force winds, their beliefs about their chances of experiencing a Category 1 storm or greater would likely change.

We use a categorical distribution to describe each agent's belief for both hurricane frequency and intensity. We define X as a categorical random variable,

$$X \sim Cat(p_1, p_2, \dots p_7), \qquad (3)$$

where p_1 through p_7 correspond to the probability of no hurricane, a tropical storm, and Category 1-5 hurricanes. While each of these divisions has a wide range of windspeeds, we select them because an individual is unlikely to know the exact windspeed they experienced, but rather an approximate wind speed such as the intensity of the storm on the Saffir-Simpson scale. The Dirichlet distribution is the Categorical distribution's conjugate prior. $\alpha_1, \alpha_2, \dots, \alpha_7$ are the Dirichlet's support parameters and represent the number of observations for each of the windspeed divisions. Thus,

$$p_1, p_2, \dots p_7 \sim Dir(\alpha_1, \alpha_2, \dots, \alpha_7) \qquad (4)$$

If we treat the categorical distribution parameters as random variables, we can leverage each new windspeed observation, X_{new}, to update these parameters using Bayes rule and the Dirichlet distribution. The posterior distribution will still be a Dirichlet distribution with parameter $\boldsymbol{\alpha}'$ given by,

$$\alpha_i' = \alpha_i + \mathbb{I}(X_{new} = i), i = 1,2, \dots, 7, \qquad (5)$$

`

where $\mathbb{I}$ is the indicator function. The posterior predictive distribution for this model is given by equation (6).

$$f(X = i|\alpha) = \frac{\alpha_i'}{\sum \alpha_{i'}}, i = 1,2,\ldots,7 \qquad (6)$$

Before the ABM is run, priors – meaning initial beliefs over the intensity and frequency of hurricanes, $p_1, \ldots p_7$ – need to be assigned for each agent. We iteratively assign different starting priors in different ABM runs to test how initial beliefs influence the agents' decision process and the long-term vulnerability of the region.

We also consider the impact of memory decay in the learning models. Memory decay or "memory effects" is a phenomenon that occurs when an individual gives more weight to more recent events[(58)]. We model the memory effects using

$$\alpha_i' = \alpha_i * d + \mathbb{I}(X_{new} = i), \qquad (7)$$

where $d < 1$, represents a memory decay. More specifically, $d < 1$ forces less weight on prior experiences and more weight on the agent's experience in that year.

Finally, we consider the confidence that an agent has in its beliefs when making decisions by leveraging the parameters that support the Dirichlet distribution. An individual who has significant experience may be more likely to act than an individual who has much less[(57)]. As example, consider two agents with different priors: (1,1,0,0,0,0,0) and (100,100,0,0,0,0,0), respectively. Although both indicate the agents believe there is a 50% probability of having no hurricanes and 50% probability of having a tropical storm in a given year, the underlying confidence for these two agents is distinct. If the two agents had priors that were (1,1,0,0,0,0,1) and (100,100,0,0,0,0,1), the agent with less experience is more likely to consider preparing for a

`

more extreme event (i.e., a Category 5 hurricane). In the ABM, we incorporate confidence by giving the agents different initial beliefs. For example, in one scenario the agents begin with a prior of (1,1,0,0,0,0,0), where in another scenario they will hold priors of (100,100,0,0,0,0,0).

2.4.4.2 Mitigation Decision Model

After agents learn, they decide whether and how to act. A detailed mathematical description of our decision model is presented in Appendix B. A brief overview follows. The agents choose from the following alternatives in each year that the model is run: installing roof-wall straps, installing storm shutters, improving the roof deck attachment, installing second level water resistance, changing the shape of the roof, or simply doing nothing. Each option corresponds to a specific resistance level. Houses that are mitigated have a lower probability of being damaged when they are upgraded which is represented by differing fragility curves. However, the degree of improvement and the mitigation costs vary considerably based on the intervention. We assume agents know the probability of their house being damaged for each category of hurricane and for all housing mitigation options. Future iterations of the work could explore relaxing this assumption.

In each simulated year, we calculate each agent's expected utility for each mitigation alternative by combining the agent's probability distributions over hurricane frequency and intensity and the likelihood of damage for each mitigation option together with the costs of the options and the agent's utility function. An exponential risk averse utility function is used[59]. The agent then chooses the alternative that maximizes their expected utility.

2.5 Case Study

`

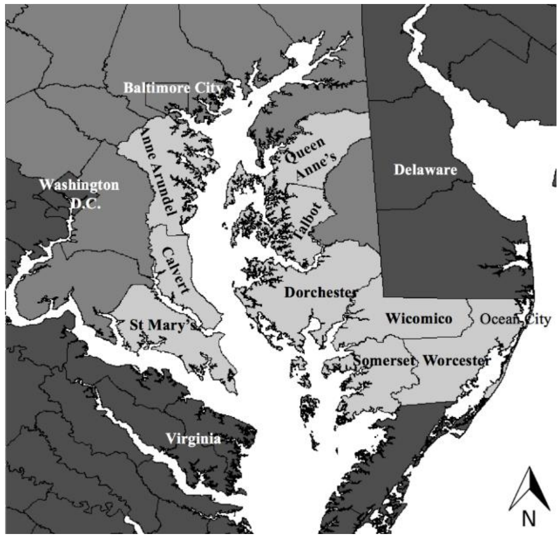


Figure 2-3 Map of region. Counties in light gray are used in the study.

Our case study is based on nine counties in the state of Maryland, U.S. - Anne Arundel, Calvert, St. Mary, Dorchester, Talbot, Queen Anne's, Wicomico, Somerset, and Worcester. These counties have coastlines along the Chesapeake Bay or the Atlantic Ocean (Figure 2-3). There are 357,120 single-family houses in this region. We exclude multi-family housing as the mitigation decision is largely not to be controlled by the residents. This leaves 11 building classes based on FEMA's residential building designations. The most common building class is two-story wood-framed single-family houses. Then for each type of building, different mitigations, such as installing hurricane straps or change roof shapes, can be invested to improve the building's resistance level. A detailed description of all mitigation strategies can be found in Reilly et al[27]. Publicly available tax-assessor data are used to spatially-locate each house and to assign each house a value. The value (i.e., "improved value" from tax assessor data) combined with the building's fragility and the hurricane windspeed provide an estimate of the loss. Anne Arundel County is the most populated and is mostly suburban. Ocean City, in Worcester County, also has dense housing. The remainder of the region is either suburban or rural.

`

Because we do not know either income or wealth at the household level, we use property value as a proxy for wealth (capital) to support mitigation decisions. This is an imperfect measure. However, those with higher home values are arguably more likely to have access to capital for mitigation through loans using their home's value as collateral.

The case study area experiences a hurricane every 11 to 13 years on average. Most storms are Category 1 strength or below by the time they make impact in the region. The hurricane intensity tends to be greatest in the southern-most counties, such as Somerset and Worcester Counties.

We are not attempting to model this region in all of its detail. That is, we are not trying to model all details from these nine counties in order to support specific policy recommendations for this location. Rather, we used data from this region to parameterize our model to provide a reasonable degree of reality and complexity to gain broader insights on the interactions among hazard frequency, damage experience, mitigation, and regional vulnerability over time.

2.6 Measure of Community Vulnerability

To compare different scenarios, it is helpful to have a measure of the overall vulnerability of the region to more simply see the effect of different scenarios. We do this through an additional calculation outside of the ABM. That is, we assessed the vulnerability of each household to six different intensities of tropical storms in that year and integrated these vulnerabilities into an overall vulnerability score for our study region. This score does not feed back to the ABM in any way. Instead, this was an extra assessment done only to provide an integrated picture of the overall vulnerability of the houses in the community. This section provides the details of how this was done.

`

In each simulated year, we record the state of each house in terms of its resistance level (i.e., its degree of mitigation). Then, separate from the ABM, i.e., not providing any feedback into the ongoing ABM, we uniformly applied six different wind speeds (the median wind speed of each hurricane category) to each house and calculated the expected damage (in dollar value) for each house and for each windspeed. We then summed the expected damage across all of the houses to get d_i, the aggregate vulnerability measure for windspeed i.

Given that there is an upper limit of the achievable resistance of houses to hurricane winds, we also calculate the maximum and minimum expected damage, $d_i^{\max}$ and $d_i^{\min}$, which describe the most and the least damage a community could experience under wind speed level i. The maximum expected damage, $d_i^{\max}$, corresponds to all houses being in the lowest resistance level (i.e., no mitigation has been conducted). Similarly, the minimum expected damage, $d_i^{\min}$, corresponds to all houses being in the highest resistance level (i.e., no mitigation has been conducted). We then normalize and aggregate according to equation (8):

$$\mathbf{VI} = \frac{1}{6} \sum_{i \in wind\ level} \frac{d_i - d_i^{\min}}{d_i^{\max} - d_i^{\min}} \qquad (8)$$

Equation (8) normalizes the range of vulnerability measure to be between 0 and 1. For example, the vulnerability in Year 1 is 1, which assumes that no households have conducted mitigation and are in their initial resistance levels. If every household in the system chooses to mitigate their house to the fullest extent, the community vulnerability will be 0. This is for the purpose of best visualizing the output results but it does not mean houses become invincible under such events. The simulation for each replication of the full history always starts with no house upgrades. The table below shows the maximum and minimum damage (in US dollars) for

`

the case study region. We use **VI** over time as our set of overall community vulnerability measures.

Table 2-1 Community Damages from each wind speed

Wind Speed Level	50 mph (~TS)	85 mph (~Cat 1)	103 mph (~Cat 2)	120 mph (~Cat 3)	143 mph (~Cat 4)	160 mph (Cat 5+)
$d^{\max}$ ($)	1.04×10^6	8.50×10^8	7.96×10^9	2.67×10^{10}	4.87×10^{10}	5.37×10^{10}
$d^{\min}$ ($)	1.43×10^5	4.18×10^8	2.61×10^9	7.76×10^9	1.94×10^{10}	3.04×10^{10}

2.7 Results

This section presents the results of our analysis. We start with the influence of hurricane frequency and intensity given weak priors (i.e., low confidence of their assessment of the hazard). In the second subsection, we then vary the intensity of the priors. The priors are varied by iteratively perturbing the support parameters of the Dirichlet distribution that reflect each agent's knowledge and level of confidence when each run of the ABM is initialized. This helps us to isolate the influence of individuals who learn and gain confidence in their knowledge on regional vulnerability verses those who learn little from their experience. Similarly, we iteratively perturb the memory decay parameter to quantify the effect that an emphasis on more recent events has in forming beliefs on mitigation decisions.

2.7.1 Influence of Hurricane Intensity and Frequency

In this section, we first show how community vulnerability evolves in our model under different hurricane climate scenarios. We tested multiple intensities and frequencies of hurricanes. The intensity and frequency are modeled as 0.75, 1, 1.25, 1.5, and 1.75 times the historical fitted hurricane frequency and intensities. We initialized the agents with a weak prior

`

of (1,1,0,0,0,0,0), which assumes the agents have little knowledge with the environment, and that they believe they are equally likely to experience a tropical storm as a Category 1 storm. This is later referred to as the "baseline" priors. These priors are unlikely to induce mitigation before an agent observes additional hurricanes because an agent holding these beliefs does not believe a hurricane with force significant enough to cause substantial damage their house will occur. The results of this initial analysis are shown in Figure 2-4.

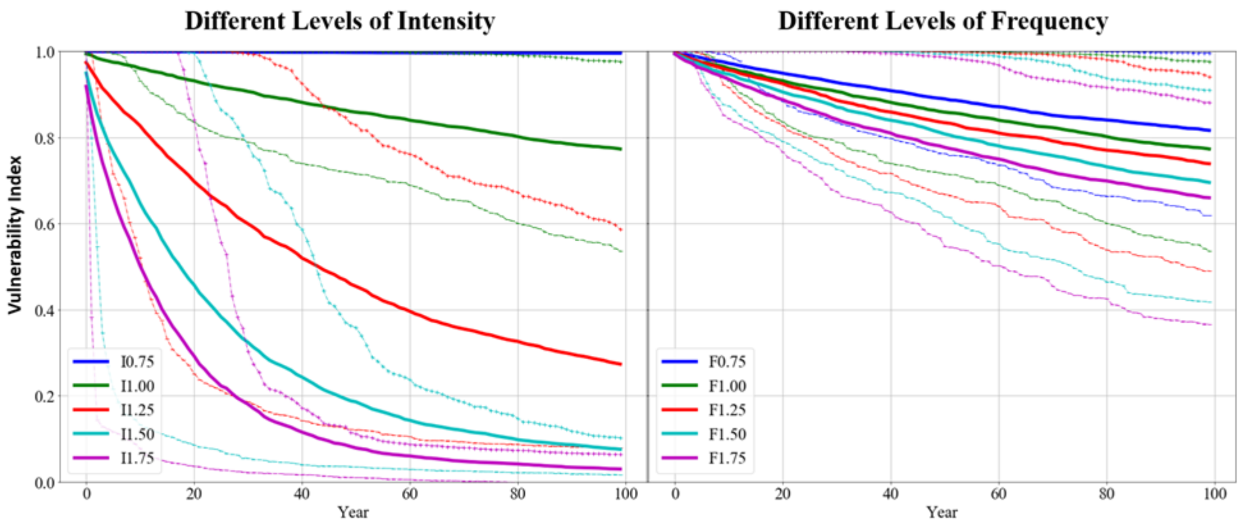


Figure 2-4 The baseline evolution of community vulnerability over a century under different hurricane scenarios. The agents are initialized with a (1,1,0,0,0,0,0) prior. Each solid line represents the mean vulnerability index across all replications in each simulation year. Dashed line with the same color with '+' sign are the 85% upper bound of the output samples. Dashed line with the same color with '-' sign are the 15% lower bound of the output samples.

In Figure 2-4, for both plots, the X-axis is the year within simulation, and the Y-axis is the vulnerability index (VI). I0.75 means hurricane intensity is 0.75 times the intensity generated from the current and baseline climate, and F0.75 means the frequency is 0.75 times the frequency in the baseline climate. The orange lines for both plots are the same baseline scenario with intensity and frequency matching historical observations. Each line represents the mean vulnerability index across all replications in each simulation year. We see that the community vulnerability decreases over time. This occur because we start the agents with no mitigation and any mitigation decision the agents are making will reduce the probability of damaged. We also

`

see that the intensity and frequency of hurricanes also substantially affect how community vulnerability evolves. As the intensity or frequency of hurricanes increases, community vulnerability to a storm of a given intensity decreases, and the higher the increase in intensity or frequency, the greater the reduction in vulnerability. This is because more mitigation takes place in response to more realized damage and this mitigation occurs sooner, meaning the benefits compound.

Based on the rate of change over time in the VI, the agents are more responsive to intensity changes than frequency changes. That is, for a given level of increase intensity, there is a greater reduction in vulnerability than for the same level of change in frequency. Furthermore, for the I1.75 scenario, the vulnerability of the region asymptotically approaches the minimum vulnerability (most resistant to hurricanes) over time. Hurricanes that are more intense tend to stimulate substantially more mitigations and reduce community vulnerability than more frequent but mild hurricanes.

2.7.2 Impact of Initial Knowledge on the Evolution of Community Vulnerability Under Different Climate Scenarios

To examine the effect of initial beliefs held by agents on the evolution of community vulnerability under different hurricane environments, we tested different sets of initial priors that all the agents hold at the beginning of the simulation. Table 2 below shows the priors we use and what each prior implies about their beliefs. We compare the results from starting with each prior with the results from starting with the baseline priors (provided in Table 2) to examine the effects of initial knowledge under each climate scenario.

`

The baseline evolution of vulnerability is shown in Figure 2-4 and is used to compare with each different prior scenario under the same climate. As before, we changed the intensity or the frequency, but not both together. We also controlled the hurricane records input for each hurricane climate scenario (e.g., a certain frequency or intensity comparing to the baseline scenario) with common random numbers to reduce the variance and make the results comparable for each scenario. That means the area will always be impacted by the same hurricanes over 100 years in each replication for each initial knowledge priors. The common random number approach helps decrease the variance and reduce the number of replications needed for convergence when comparing different scenarios[60]. Figures 2-5 and 2-6 show the difference between the mean VI from the baseline-priors case and the mean VI for a scenario using modified priors (but the same intensity-frequency combination). Also shown are the 95% confidence intervals that are constructed by bootstrapping of all VI outputs from each replication in that scenario.

Table 2-2 Initial knowledge explanations

Priors Name	Priors (No Hurr – Cat 5+ Hurr)	Implication	Legend
Partially Uninformative 1	(1,1,0,0,0,0,0)	Little knowledge for hurricanes events stronger than tropical storm	(**Baseline**)
Partially Uninformative 2	(1,1,1,0,0,0,0)	Little knowledge for events stronger than category one storms	
Partially Uninformative 3	(1,1,1,1,0,0,0)	Little knowledge for events stronger than category two storms	
Partially Uninformative 4	(1,1,1,1,1,0,0)	Little knowledge for events stronger than category three storms	
Partially Strong 1	(10,10,0,0,0,0,0)	Moderately strong belief that there will be no events stronger than tropical storm	
Partially Strong 2	(100,100, 0,0,0,0,0)	Strong beliefs that there will be no events stronger than tropical storm	
Wrong Prior (Severe Hurricanes)	(1,1,1,1,100,100,100)	Strong belief in intense hurricane events	

`

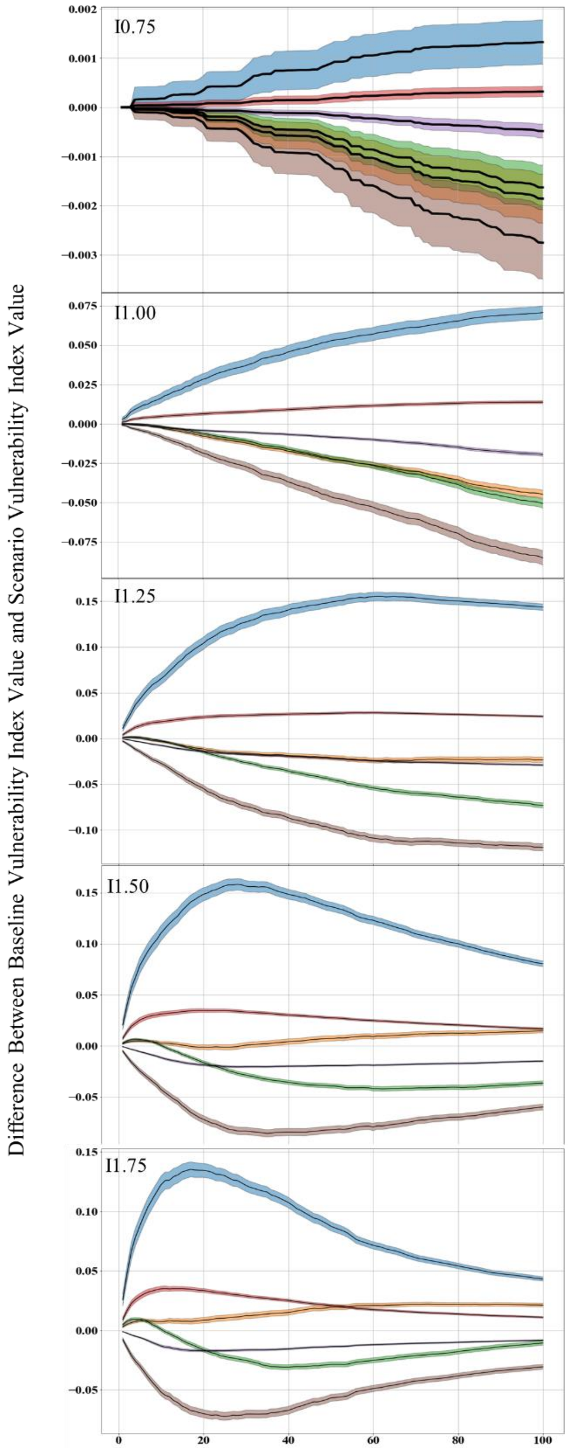


Figure 2-5 Differences in vulnerability for different initial priors compared with the baseline initial knowledge scenario over 100 years. Hurricane intensity scenario changes from 0.75, 1.00, until 1.75 times the baseline climate.

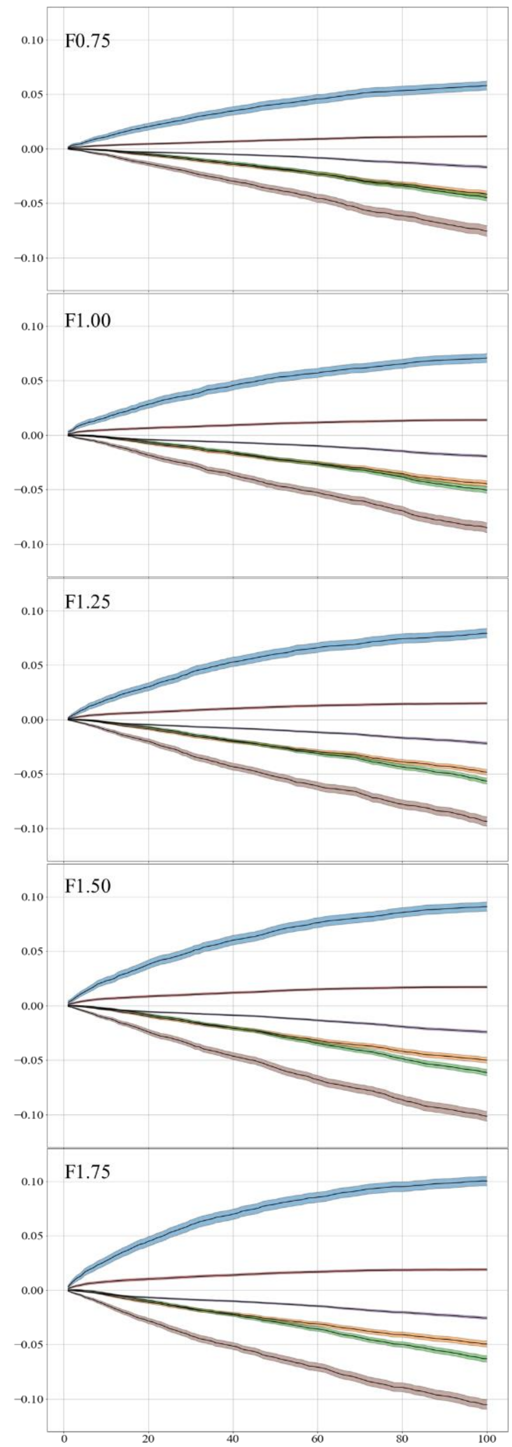


Figure 2-6 Differences in vulnerability for different initial priors comparing with the baseline initial knowledge scenario over 100 years. Hurricane frequency scenario changes from 0.75, 1.00, until 1.75 times the baseline climate.

`

The results of the influence of priors under different hurricane intensity scenarios are shown in Figure 2-5. The x-axis of each plot is time (years), and the y-axis is the difference between the mean VI for the baseline-priors case and the mean VI for a scenario using modified priors. When this difference is positive (negative), it means that this case resulted in a more (less) vulnerable community than the baseline-priors scenario. This difference can be substantial among different initial beliefs and climate scenarios.

Initial beliefs can influence agents' behaviors and community vulnerability substantially in the early years. For example, we found that under the climate scenario with hurricanes of intensity 1.75 times the initial climatic conditions, the largest difference between any two priors was more than 0.2, which is a large difference in the outcome when facing a same hurricane. When the climate intensity increases from the initial climatic conditions, cases with different initial priors begin to converge to a similar vulnerability level given a sufficient amount of time. Vulnerability levels initially diverge in earlier years and start to converge after the agents have time to be influenced by the environment. That is, we see the effects of learning over time. Stronger hurricanes speed up the effect of learning and reduce the effect of initial knowledge more quickly.

