Title
Seismic Risk and Resilience Modeling of Water Distribution Systems

Author
Tomar, Agam

Publication Date
2020

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

Seismic Risk and Resilience Modeling of Water Distribution Systems

A dissertation submitted in partial satisfaction

of the requirements for the degree

Doctor of Philosophy in Civil Engineering

by

Agam Tomar

2020

ABSTRACT OF THE DISSERTATION

Seismic Risk and Resilience Modeling of Water Distribution Systems

by

Agam Tomar

Doctor of Philosophy in Civil Engineering

University of California, Los Angeles, 2020

Professor Henry V. Burton, Co-Chair

Professor Ali Mosleh, Co-Chair

Water distribution systems are vital to the well-being of communities because they contribute to the functionality of all other infrastructure and lifeline systems. Earthquakes and other natural hazards can cause damage to the components of a water distribution system, causing far-reaching socioeconomic consequences. This research begins with the development of an end-to-end simulation framework to model post-earthquake functional loss and restoration of a water system, which encompasses seismic hazard characterization, component damage assessment, hydraulic performance evaluation, and network restoration modeling. The modeling framework is validated using data from the 2014 South Napa Earthquake and extended to a hypothetical scenario. The end-to-end simulation framework is then extended to consider stochastic event set assessments of the water network using the UCERF2 (Uniform California Earthquake Rupture Forecast, Version 2) earthquake rupture forecast model. Given that the end-to-end performance evaluation of distributed infrastructure for a large set of events is computationally expensive, a framework that uses Active learning to select a subset of ground motion maps and associated occurrence rates that reasonably estimates the water

network risk is also developed. To deal with the temporal complexities that are embedded in the post-earthquake restoration process, a dynamic updating methodology is developed to reduce uncertainties in the outcomes of post-event recovery forecasts using Bayesian Inferencing, by exploiting real-time data. The specific example of updating predictions (post-earthquake functional recovery forecasts including total recovery time and complete recovery trajectory) is presented and validated on a real pipe network (Napa water system) and event (2014 earthquake and recovery). Ultimately, the frameworks and models developed as part of this work can inform risk-based decision making and resilience planning of water networks and other lifeline systems.

The dissertation of Agam Tomar is approved.

Scott Brandenberg

Jonathan Kao

Ali Mosleh, Committee Co-Chair

Henry V. Burton, Committee Co-Chair

University of California, Los Angeles

2020

To my family …

For their endless love, support and encouragement

Table of Contents

CHAPTER 1 INTRODUCTION 1

1.1 MOTIVATION AND BACKGROUND 1

1.2 OBJECTIVES 3

1.3 ORGANIZATION AND OUTLINE 4

CHAPTER 2 HINDCASTING THE FUNCTIONAL LOSS AND RESTORATION OF THE NAPA WATER SYSTEM FOLLOWING THE 2014 EARTHQUAKE USING DISCRETE EVENT SIMULATION 6

2.1 INTRODUCTION 6

2.2 DESCRIPTION OF NAPA WATER SYSTEM AND IMPACTS FROM 2014 EARTHQUAKE 8

2.2.1 Description of Napa Water Network and Services 8

2.2.2 Impacts from 2014 Earthquake 10

2.3 DISCRETE EVENT SIMULATION MODEL FOR NAPA WATER SYSTEM 13

2.3.1 Overview of Modeling Framework 13

2.4 REPLICATING THE RECOVERY FOLLOWING 2014 EARTHQUAKE 16

2.5 DEVELOPMENT OF PIPE DAMAGE FRAGILITY FUNCTIONS USING DATA FROM THE 2014 SOUTH NAPA EARTHQUAKE 19

2.6 SCENARIO-BASED NETWORK DAMAGE AND WATER SERVICE RESTORATION ASSESSMENT 21

2.6.1 Scenario Description 21

2.6.2 Ground Motion Intensities and Network Damage 22

2.6.3 Modeling the Inspection and Repair Following the M 6.7 Scenario 24

2.6.4 Modeling Water Service Restoration Following the M 6.7 Scenario 28

2.7 SUMMARY AND CONCLUSION 32

CHAPTER 3 RISK-BASED ASSESSMENT OF THE POST-EARTHQUAKE FUNCTIONAL DISRUPTION AND RESTORATION OF A WATER DISTRIBUTION SYSTEM 34

3.1 INTRODUCTION 34

3.2 OVERVIEW OF RISK-BASED ASSESSMENT FRAMEWORK 36

3.2.1 Hazard Characterization 37

3.2.2 Damage modeling 38

3.2.3 Restoration Modeling 38

3.2.4 Network Performance 40

3.3 CASE STUDY 41

3.3.1 Description of the Napa Water Network 41
3.3.2 Generating the Earthquake Stochastic Event Set Catalogue for the Napa Region 41
3.3.3 Damage modeling 42
3.3.4 Restoration Modeling and Network Performance 44
3.4 Disaggregation of Dispersion by Source of Uncertainty 47
3.5 Annual Exceedance Rates 51
3.6 Ground Motion Intensity and Damage Map Selection using Multi-Metric Optimization 53
3.7 Summary and Conclusion 55
CHAPTER 4 ACTIVE LEARNING FRAMEWORK FOR RISK ASSESSMENT OF DISTRIBUTED INFRASTRUCTURE SYSTEMS 58
4.1 Introduction 58
4.2 Overview 60
4.3 Gaussian Process 61
4.3.1 Overview 61
4.3.2 Weight-space view 62
4.3.3 Function-space view 64
4.3.4 Advantages and disadvantages of $\mathcal{GP}$ 68
4.4 Active Learning 69
4.4.1 Query Approaches in Active learning 69
4.4.2 Evaluating the informativeness of queries via an acquisition function 70
4.4.3 Stopping Criterion 73
4.5 Case Study 73
4.5.1 Napa water network 73
4.5.2 Risk assessment of the Napa water network using an end-to-end simulation framework 74
4.5.3 Map selection algorithm using active learning 75
4.6 Results and Discussion 77
4.7 Summary and Conclusion 83
CHAPTER 5 DYNAMIC UPDATING OF POST-EARTHQUAKE DAMAGE AND FUNCTIONAL RESTORATION FORECASTS OF WATER DISTRIBUTION SYSTEMS USING BAYESIAN INFERENCING 85
5.1 Introduction 85
5.2 Overview of the Proposed Methodology 88
5.3 Description of the Water Network 91
5.4 Modeling Post-Earthquake Pipe Repairs using Discrete Event Simulation 91

5.5 PROBABILISTIC GRAPHICAL MODELS (BAYESIAN NETWORKS) 92
5.5.1 Background 92
5.5.2 Bayesian Networks 93
5.6 BAYESIAN INFERENCING 95
5.6.1 Background 95
5.6.2 Application of Bayesian Parameter Estimation to Update PBDES Event Durations 97
5.7 BAYESIAN NETWORKS FOR COMPUTING THE TOTAL RECOVERY TIME FOR THE NAPA WATER PIPE SYSTEM 97
5.8 DYNAMIC UPDATING OF THE POST-EARTHQUAKE TOTAL RECOVERY TIME AND TRAJECTORY 100
5.9 SUMMARY AND CONCLUSION 104
CHAPTER 6 CONCLUSION, LIMITATIONS AND FUTURE WORK 106
6.1 OVERVIEW 106
6.2 FINDINGS 106
6.2.1 Chapter 2: Hindcasting the Functional Loss and Restoration of the Napa Water System Following the 2014 Earthquake using Discrete Event Simulation 106
6.2.2 Chapter 3: Risk-Based Assessment of the Post-Earthquake Functional Disruption and Restoration of a Water Distribution System 107
6.2.3 Chapter 4: Active Learning Framework for Risk Assessment of Distributed Infrastructure Systems 108
6.2.4 Chapter 5: Dynamic Updating of Post-Earthquake Damage and Functional Restoration Forecasts of Water Distribution Systems using Bayesian Inferencing 109
6.3 LIMITATIONS AND FUTURE WORK 110
REFERENCES 112

List of Tables

Table 2-1 Inspection and repair crew reduction during the first 30 days after the 2014 event 18

Table 2-2 Event Duration Triangular Distribution.....24

Table 2-3 Event duration mean and standard deviation (σ).....25

Table 3-1 Mean (μ) and coefficient of variation (δ) of the network performance metrics for a **M** 6.85 earthquake scenario with all sources of uncertainty considered 48

Table 3-2 Mean (μ) and coefficient of variation (δ) values for the time to restore 25%, 50%, 75%, and 100% of the pre-event SSI levels for the **M** 6.85 earthquake scenario 50

Table 3-3 Mean performance measure curve error (MPMCE) for various target performance measures 55

Table 5-1 Pipe repair duration mean and standard deviation (STD) [55].....97

List of Figures

Figure 2-1 Schematic illustration of the interaction between the water network components, earthquake hazard, post-event restoration, and hydraulic simulation 8

Figure 2-2 City of Napa Water Service Area with 2014 Earthquake Epicenter and Napa Fault System 10

Figure 2-3 (a) peak ground velocity (PGV in cm/s) and (b) pipes damaged (leaks) during the 2014 South Napa Earthquake 12

Figure 2-4 Percentage of total damaged pipes based on material 12

Figure 2-5 Schematic representation of DES modeling framework 15

Figure 2-6 Schematic illustration of the timing and occurrence of a set of discrete events (and their associated processes) for a hypothetical two-pipe water distribution system 16

Figure 2-7 Comparing the simulated and observed progression of repairs following the 2014 event 18

Figure 2-8 Mean repair rates (per km) based on pipe material 20

Figure 2-9 Empirical repair rate (RR) model developed using data from the 2014 event 21

Figure 2-10 Spatial distribution of (a) PGVs and (b) pipe damage for **M** 6.7 scenario 23

Figure 2-11 Simulated progression of pipe repairs following the **M** 6.7 event 26

Figure 2-12 Probability density function of time to repair 90% of damaged pipes 27

Figure 2-13 Cumulative distribution function (cdf) of time to repair 90% of damaged pipes 27

Figure 2-14 Time series of system serviceability index following the **M** 6.7 event 31

Figure 2-15 Empirical cumulative distribution function (CDF) for time to repair 90% of damaged pipes and time to achieve 90% SSI 31

Figure 3-1 End-to-end simulation framework for performing risk-based assessment of post-earthquake functional impacts of a water distribution system 37

Figure 3-2 Spatial distribution of median (a) PGA (in g) and (b) PGV (in cm/s) values considering all 117,150 ground motion maps 42

Figure 3-3 Magnitude distribution for damaging scenarios 43

Figure 3-4 Spatial distribution of median (a) PGA (in g) and (b) PGV (in cm/s) values considering the 135 scenarios that caused damage 44

Figure 3-5 Mean pipe repair rate per km distribution 44

Figure 3-6 Histogram showing the distribution of SSI_{IL} for the damaging events 46

Figure 3-7 Histogram showing the distribution of $T_{100\%SSI}$ for the damaging events.....46
Figure 3-8 Histogram showing the distribution of SSI_{CL} for the damaging events 47
Figure 3-9 Sources of uncertainty considered in the risk-based assessment framework ..48
Figure 3-10 Mean and one-standard deviation bounded SSI curve for the **M** 6.85 earthquake scenario with all sources of uncertainty considered 49
Figure 3-11 Mean and one-standard deviation bounded SSI curve for the **M** 6.85 earthquake scenario while considering the uncertainty in the (a) ground shaking intensity, (b) component damage fragility, (c) crew sizes and (d) event duration parameters 50
Figure 3-12 Annual exceedance curve for SSI_{IL} 52
Figure 3-13 Annual exceedance curve for $T_{100\%SSI}$ 52
Figure 3-14 Annual exceedance curve for SSI_{CL} 53
Figure 4-1 Active learning framework 61
Figure 4-2 Query approaches in active learning 70
Figure 4-3 Generic pool-based uncertainty sampling algorithm 72
Figure 4-4 Distribution of time to restore 100% pre-earthquake SSI 75
Figure 4-5 Distribution of cumulative loss in SSI 75
Figure 4-6 Map selection algorithm using active learning 77
Figure 4-7 Histogram of empirical and predicted $T_{100\%SSI}$ distribution 78
Figure 4-8 Histogram of empirical and predicted SSI_{CL} distribution 79
Figure 4-9 $T_{100\%SSI}$ annual exceedance curve for the empirical and selected subset of ground motion maps 80
Figure 4-10 SSI_{CL} annual exceedance curve for the empirical and selected subset of ground motion maps 80
Figure 4-11 $T_{100\%SSI}$ annual exceedance curve for the empirical and selected subset of ground motion maps in the MCS 82
Figure 4-12 Time series of the stopping criterion (SC_{SU}) for every case of MCS with δ_{SU} 82
Figure 5-1 Overview of the proposed methodology 90
Figure 5-2 Sources of uncertainty in post-earthquake functional restoration forecast.....91
Figure 5-3 Bayesian network for a simple “pipe repair time” problem 94
Figure 5-4 Schematic illustration of the Bayesian updating framework for pipe repair duration 96

Figure 5-5 Schematic illustration of the Bayesian network embedded in pipe recovery plate model.....99
Figure 5-6 Bayesian network for the entire Napa water pipe system 100
Figure 5-7 Predicted total pipe recovery time distribution at different time-points after the earthquake.....102
Figure 5-8 Comparing the actual pipe repair trajectory with dynamic predictions generated using the RT-DP model 103
Figure 5-9 Residual sum of squares between predicted and actual pipe repair curves versus the number of days following the earthquake.....104

ACKNOWLEDGEMENTS

First and foremost, I would like to thank my Ph.D. advisor, Dr. Henry J. Burton for his wisdom, patience, encouragement, and continuous support. Week by week he has cultivated me as a researcher; without him, this dissertation would not exist. Besides my advisor, I would like to express my sincere gratitude to Dr. Jonathan Chau-Yan Kao for his lectures on deep learning and neural signal processing which motivated me to pursue courses across departments, particularly in Electrical and Computer Engineering, and Statistics. Also, I would like to thank Dr. Ali Mosleh, Dr. Scott Joseph Brandenberg, and Dr. Ji Yun Lee for their encouragement, insightful comments, and technical guidance.

Outside of my committee, I am grateful to Joy Eldredge and Douglas DeMaster from Napa Water Division (NWD) for providing most of the case study data for this dissertation, facilitating my trip to Napa and showing continuous support for my research progress. I am thankful to Katherine Klise (Sandia National Laboratories) for providing help and support regarding incorporating the hydraulic simulator, Water Network Tool for Resilience (WNTR), to my research. I am also thankful to Tomasz Sowinski (Bayes Fusion) for providing me python binaries of Structural Modeling, Inference and Learning Engine (SMILE), and guidance on my work on Bayesian inferencing.

My sincere thank also goes to Dr. Mehrdad Shokrabadi for offering help and guidance in every stage of my Ph.D. work. I am thankful to my friends at UCLA: Kheder Alrazza, Denrick Bautista, Srinivas Prasad, Vijay Ravi, Xingquan Guan, Zhengxiang Yi, Peng Yu Chen, Bakari Hassan for the stimulating discussions, spontaneous road trips, and for all the fun we had in the last four years. Also, I want to thank Prachi Singh, who has provided never-ending guidance and moral support during my college studies.

Funding for this research is provided by the Nuclear Regulatory Commission (NRC) Research Grant No. NRC-HQ-60-17-G-0028. Any opinions, findings, and conclusions expressed in this dissertation are those of the authors and do not necessarily reflect the views of the NRC.

Last but not the least, I would like to thank my family: my parents Mr. Mahesh Chandra Tomar and Mrs. Raj Kumari Tomar, and my younger brother Dr. Shubham Tomar for their love, kindness and unconditional support that shaped me to who I am today.

VITA

2011-2016	Integrated Dual Degree (B. Tech. in Civil Engineering and M. Tech. in Structural Engineering, Indian Institute of Technology (Banaras Hindu University), Varanasi, India
2019-2020	M.Sc. in Electrical and Computer Engineering, University of California Los Angeles
2016-2020	Ph.D. Candidate in Structural and Earthquake Engineering, University of California Los Angeles

CHAPTER 1 Introduction

1.1 Motivation and Background

Critical infrastructure systems, such as electric power, water, gas and transportation networks (so-called lifeline networks) play a vital role in modern society by supporting all sectors of the economy. A common feature of lifeline systems is that often comprised of many physically and functionally interdependent components that are distributed over large geographic areas. As such, damage to one or a few components can have cascading effects on the functionality of the system as a whole. For example, the widespread blackout in India in 2012 interrupted train and retail operation, crippled potable water delivery, trapped 200 miners, and impacted health care services [1]. Along with the cascading effect, prolonged disruption of networks is considered one of the primary factors for socio-economic losses in disasters [2–4]. In light of this, modeling is one critical research advancement necessary for understanding and quantifying complex processes driving post-disaster recovery and a myriad of influences on the decision and outcome trajectories over time.

Natural hazards are one of the primary threats to distributed infrastructure systems. Some hazards such as tornados, which tend to have a relatively small spatial footprint, can damage one or a few components that may or may not be critical to the system-level performance. On the other hand, hazards that have regional-scale impacts (e.g. earthquakes) can damage large numbers of components and are therefore more likely to significantly disrupt the functionality of the system. Regardless of the type of event, risk assessment is a key part of the plans to mitigate the disruptive impacts of natural hazards on infrastructure systems.

Risk assessment studies aim to answer three core questions: (i) what can go wrong?, (ii) what is the likelihood of such a disruptive scenario?, and (iii) what are the consequences of such a scenario? [5]. Risk management strategies have traditionally focused on reducing the likelihood of disruptive events and reducing the potential consequences of the event, as well as the synthesis of both. Broadly, these strategies emphasize on two mitigation options: (i) prevention: designing systems to avoid or withstand disrupting events, and (ii) protection: designing systems to detect the disruptive event early enough to execute an appropriate response. In the case of natural hazards

like earthquakes, detection systems following protection strategy are not efficient enough to execute appropriate response as they provide few to tens of seconds warning. The potential use of these warning systems at personal and institutional levels is discussed in Kanamori (2005). The advent of recent large-scale events, Hurricane Sandy [7], Hurricane Isabel [8], 2011 9.0 magnitude earthquake and Tsunami in Japan [9] have shown the loss of life and global disruption to lifeline systems. Department of Homeland Security, along with other institutions, has placed emphasis on resilience through preparedness, response, and recovery [10,11].

The word resilience is derived from the Latin word "resiliere", which means to "bounce back". It is defined by the National Academies of Science as "the ability to prepare and plan for, absorb, recovery from, and successfully adapt to adverse events" [12]. Readers are referred to Hosseini et al. (2016) for definitions of resilience across multiple disciplines. Hosseini et al. (2016) in his review paper on resilience metrics, identified four domains of resilience: (i) organizational: organizational resilience is the ability of an organization to absorb strain and improve functioning despite the presence of adversity [14], (ii) social: social resilience is comprised of coping capacities, adaptive capacities and transformative capacities [15], (iii) economic: economic resilience is the inherent ability and adaptive response that enables firms and regions to avoid maximum potential losses [3], (iv) engineering: engineering resilience is defined as the sum of the passive survival rate (reliability) and proactive survival rate (restoration) [16].

For engineering systems, quantitative resilience assessment approaches are of more interest than qualitative resilience assessment approaches. General resilience measures provide a quantitative approach to assess resilience of a system by measuring its performance. These general measures are broadly characterized as deterministic and stochastic, each of which have been used to describe static and dynamic system behavior [13]. For road networks, two common performance metrics are connectivity [17,18] and flow capacity [19]. For power networks, widely used performance measures are connectivity [20], serviceability ratio [21], power system flow [22], and recovery time of electrical network [23]. For water networks, hydraulic power capacity [24] which is an indicator of system performance, minimum surplus head [25] which is an indicator of system reliability, resilience index [26], modified resilience index [27], connectivity [28,29], flow capacity [21], entropy-based measures, nodal-demand, and the total number of component failures system-wide [30–32].

This dissertation focuses on the seismic risk assessment of water distribution networks. In the last three decades, various models have been developed for characterizing and addressing the uncertainties in the seismic performance of water distribution networks. A widely used methodology is to utilize topological models of these networks in which nodes and links of the network are assigned a probability of failure conditioned on the ground motion intensity [21,28,33–37]. Failed nodes and links are removed from the network topology, and performance is assessed according to the connectivity between source and sink nodes. Another methodology is to use physics-based models which simulate the distribution of resources through the network according to the physical attributes of network components as well as the governing physical equations [38–42]. Network performance in these models is measured through the satisfaction of water demand at consumer modes in the network. Other models include the application of machine learning methods to predict network performance, which aims to be more aligned to physics-based models while reducing computational expense [43,44].

