Multi-Scale Community Resilience Modeling for

Natural and Manmade Hazards

by

Xilei Zhao

A dissertation submitted to The Johns Hopkins University in conformity with the requirements for the degree of Doctor of Philosophy.

Baltimore, Maryland

May, 2017

© Xilei Zhao 2017

All rights reserved
Abstract

In an increasingly urbanizing world with growing threats of climate change and terrorism, hazards occur more frequently with more severe consequences, bringing significant long-term impacts and requiring years for a community to recover. In order to be better prepared and reduce the impacts of adverse events, communities should conduct effective emergency and mitigation planning. This requires engineers and planners to model pre-event community resilience and consequently come up with strategies to enhance it.

A novel framework will be introduced based on multi-scale community resilience modeling. The framework will emphasize “macroscopic” modeling of communities, “mesoscopic” modeling of interdependent infrastructures providing critical services to communities, and “microscopic” modeling of a single critical infrastructure. At the macroscopic level, I will introduce a dynamic county-based resilience index and a hazard-specific weighting scheme for resilience indicators by using a data-driven approach. At the mesoscopic level, I will show a risk assessment tool for analyzing urban food security of Baltimore City under acute (e.g., earthquake, hurricane, ...
ABSTRACT

flooding, etc.) and chronic (e.g., climate change) stressors; also, I will discuss how to use our model to engage different stakeholders of urban food systems. At the microscopic level, I will describe two new models of evaluating performance of single distributed networks under hazards: transportation and cyberinfrastructure. A novel Markovian framework is developed to analyze the transportation performance before and after disruptions. Furthermore, I will describe a risk assessment tool for the cyberinfrastructure and Medical Records Services in healthcare systems.

This dissertation explores multi-scale modeling of community resilience by using methods and tools in applied mathematics, statistics, and systems engineering. This work presented, is to our knowledge, the most comprehensive and multidisciplinary effort to analyze community resilience as multi-level systems.

Primary Reader: Judith Mitrani-Reiser

Secondary Reader: Takeru Igusa and Benjamin W. Schafer
Acknowledgments

I would like to thank my advisor, Judith Mitrani-Reiser for guiding me through this difficult but fascinating journey. I especially appreciate her patience, wisdom, care, and kindness. I will never forget the late nights she worked with me side by side, the tough times she led me through, and the joyful moments we shared together. She is my great mentor and close friend; I cherish our special bond and look forward to collaborating with her in the future.

I also want to thank my Master’s thesis advisor James C. Spall for his unwavering support of my work and future career. He taught me how to become a rigorous and independent researcher. Thanks to my committee members for their insightful suggestions to my dissertation. I also want to thank my collaborators, who inspired me with great ideas, useful tools, and novel perspectives: Roni Neff, Erin Biehl, Jon Links, Tom Inglesby, Elise Miller-Hooks, and Rachel Davidson, among many others. Thanks to the Department of Civil Engineering and the Department of Applied Mathematics and Statistics for their support of my multidisciplinary research interests.

I would like to thank my group members, Zhang Liu, Caitlin Jacques, Megan
ACKNOWLEDGMENTS

Boston, Fardad Haghpanah, and Mike Mieler for their generous help to my research.

I also want to thank my friends here at Hopkins, who have never left me during ups and downs throughout these years: Sen Lin, Xiaohui Tu, Zhiye Li, Qian Ke, Zhixuan Yang, Luoluo Liu, Xinyi Lin, Zhaohao Fu, Qi Wang, Yang Yang, Ian Miers, Gwen Chodur, Karla Hernandez, Sriram Sankaranarayanan, Jimi Oke, Gary Lin, Mikhail Osanov, and many others. I especially want to thank my dear friends back in China, who have loved me, missed me, and cheered me up when I felt down: Xin Pan, Cheng Chen, and Lili Wang.

I would like to thank my family members. I am extremely grateful to my parents, Xiaolin Shen and Guoquan Zhao, for raising me and loving me unconditionally. They have worked very hard to nurture me and provide me all the resources I needed for a better education. I would like to thank my late grandpas, Tiehan Shen and Rongzu Zhao, for their deep love and great support for my choices since I was a little girl. I also want to thank my other family members, especially my aunts, uncles, cousins, and grandmas, for bearing my bad temper and loving me anyway.

Finally, I would like to thank Xiang (Jacob) Yan for being my boyfriend and “editor-in-chief” for my writings. Since four years ago when we first met each other, he has always been nothing but supportive, inspirational, and encouraging. He completes me: his wit and patience show me the way of being a better researcher, while his love and empathy make me become a better human-being.
Dedication

This dissertation is dedicated to my parents, Xiaolin Shen and Guoquan Zhao.
Contents

Abstract ii

Acknowledgments iv

List of Tables xi

List of Figures xii

1 Introduction 1

1.1 Motivation .............................................. 1

1.2 What is community resilience? .......................... 3

1.3 Multi-scale modeling for community resilience .......... 6

1.4 Objectives of dissertation ............................... 9

1.5 Summary of dissertation contents ........................ 11

2 Macroscopic Model: CoPE-WELL 12

2.1 Introduction .............................................. 12
CONTENTS

2.2 Motivation of hazard-specific weighting scheme ................. 19
2.3 Methodology ................................................. 21
2.4 Weighting scheme for hurricanes ................................. 28
2.5 Model comparison for hurricanes ................................. 33
2.6 Discussion ..................................................... 37
2.7 Conclusion .................................................... 42

3 Mesoscopic Model: Food System 44
3.1 Introduction .................................................... 44
3.2 Methodology .................................................... 49
3.2.1 Identifying food system functionality ......................... 50
3.2.2 Populating subsystems .................................... 51
3.3 Results ......................................................... 53
3.3.1 Accessibility .............................................. 54
3.3.2 Availability ............................................... 57
3.3.3 Acceptability .............................................. 58
3.4 Case studies ................................................... 59
3.4.1 Flood event ............................................... 59
3.4.2 Drought event ............................................ 62
3.5 Stakeholder engagement ....................................... 65
3.6 Discussion ..................................................... 66
3.7 Conclusion .................................................... 68

viii
4 Microscopic Model: Transportation

4.1 Introduction .............................................. 70
4.2 Methodology .............................................. 76
  4.2.1 Travel time estimation ................................ 76
  4.2.2 Baseline transportation model ........................ 84
  4.2.3 Modeling transportation performance after disruptions 88
4.3 Case study ................................................. 91
  4.3.1 Uncertainty bounds for MLEs .......................... 93
  4.3.2 Markovian simulation .................................. 96
    4.3.2.1 Baseline model .................................... 96
    4.3.2.2 Modeling disruptions due to flooding .......... 97
  4.3.3 Sensitivity analysis ................................... 99
4.4 Conclusion ................................................. 101

5 Microscopic Model: Cyberinfrastructure

5.1 Introduction .............................................. 103
5.2 Methodology .............................................. 108
  5.2.1 Fault tree of Medical Records Services ............... 108
  5.2.2 Self-protecting EMRs .................................. 112
  5.2.3 Integration of self-protecting EMR technique to Medical Records Services 113
  5.2.4 Dependencies on IT Department and computers .......... 116
CONTENTS

5.3 Applications .......................................................... 118

5.3.1 Deterministic analysis ............................................. 119

5.3.2 Probabilistic analysis .............................................. 127

5.4 Conclusion .............................................................. 128

6 Conclusion and Future Work ............................................ 131

6.1 Major contributions .................................................. 131

6.1.1 Macroscopic/community level ................................... 132

6.1.2 Mesoscopic/CIbSS level ......................................... 132

6.1.3 Microscopic/critical infrastructure level ....................... 133

6.2 Limitations .............................................................. 134

6.3 Future work ............................................................. 135

A Fault Trees of Food System ............................................. 139

B Estimation Results for Network in Downtown Baltimore ........... 154

C Scenario Analyses of Medical Records Services ...................... 156

Bibliography ............................................................... 164

Vita ................................................................. 179
# List of Tables

2.1 Indicator descriptions and data sources in the engineered systems domain 24  
2.2 Relative weights and direction of indicators (six indicators) . . . . . . 32  
2.3 Cross validation results for six candidate models . . . . . . . . . . . . 35  
2.4 Relative weights and direction of indicators (three indicators). . . . . . 37  

4.1 MLEs and 95% confidence intervals for routes within the network . . 95  
4.2 Sensitivity analysis of transportation performance against flooding . . 100  

B.1 Sample means and MLEs for links’ success probabilities . . . . . . . . 155  

C.1 Scenario analysis of Case 1: earthquake . . . . . . . . . . . . . . . . . 157  
C.2 Scenario analysis of Case 2: cyber attack against hospital computers . 158  
C.3 Scenario analysis of Case 3: cyber attack against central server . . . . 159  
C.4 Scenario analysis of Case 4: cyber attack against local server . . . . . 160  
C.5 Scenario analysis of Case 5: earthquake coupled with cyber attack against hospital computers . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 161  
C.6 Scenario analysis of Case 6: earthquake coupled with cyber attack against central server . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 162  
C.7 Scenario analysis of Case 7: earthquake coupled with cyber attack against local server . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 163
List of Figures

1.1 Graphical representation of community resilience .......................... 4
1.2 Multi-scale community resilience modeling framework ....................... 7

2.1 CoPE-WELL systems dynamics model ........................................... 15
2.2 Illustration of methodology of determining hazard-specific weighting scheme ................................................................. 21
2.3 Hurricane Sandy composite surge/precipitation/wind map: Very High (purple): greater than 10,000 of county population exposed to surge; High (red): 500 – 10,000 of county population exposed to surge, or modeled wind damages > $100 million, or high precipitation (> 8”); Moderate (yellow): 100 – 500 of county population exposed to surge, or modeled wind damages $10 – $100 million, or medium precipitation (4” to 8”); Low (green): no surge impacts, or modeled wind damages < $10 million, or low precipitation (< 4”), adapted from [FEMA MOTF, 2015] .......................................................... 29
2.4 July 2013 unemployment rate in New Jersey, New York, and Connecticut counties that experienced high damage impacted by Hurricane Sandy 31
2.5 Adjusted $R^2$, $C_p$, BIC are shown for the best models from the six candidate models: red dots indicate the best models .......................... 36
2.6 Values of engineered systems domain using unit and data-driven weighting methods (hurricane case) ..................................................... 41

3.1 Simple illustration of fault tree structure ........................................ 47
3.2 Definitions for food system failure, inaccessibility, unavailability and unacceptability ................................................................. 50
3.3 Model-building framework ............................................................ 52
3.4 Main food system fault tree, with supply chain subtree ....................... 56
3.5 Propagating failures of Subtree 3 .................................................. 61
3.6 Propagating failures of Subtree 1 .................................................. 63
3.7 Propagating failures of Subtree 5 .................................................. 64
LIST OF FIGURES

4.1 Markovian framework for modeling traffic dynamics before and after disruptions due to adverse events such as hazards .......................... 72
4.2 Conceptual illustration of the transportation network system from origin to destination through a specific route (highlighted in bold orange) with four subsystems ($p = 4$), each having binary output, and one full system, having output (travel time) $T_k$ log-normally distributed .... 78
4.3 A general transportation network: solid lines denote network of interest (A-B-C-D-E-F-G-H-I-J); dashed lines denote boundary layer .... 80
4.4 Illustration of transition matrix: submatrix $P$ for links within the network and remaining elements in $P'$ for incoming and outgoing traffic streams ................................................................. 86
4.5 A transportation network with road DE disrupted: grey dashed lines indicate closed links; dark red lines denote links whose corresponding rows in the transition matrix are modified .................. 89
4.6 A transportation network in downtown Baltimore (approximately 1 mile east of the center of the Inner Harbor area): solid lines denote network of interest (A-B-C-D-E-F-G-H-I-J-K-L-M-N-O-P); dashed lines denote boundary layer .................. 93
4.7 Flood risk map of Inner Harbor area of Baltimore City, adapted from [Open Baltimore, 2014] ................................................................. 97
4.8 Histogram of travel time (origin: link 70; destination: link 54) .................. 98
5.1 Fault tree of Medical Records Services (red: staff; blue: space; green: communications infrastructure) ................................................ 110
5.2 Fault tree modification by including self-protecting EMR technique .. 115
5.3 Illustration of hospital services relying on IT Department and computers 117
5.4 Propagating failures of the original fault tree (Case 1) .................. 123
5.5 Propagating failures of the cyberinfrastructure subtree (Case 1) .... 124
5.6 Propagating failures of the revised main tree (Case 1) .................. 125
A.1 Subtree 1: high food price .................................................. 140
A.2 Subtree 2: sufficient decrease in net income .................................. 141
A.3 Subtree 3: food purveyors are not accessible .................................. 142
A.4 Subtree 3-1: public transit is not available .................................. 143
A.5 Subtree 4: unable to leave home .................................................. 144
A.6 Subtree 5: production failure .................................................. 145
A.7 Subtree 6: production failure .................................................. 146
A.8 Subtree 7: wholesale is disrupted .................................................. 147
A.9 Subtree 8: distribution is disrupted .................................................. 148
A.10 Subtree 8-1: distribution centers are disrupted .................................. 149
A.11 Subtree 9: retail is disrupted .................................................. 150
A.12 Subtree 10: food bank donation failure .................................................. 151
LIST OF FIGURES

A.13 Subtree 11: other food assistance organization donation failure . . . . 152
A.14 Subtree 12: supply chain failure . . . . . . . . . . . . . . . . . . . . . . 153
Chapter 1

Introduction

1.1 Motivation

In an increasingly urbanizing world with growing threats of climate change and terrorism, hazards occur more frequently with more severe consequences, bringing significant long-term impacts and requiring years for a community to recover. In 2016 alone, multiple major disasters struck the United States. In January, Winter Storm Jonas, with historic amounts of snow, hit the east coast and left 48 people dead. In August, Louisiana floods caused 13 fatalities across five parishes and left thousands of people homeless; this was called the worst U.S. disaster since Hurricane Sandy by the Red Cross. In October of 2016, Hurricane Matthew, a historic destructive Category 5 Atlantic hurricane, swept the Caribbean and the Southeast U.S., killing up to 1,600 people in Haiti and 47 people in the U.S. Additionally, California continued to
CHAPTER 1. INTRODUCTION

have a year-round severe drought after five years of experiencing the drought, and experienced a series of wildfires: in total, 6,986 fires have occurred and burnt 564,835 acres in California [CAL FIRE, 2016] in 2016 alone. These events have impacted large geographic areas and many communities, resulting in many injuries and deaths, population dislocation, disruption of businesses, damage of critical infrastructure, job losses, and greater demands for external resources [The National Academies, 2012].

Hazards can happen anytime, anywhere, with every person under the threat of being homeless, out of job, injured, or even dead. Our communities confront these challenges by having emergency and mitigation plans to reduce the impacts of these adverse events. One effective way to conduct long-term planning is to incorporate the concept of community resilience. In fact, the concept of resilience has been applied in many disciplines, including engineering, ecological, physical, and social sciences. Recently, resilience has been applied to communities, societal systems, and critical infrastructure systems, particularly in the context of hazards and climate change. This concept is essential to build smart cities, which can benefit the overall well-being of all the individuals who live in the communities. Therefore, community resilience has become a top priority on the nation’s agenda; multiple recent Federal policies and plans were published to enhance the resilience of our communities, such as Presidential Policy Directive PPD-8 (national preparedness) [U.S. Department of Homeland Security, 2011], Presidential Policy Directive PPD-21 (critical infrastructure security and resilience) [The White House, 2013], the Department of Homeland Security National
CHAPTER 1. INTRODUCTION


In this dissertation, we use probabilistic and statistical methods and systems tools to model, measure, and analyze resilience of our communities, societal systems, and critical infrastructure systems against natural and manmade disasters, which speaks to the national efforts of building resilient, sustainable, and smart cities.

1.2 What is community resilience?

Community resilience is defined as the ability of a community to prepare for, respond to, and recover from adverse events. There are three critical elements of this concept, i.e., preparedness, response, and recovery. According to FEMA’s terminology [FEMA, 2003],

- “Preparedness includes plans and preparations made to save lives and property and to facilitate response operation.”

- “Response includes actions taken to provide emergency assistance, save lives, minimize property damage, and speed recovery immediately following a disaster.”

- “Recovery includes actions taken to return to a normal or improved operating condition following a disaster.”
CHAPTER 1. INTRODUCTION

Therefore, community resilience is a holistic concept which integrates these three vital complementary elements; ineffective planning in any one of these three areas can result in poor resilience of our communities.

Considering the dynamic nature of resilience, it is difficult yet critical to analyze resilience quantitatively. In 2003, Bruneau et al. proposed to quantitatively assess and enhance the seismic resilience of communities [Bruneau et al., 2003]. They integrated quantitative measures of robustness, rapidity, resourcefulness, and redundancy into the four dimensions of community resilience, including technical, organizational, social, and economic. In this dissertation, we adapt the basic concept from [Bruneau et al., 2003, Dorbritz, 2011] and extend it to a broader quantitative definition of community resilience.

![Graphical representation of community resilience](image.png)

Figure 1.1: Graphical representation of community resilience
CHAPTER 1. INTRODUCTION

As shown in Figure 1.1, relative community resilience, $R$, is measured by the area under the curve for relative community functioning $F$ over time $t$ during the “resistance” and “recovery” phases, which can be presented mathematically as follows:

$$R = \int_{t_0}^{t_2} F(t)dt = \int_{t_0}^{t_1} F(t)dt + \int_{t_1}^{t_2} F(t)dt,$$

(1.1)

where $t_0$ represents the time when the event occurs, $t_1$ indicates the time when the relative community functioning reaches the lowest point immediately after the event, and $t_2$ is time when the community fully recovers from the disaster (here, we assume without losing generality that an impacted community is able to fully recover from the adverse event, though we acknowledge the fact that some communities may never fully recover in individual cases). This process includes three important phases: 1) pre-event functioning (when $t \leq t_0$); 2) resistance (when $t_0 < t \leq t_1$); and 3) recovery (when $t_1 < t \leq t_2$).

As shown in Figure 1.1, $R$ is the green area under the curve; a larger $R$ (or smaller loss in functioning) represents a more resilient community against hazards. Here, we use relative community functioning, because a community with higher level of absolute pre-event functioning can expect higher resilience. The first part of integral in eqn.(1.1) measures the resistance of a community: higher $\int_{t_0}^{t_1} F(t)dt$ indicates higher resistance and thus higher resilience of a community during the short-time period after the event. From $t_1$ to $t_2$ is the recovery phase after the disaster, so larger $\int_{t_1}^{t_2} F(t)dt$ shows faster and more efficient recovery. Note that, with proper planning and investments, a community can have enhanced community functioning (i.e., over...
CHAPTER 1. INTRODUCTION

1.0) after the recovery, which is actually more desirable and more realistic especially for communities with low pre-event functioning [Bruneau et al., 2003]. Creating, calibrating, and validating the resilience curve still remains a significant challenge for researchers, urban planners, and policy makers, considering the complexities and uncertainties of our communities, societal systems, and critical infrastructure systems and the hazards they face.

1.3 Multi-scale modeling for community resilience

In order to better understand and quantify community resilience, we propose a novel framework, called *multi-scale* community resilience modeling, to measure and assess the risk and vulnerabilities of our communities, societal systems, and critical infrastructure systems.

Figure 1.2 shows our multi-scale community resilience modeling framework, where we divide the community resilience modeling process into three different levels:

- *Macroscopic* level: communities
- *Mesoscopic* level: critical infrastructure-based societal systems (CIbSSs)
- *Microscopic* level: critical infrastructure systems
At the microscopic level, critical infrastructure systems are often defined as systems whose disruption or destruction would have debilitating impacts on security, national economic security, national public health or safety, or any combination
CHAPTER 1. INTRODUCTION

thereof [U.S. Department of Homeland Security, 2016]. As shown on the lower layer in Figure 1.2, there are six major critical infrastructure systems considered at the microscopic level, including water, wastewater, power, natural gas, cyber, and transportation infrastructures. All these critical infrastructure systems are interdependent, and together support the daily functioning of a CIbSS.

At the mesoscopic level, the CIbSSs (shown in the bubbles in the middle layer in Figure 1.2) refer to systems that consist of interdependent buildings, that together, serve a vital community function and that are dependent on the networks of critical infrastructures shown on the lower layer of Figure 1.2. A CIbSS (such as a food system, a healthcare system, a school district, etc.) serves as one of the critical community functions [Norris et al., 2008]. Here, we conceptualize community functioning as the ability and capacity of a community to provide a range of essential services to its inhabitants; we, therefore, consider the following ten major societal systems to measure pre-event community functioning: 1) communication, 2) economy, 3) education, 4) food and water, 5) government, 6) housing, 7) healthcare and public health, 8) nurturing and care, 9) transportation, and 10) well-being. All these critical infrastructure-based societal systems are of great importance in measuring community resilience, especially in measuring the preparedness of a community against disasters.

At the macroscopic level, the community is illustrated as the whole middle layer in Figure 1.2. The community consists of multiple interdependent CIbSSs (also known as community functioning domains), which are dependent on critical infrastructure
systems. Therefore, a community is a multi-layer system with high interdependencies and uncertainties.

With a clear understanding of the interdependencies and dependencies of these three levels, we implement probabilistic and statistical methods and systems tools to conduct multi-scale community resilience modeling.

1.4 Objectives of dissertation

The concept of “resilience” is recognized in the field of disaster planning and management, and lots of efforts [Bruneau et al., 2003, Norris et al., 2008, Cutter et al., 2008, Chandra et al., 2011, Foster, 2011, CARRI, 2011] have been devoted to modeling and quantifying community resilience as it applies to disasters. However, current community resilience models still have multiple limitations, for example,

- Many community-level models are still in the stage of “conceptual framework,” lacking convincing parameter estimation, model calibration, and validation. In particular, for the famous resilience index work, such as [Cutter et al., 2003, Burton, 2012, Foster, 2011], people usually equally aggregate all the indicators together, without using real data to determine how to weigh, how to calibrate, and how to validate the indices.

- Many current models focus on resilience and sustainability of a single building or bridge, but we need to understand how the performance of many buildings and
CHAPTER 1. INTRODUCTION

bridges, together, under hazards can impact the functionality of larger societal systems, such as a healthcare system and a food system, since this is important for communities to understand regional impacts and resilience.

- Most resilience models of transportation infrastructure and/or supply chains do not consider traffic dynamics after disruptions, making these models less practical and realistic for effective emergency and mitigation planning.

Therefore, this dissertation is aimed at resolving or at least partially resolving these three major limitations in the research theme of community resilience modeling.

One of the major objectives is to develop a hazard-specific weighting scheme (see Chapter 2) in order to weigh different community resilience indicators, and to estimate the importance and validate the directionality of these indicators when we aggregate them together.

In addition, we aim at developing novel risk assessment and decision-support tools using systems approaches to model the resilience of important community systems: urban food systems (see Chapter 3) and cyberinfrastructure systems (see Chapter 5), by taking into account the downtime performance of critical infrastructure, building, equipment, supply chain, staff, etc. together.