Another interesting finding is that holding strong beliefs about intense hurricanes initially (i.e., the "wrong priors case") does not necessarily imply the least vulnerability compared to other initial beliefs. The reason is that strong hurricane beliefs make the agents realize that, according to their beliefs, no matter what mitigation they undertake, damage is inevitable and

`

cannot be reduced by much given the mitigation options we modeled. Their investment on house upgrades then would not be worth the cost. On the other hand, holding strong beliefs about less intense hurricanes (i.e., the “Partially Strong 2 case”) impedes mitigation as well. This is because even an additional experience of a strong hurricane makes it difficult for agents to overcome their strongly held prior belief that hurricanes in the region are usually not intense. Only when the agents hold more neutral beliefs, i.e. Partially Uninformative 3 and 4 cases does the community become more hurricane-resistant.

Different hurricane frequency scenarios have similar rankings for different priors as the rankings for different hurricane intensities (Figure 2-6). However, the differences become larger when the frequency increases (note that the range on the y-axis differs between scenarios, so that we can zoom in and show differentiation among some results). The reason is that when more hurricanes strike the area, regions that are less likely to be impacted will be influenced at some point, and the agents will be more likely to mitigate, as in their perceived risk, the chance of no hurricane happening drops significantly. This reflects on each initial belief scenario and enlarge the gap when comparing high frequency and low frequency scenarios. In the frequency cases, initial knowledge held by the agents takes a bigger role in determining the patterns the evolution of vulnerability over time relative to the cases in which intensity was varied. The initial differences in beliefs create divergence in community vulnerability that does not converge by year 100. Therefore, the initial knowledge the agents remain important in the long run with changes in frequency.

`

2.7.3 Impact of Memory Effects on the Evolution of Community Vulnerability Under Different Climate Scenarios

Finally, we model the memory effects of agents as a decaying weight on their past experiences as discussed above. We chose the memory decay ratios to be 1 (no decay), 0.9, 0.8, 0.7, and 0.6. Such a process can be regarded as a result of memory fading and generation replacement. This process limits the agents to make decisions based on partial information from recent events instead of using a more complete record of events. As a result, we find that this memory process always promotes more mitigation and less vulnerability for any climate scenario. An example of this is shown in Figure 2-7. The results for other frequency and intensity scenarios are highly similar and are omitted for brevity. The blue line in both plots are corresponding to the green lines in Figure 2-4. We are using the same partially uninformative prior as the baseline scenario.

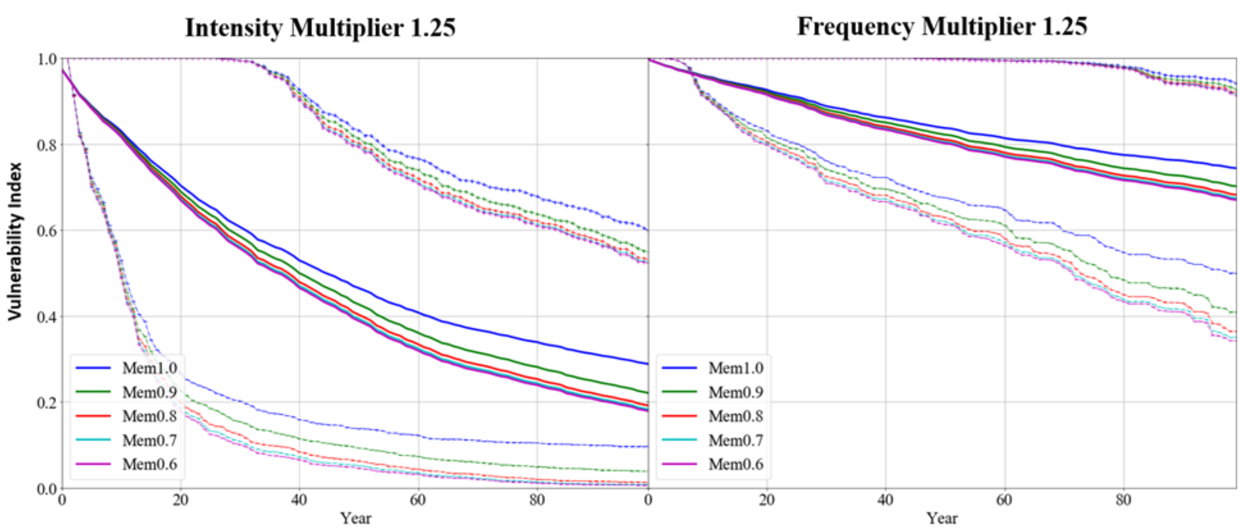


Figure 2-7 The evolution of community vulnerability under 1.25 intensity scenario (left) and 1.25 frequency scenario (right) with different memory decaying parameter. Each solid line represents a different memory decay parameter. Dashed line with the same color with '+' sign are the 85% upper bound of the output samples. Dashed line with the same color with '-' sign are the 15% lower bound of the output samples.

We see in Figure 2-7 that the faster the agents' past memory decays, the more mitigation that occurs and less vulnerable the community is. The explanation for this phenomenon is that most of the time no hurricane damage occurs for a given agent, and, if they have a high memory

`

decay, this experience of no hurricanes has a reduced impact in their updating. Then, when hurricane damage does occur, their beliefs change far more than if there was no memory decay. This in turn dramatically increases their perceived risk of hurricane damage, increasing the appeal of mitigation actions. That is, with the memory effect, agents are much more sensitive to recent hurricanes. This emphasizes the importance of better understanding the degree to which limited memory is a factor in learning from the effects of past hazards.

2.8 Conclusions

In this chapter, we analyzed the importance of individual learning, initial knowledge, and the effect of the strength of memory on the evolution of community vulnerability under different climate scenarios using an ABM. We found that incorporating the effect of learning is critical when simulating how community vulnerability may evolve under different adaptation and climate scenarios. In our model, more intense and frequent hurricane scenarios stimulated more mitigation and resulted in a less vulnerable community overall. Different initial knowledge held by the agents had an important role in determining the region's vulnerability, especially in early years. Limited memory can induce more mitigation because agents are more sensitive to recent experiences. A better understanding of the learning process of individual homeowners will lead to a better understanding of their behaviors, which will benefit decision-makers and policy-makers in long-term community vulnerability mitigation decisions. For example, when a community is anchored in their beliefs of few storms (i.e., partially strong priors – the blue line) they are more vulnerable. Getting people to realize that their prior beliefs are wrong – admittedly a challenging task – may be useful in reducing VI. On the other hand, as people are more likely to make mitigation decisions right after hurricanes, it can be more effective for government to provide subsidies on mitigation to help boost mitigation for vulnerable regions and reduce future losses.

`

More generally, the results in this chapter highlight the importance of understanding and modeling learning at an individual level when conducting a vulnerability assessment or risk analysis for a community facing the potential for repeated hazard events. Individual actions can substantially alter community vulnerability and learning from events is a critical part of this. Ignoring this learning effect can lead to substantial mis-estimation of future risk.

2.9 References

1. Reuters. Hurricane Harvey Damages Could Cost up to $180 Billion. Fortune, 2017.
2. White MC. Why Irma's Impact on the Economy Will Be Worse than Harvey's. NBC News, 2017.
3. Zahran S, Weiler S, Brody SD, Lindell MK, Highfield WE. Modeling national flood insurance policy holding at the county scale in Florida, 1999-2005. Ecological Economics, 2009; 68(10):2627–2636.
4. Dixon L, Clancy N, Seabury S, Overton A. The National Flood Insurance Program’s Market Penetration Rate: Estimates and Policy Implications. The National Flood Insurance Program’s Market Penetration Rate: Estimates and Policy Implications, 2018.
5. Hamel L, Wu B, Brodie M, Family K, Sim S-C, Marks E. One Year After the Storm: Texas Gulf Coast Residents' Views and Experiences with Hurricane Harvey Recovery The Kaiser Family Foundation/Episcopal Health Foundation Harvey Anniversary Survey., 2018.
6. Jain VK, Davidson RA. Forecasting changes in the hurricane wind vulnerability of a regional inventory of wood-frame houses. Journal of Infrastructure Systems, 2007; 13(1):31–42.
7. Jain VK, Davidson R, Rosowsky D. Modeling changes in hurricane risk over time. Natural Hazards Review, 2005; 6(2):88–96.
8. Sadri AM, Ukkusuri S V., Gladwin H. Modeling joint evacuation decisions in social networks: The case of Hurricane Sandy. Journal of Choice Modelling, 2017; 25:50–60.
9. Lien C, Raimo J, Abramowitz J, Khanijo S, Kritharis A, Mason C, Jarmon CH, Nash IS, Carney MT. Community healthcare delivery post-hurricane sandy: Lessons from a mobile health unit. Journal of Community Health, 2014; 39(3):599–605.
10. Smallegan SM, Irish JL, Van Dongeren AR, Den Bieman JP. Morphological response of a sandy barrier island with a buried seawall during Hurricane Sandy. Coastal Engineering, 2016; 110:102–110.
11. Reilly AC, Tonn GL, Zhai C, Guikema SD. Hurricanes and Power System Reliability-The Effects of Individual Decisions and System-Level Hardening. Proceedings of the IEEE, 2017; 105(7):1429–1442.

`

12. Logan TM, Guikema SD, Bricker JD. Hard-adaptive measures can increase vulnerability to storm surge and tsunami hazards over time. Nature Sustainability, 2018; 1(9):526–530.

13. Staid A, Guikema SD, Nateghi R, Quiring SM, Gao MZ. Simulation of tropical cyclone impacts to the U.S. power system under climate change scenarios. Climatic Change, 2014; 127(3–4):535–546.

14. An L. Modeling human decisions in coupled human and natural systems: Review of agent-based models. Ecological Modelling, 2012; 229(0):25–36.

15. Peacock WG. Hurricane Mitigation Status and Factors Influencing Mitigation Status among Florida’s Single-Family Homeowners. Natural Hazards Review, 2003; 4(3):149–158.

16. Burnside R, Miller DMS, Rivera JD. The impact of information and risk perception on the Hurricane evacuation decision-making of greater New Orleans residents. Sociological Spectrum, 2007; 27(6):727–740.

17. Tversky A, Kahneman D. Availability: A heuristic for judging frequency and probability. Cognitive Psychology, 1973; 5(2):207–232.

18. Liu F, Pang W. Projections of future US design wind speeds due to climate change for estimating hurricane losses. ProQuest Dissertations and Theses, 2014:275.

19. Orooji F, Friedland CJ. Cost-Benefit Framework to Generate Wind Hazard Mitigation Recommendations for Homeowners. Journal of Architectural Engineering, 2017; 23(4):1–11.

20. Pinelli JP, Simiu E, Gurley K, Subramanian C, Zhang L, Cope A, Filliben JJ, Hamid S. Hurricane damage prediction model for residential structures. Journal of Structural Engineering, 2004; 130(11):1685–1691.

21. Chung Yau S, Lin N, Vanmarcke E. Hurricane damage and loss estimation using an integrated vulnerability model. Natural Hazards Review, 2011; 12(4):184–189.

22. Van De Lindt JW, Nguyen Dao T. Loss analysis for wood frame buildings during hurricanes. II: Loss estimation. Journal of Performance of Constructed Facilities, 2012; 26(6):739–747.

23. Wang C, Zhang H, Feng K, Li Q. Assessing hurricane damage costs in the presence of vulnerability model uncertainty. Natural Hazards, 2017; 85(3):1621–1635.

24. Pinelli JP, Torkian BB, Gurley K, Subramanian C, Hamid S. Cost effectiveness of hurricane mitigation measures for residential buildings. 11th Americas Conference on Wind Engineering, 2009.

25. Rose A, Porter K, Dash N, Bouabid J, Huyck C, Whitehead J, Shaw D, Eguchi R, Taylor C, McLane T, Tobin LT, Ganderton PT, Godschalk D, Kiremidjian AS, Tierney K, et al. Benefit-cost analysis of FEMA hazard mitigation grants. Natural Hazards Review, 2007; 8(4):97–111.

26. English EC, Friedland CJ, Orooji F. Combined Flood and Wind Mitigation for Hurricane Damage Prevention: Case for Amphibious Construction. Journal of Structural Engineering (United States), 2017; 143(6):06017001.

27. Reilly AC, Guikema SD, Zhu L, Igusa T. Evolution of vulnerability of communities facing repeated hazards. Linkov I (ed). PLoS ONE, 2017; 12(9):e0182719.

28. Colten CE, Sumpter AR. Social memory and resilience in New Orleans. Natural Hazards, 2009; 48(3):355–364.

29. Bonabeau E. Agent-based modeling: Methods and techniques for simulating human systems. Proceedings of the National Academy of Sciences of the United States of America, 2002; 99(SUPPL. 3):7280–7287.

30. MacAl CM, North MJ. Tutorial on agent-based modelling and simulation. Journal of Simulation, 2010; 4(3):151–162.

31. Gorman DM, Mezic J, Mezic I, Gruenewald PJ. Agent-based modeling of drinking behavior: A preliminary model and potential applications to theory and practice. American Journal of Public Health, 2006; 96(11):2055–2060.

32. Billari FC, Fent T, Prskawetz A, Scheffran J. Agent-Based Computational Modelling: Applications in Demography, Social, Economic and Environmental Sciences., 2006.

33. Reilly AC, Dillon RL, Guikema SD. Agent-Based Models as an Integrating Boundary Object for Interdisciplinary Research. Risk Analysis, 2018.

34. Chen X, Zhan FB. Agent-based modelling and simulation of urban evacuation: Relative effectiveness of simultaneous and staged evacuation strategies. Journal of the Operational Research Society, 2008; 59(1):25–33.

35. Chen X, Meaker JW, Zhan FB. Agent-based modeling and analysis of hurricane evacuation procedures for the Florida Keys. Natural Hazards, 2006; 38(3):321–338.

36. Pan X, Han CS, Dauber K, Law KH. A multi-agent based framework for the simulation of human and social behaviors during emergency evacuations. AI and Society, 2007; 22(2):113–132.

37. Tonn GL, Guikema SD. An Agent-Based Model of Evolving Community Flood Risk. Risk Analysis, 2018; 38(6):1258–1278.

38. Breen R. Beliefs, rational choice and Bayesian learning. Rationality and Society, 1999; 11(4):463–479.

39. Panait L, Luke S. Cooperative multi-agent learning: The state of the art. Autonomous Agents and Multi-Agent Systems, 2005; 11(3):387–434.

40. Farmer JD, Foley D. The economy needs agent-based modelling. Nature, 2009; 460(7256):685–686.

41. Johnson MK, Hasher L. Human Learning and Memory. Annual Review of Psychology, 1987; 38(1):631–668.

42. Du E, Cai X, Sun Z, Minsker B. Exploring the Role of Social Media and Individual Behaviors in Flood Evacuation Processes: An Agent-Based Modeling Approach. Water Resources Research, 2017; 53(11):9164–9180.

43. Krause T, Beck E V., Cherkaoui R, Germond A, Andersson G, Ernst D. A comparison of Nash equilibria analysis and agent-based modelling for power markets. International Journal of Electrical Power and Energy Systems, 2006; 28(9 SPEC. ISS.):599–607.

44. Tsurushima A. Modeling Herd Behavior Caused by Evacuation Decision Making Using Response Threshold. Pp. 138–152 in Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). Vol 11463 LNAI. Springer Verlag, 2019.

45. Kahneman D, Tversky A. Prospect theory: An analysis of decision under risk. Pp. 143–172 in Experiments in Environmental Economics. Vol 1. 2018.

`

46. Friedman D, Rubinstein A. Modeling Bounded Rationality., 1998.

47. Burchfield EK, Gilligan JM. Dynamics of individual and collective agricultural adaptation to water scarcity. Proceedings - Winter Simulation Conference, 2016; 0:1678–1689.

48. Bouris E. Overconfidence and War: The Havoc and Glory of Positive Illusions Overconfidence and War: The Havoc and Glory of Positive Illusions . By Dominic D.P. Johnson. (Harvard University Press, 2004.) Balancing Risk: Great Power Intervention in the Periphery . B., 2006.

49. Rubinstein A. Similarity and decision-making under risk (is there a utility theory resolution to the Allais paradox?). Journal of Economic Theory, 1988; 46(1):145–153.

50. Walsh KJE, Camargo SJ, Vecchi GA, Daloz AS, Elsner J, Emanuel K, Horn M, Lim YK, Roberts M, Patricola C, Scoccimarro E, Sobel AH, Strazzo S, Villarini G, Wehner M, et al. Hurricanes and climate: The U.S. Clivar working group on hurricanes. Bulletin of the American Meteorological Society, 2015; 96(6):997–1017.

51. Meehl GA, Tebaldi C. More intense, more frequent, and longer lasting heat waves in the 21st century. Science, 2004; 305(5686):994–997.

52. Mendelsohn R, Emanuel K, Chonabayashi S, Bakkensen L. The impact of climate change on global tropical cyclone damage. Nature Climate Change, 2012; 2(3):205–209.

53. Knutson TR, Tuleya RE. Impact of CO2-induced warming on simulated hurricane intensity and precipitation: Sensitivity to the choice of climate model and convective parameterization. Journal of Climate, 2004; 17(18):3477–3495.

54. Holland GJ. An analytic model of the wind and pressure profiles in hurricanes. Monthly Weather Review, 1980; 108(8):1212–1218.

55. Vickery PJ, Skerlj PF, Lin J, Twisdale LA, Young MA, Lavelle FM. HAZUS-MH hurricane model methodology. II: Damage and loss estimation. Natural Hazards Review, 2006; 7(2):94–103.

56. Plotner SC, R.S. Means Company. Building Construction Costs with RSMeans Data 2017., 2017.

57. Siegrist M, Gutscher H. Natural Hazards and Motivation for Mitigation Behavior: People Cannot Predict the Affect Evoked by a Severe Flood. Risk Analysis, 2008; 28(3):771–778.

58. Esteban M, Valenzuela VP, Matsumaru R, Mikami T, Shibayama T, Takagi H, Thao ND, De Leon M. Storm Surge Awareness in the Philippines Prior to Typhoon Haiyan: A Comparative Analysis with Tsunami Awareness in Recent Times. Coastal Engineering Journal, 2016; 58(1).

59. Anand P. Foundation of Rational Choice under Risk / Paul Anand. Oxford University, 1993.

60. Nelson BL, Matejcik FJ. Using Common Random Numbers for Indifference-Zone Selection and Multiple Comparisons in Simulation. Management Science, 1995; 41(12):1935–1945.

61. Han SR, Guikema SD, Quiring SM, Lee KH, Rosowsky D, Davidson RA. Estimating the spatial distribution of power outages during hurricanes in the Gulf coast region. Reliability Engineering and System Safety, 2009; 94(2):199–210.

Chapter 3: Evolution of Community Vulnerability Under Repeated Heat Waves

3.1 Abstract

Repeated heat waves impact vulnerable populations such as the elderly and lower income groups that are less able to respond to high heat events. The vulnerability of these and other groups to heat waves changes over time in response to social connections, collective action, utility coordination, and government policies. This chapter presents the results of an agent-based modeling framework that integrates a heatwave model, dynamically evolving agents, social connections, and a health mortality model to study the impact of repeated heatwaves on society. We use a case study based on Baltimore, MD, USA to examine the regional dynamics and the evolution of regional vulnerability in the face of repeated heatwave hazards in a realistic setting. We find that a change of climate can influence the long-term vulnerability of a community in surprising ways. If daily average temperatures increase, it can stimulate more individual mitigation behaviors and reduce the number of people that die from future heat waves. Social networks can also have a substantial impact in helping decrease mortality by providing options for cooling at friends' houses. On the other hand, cooling centers can be used effectively in helping decrease mortality for a given event, but they may also increase the long-term vulnerability by decreasing people's motivation to adopt mitigation, which increases the risks when facing an unexpected or extreme event. Only by modeling adaptive behavior and social

`

interactions over time can we start to understand how community vulnerability evolves over time.

3.2 Introduction

Extreme heat events can cause severe consequences to the economy, infrastructure, and human lives. In 1995, a deadly five-day long heatwave struck Chicago causing 739 heat-related deaths. Many factors aggravated the catastrophe such as the lack of emergency response planning, an aging population, poverty, and power outages [(1)]. The majority of deaths occurred among older people who were socially isolated [(2)]. In 2020, multiple high temperature events in California aggravate wildfires that burned millions of acres, put pressure on power system, and created an unpleasant and unhealthy work-from-home environment coupled with Covid-19. Heat waves have become a hazardous event we cannot neglect in this century.

One of the most effective mitigation measures to prepare for heat waves is to invest an air conditioning (A/C). However, this option is not available to all individuals, especially lower-income people. Furthermore having air conditioning does not necessarily mean the owner can afford the power to operate it [(3)]. Both power outages and utility demand management programs can also block A/C functionality and cause extensive and unexpected exposure to prolonged heat. On the other hand, the government can open cooling centers and extend public pool hours during extreme heat to help mitigate the impacts. While there are studies discussing the vulnerability of communities facing heat waves, most are (1) focused on a single hazard event, i.e., the 1995 Chicago heat wave or the 2003 Europe heat wave (e.g., Changnon et al. 1996; De Bono et al. 2004) and (2) ignore the dynamics of social networks and individual behaviors that change the vulnerability of communities. The long-term impact from repeated heat waves to

`

individuals, communities, and infrastructures is obscure and there is a lack of quantitative studies that examine how adaptive behavior may change vulnerability to future events. This introduces difficulty for individual and government decision makers as they seek to make better preparation for future heat wave events, especially in regions where the climate is shifting towards higher temperatures.

There have been numerous studies of single heat wave events in terms of their impacts, outcomes, and mitigation actions. As an example, in response to the extreme heat wave that happened in Chicago in 1995, there have been studies discussing the cause and relationship to climate [(5)], the possibility of recurrence [(6)], the importance of social connections to individual vulnerability [(2)], and the preparation and response to the event [(7)]. These studies diagnosed the event in depth from a specific perspective and potentially helped decrease the vulnerability of Chicago region when facing heat waves again. However, the vulnerability changes induced by individual behaviors and local climate, which are both dynamic, are hard to capture from studies of a single event.

To understand the long-term vulnerability of a community facing heat waves, we need to consider three major mitigation sources. The first one is individual mitigation decisions. The most effective action one can take is to invest in A/C units [(8)]. However, the reasons for and timing of A/C purchases have not been well-studied. Another mitigation source that is non-neglectable during heat waves is the impact of social capital and social networks. Social capital is an important factor in disaster response, but its mechanism and influence remain unclear [(9)]. It remains unclear in which circumstances social capital may decrease vulnerability. Previous studies indicate associations between social capital and aggregate health [(9–11)]. Some studies, however, suggest that social capital does not have an influence on individual health [(12,13)]. In this

`

study, we focus on a subset of social capital and explore how social networks may influence health risks under repeated heat waves.

Lastly, government response efforts can be crucial in helping mitigation. The government can open cooling centers and extend public building hours to decrease heat exposure. However, Naughton et al. suggest that cooling centers were not used effectively during the heat wave in Chicago 1999 and needed better explanation to the residents [(14)].

The purpose of this chapter is to examine the compound effects of individual mitigation decisions, social networks, climate change, and power outages on the evolution of community vulnerability and heat-related mortality under repeated heat waves. This work also provides a framework for studying the impact of government policies, market incentives, and power demand reduction programs in combating heat wave losses.

Our analysis is based on an agent-based model (ABM) which incorporates a heat wave model, individual learning and mitigation models, a social network model, and a heat wave mortality model. In the model, each individual has social connections and makes mitigation decisions (e.g., invest an A/C) with their household members. The approach is place-based, with a case study area based on 594,077 residents in Baltimore city, Maryland. However, both the models and results are generalizable to other areas with similar data inputs.