1.2 Objectives

The objective of this study is to develop a probabilistic seismic risk analysis framework for water distribution systems that combines robust assessment of functionality with characterization of pre- and post-earthquake decisions and actions, socioeconomic vulnerability, and temporal processes. More specifically, the main objectives of the current study are outlined below:

1. Create an end-to-end (hazard--damage--recovery--pressure-driven hydraulics) simulation platform for modeling the post-earthquake functional loss and recovery of water systems
2. Develop pipe damage fragility functions using pipe damage data from the 2014 South Napa Earthquake
3. Validate the post-earthquake restoration framework using pipe break and repair data from the 2014 South Napa Earthquake
4. Formulate a methodology for conducting stochastic event-set based functional loss and restoration of water systems

5. Develop a framework for subset selection of hazard scenarios for infrastructure risk assessment
6. Establish a framework for real-time updating of post-earthquake damage and restoration forecasts using Bayesian networks

1.3 Organization and Outline

The main body of the current study consists of five chapters. Most chapters are adopted from a research paper which is cited at the beginning of the chapter.

Chapter 2 focuses on models for simulating post-earthquake functional loss and restoration of water distribution systems supporting the development and implementation of more informed resilience-enhancing strategies. The Discrete Event Simulation (DES) model along with pipe damage and repair data is used to hindcast the water system damage and functional loss, and restoration following the 2014 South Napa earthquake. By propagating the uncertainties in the damage to the network components (e.g., pipe) conditioned on the shaking intensity and the temporal (e.g., repair time) and resource-related (e.g., the number of inspection crews) variables, probabilistic descriptions of network component repair and water service restoration are generated.

Event-based methods have been used to assess the risk of infrastructure systems by modeling the hazard, corresponding damage, and, in some cases, functional restoration. Chapter 3 discusses a stochastic event-set-based framework for simulating the post-earthquake functional loss and restoration of water distribution systems. The hazard-characterization stage of the methodology produces a set of maps with randomized realizations of the joint distribution of shaking intensity at locations coinciding with the network components. By coupling these ground motion intensities with component fragility functions, damage maps reflecting the physical impact to the pipe segments, pump stations, and tanks in the network, are produced. The post-earthquake functional restoration corresponding to each damage map is assessed by combining a process-based discrete event simulation (PBDES) model with pressure-driven hydraulic analysis. The outcome of this stage is the full recovery trajectory of the system measured in terms of the system serviceability index (SSI), which is used to quantify network performance in terms of rapidity and the immediate and cumulative loss of functionality. Dispersion disaggregation is used to isolate the individual

contributions of multiple random variables to the uncertainty bounds in the restoration trajectory. The framework is applied to the City of Napa's water distribution system to assess its performance under a wide range of possible events. The proposed framework can be used as a resilience-based risk-informed decision-making tool to provide much-needed information to stakeholders and decision-makers.

Stochastic event-based methods consider all hazard scenarios that could adversely impact the infrastructure. However, in many cases, such a comprehensive consideration of the possible hazard events requires high computational effort. Chapter 4 discusses an active learning framework to select a subset of hazard scenarios for infrastructure risk assessment. Uncertainty sampling is used to sample hazard scenarios that reasonably estimate the system risk, in turn reducing the computation expense. Active learning enables the efficient training of a machine learning algorithm by choosing the data from which it learns. The active learning framework is illustrated with a case study of the Napa water distribution system by sequentially selecting a subset of earthquake scenarios and using a variance reduction stopping criterion. The full probability distribution and annual exceedance curves of the network performance metrics are shown to be reasonably estimated.

Motivated by the need to reduce the inherent uncertainties in the outcomes of DES model, Chapter 5 discusses a framework for dynamically updating post-earthquake functional recovery forecasts using Bayesian inferencing. The framework comprises of two models: (i) a Bayesian Network (BN) model which is used to provide estimates of the total recovery time and (ii) a process-based discrete event simulation (PBDES) model which is used to generate forecasts of the complete recovery trajectory. Both models rely on component damage and temporal input parameters that are dynamically updated using Bayes theorem, as information becomes available throughout the recovery process. The effectiveness of the proposed framework is demonstrated through an application to the pipe network of the City of Napa water distribution system. More specifically, pipe damage and repair data from the 2014 earthquake are used as a point of comparison for the dynamic forecasts.

Chapter 6 summarizes the findings of the previous chapters and discusses the limitations of the current study and opportunities to improve the methodologies and frameworks presented in the previous chapters.

CHAPTER 2 Hindcasting the Functional Loss and Restoration of the Napa Water System Following the 2014 Earthquake using Discrete Event Simulation

---

This chapter is adopted from the following study:

Tomar, A., Burton, H. V., Mosleh, A., and Lee, J. Y. (2020). "Hindcasting the Functional Loss and Restoration of the Napa Water System Following the 2014 Earthquake using Discrete Event Simulation," *ASCE Journal of Infrastructure Systems.* 10.1061/(ASCE)IS.1943-555X.0000574

2.1 Introduction

Lifeline infrastructure systems consist of a complex network of civil, structural, geotechnical, mechanical and electrical components. One type of lifeline, water systems, is especially vital to the well-being of communities since they are essential to public health and contribute to the functionality of all building clusters as well as other lifeline systems (e.g., energy and transportation systems). Earthquakes and other natural hazard events can cause damage to individual components of a water system leading to loss of service. The ensuing disruption has far-reaching consequences to the socioeconomic and overall well-being of a community. Prior studies on the economic losses caused by water service disruption suggest that the duration of functionality loss is just as critical as the geographical extent and severity of damage [45–50]. Therefore, models that capture the spatial distribution of component-level physical damage, and loss and restoration of functionality, are needed to develop effective risk mitigation strategies and emergency management plans for water systems.

The earliest efforts to model damage, functionality disruption and restoration of water systems used analytical representations of earthquake-induced impacts [51]. These "resource-constrained" models considered the fragility of physical components in the damage assessment and the effect of the number, efficiency, and schedule of workers on service restoration. The primary model output was the number of customers without water service as a function of time. The resource-constrained modeling approach was later extended to include the prioritization of component repair sequences in earthquake-impacted water systems [52]. Stochastic process [53]

and network models [54] have also been used to assess post-earthquake restoration of lifeline systems.

Due to the increasing number of open source software tools and programming libraries, discrete event simulation (DES) has been used in several recent studies to model post-earthquake restoration of water systems. DES provides a customizable modular framework for representing multiple precedent-driven processes (e.g., component inspection, material acquisition, component repairs), resource constraints (e.g., the number of available inspectors), and conditional event triggering. The first DES model of a water system was developed by [55]. The model was used to simulate the post-earthquake restoration of the Los Angeles Department of Water and Power (LADWP) water distribution system. A calibration was performed using damage and restoration data from the 1994 Northridge earthquake. In a later study [37], the same authors used the DES model to evaluate the following strategies for reducing the duration of post-earthquake water restoration: maximizing groundwater pumping, connecting raw emergency water storage reservoirs and rationing water use. Luna et al. (2011) also used DES to model the post-earthquake recovery of water systems. The model was divided into four modules including damage simulation, capacity loss, resource estimation, and resource allocation. The methodology was applied to the trunk network of the Tokyo water system. Recovery curves describing the percent of customers with restored service as a function of time following the earthquake was the main model output.

The current study uses pipe damage, inspection and repair data from the 2014 South Napa earthquake to construct a DES model of the city's water system. The model was developed in collaboration with the Napa Water Division (NWD). The acquired information includes the water network database, the pipe leaks, and breaks caused by the 2014 earthquake and the timing and sequencing of inspection and repairs. The DES model of the water network was used to hindcast the post-earthquake inspection and repair processes (Figure 2-1). Network-specific fragility functions were developed using pipeline damage data. Using the newly developed fragility functions and the validated DES framework, the damage, and restoration following a hypothetical magnitude 6.7 earthquake was assessed. The time-dependent performance of the network was quantified using the system serviceability index.

The integration of the employed dataset (from the 2014 Napa event), coordination with the agency in charge of managing the recovery and time-dependent pressure-driven hydraulic analyses, are what distinguishes the work presented in this paper from prior DES models of

earthquake impacts to water distribution systems [37,39,55–57]. Moreover, while [50] also conducted a hindcast study of the damage and restoration of the Napa water system following the 2014 event, his model was based on a series of analytical equations and did not incorporate an explicit hydraulic analysis. The dependence of the water network performance on other lifeline systems (e.g. power) [57,58] is not considered in the current study. Also, this paper only focusses on the technical dimension of community resilience [59]. The interaction with other dimensions (e.g. social) [60,61] is not considered. The developed framework will be used as a planning tool by the Napa Water Division to evaluate a suite of "what-if" scenarios that capture variations in the scale of earthquake damage and the effect of pre- and post-earthquake mitigation measures.

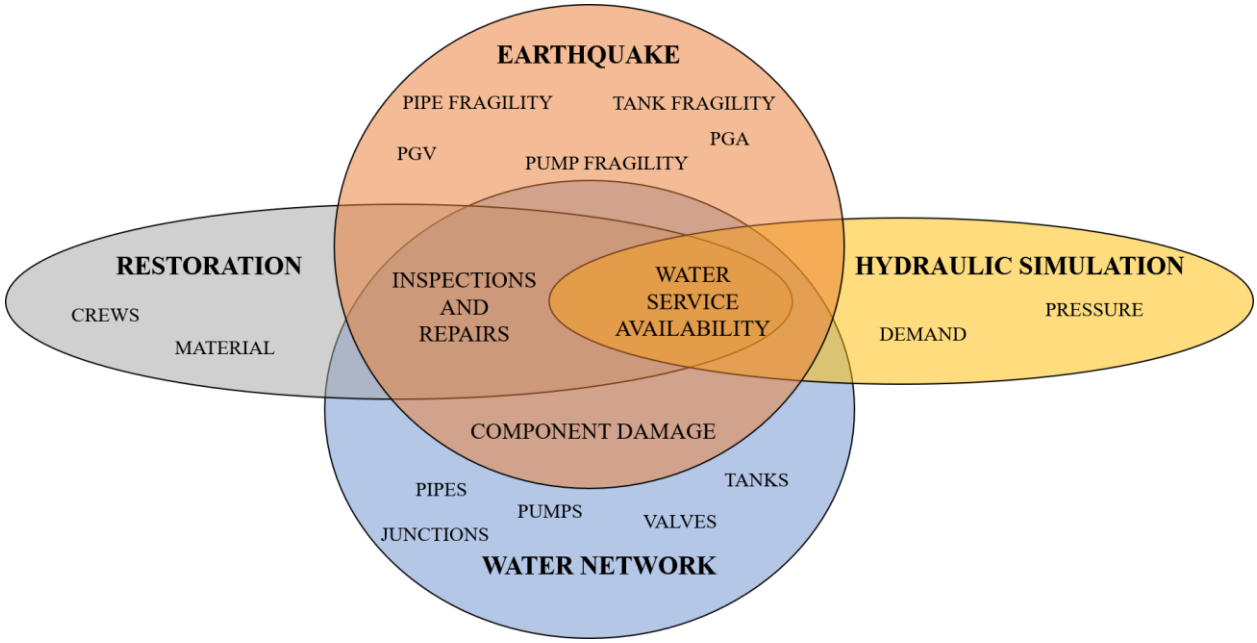


Figure 2-1 Schematic illustration of the interaction between the water network components, earthquake hazard, post-event restoration, and hydraulic simulation

2.2 Description of Napa Water System and Impacts from 2014 Earthquake

2.2.1 Description of Napa Water Network and Services

The NWD serves a population of about 88,000 people, via 25,000 service connections. A map of water service area showing key components of the system, the epicenter location for the 2014 Napa earthquake and the West Napa fault geometry [62] are shown in Figure 2-2. The system includes 3 water treatment plants, about 612 km (380 mi) of pipeline (~7400 pipe segments), 15 storage tanks with a total of 113,562 m^3 (30-million-gallon) storage, 10 pump stations, 14 pressure

regulating stations, and is operated in 5 pressure zones. The distribution lines include main (762 mm (30 inch) diameter), transmission (152.4 – 1066.8 mm (6-42 inch) diameter) and distribution pipes (203.2-762 mm (8-30 inch) diameter). Most pipes in the system are made from cast iron (CI) and ductile iron (DI), which make up 42% and 34% of the network, respectively. The other materials include asbestos cement (AC) (10%), steel (STL) (9%), polyvinyl chloride (PVC) (2%) and reinforced concrete (RC) (1%) [63].

The underground tanks (Hennessey and Lake View) are constructed using reinforced concrete, and steel tanks are used above ground. The capacity of the aboveground tanks ranges from 2.65 m^3 (700 gallons) to 18,927 m^3 (5 million gallons). The 10 pump stations serve about 10% of the population and the remaining 90% of the population is served by gravity flow coming from the water treatment plants [63].

The NWD relies on three major sources of water: Lake Hennessey (38,237,880 m^3 (31000 acre-ft)), the Milliken Reservoir (1726872 m^3 (1400 acre-ft)) and the Storage Water Project (SWP) (27013212 m^3 (21900 Acre-ft)). Lake Hennessey and the adjacent treatment plant are located approximately 21 km (13 mi) north of the city. The Milliken reservoir served as the City's only water source until the construction of Conn Creek Dam (which retains Lake Hennessey) in 1946. However, the Milliken reservoir is only currently used during the high demand summer months. The SWP was established in 1966 and delivers water through the North Bay Aqueduct (NBA) by diverting water from the Sacramento-San Joaquin Delta to two NBA terminal 5 MG reservoirs (built 2008) located at the Jamieson water treatment plant [63].

All potable water is extracted from the three surface water sources, and no groundwater is used. Of the total demand, approximately 53% is single-family residential, 16% is multifamily residential, and 15% is commercial. Approximately 14% is used by institutional facilities including landscaping, agriculture and construction and 2% is exported to the town of St. Helena. Water demand in the system peaks at about 94,635 m^3 per day (25 million gallons per day (MGD)) during a hot spell in July and drops to about 26,498 m^3 per day (7 MGD) during the winter months [63].

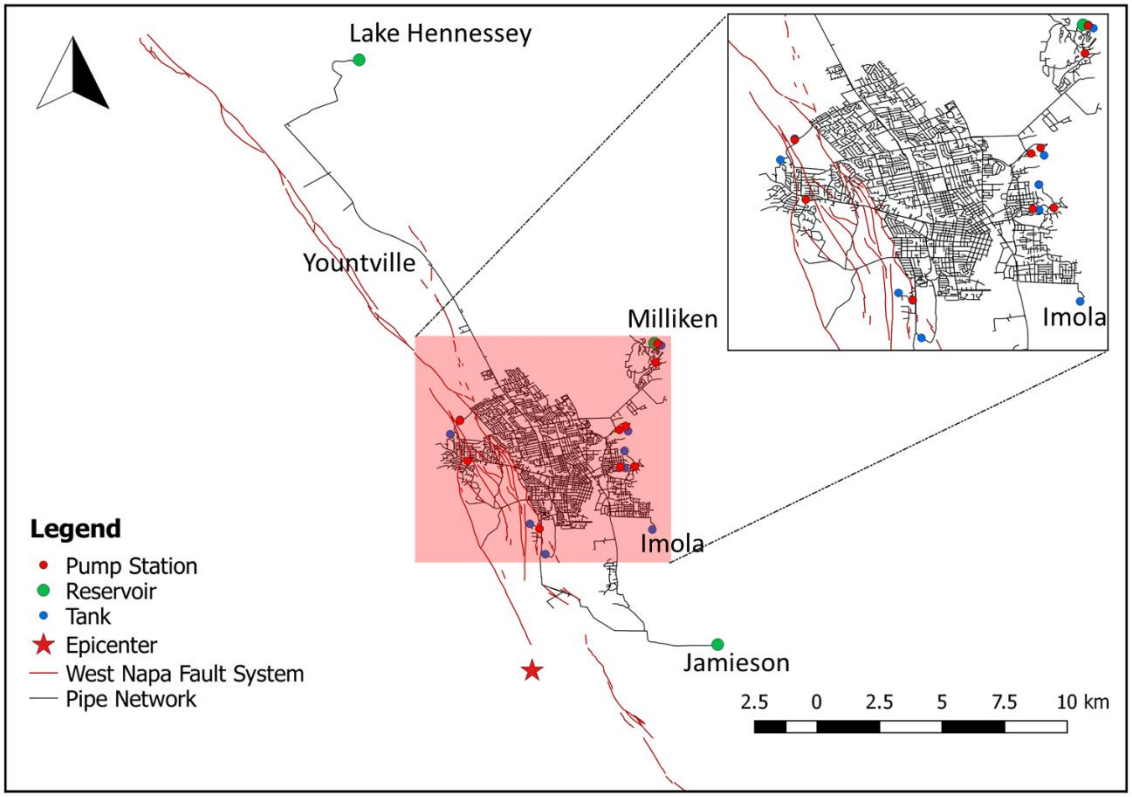


Figure 2-2 City of Napa Water Service Area with 2014 Earthquake Epicenter and Napa Fault System

2.2.2 Impacts from 2014 Earthquake

On August 24, 2014, a **M** 6.0 earthquake struck South Napa, causing damage to buildings and lifeline systems in the cities on Napa, American Canyon, and Vallejo. The epicenter was located approximately 8 km south-southwest at N 38.22 W 122.13, and the hypocenter was 11.3 km deep. Two fatalities were directly attributed to the earthquake, and there were three hundred injuries. Residential buildings accounted for most of the approximately 2,000 structures that experienced damage that could be described as moderate to severe. An estimated $500 million to $1 billion in economic losses was attributed to the earthquake [64].

According to data provided by the NWD, 250 pipe leaks (211 pipe segments since some pipes suffered multiple leaks) occurred as a result of the 2014 earthquake. Figure 2-3 shows the distributions of peak ground velocity (PGV) and pipes damaged during the 2014 event. Kriging was used to interpolate the ground motion intensities from the recorded sites (www.strongmotioncenter.org) [65] to the pipe locations [66]. The PGV values shown in Figure

2-3a range from 2.5 cm/s to 88 cm/s. The spatial distribution of pipe leaks is shown in Figure 2-3b, where it can be observed that much of the damage is concentrated in the west side of downtown Napa. Prior reports summarizing the effects of the earthquake on the water system have suggested the possibility that some pipe damage was caused by ground strains and permanent ground deformation [64,67,68], specifically in Browns Valley and North of Napa Valley, which are areas with moderate to high susceptibility to liquefaction. However, only two cases of damage induced by permanent ground deformation were definitively identified, one of which was the main transmission line from the Milliken source that suffered damage due to a rockslide. The inconsistency between the spatial distribution of PGV and pipe damage is attributed to the fact that pipe type/age, soil conditions and the occurrence of rockslide, which are not explicitly considered in this study, also affected the damage distribution. One water main leak impaired the firefighting efforts at the largest of six fires that occurred due to the earthquake [69]. Figure 2-4 shows the distribution of pipe damage based on material. Approximately 75% of the damaged pipes were made from CI and the remaining 25% are made from AC (7%), PVC (less than 1%), DI (12%), STL (5%) and other unknown materials (mostly reinforced concrete) (less than 1%). A one million-gallon capacity steel tank located in Browns Valley suffered damage to the roof due to the sloshing of water during earthquake shaking. The tank also drained rapidly after the earthquake due to a leak in a nearby pipeline. No damage to pumping stations and treatment plants were reported [67].

The city of Napa benefitted from the redundancy in the water supply system during the periods following the 2014 earthquake. More specifically, the City's decision to maintain the flow from both Lake Hennessey and the SWP meant that, despite the damage to pipelines and storage tanks, there was minimal disruption to potable water service. Pipe leaks were repaired relatively quickly with 120 leaks fixed in 5 days, and during that time, only 5% of the customers were without water. The city of Napa was aided by regional utilities through the CalWARN system. The participating utilities included Alameda County Water District (1 truck/crew), City of Fairfield (1 truck/ 2 crews), Contra Costa Water District (1 truck/crew) and East Bay Municipal Utility District (5 truck/crew). The crews arrived with their trucks and equipment, fully stocked with spare parts and NWD employees guided them during the process. Within a day of the earthquake, NWD was able to assemble repair materials (e.g., pipe, repair clamps, fittings) valued at approximately

$250,000. Within a month after the earthquake, the roof system of the damaged tank was replaced [64].