Lastly, this dissertation is aimed at incorporating recently developed statistical methods and new data sources to model the transportation system (critical in the supply chain of all CIbSSs) performance before and after disruptions due to adverse events like hazards (see Chapter 4), in order to facilitate the efforts of modeling
CHAPTER 1. INTRODUCTION

resilient, sustainable, and smart transportation systems.

1.5 Summary of dissertation contents

This dissertation presents novel multi-scale community resilience modeling approaches by using probabilistic and statistical methods and systems tools. In this dissertation, the first part of Chapter 2 introduces a novel county-level community resilience index which captures the dynamic nature of community resilience over time using systems dynamics modeling. Then, we present a hazard-specific weighting scheme for community resilience indicators using a statistical learning approach. In Chapter 3 of this dissertation, we conduct risk assessment of urban food systems by using fault tree analysis (FTA), and conduct scenario analysis to inform planning of more resilient and sustainable food systems under the threat of climate change-induced phenomena. In Chapter 4, we assess transportation system performance before and after disruptions by modeling and simulating dynamic traffic networks using Markov chain and Google Maps. Chapter 5 analyzes the risk of cyberinfrastructure and medical records in hospitals by using FTA; we also integrate a self-protecting electronic medical record (EMR) technique into the fault tree model to show enhanced cyber capabilities of hospitals. This dissertation concludes with a summary of major contributions, limitations, and future work, especially in integration of models at multiple scales, in Chapter 6.
Chapter 2

Macroscopic Model: CoPE-WELL

Some contents of the Chapter are based on an accepted journal paper [Links et al., 2017], and a published conference paper [Zhao and Mitrani-Reiser, 2017].

2.1 Introduction

Community resilience refers to the ability of a community to resist and recover from adversity, such as natural disasters, terror attacks, and pandemics. Quantifying community resilience can help communities better assess their strengths and weaknesses, prepare for different types of hazards, estimate losses in case of adverse situations, take effective measures to reduce losses, and speed post-event recovery. However, such a task is extremely challenging, because community resilience is essentially a comprehensive and complex concept with entrenched difficulties in defining
appropriate criteria for its quantification. The commonly-used approach usually considers multiple domains of a community, such as ecological, social, economic, institutional, infrastructure, and community competence [Cutter et al., 2008], and selects some indicators to capture features of each domain. These indicators are equally weighted across the domain and aggregated together to come up with an overarching index to quantify community resilience.

Existing disaster resilience efforts are an important foundation of our work. In 2003, Bruneau et al. proposed a framework to quantitatively analyze the seismic resilience of communities [Bruneau et al., 2003]. Norris et al. proposed a conceptual model of community resilience for disaster readiness, which emerges from four primary sets of adaptive capacities, namely, economic development, social capital, information and communication, and community competence [Norris et al., 2008]. Then, Cutter et al. put forward the disaster resilience of place (DROP) model to quantify community resilience exposed to natural hazards; the methodology of DROP model included 3 major steps, namely, variable selection, weighting, and aggregation [Cutter et al., 2008]. However, when it comes to weighting, they found no theoretical or practical justification for allocating different weights across indicators, so they used an equally weighted index [Cutter et al., 2010]. In 2011, RAND [Chandra et al., 2011] identified eight “levers” to define community resilience, including wellness, access, education, engagement, self-sufficiency, partnership, quality, and efficiency. Around the same time, a resilience capacity index (RCI) was developed based on U.S. metropolitan
CHAPTER 2. MACROSCOPIC MODEL: COPE-WELL

regions [Foster, 2011], which is a composite statistic summarizing a region’s score on 12 equally weighted indicators, where four indicators are in each of three domains, i.e., regional economic capacity, demographic capacity, and community connectivity capacity. Also in 2011, the Community and Regional Resilience Institute (CARRI) published the Final Report of the Community Resilience System Initiative (CRSI) (see [CARRI, 2011]). This product has contributed to helping communities to understand their vulnerabilities, and guiding them in mitigation planning, in order to enhance their overall resilience against hazards.

In 2017, a new community resilience index, called Composite of Post-Event Well-being (CoPE-WELL) \footnote{CoPE-WELL, as a “community resilience index” for the entire U.S. in the context of disaster preparedness for all hazards, has been developed as requested by the U.S. Centers for Disease Control and Prevention (CDC). The research group includes Jonathan M. Links, Brian S. Schwartz, Sen Lin, Norma Kanarek, Judith Mitrani-Reiser, Tara Kirk Sell, Crystal R. Boddie, Doug Ward, Cathy Slemp, Robert Burhans, Kimberly Gill, Tak Igusa, Benigno Aguirre, Joseph Trainor, Joanne Nigg, Thomas Ingelsby, Eric Carbone, James M. Kendra, and me.}, under development for many years was published in [Links et al., 2017]; this index was intended to help practitioners and policymakers frame high-level policy discussions about community resilience towards hazards. This project is sponsored by CDC, and the research team includes Johns Hopkins University, Johns Hopkins School of Public Health Center for Health Security, and University of Delaware. Unlike previous index work, which was mostly static, CoPE-WELL was developed to model the dynamic process inherent in community functioning and resilience after an emergent event. Unlike the past conceptualization and modeling efforts, our work is the first dynamic community resilience model based on real places.
and data in the entire U.S.

![CoPE-WELL systems dynamics model](image)

Figure 2.1: CoPE-WELL systems dynamics model

CoPE-WELL is a county-based national model which considers many indicators from various domains to represent the pre-event functioning, short-term post-event functioning, and long-term post-event functioning of the community. The systems dynamics model of CoPE-WELL is represented as a stock-and-flow diagram in Figure 2.1. The central blue tank is the “stock,” representing community functioning (note:
the amount of fluid in the tank represents functioning at a single time point). The initial level of the liquid in the central tank is set by the orange box on the left, which represents pre-event functioning. When an event (i.e., the red box on the bottom) occurs, community functioning is drawn from the central tank, although the impacts of the event are modified by the important factors of resistance (i.e., population factors and prevention/mitigation factors) shown in the green box on the left. Then, the modified event impacts are applied to Valve 1, resulting in the liquid from central tank depleting to the sink (i.e., the grey box on the right). After the event, the three tanks on the top, representing recovery, replenishes the central tank until it returns to pre-event functioning (in CoPE-WELL model, we assume an impacted community can return to the pre-event functioning level). The resilience curve on the right is plotted according to the level of liquid in the central tank over time.

Specifically, the pre-event functioning (in the orange box) is quantified by multiple community functioning domains [Norris et al., 2008], including communication, economy, education, food & water, government, housing, health care & public health, nurturing & care, transportation, and well-being. The level of pre-event functioning ($CF_0$) sets the initial height of the liquid in the central tank. When an event occurs, the impacts of the event will diminish the functioning in a community (e.g., disruption of transportation, damage of hospitals, shortage of food and water, etc.), so Valve 1 will be open to deplete liquid from the central tank to the sink immediately after the event. However, the resistance (in the green box) will modify the impacts of the event,
which contains two major components: population factors and prevention/mitigation factors. Population factors contain vulnerability, inequality, and deprivation domains; prevention/mitigation factors are comprised of natural systems, engineered systems, and countermeasures domains. Therefore, flow through Valve 1 is controlled by the modified impacts of the event. After liquid depletion finishes, the community begins to recover with the aid of three stocks: social cohesion, preparedness & response, and external resources (in the blue box). During the recovery phase, Valves 2, 3, and 4 open, so the three tanks on the top replenish the central tank over time until the community functioning reaches constant status.

In the CoPE-WELL model, we need to identify candidate measures for each domain at a regional scale (such as county, metropolitan, and state) in order to best measure resilience for local communities. Community is defined as “a group of people who live in the same area (such as a city, town, or neighborhood)” [Merriam-Webster, nd]. Ideally, we would like to choose measures and collect data directly from communities, but community is such a vague concept without a consistent definition throughout the country or even distinct geographical boundaries. Therefore, we choose county, the smallest administrative division with most data available, as the regional scale. As of 2016, there are 3,141 counties in the U.S., so on average, the population per county is over 100,000. Hence, a county usually consists of multiple communities and the statistics of a county can be thought as the average of communities that locate within county boundaries. In this study, all the candidate measures are identified at
CHAPTER 2. MACROSCOPIC MODEL: COPE-WELL

the county level (but also because model is developed in this way).

Then, our team of experts select measures that meet the following criteria: 1) having face validity and predictive construct validity for post-event community functioning; 2) publicly available for almost all U.S. counties; 3) having variation across counties; 4) easy to implement; and 5) measuring processes instead of outcomes. Our team is very multi-disciplinary and includes 19 experts in behavioral health, civil engineering, criminal justice, community health, computational modeling, education, disaster and emergency management, emergency medicine, environmental epidemiology, environmental health sciences, geography, health behavior, health communication, law, mental health, program evaluation, public health practice, public policy, public safety, risk management, social epidemiology, sociology, systems modeling, urban affairs, and urban health. In addition to utilizing published scientific literature, we also held multiple meetings and panels in the past six years to discuss, investigate, and evaluate the choice of measures in the engineered systems domain and the direction of these indicators.

The CoPE-WELL model is able to capture the dynamic process of community functioning and resilience over time, but has its own limitations: 1) the systems dynamics model has not been fully validated; 2) the disasters are only considered as purely physical events, rather than understanding that disasters are the combination of physical and psychosocial characteristics [Jacob et al., 2008]; and 3) there is no weighting scheme for the indices, which will likely vary by hazard type. The rest of
this Chapter is devoted to addressing the third limitation. Specifically, in Section 2.2, we discuss the motivation of a hazard-specific weighting scheme; in Section 2.3, we describe the methodology, and also propose an algorithm including 8 steps as a guidance to determine the relative weights and directionality of the indicators; in Section 2.4, we choose hurricanes as a case study to show how to develop a weighting scheme for a given hazard type. In Section 2.5, we use model comparison techniques (both analytical and numerical methods) to identify a best model in terms of predictive accuracy. In Section 2.6, we discuss the results obtained by the hurricane model, and draw inference from the hazard-specific weighting scheme suggested by the hurricane model. We conclude the Chapter with a summary and suggestions of future work in Section 2.7.

2.2 Motivation of hazard-specific weighting scheme

Nearly all the community resilience indices (i.e., DROP, RCI, CoPE-WELL, etc.) use a unit weighting scheme when aggregating indicators from different domains of community functioning. Admittedly, there are some good reasons behind this approach. For example, unit weights can avoid sampling error and are robust to outliers [Bobko et al., 2007], which makes it a valid approach to aggregating data.

However, when it comes to natural hazards, we may notice that some community
resilience indicators are probably more important than others, so if we put equal weights to all the indicators, the corresponding index may not correctly represent the resilience of the community. Also, we need to point out that unit weights may bring bias to the prediction model, especially when some indicators are much more important than the others. In addition, we may also find that the weighting scheme should vary by disaster type. Therefore, to address this weighting issue, we need to develop a hazard-specific weighting scheme.

Critical infrastructure plays an essential role before, during and after the disaster. While critical infrastructure breakdowns are rare, recent events, such as Hurricane Katrina, have demonstrated the catastrophic consequences of such breakdowns [Boin and McConnell, 2007]. Therefore, in this Chapter, in order to better evaluate, measure, and enhance resilience of critical infrastructure systems, we choose a set of commonly used indicators in the engineered systems domain (also considered in CoPE-WELL [Links et al., 2017]), and develop a hazard-specific weighting scheme for these indicators by using a data-driven approach. To be specific, our study uses statistical learning methods and real data from publicly available sources, to quantify the importance of indicators of the engineered systems domain, to validate the expert judgment of indicators’ directionality, and to improve the credibility and accuracy of community resilience indices.
2.3 Methodology

In this section, we present the methodology for developing a hazard-specific weighting scheme using linear regression model. One major reason is that most resilience indices provide users a composite with linear aggregation of all indicators from different domains, so linear regression is a natural and good choice to estimate the weights of indicators, which can be directly used to aggregate indicators to compute the resilience indices. Another reason is that linear regression has simple model structure, making it easy to interpret and draw inferences.

Figure 2.2: Illustration of methodology of determining hazard-specific weighting scheme

Figure 2.2 illustrates the methodology of determining the weighting scheme, which includes eight steps with Step 4 as an optional step for users, including (1) identifying indicators and acquiring data; (2) identifying the response variable and acquir-
CHAPTER 2. MACROSCOPIC MODEL: COPE-WELL

ing data; (3) deleting observations with missing data; (4) deleting some variables when lacking observations (optional step); (5) normalizing indicators; (6) removing collinearity; (7) determining final model; and (8) obtaining final results (weights and direction).

In the following analysis, we will follow the regime shown in Figure 2.2 to determine the relative weights of indicators in the engineered systems. So, first, we need to identify the indicators required to represent engineered systems in the community resilience model and acquire the data for these indicators. Specifically, the engineered systems, also referred to the infrastructure system, contains multiple subsystems including buildings, communications/cyber, transportation, water, wastewater, power, natural gas, etc. Note that we only choose indicators with publicly available data, so all the stakeholders, including emergency managers, urban planners, policy makers, and researchers can easily gain access to the data and reproduce the work. Since we could not find publicly available data (for all U.S. counties) for appropriate indicators to represent wastewater, power, and natural gas, we finalize seven indicators in the aspects of buildings, communications/cyber, transportation and water to serve as proxies to measure the resilience of the engineered systems.

The indicators’ categories, descriptions, directions, and sources in the engineered systems domain are listed in Table 2.1. We consider seven indicators from multiple subdomains, including buildings, communications/cyber, transportation, and water. Note that in the weighting scheme modeling, we follow the general indicator selec-
CHAPTER 2. MACROSCOPIC MODEL: COPE-WELL

tion criteria of CoPE-WELL with some exceptions (i.e., we develop algorithms for assessing API to acquire county-level data for No.3 and No.4, and use GIS to pre-process county-level data for No.5, which may make data acquisition hard for some users without corresponding expertise), aiming at including more measures for the engineered systems domain than CoPE-WELL model does to capture the domain performance more comprehensively. Further, the directions of indicators show their effect on the resilience of engineered systems and are pre-determined by our team of experts.

Then, we need to identify a reasonable response variable that can quantify resilience (as a proxy) after the event and collect corresponding data. In this study, we choose an economic response variable to serve as a proxy to community resilience, which would apply to all natural and manmade hazards. One reason is that major disasters always have considerable economic losses and the behavior of economic sector can be a good indicator of measuring the performance of a community’s resilience. As pointed out in [Rose and Krausmann, 2013], at least in the short term (i.e., the first year after a major disaster), business behavior is most important to economic recovery and resilience. Note that it is also possible to use hazard-specific response variable, as long as it can quantify resilience for the specific type of hazard.
Table 2.1: Indicator descriptions and data sources in the engineered systems domain

<table>
<thead>
<tr>
<th>#</th>
<th>Category</th>
<th>Description</th>
<th>Direction</th>
<th>Data Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Buildings</td>
<td>Average age of housing stock</td>
<td>NEG</td>
<td>American Housing Survey by the U.S. Census Bureau</td>
</tr>
<tr>
<td>2</td>
<td>Buildings</td>
<td>Percentage of housing units that are not mobile homes</td>
<td>POS</td>
<td>American Housing Survey by the U.S. Census Bureau</td>
</tr>
<tr>
<td>3</td>
<td>Communications/cyber</td>
<td>Median <strong>residential</strong> download speed</td>
<td>POS</td>
<td>National Broadband Map created and maintained by the National Telecommunications and Information Administration and in collaboration with the Federal Communications Commission</td>
</tr>
<tr>
<td>4</td>
<td>Communications/cyber</td>
<td>Median <strong>mobile</strong> download speed</td>
<td>POS</td>
<td>Same as #3</td>
</tr>
<tr>
<td>5</td>
<td>Transportation</td>
<td>Road miles per square mile</td>
<td>NEG</td>
<td>Topologically Integrated Geographic Encoding and Referencing by the U.S. Census Bureau</td>
</tr>
<tr>
<td>6</td>
<td>Transportation</td>
<td>Number of bridges per 100 square miles that are structurally deficient or functionally obsolete</td>
<td>NEG</td>
<td>National Bridge Inventory by the U.S. Department of Transportation Federal Highway Administration</td>
</tr>
<tr>
<td>7</td>
<td>Water</td>
<td>Percentage of population affected by water violation of those served by public water systems</td>
<td>NEG</td>
<td>Safe Drinking Water Information System by the U.S. Environmental Protection Agency</td>
</tr>
</tbody>
</table>

After obtaining all the data of indicators and response variable, we remove observations with missing data. The reason why we do not use missing data imputation here is because it is difficult to determine whether the data are (completely) missing
at random. Therefore, removing observations with missing data is the simplest and most straightforward way to deal with missing data problem. Afterwards, it is an optional step – remove “not-so-important” variables mainly based on the expert judgment, if we have too few observations compared to the number of indicators. This step is not ideal, but we have to admit the fact that the damage/impact data are so limited for certain types of hazards, especially for earthquakes, in the U.S. However, when the number of observations is less than \( n + 2 \), where \( n \) represents the number of variables in the engineered systems domain, we might need to remove some “irrelevant,” or least important, variables before conducting linear regression. Next, we need to normalize all the variables. Normalization can adjust variables measured on different scales to a notionally common scale, so after normalization, the importance of variables obtained by the regression model can be compared directly. After this step, we need to remove highly correlated variables, in that collinearity among indicators will increase the sampling variation of regression weights [Bobko et al., 2007], making the model less powerful and convincing. Therefore, the first six steps provide us a data set with parsimonious indicators and an appropriate response variable. A similar multivariate analysis of determining a parsimonious indicator set was also conducted in [Burton, 2012].

Lastly, we fit a linear regression model to the remaining variables and obtain the direction and weights of each indicator accordingly. We also compare the direction of indicators from the regression model with the direction pre-determined by the
experts (see Table 2.1), which can help us validate the direction choices. Sometimes, it is difficult for experts to agree to a direction on some indicators, so this framework can help us to determine the direction of some “not-so-sure” indicators. For example, when determining the direction of No.5, road miles per square mile, our team of experts could not reach a consensus about its direction. Some people argued that more roads indicated more accessibility, and thus more resilience. Nevertheless, others held the opinion that transportation infrastructure is very vulnerable towards hazards, so more roads in a community may bring more disruptions to the community after the hazard. In this case, this framework can help us determine the direction of this indicator by using real data and statistical learning techniques. Here, the tentative experts’ choice for this indicator’s direction is negative.

The algorithm of the hazard-specific weighting scheme is as follows:

**Step 1:** Identify the indicators in the engineered systems domain and acquire the data for each of the indicators;

**Step 2:** Identify the response variable which can quantify resilience after the event and acquire the corresponding data for the response variable;

**Step 3:** Remove observations with missing data;

**Step 4 (Optional):** Remove certain “not-so-important” variables using expert judgment, if the number of observations is less than \( n + 2 \), where \( n \) represents the number of variables in the engineered systems domain;

**Step 5:** Normalize all the variables and fit a linear regression model with all the
Step 6: Compute the variance inflation factor (VIF) to remove collinear variables and repeat this step until all the variables’ VIF are less than 10;

Step 7: Fit a linear regression model with the remaining variables;

Step 8: Obtain the relative weights and direction of each indicator.

Note that for Step 4, we need to guarantee the number of observations is no less than \( n + 2 \), because we want to make sure the \( t \)-test for an estimator has at least one degree-of-freedom. Besides, VIF is a commonly-used method to determine the multi-collinearity in a linear regression model. VIF provides a statistic that measures how much the variance of the parameter estimates is increased due to collinearity. A common choice of VIF threshold is 10; a VIF of 10 represents that with other things being equal, the variance of the \( i \)th regression coefficient is 10 times greater than it would have been if the \( i \)th variable had been linearly independent of other variables [O’Brien, 2007]. In this Chapter, we maintain every indicator in the regression model with VIF less than 10.

Overall, we develop a novel concept of hazard-specific weighting scheme. One of the strengths of our methodology is transparency. By following the 8 steps described above, any stakeholders, such as urban planners, policy makers, and researchers, can determine a hazard-specific weighting scheme for different community resilience indicators from different domains. Although we only focus on indicators in the engineered systems domain, the algorithm can be easily adapted to other domains of the com-
CHAPTER 2. MACROSCOPIC MODEL: COPE-WELL

...munity. Moreover, although we only include the hurricane case here, researchers can easily apply this method to other hazards, such as drought, flood, terrorist attack, etc., by following the algorithm proposed in this Chapter.

2.4 Weighting scheme for hurricanes

In this section, we introduce a proposed weighting scheme for hurricanes. To conduct the analysis for hurricanes, we need real data from historical storms, which caused significant disruptions to the impacted communities. As an example for choosing a weighting scheme, we chose Hurricane Sandy, which occurred on October 29, 2012, because: 1) Sandy is a storm with significant damage to the local communities, including many counties on the east coast (CoPE-WELL model has been applied to New York City to help local communities assess and enhance resilience after Hurricane Sandy); 2) there are a lot of damage/impact data of Sandy available in the open sources, which significantly facilitates our research.

First, Step 1 is completed as mentioned in Section 2.3. Seven indicators are considered to quantify the resilience of the engineered systems exposed to a hurricane. The other important factor is the “year” of the data. Since Hurricane Sandy occurred in October, 2012, ideally, we would like to use 2011 data to represent the pre-event functioning of these critical infrastructure systems. However, for No.3, No.4, and No.7, there is no data available for the year of 2011. Therefore, we have decided to use the 2014 data as a proxy for pre-event functioning level, since Sandy took place...
in late 2012, and major recovery efforts, especially the recovery of critical infrastructure systems, happened in the year of 2013 [U.S. Department of the Interior, 2013]. Therefore, in this case, we assume in the year of 2014, these critical infrastructure systems mainly recovered from the damages and can function at the pre-event level.

Figure 2.3: Hurricane Sandy composite surge/precipitation/wind map: Very High (purple): greater than 10,000 of county population exposed to surge; High (red): 500 – 10,000 of county population exposed to surge, or modeled wind damages > $100 million, or high precipitation (> 8”); Moderate (yellow): 100 – 500 of county population exposed to surge, or modeled wind damages $10 – $100 million, or medium precipitation (4” to 8”); Low (green): no surge impacts, or modeled wind damages < $10 million, or low precipitation (< 4”), adapted from [FEMA MOTF, 2015]
Before choosing the response variable, we need to obtain the damage map or impact map of the disaster, since we need to identify counties with the same level of impact for analysis; in other words, the counties we choose need to have the similar damage or impact under the hazard. Figure 2.3 is a Hurricane Sandy composite surge/precipitation/wind map based on county impact assessment by the Federal Emergency Management Agency (FEMA) Modeling Task Force (MOTF) [FEMA MOTF, 2015], where purple represents very high exposure (with greater than 10,000 of county population exposed to surge). In this case, we choose the counties with purple color as the target counties. Since these counties have the most severe impact/damage during Hurricane Sandy, we hypothesize that they will have the greatest drop in functioning after the event, and thus provide the most interesting examples for recovery and resilience.