The structure of the remainder of this chapter is as follows. We first review relevant studies from two different areas: (1) heat waves and their health impacts, and (2) agent-based modeling of community vulnerability. Next we describe the model used in this study and discuss its calibration to real-world data. Then we analyze how climate, social networks, and power outages

`

change the evolution of community health risk from heat waves in our study region. We close with concluding thoughts and suggestions for future research.

3.3 Literature review

3.3.1 Heat wave vulnerability and mitigation

In many years, heat waves are the nation's most lethal form of extreme weather [15]. Repeated heatwaves, defined as heatwaves repeatedly happening in the same geographic region, have the potential to change hazard response behavior of residents, affect long-term investments in heat wave preparation, and alter emergency response policy. Repeated events potentially lead to profound effects on the vulnerability of a region. Repeated exposure to these extreme weather conditions could be harmful to individual health [16]. However, studies of the impacts of repeated hazards, especially heat waves, are limited. Previous studies on repeated impacts focus on discipline-specific prospective on the problem rather than addressing the problem in an integrated, interdisciplinary way [16].

Heat waves can cause substantial public health impacts within the communities they hit. Very extreme heat waves, while rare, can cause substantial increases in mortality[17]. Examples include a July 1995 heat wave in Chicago, associated with nearly 800 excess deaths and with significant increases in heat related illness straining limited public health and hospital resources[18,19]. Others include the 2003 heat wave in France and the 2010 heat wave in Russia, estimated to have been associated with 15,000 and 55,000 excess deaths, respectively [20]. Slightly less severe, but more frequent, heat waves, though associated with lower overall risk of mortality, likely also contribute to an important cumulative burden over time[17].

`

Heat waves can affect certain subpopulations more than others, and the elderly have consistently been identified as a higher risk group for these effects. For example, in a systematic review including 61 studies on heat wave vulnerability and mortality, Benmarhnia et al. found significantly increased risk in older age groups[21]. Elderly populations continue to be included as an at-risk population in heat action plans around the world[22].

However, studies have suggested various public health interventions including increasing access to air-conditioned shelters [23,24] and increased social contact during extreme heat events[24] as protective during heat wave events. Of these protective factors, having a home air conditioning unit has consistently been identified as the single most protective factor during extreme heat events[14,18,24–27].

3.3.2 Agent-based model (ABM) of community vulnerability

An ABM is a bottom-up approach used to simulate a complex system where heterogeneous agents can make decisions and interact with each other and the environment dynamically [28,29]. The primary benefit of an ABM framework is that it allows researchers to create a comprehensive collection of scenarios to conduct sensitivity analysis which facilitates answering what-if questions under different circumstances [30]. Compared to top-down approaches, another benefit of an ABM is its revelation of compounding effects and interactions [31]. However, results from an ABM could be challenging to validate and verify due to untraceable human behaviors and unprecedented scenarios [30]. Output analysis can also be complex, because ABM output is often high-dimensional and requires sophisticated analytical approaches [32]. Despite these challenges, an ABM can be a useful tool to promote understanding of how a complex system operates as a result of different factors.

`

ABMs have seen increasing use in modeling natural disaster-related issues in recent years. Most work has centered on analyzing short-term actions after the hazard happens to demonstrate how agents would evacuate, take shelter, or access primary care services and discuss the potential flaws and improvements [33–35]. On the other hand, ABMs have more recently been used to model long-term effects caused by natural disasters [31,36,37]. For example, Reilly et al. focuses on repeated hurricane events and shows that individual mitigation actions (purchasing generators) can lead to less system-level investment in system hardening, leaving those less able to afford individual action worse off[37].

The composition of an ABM in longer-term vulnerability studies of natural disasters takes five basic components: a definition of agents, an environment model, a learning or risk perception model, a behavioral model, and an interaction model. We briefly review how each component is modeled in the literature. Agents are heterogeneous units that can make decisions autonomously [38]. Agents can be parcels [39], single homeowners [31], or any other decision making unit. The agents can learn from the environment to update their perception of the risks of the hazard [36,37]. The process can be viewed as a learning process with new information acquired by the agents[31,36]. The agents then make mitigation decisions based on their assessment of the risk and their characteristics, typically including their wealth as a constraint. There are three main classes of decision-making models (1) prescriptive, (2) descriptive, and (3) expert or data based if-then models. A detailed review of these models can be found in Chengwei et al. 2020. Among these models, an if-then model is particularly flexible and can be incorporated if properly tuned [36,37].

There are generally two types of agent interactions that must be modeled in a hazard-focused ABM, the interactions of the agents with the environment and the interactions of the agents with

`

each other. One of the key agent-environment interactions in a hazard-focused ABM is how the agents experience the hazard, the damage they experience, the information they receive about that damage, and related issues. Generally, agent actions can modify these impacts through, for example, pre-event mitigation or post-event response. The second key interaction is of the agents with each other, though this is also a key to non-hazard ABMs [40–42].

3.4 Methodology

3.4.1 Overview and model assumptions

We model the impacts of repeated heatwaves on residents in a city based on but not identical to Baltimore, MD in the U.S. This helps ground the model in reality. Our agents are individuals in the city, each living in an assigned household. We simulate how the vulnerability of the region, defined more precisely below, may evolve under repeated heatwaves and how individuals make decisions to mitigate heatwaves over time. The agents living in the same household make annual mitigation decisions. These decisions are driven by the perception of heatwave risk and have the potential to reduce the household's future exposure to extreme heat. We focus on the purchase of air conditioning as the household-level mitigation decision as this is a main way to mitigate the effects heatwaves at the level of an individual house. Agents also may have access to cooling by being invited to the home of someone in their social network who has air conditioning, and we model this process probabilistically. Air conditioning thus serves as a modifier of risk both for an individual household and by influencing the likelihood that they have access to air conditioning in their social network.

In this chapter we build from the general framework for agent-based modeling of repeated hazards [36,37]. In this framework we need to identify five key elements to implement the model:

`

(1) agents, (2) environment, (3) adaptation, (4) behaviors, and (5) interactions. Our model is based on Baltimore City. That is, we model each notional household in Baltimore, assigning each synthetic household a set of agents with ages such that (1) the distribution of agent ages matches that of the census tract and (2) the age structure of households is reasonable (e.g., all minors must have at least one adult in the same household). We then develop a synthetic social network for each agent, with details provided below.

We model the heatwave environment based on the historical weather data for Baltimore and apply daily temperature uniformly to each agent. That is, we do not explicitly model within-day spatial temperature variability because this data was not available historically, though we acknowledge that this variability can be important [(43)].

We also model power outage events, which can prevent those homeowners that have air conditioning from being able to use it. In the simulation, power outages happen as rare events to account for the reality that during hot weather power outages can occur as a result of overloading and heat-induced equipment failures. We model how agents learn from heat wave events and how they make mitigation decisions. Finally, we estimate mortality in the community as a result of exposure to high temperatures given each heat wave, corrected for air conditioning.

The flowchart of the simulation is shown in Figure 3-1. We start with initializing parameters and agents. The simulation model then enters an annual loop for each replication. For each year, we first determine if there is a heat wave in the given year. If there is a heat wave, then for each heat wave we first determine if there is a power outage and who will lose power. If there is no heat wave, the simulation proceeds to the next year.

`

If there is a heatwave in a given year, the model first determines if each agent will be exposed to the heat and, if so, how many days each of them will be exposed. An agent will be exposed to a day of heat wave if the household does not own an air conditioning, if they do not get invited by to the home of someone in their social network that has air conditioning, and they do not get a spot in the cooling center, an occurrence that is modeled stochastically. Next, we determine the number of deaths from the heat wave based on the exposure of individuals. At the end of each simulated year agent household decide if they want to purchase air conditioning based on a decision model described below. Later in this section, we will explain each of these portions of the model in detail.

Throughout this chapter, it is important to keep in mind that we are not trying to model the actual development or functionality of a particular city or its demographic evolution. We created a simplified version of the real Baltimore by only including elements that are most essential to answer our proposed research question regarding the vulnerability of community under heatwaves and the impact of social interactions. This allows us to keep the model grounded in a realistic case study.

`

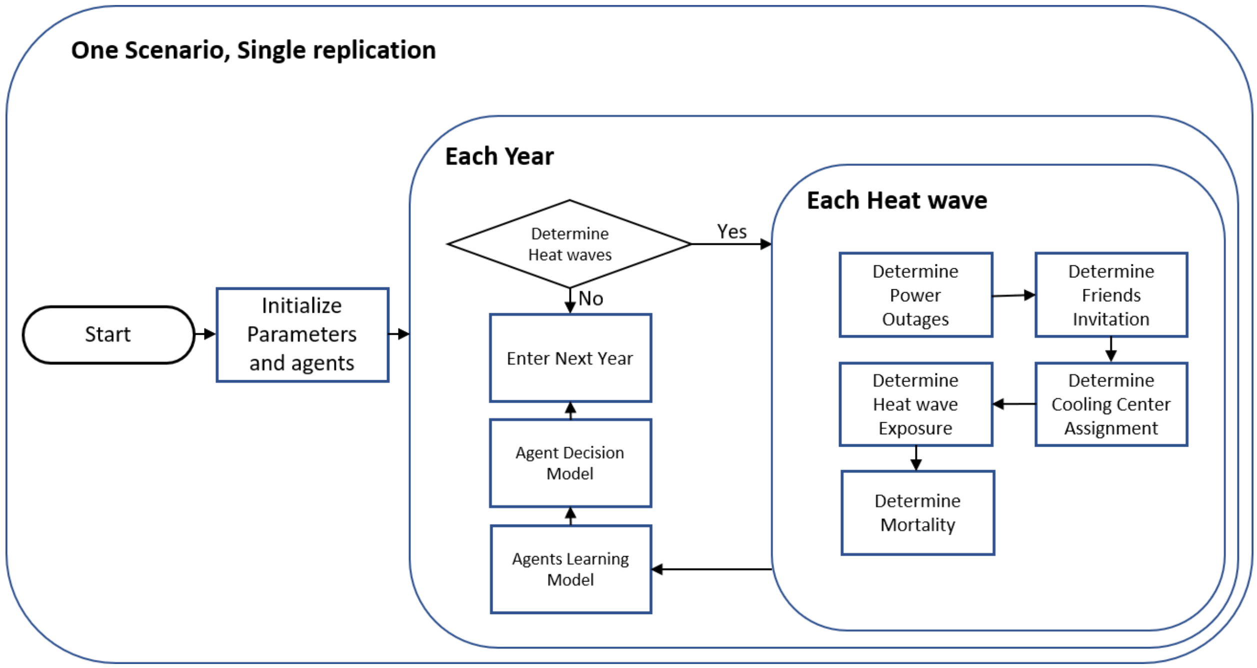


Figure 3-1 Simulation framework of the Agent Based model

There are three key areas in which we make assumptions – demographic changes, heat wave definition, and individual decision making. The first area in which we make simplifying assumptions is that the demographics and social networks remain unchanged over the 30 years in each simulation run. This means that there is no aging, births, deaths, immigration or emigration. However, as the A/C decision made at each house is inherited within each simulation, the primary assumption is that the distribution of each age group remains the same over the simulation period. This allows us to focus on the effects of social networks and individual mitigation decisions, holding demographics, which are difficult to forecast, constant.

The second area in which we make an assumption is in the definition of a heat wave. Defining a heat wave is a known research challenge [(44)]. In our study, we define a heatwave as two or more consecutive days when the average temperature is higher than the 97th quantile of historical temperature of a city as it gives us a sufficient number of events while still having each one to be a significant event.

`

We use Baltimore city, Maryland as the study region to collect input data for our simulation. In 2016, the population of Baltimore city was 616,958 and the number of single-family households was approximately 250,000. In our study, one of the most important population groups is those who are 65 and above because they are more at risk from heat events. The proportion of the Baltimore population over 65 years old has remained stable, varying from 13.7% in 1990[2] to 13.2% in 2010[3], which is also a justification of our simplification of the agent's aging model. The mortality model that we use accounts for resident age. A summary of the high temperature events we found for Baltimore city is shown in Table 3-1.

Table 3-1 Summary of High Temperature Events in Baltimore from 1987 to 2017

Number of High Temperature Events	Average Length (days)	Average mean temperature
60	4.41	85.0 F

The last area in which we make assumptions is in the area of how individuals make decisions. We assume that a household will decide to buy an air conditioner while considering the wealth of each individual in the family, their experiences from heat waves, and a majority vote from the family members. The details of this model are provided below.

[2] https://planning.maryland.gov/MSDC/Documents/pfa/1990Census/sf1b/baci_90pfap1.pdf
[3] https://www.census.gov/quickfacts/fact/table/baltimorecitymaryland,US/POP010210

`

3.4.2 Agents

We define the agents in our model to be each resident living in the city. Each agent in our model is assigned a home location, age, gender, race, education level, and income, each randomly generated based on tax assessor information and census data. The data generation scheme is illustrated in Figure 3-2.

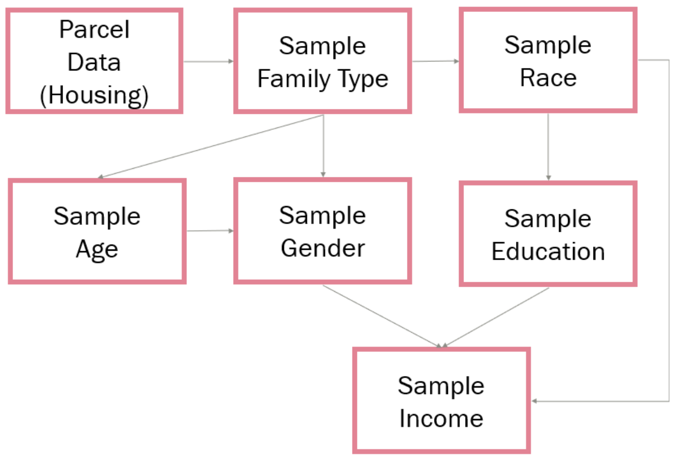


Figure 3-2 Overview of agent characteristic generation process

We start by sampling the type of family living in each home where family type includes the size of the family and the combination of family members living in the home (e.g., single-member family, family with kids, single mom with kids). From the size of each family, we then sample the age and gender for each family member based on the distributions of age and gender in the census tract. Given the spatial location of houses, we then sample the race of each agent based on census tract information. Education levels (high school degree, bachelor's degree, and master's degree and higher) of the household residents are then sampled based on census data. Finally, the income for each individual is sampled based on their gender, education, and race from the census distribution. The importance of generating these attributes is that people who share a similar demographic are more likely to be socially connected in our social network

`

model. Moreover, people with similar backgrounds tend to make similar decisions, which are influential to the vulnerability of them as a group, and such similarity can lead to very different heatwave experiences.

3.4.3 Social Network Model

To generate a social network, we used Exponential Random Graph Models (ERGM), where the probability of the existence of a connection between a pair of vertices is assumed to depend only on the states of the network [(45)]. We use Y_{ij} to represent the state of the dyadic connection of two individuals i and j, where $Y_{ij} = 1$ if there is a connection between these two persons and $Y_{ij} = 0$ otherwise. Individuals in the same household are assumed to be all connected with each other.

In an ERGM the probability that two vertices are connected is a function of the network structure and nodal attributes [(46)]. We let X be the set of nodal attributes (e.g., age, sex, wealth) for each vertex in the network. In ERGM, the relationship between a graph realization y and a data set X is modeled as:

$$P(Y = y|X, \theta) = \frac{e^{\theta^T s(X,y)}}{c} \tag{1}$$

where $s(X, y)$ is a user-defined function that depends on the observed network and nodal attributes, θ is a vector of model parameters associated with $s(X, y)$ and c is a normalizing constant such that the sum of all probabilities is 1. Then we use this probability as the parameter of a Bernoulli distribution to sample if two agents are connected.

Due to the large number of agents, the calculation of the probability of all pairs of agents is computationally infeasible. To overcome this, we first use K-means to cluster the agents based

`

on their spatial distribution (Figure 3-3). Although an agent can have more distant social connections, we assume that during a heat waves the agent will be more likely to seek help from friends who live nearby. We set the number of clusters of each social network group to be 60, which means each cluster has an average 10,000 agents as this gives a relatively fast generation speed. Within each cluster, we then used the ERGM network generation algorithm to estimate the probabilities of each agent being connected with each other agent in that cluster and used these probabilities to sample the social networks.

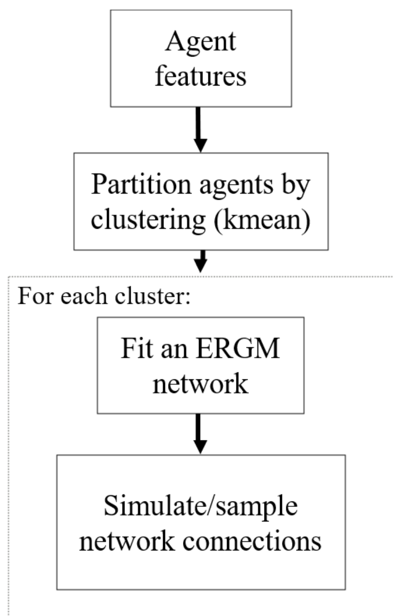


Figure 3-3 Generate social network

3.4.4 Learning model

We use a Bayesian updating model to represent the learning process of each agent in a manner similar to Reilly et al. (2017) . We use the number of unmitigated heat wave days that an agent is exposed to as the key variable they have beliefs about and learn about. We assume the agent's belief about the length of an agent's heatwave exposure X follows a Poisson distribution

`

$Poisson(\lambda)$ and we use a $Gamma(\alpha, \beta)$ prior on λ. Let X_{new} be the number of heat wave days an agent experienced without air conditioning (their own or at a social connection's house) in a given year. Then we can estimate the posterior mean as $\hat{\lambda}_{new} = \frac{\alpha_{old}+X_{new}}{\beta_{old}+1}$, where α_{old} and β_{old}represent the agent's beliefs prior to the given year. This approach provides an efficient approach for updating agent beliefs about the distribution of the number of heat wave days they will experience in a given year.

3.4.5 Behavioral model

The key decision modeled in our ABM is the agent-level decision of whether or not to purchase air conditioning for their home. We model this as a household-level decision considering the wealth of the household. However, there are no existing studies that we are aware of that explain how households make this purchasing decision. Our assumption is that the purchase decision is related to (1) past heat wave experiences, (2) financial status of the family, and (3) the opinion of each adult family member. We use an if/then model for air conditional adoption, calibrated to historical air conditioning coverage for our study region with American Housing Survey[4]. This model is summarized in Figure 3-4 and described in detail below.

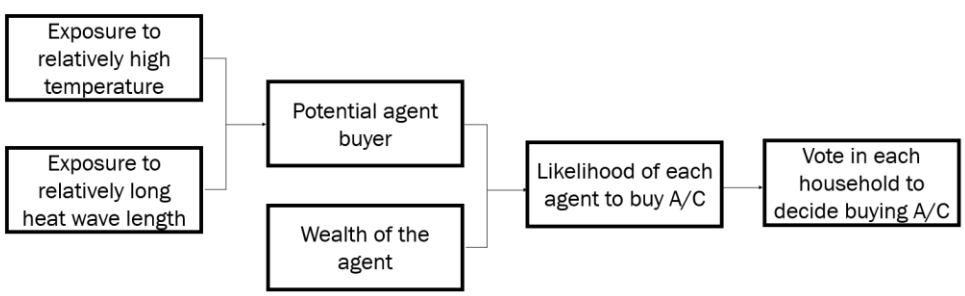


Figure 3-4 Air conditioner purchase decision model

[4] https://www.census.gov/programs-surveys/ahs.html

`

In our model, the decision of whether or not to buy air conditioning is made at the household level, and we assume that this decision is made based on a vote of the adult agents in the household. An agent in a household becomes a potential buyer when their perception of heatwave length and temperature pass a threshold. Only potential buyers can vote yes. The probability of voting yes is modeled as a function of income as follows,

$$\mathrm{p}_{\mathrm{vote}} = \frac{I}{MI_{median}} \tag{2}$$

where $\mathrm{p}_{\mathrm{vote}}$ is the probability the agent will vote to purchase, I is the income of the agent, and m is a parameter that we call the motivation parameter. We tune M to calibrate the model with historical records. I_{median} is the median income of the region. The wealthier the agent is, the more likely they will want to buy air conditioning. Next, we sample in each household using the probability each agent will vote YES to decide if the household will buy air conditioning given the majority votes.

3.4.6 Social Invitation Model

For agents who do not have air conditioning in their houses, they can be invited by their friends who do have air conditioning, and this serves to reduce their exposure to high heat. The interaction between households within a social network then is a critical aspect of the risk that agents face. The model for creating the social networks was described above. Here we describe our model for agent interactions within their network.

The likelihood of an agent being invited to the house of someone who has air conditioning is $wN_{A/C}/N_{Total}$, where $N_{A/C}$ is the number of friends with air conditioning and N_{Total} is the total number of friends, and w is the willingness of invitation, which is uniform in the model for all

`

agents. To be noted, the total number of days an agent can be invited cannot exceed the maximum number of friends in their network.

3.4.7 Government mitigation aids (Cooling Center)

Cooling centers are one of the most effective government measures to mitigate heat wave risk [47]. In the ABM, the use of cooling centers is one of the scenarios we test to examine the impact on individual behavior. We do this through a simplified, exogenously-imposed cooling center model. For each heat wave day, we assume a fixed number of spaces N_c every day that are open to the public from the cooling center, and cooling center will always have power. We then randomly select agents who do not own air conditioning and do not get invited to someone else's house to go to the cooling center. While cooling centers can effectively reduce the number of residents exposed to heat waves in short term, they may decrease the motivation to buy air conditioning, which in long term could have an impact to the vulnerability of society.

3.4.8 Power outage model

Power outages during heatwaves can be fatal. Losing power means that the primary individual mitigation, air conditioning, is no longer helpful. During heat events, power outages can happen from equipment failures due to high temperatures or because of power system overloads from increased demand due to air conditioner operation. There is thus a potential tradeoff between air conditioning prevalence and power outages during heat waves. We model

`

power outages during heatwaves from two different spatial levels: (1) distribution transformer level and (2) distribution substation level.

A transformer outage occurs when a local distribution system transformer is either overheated or overloaded, and such an event leads to loss of power to a relatively small number of customers in a localized area. Based on a previous project with a large utility that wishes to remain anonymous, we estimated that the likelihood of any one transformer failing due to prolonged high temperatures of 100 F is approximately $1x10^{-5}$. However, this value is not critical, and sensitivity analysis can be used to examine its impact. We assume 10-15 houses are served by each distribution transformer, a typical value in many urban power systems.

Larger scale power outages can be triggered by the failure of distribution substations. They may loose functionality due to the high demand from households or due to protective actions to preserve power system stability. Such events have happened in the past, and when they do large numbers of customers lose power. However, there is limited data on the frequency with which such outages occur as a function of temperature, and modeling this type of failure mechanistically would require a full AC power flow model for the region. This beyond the scope of this study. Instead, we assume a probability of failure for each substation and then conduct sensitivity analysis to examine the influence of this assumption.

3.4.9 Health Mortality model

In the ABM, we included heat-related excess mortality through a model that estimates the number of deaths among the model agents during each heatwave based on characteristics of the agent (age and access to air conditioning) and of the heatwave (duration and intensity). This is a machine learning model trained with past data on the impacts of heat waves.