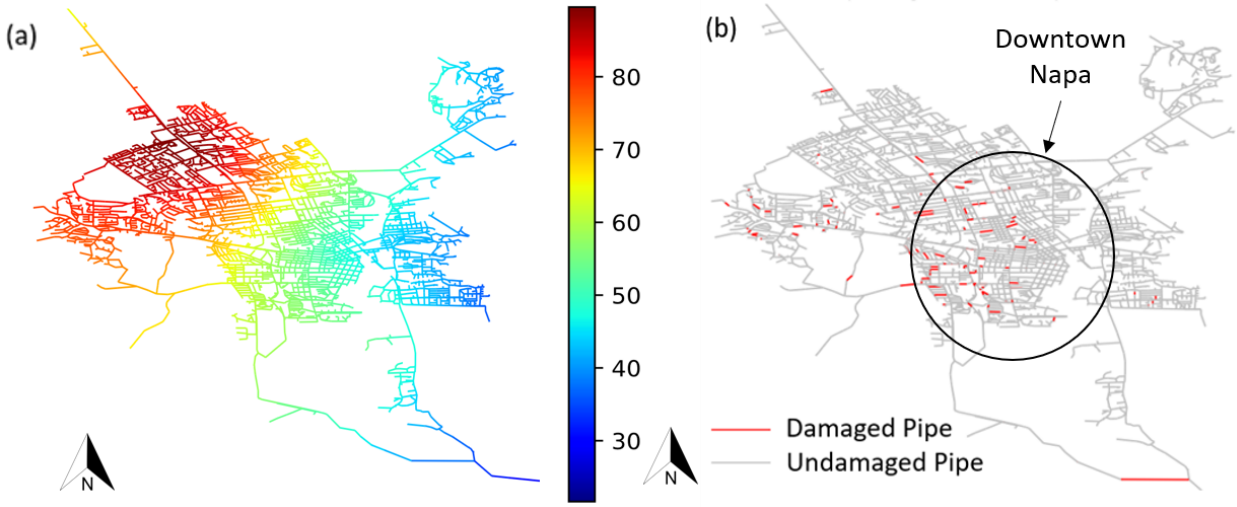


Figure 2-3 (a) peak ground velocity (PGV in cm/s) and (b) pipes damaged (leaks) during the 2014 South Napa Earthquake

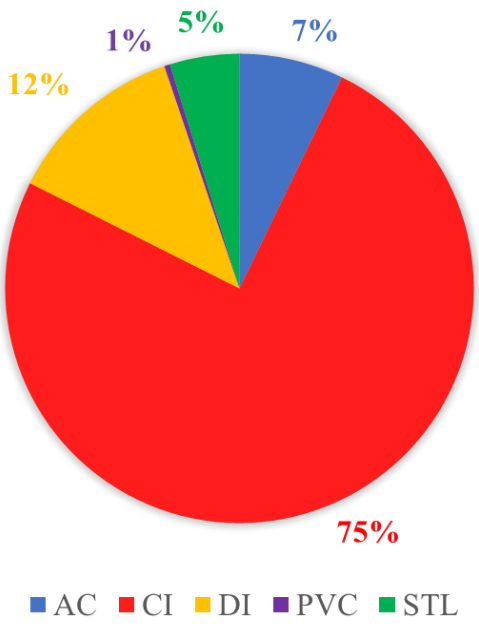


Figure 2-4 Percentage of total damaged pipes based on material

2.3 Discrete Event Simulation Model for Napa Water System

2.3.1 Overview of Modeling Framework

Discrete event simulation (DES) is used to model the earthquake-induced disruption and functional restoration of the Napa Water system where full or 100% functionality is taken as the state where the repairs to all damaged components have been completed [59]. DES models represent the behavior of a complex system as a sequence of events that occur at discrete points in time (Fishman 1978). The core elements of DES models include entities, attributes, events, resources and time. Entities are used to represent specific objects (or concepts) within the system, which can be animate or inanimate. These entities have attributes, experience events and consume resources over time. Attributes are a set of features that are specific to each entity, the accumulation of which defines the state of that entity at any given point in time. Events are occurrences that can affect the state of an entity and resources are used to represent finite or infinite goods that provide services to entities [72]. It is worth noting that, since no state changes occur between events, no computations are needed between consecutive events, which makes the DES approach computationally efficient relative to continuous simulation methods [70]. All entities reside in an environment, where they interact with events, processes, and resources. The DES modeling approach was adopted for the current study because it enables explicit representation of the resources, processes (a series of events), and variables needed to restore water service. The introduction of external resources from nearby areas or mutual aid agreements can also be included in the proposed DES model [73].

The components of a water network (e.g., pipes, tanks, and pumps) are represented as entities in the DES model. These components can be damaged as a result of ground shaking and displacement during an earthquake, which is modeled as an event. The first main step in the overall recovery is the initiation of component inspections (modeled as a process), whereby the crews (resources) travel to various network-components and perform inspections within some duration (inspection timeout event). If the damage is encountered, the component is reported to a repair crew who performs the repairs, which are also modeled processes. The crew then travels to the damaged component and performs the necessary repairs within some duration (repair timeout event). The rate of inspections and repairs is generally high during the period immediately following the earthquake because the affected jurisdiction receives assistance from those nearby.

As time progresses, this assistance is reduced (modeled as a reduction in resources, i.e., decrease in inspection crews and repair crews after the earthquake).

A schematic representation of the Napa water system DES model is shown in Figure 2-5. The physical components of the system (e.g., pipes, tanks, reservoirs, pumps) are represented as entities, and inspection crews, repair crews, and repair material are represented as resources. Variables are used to define the state of the system as a whole or are attributes that are associated with specific entities. Examples of pipe variables include its damage status, the date and time that damage, inspection, and repairs occur, and the pipe material. During the simulation, variable changes are tracked as a function of time. An event (inspection or repair) updates the value of variables. This one-to-one mapping between the physical entities and their abstraction within the simulation model minimizes the use of simplifying assumptions. The Python-based SimPy [74] process-based DES (PBDES) is implemented to model post-earthquake restoration of the water system. Within the PBDES framework, a sequence of events is represented as a process. For example, the start and end of a pipe inspection represent two events that, together, make up a single process. Repair processes are defined in a similar manner. Figure 2-6 illustrates the timing and occurrence of a set of discrete events for a hypothetical two-pipe water system with two inspection crews and one repair crew. Each state represents a snapshot of the attributes of the entities at different points in time. At State 1, an earthquake occurs, which causes no damage to pipe 1 and a leak in pipe 2. The inspection process begins to interact with pipes and to check for damage at State 2. State 3 represents the completion of the inspection process, and it is reported that there is no damage to pipe 1 and a leak in pipe 2. At State 4, the repair process begins to interact with pipes to repair the leak in pipe 2. The repair process is completed at State 5, after which, both pipes have no damage.

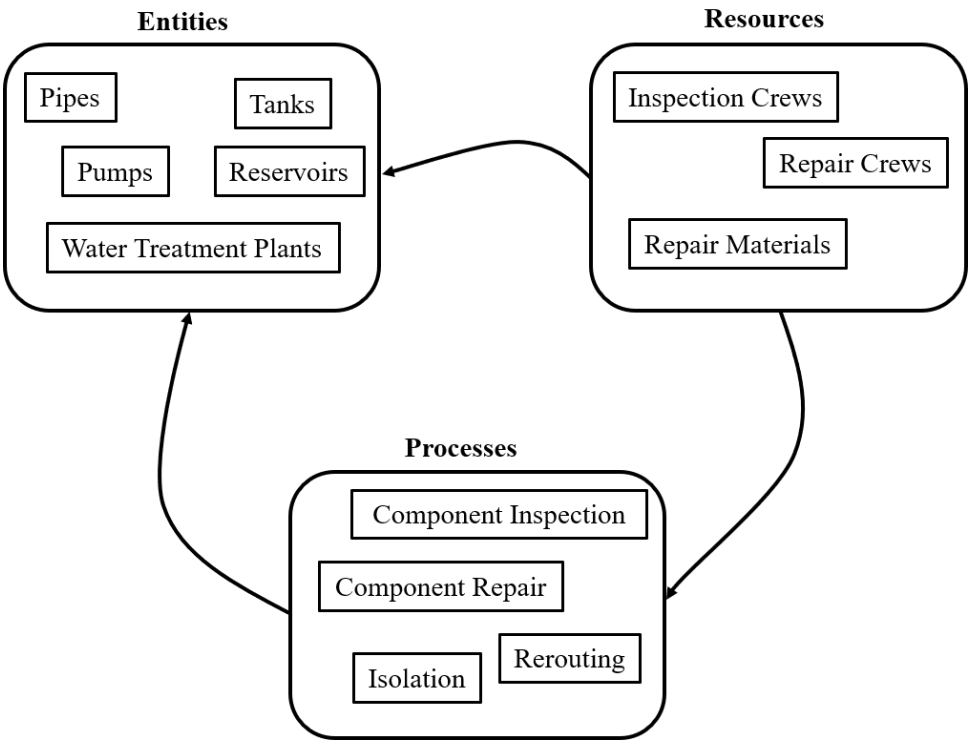


Figure 2-5 Schematic representation of DES modeling framework

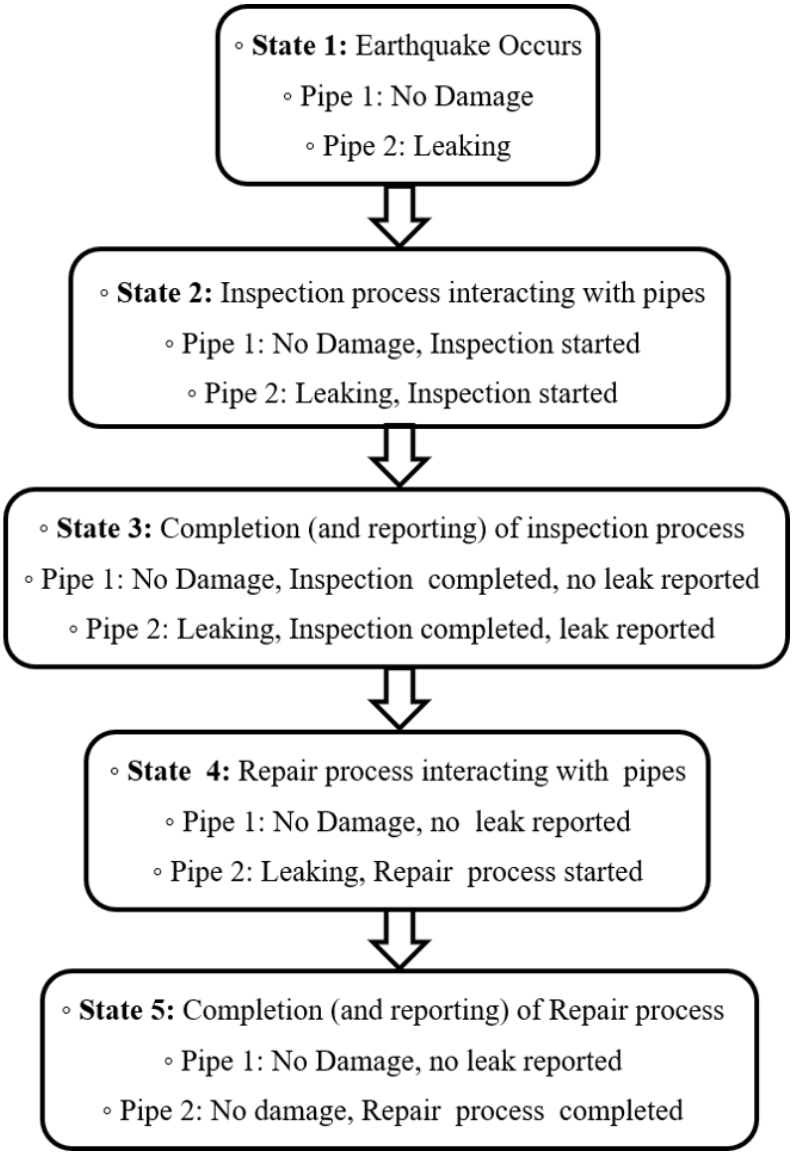


Figure 2-6 Schematic illustration of the timing and occurrence of a set of discrete events (and their associated processes) for a hypothetical two-pipe water distribution system

2.4 Replicating the Recovery following 2014 Earthquake

Using the data provided by NWD, the PBDES (inspection and repairs) model was tuned to obtain a reasonable match between the observed and simulated recovery. More specifically, NWD provided both qualitative and quantitative information regarding the type of damage, scheduling of inspection and repair crews, duration of pipe inspection and repairs and hydraulic performance during the 2014 event. The DES model was then constructed using the relevant NWD information as inputs (e.g., average pipe inspection and repair duration) and manually tuned (by varying the

crew reduction scheduled) to match the observed the pipe inspection and repair trajectories. NWD noted that the sequencing of inspections was informed by public outreach (i.e. residents reported leaking pipes and low water pressure) and the repairs are based on the observed distribution of pipe damage shown in Figure 2-3b. They also noted that $250,000 worth of repair material was ordered on the day immediately following the earthquake. Therefore, material availability was not a constraint during the restoration process and therefore not considered in the current DES model.

At the start of the DES, the inspection process interacts with all pipes and reports the ones that are damaged in terms of leaks and breaks to the repair process. It is assumed that the pipe damage is communicated to the repair process immediately after it is encountered without any time-lag. The inspection process proceeds by randomly sampling from the set of uninspected pipes and reporting its damage status. Note that this random sampling doesn't account for spatial proximity. However, the assumption is based on discussions with the NWD, who indicated that the sequencing of the pipe inspections and repairs after the 2014 event was largely driven by calls from customers. If a pipe is found to be damaged, it is added to the queue of reported pipe damage. The inspection process is constrained by the number of crews and for a single pipe, can only begin between the hours of 9 a.m. to 6 p.m. However, it is possible for an inspection to extend outside of this time-window if the start-time is prior to but close to 6 p.m. The repair process begins immediately after the first pipe is added to the queue of reported pipe damage and starts the repair of that pipe. Like inspection, the repair process is constrained by the number of crews and the start-time for a single pipe is always within the hours of 9 a.m. and 6 p.m. As all the pipe damage in the 2014 South Napa earthquake were leaks, for the calibration exercise, the pipe inspection and repair times for leaks are taken as 1 hour and 3 hours, respectively. It was further assumed that 100 inspection crews and 10 repair crews are available immediately following the earthquake. These assumptions are intended to capture the unplanned timing of inspections and repairs following the event, which were largely driven by customer notifications about leaks and service outages. Then, the crew-numbers are incrementally reduced after each 5-day interval until only 2 of each remain as shown in Table 2-1. As noted earlier, this crew-reduction schedule serves as the primary means of iteratively "tuning" the PBDES model to match the observed inspection and repair trajectory. Again, based on discussions with NWD, the final schedule shown in Table 2-1 is generally consistent with the actual occurrences following the 2014 event.

Table 2-1 Inspection and repair crew reduction during the first 30 days after the 2014 event

Days After the Earthquake	Number of Inspection Crews	Number of Repair Crews
Day 0	100	10
Day 5	90	10
Day 10	72	6
Day 15	42	6
Day 20	21	3
Day 30	8	2

Figure 2-7 compares the actual and simulated recovery, which is quantified based on the number of pipes repaired as a function of time. The higher slope of the curve during the initial days of the recovery represents the higher number of participating inspection and repair crews during that period. As noted earlier, the number of crews is gradually reduced over a thirty-day period, which in turn reduces the rate of recovery. This is evidenced by the lower slope of the recovery curve in later periods following the earthquake. The differences between the actual and simulated recovery curve represent the differences between the actual and modeled inspection and repair time and crew reduction schedule.

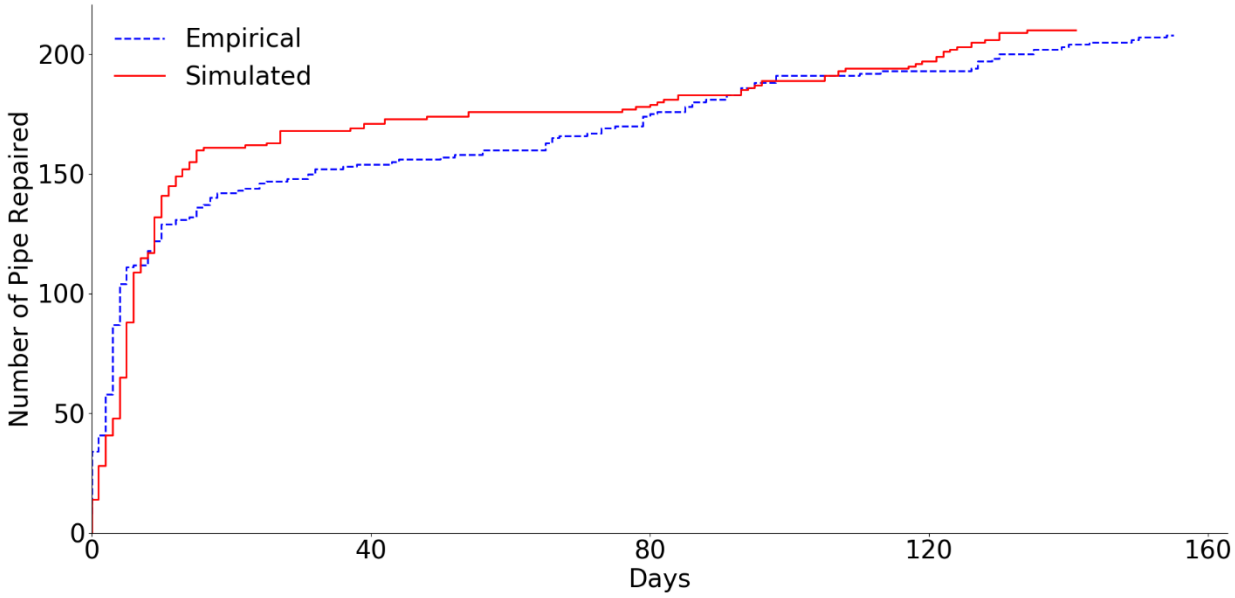


Figure 2-7 Comparing the simulated and observed progression of repairs following the 2014 event

2.5 Development of Pipe Damage Fragility Functions using Data from the 2014 South Napa Earthquake

Damage to the water system components can be caused by permanent ground deformation or seismic wave propagation. Seismic wave propagation causes (transient) ground strain and curvature in pipes, which can lead to damage. Most pipe damage is caused by transient ground strain unless the pipelines are in locations where there are unstable slopes, poorly compacted fill or liquefiable soil [75]. Examples of permanent ground deformation include landslides, surface faulting, settlement, or liquefaction-induced lateral spreading. Among the various seismic intensity measures, PGV has been found to have the highest correlation with wave propagation induced pipe damage [76–78]. Moreover, PGV has been shown to be correlated with axial soil strains caused by seismic wave propagation [79]. Like most prior studies on the seismic performance of water distribution systems [21,78,80–82], PGV is used as the intensity measure in the current study and permanent ground deformation (e.g. liquefaction, landslides) is not considered.

Fragility curves are used in seismic risk assessment to provide a probabilistic link between the predicted ground motion intensities and the damaging effects on the built environment [83]. Earthquake-induced damage to the components of a water network can be probabilistically estimated using fragility curves. The possibility of component damage can be described using two or more mutually exclusive and collectively exhaustive states. For instance, a pipe subjected to earthquake shaking can be described as undamaged, broken or having a major or minor leak. For pipe fragility functions, empirical relations are developed that give the expected repair rates due to ground motion (quantified in terms of PGV) The American Lifelines Alliance (ALA) has formulated seismic fragility functions for water system components. Damage to above ground pumps, tanks, and tunnels is often estimated using PGA, while PGV and repair rate (RR) (the number of repairs needed per unit length) are most used for damage to buried pipes caused by seismic wave propagation [84]. For a pipe segment of length L with a given RR, the probability that n repairs are needed is estimated using a Poisson distribution as shown in Equation 2.1:

$$P(N = n) = \frac{(RR.L)^n}{n!}.e^{-RR.L} \tag{2.1}$$

The repair rate does not provide any information about the severity of damage or the type of repair since most fragility relations for pipelines rely on empirical data that does not include the nature of repairs [85]. However, according to the HAZUS-MH 2.1 user manual [86], the type of

repair or damage depends on the type of hazard. Based on the HAZUS recommendation for damage in pipes due to wave propagation, 80% leaks and 20% breaks are considered.

To support the scenario-based assessment described in the next section, RR-based pipeline fragility curves are developed using the observed damage (Figure 2-3b) dataset from the 2014 event. These curves relate the pipe RR to the ground motion intensity measure (i.e., PGV). To enable the RR calculation, the Napa area is divided into 252 (1.45 km X 1.85 km) (1 minute) grids. For each grid, the mean intensity measure (PGV) value, the total pipe length for each material and the number of damage points are obtained. The repair rate for a single grid is taken as the number of damaged points on each pipe segment divided by the total length of pipe in that grid. Figure 2-8 shows the computed RRs (per km) based on pipe material. The values shown in Figure 2-8 are the mean RRs considering all grids. The CI pipes, which comprise of 42% of the network, have the highest mean RR of 0.79 per km. The mean RR of the AC pipes is 0.32 per km, which is the second highest of the different pipe materials in the network. The STL and DI pipes have comparable mean RRs and PVC has the lowest.