In addition, as discussed in the previous section, in this study, we choose an economic response variable to proxy community resilience, since major disasters always accompany considerable economic losses and business behavior plays an important role in post-event economic recovery and resilience [Rose and Krausmann, 2013]. For Hurricane Sandy, the communities we are focusing on have an economy dependent on tourism (especially in summer). Therefore, we choose July 2013 unemployment rate (UR), a typical economic metric, as the response variable to measure the impact of disruption of economic activity caused by the storm especially in the Travel and Tourism industry [U.S. Department of Commerce, 2013]. The UR for July 2013 is
CHAPTER 2. MACROSCOPIC MODEL: COPE-WELL

shown Figure 2.4 for counties impacted by Sandy. The data were collected by the Bureau of Labor Statistics (BLS) and the unemployment data is preliminary and not seasonally adjusted. Note that we also looked into the UR for the entire states of New Jersey and New York, and we found that the counties we chose for analysis (see Figure 2.4) tend to have higher UR in July 2013 compared to other counties in these two states, which may be interpreted as the business disruptions caused by Hurricane Sandy.

Figure 2.4: July 2013 unemployment rate in New Jersey, New York, and Connecticut counties that experienced high damage impacted by Hurricane Sandy

Third, we delete observations with missing data. We skip the fourth step, because we have adequate observations to compute the linear regression model. Fifth, all the indicators and response variable are normalized. Sixth, we fit a linear regression model with all variables included and compute the VIF, and we find that the fifth indicator
(i.e., road miles per square mile) has the highest VIF (i.e., 10.8). So we remove the fifth indicator, and then fit another linear regression model with remaining variables and compute VIF. After the second round of linear regression, all the indicators have a VIF less than 10, providing a convincing statistical evidence that the collinearity among indicators has been removed.

Table 2.2: Relative weights and direction of indicators (six indicators)

<table>
<thead>
<tr>
<th>#</th>
<th>Description</th>
<th>Relative Weights</th>
<th>Effect on UR</th>
<th>Effect on Resilience</th>
<th>Expert Judgment</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Average age of housing stock</td>
<td>1.0</td>
<td>POS</td>
<td>NEG</td>
<td>NEG</td>
</tr>
<tr>
<td>2</td>
<td>% of not mobile homes</td>
<td>54.0</td>
<td>NEG</td>
<td>POS</td>
<td>POS</td>
</tr>
<tr>
<td>3</td>
<td>Median residential download speed</td>
<td>5.0</td>
<td>NEG</td>
<td>POS</td>
<td>POS</td>
</tr>
<tr>
<td>4</td>
<td>Median mobile download speed</td>
<td>14.5</td>
<td>NEG</td>
<td>POS</td>
<td>POS</td>
</tr>
<tr>
<td>6</td>
<td># of deficient or obsolete bridges</td>
<td>35.1</td>
<td>POS</td>
<td>NEG</td>
<td>NEG</td>
</tr>
<tr>
<td>7</td>
<td>% of population affected by water violation</td>
<td>1.8</td>
<td>POS</td>
<td>NEG</td>
<td>NEG</td>
</tr>
</tbody>
</table>

Seventh, the model is thus finalized with 6 indicators. Eighth, the relative weights (relative to the indicator with the smallest weight) and direction of indicators are obtained, shown in Table 2.2. It is apparent that the indicators’ effect on UR is opposite to their effect on resilience, because a community with lower UR tends to be more resilient to adverse situations. Therefore, we find that the direction of
indicators obtained by our model is same as the direction identified by experts in terms of effect on resilience. We may draw the conclusion that the results obtained by the regression model are consistent with the expert judgment, which validates the independent choices made about directionality of indicators.

With the consistency of indicators’ direction obtained by the regression model and expert judgment, we may stop here and output the final model to represent the pre-event functioning of the engineered systems. However, as shown in Table 2.2, we notice that No.2, No.4, and No.6 have much higher relative weights compared to No.1, No.3, and No.7, which might indicate that No.1, No.3, and No.7 are not important or even redundant in terms of predictive accuracy. Therefore, in the next section, we will use model comparison techniques (including analytical and numerical methods) to compare a set of candidate models with different number of indicators.

2.5 Model comparison for hurricanes

In this section, we first use best subset selection to draw a set of candidate models with different number of indicators. Then, we compare these candidate models using both analytical and numerical methods. Last, we identify the best model in terms of predictive accuracy.

First, we use best subset selection (see [James et al., 2013, pp. 205 – 207]), to select the “best” subsets of indicators. To be specific, to conduction best subset selection, we fit separate regression models for all the possible combinations of the 6 indicators.
CHAPTER 2. MACROSCOPIC MODEL: COPE-WELL

obtained from the final hurricane model. In other words, we fit overall \( \binom{6}{1} = 6 \) models that contain one indicator; we then fit \( \binom{6}{2} = 15 \) models that contain two indicators; ...; we finally fit \( \binom{6}{6} = 1 \) model that contain all the indicators. We then identify the best model within each subset using \( R^2 \) as the criterion; as a consequence, we obtain 6 candidate models (shown in the second column of Table 2.3) with \( k \) number of indicators, where \( k = 1, 2, ..., 6 \). Best subset selection can serve as a screening tool, which gives us a reasonable number of “better” models that we can start with.

Next, we use analytical and numerical methods to choose a single best model from all the candidate models as shown in Table 2.3. For the analytical methods, we plan to use adjusted \( R^2 \), Mallows’ \( C_p \) [Mallows, 1973], and Bayesian information criterion (BIC) [Schwarz, 1978]. Note that in case of linear regression, \( C_p \) is equivalent to Akaike information criterion (AIC) [Akaike, 1974]. The generic forms for \( C_p \), AIC, and BIC are [Hastie et al., 2008]:

\[
C_p = \frac{1}{N} (RSS + 2d\hat{\sigma}^2), \tag{2.1}
\]

\[
AIC = \frac{1}{N} (-2 \cdot \text{loglik} + 2d), \tag{2.2}
\]

\[
BIC = -2 \cdot \text{loglik} + (\log N) \cdot d, \tag{2.3}
\]

where RSS is the residual sum of squares on a training set of data, \( d \) is the number of indicators/variables, \( \hat{\sigma}^2 \) is an estimate of the variance associated with each response in the linear regression model, \( N \) is the sample size of training set, and loglik is the maximized value of the log-likelihood function for the model. Therefore, according
CHAPTER 2. MACROSCOPIC MODEL: COPE-WELL

to eqn.(2.1) – (2.3), it is clear that the lower the values of $C_p$, AIC, and BIC are, the better the model is. For the numerical methods, we decide to use hold-one-out cross validation to compare the prediction error/mean squared error (MSE). The best model is the one with the smallest prediction error/MSE. These statistical model comparison methods can help us to decide which model works best in terms of predictive accuracy for resilience of the engineered systems in case of hurricanes.

In Figure 2.5, we show the best models drawn from the six candidate models based on adjusted $R^2$, $C_p$, and BIC, where red dots indicate the best models. We find that the model with three indicators (i.e., No.2, No.4, and No.6) is the best model according to adjusted $R^2$ and BIC, while the model with two indicators (i.e., No.2 and No.6) is the best one based on $C_p$.

The cross validation results are shown in Table 2.3. We find that the model with three indicators (i.e., No.2, No.4, and No.6) is the best model, which has the smallest prediction error/MSE among these six candidate models.

<table>
<thead>
<tr>
<th># of indicators</th>
<th>Model description</th>
<th>MSE</th>
<th>Choice</th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
<td>#1 + #2 + #3 + #4 + #6 + #7</td>
<td>1.51</td>
<td>No</td>
</tr>
<tr>
<td>5</td>
<td>#2 + #3 + #4 + #6 + #7</td>
<td>0.97</td>
<td>No</td>
</tr>
<tr>
<td>4</td>
<td>#2 + #3 + #4 + #6</td>
<td>0.64</td>
<td>No</td>
</tr>
<tr>
<td>3</td>
<td>#2 + #4 + #6</td>
<td>0.50</td>
<td>Yes</td>
</tr>
<tr>
<td>2</td>
<td>#2 + #6</td>
<td>0.53</td>
<td>No</td>
</tr>
<tr>
<td>1</td>
<td>#2</td>
<td>1.01</td>
<td>No</td>
</tr>
</tbody>
</table>

35
Taking into consideration all the results obtained by analytical and numerical methods, we choose the model with three indicators (i.e., No.2, No.4, and No.6) as the best model in terms of predictive accuracy. In Table 2.4, we show the relative weights and direction of indicators for the best model, where the direction obtained
by the linear regression model is consistent with the direction pre-determined by our team of experts.

Table 2.4: Relative weights and direction of indicators (three indicators)

<table>
<thead>
<tr>
<th>#</th>
<th>Description</th>
<th>Relative Weights</th>
<th>Effect on UR</th>
<th>Effect on Resilience</th>
<th>Expert Judgment</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>% of not mobile homes</td>
<td>3.5</td>
<td>NEG</td>
<td>POS</td>
<td>POS</td>
</tr>
<tr>
<td>4</td>
<td>Median mobile download speed</td>
<td>1.0</td>
<td>NEG</td>
<td>POS</td>
<td>POS</td>
</tr>
<tr>
<td>6</td>
<td># of deficient or obsolete bridges</td>
<td>2.2</td>
<td>POS</td>
<td>NEG</td>
<td>NEG</td>
</tr>
</tbody>
</table>

2.6 Discussion

In this section, we summarize the results obtained by the hurricane model before and after model comparison. Also, we identify the possible reasons behind the results and draw inferences from them. Last, we compare the values of engineered systems domain computed by unit weighting scheme and data-driven weighting scheme.

In Section 2.4, we obtain the final model with six indicators after finishing the eight steps of the algorithm, and we may stop there and output the final model. In the general case, if the relative weights obtained are on the similar scale, the results are acceptable and applicable. However, as shown Table 2.2, No.2 has relative weight equal to 54.0, while No.1 only has relative weight equal to 1.0, in which the
contribution of No.1 to the overall domain is negligible. In this case, we conduct an optional model comparison based on more complex statistical learning techniques. As expected, after conducting the model comparison, the model with three parameters, including No.2, No.4, and No.6, is actually better than the model with six parameters in terms of predictive accuracy.

Among these three indicators, we find that No.2, the percentage of housing units that are not mobile homes, is the most important indicator, with relative weight equal to 3.5. This suggests that for communities in the hurricane zone, mobile homes are most vulnerable to hurricanes, since high winds and extreme flooding may easily damage mobile homes (note that Hurricane Sandy was not a strong wind event). This result is very coherent with expert judgment, because the prevalence of structures that are not mobile homes directly link to the number and type of structures in damage [Cutter et al., 2003].

Then, No.6, the number of bridges per 100 square miles that are structurally deficient or functionally obsolete, is the second important indicator in the hurricane model, with relative weight equal to 2.2. This result is actually rather reasonable, since for hurricanes, bridges are very important – bridges are critical nodes in the transportation system, whose failure can significantly undermine the accessibility of local communities. Therefore, the number of deficient or obsolete bridges can help communities quantify the resilience of their transportation infrastructure in case of hurricanes.
Also, interestingly, median mobile download speed can serve as a good proxy for communications system functioning. One possible explanation is that median mobile download speed may show the operations and development of mobile carriers in the local communities; higher speed indicates more advanced development, which may connect to the resilience of communications system.

In Figure 2.6, we show the values of engineered systems domain (Z-score) computed by two weighting methods (i.e., unit weighting and data-driven weighting) in the hurricane case. The results are interesting: 1) we find that many counties share similar values under two weighting schemes, so when disaster damage data is not available, the index computed from unit weighting scheme could still be meaningful for many counties; 2) some counties do experience changes when we use data-driven weighting scheme, for example, Cape May County tends to have lower resilience of engineered systems while Middlesex County and Fairfield County show higher values of the engineered systems domain. Therefore, the hazard-specific weighting scheme is a scientific and sophisticated approach to helping people evaluate and quantify community resilience by using the “index” approach.

Compared to the unit weighting method, our method prevails in the way that more important variables are correctly accounted for and more accurate quantification of the engineered systems’ resilience is obtained. Notably, the response variable choice needs to be convincing and justifiable. In this study, even though the choice for response variable, namely, the July 2013 UR, is not perfect, and Hurricane Sandy
is only one typical case of hurricanes, the results are encouraging – the indicators’ direction obtained by the regression models coincides with the direction obtained by the expert judgment, which may serve as a justification for our response variable choice.

Further, it is worthwhile to mention that there are other indicators (not available for all counties across the U.S.) that would be better indicators for predicting hurricane damage/impact, such as building type, roof type, envelop type, etc. Also, since publicly available data significantly limits our choices for measures in the engineered systems domain, we would like to encourage federal and local government to collect more indicators at different levels, which directly relate to infrastructure functioning, such as percentage of population affected by power outage, percentage of buildings that are structurally deficient, number of communications failure incidents, etc.
CHAPTER 2. MACROSCOPIC MODEL: COPE-WELL

(a) unit weighting method

(b) data-driven weighting method

Figure 2.6: Values of engineered systems domain using unit and data-driven weighting methods (hurricane case)
CHAPTER 2. MACROSCOPIC MODEL: COPE-WELL

2.7 Conclusion

In the final analysis, we first introduce a dynamic county-based community resilience index, i.e., CoPE-WELL, to quantify community resilience over time for the entire U.S. Then, we put forward a novel method to develop a hazard-specific weighting scheme for indicators used in any community resilience modeling framework. Specifically, we choose a set of commonly-used indicators in the engineered systems domain, and use hurricanes as a case study to show the advantages of this method and validate directionality of indicators in the model, where we also use model comparison techniques to identify the best model (including three indicators, i.e., percentage of not mobile homes, median mobile download speed, and number of deficient or obsolete bridges) in terms of predictive accuracy.

The results obtained by the hurricane model support our hypothesis that indicators have different importance for different hazards. In particular, the percentage of housing units that are not mobile homes is the most important indicator when we estimate the resilience of the engineered systems domain in case of hurricanes (admittedly, this model is developed only based on data from Hurricane Sandy, but Hurricane Sandy is a typical representative of hurricanes as discussed in Section 2.4).

However, this method is not perfect. We have to admit that the sample size is relatively small. Also, the choice for response variable may not be optimal. In addition, indicators for the engineered systems are not very comprehensive, since we only include indicators with publicly available data. If the federal and local government
can collect more data at different levels directly related to critical infrastructure sys-
tems, we will be able to build up a more accurate model to quantify the resilience of
the engineered systems.

In the future, we plan to extend current work to other domains and consider more
measures/indicators. We also want to consider conditioning weighting scheme on haz-
ard intensity parameters (e.g., magnitude for earthquakes, wind speed for hurricanes,
water depth for floods, etc.). Moreover, we plan to collect more data from more
historical events to standardize the weighting scheme. Also, based on the weighting
scheme work, we plan to develop a decision-support tool for local communities to
better predict their resilience against natural and manmade hazards.
Chapter 3

Mesoscopic Model: Food System

This Chapter is based on a journal paper that is under development with co-authors, Gwen Chodur, Roni Neff, Erin Biehl, and Judith Mitrani-Reiser.

3.1 Introduction

According to the United Nations Food and Agriculture Organization’s (FAO) definition, a food system consists of all the vital actors, processes, and infrastructure involved in growing, harvesting, transporting, packing, processing, transforming, marketing, selling, acquiring, consuming, and disposing of food. The food system thus is a very complex system with multiple interdependent subsystems from global to local levels. A subsystem failure may impact other subsystems and/or even trigger the whole food system failure.
As we discussed in Chapters 1 and 2, food system is an essential domain of pre-event community functioning (see Figures 1.2 and 2.1), which plays a crucial role in modeling and quantifying community resilience. The functioning of the food system is very vulnerable to potential disruption before, during, and after hazards. However, disruption to the food provisioning processes can lead to decline in food security in impacted communities. Therefore, in this Chapter, we conduct the risk assessment of urban food systems in order to understand the failure mechanism, address the vulnerabilities, and enhance the resilience of food systems.

Recently, various research efforts have been devoted to identifying and characterizing the resilience of food systems under natural and manmade hazards. For example, James and Friel proposed an integrated approach to address resilient urban food systems in order to promote population health in the context of climate change [James and Friel, 2015]. The Initiative for a Competitive Inner City used Boston as a case study to analyze resilient food systems in order to provide useful and practical recommendations [The Initiative for a Competitive Inner City, 2015]. Also, some tools have been developed recently to model the food system resilience. For instance, National Research Council proposed a framework for assessing the health, environmental, and social effects of the food system in 2015 [National Research Council, 2015]. In 2016, Toth, Rendall, and Reitsma put forward a qualitative tool for measuring food resilience [Toth et al., 2016]. However, all these efforts are mainly focused on short-term acute hazards, without considering long-term climate change-induced weather
CHAPTER 3. MESOSCOPIC MODEL: FOOD SYSTEM

phenomena, which actually pose serious threats to our urban food security.

In this Chapter, we present a novel approach to assessing food system vulnerabilities: FTA. FTA is a widely-used tool in risk analysis of complex systems. FTA is applied to illustrate paths by which events can affect food system functioning and identifies the range of factors that could lead to system failure, enabling both clearer understanding and future modeling efforts to address key vulnerabilities within the entire system [Risebro et al., 2007; Jacques et al., 2014]. Watson at Bell Telephone Laboratories first introduced FTA in the early 1960s as a means to conduct safety evaluations of complex systems [Watson, 1961]. Haasl further developed this method by introducing the fault tree structuring process, which marked the beginning of a wider interest of applying FTA in engineering [Lee et al., 1985]. Since then, FTA has been applied to many fields, and to public health issues such as water contamination and hospital system resilience post-earthquake [Risebro et al., 2007; Jacques et al., 2014]. Moreover, there are multiple applications of FTA in the field of food security. For instance, Hope [Hope, 2004] applied FTA to analyze bioterrorist risks to the food supply in the U.S. Domenech, Escriche, and Martorell applied FTA to analyze the potential failures of food supply chain under normal and abnormal conditions [Doménech et al., 2010].

There are various systems level methods that could be used to estimate the failures of complex systems. In this Chapter, FTA is chosen to model food system failure, because:
they can analyze the overall performance of a complex system that consists of various multi-level subsystems;

- the underlying Boolean ("yes" and "no" states) logic used to combine these lower levels is transparent;

- the graphical representation of FTAs can also be used as an interactive tool with stakeholders to better understand how mitigation efforts that prevent specific failures in the food system can be used to prevent larger cascading failures.

Figure 3.1: Simple illustration of fault tree structure
CHAPTER 3. MESOSCOPIC MODEL: FOOD SYSTEM

As shown in Figure 3.1, a fault tree is structured with an overall system “failure” on the top (i.e., top event), and beneath it, all of the intermediate and basic factors that could cause failures. A failure is defined as the improper functioning of the overall system. A basic event refers to failure in a basic component of the system, which may be easier to predict in practice. An intermediate event is a failure caused by a combination of lower level failures; in this case, the intermediate event is determined by basic events 1 and 2. FTA uses logic gates, which implement Boolean functions (output: “0” or “1”) to combine event failures across levels. The “or” gate signifies that the output is true if any of the inputs are true. For example, the intermediate event can fail due to a disruption at any one of its components, such as basic event 1 OR basic event 2. The “and” gate, in contrast, indicates that the output is true (happens) if all inputs from lower level subsystems are true. If any one of the lower level subsystems can still compensate for the loss of another, a failure does not occur. For example, the top event is only true if both intermediate event AND transfer gate are true, where the “transfer” gate indicates that this part of the fault tree transfers from/to another part of the fault tree. In this Chapter, “transfer” gates connects the main tree with subtrees.

Therefore, we propose a novel model using FTA to analyze the functionality (and subsequently, inform assessments of resilience) of a comprehensive food system under natural and manmade hazards. The fault tree model, developed by a cross-disciplinary team of systems engineers (i.e., Judith Mitrani-Reiser and Xilei Zhao)
and public health professionals (Roni Neff, Erin Biehl, and Gwen Chodur), to identify the ways that an urban food system could fail in the U.S. The main fault tree is comprised of 12 subtrees, with implications for food accessibility, availability, and acceptability for different populations in the cities. To be specific, in Section 3.2, we identify the key components of food system functionality, and populate the subsystems based on a novel modeling-building framework. In Section 3.3, we illustrate how to build the fault tree model for urban food systems. Then, in Section 3.4, the fault tree model is applied to case studies of potential flooding (short term event) and drought (long term event). In Section 3.5, we introduce how to use fault trees to engage stakeholders in local communities. We discuss the strengths and limitations of this model in Section 3.6, and conclude this Chapter in Section 3.7.

3.2 Methodology

In this section, we introduce the overall methodology for modeling urban food security after disasters by using a systems tool. In Subsection 3.2.1, we identify food system functionality by splitting this concept into three major components, including accessibility, availability, and acceptability. Then, in Subsection 3.2.2, we introduce a top-down approach to populate the subsystems of the urban food system in order to develop the fault tree model.
3.2.1 Identifying food system functionality

A critical first step in developing a model to assess food system functionality is to identify the terminology used by key stakeholders to define failures in the food system. This terminology would subsequently drive the structure of the fault trees used to assess cascading failures in this complex system. Therefore, we adopted the FAO’s definition of food security to describe well-functioning food systems: “all people, at all times, have physical, social, and economic access to sufficient, safe and nutritious food that meets their dietary needs and food preferences for an active and healthy life.” [FAO, 2001]

<table>
<thead>
<tr>
<th>Food System Failure:</th>
<th>Inaccessibility:</th>
<th>Unavailability:</th>
<th>Unacceptability:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Food is present but barriers exist to prevent its acquisition by the community.</td>
<td>Food is not present at provisioning points.</td>
<td>Food is not safe, nutritious to meet dietary needs, or culturally/religiously acceptable.</td>
<td></td>
</tr>
</tbody>
</table>

Figure 3.2: Definitions for food system failure, inaccessibility, unavailability and unacceptability

Figure 3.3 illustrates the functions of the key components of a food system based on the principal components of the FAO’s definition of food security: availability, accessibility, and acceptability [FAO terminology: utilization]. These three components are equally important, and the occurrence of any one of these three (i.e., inaccessibilit-
ity, unavailability, and unacceptability) can trigger the top-level event, food system failure.

Notably, this model is a community-based model; that is to say, we mainly focus on the general failures/impacts across an entire community, city, or region, without considering the failures of individuals or households. For example, as shown in Figure 3.4, when food is not nutritionally adequate only for one person or one household, this basic event will not be triggered; in contrast, when the food becomes not nutritionally adequate for a community, this basic event will be triggered.

### 3.2.2 Populating subsystems

After determining the possible causes of the food system failure, we use a top-down approach to populate the lower-level events/subsystems of the full system (i.e., urban food system). As illustrated in Figure 3.3, we show the model-building framework of developing and validating a fault tree model of urban food system.