`

We developed the heat wave mortality model using data from 83 U.S. communities between 1987 and 2005. We next used these estimates of heatwave-specific health effects, in conjunction with characteristics of each heatwave including its length and intensity, to build predictive models of the health effects of a specific heatwave based on its characteristics. Next, we adjusted the predicted effects based on each agent's access to air conditioning during the heatwave (as modeled by other components of the ABM), based on results from previous research and estimates from the US Census's American Housing Survey of baseline prevalence of A/C in Baltimore. The heat-mortality module of the ABM uses these predicted health risks to generate predictions of risk during the heatwave for each agent and use these to estimate numbers of heat-related deaths among the agents using a standard health impact assessment approach, applying the predicted relative risk of mortality associated with each heatwave to underlying baseline mortality rates in Baltimore. A detailed description of this model can be found in Appendix C.

3.5 Methodology

3.5.1 Baseline Scenario and heatwave characteristics

We test our model in our case study area to simulate how the agents evolve from 1988 to 2017, a total of 30 years in multiple scenarios. This section defines the baseline scenario which is used for comparison with other scenarios. We ran the model for 100 iterations for each scenario as this provides an acceptable level of convergence for the total mortality and A/C coverage ratio over time. We use estimates the number of iterations needed for convergence based on the approach shown in equations (1) and (2). First, we calculated the coefficient of variation shown in equation (2) for the simulation output measure. We then use equation (1) to estimate the number of replications $n_{\min}$ we need to achieve our desired level of convergence defined by E.

`

$$n_{\min} = argmax_n |c_V^n - c_V^m| < E, \forall m > n \qquad (1)$$

$$c_V = \frac{\sigma}{\mu} \qquad (2)$$

The Baseline scenario is calibrated with the historical A/C adoption percentage for the area by tuning the purchase motivation parameter M of all households to invest in A/C units. Based on the American Housing Survey for metropolitan Baltimore, the coverage ratio for housing units with central AC or one room A/C unit are used for calibration in our model. We exclude houses with multiple room units as we are most interested in the trend of investing AC while preserving the space for the region to evolve. In the left plot of Figure 3-5, we show the estimated mortality during each high-temperature event given our definition from table 1 for the resulting base-case model. We found that for most high temperature events mortality is fairly low, consistent with historic events. However, for some extreme heat waves where the mortality is significant, the risk may decrease through mitigation efforts. The right plot of figure 5 shows the baseline AC coverage growth over time and the AHS data we use to calibrate the model. Defining the baseline scenario enables us to test the sensitivity of the output to different parameters and discuss the interplay among individual behaviors, social networks, power outages, and government mitigation aids.

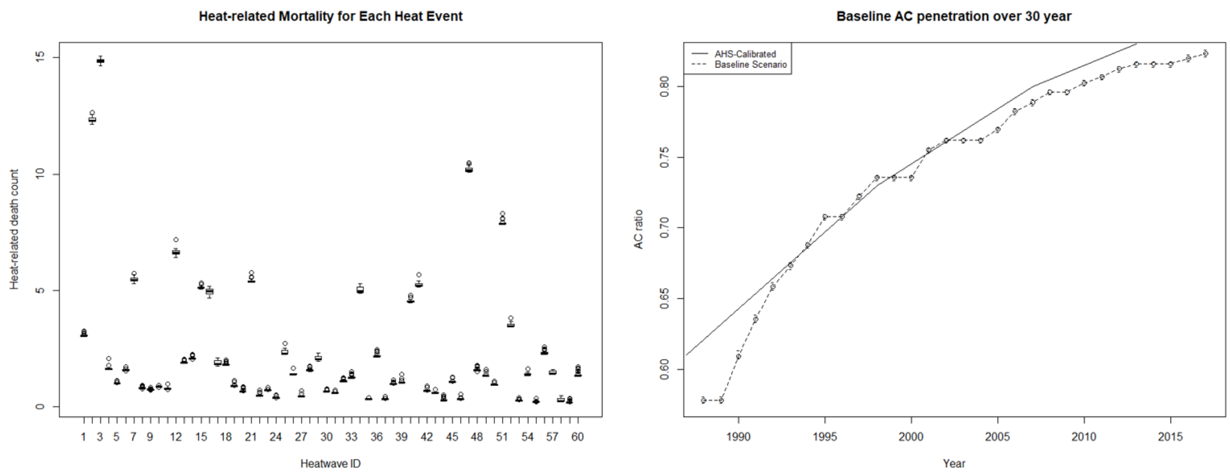


Figure 3-5 Baseline heat-related mortality and AC penetration. The box plot is showing the 5% and 95% quantile of the output samples.

3.5.2 Impacts of Different Heat Wave Environments

Climate change is generally predicted to lead more frequent and intense heat wave events[48]. These changes in heat wave intensity may interact with individual mitigation measures to change heat wave risk in ways that we may not anticipate. We used two different temperature scenarios to investigate the potential of this interaction. We assume the historical temperature becomes different for our study region. One scenario with 3 °Fwarmer and one with 1 °F lower for all year temperature. These two scenarios can be interpreted as either different locations in the US, or different climate in the future. In Figure 3-6, we compare the total mortality due to heat waves and the air conditioning coverage over time for these two scenarios and the baseline scenario.

`

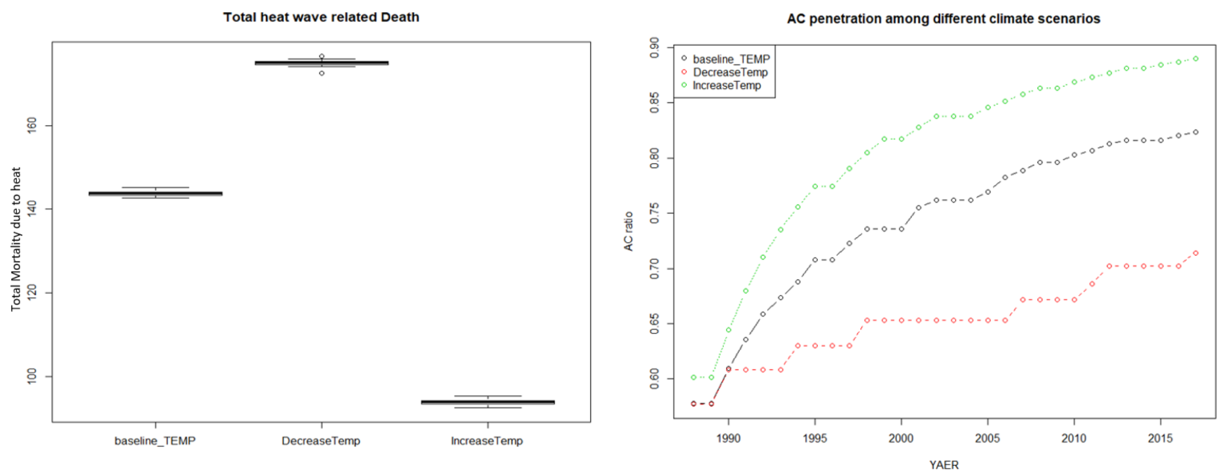


Figure 3-6 Impacts of different heat wave environment

The results of this comparison are perhaps surprising. As we can see, compared to the baseline scenario, the increased temperature case has lower heat wave mortality and the decreased temperature case has higher heat wave mortality. The underlying reason is that when we increase or decrease the temperature, it changes the heat wave experiences of the agents, which in turn changes their air conditioning purchasing decisions. When heat waves are frequent and more people are exposed to higher temperatures, individual mitigation choice will take place and result in more houses investing in air conditioning. On the other hand when the temperature is slightly lower, the agents have less incentive to purchase air conditioning which, in the long term, resulting in more casualties during heat waves that are a bit less strong but still detrimental to at-risk populations without air conditioning.

3.5.3 Impacts of Mitigation

We test the effect of three types of mitigations: (1) individual mitigation decisions, (2) access to air conditioning via an agent's social network, and (3) government interventions in the form of cooling centers. We tested each mitigation scenario with different parameter sets which alter the

`

strength of each mitigation. The parameters of each case are shown in Table 2. Parameters in bold mean they are perturbed compared to the baseline scenario. In Figure 3-7, we compare the total number of lives saved during heat waves from mitigations and the A/C coverage evolution for these scenarios. Lives saved from mitigation is calculated by comparing with the Total Estimated Mortality w/o mitigation, where it assumes every agent is being exposed for all heat waves.

Table 3-2 Definition of Mitigation Scenarios.
M is purchase motivation, w is invitation willingness, and N_C is total cooling center capacity. Parameters in bold are what changed in each scenario comparing to baseline

Scenario ID	Scenario Name	Parameters
1	Baseline	$M = 3, w = 1, N_C = 0$
2	High Social Invitation	$M = 3, \boldsymbol{w = 1.3}, N_C = 0$
3	Low Social Invitation	$M = 3, \boldsymbol{w = 0.5}, N_C = 0$
4	No Social Invitation	$M = 3, \boldsymbol{w = 0}, N_C = 0$
5	Low capacity Cooling Center	$M = 3, w = 1, \boldsymbol{N_C = 3000}$
6	Medium Capacity Cooling Center	$M = 3, w = 1, \boldsymbol{N_C = 10000}$
7	High Capacity Cooling Center	$M = 3, w = 1, \boldsymbol{N_C = 30000}$
8	Higher Purchase Motivation	$\boldsymbol{M = 9}, w = 1, N_C = 0$
9	Lower Purchase Motivation	$\boldsymbol{M = 2}, w = 1, N_C = 0$

We start by interpreting each factor. The most straightforward one is by changing the motivation of agents to purchase A/C units. This scenario serves as a sanity check for the validity of the individual behavioral model. The result is obvious. By increasing the motivation of buying A/C units, more agents investment in A/C which effectively saves more lives. On the other hand, if people have low motivation to purchase A/C units, the vulnerability of the region increase substantially. As an interesting finding, from the cumulative mortality curve, we find that although the total mortality is not as much as the worst cases, when decreasing people's motivation in self-mitigation it can be detrimental to the region's hazard coping ability. In the

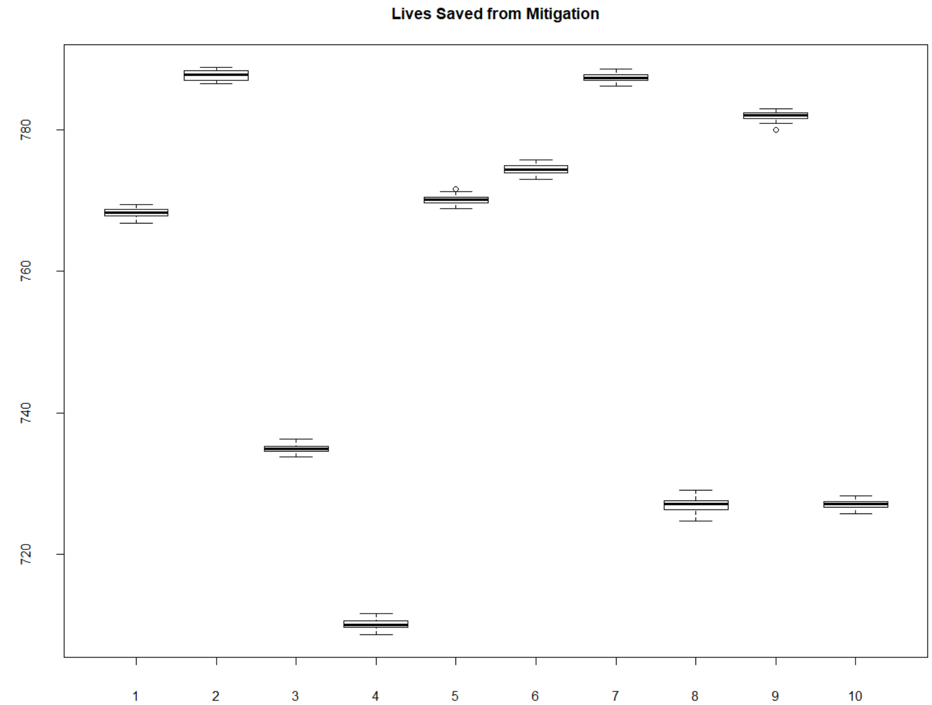
`

last few years, the mortality of this scenario is indeed uprising and surpasses scenarios with less social interactions.

Then we focus on changing the social invitation ratio. For these three cases, if there is no social network, meaning that every household can only rely on themselves, significantly more people would purchase A/C units. However, though the coverage of A/C is a higher, less lives are saved compared to any other scenario with social invitations. The reason are two-fold. On one hand, people are learning their lessons after an exposure to the heat waves, which inevitably increase the health risks of the agents. On the other hand, with a high coverage of A/C units, the chances of losing power increase for the region. Meanwhile, when we increase the chance of friend invitations, it describes one of the best scenarios where most people can be saved – even without any government mitigation like cooling centers. This scenario has the lowest A/C coverage ratio except for decreasing agent's A/C purchase motivation. It has the lowest chance of losing power and saves the most in terms of both resources and lives.

Finally, we analyze the effect of opening cooling centers in the city. Cooling centers can help the agents mitigate and save lives. Obviously, more availability spaces from the cooling center mean more lives can be saved, while the coverage of A/C decreases accordingly. This shows the importance of instructing vulnerable population about the availability of cooling centers, which is in general the most effective way of saving lives.

`



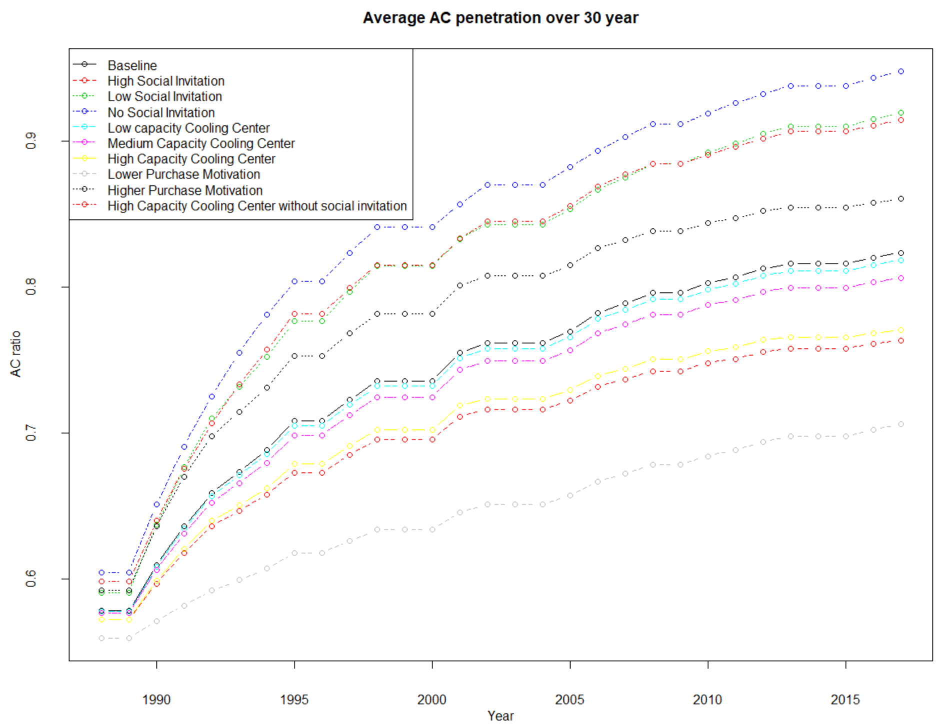


Figure 3-7 Lives saved from mitigation and AC penetration ratio for multiple scenarios. Scenario parameters can be found in Table 2.

`

3.6 Conclusion

In this work, we built an Agent Based Model as the platform to study the evolution of heat wave vulnerability of a community and increase our understanding of how individual mitigation behaviors, social connections, climate change, power outages, and government mitigation aid interact with each other and change the evolution of community vulnerability. We calibrated agents learning, decision making, and interactions in a historical weather climate of Baltimore City, MD. Then we conducted a sensitivity analysis on key parameters, such as the temperature, purchase motivation of AC units, available slots in cooling centers, and social invitation strength.

We found that higher temperatures will trigger more mitigation, which in turns lower the mortality over time. Meanwhile, we found that social invitation and cooling centers are the most effective mitigation measures. Without social intervention, though with a very high A/C penetration ratio, most mortality would occur while the likelihood of losing power is high and people in poverty could not afford to invest in A/C units. On the other hand, cooling centers do have the potential to further reduce the mortality targeting vulnerable populations, but it also means that people should have a correct knowledge regarding the heat wave events and the availability of cooling centers from the information spread by the government. Furthermore, if power outage happen less frequently from the effort of utility companies, this would also be beneficial for the region's heat resistance. These findings are valuable to achieve the most resilient community with all parties' efforts, to increase social bonding, AC coverage, robustness of the power system, and the availability and accessibility of cooling centers.

3.7 References

1. Changnon SA, Kunkel KE, Reinke BC, Changnon SA, Kunkel KE, Reinke BC. Impacts and Responses to the 1995 Heat Wave: A Call to Action. https://doi.org/10.1175/1520-0477(1996)077<1497:IARTTH>2.0.CO;2, 1996.
2. Klinenberg E. Dying Alone: The Social Production of Urban Isolation. Ethnography, 2001; 2(4):501–531.
3. Anderson GB, Bell ML. Lights out: Impact of the August 2003 power outage on mortality in New York, NY. Epidemiology, 2012; 23(2):189–193.
4. De Bono A, Peduzzi P, Kluser S, Giuliani G. Impacts of Summer 2003 Heat Wave in Europe. United Nations Environment Programme, 2004.
5. Kunkel KE, Changnon SA, Reinke BC, Arritt RW, Kunkel KE, Changnon SA, Reinke BC, Arritt RW. The July 1995 Heat Wave in the Midwest: A Climatic Perspective and Critical Weather Factors. https://doi.org/10.1175/1520-0477(1996)077<1507:TJHWIT>2.0.CO;2, 1996.
6. Karl TR, Knight RW, Karl TR, Knight RW. The 1995 Chicago Heat Wave: How Likely Is a Recurrence? https://doi.org/10.1175/1520-0477(1997)078<1107:TCHWHL>2.0.CO;2, 1997.
7. Weisskopf MG, Anderson HA, Foldy S, Hanrahan LP, Blair K, Török TJ, Rumm PD. Heat Wave Morbidity and Mortality, Milwaukee, Wis, 1999 vs 1995: An Improved Response? American Journal of Public Health, 2002; 92(5):830–833.
8. Bernard SM, McGeehin MA. Municipal Heat Wave Response Plans. American Journal of Public Health, 2004; 94(9):1520–1522.
9. Pollack CE, Von Dem Knesebeck O. Social capital and health among the aged: Comparisons between the United States and Germany. Health and Place, 2004; 10(4):383–391.
10. Hyyppä MT, Mäki J. Social participation and health in a community rich in stock of social capital. Health Education Research, 2003; 18(6):770–779.
11. Rose R. How much does social capital add to individual health? A survey study of Russians. Social Science and Medicine, 2000; 51(9):1421–1435.
12. Veenstra G. Social capital, SES and health: An individual-level analysis. Social Science and Medicine, 2000; 50(5):619–629.
13. Kawachi I, Kennedy BP, Glass R. Social capital and self-rated health: A contextual analysis. American Journal of Public Health, 1999; 89(8):1187–1193.
14. Naughton MP, Henderson A, Mirabelli MC, Kaiser R, Wilhelm JL, Kieszak SM, Rubin CH, McGeehin MA. Heat-related mortality during a 1999 heat wave in Chicago. American Journal of Preventive Medicine, 2002; 22(4):221–227.
15. Klinenberg E. Heat Wave: A Social Autopsy Of Disaster In Chicago., 2015.
16. Hajat S, Armstrong B, Baccini M, Biggeri A, Epidemiology LB-, 2006 undefined. Impact of high temperatures on mortality: is there an added heat wave effect? JSTOR,.
17. Anderson GB, Oleson KW, Jones B, Peng RD. Projected trends in high-mortality heatwaves under different scenarios of climate, population, and adaptation in 82 US communities. Climatic Change, 2018; 146(3–4):455–470.
18. Semenza JC, Rubin CH, Falter KH, Selanikio JD, Flanders WD, Howe HL, Wilhelm JL. Heat-Related Deaths during the July 1995 Heat Wave in Chicago. New England Journal of Medicine, 1996; 335(2):84–90.
19. Anon. Heat-related mortality--Chicago, July 1995 - PubMed.

20. CENTRE FOR RESEARCH ON THE EPIDEMIOLOGY OF DISASTERS (CRED). Annual disaster statistical review 2010: the numbers and trends | PreventionWeb.net.
21. Benmarhnia T, Deguen S, Kaufman JS, Smargiassi A. Vulnerability to heat-related mortality: A systematic review, meta-analysis, and meta-regression analysis. Epidemiology, 2015; 26(6):781–793.
22. Matthies F, Bickler G, Mar\'\in NC. Heat-Health Action Plans: Guidance. World Health Organization, 2008.
23. Gronlund CJ, Zanobetti A, Schwartz JD, Wellenius GA, O'Neill MS. Heat, Heat Waves, and Hospital Admissions among the Elderly in the United States, 1992–2006. Environmental Health Perspectives, 2014; 122(11):1187–1192.
24. Bouchama A. Prognostic Factors in Heat Wave–Related Deaths<subtitle>A Meta-analysis</subtitle> Archives of Internal Medicine, 2007; 167(20):2170.
25. O'Neill MS, Zanobetti A, Schwartz J. Disparities by race in heat-related mortality in four US cities: The role of air conditioning prevalence. Journal of Urban Health, 2005; 82(2):191–197.
26. Ostro B, Rauch S, Green R, Malig B, Basu R. The effects of temperature and use of air conditioning on hospitalizations. American Journal of Epidemiology, 2010; 172(9):1053–1061.
27. Medina-Ramón M, Schwartz J. Temperature, temperature extremes, and mortality: A study of acclimatisation and effect modification in 50 US cities. Occupational and Environmental Medicine, 2007; 64(12):827–833.
28. Bonabeau E. Agent-based modeling: Methods and techniques for simulating human systems. Proceedings of the National Academy of Sciences of the United States of America, 2002; 99(SUPPL. 3):7280–7287.
29. MacAl CM, North MJ. Tutorial on agent-based modelling and simulation. Journal of Simulation, 2010; 4(3):151–162.
30. An L. Modeling human decisions in coupled human and natural systems: Review of agent-based models. Ecological Modelling, 2012; 229(0):25–36.
31. Reilly AC, Guikema SD, Zhu L, Igusa T. Evolution of vulnerability of communities facing repeated hazards. Linkov I (ed). PLoS ONE, 2017; 12(9):e0182719.
32. Lee JS, Filatova T, Ligmann-Zielinska A, Hassani-Mahmooei B, Stonedahl F, Lorscheid I, Voinov A, Polhill G, Sun Z, Parker DC. The complexities of agent-based modeling output analysis. Jasss, 2015; 18(4).
33. Pan X, Han CS, Dauber K, Law KH. A multi-agent based framework for the simulation of human and social behaviors during emergency evacuations. AI and Society, 2007; 22(2):113–132.
34. Chen X, Zhan FB. Agent-based modelling and simulation of urban evacuation: Relative effectiveness of simultaneous and staged evacuation strategies. Journal of the Operational Research Society, 2008; 59(1):25–33.
35. Chen X, Meaker JW, Zhan FB. Agent-based modeling and analysis of hurricane evacuation procedures for the Florida Keys. Natural Hazards, 2006; 38(3):321–338.
36. Tonn GL, Guikema SD. An Agent-Based Model of Evolving Community Flood Risk. Risk Analysis, 2018; 38(6):1258–1278.
37. Reilly AC, Tonn GL, Zhai C, Guikema SD. Hurricanes and Power System Reliability-The Effects of Individual Decisions and System-Level Hardening. Proceedings of the IEEE, 2017; 105(7):1429–1442.