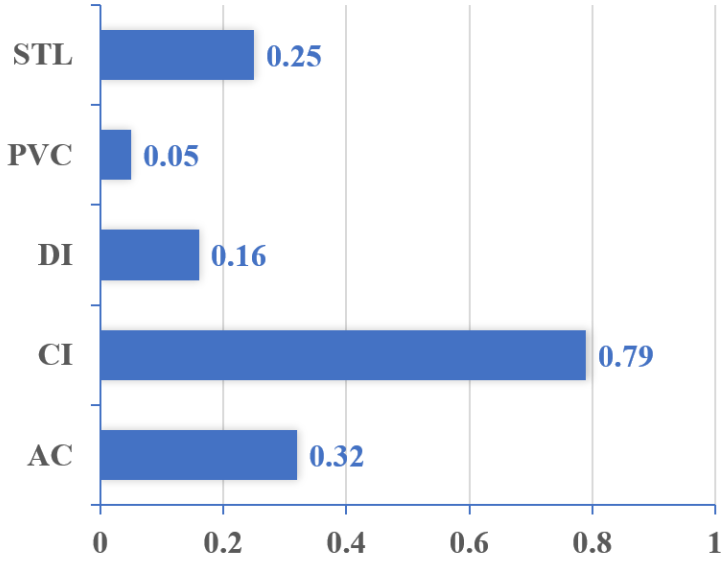


Figure 2-8 Mean repair rates (per km) based on pipe material

Figure 2-9 shows a plot of the repair rate versus PGV for the entire damage dataset (i.e., there is no disaggregation based on pipe material) where several fitted models are shown in addition to the empirical data. Linear and exponential models are constructed from the empirical data using least squares regression. ALA-based linear and exponential models [84] are also shown in Figure 2-9. It can be observed that the fitted linear model performs much better than the exponential

model. The residual-sum-of-squares (RSS) values, which serve as measures of the discrepancy between the data and the estimation models, are computed to be 7.41 per sq. km and 11.85 per sq. km for the linear and exponential models, respectively. The two ALA-based models (linear and exponential) perform comparably in terms of being able to represent the damage observed during the 2014 event. However, as expected, these two models are also outperformed by the fitted linear model. The RSS values for the ALA-based linear and exponential models are 9.31 per sq. km and 9.21 per sq. km, respectively. The parameters of the linear model are shown in Equation 2.2. This model is adopted for the scenario-based assessment described in the next section. Note that because Equation 2.2 has a negative intercept, a negative value of repair rate is obtained for very small PGVs. These unrealistic negative RR values are taken as 0.0.

$$\text{RR (per km)} = 0.01053 \times \text{PGV (cm/s)} - 0.33649 \tag{2.2}$$

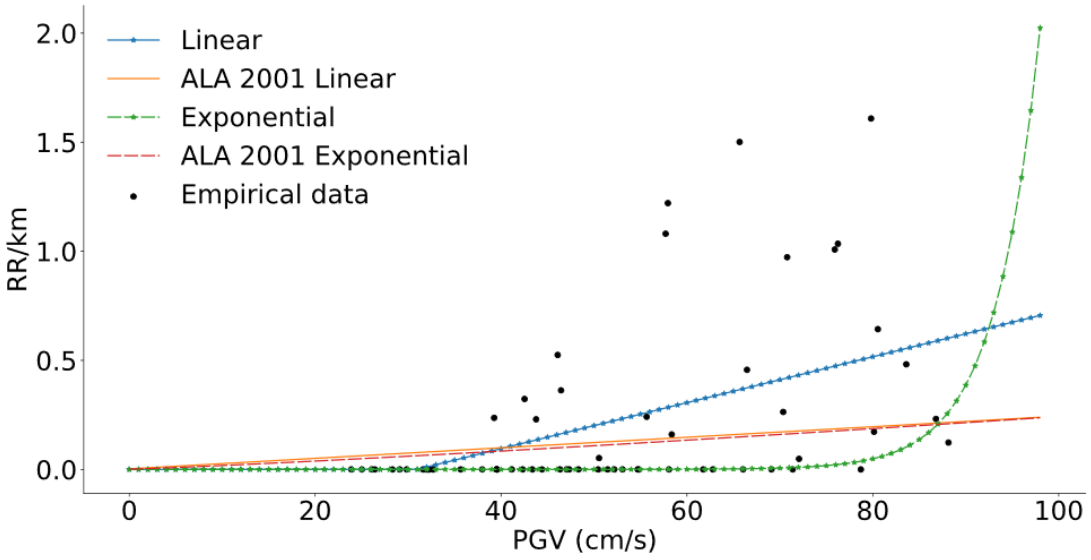


Figure 2-9 Empirical repair rate (RR) model developed using data from the 2014 event

2.6 Scenario-Based Network Damage and Water Service Restoration Assessment

2.6.1 Scenario Description

A scenario earthquake represents one realization of a potential future event based on a particular magnitude, location, and fault-rupture geometry and estimating shaking using a variety of strategies [87]. Given these parameters, a set of spatially correlated shaking intensities over a predefined geographic area is obtained using the appropriate ground motion model(s). Scenario

earthquakes are useful for planning and coordinating the emergency response of local governments and other organizations based on estimates of the physical and socioeconomic impacts.

Using the pipe repair rate and tuned DES restoration model developed in previous sections, the damage and service restoration are assessed for the Napa water system. A **M** 6.7 earthquake occurring on the West Napa fault rupturing the entire fault with an epicentral location of N 38.22 W 122.13 is used as the scenario event which is modeled in OpenSHA [88]. The **M** 6.7 is chosen based on the magnitude-area relationship for a given fault [89,90]. The spatial distribution of shaking is obtained using the Scenario ShakeMap Calculator application in OpenSHA. The inputs to OpenSHA include the latitude and longitude of the earthquake and sites of all considered water system components, and the earthquake rupture parameters (method of rupture, rupture type, magnitude, fault surface). The Campbell and Bozorgnia [91] ground motion attenuation relationship is used to obtain the median shaking intensities. Assuming linear soil conditions, the inter- and intra-event residuals are taken to be 0.322g and 0.576g, respectively for PGA and 0.297 cm/s and 0.578 cm/s, respectively for PGV [91]. Spatially correlated shaking intensities are generated using the model by Jayaram and Baker [92]. Note that the cascading effects of fault failure (multiple fault section ruptures) recommended in the Uniform California Earthquake Rupture Forecast, Version-3 (UCERF 3) are not considered.

2.6.2 Ground Motion Intensities and Network Damage

Figure 2-10a shows the PGVs generated for the **M** 6.7 scenario, which ranges from 52.2 cm/s to 377.12 cm/s. This is significantly higher than the 2.5 cm/s to 88 cm/s PGV range from the 2014 event. There are also differences in where the highest levels of shaking are concentrated. The highest shaking intensities for the 2014 event occurred in the Northwest region of the city, whereas for the **M** 6.7 event, the strongest shaking is concentrated in the central business district area. The maximum PGV generated by the OpenSHA model (107.2 cm/s) occurs west of the Napa River in the community of Stanley.

Using the ground motion intensities (PGA, PGV) generated from the **M** 6.7 event, pipeline damage within the Napa water network is assessed. The aggregated (not specific to the material) RR model developed in the previous section is used for the damage simulation. The 1.45 km X 1.85 km grids used to develop the RR model was also utilized for the scenario damage assessment. Given the PGV value within each grid, the RR is computed using Equation 2.2 and used as the

input into Equation 2.1 to compute the probability of a "leak" or "break" in a segment of pipe with length L. For pipes fragility curves, an exponential distribution is used to represent two damage states (DS): leak (DS1) and break (DS2). 80% of pipe damage is assumed as leaks and 20% is assumed as breaks as recommended by HAZUS [86]. It is worth noting that, since permanent ground deformations are not captured, the spatial distribution of pipe damage is likely underestimated. For tanks and pump stations, the PGA-based fragility functions provided by HAZUS [86] are used. Different types of damage have different repair times, which are used in the DES restoration model. The rate at which water escapes from a leaking or broken pipe is also different. This information is incorporated in the hydraulic simulation described later in the paper.

For the restoration modeling (described in the next section), Monte Carlo simulation is performed where 1000 realizations of spatially distributed network component damage are generated for the event. The number of realizations was chosen to minimize the dispersion across the Monte Carlo simulation estimates. Figure 2-10b shows the distribution of pipe damage for a single realization. For a typical damage realization, approximately 560 leaks and 145 breaks occurred as a result of the **M** 6.7 event. Compared to the 2014 earthquake, the pipe damage for the **M** 6.7 event is much more dispersed throughout the network. However, like the 2014 earthquake, CI pipes comprise the majority of the damage for the **M** 6.7 event, accounting for 32.8% of the leaks and 34.4% of the breaks in the network.

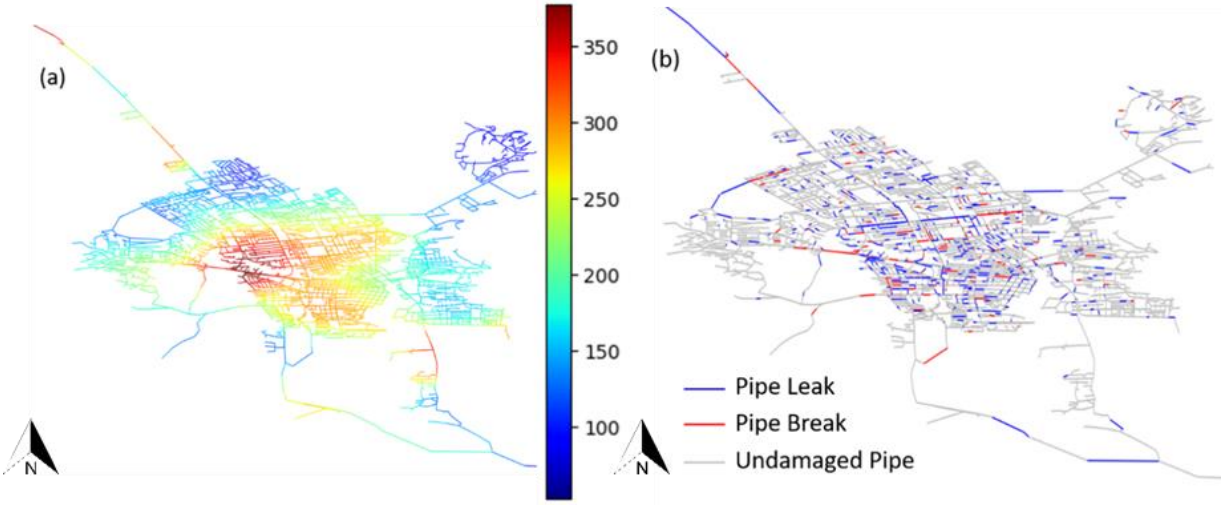


Figure 2-10 Spatial distribution of (a) PGVs and (b) pipe damage for **M** 6.7 scenario

2.6.3 Modeling the Inspection and Repair Following the M 6.7 Scenario

The tuned DES model is used to simulate the inspection, repair, and water service restoration following the **M** 6.7 earthquake. The previously described DES modeling procedure is also adopted. However, for the scenario-based assessment, the uncertainty in the inspection and repair processes are captured by assigning probability distributions and associated parameters (e.g., central tendency and dispersion) to the number of crews and repair durations. The initial number of inspection and repair crews for pipes are sampled from a uniform distribution whose values can range from 10 to 120 and 20 to 40, respectively. Similarly, the initial number of inspection and repair crews for tanks and pumps are sampled from a uniform distribution ranging from 10 to 20 and from 2 to 10, respectively. These value ranges are adopted based on discussions with the NWD, and the uniform distribution has been used in prior studies on DES modeling of water systems [55]. The crew-reduction schedule for pipe inspection and repair crews shown in Table 2-1 is also included in the model. A triangular distribution is used to probabilistically represent the durations of inspections and repairs of pipes [55], which are shown in Table 2-2. The normal distribution is used for the tank and pump station repair times [86], which is shown in Table 2-3. The values differ based on the type of component (e.g. inspection time for a reservoir is different from a pipe). It is worth noting that the deterministic initial values for the number of inspection and repair crews used in the calibration exercise also lie within the uniform distribution range used for the scenario-based assessment.

Table 2-2 Event Duration Triangular Distribution

Event		Min	Mode	Max
Inspections	Trunk or Distribution Damage Tank Pump	0.5 hour	1 hour	2 hours
Repairs	Distribution Leak	3 hours	4 hours	6 hours
	Distribution Break	4 hours	6 hours	12 hours
	Trunk Leak	4 days	4 days	6 days
	Trunk Break	6 days	8 days	10 days

Table 2-3 Event duration mean and standard deviation (σ)

Event			Mean (days)	σ (days)
Repairs	Pumping Plants	slight/minor	0.9	0.3
		moderate	3.1	2.7
		extensive	13.5	10.0
		complete	35.0	18.0
	Water Storage Tanks	slight/minor	1.2	0.4
		moderate	3.1	2.7
		extensive	93.0	85.0
		complete	155.0	120.0

The uncertainties embedded in the simulated post-earthquake pipe inspection/repair processes include the damage conditioned on the shaking intensity, the number of available inspection and repair crews and the duration of the inspection and repair processes. Figure 2-11 shows a "pipe repair cloud" for the **M** 6.7 scenario. The term "cloud" (in lieu of curve) is adopted because it shows the full range of trajectories when the various sources of uncertainty are propagated in the restoration model. Monte Carlo simulation is utilized, whereby a single repair curve is generated by sampling the damage state of each pipe, the number of crews and the processes duration. In Figure 2-11, each grey line represents a single realization of a repair trajectory and the black lines represent the median (50^{th} percentile) (solid line), 16^{th} and 84^{th} percentile trajectories (dashed lines). All probabilistic restoration outcomes (including the results from the hydraulic simulation presented in the next section) are based on 1000 realizations.

The vertical axis in Figure 2-11 represents the cumulative percentage of pipes repaired after the earthquake. From the median repair curve, it can be observed that the rate of repaired pipes is very high in the first two weeks after the earthquake. This is a period of high levels of activity when the emergency protocols of the NWD are being implemented, including the acquisition of additional resources (inspection/repair crews) from neighboring counties. After the first two weeks, there is a significant decrease in the rate of repairs due to the decrease in the number of inspection/repair crews. The mean time needed to repair 90% of all pipes is 205 days, which is 209% more than the actual time it took to repair 90% of the pipes damaged during the 2014 event. The mean plus and minus one standard deviation for the same metric is 354 days and 56 days, respectively.

For the **M** 6.7 scenario, the impact to tanks and pumps was minimal (i.e. damage was mostly limited to the slight and moderate states). For the 1000 realizations, the median repair time for tanks and pumps was 11 days and 21 days, respectively, indicating that these two components were not as critical as the pipes in the restoration of hydraulic performance.

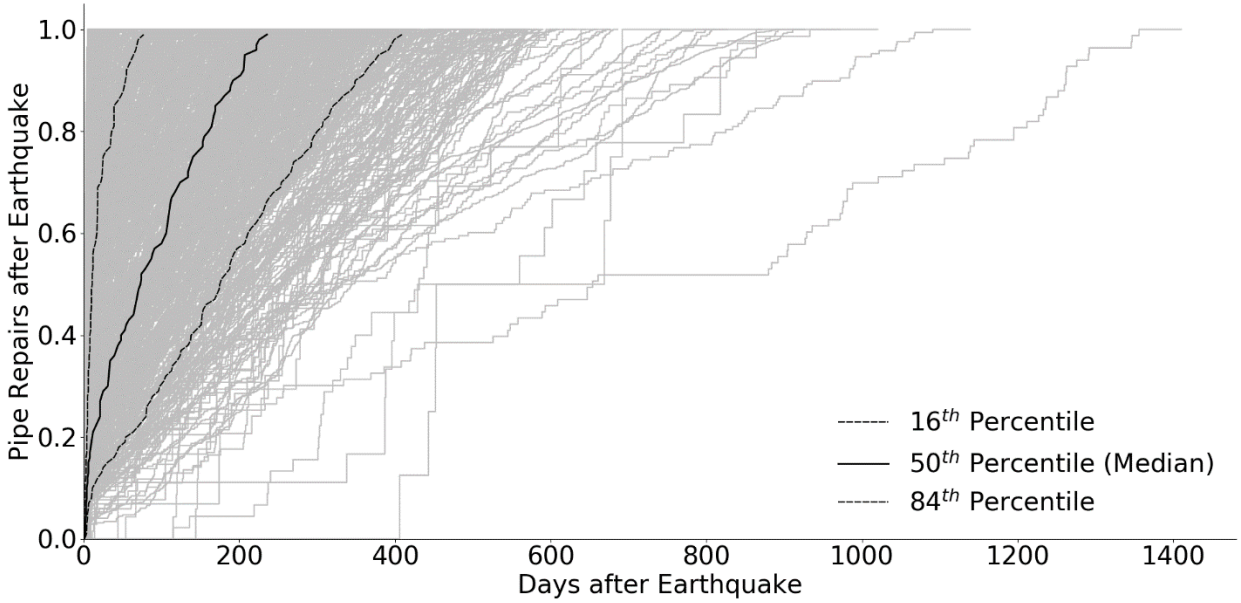


Figure 2-11 Simulated progression of pipe repairs following the **M** 6.7 event

Figure 2-12 shows the probability mass function (pmf) and fitted normal probability density function (pdf) of the time to repair 90% of the damaged pipes. The normal and lognormal distributions are evaluated by performing the two-sample Kolmogorov-Smirnov (KS) test [93] where the null hypothesis is that the pdf of the time to repair 90% of damaged pipes follows a particular distribution. *p-values of 0.046 and* 10^{-6} are obtained for the normal and lognormal distributions, respectively. Even though both are below the assumed 5% threshold, the normal distribution is adopted because the *p-value* is much higher than the value associated with the lognormal distribution. The coefficient of variation (COV) for the time to repair 90% of the damaged pipes is 0.72. In general, the repair time dispersion decreases with the cumulative fraction of the repaired pipes. For example, the COV for the time to repair all damaged pipes is 0.69, compared to 0.96 for 50% of the damaged pipes. Figure 2-13 shows the empirical and theoretical (based on a normal distribution) cumulative density function (cdf) of the time to repair 90% of damaged pipes, which can be used to probabilistically evaluate repair duration targets for the water system. For instance, if the goal is to repair 90% of damaged pipes within a six-month period, the

associated probability of success is only 0.48, whereas the probability of achieving the same repair level target within one year of the earthquake is 0.87. Note that in the empirical distribution, 120 realizations (approximately 10% of all realizations) had 90% of all pipes repaired within the first 50 days, which explains the high initial slope of the cdf.