In the first step, we start by reviewing previous studies and literature on food security and food system resilience. The model is also informed by findings from 36 qualitative interviews with stakeholders throughout the Baltimore City food system conducted by our public health colleagues. In addition, we harness our team members’ expertise in food systems and collect useful empirical data of urban food security. The following step is to apply the knowledge from Step 1 to build a conceptual fault tree, and to define quantifiable indicators for the basic and intermediate failure events of
the tree and choose proper failure thresholds for these indicators. The third step is to collect empirical data for the tree’s indicators to assess the validity of the tree structure. The final step is to check whether the case study supports the conceptual model; Steps 2 – 4 are repeated until the case study supports the overall structure of the conceptual model. Finally, a validated model can be shared with stakeholders to inform planning, policies, and programs of food security.

Figure 3.3: Model-building framework

In this Chapter, the fault tree model is developed by using Smartdraw online software [SmartDraw Software LLC [US], 2017]. We consulted expert opinions from
five food systems experts who reviewed the fault trees and helped us tackle challenging questions with respect to how best to structure it in order to reveal the food system failure in a comprehensive way. Engineers on the team built and reviewed the model from an engineering perspective to assure the correctness of logic flow. The fault tree model can be readily applied to analyze food system resilience by assessing food system functioning and identifying potential vulnerabilities within the system. Note that the fault tree model developed in this Chapter has not been fully validated yet; that is, we only finish the first two steps of the model-building framework. In the future, we plan to collect historical data to validate our model and then output it to the city in order to plan interventions.

3.3 Results

In this section, we show how to build a fault tree model for urban food system for different populations in the cities by splitting the tree into three major branches, i.e., food being not accessible, food being not available, and food being not acceptable. Under these three branches, 12 subtrees are considered to show lower-level systems/subsystems of the urban food system.

This fault tree model is a system model, which means that it can only be considered as a system failure if a high number of people are affected. Note that in 2015, there are 12.7% of American households experiencing food insecurity [Coleman-Jensen
et al., 2016], so the national food system has already failed for these people. In our fault trees, especially in Subtrees 1 and 2 (see Figure A.1 and A.2), we have already shown some root causes (i.e., high food prices and significant decrease in net income) of failures for this portion of the overall population. However, our fault trees are not specifically designed to focus on current food system failures; instead, we want to use FTA to investigate potential events which could lead to population-wide impacts.

The food system failure mechanism is very complex, which can be elaborated by a main tree and 12 subtrees. In Figure 3.4, we show the main food system fault tree (with three major branches, i.e., inaccessibility, unavailability, and unacceptability) and the supply chain subtree. The detailed 12 subtrees are shown in Figure A.1 – A.14 in Appendix A. Note that the fault trees are not fully comprehensive, but aim at capturing the principal factors of concern.

### 3.3.1 Accessibility

According to Figure 3.2, *inaccessibility* is defined as food being present but with existing barriers that prevent its acquisition by the community. In this Chapter, we consider two major barriers that may cause food to become inaccessible to the community: 1) economic barriers: they make food available for purchase but unaffordable to the community; and 2) physical barriers: they disrupt ability to travel to the food provisioning points in the community.

Food being not economically accessible may result from high food prices (see
CHAPTER 3. MESOSCOPIC MODEL: FOOD SYSTEM

Figure A.1) or significant decrease in net income (see Figure A.2). On the one hand, high food prices may have multiple reasons, including decreased supply, increased transport costs, and increased production costs. On the other hand, decrease in population making a living wage, higher unemployment rates and safety nets failure can cause significant decrease in net income.

Food being not physically accessible may be related to the food purveyor being not accessible (see Figures A.3 – A.4) and community members being unable to leave home (see Figure A.5). The food purveyor being not accessible may have two major causes: 1) food purveyors are not accessible by foot; and 2) food purveyors not in walking distance are not accessible. In other words, food provisioning points may be inaccessible because of events containing transportation barriers, lack of proximity to any provisioning points, or disruptions of ordinary means of transit (including private transit and public transit). Moreover, community members may be unable to leave home due to various reasons, i.e., restriction of movement, acute health issues, and safety concerns.
Figure 3.4: Main food system fault tree, with supply chain subtree
3.3.2 Availability

We consider two major causes of food becoming unavailable at provisioning points to the community, including supply chain failure and donation failure. Note that the gate that connects these two subsystems is an “OR” gate for people who are mainly dependent on food donation, because for those people, the mere failure of donation systems can make them lose their major food supplies. On the other hand, the gate is an “AND” gate for the rest of the population, since only both failures of supply chain and emergency backup system will leave them hungry.

The first set of “unavailability” is in food supply chain failures. The food supply chain is a complex system with multiple essential nodes, and the failure of any of the essential nodes can paralyze the entire food supply chain. Here, we show five critical nodes within the food supply chain system (see Figure A.14), including production (see Figure A.6), processing (see Figure A.7), wholesale (see Figure A.8), distribution (see Figures A.9–A.10), and retail (see Figure A.11). The types of events that could fail these subtrees share common features, such as staff, essential resources, critical infrastructure, business management, weather, etc. All these factors result from the intrinsic characteristics of food itself, including its requirement for temperature control, storage, and packaging to ensure its quality for sales.

The second set of “unavailability” is in food donation failures by private and/or government donors. Donation failure mainly results from food bank donation failure (see Figure A.12), donation failure from other food assistance organization (e.g., food
CHAPTER 3. MESOSCOPIC MODEL: FOOD SYSTEM

pantries, soup kitchens, shelters, and emergency government assistance programs) (see Figure A.13), and the failure of supply chains that support the normal function of the food donation (see Figure A.14). Note that we assume the donation food and normal food rely on the same supply chains, since they are dependent on many common features, especially staff, essential resources, critical infrastructure, etc. Besides, the failure of food assistance benefits due to unavailability of the cyberinfrastructure required for benefit card use, shows in Subtree 2 (see Figure A.2) instead of here, in that this, essentially, is an income failure.

3.3.3 Acceptability

Even though food is accessible and available to the community, food unacceptability still poses a threat to food security. In this Chapter, we split the concept of food unacceptability into three major causes, including food being medically contra-indicated, food being not nutritionally adequate, and food being not religiously/culturally appropriate. Any occurrence of each of the three causes can fail part of or even all of the community members, resulting in a significant decline in food security for these vulnerable populations.
CHAPTER 3. MESOSCOPIC MODEL: FOOD SYSTEM

3.4 Case studies

Food system resilience is the continuation of food security in the face of disruptive events [Candy et al., 2015]. Therefore, although this model highlights potential vulnerabilities in a food system at different failure points or how well a current system functions, assessing food system resilience requires evaluating other factors necessary for resilient systems, such as the capacity of a system to adapt to changes. The use of FTA to assess food functionality and resilience is demonstrated through two applications of the model described below.

3.4.1 Flood event

On July 31, 2016, a serious flash flood attacked Ellicott City with six inches of rainfall within two hours, and the Ratapsco River swelled, rising six feet in just one hour. The flood devastated downtown Ellicott City, Maryland, killing two people and destroying or damaging at least 25 buildings. This event is just one of the devastating floods which occurred in the U.S. this summer. According to the literature [Kay et al., 2009, Hirabayashi et al., 2013], climate change is among the most significant factors that contribute to the frequency and severity of flooding. Furthermore, Baltimore City has high flooding risks, and more importantly, recent studies [Adger et al., 2003, Hallegatte, 2009] find that rising sea level from climate change may double the risk of flooding in coastal communities (many located in Baltimore City) in Maryland.
CHAPTER 3. MESOSCOPIC MODEL: FOOD SYSTEM

Therefore, in this Chapter, we consider a flood scenario in Baltimore City as a typical short-term hazard to impact the urban food security.

Suppose a super storm (such as Hurricane Sandy) occurs in the Inner Harbor area of Baltimore City, cutting off critical traffic links for several local communities, and food purveyors are becoming not accessible by foot for those communities. According to the propagating failures of Subtree 3 (shown Figure 3.5), roads are obstructed due to the flood, and transportation links to food purveyors are closed, resulting in cars, bikes and bus services becoming unavailable. Additionally, the metro and light rail (i.e., public transit) in this area also become unavailable after the flood event. Unavailability of cars, bikes and public transit gives rise to the inaccessibility of food purveyors that are not in walking distance for these local communities. Therefore, the failure of Subtree 3 triggers the upper level intermediate events, namely, “food is not physically accessible” and “food is not accessible” (see Figure 3.4). Therefore, by following the logic of the main tree, the urban food system of Baltimore City would fail during the flood event due to food being inaccessible for these local communities.
CHAPTER 3. MESOSCOPIC MODEL: FOOD SYSTEM

Figure 3.5: Propagating failures of Subtree 3
3.4.2 Drought event

Even though short-term hazards are usually acute and can significantly impact the local communities, long-term hazards can also threaten urban food security. For example, California has experienced a historic drought since 2011, characterized by frequent wildfires and dying forests, and supported by climate change [Swain et al., 2014]. California’s farm lands produce 96% of the broccoli, 91% of the tomatoes, 72% of the lettuce, 91% of the strawberries, and 88% of the grapes to the entire U.S. [U.S. Department of Agriculture Economic Research Service, 2016]. Therefore, California’s drought could significantly affect the city’s supply of vegetables, fruits, and nuts. This case study assesses the potential impacts of severe California drought causing a high percentage of crop failure on Baltimore’s food system. This single season crop failure could also have broader ramifications if farm businesses fail.

In the initial season of drought, shortages of essential crops such as broccoli, tomatoes, and lettuces may happen, giving rise to higher food prices. As shown in Figures 3.4 and 3.6, high food prices can lead to the entire food system failure. However, the populations with lower incomes are particularly vulnerable in this situation, because these essential food becomes less affordable for them. When the production failure in California is prolonged and becoming more severe, this can trigger long-term impacts for urban food security, as shown in Figure 3.7. In this case, the food system in Baltimore City may fail, if some major types of food largely produced in California become unavailable and alternative production sources do not exist for Baltimore.
Figure 3.6: Propagating failures of Subtree 1
CHAPTER 3. MESOSCOPIC MODEL: FOOD SYSTEM

Figure 3.7: Propagating failures of Subtree 5
CHAPTER 3. MESOSCOPIC MODEL: FOOD SYSTEM

Note that when local markets fail, we may expect global markets to respond to this opportunity, which can be seen as another source of resilience. We may treat the response of global markets as a backup to domestic markets, but in this study, we assume no backup production available as a worst case scenario.

3.5 Stakeholder engagement

In this Chapter, we use FTA to model urban food system failure mechanism, which has many useful applications, such as scenario analysis (discussed in Section 3.4) and stakeholder engagement, in planning resilient and sustainable urban food systems and enhancing urban food security. These applications underscore assessing potential risks and vulnerabilities in a food system at different levels.

FTA can be easily transferred into a decision-support tool in order to engage different stakeholders (such as urban planners, policy makers, food program managers, engineers, etc.) in local communities. FTA is a transparent and straightforward risk assessment tool with clear tree structure and simple logic. Fault trees can be very general to show the big picture of the problem of interest (see the main tree in Figure 3.4); they can also include many details which are commonly represented as subtrees with multiple layers of intermediate events and basic events (see the subtrees in Appendix A). Therefore, FTA can be adapted into an ideal decision-support tool, which allows different stakeholders to sit on the same table to understand
CHAPTER 3. MESOSCOPIC MODEL: FOOD SYSTEM

the big picture of the food system failure structure, and then work together to dig into different parts of the trees to identify key components in their services and find possible ways these complex systems can fail. These efforts can help urban planners and policy makers to evaluate the resilience of urban food systems, to develop more effective strategies and policies, to speed post-event recovery, and to enhance the ability of a food system to adapt to changes, especially in face of climate change.

We have already applied our current fault tree model (even though it has not been fully validated yet) to Baltimore City, aiming at engaging different stakeholders, especially the planners and policy makers, to investigate potential vulnerabilities within the food system and develop plans to enhance its resilience.

3.6 Discussion

Using FTA to assess food system vulnerabilities and resilience has multiple advantages. By taking into account a system’s complexity, FTA focuses on a variety of subsystems and relationships among them. An urban food system is such a complex system with multiple interdependent subsystems, so FTA is an ideal method to conduct risk assessment of food systems. The fault trees of urban food systems can capture the long-term and short-term effects of hazards in a single framework.

Furthermore, by using Boolean logic to come up with a tree structure, FTA is transparent and enables effective discussion among model developers, experts, and
stakeholders. FTA also provides a tangible systematic framework to combine basic events, intermediate events, and top events, which allows urban planners and policy makers to see the big picture of the system and to understand all potential failure points of the overall system. Moreover, FTA can be easily transferred into an interactive decision-support tool to characterize systems, assess mitigation plans, and enhance food system resilience for the local communities.

Admittedly, the fault tree model of urban food system has its own limitations. Fault trees developed in this Chapter are not exhaustive of all the potential basic events, and the basic events may not be on the same scale. Some basic events are not unique and may fit in the model at different places; for example, in Subtree 1 (see Figure A.1), peak oil price leads to both increased transport costs and increased production costs. In addition, the detailed fault trees tend to grow very large, which limits the ability of communicating this framework to stakeholders. Furthermore, the impacts of some events may be iterative, which is difficult to model using FTA. For instance, in Subtree 2 (see Figure A.2), economic crisis and higher unemployment rate are iterative instead of cascading.

To address the first three limitations, we could choose proper indicators and corresponding thresholds to characterize and measure intermediate events. When the model is populated with real data, FTA can be applied to conduct quantitative analysis to populate empirical food system failure distribution using Monte Carlo simulation.
CHAPTER 3. MESOSCOPIC MODEL: FOOD SYSTEM

3.7 Conclusion

The normal functioning of the food system plays a critical role before, during, and after any natural or manmade hazards. Disruption of the food system may lead to decreases or even failures in food security in impacted areas. Therefore, both researchers and practitioners start to develop novel approaches to enhance food system resilience. In this Chapter, we use FTA to model and analyze the failure mechanism of a complex urban food system, which could help policy makers, urban planners, and stakeholders to understand and mitigate the threats, and to identify planning intervention points.

The fault tree of urban food system is created to characterize the potential failure mechanism of urban U.S. food system, which includes three important branches (i.e., inaccessibility, unavailability, and unacceptability). Also, our model is designed for different populations of the urban food system. This model has multiple applications in the field of public health, including identifying vulnerabilities in existing urban food systems, guiding mitigation planning and decision making, and enhancing resilience of the U.S. urban food systems.

In the future, we plan to validate this model by determining indicators and their thresholds, and collecting local data from historical events, such as the 1984 Rajneeshee incident in Oregon [Hope, 2004] and the continued food insecurity post-Hurricane Katrina [Papas et al., 2015]. After validating this model, we plan to transfer this model into an interactive decision-support tool to facilitate decision-making.
and plan interventions. Also, we want to extend this Baltimore-based model to other cities in the U.S. in order to come up with a more general model, which may be used to advance national standards and policies.
Chapter 4

Microscopic Model:

Transportation

Some contents of the Chapter are based on my Master’s thesis [Zhao, 2017], a published conference paper [Zhao and Spall, 2016], and a submitted journal paper (to Transportation Research Part C: Emerging Technologies) with James C. Spall.

4.1 Introduction

As shown in Figure 1.2 and discussed in Chapter 3, transportation infrastructure plays a critical role in CIbSSs and communities. However, urban transportation networks are quite complex, consisting of many components (such as traffic signals, cars, buses, pedestrians, etc.), which makes modeling resilient transportation systems
very challenging. In this Chapter, we choose to use the origin-destination (O-D) travel time, one of the most important metrics in transportation engineering, to assess the performance of transportation networks before and after disruptions (due to emergencies such as natural and manmade hazards). The two main reasons of choosing the O-D travel time are: 1) it is affected by multiple important factors of traffic dynamics (including traffic conditions, signal timings, road attributes, driver behavior, etc.); and 2) it can produce intuitive and straightforward results.

A lot of studies have been conducted for estimating travel times under baseline conditions. In 2012, Ho\neitner et al. proposed to combine free-flow travel time with stopping time to formulate an overall travel time distribution [Ho\neitner et al., 2012]. Then, based on [Ho\neitner et al., 2012], Cao et al. used a truncated distribution to quantify free-flow travel time, and achieved more accurate overall travel time distribution [Cao et al., 2014]. Furthermore, Bayesian analysis has also been applied to similar mixture models [Jintanakul et al., 2009].

In addition, multiple recent works have been devoted to analyzing the performance of disrupted transportation network and estimating corresponding O-D travel time in emergencies such as natural and manmade hazards. Murray-Tuite and Mahmassani created a disruption index that measured the damage/disruption to the network by which the “evil entity” might rank links as its striking targets [Murray-Tuite and Mahmassani, 2004]. Suarez et al. analyzed the impacts of flooding and climate change on the system-wide performance of urban transportation infrastructure by
CHAPTER 4. MICROSCOPIC MODEL: TRANSPORTATION

using the Boston Metro Area as a case study [Suarez et al., 2005]. Shen et al. modeled transportation networks and travel times during hazards and emergency evacuations [Shen et al., 2008].

While significant results exist in the modeling of dynamic traffic network and travel time estimation, existing efforts usually suffer from high cost in data acquisition, impractical assumptions in simulation, or overly complex methods for modeling the physics of traffic.

Figure 4.1: Markovian framework for modeling traffic dynamics before and after disruptions due to adverse events such as hazards

Building on previous work, we propose to address the issues above by inventing a
novel model to assess transportation performance and O-D travel time before and after disruptions using Markov chains, maximum likelihood estimate (MLE) and Google Maps. In Figure 4.1, we show the Markovian framework for modeling traffic dynamics before and after disruptions. The overall framework is based upon Markov chain theory. First, we determine the model input, including network boundary selection, time-of-day selection, and time unit selection. After determining the model input, we estimate the travel time of links and turning probabilities at intersections. For travel time estimation, we integrate the novel data technologies (i.e., Google Maps) with the recently developed methods in MLE for full systems of multiple subsystems [Spall, 2014, Zhao and Spall, 2016]. Also, we derive the Fisher information matrix (FIM) of our MLEs. On the other hand, the turning probability estimation, which usually includes two major steps, including collecting turning flow data and applying pre-existing estimation methods, has been discussed intensively in the previous literature, for example, [Maher, 1984, Mirchandani and Head, 2001, Chen et al., 2012], so in this Chapter, we do not discuss turning probability estimation in detail. After estimating travel time and turning probability, we are able to model baseline Markov chain for traffic dynamics with incoming/outgoing traffic flows considered. We identify transition matrix of the Markov chain, and apply FIM to compute asymptotic uncertainty bounds of major route’s travel time estimation, and run Monte Carlo simulation to compute the typical travel time for any random O-D pairs within the network. Then, after an event occurs (such as flood, blast, fire, etc.) causing disrup-
tions to the network, we modify the transition matrix of the Markov chain to model the link disruptions, and run Monte Carlo simulation again to simulate traffic dynamics, estimate the O-D travel time after disruptions, and identify vulnerable links within the network.

A major reason for using the MLE-based full-system/subsystem technique is that complicated connections exist between the full system traffic behavior (routes’ travel times) and the subsystem traffic flow (links’ success rates). That is, the complexities of network traffic (e.g., traffic incidents, work zones, bad weather, poor traffic signal timing, etc.) cannot be readily modeled mathematically, but the MLEs based on test data can make full use of information at both link levels and O-D levels to properly represent these connections and implicitly capture the physics of traffic. The full-system/subsystem technique as applied to a single route has been proved simple and easy-to-implement [Zhao and Spall, 2016]; this thesis extends the idea to full networks.

To implement the MLE-based idea above, we need data to populate the model for traffic networks. Ideally, any data source that can output traffic condition on links and travel time for routes is suitable for this study. In the field of transportation engineering, real data usually come from sensors, cameras, and probe vehicles; however, these traditional data sources have their own limitations, such as being hard to acquire, expensive to purchase, limited in quantity, biased in data sampling (such as using taxi data to represent all travelers’ behavior), etc. The real-time traffic data available on Google Maps overcomes these limitations. Most importantly, Google
CHAPTER 4. MICROSCOPIC MODEL: TRANSPORTATION

Maps is an open-source platform with real-time traffic condition information and O-D travel time estimates provided by its users within the network. As an open-source platform, data acquisition and high cost problems of traditional data sources are solved by using Google Maps. Furthermore, instead of hiring drivers to collect probe data or using taxi GPS data, Google Maps has a large quantity of users. The users include common car drivers, Uber/Lyft drivers, truck drivers, and others, who provide free GPS data to Google Maps in real time, helping to solve the problems of limited data points and biased data. Compared to the traditional data sources, Google Maps has another advantage: it provides traffic data over small, medium, or large towns or municipalities throughout the country, providing the data needed to implement the Markov approach here. Moreover, in terms of data quality, some previous studies (e.g., [Ozimek and Miles, 2011, Wang and Xu, 2011]) have verified the applicability and accuracy of Google Maps data.

Another major contribution of this work is to propose a novel approach to simulate traffic dynamics after disruptions by modifying the transition matrix derived in the baseline condition. We also apply the modified transition matrix and Monte Carlo simulation to conduct sensitivity analysis in order to identify vulnerable links within the network when confronted with link blockage due to hazards.

This work extends the theoretical results from the previous studies to a complex urban network system to assess the transportation performance before and after disruptions. We apply the full-system/subsystem technique with integrated route and
link data collected from Google Maps to come up with a more powerful mathematical model. In Section 4.2, we briefly introduce the mathematical modeling process, including the maximum likelihood (ML) formulation and estimation, Markov chain identification, and assessment of transportation performance under hazards. In Section 4.3, we use a case study in downtown Baltimore to illustrate the approach and show the applicability of this model. We also conduct sensitivity analysis for the impacted network in order to identify the most vulnerable link in case of flooding. In Section 4.4, we conclude this Chapter by summarizing the strengths and limitations of the model and suggest the future work.

4.2 Methodology

In this section, we first introduce the baseline Markovian model in two major steps, including travel time estimation by using MLE and data from Google Maps, and Markov chain formulation through system identification. Then, we discuss how to extend the baseline Markovian model to measure and simulate transportation performance after disruptions due to adverse events such as hazards.

4.2.1 Travel time estimation

The conceptual illustration of the transportation network system is shown in Figure 4.2. We represent the transportation network as nodes and links. Nodes are
CHAPTER 4. MICROSCOPIC MODEL: TRANSPORTATION

intersections to the transportation network; links are the connections between the adjacent nodes [Shen et al., 2008]. Then, we identify the transportation network from origin to destination through a specific route as a full system, and traffic links as subsystems. The full system output is the O-D travel time through a specific route. According to previous studies, the travel times are commonly assumed to follow distributions such as normal, log-normal, or gamma [El Faouzi and Maurin, 2007, Uno et al., 2009]. In this study, we assume the full system outputs follow the log-normal distribution. Note that a log-normal distribution is defined on the domain of positive real numbers, which is consistent with the definition of time, and that a log-normal distribution has a long tail to the right, so the shape can well characterize the travel time distribution (there is a minimum travel time for a specific route, and the long tail can practically model the features of traffic delay). Also, as shown in [Zhao and Spall, 2016], we applied the Lilliefors goodness-of-fit test to statistically support the conclusion that the full system outputs are log-normally distributed.
CHAPTER 4. MICROSCOPIC MODEL: TRANSPORTATION

Figure 4.2: Conceptual illustration of the transportation network system from origin to destination through a specific route (highlighted in bold orange) with four subsystems \( p = 4 \), each having binary output, and one full system, having output (travel time) \( T_k \) log-normally distributed.