38. Dignum V, Gilbert N, Wellman MP. Introduction to the special issue on autonomous agents for agent-based modeling. Autonomous Agents and Multi-Agent Systems, 2016; 30(6):1021–1022.
39. Logan TM, Guikema SD, Bricker JD. Hard-adaptive measures can increase vulnerability to storm surge and tsunami hazards over time. Nature Sustainability, 2018; 1(9):526–530.
40. El-Sayed AM, Scarborough P, Seemann L, Galea S. Social network analysis and agent-based modeling in social epidemiology. Epidemiologic Perspectives and Innovations, 2012; 9(1):1–9.
41. Yang Y, Mao L, Metcalf SS. Diffusion of hurricane evacuation behavior through a home-workplace social network: A spatially explicit agent-based simulation model. Computers, Environment and Urban Systems, 2019; 74:13–22.
42. Ko TH, Berry NM. Agent-Based Modeling with Social Networks for Terrorist Recruitment.
43. Logan TM, Zaitchik B, Guikema S, Nisbet A. Night and day: The influence and relative importance of urban characteristics on remotely sensed land surface temperature. Remote Sensing of Environment, 2020; 247:111861.
44. Robinson PJ, Robinson PJ. On the Definition of a Heat Wave. http://dx.doi.org/10.1175/1520-0450(2001)040<0762:OTDOAH>2.0.CO;2, 2001.
45. Frank O, Strauss D. Markov graphs. Journal of the American Statistical Association, 1986; 81(395):832–842.
46. Hunter DR. Curved exponential family models for social networks. Social Networks, 2007; 29(2):216–230.
47. Berisha V, Hondula D, Roach M, White JR, McKinney B, Bentz D, Mohamed A, Uebelherr J, Goodin K. Assessing adaptation strategies for extreme heat: A public health evaluation of cooling centers in Maricopa County, Arizona. Weather, Climate, and Society, 2017; 9(1):71–80.
48. Meehl GA, Tebaldi C. More intense, more frequent, and longer lasting heat waves in the 21st century. Science, 2004; 305(5686):994–997.

Chapter 4: Power Outage Prediction for Natural Hazards Using Synthetic Power Distribution Systems

4.1 Abstract

Power outage prediction for natural hazards usually relies on one of two approaches, statistical models or fragility-based methods. Statistical models have provided strong predictive accuracy, but only in an area-aggregated manner. Fragility-based approaches have not offered strong prediction accuracy and have been limited to systems for which system topology or performance models are available. In this chapter, we create an algorithm that (1) generates a synthetic power system layout for any U.S. city based only on public data and then (2) simulates power outages at the level of individual buildings under hazard loading using fragility functions. This approach provides much more localized, building-level estimates of the likelihood of losing power due to a natural hazard. We validate our model by comparing the network properties and power outage events based on our approach with data from a real power system in Ohio. We find that our model relies on less input data comparing to statistical learning approaches yet can make accurate predictions, provided accurate fragility curves are available.

4.2 Introduction

`

Power outages regularly cause large economic losses, impact other critical infrastructure systems and significantly disrupt daily life. The leading cause of large-scale power outages in the U.S. are severe weather events, though other natural hazards such as earthquakes can lead to significant outage events as well [(1)]. By some estimates, 75% of power outages are either directly caused by weather-induced faults, or indirectly through failure of aging equipment exposed to significant weather events[(2)]. For example, in 2012, a powerful derecho struck the Midwestern United States and caused 4.2 million customers across 11 states to lose power. Power restoration took up to 10 days for some areas [(3)].

Power outage estimation can help power utilities, the managers of other critical infrastructure that are dependent on power, governments, and private organizations with both short-term event response and longer-term resilience planning. In the short term, pre-event power outage estimation can help utilities better plan their response and thus better balance cost and restoration speed. For large-scale outage events, utilities rely heavily on outside personnel and restoration material through mutual aid agreements. These resources are critical in restoring power quickly but are very costly, with restoration costs in the tens of millions of dollars per day for larger utilities [(4)]. If a utility underestimates the outages from a forecast weather event and does not bring in enough mutual aid, their customers face prolonged outages. On the other hand, if a utility overestimates outages and brings in more mutual aid than is needed, they incur unnecessarily high costs which must then be borne by either rate-payers or corporate owners or shareholders. In the longer term, being able to estimate the likelihood of power outages at the individual building level from both weather events and other hazards such as earthquakes and floods can help utilities better understand where system strengthening may be needed. Building-level estimates of the likelihood of losing power can also help businesses and homeowners better

`

understand and improve their resilience locally through considerations of back-up power, power disruption insurance (for a commercial entity), and other measures.

There are two main approaches for estimating power outages in the literature, statistical models and fragility-based models [5,6]. These will be discussed in more details in Section 2, but briefly, statistical models have been implemented successfully in practice and can provide accurate estimates of power outages due to a forecast weather event at a spatial unit scale (e.g., census tract, zip code, or county). However, to date, they have not provided building-level estimates. On the other hand, fragility-based approaches do provide building-level outage estimates but they (1) rely on data about the power system layout that is not usually available, (2) have not shown strong accuracy to date and (3) can be computationally difficult and require a non-negligible amount of data collection effort to scale up to national or global level analysis.

The work in this chapter significantly advances the second of these approaches – fragility-based outage estimation – by developing a new framework to accurately estimate the probability of losing power at the individual building level using only publicly available data. The fragility-based approach simulates how each component of the system fails given fragility curves and determines the functionality of the system given these failure components. However, there is a fundamental challenge. The topology of power distribution systems is not publicly available in the U.S. due to security concerns. Because of this, we first develop a method to generate a synthetic power distribution system based on only publicly available data. The method is generalizable throughout the US and can be scaled from the city level to a larger spatial level, such as county level or state level. We then develop a method to simulate power outages at the building level based on this synthetic system in combination with hazard loading and fragility information. This approach can be used to simulate many types of natural disasters, e.g.

`

hurricanes, flooding, and earthquakes, provided that relevant hazard loading maps and fragility curves are available.

The structure of this chapter is as follows. We first discuss previous research on power outage prediction, synthetic power grid generation, and infrastructure vulnerability analysis. We then introduce our methodology to generate a synthetic distribution system and how to simulate its performance under hazard loading. Next we validate our synthetic network layouts against actual distribution systems and at last we test the outage simulations against historical large-scale power outage events to demonstrate the accuracy and functionality of the system.

4.3 Literature Review

Before giving an overview of the related literature, one point of clarification is needed. Most outages due to severe weather events occur in the power distribution system, while for earthquakes there is often also damage to substations (and sometimes the transmission system and generators) (7). In comparison to the power transmission system, the power distribution system, which delivers power from local substations to each customer, is more vulnerable and more likely to be damaged and cause customers to lose power during adverse weather events (8). For the distribution system, when a line breaks, a pole falls, or a distribution substation fails, customers that are downstream of the failed devices will be isolated and lose power unless there is an alternate set of lines that can serve that customer (9). This redundancy is rare in power distribution systems; most power distribution systems are dominantly radial in design (10,11). The method we present in this chapter focuses on low voltage substations and the distribution system, though it can incorporate the transmission system as well provided that network layout

`

information is available. Our literature review covers both systems. We first review previous work on statistical approaches for power outage prediction. Then we review methods of generating synthetic power systems followed by a review of fragility curves and outage estimation approaches based on fragility curves.

4.3.1 Statistical power outage predictions

One of the main approaches in both the research literature and in use in practice for estimating power outages are statistical methods. Many, though not all, of these studies focus on tropical cyclones [5,12–18]. Others focus more broadly on a range of weather conditions [19]. These approaches use data from past events together with a wide array of explanatory variables to develop statistical and machine learning models to predict power outages due to a forecast weather event.

The statistical models used for outage and damage forecasting have varied from relatively simple generalized linear and generalized additive models in early work [13,20] to regression trees [15] to ensembles of trees and other machine learning methods more recently [5,16,18,19,21,22] . One consistent challenge in using statistical approaches for outage forecasting is zero-inflation of the outage data, meaning that even for significant adverse events, many more areas experience zero outages than experience outages if the spatial units used are small. Guikema et al. developed a two-stage process that combines a classification model and a regression model to first predict in a given area whether power outages will occur or not, then proceed to estimate the severity of the outages (e.g. number of outages) [23]. Shashaani et al. (2018) and Kabir et al. (2019) developed three-stage approaches that introduced a new stage to predict the severity class of power outages followed by a quantile regression forest to provide a probabilistic prediction [18,19].

`

Statistical outage forecasting models have focused on severe weather events, and the input data used has consequently focused on features that may help predict the number of outages. This has included weather forecast information from numerical weather forecasting models [19,24] or, for hurricanes, from hurricane wind field models [5,16,20]. It has also included information about the assets exposed to the hazard such as the numbers of poles, transformers, and wire spans in each spatial unit [17–20]. In addition, it has also included a range of features that describe local geography, plant species, utility vegetation management, pre-storm soil moisture levels, and plant species in each area [17,25]. These features have been found to offer improved predictive accuracy. Other approaches, particularly those of Guikema et al. and Nateghi et al. have sought to use a reduced set of input features to predict power outages due to hurricanes [16,26]. The advantages of easier implementation in practice and potentially greater generalizability, but at the cost of some loss of predictive accuracy. Regardless of the details, all of these approaches provide outage estimates at the level of an aggregated spatial level, which may vary from relatively small (e.g., 5km by 5km square) grid cells to census tracts, counties, or utility operating districts. They do not provide building or facility-level estimates.

4.3.2 Fragility curves and fragility curve based methods

The other main approach for estimating loss of power and power system damage due to natural hazards is one based on fragility curves. This is the approach implemented in HAZUS, a natural hazards loss estimation software package supported by the U.S. Federal Emergency Management Agency (FEMA). A fragility curve is a function giving the probability of a particular infrastructure asset or building reaching or exceeding a specific damage state given a

`

quantified disaster intensity metric. For example, a fragility curve could give the probability that a substation is in each of four damage severity states as a function of the ground shacking, measured by peak ground acceleration, due to an earthquake at that substation's location. One of the more widely used set of fragility curves are those in HAZUS for damage to infrastructure components and buildings from earthquakes, wind events, floods, and tsunamis (27,28).

Fragility functions are of course hazard-dependent, and some hazards are more well-studied than others. Fragility functions for power system components for seismic events have been particularly well-studied. Power system components can be divided into micro-components (e.g., coil support, circuit breaker, transformer) and macro-components (a combination of micro-components). Vanzi developed some of the earliest fragility functions for electric power system components for seismic events based on a functional form given by a cumulative lognormal distribution (29). HAZUS includes fragility curves for a broader set of power system assets, including generation plants, substations, and distribution. For example, HAZUS classifies substations into low voltage, medium voltage and high voltage. For each voltage level, the substation can be anchored or unanchored. It then describes the damage of electric power substation from earthquake with five different severity states and provides the lognormal parameters of the probability the substation exceeds each damage state given the peak ground acceleration.

Fragility curves for strong wind events are not as developed as for seismic events, though significant progress has been made in the last several years. The most critical asset for wind events is the poles in the system as these are the major locations of failures and thus cause of outages during wind events. An early approach is that of Han et al. which used a structural reliability model to estimate a fragility curve for power poles, used this to formulate a prior

`

probability distribution, and then updated this with observed pole failure data [30]. However, they were hampered by insufficient observations of pole states after events. Mohammadi et al. provide a more recent and comprehensive study on utility poles fragility curves for strong wind events [31]. They consider pole age, conductor area, height and wind direction as variables in the lognormal fragility function. However, none of this prior work incorporates the impact of trees falling and shedding branches onto power lines or power poles in their functions, and this can be a major cause of failure [16,32].

A set of fragility curves on its own is not a complete model for estimating power outages. The fragility curves must be coupled with a method for simulating realizations of damage states of the set of assets and then for translating each of these realizations into an estimate of which customers lose power. If the actual power distribution system layout is known, then this is a relatively straight-forward task. Typically, a Monte Carlo simulation is used to generate N replications of asset damage states, where one replication includes a damage state for each asset in the system. Then either a power flow or, more often for a distribution system level analysis, a connectivity-based model is used to determine whether or not each customer has power based on the set of assets damaged between that customer and the substation serving them. If there is redundancy in the system, then a power flow analysis is likely necessary instead of a simple connectivity model [33].

A challenge is that the actual power system layout is rarely available at the distribution level, which is the level that is most needed for outage estimation. Power utilities do not share this data publicly due to security concerns. While specific researchers do get access to the layouts for specific utilities for specific projects, this to date has made it very difficult to apply fragility-based approaches widely. One way around this has been to assume a system topology. For

`

example, Winkler et al. (2010) assumed that all roads in a city had a power line and then used this to simulate outages. This, however, does not approximate the topology of actual distribution systems well, particularly in how each building is connected through the network to a substation. An alternative, the one we develop further in this chapter, is to generate synthetic layouts that better represent the layout of power systems.

4.3.3 Synthetic power system generation

There has been some previous work on synthetic power grid generation. Most of this previous research has focused on the transmission system [34–37]. For example, Birchfield et al. propose a method to generate synthetic transmission systems together with validation criteria for these systems [34]. They first place high voltage substations with a clustering algorithm considering the spatial distribution of customers. Then they add in transmission lines that would meet power flow constraints. They test their model with a 2000-bus public test case. Soltan and Zussman (2016) present a Geographical Learner and Generator Algorithm (GLGA) to generate synthetic transmission networks similar to a given network comparing similar structural and spatial properties such as average path length, clustering coefficient, degree distribution of the nodes, and length distribution of the lines. Some recent work also tries to model synthetic distribution networks [38,39]. Pisano et al. use georeferenced information and other publicly available open data to estimate the energy consumption of a region. Based on locations of primary substations and territory segmentation they create the layout of the distribution system which can be used further in optimization studies. However, these generated synthetic distribution networks cannot serve as a good representation of the actual system when

`

determining power outages during natural disasters because they, (1) lack critical details such as the locations of poles undergrounding and (2) lack validation against actual distribution systems.

For the distribution grid, considerable work has been focused on optimal layouts of distribution networks [(40–42)].Miranda et al. (1994) used a genetic algorithm to plan the placement of new distribution network to expand from an existing system. They assumed the network to be radial and optimized the system to minimize new facility installation costs and operation costs under constraints such as power flow, voltage drop, and power demand etc. Their approach starts with possible sites for substations and potential power line locations and uses the genetic algorithm to find an optimal solution for this binary integer optimization problem. Valenzuela et al. (2019) used a Minimum Spanning Tree model to create a distribution network with georeferenced data (e.g. customers' locations, street point positions). They focus on the optimal allocation of distribution transformers and assumed undergrounding lines only, a situation that would be quite uncommon in the U.S. The goal is to create a distribution network that minimizes the total load shed during extreme events. Overall, these approaches provide starting points, but there does not yet exist an approach that allows us to create a synthetic power distribution network that is representative of power distribution systems and then to simulate hazard-induced failures and estimate the probability of loss of power at the individual building level. This is the challenge we address in this chapter.

`

4.4 Methodology

4.4.1 Assumptions and information needed

Even though distribution network layouts are not publicly available, two critical aspects that help define a distribution network can be acquired or can be proxied. One of these is the location of customers (power meters). We know that every building in a developed nation can receive electric power so we can assume the locations of meters are the locations of buildings and that each building must thus be served by the synthetic system. This is not an exact representation of the number of customers because, for example, some multi-unit buildings have multiple meters. However, the spatial distribution of building locations is a good proxy for the actual customer locations for the purpose of creating locations for synthetic power lines. Such information can be retrieved from building footprints that are publicly available for most cities in the US.

Another critical component in developing a synthetic power distribution system is the location of distribution substations. The locations of distribution substations are publicly available from open-source map platforms[5]. We can view these substations as power supply points that deliver power to the distribution feeders leading to each customer.

Three key assumptions underlie our approach for generating synthetic power distribution systems. These assumptions are:

1) All powerlines are within roads' rights-of-way,

---

[5] *e.g.* *overpass-turbo.eu*

`

2) the system is designed in manner that at least approximately the least-cost method of connecting all customers in terms of total line miles, and

3) the network is radial. That is, it has a tree-structure.

None of these are strictly true for all real systems. However, they are true for large portions of many U.S. power distribution systems (43), and, as we show below, they allow us to achieve our goal of accurately estimating the likelihood of power outages at the individual building level based on only publicly available information. We will elaborate further on these below. With these two sets of locations, customers and substations, and three key assumptions, we can create a synthetic distribution power network to simulate power outages under extreme hazard events.

We focus our approach on the major power infrastructure components that could be damaged by natural hazards. That is, rather than trying to model every component, we focus on the asset classes that are the main sources of loss of power and that correspond to the available fragility curves, namely substations, poles, and power lines. Different types of power system components have different responses to natural disasters. Substations, which transform voltage from high to low, are more likely to be damaged during earthquakes and flooding events than during wind events. In a radial system, damage to substations may lead to loss of entire feeders (a set of distribution lines serving one portion of a substation's service area), creating an outage that impacts many customers simultaneously. Poles are more likely to be damaged during strong wind events and, if there is liquefaction, during earthquakes. Pole damage in a radial system cuts power to all downstream customers. Above-ground power lines can be damaged from falling trees and limbs and blown debris, and below-ground lines can be damaged due to flooding. Fragility curves, e.g., those in HAZUS, often aggregate the line damage probabilities and assign it to the closest upstream pole. We adopt this approach and do not assign separate failures probabilities to lines.

`

We create network layouts for a power distribution system, each of which describes how each customer gets power from distribution substations through power lines. Each layout is a graph with the customers, substations, and poles as vertices and the power lines as the edges. This introduces three questions we need to answer: 1) which substation each customer gets power from, 2) the route from each substation to each customer, and 3) whether the power lines in each location are overhead or underground.

As discussed above, one of our key assumptions is that power lines are strictly along the roads. Based on our observation from the actual distribution layout in our case study system, that of Franklin County, Ohio, the average distance from a powerline to the nearest road is less than 100m for 92% of power lines are within 100m to roads. We calculate the Euclidean distance from each end point of each segment of power line polyline to the closest road intersections and average the value for each polyline to make this calculation.

One of the major advantages of our approach is the low requirement of data collection in comparison to statistical learning approaches [(18)]. We use only open-source data, i.e., road layouts, building locations, partial building information, and substation locations, to generate the power system network. Road layouts are shapefiles available from United States Census Bureau at the county level[6]. Power demands locations are the coordinates of each customer within the city/region boundaries. We extract these coordinates from any given city's building footprints and approximate each customer location with the centroid of each building[7]. For supply points, we focus on the distribution system because this is the portion of the system most likely to be damaged

---

[6] E.g., *https://catalog.data.gov/dataset/tiger-line-shapefile-2015-county-franklin-county-oh-all-roads-county-based-shapefile*

[7] E.g. *https://geo.btaa.org/catalog/6a0036b36c004d53b747d322265df751_1*

`

by many types of hazards [8]. We use open-source map query website *overpass.turbo* to identify all the substation locations and download them and we use *Zillow.com* to retrieve building information. With this set of information, we can then create our synthetic power network that provides power connectivity from substations to all customers.

4.4.2 Network Generation Methodology

In this section, we introduce the process of generating a synthetic power system layout with different approaches and validate those layouts. In general, the process can be divided into 6 steps as the flow chart in Figure 4-1 shows.

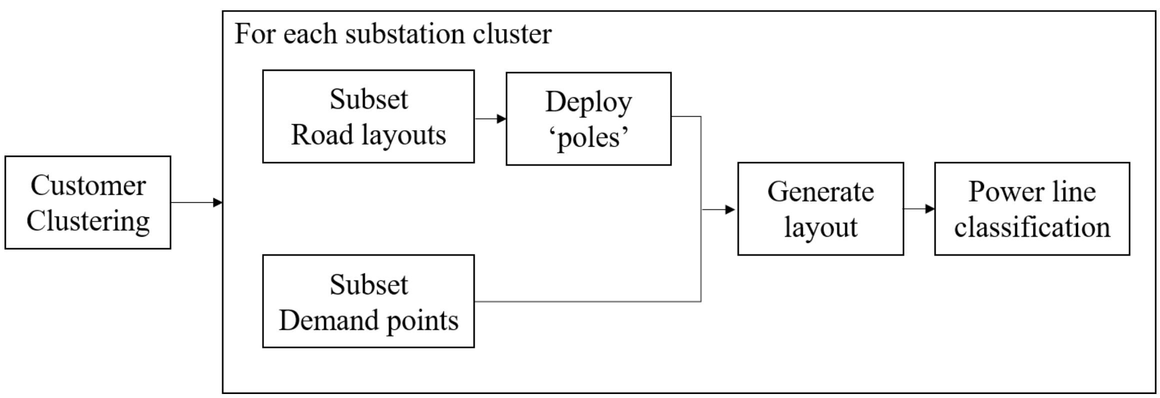


Figure 4-1 The six steps of our synthetic power distribution system network generation algorithm

4.4.2.1 Customer clustering

The first step in generating the distribution system is to create a cluster of customers served by each substation. This process answers our first question in synthetic distribution network generation – which substation a customer gets power from. In the U.S., where distribution systems

`

are typically radial, each customer is served by only one substation. In practice, many assign customers to substations based on closest euclidean distance. However, this is not necessarily the case. Sometimes for a customer, the closest substation following the path of power lines is not the the substatuion with the closest euclidian distance. For example, topological effects, water bodies, and other considerations can lead a customer to be served by a substation other than the closest Euclidean distance substation. The assumption that customers are served by the closest substation by euclidian distance can lead to substantially different results for risk assessment purposes . It eliminates the possibility the customer is actually served by another substation which experiences a very different damage condition. It is crucial to determine the potential substation service territory based on network distance, not euclidian distance.

Our method uses the following approach to determine the service territory of each substation. We first calculate the distances from each customer to all the substations along the road network. Then for each customer, we find the closest substation based on network distances. This gives us a basic customer cluster for each substation. This is helpful to determine the service substation for customers that are only close to one substation. However for customers that have a similar network distances to multiple substations, we generate multiple realizations of the layout that allow these customers to be in different clusters. This acknowledges that we are uncertain about which substation they are served by. We define c_i as the cluster that customer i belongs to and $d_n(i,j)$ as the network distance between customer i, $i \in [1, N]$ with substation $j \in [1, M]$. Then we have

$$c_i = j, if \quad \frac{d_n(i,j) - \min\limits_{\mathrm{j=1:M}} d_n(\mathrm{i,j})}{\min\limits_{\mathrm{j=1:M}} d_n(\mathrm{i,j}))} < \mathrm{b} \qquad (1)$$

where b is the threshold (10-20%) to constrain the relative distances from customers to other substations comparing to their closest substation. This means that customers on the boarders of clusters can appear in different substation clusters in different realizations of the network layout.

`

As an example shown in Figure 4-2, where there are two substations A and B. Orange customers and yellow customers are served by substation A and B respectively as they are very close to their substation. The two green customers have a similar distances to both substations, so they are considered to be in both substation clusters. This does not mean green customers are served by two substations at the same time but they are simulated independently in each substation cluster that has them, generating multiple network layouts that form an ensemble. As the output from the simulation, depending on the purpose of the study, we can choose to report the average probability across the ensemble or the highest probability across the ensemble of the green customers losing power from substation cluster A and B.