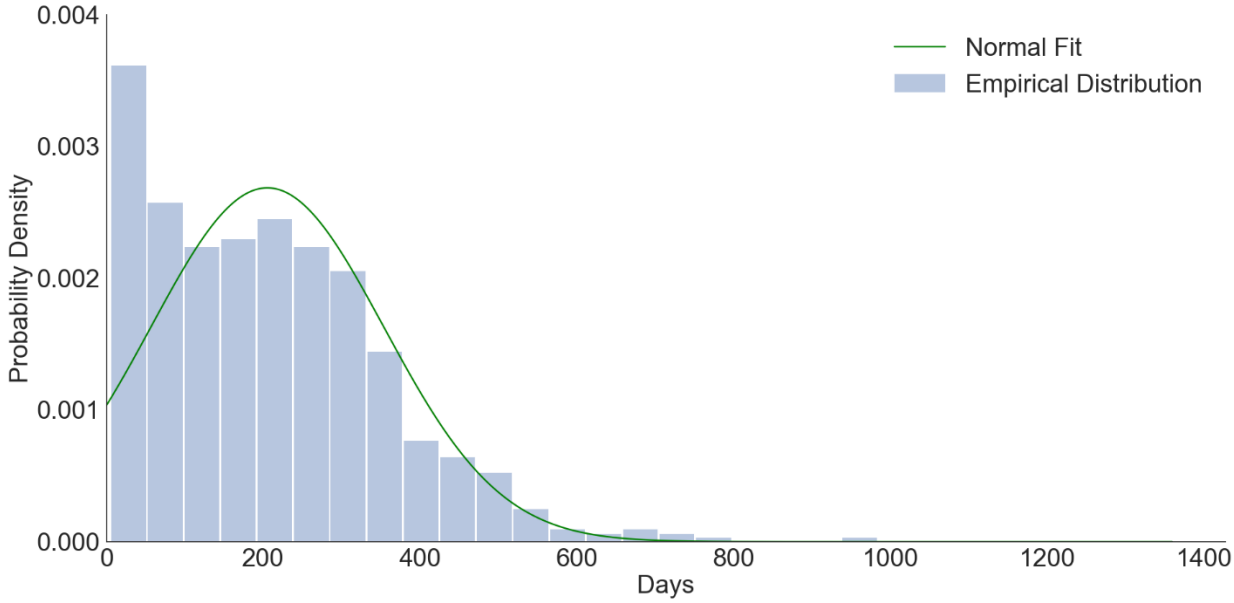


Figure 2-12 Probability density function of time to repair 90% of damaged pipes

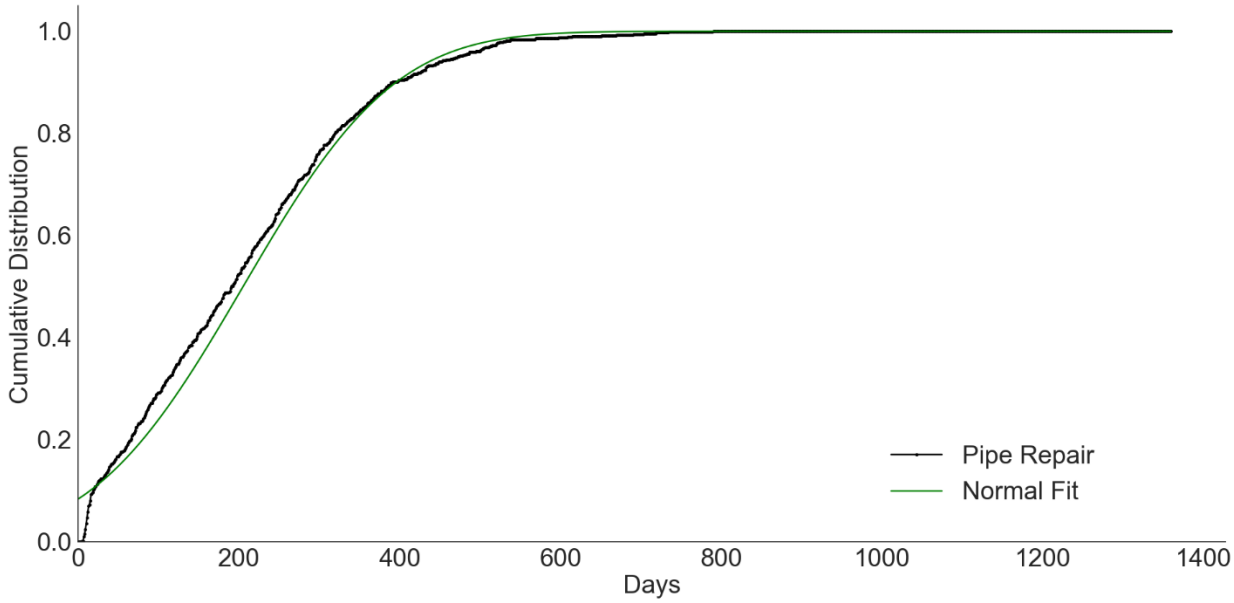


Figure 2-13 Cumulative distribution function (cdf) of time to repair 90% of damaged pipes

2.6.4 Modeling Water Service Restoration Following the M 6.7 Scenario

The completion of pipe repairs does not always ensure the satisfaction of water demands for households and other critical infrastructure (e.g., hospitals, fire-stations). To provide further insight into this issue, a hydraulic simulation is performed at each restoration time step to quantify the availability of water at each demand node. This information can enable a water agency to compare and design more effective pre-event mitigation strategies.

Existing hydraulic simulation models do not support sudden failures in the water network that cause pressure variations and rapid changes in system operation. Moreover, most commonly used hydraulic simulators (e.g., EPANET) employ demand-driven analysis (DDA), which assumes that customer demands are always satisfied even if the pressure is insufficient. In reality, extreme events can cause a considerable reduction in water pressure, and eventually, customer demands will not be met [42]. Extreme events can also lead to changes in the demand on the water network [94], which is not considered in this study.

Given the drawbacks of DDA, several alternatives have been proposed in recent years. [95] developed pressure-driven (hydraulic) analysis (PDA) for water distribution systems in which the demand supplied to a node is a function of the pressure at that node. In other words, for low-pressure conditions, only a fraction of customers will have their demands met. Examples of simulators running PDA include WaterNetGen [41,96], which is an open-source tool, and WaterGEMS™ [97], which is a commercial software. Quasi-PDA simulators (or semi-PDA) run DDA simulations in an iterative manner where the nodes are switched between having constant-demand, zero-demand, and (sometimes) emitters (devices associated with junctions that model the flow through a nozzle or orifice that discharges to the atmosphere [98]), depending on the pressure domain [32,40,99]. The difference between using DDA and PDA simulation can be significant for extreme events [100].

At its core, a hydraulic simulator represents the water distribution system as a network model consisting of nodes and links. Pipes and pumps are represented as links, and junctions, tanks, and reservoirs are nodes. Consumer water demands are modeled as nodal attributes, and available water storage is modeled as an attribute of tanks. The simulator performs a network analysis at each time step, which is based on conservation of energy and mass balance and the pressure and flow of water from tanks through pipes to demand nodes [98]. Pipe damage is represented as the splitting of a segment to create a junction at its midpoint and assigning a demand (leak demand or

break demand) to that junction. All pipe damage is assumed to happen right after the earthquake and the inspection plus repair time for damaged pipes are obtained from the PBDES model. Similarly, tank damage is modeled by adding a leak at the base with an area that is defined in terms of the tank diameter (e.g., for damage states 1 through 4, the leak area is equal to the area of a circular orifice with diameter 0.1%, 1%, 10% and 50% of the tank diameter respectively). The pressure reduction in the tank is a function of the leak area. Also, when this pressure falls below some threshold, the tank is taken out of operation. Damage to pump stations is represented as an outage (i.e., duration where the pump is not operational).

For the hydraulic simulation, this study uses Water Network Tool for Resilience (WNTR) developed by The United States Environmental Protection Agency in partnership with Sandia National Laboratories [42]. WNTR is an open-source package that can perform many complex analyses by exploiting the object-oriented programming capabilities of python. Several aspects of resilience modeling for water distribution systems are integrated into a single software framework, which can be used to model the effects of disruptive incidents and repair strategies. The advancements over other hydraulic simulation platforms (e.g., EPANET) include the representation of PDA, the ability to add and repair leaks, and the ability to change operations and/or response strategies during a simulation. For the **M** 6.7 scenario, the time points associated with the start and end of the inspection and repair processes for each pipe are used as inputs into the WNTR hydraulic simulation model. A pipe damage is modeled by assigning leakage discharge at the downstream node. The discharge is assumed to be a hole in the pipe that is 20% and 95% of its diameter for a leak and break, respectively.

Community resilience is the ability to prepare for anticipated hazards, adapt to changing conditions, and withstand and recover rapidly from disruptions [101]. Metrics that quantify resilience need to account for system function before, during, and after the disruption. Several resilience metrics have been used in prior studies including robustness, redundancy, resourcefulness, and rapidity [59]. For water distribution systems, resilience metrics fall into four categories: topographic, hydraulic, water quality/security, and economic [42]. In this study, the focus is on quantifying the effect of pipe damage and the inspection/repair processes on the satisfaction of nodal demand. In this context, system performance is defined as the ratio of available water supply to the required water demand. More specifically, nodal serviceability is defined as the demand satisfaction ratio at a given node while the system serviceability index (SSI)

is defined as the ratio of post-earthquake water supply to pre-earthquake demands in the system [45] as shown in Equation 2.3.

$$SSI_j = \frac{\sum_{i=1}^{n} q_{i,j}}{\sum_{i=1}^{n} d_i} \tag{2.3}$$

where SSI_j is the system serviceability index for a damage realization j, n is the number of demand nodes in the system, $q_{i,j}$ is the actual water flow supplied to the user at node i under the j^{th} damage realization, and d_i is the water demand at node i. The water demand at a node is obtained from the NWD.

Figure 2-14 shows an SSI restoration cloud, which is generated by performing hydraulic simulation for each restoration (damage, inspection, and repair) realization. The vertical axis represents the SSI value on a given day after the earthquake. Each grey line represents a single realization where the SSI is computed at each time step from the hydraulic simulation output obtained from the WNTR PDA simulator. The 50^{th} percentile (solid line) and 16^{th} and 84^{th} percentile (dashed lines) are also shown in Figure 2-14. It can be observed that the SSI plummets immediately following the earthquake indicating the reduction in satisfied demand nodes. Over the course of the inspection and repair activities, the SSI recovers and eventually returns to a value of 1.0, which indicates that all demands are satisfied with adequate pressure. One observation that is common for all realizations is the eventual sudden increase in the SSI value, which can be explained by the "betweenness centrality" of the network. Betweenness centrality measures the extent to which a vertex (pipe) lies on paths between other vertices (pipes) [102]. After the earthquake, the pipe damage divides the water network into clusters. These clusters are reconnected when the pipes having the highest value of betweenness centrality are repaired, resulting in the sudden increase in SSI. The SSI cloud shown in Figure 2-14 can be used to make informed decisions regarding pre- and post-event mitigation measures that are implemented with the goal of reducing the overall loss of water service to the network.

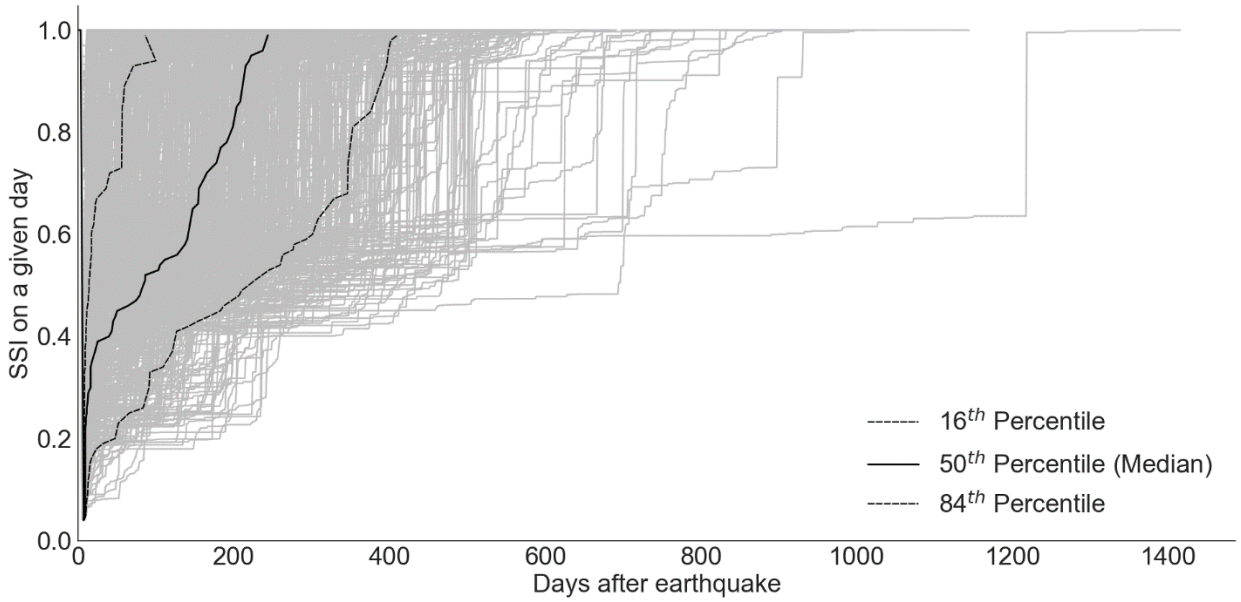


Figure 2-14 Time series of system serviceability index following the **M** 6.7 event

Figure 2-15 shows the empirical cdf for the time to repair 90% of damaged pipes overlaid with the empirical cdf of the time to achieve 90% of SSI. It is observed that there is a lag between pipe repairs and the SSI at CDF values higher than approximately 70%. For instance, the median (50th percentile) pipe repair and SSI durations are almost the same (191 and 190 days, respectively). Whereas the SSI 95th percentile duration is 24.4% higher than for pipe repairs.

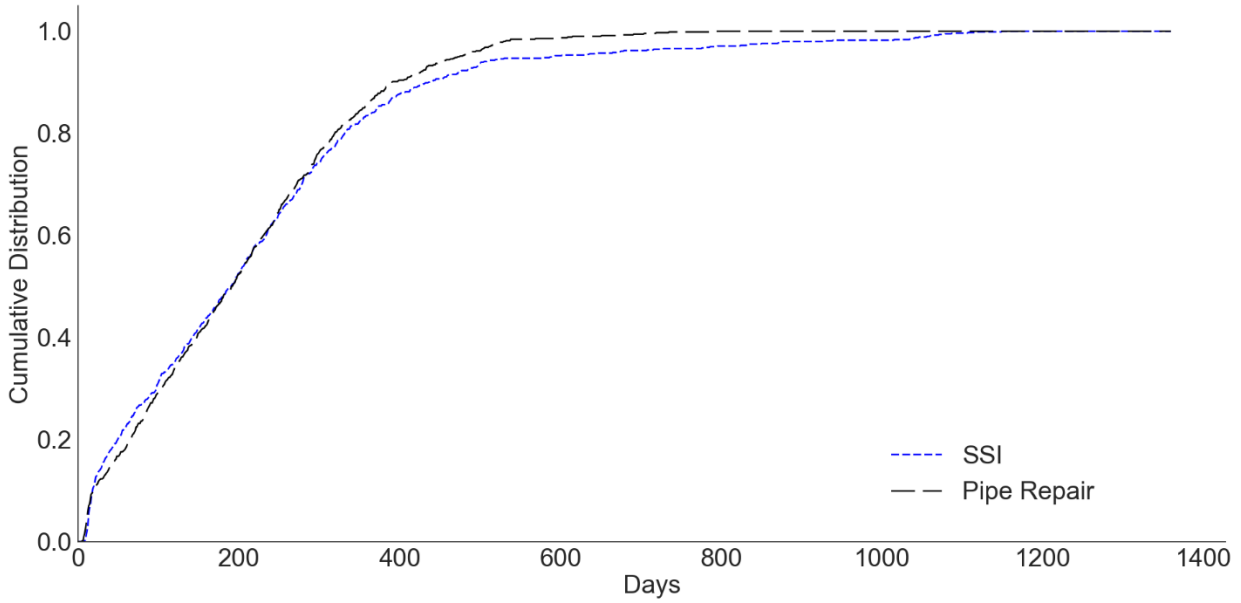


Figure 2-15 Empirical cumulative distribution function (CDF) for time to repair 90% of damaged pipes and time to achieve 90% SSI

2.7 Summary and Conclusion

In this paper, a discrete event simulation (DES) model is used to hindcast the damage and functional impacts to the Napa Water system following the 2014 earthquake. The pipe damage and restoration (inspection and repair) data from the 2014 event was used to tune the DES model. The tuned model was then extended to consider a hypothetical scenario. The overarching framework, key assumptions, and model parameters were established based on discussions with the Napa Water Division. Model tuning was performed by varying the resource-related (inspection and repair crews) parameters. Using the pipe damage data from the 2014 earthquake, new fragility functions were developed and compared with the industry standards (American Lifeline Alliance (2001) recommendations).

After tuning, the DES along with the new fragility functions were used to model the pipe damage, inspection, and repair processes and hydraulic performance of the Napa Water network following a magnitude 6.7 scenario earthquake occurring on the West Napa Fault. The distribution of peak ground velocities (PGV) from the scenario event was obtained from OpenSHA. Monte Carlo simulation was employed to generate the distribution of pipe damage based on the PGV map and newly developed fragility functions. Damage from permanent ground movements was not explicitly modeled. The Monte Carlo realizations of pipe damage were then used as inputs into the DES model, and the corresponding pipe repair curves were generated. Additionally, the overall availability of water at the demand nodes was quantified by performing pressure-driven hydraulic simulation for all pipe damage and repair trajectory realizations. The system serviceability index (SSI) was used to assess the time-dependent availability of water in the network.

Multiple sources of uncertainty were propagated within the scenario-based assessment including the level of pipe damage conditioned on the shaking intensity, resource scheduling (number of repair and inspection crews), and the pipe inspection and repair times. The probabilistic outcomes were described using restoration “clouds”, which show single-realization and central tendency trajectories. The results from a two-sample Kolmogorov-Smirnov test showed that the normal distribution is appropriate for probabilistically describing the time needed to achieve specific levels of cumulative pipe repair and SSI. Associated cumulative density functions are then developed to enable a probabilistic evaluation of restoration-based performance targets.

The developed model can be used to inform decision-making in the pre-and post-earthquake environment. More specifically, for pre-earthquake planning, the damage and functional loss and

restoration forecasts can be used to prioritize the replacement of deteriorating components. Following a real earthquake, the DES model can inform the sequencing of inspection and repairs with the goal of minimizing the cumulative hydraulic impacts to the system. The DES model in the current study was manually tuned using a dataset from a single (moderate) earthquake and characteristic region (small city). This is one of the main limitations of the current study. For example, the operational (e.g. stockpiling repair materials) and physical changes (retrofitting vulnerable pipes) to the studied system that may have been spurred by the 2014 earthquake, were not considered when forecasting the impacts of the hypothetical M6.7 event. To develop a more generalized model, further studies are needed where variations in the regional context and size and impact of the earthquake are incorporated. The random sampling of the pipe inspection process in the PBDES model can be improved by implementing strategies that are currently being used by water utilities (e.g. prioritizing the inspection and repair of larger and/or more important pipes). Additionally, the explicit consideration of operational adaptations, which were beyond the scope of the current study, could provide a more realistic (possibly less conservative) representation of the restoration trajectories. Finally, other network performance indicators such as “betweenness centrality” can be further explored.

CHAPTER 3 Risk-Based Assessment of the Post-Earthquake Functional Disruption and Restoration of a Water Distribution System

3.1 Introduction

Water distribution systems are critical to the well-being of modern society as they support the functionality of most other types of infrastructure. Earthquake damage to the individual components of a water distribution system can trigger functional disruption, which can lead to a reduction in the network supply. Key to developing effective resilience plans for water networks is the ability to quantify the temporal disruption caused by different events and evaluate the effectiveness of different interventions that seek to minimize the cumulative loss of service. Models that assess the immediate functional disruption and restoration following a hazard event can support these types of decision-making processes. One major challenge in this regard is the selection of the event(s) to use as the basis of the assessment. For spatially distributed infrastructure systems, it is often impossible to find one such event. Additionally, even if a single "worst-case" (in terms of consequences) scenario is determined, it is important to consider its probability of occurrence. Risk-based assessment methodologies are useful in this regard because they consider the full spectrum of possible consequential scenarios and their associated occurrence probabilities (or rates).

One of the earliest models to incorporate the time-dependent functionality of water distribution systems following an earthquake utilized simple analytical equations [51]. Described as a "resource-constrained" model, the equations included variables to capture the effect of the severity and number of damaged components and the number, efficiency, and schedule of workers, on service restoration. This resource-constrained modeling approach was later extended to include the prioritization of component repair sequences [52]. Spurred by an increase in access to advanced computing resources, recent efforts to model the earthquake-induced loss and restoration of water network service have embraced simulation-based approaches [37,39,50,56,57,103,104]. A common theme among these studies is that only a single event was considered in their assessments. Single-scenario seismic evaluations are desirable because they require less computational effort and can be useful under some circumstances such as for model validation using data from past

events or communicating potential impacts to the public. However, for other applications such as in the insurance industry and regional resilience planning, it is necessary to estimate the effects of many or even all possible future earthquake scenarios that can cause significant impacts.

The goal of seismic risk assessment for spatially distributed infrastructure is to estimate the annual rate of exceedance or return period of different levels and types (i.e. different performance metrics) of impacts. Stochastic event-set models, which consider most or all significant events and their annual rates of exceedance, have been used for this purpose [105–110]. For seismic risk quantification, this requires the probabilistic generation of a set of ground-motion intensity maps considering all of the significant rupture scenarios that could occur in the region. To consider the uncertainty in the spatial distribution of shaking, multiple maps are produced for each scenario. Conditioned on a single ground-motion map, damage to the individual components that comprise the system is probabilistically sampled using the appropriate fragility relationships. For each damage scenario, the infrastructure performance is evaluated using an appropriate network performance measure. This process is repeated for all ground motion maps and the associated network performances are probabilistically combined to obtain their annual exceedance rates. To reduce the computational expense of stochastic event-set approaches, linear programming has been used to select a subset of damage maps, corresponding ground-motion intensity maps, and associated occurrence rates, and provide a reasonable estimate of the full probability distribution of the network performance [111]. It is worth noting that all of the prior studies on stochastic-event-set assessment of distributed infrastructure utilized immediate-impact performance metrics. In other words, the functional restoration of the system was not considered.