Moreover, we assume the subsystem output is the traffic condition shown on Google Maps (blue or yellow: “1”; red or dark red: “0”) on each link. There are several reasons why we restrict to binary outputs on the links: 1) in the urban planning point of view, urban planners always care about the overall travel time from origin to destination without paying attention to the small speed differences in links; 2) in real-time decision-making (like an emergency), a person is not going to care...
about subtle differences in speed, but is only going to care if it is “good” or “not good” to include the travel path; and 3) two different outputs from the subsystems can effectively simplify the mathematical modeling process and maintain a good accuracy, which is shown in [Zhao and Spall, 2016].

In [Zhao and Spall, 2016], we discuss the modeling process for a full system with multiple subsystems (as shown in Figure 4.2). However, for a general network, multiple full systems and subsystems need to be considered for analysis. Therefore, we extend this full-system/subsystem concept to a general network as shown in Figure 4.3. Taking a small general transportation network as an example (see Figure 4.3), we first define the boundary of network, and in this example, the boundary is Square ACIG. All the traffic links within Square ACIG are considered for analysis. Note that the different directions of a road are considered as two distinct links. For example, in Figure 4.3, link AB, from west to east, and link BA, from east to west, are treated as two separate links. Specifically, as shown in Figure 4.3, links within the network are shown in orange, incoming traffic streams are denoted as green arrows, and outgoing traffic streams are identified as red arrows. The links and nodes in the dashed border (“boundary layer”) will play a role in modeling entry and exit to the network of interest, shown in solid lines.

Consider a transportation network system that consists of $p$ links (subsystems) with binary output (“0” for congested links and “1” for non-congested links). We assume that test data for all the links, including within and across the links, are
independent. The test data for link \( j \), where \( j = 1, 2, \ldots, p \), are independent and identically distributed (i.i.d.), because we suggest collecting one data point for link \( j \) at a specific time on one day; that is to say, for link \( j \), test data collected on day 1 is i.i.d. from test data collected on day 2. (We do not assume data across links are identically distributed.) For data across links at a given time and day, distant links can be viewed as independent, whereas the traffic conditions of adjacent links may influence each other; a novel link data collection strategy was proposed in [Zhao, 2017] to solve data dependence problem, where empirical evidence was also provided.

Figure 4.3: A general transportation network: solid lines denote network of interest (A-B-C-D-E-F-G-H-I-J); dashed lines denote boundary layer
A full system is defined as the travel time from origin to destination through a specific route. However, it is hard to include all the links in one route, so we need to collect data for several full systems in order to cover all the traffic links within the network. We assume that test data for all the full systems are independent. Even though full system data might have some correlation issues among different routes, we try to minimize the correlation by properly choosing full systems. Note that formal experimental design for full system data collection could be used here for collecting data efficiently and optimally, but we do not consider that in this paper. It is also worth pointing out that full system data and subsystem data are not collected on the same day in order to ensure independence.

Let us now define $\Theta$ and describe our notation for the data. We use semicolon to represent a separate row for convenience (e.g., $[a, b; c, d]$ denotes a $2 \times 2$ matrix with rows $a, b$ and $c, d$). Suppose that data are collected for $r$ full systems in the network. Let $\zeta = [\omega_1, \sigma_1^2; \omega_2, \sigma_2^2; \ldots; \omega_r, \sigma_r^2]$ represent an $r$-by-2 matrix with $\omega_i$ and $\sigma_i^2$ representing unknown means and variances of the normally distributed logarithm of the outputs of the $r$ full systems. Let $\rho_j$ represent the success probabilities for subsystem $j$, $j = 1, 2, \ldots, p$. The parameter vector $\Theta \equiv [\rho_1, \rho_2, \rho_3, \ldots, \rho_p]^T$; elements in $\zeta$ are not included in the parameter vector to be estimated because they are uniquely determined by $\Theta$ and relevant constraints. Let $T = \{T_{11}, T_{12}, \ldots, T_{1,k(1)}; T_{21}, T_{22}, \ldots, T_{2,k(2)}; \ldots; T_{r1}, T_{r2}, \ldots, T_{r,k(r)}\}$ indicate the collection of observed, scalar-valued full system output $T_{qi}$ from $k(q)$ i.i.d. experiments on the
full system \( q, q = 1, 2, \ldots, r \), representing the O-D travel times through route \( q \) in the traffic network. Because we assume the full system outputs follow log-normal distribution as described above, we let 

\[
Z = \{Z_{11}, Z_{12}, \ldots, Z_{1,k(1)}; Z_{21}, Z_{22}, \ldots, Z_{2,k(2)}; \ldots; Z_{r1}, Z_{r2}, \ldots, Z_{r,k(r)}\} = \{ \log(T_{11}), \log(T_{12}), \ldots, \log(T_{1,k(1)}); \log(T_{21}), \log(T_{22}), \ldots, \log(T_{2,k(2)}); \ldots; \log(T_{r1}), \log(T_{r2}), \ldots, \log(T_{r,k(r)})\}
\]

represent the normally distributed collection of log-transformed full system outputs, which can facilitate the following derivation.

Here, we omit the detailed derivation of the log-likelihood function, which can be found in [Zhao, 2017]. The log-likelihood function for the entire system, including all the full system test data and the subsystem test data, is:

\[
\log L(\Theta) = \sum_{q=1}^{r} \left[ -\frac{k(q)}{2} \log(\sigma_q^2) - \frac{1}{2\sigma_q^2} \sum_{j=1}^{k(q)} (Z_{qj} - \omega_q)^2 \right]
+ \sum_{j=1}^{p} \left[ S_j \log(\rho_j) + (n(j) - S_j) \log(1 - \rho_j) \right] + \text{constant.} \quad (4.1)
\]

where \( S_j = \sum_{i=1}^{n(j)} X_{ji} \) represents the number of successes in \( n(j) \) i.i.d. experiments on subsystem \( j, j = 1, 2, \ldots, p \) and \( X_{ji} \) is the \( i \)th output of the \( j \)th subsystem for \( i = 1, 2, \ldots, n(j) \).

By differentiating the log-likelihood function shown in eqn.(4.1), we are able to
CHAPTER 4. MICROSCOPIC MODEL: TRANSPORTATION

obtain the score vector:

\[
\frac{\partial \log L(\theta)}{\partial \theta} = \sum_{q=1}^{r} \left( -\frac{k(q)}{2\sigma^2_q} h'_{q2}(\theta) + \frac{h'_{q2}(\theta)}{2\sigma^2_q} \sum_{j=1}^{k(q)} (Z_{qj} - \omega_q)^2 + \frac{h'_{q1}(\theta)}{\sigma^2_q} \sum_{j=1}^{k(q)} (Z_{qj} - \omega_q) \right)
\]

\[
+ \begin{pmatrix}
S_1 - \frac{u_1 - S_1}{1 - p_1} \\
\vdots \\
S_p - \frac{n_p - S_p}{1 - p_p}
\end{pmatrix}, \quad (4.2)
\]

where \( h_{q1}(\theta) \) and \( h_{q2}(\theta) \) are the functions that reflect the relationship between full systems and subsystems (see [Zhao, 2017] for the detailed derivation), \( h'_{q1}(\theta) \) and \( h'_{q2}(\theta) \) represent the gradients of \( h_{q1}(\theta) \) and \( h_{q2}(\theta) \) with respect to \( \theta \), for \( q = 1, 2, \ldots, r \).

Solving the score equation, \( \partial \log L(\theta)/\partial \theta = 0 \), reveals a careful balancing of information between the full system and subsystems. Generally speaking, the solution to the score equation is not unique and can only be achieved numerically.

Also, we derive the FIM for our full system and subsystem MLEs [Zhao, 2017]. In this study, our interest centers on the use of FIM for constructing confidence regions for the estimates, \( \omega_q, \sigma^2_q \). The \( p \times p \) FIM \( F_N(\theta) \) for a twice-differentiable log-likelihood function, \( \log L(\theta) \), is given by

\[
F_N(\theta) = \sum_{q=1}^{r} \left[ \frac{k(q)}{2(h_{q2}(\theta))^2} h'_{q2}(\theta)h'_{q2}(\theta)^T - \frac{k(q)}{h_{q2}(\theta)} h'_{q1}(\theta)h'_{q1}(\theta)^T \right] + J_N(\theta), \quad (4.3)
\]

where \( J_N(\theta) = \text{diag}[n(1)/(\rho_1(1 - \rho_1)), \ldots, n(p)/(\rho_p(1 - \rho_p))] \). One of the most significant properties of the MLE and FIM is asymptotic normality. Here, we are only considering uncertainty in \( \omega_q \). Based on asymptotic distribution theory described
in [Spall, 2014], we have

\[ \hat{\omega}_q \sim N(\omega_q, h_{q1}(\theta)^T F_N(\theta)^{-1} h_{q1}(\theta)), \]  
\(4.4\)

for full system \( q \) where \( q = 1, 2, \ldots, r \) and sufficiently large sample sizes. In practice, we often set \( \theta \) equal to \( \hat{\theta} \) on the right hand side of (4.4). This property of FIM is very useful, which can be readily used to compute the asymptotic uncertainty bounds for \( \hat{\omega}_q \), and can also be adapted to give uncertainty bounds on travel time reliability for any O-D pairs within the network (even those for which data were not collected).

### 4.2.2 Baseline transportation model

A discrete time Markov chain presents a random process that transits from one state to another state on a state space. First, let us recall some basic definitions about Markov chains. Consider a stochastic process with discrete state via discrete time \( Y_0, Y_1, Y_2, \ldots \). The sequence is a Markov process if the following relationship holds for all \( \tau = 0, 1, 2, \ldots \):

\[ P(Y_{\tau+1}|Y_0, Y_1, \ldots, Y_\tau) = P(Y_{\tau+1}|Y_\tau) \]

Therefore, the probability of moving to the next state at time step \( \tau + 1 \) depends only on the present state at time step \( \tau \) and not on the previous state. That is, using a Markovian model, we can predict the future state of the process solely based on its current state without knowing the full history of the process.
We let state $\mathbf{Y}_\tau = [y_{\tau1}, y_{\tau2}, ..., y_{\tau p}]^T$ be a vector that denotes the location of a given vehicle among the $p$ links of time $\tau$ (hence, $\mathbf{Y}_\tau$ is a unit vector with a 1 in one location and with 0s in all others). For example, as illustrated in Figure 2.1, if link AB corresponds to link 1, then $y_{\tau1}$ is a binary component of state $\mathbf{Y}_\tau$, which represents whether a vehicle is traveling on link AB, in the direction from node A to node B. At state $\mathbf{Y}_\tau$, if the vehicle is traveling on link AB, then $y_{\tau1}$ is equal to 1 with other elements equal to 0, i.e., $\mathbf{Y}_\tau = [1, 0, 0, ..., 0]^T$. While the Markov assumption is unrealistic from a driver’s perspective, because a driver’s behavior is based on the known origin and destination and his or her movement within the network is largely predetermined, we, as external observers, do not know what the driver’s intention is in the next moment, and thus his or her movement in the network seems random from our perspective, which makes Markov assumption reasonable to capture the stochastic nature of dynamic traffic network.

The Markov process within a transportation network with $p$ links can be completely described by a $p$-by-$p$ transition matrix $\mathbf{P}$. For example, if the chain is currently in state $\mathbf{Y}_\tau$ (corresponding to one arrangement: $y_{\tau a} = 1, a \in \{1, 2, ..., p\}$ and other elements = 0), then it moves to state $\mathbf{Y}_{\tau+1}$ (another arrangement: $y_{\tau+1, b} = 1, b \in \{1, 2, ..., p\}$ and other elements = 0) with transition probability $p_{ab}$. Note that $p_{ab}$ is one of the entries of $\mathbf{P}$.

The core of our Markovian framework is the transition matrix estimation. Figure 4.4 shows the overall transition matrix $\mathbf{P}'$ we need to estimate, including a submatrix.
In this Chapter, we skip the detailed derivation and proof of the transition matrix identification, which can be found in [Zhao, 2017]. The ultimate transition matrix $P'$ takes into account the links within the network and the incoming and outgoing traffic flows to the network. In particular, the diagonal entries are estimated by

$$
\hat{p}'_{jj} = \begin{cases} 
\frac{s_{j-1}}{s_j} & \text{for links within the network} \\
0 & \text{for entering links} \\
1 & \text{for exiting links,}
\end{cases}
$$

(4.5)

where $s_j$ represents the number of steps needed before the vehicle reaches the inter-
section at the end of the link. The off-diagonal entries can be obtained as

\[
\hat{p}_{ji} = \begin{cases} 
(1 - \hat{p}_{jj}') \times \hat{f}_{ji} & \text{for links within the network} \\
\hat{f}_{ji} & \text{for entering link } j \\
0 & \text{for exiting link } j,
\end{cases} 
\]

(4.6)

where \( \hat{f}_{ji} \) is the estimated turning probability from link \( j \) to link \( i \) for \( i \neq j \) and \( i, j = 1, 2, \ldots, p + q \).

Without assuming the stationary distribution exists for the Markov chain (which is the major assumption of previous work [Crisostomi et al., 2011, Moosavi and Hovestadt, 2013, Schlote, 2014]), we are able to identify all the parameters of the transition matrix. Identifiability is an important property that a statistical model needs to satisfy in order to draw formal conclusions and make concrete predictions. The model is identifiable if it is theoretically possible to compute the true values of the model’s parameters, when the sample size goes to infinity. Our model is locally identifiable, with full proof provided in [Zhao, 2017, Appendix A].

Previous work (e.g., [Crisostomi et al., 2011, Moosavi and Hovestadt, 2013, Schlote, 2014]) and our thesis all assume that the Markov chain is time-homogeneous. Clearly, travel times and turning probabilities are time-dependent [Skabardonis et al., 2003, Van Lint and Van Zuylen, 2005]. In practice, several different transition matrices will be needed to capture differences in traffic flow throughout the day, and each of those matrices will need to be periodically updated to capture long-term (e.g., seasonal) changes in traffic dynamics or network topology.
4.2.3 Modeling transportation performance after disruptions

After discussing travel time estimation and baseline Markov chain modeling, we move to modeling and simulating transportation performance after disruptions due to emergencies like natural and manmade hazards. This work is of great importance, because

- there is nearly no data available for measuring transportation performance under hazards, so we need to properly model the disrupted network based on the baseline model and disaster management knowledge;

- with serious threat of climate change, we may expect increased number and frequency of natural hazards, especially flood, drought, hurricane, etc., which always severely impacts the transportation infrastructure;

- transportation infrastructure plays an essential role in critical supply chains, evacuation, and urban search and rescue during and after the event, so a better model of transportation performance under hazards can significantly advance emergency and mitigation planning, and enhance the overall resilience of our CIbSSs and communities.
Figure 4.5: A transportation network with road DE disrupted: grey dashed lines indicate closed links; dark red lines denote links whose corresponding rows in the transition matrix are modified

In Figure 4.5, we show a general transportation network with road DE disrupted by an adverse event, such as flood, blast, fire, major car accident, etc. This event leads to the closure of link DE and ED. To model the link closure, we first modify the corresponding rows of link DE and ED in the transition matrix to zero vectors, indicating for these two links, probabilities of staying on the links and transferring to
neighboring links are equal to zero, since the vehicles cannot access these two links after the event.

Then, for links that can transfer to these two links (i.e., links denoted in dark red as shown in Figure 4.5), we need to modify the entries of turning probabilities to these two links to zero, because it is not possible for vehicles to turn to these two links when links are closed. However, after making these off-diagonal entries equal to zero, the sum of each row for those dark red links is not equal to one any more, which violates the core property of Markov chain transition matrix. Therefore, in this case, we assume, after modifying some off-diagonal entries to zero, other non-zero entries will be scaled up proportionally in order to ensure the sum of each row equal to one. As shown in Figure 4.5, taking link FE as an example, suppose the diagonal entry is 0.4, and the three off-diagonal entries all equal to 0.2 before the event; when the event occurs, the entry corresponding to turning to link ED is modified to 0, while other entries (link EB and EH) are scaled up proportionally, with diagonal entry equal to 0.5 and two other off-diagonal entries all equal to 0.25. That is to say, mathematically, if link $j, j = 1, 2, ..., p$ (links within the network) is disrupted due to the emergency, the modified transition matrix $M$ has following changed entries compared to $P'$:

\[
m_{jj} = m_{ji} = 0, \quad (4.7)
\]

\[
m_{ij} = 0, \text{ if } p_{ij}' > 0, \quad (4.8)
\]

\[
m_{ik} = \frac{p_{ik}'}{1 - p_{ij}'}, \text{ if } p_{ij} > 0 \text{ and } p_{ik}' > 0, \quad (4.9)
\]

90
where \( k \neq j \) and \( i, k = 1, 2, \ldots, p + q \). Except these changed entries, other entries in \( M \) are same as the corresponding entries in \( P' \). The intuition is that it is similar to a first order Taylor expansion to approximate the transportation performance after disruptions. Hence, we can not only ensure the row sum equal to one, but also keep the proportionality for non-zero entries constant for each of those dark red links (see Figure 4.5). After modifying the transition probabilities for the dark red links, we can expect a larger diagonal entry for those links, denoting higher traffic and travel time on those links due to the emergency. In addition, although the off-diagonal entries also increase after the modification, the proportionality for these entries maintain the same value; that is, for link FE, the ratio of turning probabilities to link EB and EH remains constant before and after the event occurs on link ED.

Therefore, we modify the baseline transition matrix derived in Subsection 4.2.2 to model the transportation performance after disruptions. We can directly use this Markovian model to simulate traffic dynamics and estimate travel time for the disrupted network. Also, this model can be applied to conduct sensitivity analysis to assess the vulnerable links within the network.

## 4.3 Case study

In this section, we choose downtown Baltimore as a case study to illustrate the overall framework. According to 2017 Infrastructure Report Card provided by Amer-
CHAPTER 4. MICROSCOPIC MODEL: TRANSPORTATION

American Society of Civil Engineers (ASCE) [ASCE, 2017], Maryland faces serious challenges of its own infrastructures, especially the transportation infrastructure: for example, driving roads that require repair costs each driver $550 per year in Maryland; 5.8% of bridges are structurally deficient; 82 dams are rated high-hazard potential. The vulnerable transportation infrastructure in Maryland poses serious threat to community resilience and sustainability especially in face of climate change. Therefore, we take Baltimore, the biggest city in Maryland, as an example to analyze the transportation performance before and after disruptions caused by hazards.

The transportation network in the Inner Harbor area of Baltimore City is selected as shown in Figure 4.6, where there are 16 nodes and 72 links in total with 46 links within the network and 13 entering links and 13 exiting links on the boundary layer. Note that the road between node I and node E and the road between node E and node A are one-way. So only two links (i.e., link IE and link EA) are considered for these two roads.

In this section, we introduce three major applications of the Markovian framework. First, we explain how to compute the asymptotic uncertainty bounds for MLEs of full systems (O-D travel times). Second, we simulate the Markov chain for random O-D pairs within the network before and after disruptions. Third, we conduct sensitivity analysis of the impacted network to identify the vulnerable links within the network.
Figure 4.6: A transportation network in downtown Baltimore (approximately 1 mile east of the center of the Inner Harbor area): solid lines denote network of interest (A-B-C-D-E-F-G-H-I-J-K-L-M-N-O-P); dashed lines denote boundary layer

4.3.1 Uncertainty bounds for MLEs

We considered 12 full systems in collecting data for this network (M-N-O-P-L-H-D; M-I-E-A-B-C-D; D-H-L-P-O-N-M; D-C-B-A; M-I-J-K-L-H; I-E-F-G-H-D; M-N-J-F-B-C; C-B-F-J-N-M; N-O-K-G-C-D; D-C-G-K-O-M; D-H-G-F-E; H-L-K-J-I-M). The reasons why we chose these 12 full systems are: 1) these 12 routes cover all 46 traffic links within the network; 2) these 12 routes have few overlaps of links in order to minimize correlation among full systems. Formal experimental design for how to
choose full systems to collect data can be beneficial for this study, but we do not consider it here. On the other hand, the link data collection strategy (to resolve data dependence problem) and the corresponding empirical evidence were provided in [Zhao, 2017], which is not covered here. In total, we collected 16 observations for each full system and 11 – 27 observations for subsystems from Google Maps for this network at 5pm on certain weekdays (from Monday to Friday except for U.S. legal holidays) from March 31, 2016 through December 16, 2016.

The MLEs for subsystem success probabilities are shown in Appendix B. In contrast to the indicated sample means from only link data, it is expected that the MLEs for the links better represent the true success probabilities since the MLEs implicitly incorporate link interactions via the full system data. Taking link 46 as an example, the sample mean for subsystem link data alone is 1.00, but after incorporating full system information, the MLE for the success probability in link 46 decreases to 0.80.

After obtaining the MLEs for the transportation network, we are able to use (4.4) to compute the uncertainty bounds for any routes within the network (including full systems and other routes with no data collected). As an illustration of the results that can be obtained, Table 4.1 shows the MLEs and their corresponding 95% confidence intervals for four routes: the first three are full systems, and the last one is a new route with no data collected.

According to the MLEs for \( \theta \) as shown in Appendix B, we are able to compute MLEs for the full system parameter matrix \( \zeta \). For Route 1, we have \( h_{q1} = -2.7581 \).
which corresponds to the MLE for the mean of Route 1 travel time (i.e., $\hat{\omega}_1$), when measured in units of log (hour). For the short travel time less than 1 hour, the value of $h_{q1}$ is less than zero. We can transform the MLE of $\omega_1$ back to original travel time domain, which is $60e^{\hat{\omega}_1} = 3.80$ min. By using (4.4), we are able to compute the uncertainty bounds for $\hat{\omega}_1$. Based on the asymptotic normality in (4.4), the 95% confidence bound for $\hat{\omega}_1$ is $[3.71, 3.90]$ min. Uncertainty bounds for full system travel time estimates are very useful for various stakeholders. For example, the uncertainty bounds show the estimation uncertainty of the MLEs, which can be used in transportation planning in order to be sufficiently conservative with respect to potential public sector investments or in terms of deciding any other strategy that may affect the public. In particular, it would not be good to make a decision based on an overly optimistic estimate, and the uncertainty bounds give the decision-making a formal basis for being rationally conservative.