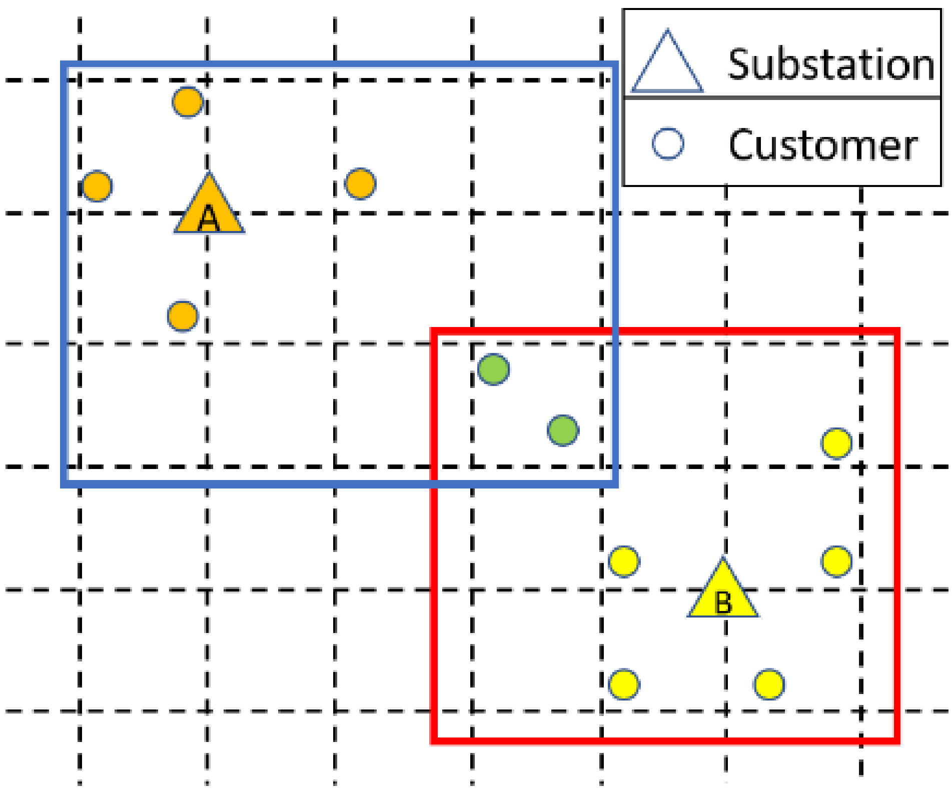


Figure 4-2 Illustration of border customers and their assignment to different substations

4.4.2.2 Create reduced problem

`

Instead of creating a fully connected distribution network that connects all the substations and customers, we can create multiple reduced distribution networks given each substation cluster. This simplifies computation and better reflects the layout of real substations. For each substation cluster, we generate a buffer that is slightly larger than the spatial locations of customers. As an example, in figure 4-2, the two substation clusters A and B, along with their customers and the roads within each square buffer compose two independent reduced problems. Then for each reduced problem, we create connectivity for each customer to get power from the substation. The benefit from creating the reduced problem is that it can reduce the computational effort relative to running algorithms on the entire system and it is a reasonable relaxation to the original problem as the actual system is typically radial and separated into substation feeders in the US.

4.4.2.3 Generate poles and road segmentation

Each customer in a typical power distribution system is connected to the substation through a combination of overhead lines (required utility poles) and underground lines. Therefore, we need to split the roads into segments such that that the distance between two nodes is similar to the actual distance between poles. For each road's polyline, we define the coordinates of all the nodes of the polyline as the sequence as $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$. We define these nodes as power nodes. The point (x_1, y_1) is the beginning node of the road segment, and (x_n, y_n) is the ending node. We define the Euclidean distance between point i and j as $d_e(i, j)$. The sequence of the polyline can be reversed. We assume the distance between poles to be a constant d_{pole} , and we assume this is 40 meters based on our observations of the pole distribution of actual systems. However, d_{pole}, can be changed for specific systems if information is available that suggests a

`

different span length for overhead lines. We define S as the set containing all the power nodes we created from the road shapefile.

The algorithm is as follows. We start by putting (x_1, y_1) into S. If the distance between (x_1, y_1) and (x_2, y_2) is less or equal to d_{pole}, we put (x_2, y_2) into S. If the distance is larger than d_{pole}, we split the line between (x_1, y_1) and (x_2, y_2) evenly and the number of nodes added is determined by the floor of $d_e(1,2)/d_{pole}$. For example, if $d_e(1,2) = 50m$, we will add in one point $(x_{1.1}, y_{1.1})$ in the middle of the polyline. We then put in nodes $(x_{1.1}, y_{1.1})$ and (x_2, y_2) into S. As a result, S should contains all the original nodes from road polylines and new nodes created when the distance between neighboring nodes are further than d_{pole}. If a power node belongs to an overhead power line, it will be viewed as a utility pole which can be exposed and damaged by wind or earthquake-induced liquification. A power feeder node is defined as the power node that is used to connect to at least one customer.

4.4.2.4 Generate distribution network

The next step is to create connectivity within each cluster to deliver power from substations to customers. We compare three different models to accomplish this goal: Steiner tree, K-mean clustering Steiner tree, and shortest path.

The first method we consider is to connect all customers and the substation in a minimum cost manner. In another words, we want to find the tree with the minimum cost that connects all the nodes of interests on an undirected graph. This is the problem of finding the Steiner tree on the graph [(44)]. Our base network is the road layouts as we constrain our power lines to be along roads. Our important vertices are power feeders, where one or more buildings receive power. We

`

calculate the closest power node for each building, and those power nodes with at least one customer near it will be considered as important nodes. The Steiner tree problem is a well-studied NP-hard problem, and many approximation algorithms have been created to reduce the difference from the optimal Steiner tree to the approximation solution [(45)]. We use an approximation algorithm created by Takahashi and Matsuyama [(44)]. We define the original road network as an undirected graph $G = (N, E)$, consisting the set N of all the potential power nodes we created from 4.4.2.3 and the set E of all the road segments that connect the power nodes. We define another undirected graph $G' = (N', E')$ to be the Steiner tree we are looking for, $N' \in N, E' \in E$. The weight of each edge is the length of the road segment. We define set $N_I \in N$ of all the power feeders. The algorithm is as following.

Algorithm 1. Steiner tree to generate distribution layout

Step 1: Select a random vertex $s \in N_I$, and find a vertex $t \in N', s \neq t$ that gives the shortest weighted path e_{st} to s. Add e to E' and all the vertices in e_{st} to N'. This gives a starting tree that connects s and t. Remove s and t from N_I.

Step 2: Search from all the vertices in N_I and find a vertex u such that the weighted path e_u from u to G' is the shortest. We then add e_u and all the vertices in e_u to N'. We then remove u from N_I.

Step 3: Repeat Step 2 until all the vertices in N_I have been connected to G' and N_I becomes empty.

The second method we examine is called K-mean clustering Steiner tree. The intuition of this method is trying to imitate the development progress of communities. We assume that the development of each substation cluster's distribution system begins with building major power lines from the substation to each neighborhood (i.e., grouping of buildings). Then the power lines

`

within each group of buildings are added, connecting to the original main power line. The algorithm changes as following.

Algorithm 2. Steiner tree with K-mean clustering to generate distribution layout

Step 1: Spatially cluster customers within each substation cluster with a K-mean algorithm and determine the best number of clusters based on silhouette score. We use the closest power node to each cluster center in G as our centers for communities, i.e. n_1, n_2, n_3.

Step 2: Then connect the substation's closest power node n_s to each substation cluster's power nodes n_1, n_2, n_3 with the shortest weighted path on G. Then add these vertices and arcs in G'. Remove substation power nodes and community cluster power nodes from N_I.

Step 3: Apply the Steiner tree algorithm to include all the important nodes and paths into G'.

The third method we examine to connect each building to the substation is a shortest weighted path approach. If we have power feeders $(n_1, n_2, n_3 \dots)$ and the substation power node n_s, then we find the shortest path from each power feeder node to a substation power node and include the path into G'. In this way, all the buildings are connected to the substation most efficiently. This method can be viewed as a special case to the second method when the number of clusters is equal to the number of customers.

`

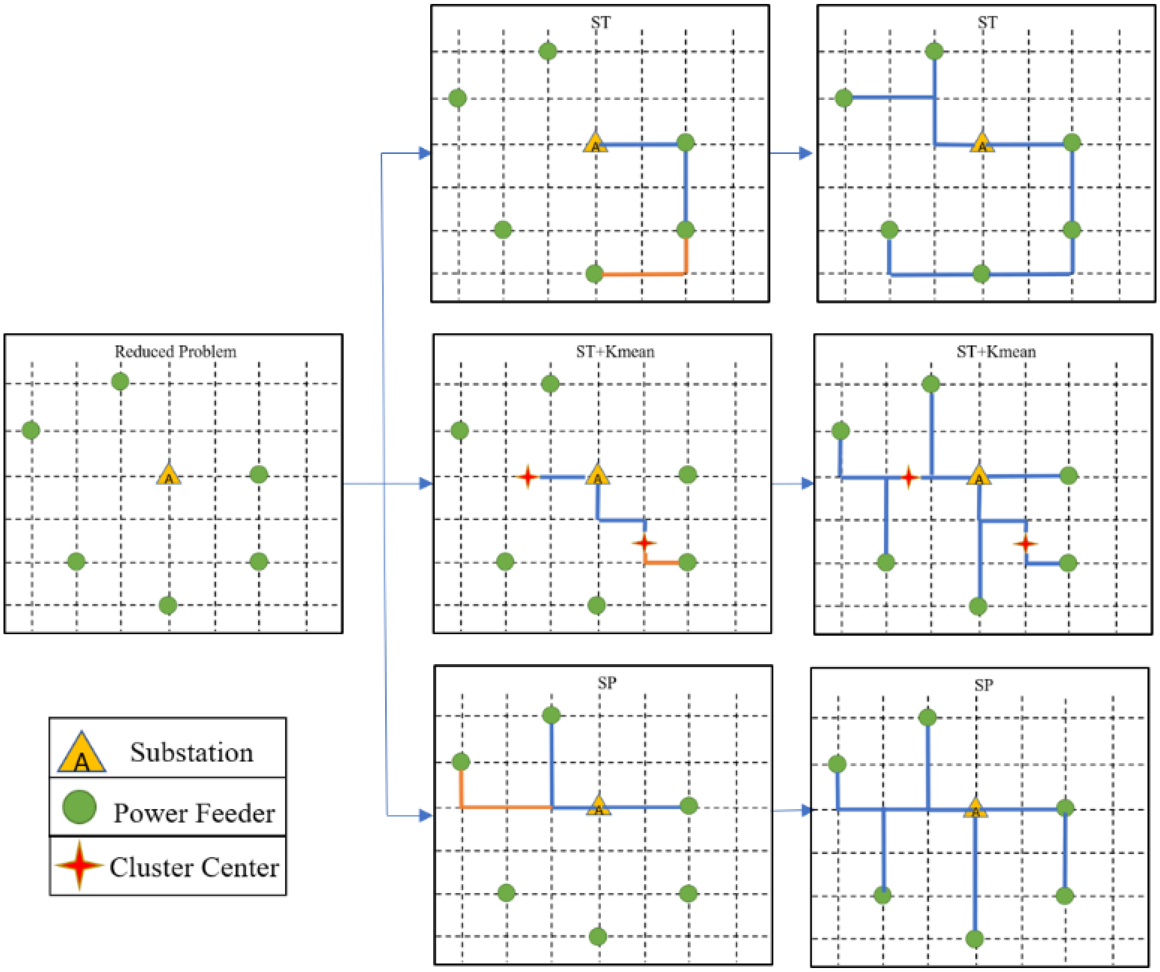


Figure 4-3 Illustrative examples of the three network connectivity generation algorithms

Each of these three methods generates a distribution layout for each substation cluster. The distribution network connects each customer to a power node, which is connected to the substation. This connectivity pathway is critical in the simulation step of our algorithm because we assume that if a customer is connected to a substation, they can receive power. That is, we ignore power flow constraints in our radial power distribution system. Therefore, the distance from the substation to a customer is crucial in determining the customer's probability of losing power. The further the customer is to the substation, the more components (e.g., poles) there likely are between the customer and the substation, increasing the possibility of a disruption on the distribution path to cut off power to the customer. Therefore, one of the validation criteria we use below is a

`

comparison of network distance from each customer to the substation in our synthetic network with that in the actual distribution network.

Simple examples of the three methods are shown in Figure 4-3. Green dots are power feeders that deliver power from the distribution network to customers, and the triangle is the substation. The goal is to connect all of them on the network grid with certain rules. We assume each dashed line represent roads, and we implement the three methods we proposed to solve the same problem. The three plots in the middle give examples of intermediate steps while solving the problem with each method. The orange lines are expected to be the next power lines to be added into the prior solution represented by blue lines. The three plots on the right are the final solution from each network generation method. The two red stars in the "ST+Kmean" plot are cluster centers.

4.4.2.5 Overhead/underground power lines classification

It is critical to determine whether each power line is overhead or underground because overhead and underground lines have substantially different vulnerabilities to hazards. We use a random forest classifier to estimate whether each line is overhead or underground, and the flowchart of this classification process for each substation cluster is shown in Figure 4-4. This approach uses the actual status – overhead vs. underground – of each line segment in the actual system we have data for together with housing characteristics of the area around the lines to develop a predictive machine learning classifier. The assumption here, based on observations working with data from multiple utilities, is that underground lines are more likely in certain types of areas such as those with newer homes with higher values.

We start by collecting real estate information for the area from *Zillow.com*. Other sources of real estate information could be used, but this website offers a convenient, free source of data. We

`

take a few factors into consideration as predictive variables and train a statistical learning model to classify the line type that connect to the building. Undergrounding technology became more prevalent in the 1970s in the U.S., and it tends to be associated with larger, more valuable homes.

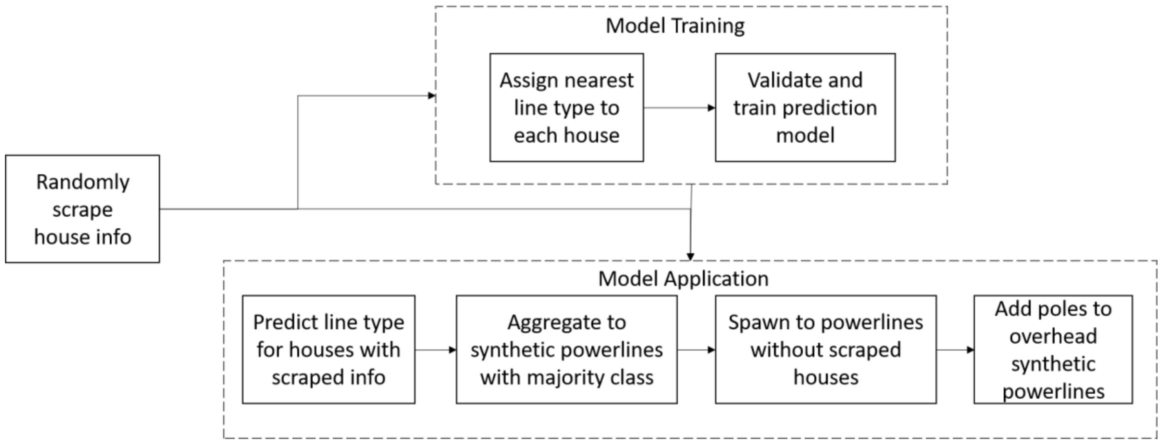


Figure 4-4 Classification process for determining overhead vs underground status of each line

For each home we query, we obtain the year built, the Zillow-estimated value, the finished size of the home, the parcel (lot) size, and the tax assessment as our covariates. The status of the power line nearest to the building (overhead vs. underground) is the response variable. We train random forest classifier and validate the model with holdout tests. We then finalize the model by training the random forest with that variable subset with the whole dataset.

In practice, due to the difficulties of scraping the housing data (Zillow API query is limited to 1000/day), we could only get house information for a small portion of houses. For example, it took us 14 days to retrieve information for 60,000 houses out of the 650,000 houses for Franklin County, Ohio. As a result, the line type of 90% of powerlines cannot be directly predicted by the model due to a lack of housing information on those roads. Therefore, we first predict the powerline type for each house with existing data. Then we aggregate each house's powerline type to its closest road by using the majority line type. For example, for a given synthetic powerline, if we scraped four houses along the powerline and three of the houses are predicted to be overhead and one to

`

be underground, then we classify the powerline type of this synthetic powerline to be overhead. If there is a tie, we choose overhead as the powerline type.

Then for synthetic powerlines without any scraped houses along them, we spawn their powerline types from powerlines that have been classified. To do this, we iterate through all the unclassified powerlines, and find their connected powerlines. The powerline type of undetermined powerlines will be randomly sampled from their neighboring and classified powerlines' types. If none of an unclassified powerline has been determined, this powerline will remain undetermined until the next round. After all the unclassified powerlines have been iterated, we start the process again for any powerlines that are still unclassified. The process ends when all the powerlines are assigned a powerline type.

4.4.3 Power outage simulation

With the synthetic power system generated, we can use it to simulate power outages due to natural hazards through use of infrastructure fragility curves. The simulation scheme is shown in Figure 4-5. This algorithm estimates the probability of each customer point losing power due to a given hazard loading scenario.

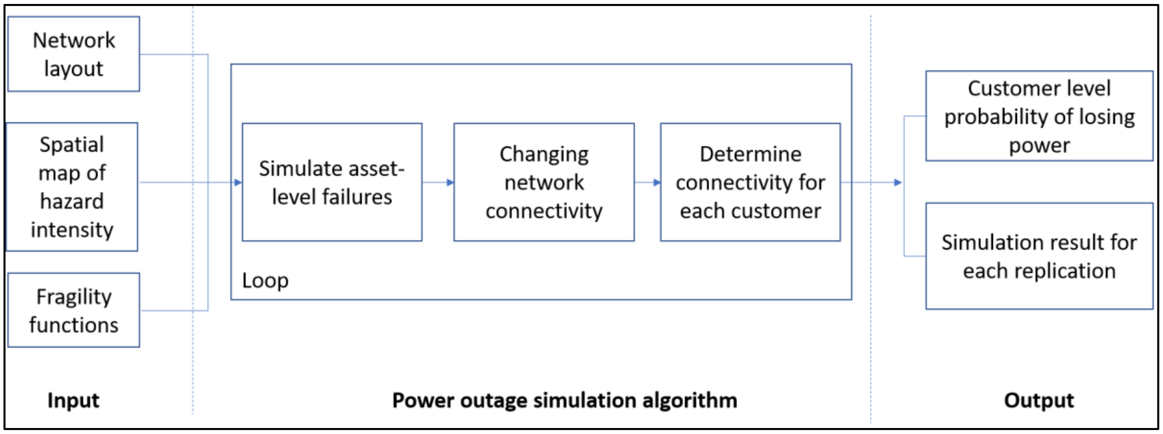


Figure 4-5 Simulation framework for estimating power outages at the household level given a network layout and asset-level fragility curves

Our synthetic power system can be used to simulate power outages for many types of natural hazards provided we have (1) a proper spatial map of the loading due to the hazard and (2) valid fragility functions for poles and substations that converts the hazard loading parameters into asset-level failure probabilities. We explain the framework with using strong wind events as an example.

As a starting point, we have as our inputs the network layouts for each substation cluster, a map of wind speeds over the study area, and fragility functions that give the probability of failure as a function of wind speed. Network layouts consist of the distribution layouts for each substation cluster of the study area. In our example we use three-second wind gust speed as our hazard loading measure. The fragility functions then give the probability of pole failures for a given wind speed. For each substation cluster, we simulate the power outages with a sufficient number of replications (1000) for the probability of power outages of each customer to converge. Within the simulation, we first simulate which infrastructures fail, i.e., poles and substations. Then we change the network structure by removing these assets. The last step is to check if each customer still has connectivity to the substation on the damaged network. If there is no route to connect them to the substation

`

because of pole or power line failures, or their substations are damaged, they will lose power. As a result, by integrating across all of the replications for each house, we can estimate the probability of a house losing power due to the simulated event, which can be informative to government, utility companies, and decision makers.

4.5 Results

4.5.1 Case study

Franklin County, Ohio is our primary case study. There are approximately 655,440 buildings in this county with a population of approximately 1.3 million. Our algorithm can conceptually be applied to any city, county, or an entire state if the required data is available and there are sufficient computational resources. We choose Franklin county, Ohio because we have the historical power outage data to compare our algorithm's output to, and we have the distribution system's actual layout which enables us to validate the layout generated by our model. We present our validations from two aspects, the similarity of our synthetic networks to the actual distribution network with multiple metrics, and the performance of our model in simulating a historical power outage event. We then show an application of the approach to a different area, Corpus Christi, Texas, for which we have some, but less spatially detailed, information to validate against. This provides at least some confidence in the ability to generalize the approach to other locations.

4.5.2 Network validation

The purpose of our synthetic model is to provide informative risk assessment of the impacts of hazards on the power system of an area. As such, comparing the end to end prediction performance using our model against power outages from historical events is important. However, we first

`

compare network structure of our synthetic power network with the actual distribution network and evaluate how accurate our overhead/underground classifier is in comparison to the actual layout. These similarity-based validation metrics are important because, if the algorithm accurately reflects key properties of real systems, we can have more confidence in the applicability of the approach to hazards beyond the limited set of validation events that we have to compare to.

4.5.2.1 Network similarity

To compare the similarity between networks, we first compare global network parameters, such as the average nodal degree, betweenness centrality, number of circles, and the total length of the network. These parameters are commonly used to represent the attributes of network graphs, but they are not necessarily informative for risk assessment purposes [(33)]. We compare these metrics for each substation cluster we create with all three methods.

Nodal degree is the number of nodes that directly connect to a given node. From Table 1 we see that the average nodal degrees we generate with all three approaches are all close to the actual distribution network. This is clearly the result of the fundamental structure of distribution network because most of the nodes (poles) in the graph are connected with only two other nodes aiming to deliver electricity. Betweenness centrality quantifies the number of times a node acts as a bridge along the shortest path between two other nodes. The average betweenness centrality is also at the same level between the two systems. The average Number of Circle describes the average number of loops in each substation cluster. We only consider the overhead system in this metric because underground systems are typically built as open loops with switches for extra robustness and from the data available to us form the utility, we cannot determine the actual connectivity of the undergrounding network. Table 4-1 shows that circles are rare in real overhead distribution networks which substantiates our assumption that the system is highly radial. Our synthetic

`

systems are, by design, completely radial. Lastly, we also compare the total length of our synthetic system versus the actual system and we find they are very close. These results show that the approximation using road network to the distribution network is reasonable in terms of network topological parameters.

Table 4- 1 Results of Comparing Network Topology Measures and Network Distance Between the Real and Synthetic distribution systems

	Average Nodal Degree	Average Betweenness Centrality	Average Number of Circle in Overhead	Total Length of Lines (Meters)	Mean Absolute Difference (m)	Pearson correlation
Actual System	2.066	0.0429	2.44	1.04×10^8	-	-
Steiner Tree Heuristic	1.999	0.0400	0	1.08×10^8	1426	0.699
Steiner Tree + Kmean	1.999	0.0389	0	1.09×10^8	882	0.745
Shortest Path	1.999	0.0363	0	1.07×10^8	693	0.842

One of the most important metrics for risk assessment purposes is the network distance from each customer to its substation. Due to the nature of radial systems, each customer is served by one substation. The probability of losing power for a customer is positively correlated to its network distance to the substation because the further the distance, the higher the number of potential failure points between the customer and the substation.

We compare the actual customer to substation network distance from the distribution layout with the synthetic networks generated by the three algorithms we propose. For each customer, on the synthetic network, we calculate the shortest path to the substation it is assigned to. If a customer is within multiple substation clusters, we use the shortest one for comparison. The result is shown in the last two columns in Table 4-1. In terms of each buildings distance to the nearest substation, the best model we find is the shortest path model. The average absolute difference in distance to the actual network is 693m. We also calculate the Pearson correlation and the shortest path model

`

outperforms the other two models as well. Pearson correlation coefficient measures the linear correlation between two variables. These results suggest that our model, while not perfect, generates synthetic systems with customer distances to substations that correlate well with the actual values.