This study presents a methodology for risk-based assessment of the post-earthquake functional disruption and restoration of water distribution systems. An end-to-end discrete event simulation (DES) framework is used to link the spatial distribution of ground shaking for each scenario in a stochastic earthquake catalogue to immediate (physical and functional) and cumulative (functional) impacts. A pressure-driven analysis (PDA), which is embedded in the end-to-end simulation, is used to compute the immediate loss and restoration of the system serviceability index (SSI). By simulating the full SSI recovery trajectory, annual exceedance rate curves are generated for multiple resilience-based performance metrics (e.g. robustness, rapidity, cumulative loss of service). The framework is applied to the water network for the City of Napa and the Uniform California Earthquake Rupture Forecast (UCERF2) model is used to generate the stochastic

catalogue of earthquake scenarios for that region. The types of results generated by the current study can provide a comprehensive understanding of the impact to stakeholders and inform risk-mitigation and resilience-enhancement planning.

3.2 Overview of Risk-Based Assessment Framework

Figure 3-1 shows an overview of risk-based post-earthquake functional disruption and restoration assessment framework for water network. First, a stochastic event-set catalogue comprised of N_s scenarios, is generated for the region of interest. The events in the catalogue vary based on magnitude, epicentral location, rupture parameters (e.g. method of rupture, rupture type), and annual rate of exceedance. For each event, a set of b ground motion maps are produced using the appropriate ground motion model(s) (GMM) and spatial correlation model. For each ground motion map, N_d realizations of damage to individual components (e.g. pipes, tanks and pump stations) are sampled using the associated fragility function and the shaking intensity at the relevant location. The ground motion map used for each damage simulation is randomly sampled from the set of b maps. In other words, considering the number of events and ground motion fields, a total of $M = N_s \times N_d$ damage maps are produced. For pipes, the functional impact of physical damage is manifested in the creation of a leak or break, which leads to the loss of water. Damage to pumps and tanks leads to a loss of hydraulic power. Component repairs are simulated on each damage map using a process-based DES (PBDES) model, which considers the available resources (inspection, repair crew, and materials), and the sequencing (e.g. broken pipe must be inspected before it is repaired) and duration of the associated activities. The Water Network Tool for Resilience (WNTR) [42] is used to perform a pressure-driven hydraulic analysis at each discrete time-steps (e.g. hourly, daily) during the repair process. Therefore, the restoration of hydraulic performance of the system, as measured by the system serviceability index (SSI) [112], is explicitly considered. From these SSI recovery trajectories, several static (e.g. robustness) and temporal (e.g. time to restore 100% of the pre-event water supply) metrics are quantified. By integrating the metric values associated with a given damage map with the occurrence rate of the event that produced it, the full probability distribution of the system performance is obtained.

Recent advancements in computing technology have made it possible to perform large scale simulations such as the ones used in the proposed risk-based assessment framework. Since one end-to-end simulation does not rely on intermediate results from others, the asynchronous parallel

programming paradigm [113] can be adopted. The duration of end-to-end simulation is primarily dominated by the hydraulic simulation, which is performed at each time step over the course of the entire restoration period. Adoption of the asynchronous parallel programming paradigm could result in significant reduction in the total simulation time.

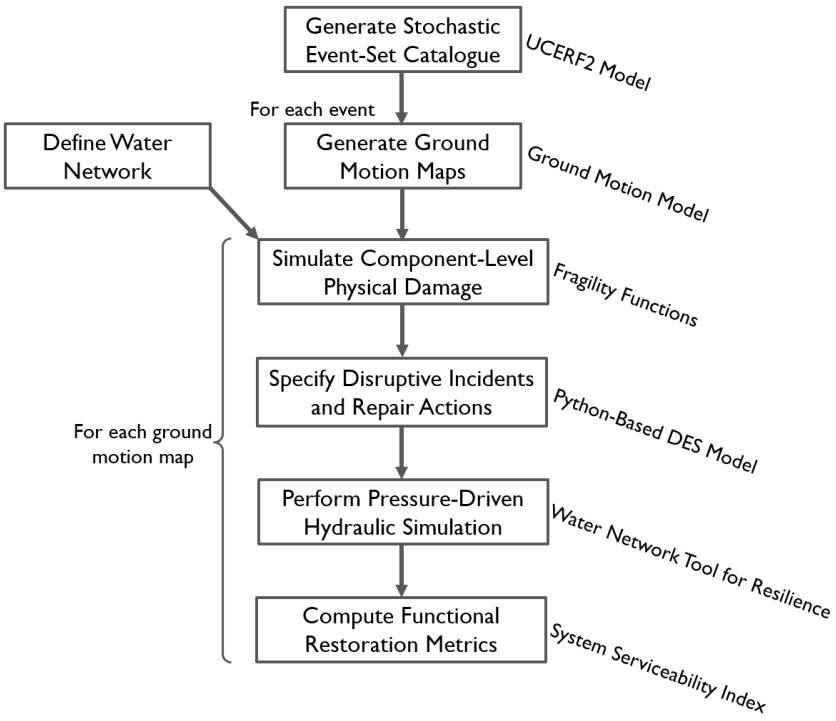


Figure 3-1 End-to-end simulation framework for performing risk-based assessment of post-earthquake functional impacts of a water distribution system

3.2.1 Hazard Characterization

The hazard characterization step produces a set of maps with ground-motion intensity realizations at each location of interest (i.e. the locations of all considered components of the network) and their corresponding occurrence rates. Each map captures the joint probability distribution of the spatially distributed ground motion intensities.

The first sub-step of the hazard assessment is to generate Q earthquake scenarios from a seismic source model, which gives the annual rates at which earthquakes of specified magnitudes, locations, and faulting types will occur. Then, for each earthquake scenario generated by the seismic source model, an empirical GMM is used to model the resulting intensity (Peak Ground Acceleration (PGA) and Peak Ground Velocity (PGV)) at each location of interest. The GMM predicts the log-mean $\left(\overline{ln\, Y}(M_q, R_{iq}, V_{s30,i}, \dots)\right)$, within-event $\left(\sigma_{iq}\right)$ and between-event $\left(\tau_q\right)$

residual standard deviations of the ground-motion intensity, for the i^{th} site ($i = 1, 2, \dots, n$) in the q^{th} earthquake scenario where ($q = 1, 2, \dots, Q$). M_q is the moment magnitude of the q^{th} scenario, R_{iq} is the closest horizontal distance from the surface projection of the fault plane to location i, and $V_{s30,i}$ is the average shear wave velocity down to $30m$ at the i^{th} location. For each of the Q earthquake scenarios, b realizations of the spatially correlated ground-motion intensity residual terms are obtained using an appropriate model. Once the residuals are sampled, the total natural log of the ground motion intensity (Y) is computed as Equation 3.1:

$$\ln Y_{ij} = \overline{ln\, Y}(M_q, R_{iq}, V_{s30,i}, \dots) + \sigma_{ij}\varepsilon_{ij} + \tau_j\eta_j \quad (3.1)$$

where j is the ground-motion intensity map index ($j = 1,2, \dots, M\ where\ M = Q \times b$), ε_{ij} and η_j is are the normalized within-event and between-event residuals for $ln\, Y$, respectively. ε_{ij} reflects the location-to-location variability and η_j represents event-to-event variability. Both ε_{ij} and η_j are normal random variables with zero mean and unit standard deviation. The vector of ε_{ij} is modeled as a spatially correlated multivariate normal distribution and η_j is a standard univariate normally distributed random variable. The result is a set of M ground-motion intensity maps. Since an equal number (b) of ground-motion maps is produced for each earthquake scenario, the annual rate of occurrence of the earthquake scenario normalized by b is represented as w_j.

3.2.2 Damage modeling

Calculating network performance requires an assessment of the earthquake-induced physical damage to the relevant components of the network. The probabilistic link between the ground-motion intensity and physical damage is provided by fragility functions in the form of $P(DS_\psi \geq ds|Y_{ij} = y)$ where DS_ψ is a discrete random variable whose value represents the damage state for the ψ^{th} component and ds is a damage state threshold of interest. The damage state is conditioned on a realization, y, of the random variable Y_{ij}, which, in this study is the shaking intensity at the i^{th} site corresponding to the j^{th} ground-motion map.

3.2.3 Restoration Modeling

The discrete event simulation (DES) approach [72] is used to conduct the restoration analysis, which mimics the behavior of a system as a sequence of events that occurs at discrete points in time. The core elements of a DES model are entities, attributes, events, resources, and time.

Entities, which are used to represent specific objects within the system, have attributes, experience events, and consume resources over time. Attributes are a set of features that are specific to each entity, the accumulation of which defines the state of that entity at any point in time. Events are occurrences that can affect the state of an entity, and resources are objects that provide services to entities. All entities reside in an environment, where they interact with events, processes, and resources. The PBDES approach utilizes a set of processes, which represent a sequence of events and activities through which a specific object moves. In the PBDES restoration model, the water network components (e.g. pipes, tanks, pump stations) are the entities, their damaged and/or functional states are the attributes, and the inspection and repair crews are the resources. The processes include network component inspection, repair, and changes in the repair crew.

Given the damage map, the first step in the PBDES water network restoration model is the initiation of component (e.g. pipes, tanks, and pump stations) inspections, which is modeled as a process. Inspection crew members, which represent a type of resource, travel to various component locations prioritized by some metric (e.g. diameter of pipes, capacity of tanks) within the network and perform inspections within some duration. The start and end of the inspection process represent two events. If an inspected component is found to be damaged, it is added to a repair process queue. Once the component gets to the "front" of the queue, its repair takes place over some duration. Similar to inspection, the beginning and end of the repair process for a component are represented as events.

Hydraulic Simulation

The restoration process for the water distribution system ultimately seeks to satisfy the water demand for households and other critical infrastructure. Because the system comprises a complex network of components, individual repairs do not always increase the overall water supply. To ensure that the model reflects this, a hydraulic simulation of the water network is performed at each restoration time step to quantify the availability of water at each demand node. This hydraulic simulation output combined with the component's recovery information can enable stakeholders to make informed decisions and design more effective pre-mitigation strategies. More specifically, a pressure-driven analysis is performed using Water Network Tool for Resilience (WNTR) [42]. Damage to a pipe is modeled by assigning a leakage discharge, which assumes that there is a hole in the pipe with a specified diameter. Similarly, tank damage is modeled by adding a leak at the

base with an area that is defined in terms of the tank diameter. The pressure reduction in the tank is a function of the leak area. Also, when the pressure in the tank falls below some threshold, it is taken out of operation. Damage to pump stations are represented as an outage, during which the pump is not operational.

3.2.4 Network Performance

The final step in the end-to-end simulation framework is to evaluate the network performance using different resilience measures. Bruneau et al. (2003) defined four dimensions of seismic resilience for civil infrastructure: (i) robustness, which describes the strength, or the ability of elements, systems, and other units of analysis to withstand a given level of stress or demand without suffering degradation or loss of function, (ii) rapidity or the capacity to meet priorities and achieve goals in a timely manner such that losses are contained and future disruptions are avoided, (iii) resourcefulness, which reflects the access to the materials (i.e., informational, technological, physical) and human resources (i.e., labor) needed to respond to a disruptive event, and (iv) redundancy, the extent to which different components of the system are substitutable. The same paper proposed a deterministic static metric for measuring the resilience loss (RL) of a system to an earthquake, which is mathematically represented in Equation 3.2, where t_0 is the time at which the disruption started, t_1 is the time at which the system returns to its normal pre-disruption state and $\mathcal{F}(t)$ is the functionality of the infrastructure at time t.

$$RL = \int_{t_0}^{t_1} [100 - \mathcal{F}(t)]dt \tag{3.2}$$

As functionality is a general concept, the RL is applicable to different systems. A larger RL indicates lower resilience and vice versa. RL can be viewed as a metric that encompasses both robustness and rapidity or a measure of the cumulative loss of functionality. However, it is worth noting that this metric is less sensitive to major changes in the network, and two different combinations of robustness and rapidity can yield the same value of RL. Also, while it is generally applicability, the RL metric could be difficult to communicate to stakeholders and decision-makers, even when given as a percentage and the assumption of 100% functionality before disruption.

Hydraulic reliability quantifies the ability of a water distribution system to supply the required quantity and quality of water at the necessary pressure to the desired locations at the appropriate

times. Calculation of this type of metric requires hydraulic simulation of the system and is based on the spatial and temporal flow of water and/or pressure variables. In the case of disruptive events, pressure-driven analysis is preferred over demand-driven analysis for the calculation of hydraulic reliability metrics [100]. Ostfeld et al. (2002) proposed a resilience metric to quantify the satisfaction of a consumer's demand by calculating the fraction of delivered demand to consumer nodes. This metric, which can also be computed as a function of time and space, is called *nodal serviceability* [45]. Masoomi et al. (2020) adapted the nodal serviceability metric to capture statistics about the population served by calculating the number of people that receive a supply of water that is less than some critical threshold. The concept of nodal serviceability is extended to the network level by computing the ratio of post-earthquake water supply to pre-earthquake demands, which is defined as the *System Serviceability Index* (SSI) [45].

3.3 Case Study

3.3.1 Description of the Napa Water Network

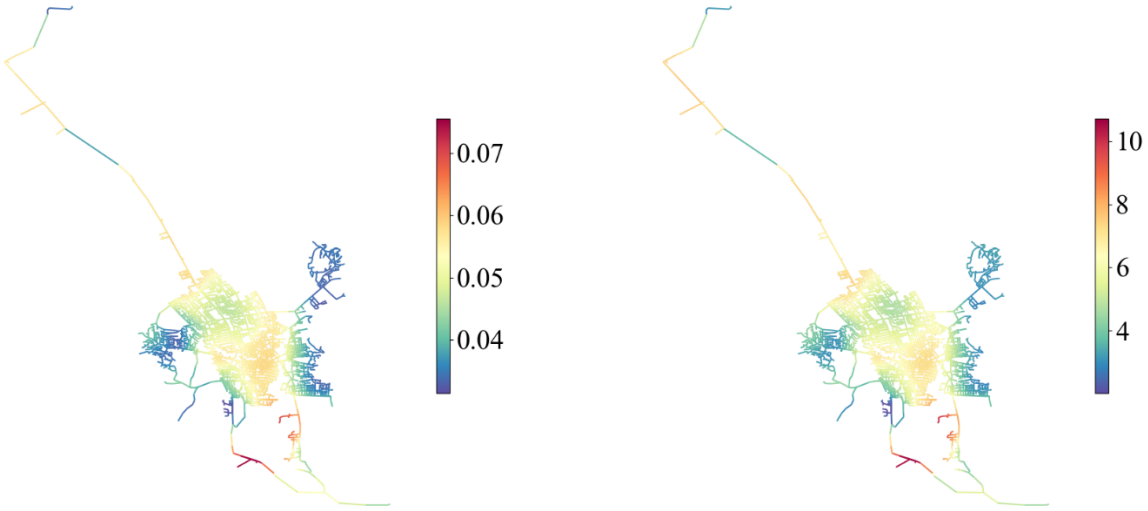
Readers are referred to Section 2.2 for a detailed discussion on the Napa water network and the damage and recovery following the 2014 South Napa earthquake.

3.3.2 Generating the Earthquake Stochastic Event Set Catalogue for the Napa Region

The suite of possible earthquake scenarios that would affect the Napa region was generated using the Uniform California Earthquake Rupture Forecast (UCERF2) model [116–118]. The Napa region is divided into 285 (1.45 km X 1.85 km) (1 minute) grids, which overlay the water network, and the center of each grid is taken as a location of interest. The OpenSHA IM Event Set Calculator [88], an open-source Java-based platform for conducting seismic hazard analysis, is used to generate a total of $Q = 2{,}343$ scenarios.

For each earthquake scenario, the ground motion models (GMM) is used to obtain the log-mean and within-event and between-event residuals for PGA and PGV (e.g. Abrahamson et al. 2014; Boore et al. 2014; Campbell and Bozorgnia 2014; Chiou and Youngs 2014; Idriss 2014). For PGA, a weight of 0.12 is applied to the Idriss model and 0.22 to all others as recommended in the 2014 U.S. National Seismic Hazard Maps and for PGV, equal weight of 0.25 is applied to all models except Idriss [123]. The Jayaram and Baker (2009) algorithm is used to generate $b = 50$ realizations of the spatially correlated ground-motion intensity residuals for each scenario. The

resulting event set consists of $M = 117{,}150$ ground-motion intensity maps. Figure 3-2 shows the median-value spatial distribution of PGA and PGV considering all spatially correlated maps and all scenarios. The associated PGA and PGV values range from 0.0317 g to 0.0741 g and from 2.83 cm/s to 10.66 cm/s, respectively.



(a) (b)

Figure 3-2 Spatial distribution of median (a) PGA (in g) and (b) PGV (in cm/s) values considering all 117,150 ground motion maps

3.3.3 Damage modeling

A damage map is generated by sampling the damage state for each component, with probabilities obtained from the fragility functions conditioned on the ground-motion intensity. For pipes, the fragility functions developed by [103] are used to estimate damage state conditioned on PGV and the repair rate (RR). For tanks and pump stations, fragility curves from HAZUS-MH [86] are used to estimate the damage state conditioned on PGA. More specifically, for each scenario, a spatially correlated ground motion map is chosen randomly from the set of 50, and the corresponding damage map is obtained. This process is repeated 1000 times for each scenario resulting in total of $N_d = 2{,}343{,}000$ damage maps.

Of the 2,343 scenarios, only 135 caused damage to the water network. These 135 scenarios occurred on the Hayward-Rodgers Creek, North San Andreas, Great Valley, Hunting Creek-

Berryessa, Maacama-Garberville, Mount Diablo Thrust, and West Napa fault. Figure 3-3 shows the magnitude distribution of the damaging scenarios. Most magnitudes range between 6.75 – 7.0 followed by 7.0 – 7.25. The median PGA (in g) and PGV (cm/s) across all ground motion maps associated with the damaging scenarios are shown in Figure 3-4. These median intensities are obtained by taking mean of intensities corresponding to same scenario and then median across all damaging scenarios. For the damaging scenarios, the median PGA value ranges from 0.092 g to 0.211 g and the median PGV value ranges from 8.37 cm/s to 34.94 cm/s. The spatial distribution of mean pipe repair rate (RR) per km across all the ground motion maps corresponding to the damaging events is shown in Figure 3-5. A **M** 6.7 on the West Napa fault resulted in the highest number of damaged pipes. With an annual exceedance rate of 8.288E-5, this event damaged 559 pipes (458 leaks and 140 breaks), 11 tanks (10 with slight damage and 1 with moderate damage). There was no damage to pump stations.