Table 4.1: MLEs and 95% confidence intervals for routes within the network

<table>
<thead>
<tr>
<th>Number</th>
<th>Route</th>
<th>MLE (min)</th>
<th>95% confidence interval (min)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>M-N-O-P-L-H-D</td>
<td>3.80</td>
<td>[3.71, 3.90]</td>
</tr>
<tr>
<td>2</td>
<td>M-I-E-A-B-C-D</td>
<td>3.85</td>
<td>[3.74, 3.96]</td>
</tr>
<tr>
<td>3</td>
<td>M-I-J-K-L-H</td>
<td>3.50</td>
<td>[3.37, 3.64]</td>
</tr>
<tr>
<td>4</td>
<td>H-L-K-J-I</td>
<td>2.44</td>
<td>[2.36, 2.53]</td>
</tr>
</tbody>
</table>
4.3.2 Markovian simulation

Based on the network model obtained in Subsection 4.3.1, we are able to simulate the Markov Chain and compute the typical travel time for a random O-D pair within the network (not just O-D pairs in the model estimation alone) before and after disruptions.

4.3.2.1 Baseline model

As shown in Figure 4.6, we take link 70 as the entrance and link 54 as the exit, meaning that we simulate vehicles that start at link 70 and ends at link 54, regardless of routes/moving trajectories. Here, we assume equal turning probabilities for each intersection, since we did not collect turning probability data from the actual network. We then run Monte Carlo simulation 5,000 times and only save the traces of the 122 vehicles that end at link 54. Because there are 13 exits of this network and vehicles can leave at any exits during the simulation; we only care about the vehicles that leave at link 54. Figure 4.8(a) shows a histogram of the travel time for those vehicles exiting at link 54.

The travel time for a random O-D pair has an overall log-normal-type shape (although not necessarily a formal log-normal distribution since the histogram represents an amalgamation of several route choices). The median and mean of the empirical distribution of travel time for these 122 vehicles are equal to 4.33 min and 4.90 min, respectively. Google Maps also provides a travel time estimate of “typically 4 min”
for this O-D pair if the vehicle leaves at 5pm on a weekday. Finally, even though we assume equal turning probabilities for this network, the typical travel time obtained by simulations is consistent with Google Maps.

### 4.3.2.2 Modeling disruptions due to flooding

Flood risk is becoming increasingly higher, especially in the face of climate change. As shown in Figure 4.7, we show a flood risk map of Inner Harbor area in downtown Baltimore, where polygon feature indicates flood risk as defined by the FEMA Digital Flood Insurance Rate Map database [Open Baltimore, 2014]. It is apparent that, for the network of interest, many links have very high flood risk. Therefore, as an example, we consider an scenario that road NO (i.e., link 24 and 36) is flooded, aiming at estimating the typical travel time (after link disruptions) from link 70 to link 54 using Markovian simulation.

![Flood risk map of Inner Harbor area of Baltimore City](image)

Figure 4.7: Flood risk map of Inner Harbor area of Baltimore City, adapted from [Open Baltimore, 2014]
CHAPTER 4. MICROSCOPIC MODEL: TRANSPORTATION

According to eqn. (4.7) – (4.9), we modify the transition matrix derived from the baseline scenario to obtain the transition matrix under this flood event. Similarly, we run Monte Carlo simulation 5,000 times and only save the traces of vehicles that stop at link 54. Figure 4.8(b) illustrates a histogram of the travel time for those vehicles exiting at link 54. The median and mean of the empirical distribution are 4.67 min and 5.30 min, respectively.

![Histogram of travel time](image)

(a) Before disruptions  
(b) After disruptions

Figure 4.8: Histogram of travel time (origin: link 70; destination: link 54)

As shown in Figure 4.8, the distribution of travel time after disruptions is shifted to the right, with larger median and mean. To compare the results in two scenarios, we find that, after road NO is flooded, the median and mean of travel time from link 70 to link 54 are decreased by 8% respectively. For such a small network with only one road disrupted, the travel time has been decreased by 8%; we could imagine, for larger urban network with multiple roads/links flooded, vehicles can experience
CHAPTER 4. MICROSCOPIC MODEL: TRANSPORTATION

much higher-than-usual traffic and a major delay in travel time, which may bring serious economic and psychosocial impacts, and undermine the resilience of CIbSSs and communities.

4.3.3 Sensitivity analysis

According to the flood risk map shown in Figure 4.7, we find that road JN, NO, OP, PL, and LH have higher risk of flooding than other roads within this network. In other words, these five roads have higher probabilities to be flooded compared to other roads in this network. Therefore, in this Chapter, we select these five roads to conduct sensitivity analysis, in order to assess this network’s transportation performance against flooding and identify vulnerable roads/links within the network.

To be specific, we use Markovian simulation to estimate the typical travel time from link 70 to link 54 by assuming one of road JN, NO, OP, PL, and LH is disrupted at a time. The results are shown in Table 4.2, where we compute the median and mean for each of these five cases.

We find that road NO is the most vulnerable one with the highest median and mean. Therefore, among these five roads with higher flood risk in the network, we identify road NO as the most vulnerable road within the network against flooding; in other words, if road NO is disrupted, we can expect the most severe travel time delay, compared to other roads within the network with the same level of flood risk. One possible interpretation is that road NO is an essential connecting road which
connects major business buildings on both sides. On the other hand, compared with the results computed from baseline model, we find that among these five roads, road PL is the most resilient one, only having a 2% increase in mean of travel time.

Table 4.2: Sensitivity analysis of transportation performance against flooding

<table>
<thead>
<tr>
<th>Disrupted road</th>
<th>Median(min)</th>
<th>Mean(min)</th>
</tr>
</thead>
<tbody>
<tr>
<td>JN (link 4 and 14)</td>
<td>4.67</td>
<td>5.21</td>
</tr>
<tr>
<td>NO (link 24 and 36)</td>
<td>4.67</td>
<td>5.30</td>
</tr>
<tr>
<td>OP (link 25 and 37)</td>
<td>4.50</td>
<td>5.13</td>
</tr>
<tr>
<td>PL (link 10 and 20)</td>
<td>4.33</td>
<td>5.02</td>
</tr>
<tr>
<td>LH (link 11 and 21)</td>
<td>4.67</td>
<td>5.19</td>
</tr>
<tr>
<td>Baseline</td>
<td>4.33</td>
<td>4.90</td>
</tr>
</tbody>
</table>

We conduct sensitivity analysis of this network against flooding and identify the most vulnerable roads/links within the network. These results and conclusions are very useful for stakeholders, which provide them quantitative evidence to address their vulnerabilities in the transportation system, and thus to improve the resilience of CIbSSs and communities. This analysis can be easily extended to other hazards and adverse situations. Therefore, by repeating the process described above for other hazards, we are able to develop a multi-hazard decision-support tool to assess different risks and obtain vulnerable roads/links within the network in order to enhance
the overall resilience of transportation infrastructure against natural and manmade hazards.

4.4 Conclusion

In summary, this paper introduces a novel method to model dynamic traffic network and assess transportation performance before and after disruptions using a Markov chain framework and Google Maps data. By extending the previous full-system/subsystem work [Zhao and Spall, 2016] to a general transportation network, we use ML formulation and data from Google Maps to estimate the baseline travel time of each link within the network. We then develop a baseline Markov chain model to incorporate MLEs from the network model and to consider incoming and outgoing traffic streams to the network of interest. Then, we extend the baseline Markovian model to measure and simulate transportation performance under hazards, in order to inform emergency and mitigation planning of transportation infrastructure and enhance the overall resilience of CIbSSs and communities.

Furthermore, we use the FIM to compute the asymptotic uncertainty bounds for the travel time of arbitrary routes within the network under the baseline scenarios. We also propose a novel method to compute the typical travel time before and after disruptions for any random O-D pairs within the network using Monte Carlo simulation (and reflecting that multiple routes are associated with a given O-D pair).
CHAPTER 4. MICROSCOPIC MODEL: TRANSPORTATION

Furthermore, we conduct sensitivity analysis of the impacted transportation network to identify the most vulnerable roads/links within the network. This work has the potential to be applied in many areas, such as traffic control [Crisostomi et al., 2011, Spall and Chin, 1997], community detection [Moosavi and Hovestadt, 2013], and disaster management [Shen et al., 2008].

The main limitations of this work include that turning probabilities need to be obtained as an important input to the Markovian framework; the full system data collected within one day may not be fully independent as discussed in the Sections 4.2 and 4.3; and the full system data collection strategy (i.e., which routes to choose and how many data to collect in each route) might not be optimal. All of the above can, in principle, be addressed with additional resources. On the other hand, the Markovian model under hazards needs to be validated, but this problem is intrinsic and requires a lot of efforts to solve, since it is very difficult to obtain historical traffic data after disruptions.

In future work, we hope to come up with an optimal full system data collection strategy using statistical experimental design. We are now collecting Google Maps data with road closure as a proxy to mimic disruptions, and then plan to use the new data set to validate the Markovian model under hazards. Furthermore, we plan to create a multi-hazard decision-support tool for stakeholders such as urban planners and emergency managers to conduct risk assessment of transportation infrastructure and develop better strategies to enhance community resilience.
Chapter 5

Microscopic Model: Cyberinfrastructure

This Chapter is based on a journal paper that is under development with co-authors, Ian Miers, Matthew Green, and Judith Mitrani-Reiser.

5.1 Introduction

In recent years, increasingly more cyber attacks against the healthcare system have raised serious concerns about the security of healthcare cyberinfrastructure. For example, in February 2016, the entire computer network of a Los Angeles hospital was locked up by hackers, and the hospital had no choice but to pay the hackers $17,000 to regain control of their computers and patient records. MedStar Health, a non-profit
healthcare company that operates ten hospitals in the Baltimore/Washington region, was also attacked in March 2016. The company’s computer network was completely shut down after the attack, severely disrupting normal function of the healthcare system and compromising the overall security of confidential data. Taking into account these recent incidents, we find that cyber attacks compromise the overall security of healthcare infrastructure and compromise patient care in hospitals. A cyber attack against the health care system often begins with “infected emails;” after a person in the healthcare system clicks on the link or opens the attachment in the malicious email, the computer or the entire network will lock up. The victims are asked to pay a high ransom to buy back the critical information, e.g., EMRs, which they lose in the cyber attack. “Computer security experts said hospitals are particularly vulnerable because some medical equipment runs on old operating systems that cannot easily be safeguarded,” as pointed out in The Seattle Times [Pritchard, 2016]. Considering the critical role that hospitals play in communities, it becomes imperative to enhance their cyber capabilities to protect their patients’ medical records and other sensitive information. Therefore, in this Chapter, we aim at using a systems approach to model the cyberinfrastructure of hospitals against natural and manmade hazards.

As we discussed in Chapter 1, a CIbSS is comprised of interdependent buildings that, together, serve a community function and that are dependent on networks of critical lifelines (i.e., water, wastewater, power, natural gas, communications and cyber, transportation, etc.). Nearly every type of community functioning, or a CIbSS,
is increasingly reliant on cyber capabilities, such as networked information systems, automation of processes, and a shift to electronic-only record keeping systems. These dependencies enhance productivity under normal conditions, but can significantly escalate the impact of a natural disaster. Escalation may have different causes. For example, loss of power at data centers (e.g., Health Information Exchanges, off-site hospital record stores) can disable information systems even when the hospital system remains functional. Loss of network connectivity can disable access to records stored at satellite offices, or to information systems that have been moved to off-site locations. Moreover, in emergency situations IT administrators may disable or bypass security protections information, making these systems vulnerable to escalations caused by follow-on attacks (e.g., deliberate cyber attacks or malware infection). A significant vulnerability we have identified in the healthcare system is the vulnerability of existing EMR systems to communications failure or cyber terrorism. Therefore, it is critical to model the failure mechanism of the EMR system to help us understand the overall resilience of hospitals and healthcare facilities.

In this Chapter, we apply FTA (introduced in Chapter 3), a commonly-used risk assessment tool, to model the failure mechanism of Medical Records Services under hazards. FTA is a top-down, deductive failure analysis tool that evaluates the occurrence of the undesired top event (i.e., Medical Records Services are disable) based on a series of basic events, which has been used in many other applications (e.g., [Barlow and Chatterjee, 1973, Ericson and Li, 1999, Clemens, 2002]). The top
event and basic events are connected through Boolean gates and intermediate events. In the process of constructing the fault tree, we can obtain a thorough understanding of the logic and major causes that give rise to the top event [Vesely, 2002].

Further, we introduce a new technique, known as the self-protecting EMR, to enhance cyber capabilities of EMRs during emergencies. EMRs are protected by access control mechanisms that are operated by software running on servers (including local server and central server), which requires the servers to be trusted in baseline and emergent scenarios. The EMRs will fail, if servers or connections to them are disrupted during disasters. The failure of EMRs can cause serious cascading failures in Medical Records Services and other essential healthcare services. In this study, we propose to utilize a self-protecting EMR technique to enforce access control with sophisticated cryptography techniques, allowing the EMRs to be distrusted without trusted servers, and allowing access to EMRs when servers or networks fail. As a consequence, we integrate self-protecting EMR technology into the fault tree model of Medical Records Services, in order to assess how this new technique will enhance the resilience of our Medical Records Services, thus, increasing the overall resilience of the healthcare system.

Unlike using FTA as a tool to analyze the failure mechanism of a CIbSS (i.e., a food system) and apply the fault tree model to engage different stakeholders in local communities as described in Chapter 3, in this Chapter, we mainly focus on using FTA to analyze the functionality of critical infrastructure (i.e., cyberinfrastr-
CHAPTER 5. MICROSCOPIC MODEL: CYBERINFRASTRUCTURE

tructure) and Medical Records Services under hazards conditioned on whether using the self-protecting EMR technique. In particular, we investigate how single events and combinations of multiple events (i.e., coupled hazards) can lead to the top event failure (i.e., Medical Records Services failure), and compare the results from hospitals with or without using the self-protecting EMR technique. We aim at applying our fault tree model to show how the self-protecting EMR technique can help U.S. hospitals enhance their resilience against natural and manmade hazards.

The remainder of this Chapter is organized as follows: in Section 5.2, we introduce the fault tree model of Medical Record Services in hospitals. We also discuss why current medical records systems are vulnerable, what the self-protecting EMR is, why this technique is important, and how this technique can improve our existing systems. Then, we explore the interdependencies of hospital departments on IT Department and computers, showing the importance of cyberinfrastructure to the entire healthcare system. In Section 5.3, we discuss the applications of the fault tree model of Medical Record Services, including deterministic analysis and probabilistic analysis. In particular, we use scenario analysis (a deterministic approach) to identify the vulnerabilities of the Medical Records Services under different types of hazards, and provide suggestions to healthcare stakeholders, emergency managers, and policy makers to improve the resilience of Medical Records Services in healthcare facilities. In Section 5.4, we conclude this Chapter, summarize the strengths and limitations of the fault tree model, and suggest future work.
5.2 Methodology

In this section, we introduce the fault tree model of Medical Records Services. Next, we introduce the self-protecting EMR technique using attribute based encryption (ABE). Then, we integrate the self-protecting EMR technique with the fault tree of Medical Records Services. Last, we analyze the interdependencies of many hospital departments on IT Department and computers in hospitals.

5.2.1 Fault tree of Medical Records Services

FTA has been introduced and discussed in Chapter 3. Here, let us briefly refresh the basic concepts and symbols of FTA. FTA is a risk-assessment method where the performance of a complex system can be examined in terms of the performance of its subsystems. The failure of the top event is determined by failure of basic events through a combination of logical gates and intermediate events. The symbols of FTA can be divided into two categories: event symbols and gate symbols. Although more types of event or gate symbols are available for complex fault trees, in this Chapter, we only use two types of event symbols (basic event and top event/intermediate event), and three types of gate symbols (“and” gate, “or” gate, and “transfer” gate). To be specific, basic event refers to failure or error in a basic system component or element; top event/intermediate event describes the consequences of a combination of lower level events; “and” gate indicates the output occurs only if all inputs occur; “or” gate
indicates the output occurs if any input occurs; “transfer” gate refers to transferring
to/from another part of the fault tree.

FTA is used in this Chapter to model Medical Records Services failure impacted
by natural and manmade hazards. In 1982, Fischhoff et al. pointed out that FTA is
crucial for stakeholders to understand how the complex systems they manage will be
impacted by a hazard and to inform decision making [Fischhoff et al., 1982]. In this
Chapter, we try to model the failure mechanism of Medical Records Services in the
hospital system when impacted by hazards. Figure 5.1 illustrates a fault tree that
generically captures a series of events that result in the failure of Medical Records
Service in the hospital. This fault tree includes two subtrees: staff (non-clinical staff)
and space (critical infrastructure, structural and non-structural elements, egress, etc.).
As shown in Figure 5.1, the staff subtree is outlined with a red dotted box and the
space subtree is outlined with a blue dotted box. Also, note that the communications
infrastructure subtree is denoted by a green dotted box. Apparently, the communica-
tions infrastructure subtree will fail, if both the internal IT network and the Internet
access fail, which is very likely to happen in both natural and manmade hazards.
CHAPTER 5. MICROSCOPIC MODEL: CYBERINFRASTRUCTURE

Figure 5.1: Fault tree of Medical Records Services (red: staff; blue: space; green: communications infrastructure)
As shown in Figure 5.1, the fault tree of Medical Records Services is deterministic, with all the basic events and intermediate events having binary ("0" or "1") outcomes. The top event is the failure of the Medical Records Services in the hospital. The intermediate events are subsystem states which describe the consequences of lower level events through gates and lead to the occurrence of upper level events, including "alternate staff cannot maintain medical records," "medical records staff are unavailable," "power infrastructure fails," "communications infrastructure fails," "critical infrastructure fails," "electronic medical records cannot be accessed," "assigned the space is severely damaged," "horizontal means of egress are compromised," "vertical means of egress are compromised," "space is not accessible vertically," "assigned space is inaccessible," "physical medical records cannot be accessed," and "medical records cannot be accessed." There are 15 basic events shown at the bottom of the fault tree, where 3 basic events belong to the staff subtree and 12 basic events belong to the space subtree. To be specific, the basic events considered in the staff subtree include: "administrators are absent," "alternate staff are absent," and "alternate staff are not trained." The basic events in the space subtree contain: "computer equipment is damaged," "municipal power fails," "backup power fails," "internal IT network fails," "Internet access fails," "severe structural damage," "severe non-structural damage," "corridors are severely damaged," "exterior exits are damaged, blocked, or non-existent," "elevators fail," "stairs fail," and "no space is on ground floor."
We want to stress here that communications infrastructure failure (denoted in a green dotted box in Figure 5.1) occurs very often in different types of hazards, such as earthquakes, hurricanes, cyber attacks, etc. Considering recent events of cyber attack against hospitals, it is imperative to improve the overall resilience of hospitals and healthcare facilities by enhancing their cyber capabilities. Next, we will introduce a new technique, called the self-protecting EMR, which can improve the resilience of cyberinfrastructure in the hospital.

5.2.2 Self-protecting EMRs

In this subsection, we introduce the notion of pre-positioned self-protecting EMRs to deal with servers and/or communications failures.

Some previous work (e.g., [Akinyele et al., 2011, Ibraimi et al., 2009a, Ibraimi et al., 2009b, Narayan et al., 2010]) proposed to use ABE to protect EMRs. According to [Goyal et al., 2006], ABEs allow for access control policies to be cryptographically implemented on encrypted records, such as EMRs. The major contribution of ABEs is that they can prevent collusion between users. For instance, when two types of users have the nurse attributes and pharmacist attributes, respectively, they cannot access the medical records that require both two attributes to decrypt.

EMRs are usually stored in operation systems administered by each hospital or healthcare system. The current record-sharing mechanism is that other users can gain access to EMRs stored on the originating hospital/healthcare system’s servers which
provide access control and logging. This mechanism actually simplifies technical and liability issues, and it becomes very crucial in hazards when patients and staff are transferred more frequently due to demands.

However, this mechanism requires availability of hospital/healthcare system’s server and working network connections. Neither of these can be guaranteed during major hazards. Hence, our team members from computer science (i.e., Ian Miers and Matthew Green) propose a novel approach to tackle the limitations of this mechanism.

They create a new method, called self-protecting EMRs, to preposition copies of EMRs, which are protected by ABE and a master key/password to local servers. In the baseline condition, the master key/password is kept secret. While a disaster prevents access to the central server, the key/password will be distributed and the ABE-protected EMRs will become available to the impacted hospitals.

\section*{5.2.3 Integration of self-protecting EMR technique to Medical Records Services}

The algorithms for protecting the EMR and making them accessible to critical staff, via the self-protecting mechanism, are integrated into the fault tree structure. Specifically, we replace the part in the green dotted box shown in Figure 5.1 with a cyberinfrastructure subtree (see Figure 5.2). There are five intermediate events included in this subtree, i.e., “communication infrastructure fails,” “central server
is unavailable,” “central server is unreachable,” “local server is unreachable,” and “cyberinfrastructure fails.” Moreover, there are eight basic events considered, namely, “internal IT network fails,” “Internet access fails,” “central server is hacked,” “central server is damaged,” “central server’s power fails,” “local server is hacked,” “local server is damaged,” and “password is lost.”

With the application of the self-protecting EMR technique, we find that it becomes more difficult to fail cyberinfrastructure, and thus the Medical Records Service is more resilient under hazards. For example, in the original fault tree, if both the internal IT network and Internet access fail, the communications infrastructure will fail, resulting in the failure of EMRs. In contrast, in the modified fault tree, although both the internal IT network and Internet access fail, the cyberinfrastructure will not fail, so EMRs will not fail in this case (as long as the local server remains intact). In other words, the resilience of the EMRs and the overall Medical Records Service is greatly improved by utilizing the self-protecting EMR technique.
Figure 5.2: Fault tree modification by including self-protecting EMR technique
5.2.4 Dependencies on IT Department and computers

Interdependencies among hospital departments or services have been recognized and stressed in many previous works (e.g., [Abernathy and Lillis, 2001, Khanna et al., 2012, Jacques, 2016]). The interdependencies among hospital departments or services strongly influence the efficiency and effectiveness of the hospital system under normal conditions, and more importantly, determine the resilience of the entire healthcare system during stressed conditions (i.e., natural and manmade hazards). Therefore, it is meaningful to identify and study dependencies of hospital departments on IT Department and computers. Note that the dependencies of many hospital services on IT Department and computers have also been discussed in [Jacques, 2016]. As shown in Figure 5.3, many hospital departments/services depend highly on IT Department and computers.

Figure 5.3 shows that there are a total of 14 services/departments in the hospital depending on IT Department and computers. To be specific, there are six clinical services/departments relying on IT Department and computers: Adult Emergency Department, Adult Respiratory Therapy, Pediatric Emergency Department, Pediatric Intensive Care, Neurological Critical Care, and Social Work. For non-clinical services/departments, Medical Records Services, Materials Management, Health Safety and Environment Services, Admitting and Bed Management, Facilities Engineering,
CHAPTER 5. MICROSCOPIC MODEL: CYBERINFRASTRUCTURE

Human Resources, Information Technology, and Patient and Visitor Services are dependent on IT Department and computers. Over 80% of all the departments/services in the hospital rely heavily on IT Department and computers. Therefore, it is obvious that if IT Department and computers are compromised, the hospital will be heavily impacted in clinical and non-clinical areas. Apparently, a cyber attack could wreak havoc in hospitals and healthcare systems.