4.5.2.2 Overhead/underground power lines classification validation

The second step is to evaluate the accuracy of our overhead/underground classifier. We first test the accuracy of the random forest model given our dataset. The in-sample prediction accuracy with the whole dataset is 100%. More meaningful is the out of sample accuracy. The average out-of-sample prediction accuracy in 30 repeated random holdouts is 91%. 10.6% of the synthetic powerlines are directly predicted from the dataset and the accuracy is 100% (due to in-sample prediction). After applying the powerline type spawning algorithm, the overall prediction accuracy for the whole network is 84.1% for our study region. One issue of our model is that it cannot capture commercial buildings because Zillow does not provide such information. One potential solution is to identify commercial area and use a pre-defined line type for power lines surrounding commercial areas.

The level of accuracy of our model is strong even given our limited dataset, and it may improve if information on more buildings or other sources of additional real estate data become available. We have also directly compared the generated systems to the actual systems visually on a map. However, for reasons of security, the data provider does not allow the actual maps to be shown. However, the systems are visually similar.

`

4.5.3 Extreme weather simulation

4.5.3.1 Franklin County, Ohio - Derecho, 2012

In this section, we use the model we developed to predict power outages the Derecho that impacted Franklin County, Ohio on June 29, 2012. The event caused more than 50% of customers in the county to lose power. We have the utility outage data, aggregated to 5km grid cells, to compare with our model outputs. During the event, wind was the driving force of power outages. We retrieved the maximum gust wind speed during the event for all the airports in or near the county and interpolate to get the gust wind speed for each pole given the pole's spatial location to those airports.

We applied the fragility curves developed by Darestani and Shafieezade [(31)]. The fragility of poles under wind events is determined by the pole class, wind speed and direction, age, diameter of conductors, heights. By inputting these parameters, we can estimate the probabilities of pole failure and use these to simulate the change of network connectivity. We repeated out analysis for different combinations of pole classes and ages and conduct sensitivity analysis to cover the gap of unknown pole information. We do not consider damage to falling trees due to lack of information, though we acknowledge that tree failures can be substantial causes of outages. We conduct a convergence test on the average customers without power for several substation clusters with 50,000 replications. We find that the relative difference for the average customers without power is converging to less than 1% after 10,000 replications.

We compare our simulation results with the actual event in 5km by 5km grids which are shown in Fig. 6. We assumed all the poles are the same throughout the system in each run, and we tested

`

different types of poles in each scenario: 60-year-old class 4 poles, 50-year-old class 5 poles, and 60-year-old class 5 poles. These pole ages were chosen based on the average age of buildings in the region. The class of poles is determined by the minimum circumference that depends on the species of tree and the length of the pole [(46)]. Higher-class poles (e.g. class five) typically can hold less horizontal load than lower-class poles (e.g. class four). Class four and five poles are typically used in distribution systems in the U.S.

From these simulation scenarios we gain insights into the relative risks of losing power. In all cases (plots (b)-(d) in Figure 4-6), the model is capturing the power outages on the northern side of the region reasonably well but underestimating outages in the southern portions of the region. Part of the reason for this is because we do not have the most accurate wind speed data for all grid cells, but only the seven major airports in this area. These airports are mostly located in the northern part of the county. Plots (b) and (d) are both for 60-year-old poles. These two model runs give a total number of customers without power relatively close to the actual 279,000. The major difficulty in simulating this Derecho event is the to a lack of information on the hazard loading (i.e., wind speeds). We use only gust wind speed at major airports to approximate the gust wind speeds throughout the system. Informal conversations with the meteorologists from the utility that provided the data revealed that wind speeds were highly spatially variable during this event, and we lack the data to capture these local differences. To improve the simulation results, a more detailed wind speed map would be helpful.

`

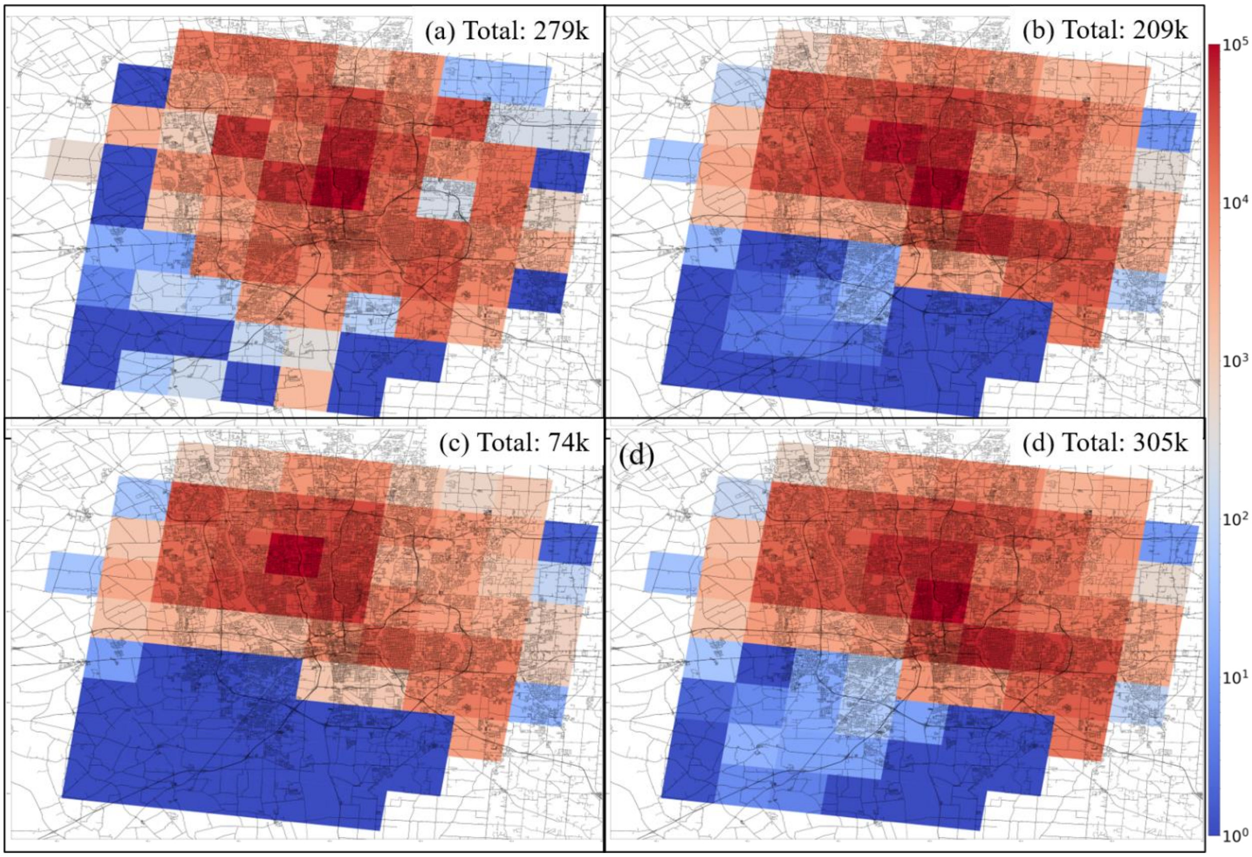


Figure 4-6 Comparison of actual and simulated power outages for the 2012 Derecho in Franklin County, OH. The legend shows the average number of customers without power in each grid cell. (a) actual power outage during the derecho, (b) If all the poles are class 4 and age 60, (c) If all the poles are class 5 and age 50, (d) If all the poles are class 5 and age 60. Total number of customers without power is also shown in each scenario.

4.5.3.2 Corpus Christi - Harvey, 2017

To test the generalizability of the approach to other areas and other events, we also simulated the impacts of Hurricane Harvey in southern Texas. We do not have the actual outage data for this area in the same level of detail but have obtained estimates of outages from media reports. We also do not have the full system topology to compare to. We should note that we are focusing on the area of Texas that experienced Harvey as a strong wind event, not primarily a flooding event (i.e., we are not considering the Houston area).

Hurricane Harvey made landfall in Texas on August 23, 2017. The hurricane caused considerable damage loss of life. During the event, widespread power outages occurred in multiple

`

major cities due to strong wind, hurricane surge, and flooding caused by rainfall. We selected Corpus Christi to test our model because the majority of power outages there were wind driven. We use the best track estimation[8] of hurricane Harvey and an existing hurricane wind field model [(47)] to estimate the 3-second gust wind speed for each distribution substation. We then apply the same wind speed from each substation to all the poles connected to that substation and simulated outages. The results are shown in Figure 4-7. Red areas are those areas where the buildings are more likely to lose power and blue are less likely to lose power. We assume the parameter for poles to be 40-year-old class 4 poles, 50-year-old class 4 poles, and 40-year-old class 5 poles based on the building stock age for the city.

As a comparison, from media reports, the peak number of customers without power reported by AEP, the utility serving the area, during Harvey for Corpus Christi was approximately 91,500. The majority of the outages were in the Midtown area and the Southside area while the Northwest area was mildly damaged. With our simulation, we estimate the average number of customers without power as 67,000, 84,000, and 82,000, where the estimates from the latter two sets of pole parameters are close to the actual value. The model estimates the risk of losing power to be high in the midtown area for all three sets of pole parameters. For the southside area, the outage estimates are centered because the model estimates that the three substations in this area have primarily underground power lines serving customers, which is generally true after checking from google streetview. The other substations' service territories are severely damaged. For the northwest area, especially in the scenario of 40-year-old class 4 poles, there are less customers damaged. While we do not have detailed ground truth data to compare to, overall the model is in

---

8 https://www.nhc.noaa.gov/gis/archive_besttrack_results.php?id=al09&year=2017&name=Hurricane%20HARVEY

`

agreement with the media reports of outages, at least at a high level. Overall, these simulation results can be a good source of information to the public, critical infrastructures, and utility companies to assess the potential chances of losing power and find a better way to mitigate.

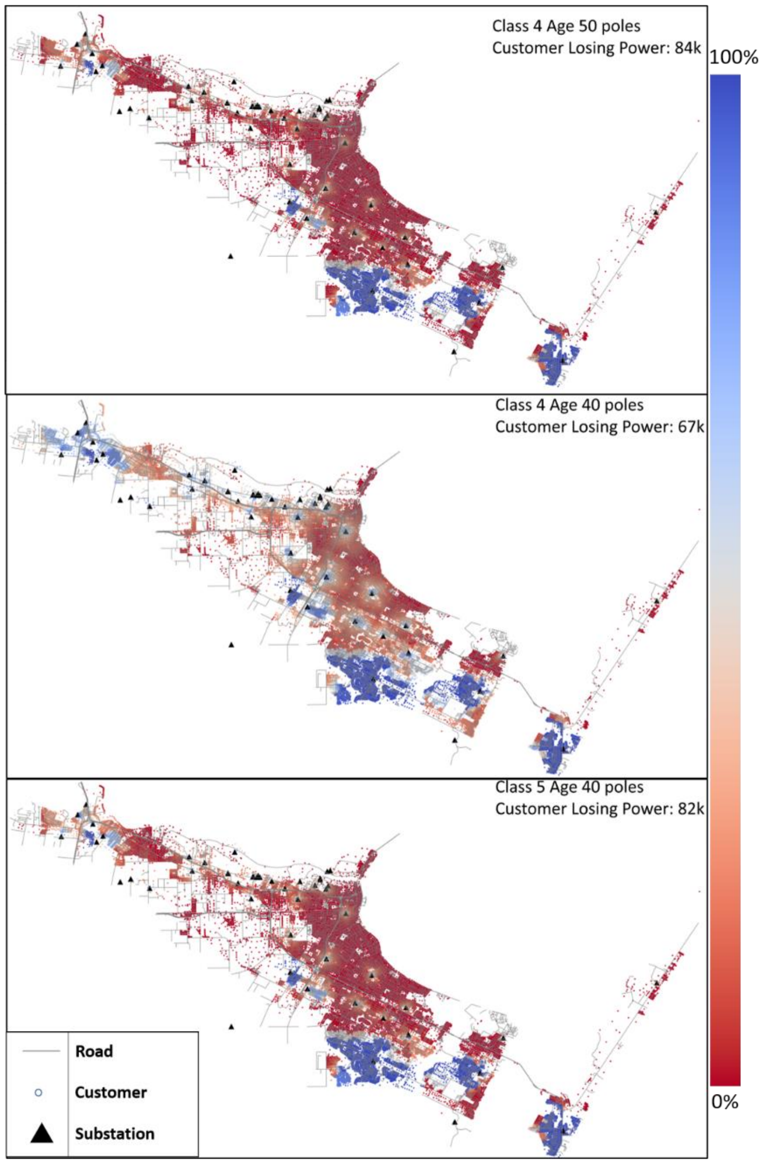


Figure 4-7 Simulation for Hurricane Harvey, 2012 for Corpus Christi. The legend shows the probability each customer with power. Closing to 100% indicates the customer will have a high probability of having power. Closing to 0% indicates the customer will have a lower probability of having power.

`

4.6 Conclusion

We develop a model to use synthetic network generation to cover a critical gap in power outage risk assessment research. We use publicly available data to create a synthetic version of the distribution system given building information, substation locations, and road networks. The method is generalizable to larger scales, from city level to county level, and potentially to state level. We tested our approach for Franklin county, Ohio, and Corpus Christi, TX. For Franklin county, we compared our network with the actual distribution network with multiple measures and found that the model approximated the actual system well in terms of topological characteristics. We also simulated two major power outage events, the Derecho in Ohio in 2012 and hurricane Harvey in Corpus Christi in 2017. The results of these simulations show that the model can produce accurate estimates of power outages provided accurate hazard loading maps are available.

However, there are several limitations in this method that needs to be addressed. What we are creating is a synthetic distribution network that can be representative of the real-life system and can be used to estimate damage to the distribution system and household likelihood of losing power in the US. For other countries, it is not clear at this point if the method is applicable because 1) certain assumptions may no longer be held (i.e., that the distribution system is radial), 2) certain input data may become harder to acquire, (i.e., building information and substation locations), and 3) the overhead and underground power line classifier may no longer work because of differences in how the system is designed. Some of these limitations can be resolved by introducing extra models to impute missing information, such as using population and power consumption information to create synthetic distribution substations. In the meantime, with better input data, the performance of certain models can be improved. For example, with a more detailed building information, the classification of lines as overhead or underground may become more accurate. In

`

addition, more information regarding the age of certain infrastructure such as poles can help improve the outage estimation. Overall, our work shows the effectiveness of the method with limited information.

Another major challenge of the proposed method is its scalability. While the method has been designed for use at the scale of a metropolitan area, there may be interest in scaling up to a state or national level. Scaling up to state-wide or nation-wide analysis with this approach would have three major challenges: data collection, computing power limitations, and fragility function limitations. The first challenge would be how to collect the necessary input data for the model if applying the model to the state or national level. Road locations, building locations, and substation locations can all be retrieved publicly for the entire U.S., but the effort required to gather this data grows substantially for larger application domains. For building information, one possibility would be to switch to a commercially available database of building locations and characteristics. The second challenge is the computational power required for both network generation and network simulation. Highly optimized code and parallelization can increase the feasibility of this application. Variance reduction and sampling strategies can be leveraged for faster convergence. Even with these strategies, computational effort will remain a challenge for very large spatial domains. The last challenge is regarding fragility functions. A comprehensive set of fragility functions for vulnerable components of the distribution system for the hazard type being modeled is crucial in improving the accuracy of the estimation. For certain components under some disasters, the fragility functions available may not exist up to date. Resolving this issue will rely on the development of improved fragility functions.

Though only tested for wind events in this chapter, we have run the model for other hazard types. We use only wind events as an example in this chapter because of the availability of data to

`

validate the performance of our model against historical data. Using the model for a different hazard requires hazard-appropriate fragility functions and potentially the inclusion of additional system components if they are vulnerable under the other hazards. For example, for earthquakes, the damage to the distribution system comes largely from damage to low-voltage substations and utility poles, requiring seismic fragility functions for these components. On the other hand, for flooding, the height of certain utility equipment, such as substation transformers should be explicitly modelled to estimate their failure probabilities.

Overall, our model provides a novel approach for estimating the building-level probability of losing power for a natural hazard event. We have shown that this approach can yield accurate predictions provided an accurate map of hazard loading is used and appropriate fragility functions are used. This approach has the potentially to improve predictive accuracy of power outage estimation and to provide substantially finer spatial detail in predictions than existing approaches provide.

4.7 References

1. Hines P, Apt J, Talukdar S. Trends in the history of large blackouts in the United States. P. in IEEE Power and Energy Society 2008 General Meeting: Conversion and Delivery of Electrical Energy in the 21st Century, PES. 2008.

2. Kezunovic M, Obradovic Z, Dokic T, Zhang B, Stojanovic J, Dehghanian P, Chen P-C. Predicting Spatiotemporal Impacts of Weather on Power Systems Using Big Data Science. Pp. 265–299 in Studies in Big Data. Vol 24. Springer, Cham, 2017.

3. Eisenbach Consulting LLC. 9 of the Worst Power Outages in United States History. Electric Choice, 2017:https://www.electricchoice.com/blog/worst-power-ou.

4. Johnson BW. AFTER THE DISASTER: Utility Restoration Cost Recovery Exhibit MP1., 2005.

5. Guikema SD, Nateghi R, Quiring SM, Staid A, Reilly AC, Gao M. Predicting Hurricane

`

Power Outages to Support Storm Response Planning. IEEE Access, 2014; 2:1364–1373.

6. Shafieezadeh A, Onyewuchi UP, Begovic MM, Desroches R. Age-dependent fragility models of utility wood poles in power distribution networks against extreme wind hazards. IEEE Transactions on Power Delivery, 2014; 29(1):131–139.

7. Salman AM, Li Y, Stewart MG. Evaluating system reliability and targeted hardening strategies of power distribution systems subjected to hurricanes. Reliability Engineering and System Safety, 2015; 144:319–333.

8. Campbell RJ. CRS Report for Congress Weather-Related Power Outages and Electric System Resiliency Specialist in Energy Policy Weather-Related Power Outages and Electric System Resiliency Congressional Research Service Weather-Related Power Outages and Electric System Resiliency Congressional Research Service., 2012.

9. Davidson RA, Liu H, Sarpong K, Sparks P, Rosowsky D V. Electric power distribution system performance in Carolina Hurricanes. Natural Hazards Review, 2003; 4(1):36–45.

10. Deka D, Backhaus S, Chertkov M. Learning topology of distribution grids using only terminal node measurements. Pp. 205–211 in 2016 IEEE International Conference on Smart Grid Communications, SmartGridComm 2016. Institute of Electrical and Electronics Engineers Inc., 2016.

11. Abdul Rahman F, Varuttamaseni A, Kintner-Meyer M, Lee JC. Application of fault tree analysis for customer reliability assessment of a distribution power system. Reliability Engineering and System Safety, 2013; 111:76–85.

12. Davidson RA, Liu H, Sarpong K, Sparks P, Rosowsky D V. Electric power distribution system performance in Carolina Hurricanes. Natural Hazards Review, 2003; 4(1):36–45.

13. Han SR, Guikema SD, Quiring SM. Improving the predictive accuracy of hurricane power outage forecasts using generalized additive models. Risk Analysis, 2009; 29(10):1443–1453.

14. Han SR, Guikema SD, Quiring SM, Lee KH, Rosowsky D, Davidson RA. Estimating the spatial distribution of power outages during hurricanes in the Gulf coast region. Reliability Engineering and System Safety, 2009; 94(2):199–210.

15. Guikema SD, Quiring SM, Han SR. Prestorm Estimation of Hurricane Damage to Electric Power Distribution Systems. Risk Analysis, 2010; 30(12):1744–1752.

16. Nateghi R, Guikema S, Quiring SM. Power Outage Estimation for Tropical Cyclones: Improved Accuracy with Simpler Models. Risk Analysis, 2014; 34(6):1069–1078.

17. McRoberts DB, Quiring SM, Guikema SD. Improving Hurricane Power Outage Prediction Models Through the Inclusion of Local Environmental Factors. Risk Analysis, 2018; 38(12):2722–2737.

`

18. Shashaani S, Guikema SD, Zhai C, Pino J V., Quiring SM. Multi-Stage Prediction for Zero-Inflated Hurricane Induced Power Outages. IEEE Access, 2018; 6:62432–62449.

19. Kabir E, Guikema SD, Quiring SM. Predicting Thunderstorm-Induced Power Outages to Support Utility Restoration. IEEE Transactions on Power Systems, 2019; 34(6):4370–4381.

20. Han SR, Guikema SD, Quiring SM, Lee KH, Rosowsky D, Davidson RA. Estimating the spatial distribution of power outages during hurricanes in the Gulf coast region. Reliability Engineering and System Safety, 2009; 94(2):199–210.

21. He J, Wanik DW, Hartman BM, Anagnostou EN, Astitha M, Frediani MEB. Nonparametric Tree-Based Predictive Modeling of Storm Outages on an Electric Distribution Network. Risk Analysis, 2017; 37(3):441–458.

22. Guikema SD. Natural disaster risk analysis for critical infrastructure systems: An approach based on statistical learning theory. Reliability Engineering and System Safety, 2009; 94(4):855–860.

23. Guikema SD, Quiring SM. Hybrid data mining-regression for infrastructure risk assessment based on zero-inflated data. Reliability Engineering and System Safety, 2012; 99:178–182.

24. Cerrai D, Wanik DW, Bhuiyan MAE, Zhang X, Yang J, Frediani MEB, Anagnostou EN. Predicting Storm Outages Through New Representations of Weather and Vegetation. IEEE Access, 2019; 7:29639–29654.

25. D’Amico DF, Quiring SM, Maderia CM, McRoberts DB. Improving the Hurricane Outage Prediction Model by including tree species. Climate Risk Management, 2019; 25:100193.

26. Staid A, Guikema SD, Nateghi R, Quiring SM, Gao MZ. Simulation of tropical cyclone impacts to the U.S. power system under climate change scenarios. Climatic Change, 2014; 127(3–4):535–546.

27. Vickery PJ, Skerlj PF, Lin J, Twisdale LA, Young MA, Lavelle FM. HAZUS-MH hurricane model methodology. II: Damage and loss estimation. Natural Hazards Review, 2006; 7(2):94–103.

28. Kircher CA, Whitman R V., Holmes WT. HAZUS earthquake loss estimation methods. Natural Hazards Review, 2006; 7(2):45–59.

29. Vanzi I. Seismic reliability of electric power networks: Methodology and application. Structural Safety, 1996; 18(4):311–327.

30. Han S-R, Rosowsky D, Guikema S. Integrating Models and Data to Estimate the Structural Reliability of Utility Poles During Hurricanes. Risk Analysis, 2014; 34(6):1079–1094.

`

31. Mohammadi Darestani Y, Shafieezadeh A. Multi-dimensional wind fragility functions for wood utility poles. Engineering Structures, 2019; 183:937–948.

32. Guikema SD, Davidson RA, Liu H. Statistical models of the effects of tree trimming on power system outages. IEEE Transactions on Power Delivery, 2006; 21(3):1549–1557.

33. Larocca S, Johansson J, Hassel H, Guikema S. Topological Performance Measures as Surrogates for Physical Flow Models for Risk and Vulnerability Analysis for Electric Power Systems., 2013.

34. Birchfield AB, Gegner KM, Xu T, Shetye KS, Overbye TJ. Statistical Considerations in the Creation of Realistic Synthetic Power Grids for Geomagnetic Disturbance Studies. IEEE Transactions on Power Systems, 2017; 32(2):1502–1510.

35. Pahwa S, Scoglio C, Scala A. Abruptness of cascade failures in power grids. Scientific Reports, 2014; 4.

36. Schultz P, Heitzig J, Kurths J. A random growth model for power grids and other spatially embedded infrastructure networks. European Physical Journal: Special Topics, 2014; 223(12):2593–2610.

37. Soltan S, Zussman G. Generation of synthetic spatially embedded power grid networks. P. in IEEE Power and Energy Society General Meeting. Vol 2016-Novem. 2016.