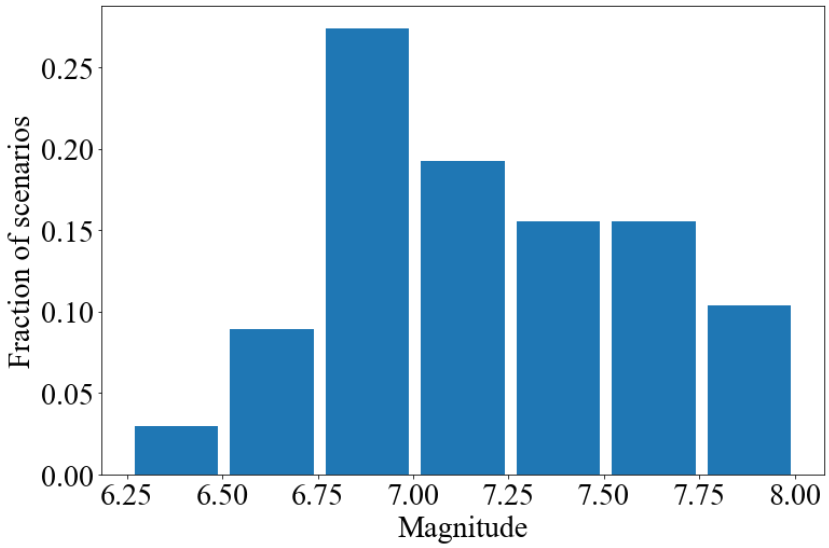
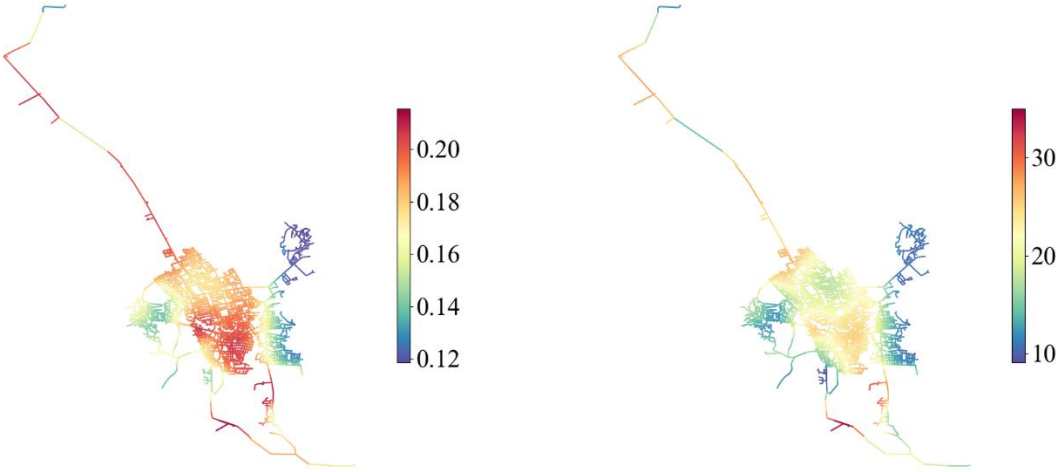


Figure 3-3 Magnitude distribution for damaging scenarios



(a) (b)

Figure 3-4 Spatial distribution of median (a) PGA (in g) and (b) PGV (in cm/s) values considering the 135 scenarios that caused damage

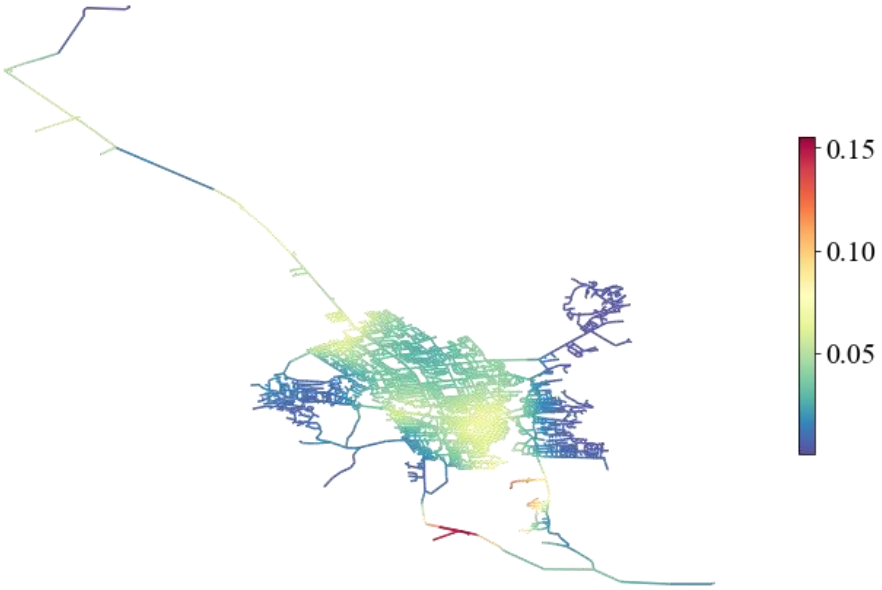


Figure 3-5 Mean pipe repair rate per km distribution

3.3.4 Restoration Modeling and Network Performance

The DES modeling approach, discussed in Section 2.6.3, is used to the simulate recovery of Napa water system. The uncertainty in the inspection and repair processes are captured by assigning probability distributions and associated parameters (e.g., central tendency and dispersion) to the

number of crews and event durations. The inspection and repair event durations and the initial distribution of crew members are taken from previous studies [37,86,103] and discussed in Section 2.6.3.

By coupling the DES model with the WNTR platform, an SSI restoration curve is generated for each damaging ground motion map. More specifically, a total of 2,343,000 SSI restoration curves are produced. Several studies have provided guidelines for developing infrastructure resilience goals based on different types of metrics (e.g. robustness, rapidity and functionality) (e.g. Cauffman 2015; Chang and Shinozuka 2004; Poland 20s09). In the current study, three performance metrics are extracted from each SSI curve. The robustness metric is taken as the immediate loss of SSI following the event (SSI_{IL}) which is defined on a scale from 0 (no loss) to 100% (complete SSI loss).The rapidity metric is taken as the time to restore 100% of the pre-earthquake SSI and is denoted as $T_{100\%SSI}$. The cumulative loss of service (SSI_{CL}) is measured as the area under the SSI curve (i.e. the resilience loss per Bruneau et al. (2003)).

Figure 3-6 shows a histogram of SSI_{IL} based only on the damaging events. The SSI_{IL} values range from approximately 0% to 88%. The 59,886 that resulted in water service disruption accounts for approximately 2.5% of the stochastic event set. A histogram of the time to restore 100% SSI is shown in Figure 3-7. Similar to Figure 3-6, the vertical axis values (fraction of cases) are only based on the damaging events. Note that the bins of the histogram are unevenly distribution to highlight the common periods of interest to decision-makers (e.g. 3 days, 1 week, 2 weeks, 1 month, 3 months, 6 months following the earthquake). The $T_{100\%SSI}$ values in Figure 3-7 range from 0 to 220 days, which indicates that, for some of the worst-case scenarios, the water service can remain disrupted for up to 7 months. The Applied Technology Council (ATC) suggested a performance target of 100% water service restoration within 3 days following a "probable" earthquake [127]. It is worth noting that 2,246,016 damage maps (~95.7% of all damage maps) meet this criterion. Figure 3-8 shows a histogram of the SSI_{CL}, which is taken as the area above the restoration curve. A 100% value corresponds to a complete loss of service over the restoration period and a 0% value means no loss. The SSI_{CL} values ranged from 0% to approximately 56% with a mean value (considering only the damaging events) of 1.31%.

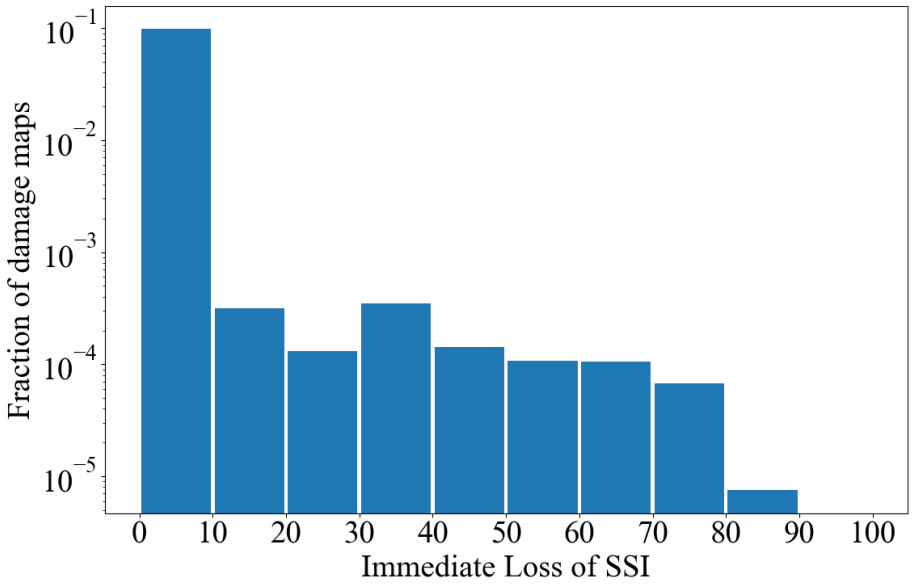


Figure 3-6 Histogram showing the distribution of SSI_{IL} for the damaging events

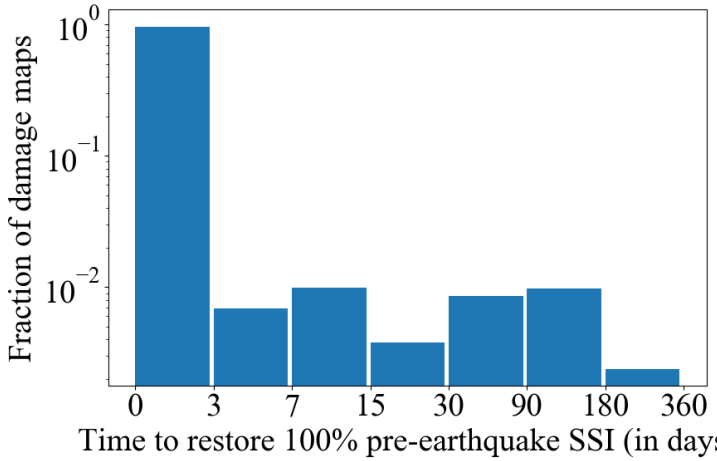


Figure 3-7 Histogram showing the distribution of $T_{100\%SSI}$ for the damaging events

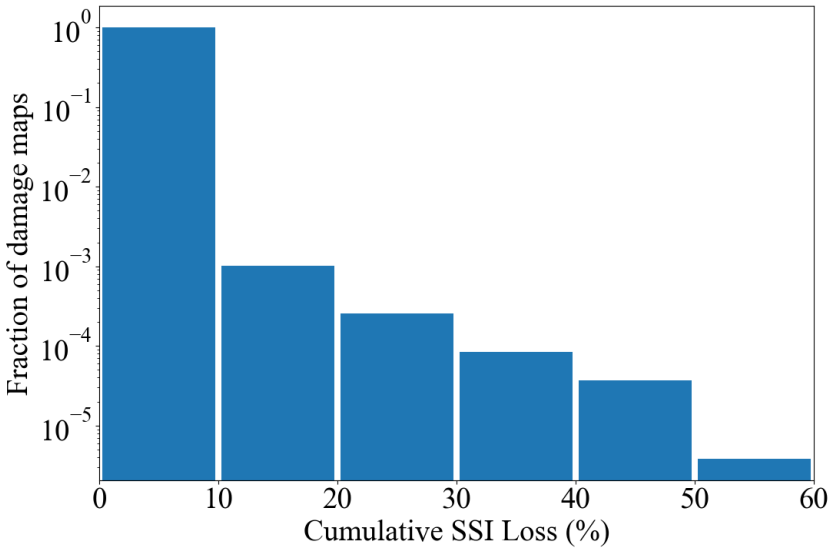


Figure 3-8 Histogram showing the distribution of SSI_{CL} for the damaging events

3.4 Disaggregation of Dispersion by Source of Uncertainty

Disaggregation of the network performance dispersion is an effective way to identify which random variables contribute the most to the uncertainty bounds of the predicted outcomes. Within the end-to-end simulation framework, there are several sources of uncertainties embedded in different stages (Figure 3-9). In the hazard characterization, there is uncertainty in the earthquake scenario (addressed by considering all possible earthquake scenarios) and the spatial distribution of shaking intensity conditioned on a single scenario (addressed by generating multiple spatially correlated ground motion maps). In the damage assessment, multiple maps are generated to propagate the uncertainty embedded in the fragility functions. Within the DES restoration model, the initial number of inspection and repair crews, temporal input parameters (i.e., inspection and repair times) and the availability of repair crews, are all considered as random variables.

Four sources of uncertainty are considered in the disaggregation: spatial distribution of shaking conditioned on the scenario, component damage fragility, the available crew members and the process durations. For a given scenario, the end-to-end simulation is performed by considering the uncertainty in the random variable of interest while using the median values for all others. An **M** 6.85 event occurring on the West Napa fault with an annual rate of occurrence of 3.01E-5 is considered for this purpose.

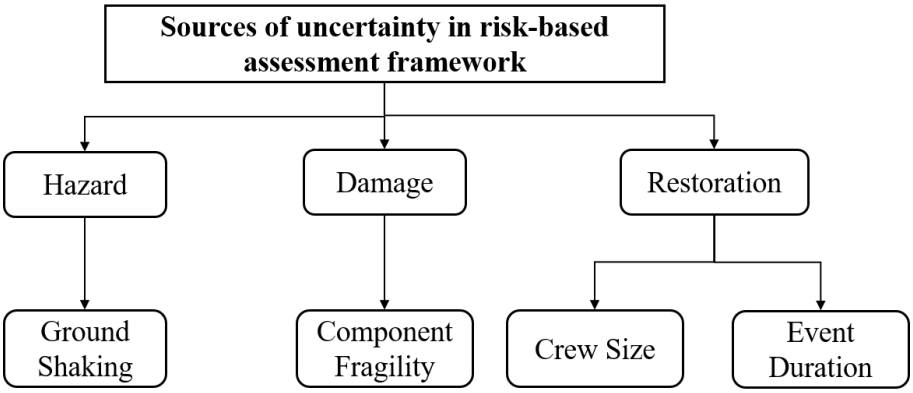


Figure 3-9 Sources of uncertainty considered in the risk-based assessment framework

To quantify the uncertainty associated with each type of parameter, 100 end-to-end simulations are performed. The mean plus and minus one standard deviation SSI curves when all sources of uncertainty are considered is shown in Figure 3-10 and Table 3-1 shows the associated mean (μ) and coefficient of variation (δ) values of the three network performance metrics. $T_{100\%SSI}$ has the highest dispersion value as measured by the coefficient of variation (δ = 0.82) and SSI_{IL} has the lowest (δ = 0.18). This result suggests that the dispersion associated with metrics related to recovery time are generally higher than the ones that only consider the initial loss of service. Additionally, the metric that considers both the initial loss of service and recovery time have a higher dispersion than initial loss of service metric but lower than the recovery time metric.

Table 3-1 Mean (μ) and coefficient of variation (δ) of the network performance metrics for a **M** 6.85 earthquake scenario with all sources of uncertainty considered

Performance Metric	Statistic Type	Statistic Value
SSI_{IL}	μ	65.6%
	δ	0.18
$T_{100\%SSI}$	μ	71.9 days
	δ	0.82
SSI_{CL}	μ	13.1%
	δ	0.69

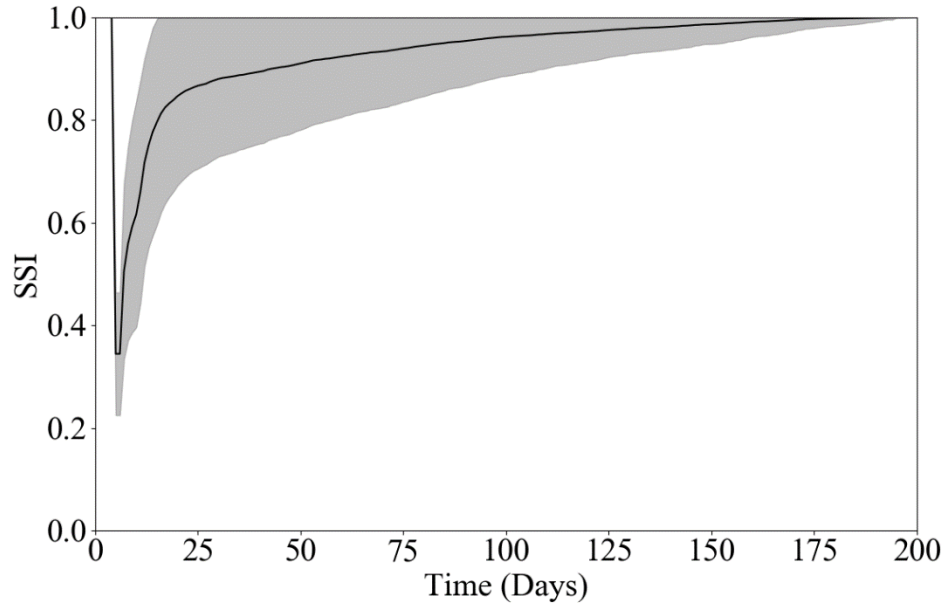


Figure 3-10 Mean and one-standard deviation bounded SSI curve for the **M** 6.85 earthquake scenario with all sources of uncertainty considered

The 100 end-to-end simulations are repeated while considering the four sources of uncertainty (ground shaking, component damage, crew size and duration parameters) individually. Figure 3-11 shows the associated SSI curves and the mean (μ) and coefficient of variation (δ) values for the time required (in days) to reach 25%, 50%, 75%, 100% of the pre-event SSI levels are shown in Table 3-2. The coefficient of variation values in Table 3-2 and plots in Figure 3-11 indicate that the crew size contributes the most to the overall uncertainty bounds for restoration times associated with higher pre-event SSI values (75% and 100%). Whereas for lower pre-event SSI levels (25% and 50%), the overall uncertainty in the restoration time is dominated by the ground shaking intensity. In general, the dispersion in the restoration time increases with the pre-event SSI level. For the process duration variables, Figure 3-11 shows that most of the uncertainty is in the tail end of the restoration curve. Further evidence of this can be found by comparing the coefficient of variation for the 100% SSI pre-event restoration time (0.88) to the next highest value (0.06), which corresponds to the 75% SSI level.

Table 3-2 Mean (μ) and coefficient of variation (δ) values for the time to restore 25%, 50%, 75%, and 100% of the pre-event SSI levels for the **M** 6.85 earthquake scenario

Percentage of Pre-Event SSI	Statistic	Restoration Time (days) Conditioned on Source of Uncertainty			
		Ground Shaking	Component Damage	Crew Size	Process Duration
25	***μ***	4.9	4.8	4.7	4.9
	δ	0.20	0.07	0.15	0.00
50	***μ***	7.9	6.9	8.4	11.3
	δ	0.33	0.11	0.31	0.05
75	***μ***	10.8	10.8	38.6	12.5
	δ	0.55	0.15	1.37	0.06
100	***μ***	102.7	108.7	138.0	115.3
	δ	0.50	0.33	1.05	0.88

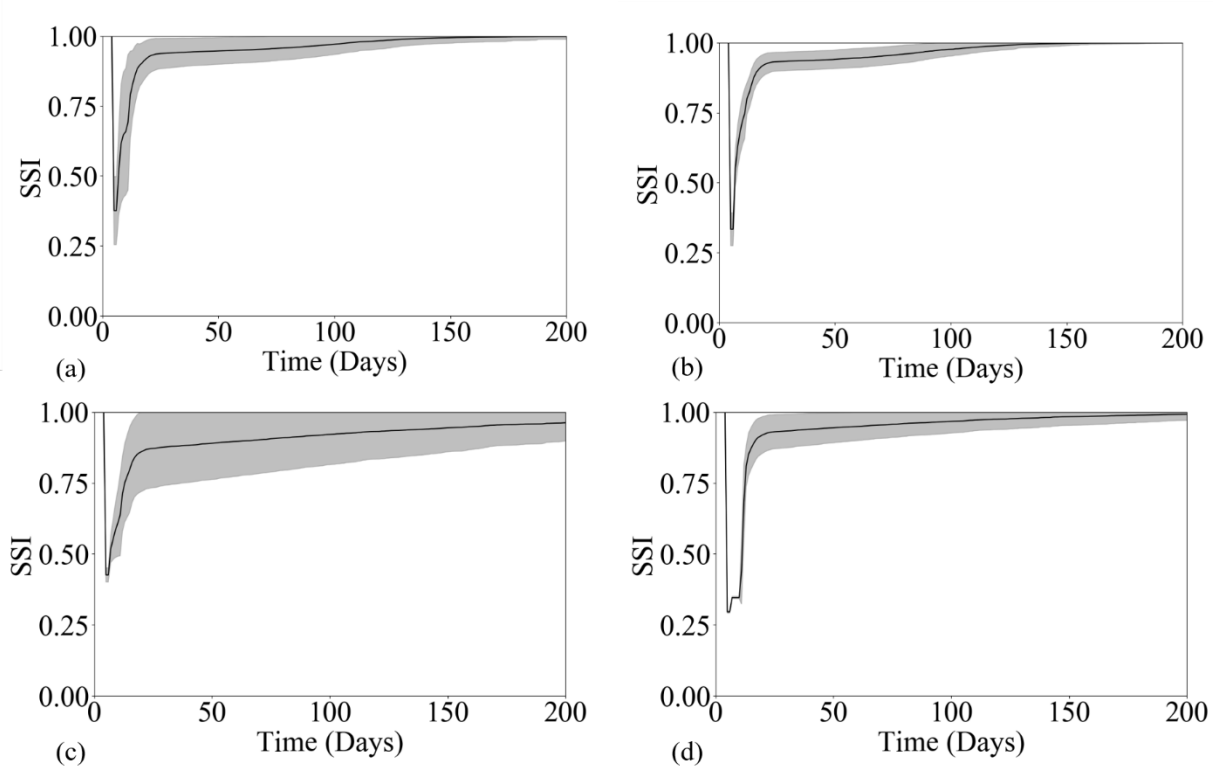


Figure 3-11 Mean and one-standard deviation bounded SSI curve for the **M** 6.85 earthquake scenario while considering the uncertainty in the (a) ground shaking intensity, (b) component damage fragility, (c) crew sizes and (d) event duration parameters

3.5 Annual Exceedance Rates

After calculating the network performance for all simulated ground-motion maps, the annual exceedance rates corresponding to each metric is calculated using Equation 3.3

$$\lambda(\hat{p}) = \sum_{j \in M} w_j \mathbb{I}[p_j \geq \hat{p}] \tag{3.3}$$

where $\lambda(\hat{p})$ is the annual exceedance rate for a specified network performance level $\hat{p}$, M is the set of all damage maps, w_j is the weight (annual occurrence rate) of damage map j, p_j signifies the network performance resulting from damage map j, and $\mathbb{I}[\blacksquare]$ represents an indicator function which is equal to 1 when the network performance resulting from damage map j exceeds the specified network performance level, and 0 otherwise. Figure 3-12, Figure 3-13 and Figure 3-14 show the annual exceedance rates for SSI_{IL}, $T_{100\%SSI}$, SSI_{CL}, respectively. These exceedance rates can be used to evaluate time-based performance targets i.e. probability of exceedance over some duration. An example of a time-based target is that a desired performance level is maintained within a duration t (e.g. 100 years). By assuming that the failure to meet this target is an event that follows a Poisson distribution, the associated probability is computed using Equation 3.4:

$$P(T < t) = 1 - e^{-\lambda(\hat{p})t} \tag{3.4}$$

A more specific hypothetical performance target is that the immediate post-event loss of SSI does not exceed 50% within a 100-year period. Using the $\lambda(\hat{p})$ in Figure 3-12 that corresponds to $SSI_{IL} = 50\%$, the probability of not meeting that target is computed to be 6%. Similarly, there may be a desire to limit $T_{100\%SSI}$ to less than 6 months within a 100-year period. From Figure 3-13, the probability of not meeting this target is computed to be 6%. Using the annual exceedance relationship in Figure 3-14, the probability that the cumulative SSI loss exceeds 20% ($SSI_{CL} < 20\%$) over a 100-year duration is 1%.