Figure 5.3: Illustration of hospital services relying on IT Department and computers

Considering so many hospital departments or services depend on IT Department and computers, it is imperative to attach more importance to the cyber security,
which may result in losing access to the computers and internal network failure, if the hospital becomes the victim of cyber attack. According to the real cases mentioned in the first section of the Chapter, after IT Department and computers are compromised, the internal computer network will be shut down, and the EMRs are the first to become inaccessible if the hospital does not use the self-protecting EMR technique. Besides, if the hospital does not have physical medical records in place when EMRs are compromised, many hospital services cannot function as normal.

With increasingly more importance attached to cyberinfrastructure, we should acknowledge more significance of the self-protecting EMR technique. According to Figure 5.2, we can easily draw the conclusion that with the self-protecting EMR technique in place, even if computers in the hospital are hacked and internal network is down, doctors and nurses can still gain access to EMRs as long as a local server remains accessible. Of course, the cyber security awareness should be raised and the resilience of cyberinfrastructure should be enhanced for most hospitals in the U.S.

5.3 Applications

In the previous sections, we have introduced the fault tree models of Medical Records Services in the healthcare system with or without using the self-protecting EMR technique. In this section, we discuss the applications of these fault tree models against natural and manmade hazards by using deterministic and probabilistic
CHAPTER 5. MICROSCOPIC MODEL: CYBERINFRASTRUCTURE

approaches.

5.3.1 Deterministic analysis

In this subsection, we use scenario analysis (a typical deterministic approach) to assess and compare the performance of Medical Records Services in hospitals with or without using the self-protecting EMR technique under (coupled) hazards.

It is notable that natural hazards will become more devastating if coupled with a cyber attack. However, hackers may see these natural hazards as opportunities to claim much higher ransom. Therefore, we consider seven different types of scenarios in this section to investigate the application of fault trees, draw inferences from the results, and provide suggestions to healthcare stakeholders, emergency managers, policy makers, and researchers. The seven different scenarios are summarized as follows:

- Case 1: earthquake
- Case 2: cyber attack against hospital computers
- Case 3: cyber attack against central server
- Case 4: cyber attack against local server
- Case 5: earthquake coupled with cyber attack against hospital computers
- Case 6: earthquake coupled with cyber attack against central server
Case 7: earthquake coupled with cyber attack against local server

The reasons to choose these seven different scenarios are:

- strong earthquakes are representative of natural hazards that are characterized with severe structural damage, fatality rates, and demand for healthcare services;

- cyber attack is a relatively new type of manmade hazards, but it has become a real threat to many critical infrastructures in the society, especially for healthcare facilities as mentioned in Section 5.1;

- it is meaningful to investigate the escalating failures in CIbSSs, when natural hazards are coupled with different types of cyber attacks, i.e., cyber attack against hospital computers, cyber attack against central server, and cyber attack against local server, where cyber attack against hospital computers is the most common case among these three scenarios (cyber attack against central server or local server is less common, but can bring more severe damage and disruption to the entire healthcare system).

Taking Scenario 1 as an example, we show how to use FTA to conduct risk assessment. To be specific, we choose a medium-level earthquake, such as the 2011 Mineral, Virginia, earthquake [USGS, 2011]. As discussed in [Boston, 2017], the ground motion of a medium-level earthquake is commonly selected as an intensity measure equivalent to the design basis earthquake (DBE), which refers to “the earthquake which the
structure is required to safely withstand with repairable damage” [Bureau of Reclamation Glossary, nd]. The DBE represents an earthquake that has a 10% chance of occurrence in 50 years, and it is the minimum intensity at which the hospitals are supposed to withstand and keep functional.

Boston discussed the outages of power and water infrastructure, and the damage of structural and non-structural components of the impacted hospital after a medium-level earthquake [Boston, 2017]. On the other hand, as pointed out in [Townsend and Moss, 2005], communications infrastructure, as the most sophisticated but fragile infrastructure, is damaged in nearly every major urban disaster; however, the size of the disaster is not the determining factor, but how its geographical distribution of damage coincides with communications facilities. After the 2011 Mineral, Virginia, earthquake (a medium-level earthquake) hit the East Coast, it was wreaking havoc on the mobile networks from Verizon, Sprint, T-Mobile, AT&T [Smith, 2011]. Therefore, after a medium-level earthquake, the communications networks are very likely to fail.

In this paper, we use some of the results from [Boston, 2017], and assume that communications infrastructure fails (both internal IT network and Internet access fail), and alternate staff are absent due to safety concerns, in order to populate the fault trees. As shown in [Boston, 2017], after this medium-level earthquake occurs, the shaking of the earthquake causes power failure, elevators failure, and severe damage of non-structural components (including hot water piping, chilled piping, steam piping, cooling tower, etc.) for both moment frame and base isolated structures. For the
moment frame structures, the non-structural damage is even more severe, such as damage of fire sprinklers, sanitary pipe bracing, chillers, etc.

Therefore, based on results and assumptions discussed above, we populate the original fault tree, and the propagating failure process is illustrated in Figure 5.4. With both electronic and physical medical records becoming inaccessible, all the medical records cannot be accessed, so the top event – Medical Records Services are disabled – will occur. In other words, the original fault tree will fail under the attack of a medium-level earthquake.

Then, for the revised fault tree with cyberinfrastructure considered, after the same events happen, the propagating failure process is illustrated in Figure 5.5 – 5.6. Under the attack of the same medium-level earthquake, the same basic events fail or get triggered. We can find that, as shown in Figure 5.5, the cyberinfrastructure will not fail, even though communications infrastructure failure leads to the central server being unreachable, because the EMRs stored in the local server remain accessible during the emergency. Figure 5.6 shows that the top event will not occur in the revised fault tree, so we may draw the conclusion that self-protecting EMRs enhance the resilience of cyberinfrastructure and Medical Records Services in hospitals.
CHAPTER 5. MICROSCOPIC MODEL: CYBERINFRASTRUCTURE

Figure 5.4: Propagating failures of the original fault tree (Case 1)
CHAPTER 5. MICROSCOPIC MODEL: CYBERINFRASTRUCTURE

Figure 5.5: Propagating failures of the cyberinfrastructure subtree (Case 1)
CHAPTER 5. MICROSCOPIC MODEL: CYBERINFRASTRUCTURE

Figure 5.6: Propagating failures of the revised main tree (Case 1)
The detailed results of these seven scenarios are shown in Table C.1 – C.7 in the Appendix C. We identify the outcomes of basic events for both the original fault tree and the revised fault tree under different scenarios, where “TRUE” means the event occurs and “FALSE” means the event does not occur. Note that “N/A” indicates these basic events are not included in the fault tree. We conduct propagating failure process for these seven scenarios, and we find that the original fault tree in Case 1, Case 5, and Case 6, and both the original and revised fault trees in Case 7 fail. One observation is apparent: the revised fault tree is more resilient to different types of disasters.

Note that many hospitals in the U.S. are switching to EMRs from paper records, due to many advantages of EMRs [Hoyt and Yoshihashi, 2014]. For those hospitals with full transition from paper records to EMRs, without using the self-protecting EMR technique, they will lose access to EMRs and thus the entire Medical Records Services will be disabled, if Case 2 occurs. But if they use self-protecting EMRs in the hospital, they can still function as normal in Case 2. For Case 5 and Case 6, namely, the earthquake coupled with cyber attack against hospital computers and earthquake coupled with cyber attack against central server, the hospital, which uses the self-protecting EMR technique, can survive the crisis and function in the emergency. However, the hospital without using self-protecting EMRs cannot. For Case 7, we find that even the hospital that uses the self-protecting EMR technique still fails in this scenario, because cyberinfrastructure fails due to both the central server and
local server becoming unreachable in this case. In contrast, in Case 5 and Case 6, although hospital loses access to central server, the EMRs stored in the local server become available in emergencies, so the top event will not happen. In other words, that local server remains accessible is the key to ensuring Medical Records Services functional after hazards. The local server plays a critical role in the resilience of cyberinfrastructure and the entire healthcare system. Therefore, we need to guarantee the local server remains functional in hazards by the following approaches:

- local server has high level cyber capability against hacking;
- local server should be housed in a building built with performance-based design with higher standards, such as base isolated structure, to minimize physical damage in dynamic loads from hazards;
- the key/password to the local server needs to be available during hazards.

5.3.2 Probabilistic analysis

The fault tree gives us a framework for both qualitative and quantitative assessment of the top event [Vesely, 2002]. In the previous subsection, we have shown how to use the fault tree model of Medical Records Services in a deterministic way. It can also be extended to probabilistic FTA.

To be specific, we first need to collect failure data for all the basic events shown in the fault tree. Taking “severe non-structural damage” for example, we need to
CHAPTER 5. MICROSCOPIC MODEL: CYBERINFRASTRUCTURE

collect the percentage of hospitals in the impacted region that experience severe non-structural damage after an event, and repeat this data collection process for a series of similar hazard events (such as earthquakes with similar magnitude). After that, we can obtain the empirical probability distribution of this basic event. In this way, we collect data and populate empirical probability distributions for all the basic events. Then, based on the empirical probability distributions of basic events and the logic provided by the fault tree, we could use Monte Carlo simulation to obtain the probability distribution of the top event, namely, the failure probability distribution of Medical Records Services.

Note that the data at the basic event level may not always be available. One way to resolve this issue is to collect data for the intermediate events that are connected to the basic events with no data available. Another way is to use pre-determined probability distributions for these basic events, according to previous literature or expert judgment.

5.4 Conclusion

In the final analysis, we propose a novel approach to model the cybersecurity of hospitals in both natural and manmade hazards. Furthermore, in order to address the cyber vulnerabilities in hospitals, we propose to use self-protecting EMRs to provide Medical Records Services in hospitals with higher cyber capabilities against commu-
communications failure and cyber terrorism. To be specific, we conduct risk assessment by developing fault tree models of Medical Records Services in hospitals with or without using the self-protecting EMR technique. Besides, we also find that over 80% of all the departments/services in the hospital rely heavily on IT Department and computers, so cybersecurity is of great significance when it comes to emergency preparedness in hospitals. Furthermore, we conduct scenario analysis for seven typical cases, including a medium-level earthquake, three different types of cyber attack, and the earthquake coupled with each one of the three different cyber attacks. The results indicate that a hospital using the self-protecting EMR technique will survive more hazards, showing higher resilience to both natural and manmade hazards. We also identify how to conduct probabilistic analysis for the fault trees. The framework proposed in the Chapter is very useful for healthcare stakeholders, emergency managers, policy makers, and researchers in the process of emergency and mitigation planning before hazards, decision making during hazards, and hospitals’ recovery after hazards.

However, we need to admit that the fault trees shown in the Chapter may not be very comprehensive, and the fault tree structure may vary from hospital to hospital. Therefore, it is strongly advised for each hospital or healthcare system use the method presented in this Chapter to develop and populate its own fault trees, and identify and address its own vulnerabilities. Moreover, self-protecting EMRs do not provide a full replacement for an EMR system after disasters. So it calls for a more sophisticated technique to better prepare EMRs from communications failure and cyber terrorism.
CHAPTER 5. MICROSCOPIC MODEL: CYBERINFRASTRUCTURE

In the future, we plan to collect data to populate the probabilistic fault trees of Medical Records Services in hospitals. In addition, we want to extend this work to dynamic FTA, as discussed in [Dugan et al., 1992]. Furthermore, we plan to apply FTA to other essential medical equipment and other departments/services in hospitals to facilitate more comprehensive and effective emergency planning for healthcare systems.
Chapter 6

Conclusion and Future Work

This dissertation is devoted to conducting multi-scale community resilience modeling against natural and manmade hazards. We apply probabilistic and statistical methods and systems tools to model, measure, and analyze the risk and resilience of our communities, CIbSSs, and critical infrastructure systems in face of threats from hazards and climate change. The major contributions, limitations, and future work summarized below.

6.1 Major contributions

In this dissertation, we build four models (i.e., CoPE-WELL, food system model, transportation infrastructure model, and cyberinfrastructure model) at three different levels to measure and quantify resilience before, during, and after disasters.
6.1.1 Macroscopic/community level

- We develop a novel county-level community resilience index (i.e., CoPE-WELL) to quantify resilience over time, which is the first dynamic index in the research theme of community resilience. CoPE-WELL can capture the dynamic behavior of resilience by splitting this concept into three major components, including pre-event functioning (contains ten domains), resistance (contains six domains), and recovery (contains three domains), and aggregating all these domains together using systems dynamics modeling.

- We create a hazard-specific weighting scheme for community resilience indicators in the engineered systems domain by using a data-driven approach. Furthermore, by applying model comparison techniques, we obtain the “best predictive model” with three indicators, including percentage of not mobile homes, number of deficient or obsolete bridges, and median mobile download speed. The results obtained in the hurricane case not only show the different importance of indicators in the engineered systems domain, but also validate the directionality of indicators that is pre-determined by expert judgment.

6.1.2 Mesoscopic/CIbSS level

- We conceptualize the urban food system failure mechanism by splitting the concept into three major components, including unavailability, inaccessibility,
CHAPTER 6. CONCLUSION AND FUTURE WORK

and unacceptability.

- A novel risk assessment tool is created to assess the resilience and identify vulnerabilities of urban food system by using FTA.

- We discuss the applications of the fault tree model of urban food systems, including case studies and stakeholder engagement, aiming at planning more resilient and sustainable food systems especially in face of climate change-induced phenomena.

### 6.1.3 Microscopic/critical infrastructure level

For the transportation infrastructure model, our major contributions are:

- We develop a novel approach to model transportation network and traffic dynamics before and after disruptions using a Markovian framework and data from Google Maps.

- The transportation model integrates novel data technologies with recently developed methods in MLE for systems of multiple subsystems.

- We model and simulate the transportation system performance after disruptions due to emergencies like hazards, and identify the vulnerable links within the network.

For the cyberinfrastructure model, our results are summarized as follows:
CHAPTER 6. CONCLUSION AND FUTURE WORK

- We conduct risk assessment of cyberinfrastructure and Medical Records Services in hospitals by using FTA.

- A new technique, called self-protecting EMRs, is integrated into the fault tree model. We conduct scenario analysis to show enhanced cyber capabilities of hospitals by using this technique.

6.2 Limitations

According to the famous quote from George E. P. Box, “essentially, all models are wrong, but some are useful,” admittedly, all the models described in this dissertation have their own pros and cons. Here, we summarize the major limitations as follows:

- For the CoPE-WELL index, the systems dynamics model has not been fully calibrated and validated, and the psychosocial impacts of the disasters have not been considered in the model.

- For the hazard-specific weighting scheme, the disaster data are not adequate and engineered systems’ indicators considered in the study are not very comprehensive (lacking indicators for waste water, power, and natural gas).

- The fault tree model of urban food system is not exhaustive of all the potential basic events, and the tree has grown very large, limiting ability of communicating this model to stakeholders.
CHAPTER 6. CONCLUSION AND FUTURE WORK

- For the transportation model, the full system data collection strategy may not be optimal. The Markovian model for transportation performance after disruptions needs to be validated with real data.

- For the cyberinfrastructure model, the presented fault tree model of Medical Records Services is deterministic and may not be applicable for all hospitals in the U.S.

6.3 Future work

In the short term, I would like to accomplish following tasks:

- I plan to develop novel modules that can be included in most community resilience models to improve their predictive capability. In addition to the hazard-specific weighting scheme, I want to develop new dynamic models to capture the impact trajectories of disasters to local communities by taking into account the physical and psychosocial characteristics of disasters.

- I want to collect more data for disasters, and identify more indicators for the engineered systems domain, in order to standardize the weighting scheme. I also plan to extend current work to other domains of community functioning. The local and federal government are encouraged to collect more data at different levels, especially the data directly related to critical infrastructures, which can essentially benefit community resilience modeling and planning.
• I plan to look into different choices for the response variable (to proxy resilience),
  by taking into account cyclical pattern of national economy and UR. I also plan to
  compare these choices using statistical tools like machine learning.

• I am working on determining indicators and their thresholds for the fault tree
  model of urban food system, and collecting data from historical events, in order
  to validate the model and quantify resilience.

• I want to transfer the fault tree model into an interactive decision-support
  tool to characterize systems, assess mitigation plans, and enhance food system
  resilience for the local communities.

• I hope to develop optimal data collection strategy for full systems by using
  proper experimental design. I am now collecting Google Maps data with road
  closure to serve as a proxy to mimic link disruption after disasters, and plan to
  use the data to validate the proposed Markovian model under hazards.

• I plan to compare our Markovian traffic model with other traffic simulation
  tools, such as Simulation of Urban MObilility (SUMO), to show strengths and
  weaknesses of our model, and to improve our model accordingly.

• I plan to develop the probabilistic fault tree model of Medical Records Services
  by using historical data from previous events. I also want to extend this work
  to other essential medical equipment and other hospital departments/services.
  Each hospital or healthcare system is strongly advised to develop its own set of
fault trees to identify its own vulnerabilities, and enhance its resilience against natural and manmade hazards.

- I want to conduct sensitivity analysis based on the probabilistic fault tree model of Medical Records Services in order to identify the most vulnerable nodes within the tree, and propose corresponding approaches to address these issues and enhance resilience of healthcare facilities.

In the long term, I have two major goals: integrating resilience models at different levels and modeling resilient and sustainable transportation infrastructure. To achieve these two goals, I plan to achieve following objectives:

- I want to build a decision-support tool for assessing the resilience for the entire community by using FTA to combine resilience models at different levels, and conduct deterministic and probabilistic analysis based on the fault trees.

- I plan to build a community resilience simulator by applying systems dynamics modeling or agent-based modeling to integrate multi-scale community resilience models.

- I want to use the transportation model to analyze transportation performance under multi-hazard impacts.

- I plan to analyze the risks and vulnerabilities of transportation systems in face of climate change.
CHAPTER 6. CONCLUSION AND FUTURE WORK

- I plan to apply the transportation model to food supply chains to plan for resilient urban food systems under multiple climate change-induced weather phenomena and to develop engineered and nature-based mitigation plans to ensure food security.

- I want to integrate the transportation model with optimization, data science, and agent-based modeling, and apply it to evacuation and urban search and rescue, aiming at developing a dynamic decision-support tool for response to hazard events.

- I plan to further model the transportation infrastructure as a cyber-physical system, and then use it to assess the impacts of electric vehicles and autonomous vehicles on the resilience of transportation infrastructure system, CIbSSs, and communities.
Appendix A

Fault Trees of Food System
APPENDIX A. FAULT TREES OF FOOD SYSTEM

Figure A1: Subtree 1: high food price
Figure A.2: Subtree 2: sufficient decrease in net income
APPENDIX A. FAULT TREES OF FOOD SYSTEM

Figure A.3: Subtree 3: food purveyors are not accessible
Figure A.4: Subtree 3-1: public transit is not available
Figure A.5: Subtree 4: unable to leave home
APPENDIX A. FAULT TREES OF FOOD SYSTEM

Figure A.6: Subtree 5: production failure
APPENDIX A. FAULT TREES OF FOOD SYSTEM

Figure A.7: Subtree 6: production failure
APPENDIX A. FAULT TREES OF FOOD SYSTEM

Figure A.8: Subtree 7: wholesale is disrupted
Figure A.9: Subtree 8: distribution is disrupted
APPENDIX A. FAULT TREES OF FOOD SYSTEM

Figure A.10: Subtree 8-1: distribution centers are disrupted
APPENDIX A. FAULT TREES OF FOOD SYSTEM

Figure A.11: Subtree 9: retail is disrupted
APPENDIX A. FAULT TREES OF FOOD SYSTEM

Figure A.12: Subtree 10: food bank donation failure
Figure A.13: Subtree 11: other food assistance organization donation failure
APPENDIX A. FAULT TREES OF FOOD SYSTEM

Figure A.14: Subtree 12: supply chain failure
Appendix B

Estimation Results for Network in Downtown Baltimore

The sample means below are the estimates of $\rho_j$ from data on link $j$ only; the MLEs are the estimates from link and route (full system) data.
## APPENDIX B. ESTIMATION RESULTS FOR NETWORK IN DOWNTOWN BALTIMORE

Table B.1: Sample means and MLEs for links’ success probabilities

<table>
<thead>
<tr>
<th>Link</th>
<th>Sample Mean</th>
<th>MLE</th>
<th>Rel. Diff.</th>
<th>Link</th>
<th>Sample Mean</th>
<th>MLE</th>
<th>Rel. Diff.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.80</td>
<td>0.77</td>
<td>−4.36%</td>
<td>24</td>
<td>0.92</td>
<td>0.93</td>
<td>−4.70%</td>
</tr>
<tr>
<td>2</td>
<td>0.50</td>
<td>0.54</td>
<td>7.25%</td>
<td>25</td>
<td>1.00</td>
<td>0.95</td>
<td>−1.93%</td>
</tr>
<tr>
<td>3</td>
<td>0.78</td>
<td>0.78</td>
<td>−0.04%</td>
<td>26</td>
<td>0.50</td>
<td>0.52</td>
<td>4.24%</td>
</tr>
<tr>
<td>4</td>
<td>0.65</td>
<td>0.61</td>
<td>−6.34%</td>
<td>27</td>
<td>0.89</td>
<td>0.87</td>
<td>−1.93%</td>
</tr>
<tr>
<td>5</td>
<td>0.89</td>
<td>0.90</td>
<td>1.05%</td>
<td>28</td>
<td>0.42</td>
<td>0.47</td>
<td>11.58%</td>
</tr>
<tr>
<td>6</td>
<td>0.88</td>
<td>0.89</td>
<td>1.16%</td>
<td>29</td>
<td>0.85</td>
<td>0.84</td>
<td>−1.94%</td>
</tr>
<tr>
<td>7</td>
<td>0.63</td>
<td>0.68</td>
<td>9.13%</td>
<td>30</td>
<td>0.81</td>
<td>0.80</td>
<td>−0.48%</td>
</tr>
<tr>
<td>8</td>
<td>0.64</td>
<td>0.68</td>
<td>5.84%</td>
<td>31</td>
<td>0.81</td>
<td>0.78</td>
<td>−4.33%</td>
</tr>
<tr>
<td>9</td>
<td>0.56</td>
<td>0.60</td>
<td>8.29%</td>
<td>32</td>
<td>0.96</td>
<td>0.96</td>
<td>0.03%</td>
</tr>
<tr>
<td>10</td>
<td>0.88</td>
<td>0.88</td>
<td>−1.07%</td>
<td>33</td>
<td>1.00</td>
<td>1.00</td>
<td>0.00%</td>
</tr>
<tr>
<td>11</td>
<td>0.89</td>
<td>0.84</td>
<td>−5.20%</td>
<td>34</td>
<td>0.96</td>
<td>0.96</td>
<td>0.12%</td>
</tr>
<tr>
<td>12</td>
<td>0.96</td>
<td>0.91</td>
<td>−5.13%</td>
<td>35</td>
<td>0.96</td>
<td>0.94</td>
<td>−2.39%</td>
</tr>
<tr>
<td>13</td>
<td>0.55</td>
<td>0.56</td>
<td>2.58%</td>
<td>36</td>
<td>0.96</td>
<td>0.98</td>
<td>1.37%</td>
</tr>
<tr>
<td>14</td>
<td>1.00</td>
<td>0.98</td>
<td>−1.58%</td>
<td>37</td>
<td>0.88</td>
<td>0.86</td>
<td>−2.52%</td>
</tr>
<tr>
<td>15</td>
<td>0.85</td>
<td>0.82</td>
<td>−2.61%</td>
<td>38</td>
<td>0.96</td>
<td>0.95</td>
<td>−0.97%</td>
</tr>
<tr>
<td>16</td>
<td>0.74</td>
<td>0.76</td>
<td>2.79%</td>
<td>39</td>
<td>0.88</td>
<td>0.88</td>
<td>−1.07%</td>
</tr>
<tr>
<td>17</td>
<td>0.69</td>
<td>0.84</td>
<td>22.02%</td>
<td>40</td>
<td>0.96</td>
<td>0.94</td>
<td>−2.09%</td>
</tr>
<tr>
<td>18</td>
<td>0.70</td>
<td>0.82</td>
<td>16.46%</td>
<td>41</td>
<td>0.81</td>
<td>0.85</td>
<td>5.83%</td>
</tr>
<tr>
<td>19</td>
<td>0.64</td>
<td>0.76</td>
<td>20.12%</td>
<td>42</td>
<td>0.67</td>
<td>0.77</td>
<td>15.02%</td>
</tr>
<tr>
<td>20</td>
<td>0.96</td>
<td>0.95</td>
<td>−1.14%</td>
<td>43</td>
<td>0.92</td>
<td>0.91</td>
<td>−1.25%</td>
</tr>
<tr>
<td>21</td>
<td>0.77</td>
<td>0.82</td>
<td>6.01%</td>
<td>44</td>
<td>0.85</td>
<td>0.81</td>
<td>−4.82%</td>
</tr>
<tr>
<td>22</td>
<td>0.70</td>
<td>0.82</td>
<td>16.93%</td>
<td>45</td>
<td>0.81</td>
<td>0.86</td>
<td>6.49%</td>
</tr>
<tr>
<td>23</td>
<td>0.89</td>
<td>0.89</td>
<td>0.53%</td>
<td>46</td>
<td>1.00</td>
<td>0.80</td>
<td>−20.19%</td>
</tr>
</tbody>
</table>
Appendix C