38. Pisano G, Chowdhury N, Coppo M, Natale N, Petretto G, Soma GG, Turri R, Pilo F. Synthetic Models of Distribution Networks Based on Open Data and Georeferenced Information. Energies, 2019; 12(23):4500.

39. Schweitzer E, Member S, Scaglione A, Monti A, Member S, Andrea Pagani G. Automated Generation Algorithm for Synthetic Medium Voltage Radial Distribution Systems. IEEE JOURNAL ON EMERGING AND SELECTED TOPICS IN CIRCUITS AND SYSTEMS, 2017; 7(2).

40. Miranda V, Ranito J V., Proena LM. Genetic algorithms in optimal multistage distribution network planning. IEEE Transactions on Power Systems, 1994; 9(4):1927–1933.

41. Valenzuela A, Inga E, Simani S. Planning of a resilient underground distribution network using georeferenced data. Energies, 2019; 12(4):644.

42. Yuan W, Wang J, Qiu F, Chen C, Kang C, Zeng B. Robust Optimization-Based Resilient Distribution Network Planning Against Natural Disasters. IEEE Transactions on Smart Grid, 2016; 7(6):2817–2826.

43. National Academy of Sciences NA of E and NRC. America's Energy Future: Technology and Transformation. National Academies Press, 2010.

44. Takahashi H, Matsuyama A. An approximate solution for the Steiner tree problem in graphs. 1980.

`

45. Garey MR, Johnson DS. The Rectilinear Steiner Tree Problem is NP-Complete. SIAM Journal on Applied Mathematics, 1977; 32(4):826–834.

46. Wolfe R. STANDARD SPECIFICATIONS FOR WOOD POLES., 1997.

47. Holland GJ, Belanger JI, Fritz A. A Revised Model for Radial Profiles of Hurricane Winds. Monthly Weather Review, 2010; 138(12):4393–4401.

Chapter 5: Conclusion

Natural disasters bring huge challenge to our lives, to the infrastructure, and to the economy. A better understanding of the long-term vulnerability of individuals and communities and the short-term vulnerability of infrastructure can give insights and support better planning, preparation, recovery, and retrofitting. This study builds models to study how the long-term vulnerability of communities evolves under repeated hurricane and heat waves, and to estimate the short-term vulnerability of power distribution systems. Simulation models, agent-based models, network optimization and analysis, and predictive modeling approaches are used to answer these questions and advance vulnerability analysis.

5.1 Summary and contributions

5.1.1 Evolution of the long-term vulnerability under repeated hurricanes

Hurricanes can repeatedly cause extensive economic losses in the U.S especially to residential building stock. However, little is understood about how repeated hazards change communities and impact their preparation for future events. Individual mitigation actions may drive how a community's resilience evolves under repeated hazards. In chapter 2, I investigate the effect that learning by homeowners has on household mitigation decisions and on how this influences a region's vulnerability to hurricanes over time. I build an Agent-Based Model (ABM) to simulate homeowners' adaptation to repeated hurricanes and how this affects the vulnerability of the regional housing stock. I apply this model in nine coastal counties in Maryland to study

`

how different initial beliefs about the hurricane hazard and how the memory of recent hurricanes could change a community's vulnerability both under current and potential future hurricane scenarios under climate change. In some future hurricane environments, different initial beliefs can result in large differences in the region's long-term vulnerability to hurricanes. I find that getting some homeowners to mitigate right after a hurricane given that this is when their memory of the event is the strongest can help decrease the vulnerability of a community most effectively.

There are three major contributions of this work that can have substantial impact to future studies in this field. First, this work proposes a general framework for how to analyze the long-term vulnerability of a community facing repeated hazards. The framework can be used to study different regions, different hazards, and different agents. Second, the created ABM for repeated hurricanes serves as a flexible boundary object that can be used in interdisciplinary studies to understand each component of the complex system. The introduction of agent learning and decision-making models are the first step into creating more comprehensive scenarios in long-term hurricane planning. Third, this work also reveals the impact of learning and memory decay from homeowners and how their decisions can alter the evolution of vulnerability of the community under repeated hurricanes.

5.1.2 Evolution of the long-term vulnerability under repeated heat waves

Repeated heat waves impact vulnerable populations such as the elderly and lower income groups that are less able to respond to high heat events. The vulnerability of these and other groups to heat waves changes over time in response to social connections, collective action, utility coordination, and government policies. In chapter 3, I create an agent-based modeling framework that integrates an environment that experiences repeated heat waves, dynamically evolving agents, social networks, and a health mortality model to study the impact of repeated heatwaves on society. In this work, I find that a change of climate can influence the long-term vulnerability of a community in surprising ways. Increasing daily average temperatures can stimulate more individual mitigation behaviors and reduce the number of people that die from future heat waves. Social networks can also have a substantial impact in helping decrease mortality by providing options for cooling at friends' houses. On the other hand, cooling centers can be used effectively in helping decrease mortality for a given event, but they may also increase the long-term vulnerability by decreasing people's motivation to adopt mitigation, which increases the risks when facing an unexpected or extreme event. Only by modeling adaptive behavior and social interactions over time can we start to understand how community vulnerability evolves over time.

There are two major contributions of this work that can have substantial impact to future studies in this field. This study creates a platform to study the long-term dynamics of community vulnerability under repeated heat waves. It can be used to incorporate different individual learning, behavior, and interaction models and test the output's sensitivity from the change of each component. This work incorporates social networks, which is something often not thoroughly considered in other disaster simulation studies. I explored how individual mitigation choices, government policies, and social connections can result in different patterns of the evolution of vulnerability. Social connection along with government intervention such as cooling center can make a great difference to the mortality outcome.

5.1.3 Weather-related power outage prediction

`

Power outage prediction can be valuable for both pre-hazard planning of resources as well as post-hazard or long-term system hardening planning. In chapter 4, I create an algorithm that (1) generates a synthetic power distribution system layout for any U.S. city based only on public data and then (2) simulates power outages at the level of individual buildings under hazard loading using fragility functions. This approach provides localized, building-level estimates of the likelihood of losing power due to a natural hazard. I validate this model by comparing the network properties and power outage events based on our approach with data from a real power system in Ohio. I then apply this model to estimate power outages for a historical Derecho event and a historical hurricane event. I find that this model relies on less input data comparing to statistical learning approaches yet can make accurate predictions, provided accurate fragility curves are available.

There are three major contributions of this work that can have a substantial impact on future studies in this field. To fill the gap of the partially unknown distribution network for power outage prediction purposes, I use publicly available information of substations to create a synthetic distribution network. I compare different network generation models and find the best model comparing to the actual distribution network. Then, because of the huge difference in response to hazards for overhead and underground distribution systems, I create a random forest classifier that can differentiate these two types of systems given characteristics of nearby customer buildings.

5.2 Limitation and future work

5.2.1 Evolution of the long-term vulnerability under repeated hurricanes

`

This is the first project I have been working on since joining the Ph.D. program. The model created has potential in studying different aspects of the long-term vulnerability of a community under repeated hurricanes. In this work, I investigated in the effect of learning particularly. I also created models to study the impact of different learning and behavioral models, different government policies, land use changes and real estate markets, hurricane surge, and flood insurance. For example, the behavioral model of the agents can be very interesting to understand either an optimal solution, or the most realistic decision making to the changing environment. These are good topics that can be addresses with extensions of the current model. The current model focuses on only residential buildings. Another interesting extension of the model can be on a different subject. Comparing damages to residential building stocks, it is also very critical to study the impact on commercial buildings and analyze how their mitigation efforts can help save their potential future losses.

5.2.2 Evolution of the long-term vulnerability under repeated heat waves

In this work, I apply a model structure that is similar to the hurricane work but on a different type of hazard. Though the framework is similar, I needed to reconsider the best way to understand this disaster and how is it different from hurricanes. For heat waves, the focus is on the interplay of social networks, individual decisions, and government intervention. Social networks and government interventions are modeled in this work to help us understand their impacts on the long-term vulnerability of the community. Similar to the hurricane work, it is also interesting to discuss how different behavioral models and government interventions can change the evolution of the community. It is also interesting to see how we can combine the hurricane work and this heat wave work together and propose metrics to measure the vulnerability of a

`

community and evaluate this complex system. This can be the first step into creating a multi-hazard simulation platform for long-term hazard planning.

5.2.3 Weather-related power outage prediction

Power outage prediction is a topic I have been very interested in since the beginning of my Ph.D. I have been working on both statistical based approach and simulation-based approach for power outage prediction. For both methods, they have restrictions that limit them from being a universally applicable model.

The primary challenge for statistical-based approach is its generalizability. We have seen from many works that with sufficient data for certain region for a certain type of hazard or weather event, the trained model can be relatively accurate in out-of-sample predictions and can be used in practice. However, if the hazard is a rare event without much data, i.e., earthquakes, or the study region's system or infrastructure characteristics are very different from the training region, then the prediction accuracy may be compromised. For a simulation-based method the challenge is its scalability. While the method has been designed for use at the scale of a metropolitan area, there may be interest in scaling up to a state or national level. Scaling up to state-wide or nation-wide analysis with this approach would have three major challenges: data collection, computing power limitations, and fragility function limitations as I have discussed in detail in the chapter. These two challenges limit the potential for an all hazards power outage prediction model, which can be very valuable in helping understand business downtime and potential mitigations that can decrease loss.

Appendix A: Synthetic Hurricane Generation

Synthetic hurricanes are generated outside of the ABM using the four-step process developed in Staid *et al*.[1]. These steps are:

a. Choose an initial windspeed and location for landfall: The US coastline is divided into 50 km bins and the bins are populated by the number of hurricanes to make landfall within its boundaries. The initial location at landfall is randomly selected from these bins in proportion to historic landfall occurrences. The windspeed is sampled from the historical record;

b. Using the initial windspeed and location, generate a hurricane track from a non-parametric random forest model trained on the suite of historical US hurricane tracks[1]. The track reports the location of the center of the hurricane in six-hour increments;

c. Compute the peak 3-second 10-m peak wind gust at the center of each parcel in the study area using a parametric hurricane wind field and decay model[2];

d. Discard any hurricane that does not impact the study region.

This is repeated until a library of 36,399 synthetic storms to impact the study region is generated. The storms range in intensity from tropical storms to Category 5 hurricanes.

`

Next, the case study area's historic hurricane intensity and frequency records are fitted to Poisson and Weibull distributions, respectively, to form baseline hurricane scenarios as shown in equations (3) and (4):

$$N_i \sim Poisson(cF) \qquad (3)$$

$$v_{max} \sim Weibull(\alpha I, \beta) \qquad (4)$$

Here N_i is the number of hurricanes that impact the case study area in year i, c is the historical annual frequency with which hurricanes make landfall in the study region and equals 0.1373, and F is the multiplier that controls hurricane frequency. $F = 1.0$ implies the baseline case. v_{max} is the maximum windspeed for each sampled hurricane. The parameters α and β describe a two-parameter Weibull distribution fitted using historical maximum hurricane windspeed in this region, where $\alpha = 67.76$ and $\beta = 3.64$. I is the multiplier that controls hurricane intensity by changing the scale parameter. $I = 1.0$ implies a baseline case. These distributions are used in the ABM to sample both the number of hurricanes that occur in a year and the intensity of each hurricane. The synthetic track in the library with the maximum 3-second peak wind gust closest to the sampled hurricane landfall windspeed is selected.

References List:

1. Staid A, Guikema SD, Nateghi R, Quiring SM, Gao MZ. Simulation of tropical cyclone impacts to the U.S. power system under climate change scenarios. Climatic Change, 2014; 127(3–4):535–546.
2. Han SR, Guikema SD, Quiring SM, Lee KH, Rosowsky D, Davidson RA. Estimating the spatial distribution of power outages during hurricanes in the Gulf coast region. Reliability Engineering and System Safety, 2009; 94(2):199–210.

`

Appendix B: Decision Model

For the decision-making model, we use a decision tree to enumerate all possible alternatives that an agent can choose given their previous upgrade history. In each simulated year with at least one hurricane, each agent chooses from a set of alternatives that increases the wind resistance of their house or do nothing. We use utility theory to model each agent's preference.

In the simulation, each agent's parcel is initially assigned a building type, which is extracted from the Maryland Department of Planning[9]. There are 11 building types, defined by the building materials and number of stories. Examples of these are a single-story wood-framed homes and a two-story unreinforced masonry homes. For each building type, different upgrading options to reduce wind damages are available (e.g., adding roof-to-wall straps). For each type of building and given upgrade alternative (if any) selected, we use the fragility curve which estimates the probabilities of that building being in each of the possible damage states (i.e., damage states 1-4) from a given windspeed from HAZUS. Each agent incorporates this fragility function with their probabilistic beliefs about the frequency and intensity of hurricanes to calculate the expected utility of each alternative and then choose the alternative with the highest expected utility. Each upgrade can only be applied once, and we assume the reliability of the

[9] http://planning.maryland.gov/OurProducts/downloadFiles.shtml

`

upgrade does not decay as time progresses in the simulation. A detailed description of build stock information can be found in Reilly et al.

We define the current state of a house as s_0, which is the combination of its building type and upgrades previously done to that house. The potential future states of the house are denoted as $s_i, 0 \leq i \leq n$, where n is the number of different alternatives the homeowner has for upgrading their house. A cost $c_i, 0 \leq i \leq n$ is associated with each choice. After a storm, the house can either be not damaged or in one of four different damage states (from minor damage to completely destroyed) with damage costs c_{d0}, c_{d1}, c_{d2}, c_{d3}, c_{d4} equal to 0, 0.05,0.2,0.45, and 0.99 times the house value W.

The agent updates their estimates of the probability of occurrence of each category of hurricanes written as $p_j, 1 \leq j \leq 7$, corresponding to the probability of no hurricane, a tropical storm, and Category 1-5 hurricanes. This is updated each year based on new observations by the agent in each simulated year. The probability of being in each damage state k, $0 \leq\ k \leq 4$, given the average wind speed for each category of hurricanes and the potential state of the house is defined as $p(d_k|s_i, Hurr\ j)$, $1 \leq j \leq 7$. We assume the agent knows these probabilities when making mitigation decisions. We use a risk averse exponential utility function shown in equation (1) to quantify the preference over outcomes

$$U(x) = 1 - e^{-x/R} \tag{1}$$

where R is the risk aversion factor. In the model we select R to be $0.05 * W$ where W is the wealth of each agent. We use the house value as a proxy for wealth because we lack household-level net worth information. As a result, wealthy people can afford more expensive upgrades and to be more risk seeking than less wealthy people.

`

The agents also consider the long-term return from their investment on house upgrades, i.e., what is their return over the next T years with a discount factor γ. $T = 10$ and $\gamma = 0.03$ in our model. Here the return on their investment is defined based on how much damage is reduced if they take that action given their assessment of the risk. Hence, we can calculate the expected utility considering long-term return using equation (2).

$$EU(s_i|s_0) = \sum_{j\in[1,7]} p(j) \sum_{k\in[0,4]} p(d_k|s_i, j) LU(c_i, c_{dk}) \tag{2}$$

Where $EU(s_i|s_0)$ is the expected utility for a house in state s_0 to upgrade to s_i, and $LU(c_i, c_{dk})$ is the long-term utility if the house is in damage state k

$$LU(c_i, c_{dk}) = U(w - c_{dk} - c_i\ ) + \sum_{t=1}^{T} U\left((w - c_{dk}) * \frac{1}{(1+\gamma)^t}\right). \tag{3}$$

$U(w - c_{dk} - c_i\ )$ is the utility value when implementing the upgrade in the current year, and $\sum_{t=1}^{T} U\left((w - c_{dk}) * \frac{1}{(1+\gamma)^t}\right)$ is the utility with this upgrade over the next T years. At the end of each year, based on the agent‘s perception of hurricane occuring, each agent identifies the state s_i that gives them the best long-term return and upgrades their house accordingly.

`

Appendix C: Mortality Model

This model was developed by Brooke Anderson at Colorado State University as part of a collaborative project funded by the National Science Foundation Hazard-SEES program. My advisor Dr. Guikema was the PI on this grant. This was not my work, but the details of the model are being reported here to provide needed details and context for the mortality model I used the Agent Based Model.

a) Developing the training dataset for the predictive health risk models

The heat mortality module of the ABM was developed in a series of steps. First, they created a dataset of the estimated mortality effects of previous heatwaves, as well as characteristics of those heatwaves like their length and intensity, to use as a training dataset to fit predictive models linking a heat wave's characteristics to its expected mortality impacts in three age categories.

To develop this training dataset, they used daily time series of health and weather data from the National Morbidity, Mortality, and Air Pollution Study (NMMAPS) data set for 83 high-population communities from 1987–2005 [1]. These data included a separate time series for each of the 83 communities, with aggregated daily mortality counts for three age groups (< 65 years, 65–74 years, and ≥ 75 years), as well as daily weather characteristics. Within these time series, they identified heatwaves as periods in which average daily temperature was at or above the 98th percentile of the year-round daily temperature values for that community for at least 2 consecutive days, following a definition used in previous research [2], [3].

`

They estimated the mortality risk associated with each of these heatwaves using an epidemiological modeling approach that is standard in heat-health research, comparing the observed rate of mortality to that expected without the heatwave using a generalized linear model framework that leveraged the full time series for each community [2], [4]. Age-specific ($<$ 65 years, 65–74 years, $\geq$ 75 years) log relative risks of mortality for each heatwave were estimated using an overdispersed Poisson generalized linear model of daily all-cause mortality regressed on the occurrence of a heat wave, controlling for long-term seasonal trends (using a natural cubic spline) and day of the week for each heat wave [2], [3]. Since some associations were estimated with greater variance (e.g., due to shorter heat wave duration or a smaller populations in the community), Bayesian pooling was used to stabilize these estimates. This was done by taking the posteriors of the estimates, drawing those estimates with greater uncertainty closer to the mean risk of mortality for a heat wave [5].

To complete the training dataset, they measured 20 characteristics of each heatwave that, *a priori*, they identified as potential modifiers of the mortality risk associated with the events. These characteristics included duration of the heat wave, average temperature during the heat wave, and maximum temperature during the heat wave, as well as a number of hybrid characteristics that combined duration and intensity, a number of characteristics that characterized intensity of the heatwave relative to the climate of the community, rather than through an absolute measure of temperature, and characteristics of the community in which the heat wave occurred. They also included the community in which a heatwave occurred as a predictive variable in this models, allowing them to predict estimates specific to Baltimore when applying the models.

b) Training the predictive health risk model

`

Using this training dataset, they trained three 1,000 tree random forest models (one for each age category) to predict the expected age-specific risk of a heatwave based on its duration, intensity, and other measured characteristics. They trained these models using the *randomForest* R package [6], tuning them using 10-fold cross-validation to optimize for the number of model parameters considered at each node split.

c) Applying predictive model within the ABM

These random forest models were incorporated into the ABM to predict the mortality risk for each heatwave for model agents in each age category. The output of the predictive model was further adjusted within the ABM to account for access to air conditioning (A/C) among the model agents during each heatwave, as modeled through other components of the ABM.

The initial predictions from the random forest models represent risk given a community's baseline A/C prevalence. During power outages, the availability of A/C among model agents decreases from this baseline prevalence, while other factors (e.g., opening cooling centers in the community or access to air conditioned refuges through social connections) can increase availability of A/C among the model agents above Baltimore's baseline A/C prevalence. To adjust the predictions from the random forest model based on each model agent's individual access to A/C, the heat-mortality module of the ABM adjusted the initial random forest predictions based on results from O'Neill, Zanobetti, and Schwartz, who found a 1.4% decrease in the association between heatwaves and mortality risk for every 10 percentage point increase in the proportion of the community with home air conditioning[7]. A similar A/C prevalence effect modification size was found for the association between extreme heat and health in another US-based study[8].

`

Based on these results, the following equation was used to adjust community- and age-specific mortality risk predictions from the random forest model based on an agent's access to A/C during the heatwave [7]:

$$\beta = C_0 - 0.14P_{AC}$$

In this equation, β is the log relative risk for mortality during a heat wave for the city population given a certain prevalence of homes with A/C within the community (P_{AC}, taking a value from 0 for a community in which no homes have central A/C to 1 for a community in which all homes have central A/C). The intercept, C_0, represents the log relative risk for a community with $P_{AC} = 0$.

In making this adjustment, we assumed that the overall mortality risk predicted by the model represented a mixture of the individual-level risks of residents with and without access to A/C in the community. Based on American Housing Survey data (United States Census Bureau 2020), the baseline prevalence of A/C in Baltimore homes over the period covered by the training data (1987–2005) was approximately 60%. Under this assumption, the heat-mortality module uses the predicted log relative risk calculated for the heatwave and age group from the random forest model as β and Baltimore's baseline A/C prevalence of 0.6 for P_{AC} to solve for C_0 for each heat wave and age category. It then estimates the agent-level risk for each heatwave separately for agents without access to A/C (setting $P_{AC} = 0$) and for agents with access to A/C (setting $P_{AC} = 1$).

Finally, the heat-mortality module within the ABM applied a standard health impact assessment approach to estimate the number of deaths expected among the model agents during each heatwave, based on each agent's predicted increase in risk of mortality associated with the

`

heat wave and its baseline probability of mortality if a heat wave had not occurred. Baseline probability of mortality for each agent was calculated as a function of the age-specific daily probability of mortality, using age-specific mortality rates for 1999 from the US Centers for Disease Control and Prevention, and the number of days in the heat wave. The probability of baseline mortality throughout each multi-day heat wave was calculated as the complement of the probability of survival on all days within the heat wave. The probability of mortality during the heat wave was then calculated by multiplying the agent's predicted relative risk of mortality associated with the heat wave by the probability of daily mortality had the heat wave not occurred. These probabilities were then applied to each agent within the ABM to estimate the number of excess heatwave-related deaths among ABM agents during each heatwave in the simulation, and those values were then aggregated by year to estimate yearly excess mortality attributable to all heatwave events generated by the ABM for that year.

d) References

[1] J. M. Samet *et al.*, "The national morbidity, mortality, and air pollution study," *Part II Morb. Mortal. from air Pollut. United States Res Rep Heal. Eff Inst*, vol. 94, no. pt 2, pp. 5–79, 2000.

[2] B. G. Anderson and M. L. Bell, "Weather-related mortality: How heat, cold, and heat waves affect mortality in the United States," *Epidemiology*, vol. 20, no. 2, pp. 205–213, Mar. 2009.

[3] K. W. Oleson, G. B. Anderson, B. Jones, S. A. McGinnis, and B. Sanderson, "Avoided climate impacts of urban and rural heat and cold waves over the U.S. using large climate model ensembles for RCP8.5 and RCP4.5," *Clim. Change*, vol. 146, no. 3–4, pp. 377–392, Feb. 2018.

[4] A. M. Vicedo-Cabrera, F. Sera, and A. Gasparrini, "Hands-on Tutorial on a Modeling Framework for Projections of Climate Change Impacts on Health," *Epidemiology*, vol. 30, no. 3, pp. 321–329, May 2019.

[5] P. J. Everson and C. N. Morris, "Inference for multivariate normal hierarchical models," *J. R. Stat. Soc. Ser. B (Statistical Methodol.*, vol. 62, no. 2, pp. 399–412, 2000.

[6] A. Liaw and M. Wiener, "Classification and Regression by RandomForest," 2001.

`

[7] M. S. O'Neill, A. Zanobetti, and J. Schwartz, "Disparities by race in heat-related mortality in four US cities: The role of air conditioning prevalence," *J. Urban Heal.*, vol. 82, no. 2, pp. 191–197, Feb. 2005.

[8] M. Medina-Ramón and J. Schwartz, "Temperature, temperature extremes, and mortality: A study of acclimatisation and effect modification in 50 US cities," *Occup. Environ. Med.*, vol. 64, no. 12, pp. 827–833, Dec. 2007.