Scenario Analyses of Medical Records Services
Table C.1: Scenario analysis of Case 1: earthquake

<table>
<thead>
<tr>
<th>Events</th>
<th>Original FT</th>
<th>Revised FT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Administrators are absent</td>
<td>FALSE</td>
<td>FALSE</td>
</tr>
<tr>
<td>Alternate staff are absent</td>
<td>TRUE</td>
<td>TRUE</td>
</tr>
<tr>
<td>Alternate staff are not trained</td>
<td>FALSE</td>
<td>FALSE</td>
</tr>
<tr>
<td>Computer equipment is damaged</td>
<td>FALSE</td>
<td>FALSE</td>
</tr>
<tr>
<td>Municipal power fails</td>
<td>TRUE</td>
<td>TRUE</td>
</tr>
<tr>
<td>Backup power fails</td>
<td>FALSE</td>
<td>FALSE</td>
</tr>
<tr>
<td>Internal IT network fails</td>
<td>TRUE</td>
<td>TRUE</td>
</tr>
<tr>
<td>Internet access fails</td>
<td>TRUE</td>
<td>TRUE</td>
</tr>
<tr>
<td>Central server is hacked</td>
<td>N/A</td>
<td>FALSE</td>
</tr>
<tr>
<td>Central server is damaged</td>
<td>N/A</td>
<td>FALSE</td>
</tr>
<tr>
<td>Central server’s power fails</td>
<td>N/A</td>
<td>FALSE</td>
</tr>
<tr>
<td>Local server is hacked</td>
<td>N/A</td>
<td>FALSE</td>
</tr>
<tr>
<td>Local server is damaged</td>
<td>N/A</td>
<td>FALSE</td>
</tr>
<tr>
<td>Password is lost</td>
<td>N/A</td>
<td>FALSE</td>
</tr>
<tr>
<td>Severe structural damage</td>
<td>FALSE</td>
<td>FALSE</td>
</tr>
<tr>
<td>Severe non-structural damage</td>
<td>TRUE</td>
<td>TRUE</td>
</tr>
<tr>
<td>Corridors are severely damaged</td>
<td>FALSE</td>
<td>FALSE</td>
</tr>
<tr>
<td>Exterior exists are damaged, blocked, or non-existent</td>
<td>FALSE</td>
<td>FALSE</td>
</tr>
<tr>
<td>Elevators fail</td>
<td>TRUE</td>
<td>TRUE</td>
</tr>
<tr>
<td>Stairs fail</td>
<td>FALSE</td>
<td>FALSE</td>
</tr>
<tr>
<td>No space is on ground floor</td>
<td>FALSE</td>
<td>FALSE</td>
</tr>
<tr>
<td>Medical Records Services are disabled</td>
<td><strong>TRUE</strong></td>
<td>FALSE</td>
</tr>
</tbody>
</table>
## APPENDIX C. SCENARIO ANALYSES OF MEDICAL RECORDS SERVICES

Table C.2: Scenario analysis of Case 2: cyber attack against hospital computers

<table>
<thead>
<tr>
<th>Events</th>
<th>Original FT</th>
<th>Revised FT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Administrators are absent</td>
<td>FALSE</td>
<td>FALSE</td>
</tr>
<tr>
<td>Alternate staff are absent</td>
<td>FALSE</td>
<td>FALSE</td>
</tr>
<tr>
<td>Alternate staff are not trained</td>
<td>FALSE</td>
<td>FALSE</td>
</tr>
<tr>
<td>Computer equipment is damaged</td>
<td>FALSE</td>
<td>FALSE</td>
</tr>
<tr>
<td>Municipal power fails</td>
<td>FALSE</td>
<td>FALSE</td>
</tr>
<tr>
<td>Backup power fails</td>
<td>FALSE</td>
<td>FALSE</td>
</tr>
<tr>
<td>Internal IT network fails</td>
<td>TRUE</td>
<td>TRUE</td>
</tr>
<tr>
<td>Internet access fails</td>
<td>TRUE</td>
<td>TRUE</td>
</tr>
<tr>
<td>Central server is hacked</td>
<td>N/A</td>
<td>FALSE</td>
</tr>
<tr>
<td>Central server is damaged</td>
<td>N/A</td>
<td>FALSE</td>
</tr>
<tr>
<td>Central server’s power fails</td>
<td>N/A</td>
<td>FALSE</td>
</tr>
<tr>
<td>Local server is hacked</td>
<td>N/A</td>
<td>FALSE</td>
</tr>
<tr>
<td>Local server is damaged</td>
<td>N/A</td>
<td>FALSE</td>
</tr>
<tr>
<td>Password is lost</td>
<td>N/A</td>
<td>FALSE</td>
</tr>
<tr>
<td>Severe structural damage</td>
<td>FALSE</td>
<td>FALSE</td>
</tr>
<tr>
<td>Severe non-structural damage</td>
<td>FALSE</td>
<td>FALSE</td>
</tr>
<tr>
<td>Corridors are severely damaged</td>
<td>FALSE</td>
<td>FALSE</td>
</tr>
<tr>
<td>Exterior exists are damaged, blocked, or non-existent</td>
<td>FALSE</td>
<td>FALSE</td>
</tr>
<tr>
<td>Elevators fail</td>
<td>FALSE</td>
<td>FALSE</td>
</tr>
<tr>
<td>Stairs fail</td>
<td>FALSE</td>
<td>FALSE</td>
</tr>
<tr>
<td>No space is on ground floor</td>
<td>FALSE</td>
<td>FALSE</td>
</tr>
<tr>
<td>Medical Records Services are disabled</td>
<td>FALSE</td>
<td>FALSE</td>
</tr>
</tbody>
</table>
Table C.3: Scenario analysis of Case 3: cyber attack against central server

<table>
<thead>
<tr>
<th>Events</th>
<th>Original FT</th>
<th>Revised FT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Administrators are absent</td>
<td>FALSE</td>
<td>FALSE</td>
</tr>
<tr>
<td>Alternate staff are absent</td>
<td>FALSE</td>
<td>FALSE</td>
</tr>
<tr>
<td>Alternate staff are not trained</td>
<td>FALSE</td>
<td>FALSE</td>
</tr>
<tr>
<td>Computer equipment is damaged</td>
<td>FALSE</td>
<td>FALSE</td>
</tr>
<tr>
<td>Municipal power fails</td>
<td>FALSE</td>
<td>FALSE</td>
</tr>
<tr>
<td>Backup power fails</td>
<td>FALSE</td>
<td>FALSE</td>
</tr>
<tr>
<td>Internal IT network fails</td>
<td>FALSE</td>
<td>FALSE</td>
</tr>
<tr>
<td>Internet access fails</td>
<td>FALSE</td>
<td>FALSE</td>
</tr>
<tr>
<td>Central server is hacked</td>
<td>N/A</td>
<td>TRUE</td>
</tr>
<tr>
<td>Central server is damaged</td>
<td>N/A</td>
<td>FALSE</td>
</tr>
<tr>
<td>Central server’s power fails</td>
<td>N/A</td>
<td>FALSE</td>
</tr>
<tr>
<td>Local server is hacked</td>
<td>N/A</td>
<td>FALSE</td>
</tr>
<tr>
<td>Local server is damaged</td>
<td>N/A</td>
<td>FALSE</td>
</tr>
<tr>
<td>Password is lost</td>
<td>N/A</td>
<td>FALSE</td>
</tr>
<tr>
<td>Severe structural damage</td>
<td>FALSE</td>
<td>FALSE</td>
</tr>
<tr>
<td>Severe non-structural damage</td>
<td>FALSE</td>
<td>FALSE</td>
</tr>
<tr>
<td>Corridors are severely damaged</td>
<td>FALSE</td>
<td>FALSE</td>
</tr>
<tr>
<td>Exterior exists are damaged, blocked, or non-existent</td>
<td>FALSE</td>
<td>FALSE</td>
</tr>
<tr>
<td>Elevators fail</td>
<td>FALSE</td>
<td>FALSE</td>
</tr>
<tr>
<td>Stairs fail</td>
<td>FALSE</td>
<td>FALSE</td>
</tr>
<tr>
<td>No space is on ground floor</td>
<td>FALSE</td>
<td>FALSE</td>
</tr>
<tr>
<td>Medical Records Services are disabled</td>
<td>FALSE</td>
<td>FALSE</td>
</tr>
</tbody>
</table>
Table C.4: Scenario analysis of Case 4: cyber attack against local server

<table>
<thead>
<tr>
<th>Events</th>
<th>Original FT</th>
<th>Revised FT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Administrators are absent</td>
<td>FALSE</td>
<td>FALSE</td>
</tr>
<tr>
<td>Alternate staff are absent</td>
<td>TRUE</td>
<td>TRUE</td>
</tr>
<tr>
<td>Alternate staff are not trained</td>
<td>FALSE</td>
<td>FALSE</td>
</tr>
<tr>
<td>Computer equipment is damaged</td>
<td>FALSE</td>
<td>FALSE</td>
</tr>
<tr>
<td>Municipal power fails</td>
<td>FALSE</td>
<td>FALSE</td>
</tr>
<tr>
<td>Backup power fails</td>
<td>FALSE</td>
<td>FALSE</td>
</tr>
<tr>
<td>Internal IT network fails</td>
<td>FALSE</td>
<td>FALSE</td>
</tr>
<tr>
<td>Internet access fails</td>
<td>FALSE</td>
<td>FALSE</td>
</tr>
<tr>
<td>Central server is hacked</td>
<td>N/A</td>
<td>FALSE</td>
</tr>
<tr>
<td>Central server is damaged</td>
<td>N/A</td>
<td>FALSE</td>
</tr>
<tr>
<td>Central server’s power fails</td>
<td>N/A</td>
<td>FALSE</td>
</tr>
<tr>
<td>Local server is hacked</td>
<td>N/A</td>
<td>TRUE</td>
</tr>
<tr>
<td>Local server is damaged</td>
<td>N/A</td>
<td>FALSE</td>
</tr>
<tr>
<td>Password is lost</td>
<td>N/A</td>
<td>FALSE</td>
</tr>
<tr>
<td>Severe structural damage</td>
<td>FALSE</td>
<td>FALSE</td>
</tr>
<tr>
<td>Severe non-structural damage</td>
<td>FALSE</td>
<td>FALSE</td>
</tr>
<tr>
<td>Corridors are severely damaged</td>
<td>FALSE</td>
<td>FALSE</td>
</tr>
<tr>
<td>Exterior exists are damaged, blocked, or non-existent</td>
<td>FALSE</td>
<td>FALSE</td>
</tr>
<tr>
<td>Elevators fail</td>
<td>FALSE</td>
<td>FALSE</td>
</tr>
<tr>
<td>Stairs fail</td>
<td>FALSE</td>
<td>FALSE</td>
</tr>
<tr>
<td>No space is on ground floor</td>
<td>FALSE</td>
<td>FALSE</td>
</tr>
<tr>
<td>Medical Records Services are disabled</td>
<td>FALSE</td>
<td>FALSE</td>
</tr>
</tbody>
</table>
### APPENDIX C. SCENARIO ANALYSES OF MEDICAL RECORDS SERVICES

Table C.5: Scenario analysis of Case 5: earthquake coupled with cyber attack against hospital computers

<table>
<thead>
<tr>
<th>Events</th>
<th>Original FT</th>
<th>Revised FT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Administrators are absent</td>
<td>FALSE</td>
<td>FALSE</td>
</tr>
<tr>
<td>Alternate staff are absent</td>
<td>TRUE</td>
<td>TRUE</td>
</tr>
<tr>
<td>Alternate staff are not trained</td>
<td>FALSE</td>
<td>FALSE</td>
</tr>
<tr>
<td>Computer equipment is damaged</td>
<td>FALSE</td>
<td>FALSE</td>
</tr>
<tr>
<td>Municipal power fails</td>
<td>TRUE</td>
<td>TRUE</td>
</tr>
<tr>
<td>Backup power fails</td>
<td>FALSE</td>
<td>FALSE</td>
</tr>
<tr>
<td>Internal IT network fails</td>
<td>TRUE</td>
<td>TRUE</td>
</tr>
<tr>
<td>Internet access fails</td>
<td>TRUE</td>
<td>TRUE</td>
</tr>
<tr>
<td>Central server is hacked</td>
<td>N/A</td>
<td>FALSE</td>
</tr>
<tr>
<td>Central server is damaged</td>
<td>N/A</td>
<td>FALSE</td>
</tr>
<tr>
<td>Central server’s power fails</td>
<td>N/A</td>
<td>FALSE</td>
</tr>
<tr>
<td>Local server is hacked</td>
<td>N/A</td>
<td>FALSE</td>
</tr>
<tr>
<td>Local server is damaged</td>
<td>N/A</td>
<td>FALSE</td>
</tr>
<tr>
<td>Password is lost</td>
<td>N/A</td>
<td>FALSE</td>
</tr>
<tr>
<td>Severe structural damage</td>
<td>FALSE</td>
<td>FALSE</td>
</tr>
<tr>
<td>Severe non-structural damage</td>
<td>TRUE</td>
<td>TRUE</td>
</tr>
<tr>
<td>Corridors are severely damaged</td>
<td>FALSE</td>
<td>FALSE</td>
</tr>
<tr>
<td>Exterior exists are damaged, blocked, or non-existent</td>
<td>FALSE</td>
<td>FALSE</td>
</tr>
<tr>
<td>Elevators fail</td>
<td>TRUE</td>
<td>TRUE</td>
</tr>
<tr>
<td>Stairs fail</td>
<td>FALSE</td>
<td>FALSE</td>
</tr>
<tr>
<td>No space is on ground floor</td>
<td>FALSE</td>
<td>FALSE</td>
</tr>
<tr>
<td>Medical Records Services are disabled</td>
<td>TRUE</td>
<td>FALSE</td>
</tr>
</tbody>
</table>
Table C.6: Scenario analysis of Case 6: earthquake coupled with cyber attack against central server

<table>
<thead>
<tr>
<th>Events</th>
<th>Original FT</th>
<th>Revised FT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Administrators are absent</td>
<td>FALSE</td>
<td>FALSE</td>
</tr>
<tr>
<td>Alternate staff are absent</td>
<td>TRUE</td>
<td>TRUE</td>
</tr>
<tr>
<td>Alternate staff are not trained</td>
<td>FALSE</td>
<td>FALSE</td>
</tr>
<tr>
<td>Computer equipment is damaged</td>
<td>FALSE</td>
<td>FALSE</td>
</tr>
<tr>
<td>Municipal power fails</td>
<td>TRUE</td>
<td>TRUE</td>
</tr>
<tr>
<td>Backup power fails</td>
<td>FALSE</td>
<td>FALSE</td>
</tr>
<tr>
<td>Internal IT network fails</td>
<td>TRUE</td>
<td>TRUE</td>
</tr>
<tr>
<td>Internet access fails</td>
<td>TRUE</td>
<td>TRUE</td>
</tr>
<tr>
<td>Central server is hacked</td>
<td>N/A</td>
<td>TRUE</td>
</tr>
<tr>
<td>Central server is damaged</td>
<td>N/A</td>
<td>FALSE</td>
</tr>
<tr>
<td>Central server’s power fails</td>
<td>N/A</td>
<td>FALSE</td>
</tr>
<tr>
<td>Local server is hacked</td>
<td>N/A</td>
<td>FALSE</td>
</tr>
<tr>
<td>Local server is damaged</td>
<td>N/A</td>
<td>FALSE</td>
</tr>
<tr>
<td>Password is lost</td>
<td>N/A</td>
<td>FALSE</td>
</tr>
<tr>
<td>Severe structural damage</td>
<td>FALSE</td>
<td>FALSE</td>
</tr>
<tr>
<td>Severe non-structural damage</td>
<td>TRUE</td>
<td>TRUE</td>
</tr>
<tr>
<td>Corridors are severely damaged</td>
<td>FALSE</td>
<td>FALSE</td>
</tr>
<tr>
<td>Exterior exists are damaged, blocked, or non-existent</td>
<td>FALSE</td>
<td>FALSE</td>
</tr>
<tr>
<td>Elevators fail</td>
<td>TRUE</td>
<td>TRUE</td>
</tr>
<tr>
<td>Stairs fail</td>
<td>FALSE</td>
<td>FALSE</td>
</tr>
<tr>
<td>No space is on ground floor</td>
<td>FALSE</td>
<td>FALSE</td>
</tr>
<tr>
<td>Medical Records Services are disabled</td>
<td>TRUE</td>
<td>FALSE</td>
</tr>
</tbody>
</table>
Table C.7: Scenario analysis of Case 7: earthquake coupled with cyber attack against local server

<table>
<thead>
<tr>
<th>Events</th>
<th>Original FT</th>
<th>Revised FT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Administrators are absent</td>
<td>FALSE</td>
<td>FALSE</td>
</tr>
<tr>
<td>Alternate staff are absent</td>
<td>TRUE</td>
<td>TRUE</td>
</tr>
<tr>
<td>Alternate staff are not trained</td>
<td>FALSE</td>
<td>FALSE</td>
</tr>
<tr>
<td>Computer equipment is damaged</td>
<td>FALSE</td>
<td>FALSE</td>
</tr>
<tr>
<td>Municipal power fails</td>
<td>TRUE</td>
<td>TRUE</td>
</tr>
<tr>
<td>Backup power fails</td>
<td>FALSE</td>
<td>FALSE</td>
</tr>
<tr>
<td>Internal IT network fails</td>
<td>TRUE</td>
<td>TRUE</td>
</tr>
<tr>
<td>Internet access fails</td>
<td>TRUE</td>
<td>TRUE</td>
</tr>
<tr>
<td>Central server is hacked</td>
<td>N/A</td>
<td>FALSE</td>
</tr>
<tr>
<td>Central server is damaged</td>
<td>N/A</td>
<td>FALSE</td>
</tr>
<tr>
<td>Central server’s power fails</td>
<td>N/A</td>
<td>FALSE</td>
</tr>
<tr>
<td>Local server is hacked</td>
<td>N/A</td>
<td>TRUE</td>
</tr>
<tr>
<td>Local server is damaged</td>
<td>N/A</td>
<td>FALSE</td>
</tr>
<tr>
<td>Password is lost</td>
<td>N/A</td>
<td>FALSE</td>
</tr>
<tr>
<td>Severe structural damage</td>
<td>FALSE</td>
<td>FALSE</td>
</tr>
<tr>
<td>Severe non-structural damage</td>
<td>TRUE</td>
<td>TRUE</td>
</tr>
<tr>
<td>Corridors are severely damaged</td>
<td>FALSE</td>
<td>FALSE</td>
</tr>
<tr>
<td>Exterior exists are damaged, blocked, or non-existent</td>
<td>FALSE</td>
<td>FALSE</td>
</tr>
<tr>
<td>Elevators fail</td>
<td>TRUE</td>
<td>TRUE</td>
</tr>
<tr>
<td>Stairs fail</td>
<td>FALSE</td>
<td>FALSE</td>
</tr>
<tr>
<td>No space is on ground floor</td>
<td>FALSE</td>
<td>FALSE</td>
</tr>
<tr>
<td>Medical Records Services are disabled</td>
<td>TRUE</td>
<td>TRUE</td>
</tr>
</tbody>
</table>
Bibliography


BIBLIOGRAPHY


BIBLIOGRAPHY


[Doménech et al., 2010] Doménech, E., Escriche, I., and Martorell, S. (2010). Ex-
BIBLIOGRAPHY


BIBLIOGRAPHY


queuing theory approach. In The 91st Annual Meeting of the Transportation Research Board, Washington D.C.


BIBLIOGRAPHY


BIBLIOGRAPHY


BIBLIOGRAPHY


BIBLIOGRAPHY


BIBLIOGRAPHY


BIBLIOGRAPHY


Xilei Zhao was born on November 11, 1990 in Nanjing, China. She attended Nanjing Jinling High School from 2006 to 2009. Then, she received her B.E. degree in Civil Engineering from Southeast University, China, in 2013, and then enrolled in the Civil Engineering Ph.D. program at the Johns Hopkins University in the same year. In pursuit of her Ph.D. in Civil Engineering, she also earned two Master’s degrees: one in Civil Engineering, the other in Applied Mathematics and Statistics. Her research focuses on community resilience and critical infrastructure modeling.

Starting in July 2017, Xilei will be working with Prof. Pascal Van Hentenryck as a Postdoctoral Research Fellow on the Reinventing Urban Transportation and Mobility (RITMO) project in Industrial and Operations Engineering at the University of Michigan, Ann Arbor.