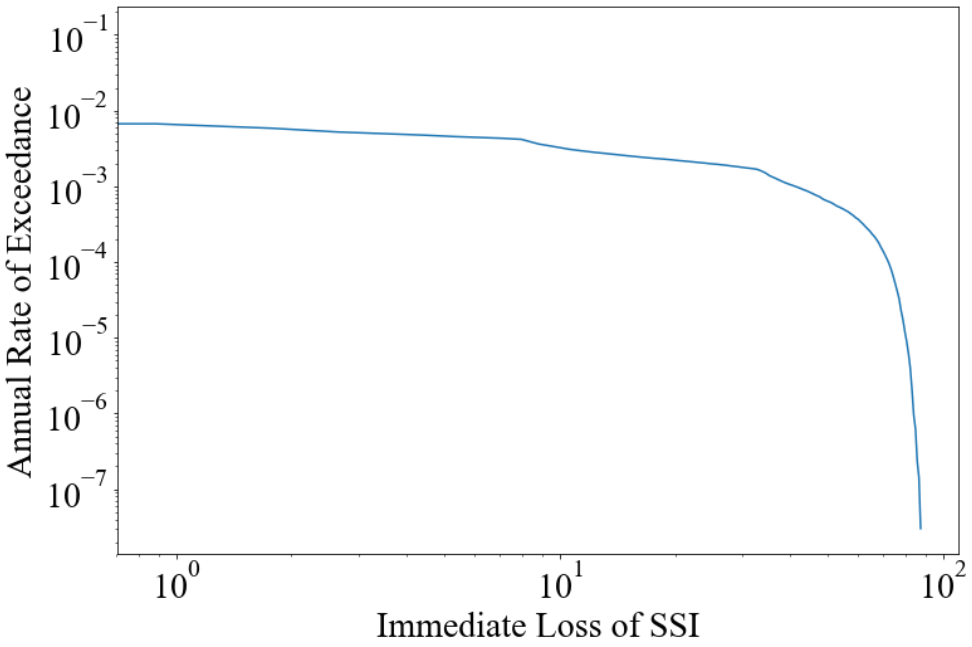


Figure 3-12 Annual exceedance curve for SSI_{IL}

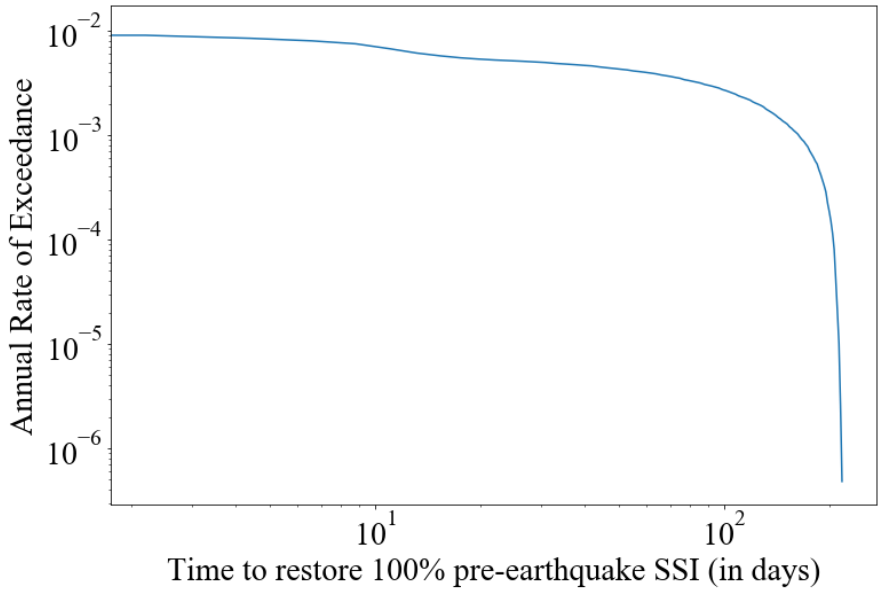


Figure 3-13 Annual exceedance curve for $T_{100\%SSI}$

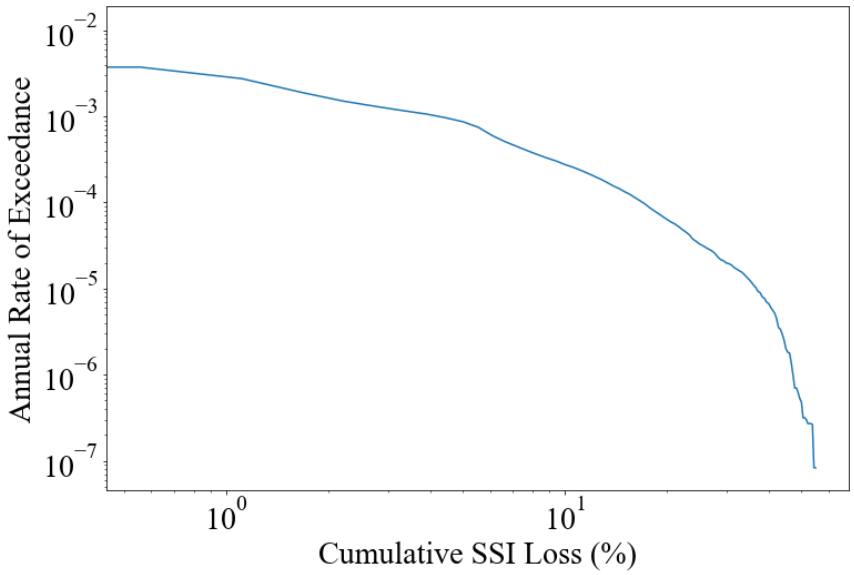


Figure 3-14 Annual exceedance curve for SSI_{CL}

3.6 Ground Motion Intensity and Damage Map Selection using Multi-Metric Optimization

Event-based methods are commonly used to assess the seismic risk to distributed infrastructure systems because the relationship between the earthquake hazard and system performance metrics cannot be expressed analytically. To reduce the computational expense associated with event-based methods Miller and Baker (2015) proposed an optimization-based approach to selecting a subset of damage and corresponding ground-motion intensity maps that provides a reasonable estimate of the full probability distribution of a single network performance. However, as noted earlier, there is often a desire to characterize the performance of a distributed system using multiple metrics. This section discusses an extension of Miller and Baker optimization-based map selection algorithm that includes multiple network performance metrics in the objective function.

Map Selection Problem

The underlying goal of the map-selection problem is to choose a subset of k damage maps and corresponding ground motion fields, each with an adjusted annual rate of occurrence w'. From an optimization perspective, the difference between the network performance (as quantified by the threshold exceedance curves) obtained from the subset k and the complete set of N_m damage maps, should be minimized. More specifically, the multiple performance metric optimization problem can be expressed as:

minimize

$$\alpha_i \sum_{i=1}^{c} \|diag(\boldsymbol{\lambda_i})^{-1}(\boldsymbol{\lambda_i} - \boldsymbol{\Psi_i w})\|_1 \quad (3.5a)$$

subject to

$$\|\boldsymbol{w}\|_1 \leq w_{total} \quad (3.5a)$$

$$\mathbf{0} \leq \boldsymbol{w} \quad (3.5b)$$

where each element of $\boldsymbol{w} \in \mathbb{R}^{M \times 1}$, $w_{j'}$, represents the adjusted annual occurrence rate for the j' damage map, $\|\blacksquare\|_1$ is the L^1-norm or sum of absolute values of the vector $\blacksquare$, w_{total} is the sum of the original occurrence rates of all M damage maps, $\mathbf{0} \in \mathbb{R}^{M \times 1}$ is a vector of zeros, $\boldsymbol{\lambda_i}$ is a vector of annualized exceedance rates for the i^{th} network performance metric corresponding to R discretized return period values. $\boldsymbol{\Psi_i} \in \mathbb{R}^{R \times M}$ is a matrix of constants with each entry $\psi_{r,j'} = \mathbb{I}\{p_{j'} \geq \hat{p}_r\}$, α_i is a scalar value that controls the weight assigned to the i^{th} network performance metric and ν is the total number of considered metrics.

The linear objective function in Equation 3.5a minimizes the difference between the exceedance rate curves for the baseline set and the selected subset while considering all network performance metrics. The first constraint in Equation 3.5b specifies that the sum of $w_{j'}$ should be less than or equal to w_{total}. In other words, the sum of occurrences of all maps in the selected subset should be less than or equal to the sum of occurrences of all maps in the baseline set. The second constraint in Equation 3.5b enforces non-negativity to every value of $w_{j'}$, which is consistent with the physical interpretation of annual occurrence rates. The objective function and constraints are linear in terms of $\boldsymbol{w}$, therefore an optimal solution is guaranteed to exist [128]. CVXPY [129], a python-embedded modeling language for convex optimization is used to solve the linear programming problem. The map subset obtained after solving above the optimization problem may result in more than k non-zero values. The heuristic method [111,130] is used to select the k largest weights, which are renormalized such that the sum of normalized weights equal to w_{total}. To evaluate the error in the network performance metric exceedance curves between the subset and baseline set, the Mean Performance Measure Curve Error (MPMCE) proposed by Miller and Baker (2015) is used shown in Equation 3.6.

$$MPMCE = \frac{1}{R}\sum_{r=1}^{R}\left|\frac{p_r - \hat{\hat{p}}_r}{p_r}\right| \quad (3.6)$$

where $\hat{\hat{p}}_r$ is the estimated performance metric value using the selected subset of maps for the r^{th} return period.

The 2,343,000 damage maps for the Napa water network is used as the baseline and a subset of 1000 maps is selected using multi-objective optimization with equal weights assigned to all performance metrics. Table 3-3 tabulates the MPMCE values obtained from optimizing on a single and multiple performance metrics with testing on the remaining others. The values suggest that if the subset is optimized for a single metric, for example, SSI_{IL}, the MPMCE value is minimum for the metric used as the basis of the optimization and yields high MPMCE values for the others. On the other hand, if the subset is chosen while considering all network performance metrics, a universally high accuracy reflected by low MPMCE values is obtained.

Table 3-3 Mean performance measure curve error (MPMCE) for various target performance measures

		Optimized For			
	MPMCE (%)	All	SSI_{IL}	$T_{100\%SSI}$	SSI_{CL}
Tested On	SSI_{IL}	1.31	1.31	1.32	1.31
	$T_{100\%SSI}$	3.47	3.54	3.48	3.61
	SSI_{CL}	3.47	3.47	3.48	3.47

3.7 Summary and Conclusion

A risk-based framework for evaluating the post-earthquake functional loss and restoration of a water distribution system using a regional stochastic event set catalogue is presented. The four main stages of the assessment include characterization seismic hazard, physical damage, recovery or functional restoration and network performance. Hazard characterization produces a set of maps that capture the joint distribution of shaking intensity at the locations of interest. By coupling these maps with fragility functions in a Monte Carlo simulation procedure, the damage to the individual components of the network is assessed. A functional restoration trajectory is generated for each damage map using process-based discrete event simulation (PBDES) and pressure-driven hydraulic analysis and multiple metrics are used to assess the network performance. Uncertainties

in the ground motion intensity, component damage, inspection/repair crew availability and the temporal restoration parameters are propagated through the various stages of the framework. A methodology for selecting a subset of damage and ground motion maps using multi-objective optimization is also established.

The proposed methodology was used to conduct a risk-based assessment of the city of Napa's water distribution system. A stochastic event set catalogue comprised of 2,343 scenarios was assembled using the UCERF2 seismic source model. For each scenario, 50 spatially correlated ground motion maps were generated to obtain shaking intensities at the locations of the individual components that make up the water network. Using Monte Carlo simulation, a set of 1000 damage maps is produced for each scenario using the randomly sampled (from the set of 50) ground motion maps. The damaged state of each component (pipes, tanks and pump stations) in the network is represented each damage map. Only 135 scenarios with magnitudes ranging from 6.25 to 8.0 caused damage to the Napa water network.

The component inspection and repair process for each damage map is simulated using a PBDES model while explicitly considering the available resources (inspection and repair crew), sequencing and duration of the associated activities. A pressure-driven hydraulic simulation was performed at regular time intervals over the entire duration of the restoration process to quantify the functional recovery in terms of the system serviceability index (SSI). This formed the basis of the network performance assessment using metrics related to the initial loss of SSI (SSI_{IL}), the time to restore the pre-event SSI ($T_{100\%SSI}$) and the cumulative SSI loss (SSI_{CL}). The results indicated that approximately 2.5% of the damage maps from the entire stochastic event set catalogue resulted in post-earthquake water service disruption. Additionally, 95% of the damage maps achieved full SSI restoration within 3 days i.e. $T_{100\%SSI} \leq 3$ days.

Disaggregation of the dispersion in the network performance outcomes was used compare the relative contribution of different sources of uncertainty including ground shaking intensity, component fragility, crew size and event duration. When all sources of uncertainty were considered, the dispersion associated with recovery time metrics (e.g. $T_{100\%SSI}$) was found to be higher than that of SSI_{IL} and SSI_{CL}. The results from the disaggregation showed that the ground shaking intensity had the highest contribution to the dispersion in the restoration time corresponding to lower SSI levels (e.g. time to restore 25% and 50% of pre-event level). Whereas at higher SSI levels (75% and 100% of pre-event level), the number of crew members dominated

the dispersion in the restoration time. In general, the dispersion in restoration time increases with the SSI level.

We also discussed an extension of Miller and Baker optimization-based map selection algorithm that includes multiple network performance metrics in the objective function. Through the case study of Napa water distribution system, we have demonstrated how to use the optimization formulation to select a subset of damage maps (with corresponding ground-motion intensity maps and occurrence rates) from a larger set of candidate maps. We have also shown that the results from the subset are a good estimate of the results from an extensively sampled baseline set of maps.

This study has advanced past studies on the seismic impact to water distribution systems by including performance measures based on restoration hydraulic analysis in a risk-based assessment framework. Future work can extend the framework to incorporate multiple hazards, operational adaptations, other lifeline systems including their interaction and interdependence, other network performance indicators (like centrality measures) and economic loss assessment.

CHAPTER 4 Active Learning Framework for Risk Assessment of Distributed Infrastructure Systems

4.1 Introduction

Critical infrastructure systems such as electric power, water, gas and transportation networks (so-called lifelines) play a vital role in modern society by supporting all sectors of the economy. A common feature of lifeline systems is that they are often comprised of many physically and functionally interdependent components that are distributed over large geographic areas. As such, damage to one or a few components can have cascading effects on the functionality of the system as whole. Natural hazards are one of the primary threats to distributed infrastructure systems. Hazards that have a relatively small spatial footprint (e.g. tornados) can damage one or a few components that may or may not be critical to the system level performance. On the other hand, hazards that have regional scale impacts (e.g. earthquakes) can damage large numbers of components and are therefore more likely to significantly disrupt the functionality of the system. Regardless of the type of event, risk assessment is key to mitigating the disruptive effects of natural hazards on infrastructure systems.

The main steps used to assess the risk to a distributed infrastructure system include, hazard characterization, component-level damage assessment and network-level performance quantification. The metrics used to quantify the performance of the system can be static (e.g. loss of service immediately following the hazard event) or dynamic (e.g. time to restore full service). The latter requires models that can capture the time-dependent effect of repair activities on network performance. One or a few scenarios can be used as the basis for the hazard characterization. This approach is desirable because it reduces the computational demands of the overall risk assessment. However, the selection of the event(s) to use in the assessment is often a challenge because, for spatially distributed infrastructure systems, it is often impossible to find one or even several such events. Additionally, even if a single “worst-case” (in terms of consequences) event is determined, its probability of occurrence should ideally be considered. Hazards can also be defined using a stochastic catalogue where the full spectrum of possible scenarios and their associated occurrence probabilities (or rates) are considered [105–110].

When performing stochastic event set assessments, the uncertainty in the spatial distribution of loading (e.g. ground shaking intensities for an earthquake or wind speed for a hurricane) is considered by employing a Monte Carlo approach. More specifically, for each event, multiple loading realizations are generated. Therefore, when coupled with component damage and network performance simulation (e.g. functional recovery assessments), where additional sources of uncertainty are considered, the overall assessment can become computationally intractable. One strategy that has been proposed to minimize the computational cost of stochastic catalogue assessments is to use optimization to select a subset of loading (in this case earthquake ground shaking intensities) and damage realizations (Miller and Baker, 2015) that provide a reasonable estimate of the network performance. A key limitation of this approach is that the network performance quantification for the complete set of loading and damage realizations is explicitly utilized in the objective function. In other words, the selection of a subset of ground motion and damage maps using optimization requires quantification of the system performance for all realizations, which, in essence, defeats the purpose of the procedure.

This study presents an active learning-based framework to select a subset of loading realizations from a stochastic catalogue. Active learning is used to incrementally query the complete dataset for individual (or subgroups of) data samples, which are selected based on how much they reduce the uncertainty in the prediction outcomes for the full probability distribution of the network performance metric. An important advantage of the active learning approach is that the full risk assessment procedure (hazard characterization, damage assessment and network performance quantification) is only performed on the subset of scenarios that are incrementally queried. To demonstrate the methodology, a risk-based stochastic catalogue assessment of the post-earthquake functional loss and recovery of the city of Napa's water distribution system is conducted. However, it should be noted that the framework is developed and presented in a generalized manner such that it could be applied to other hazards and infrastructure systems. Section 4.2 provides an overview of active learning framework. Section 4.3 discusses the Gaussian process, which is central to utilizing the active learning framework for stochastic catalogue event-subset selection. The details of the active learning framework are presented in Section 4.4 and Section 4.5 describes the use case involving the seismic risk assessment of the city of Napa's water distribution system. Section 4.6 summarizes the key findings from the study and provides recommendations for future research in this area.

4.2 Overview

Active learning is a subfield of machine learning that focuses on the development of computer systems that improve with experience and training. Traditional "passive" learning systems (e.g., linear regression, random forest) propose a hypothesis (i.e. a model) to explain and provide a mapping between the input and output data. On the other hand, an active learning system develops and tests new hypotheses as part of a continually interactive learning process. In other words, the learner is allowed ask questions about (or query) the given data as opposed to only learning from it. More specifically, the learner gets to ask "an oracle" (a human annotator) about the label of certain instances. By asking the right questions, a high level of generalizable accuracy about the mapping can be obtained with a smaller labeled dataset. A model is fitted to the currently labeled subset after each iteration and used to decide which instance(s) should be labeled next. A one-dimensional binary search algorithm (an algorithm that finds the position of a target value within a sorted array) [131] can be viewed as an example of active learning in which the next query location is decided after obtaining an answer for the previous query.

Figure 4-1 shows an overview of active learning framework which is comprised of four main steps. It starts by specifying the learner, which is a model (here, a Gaussian process) fitted to a small number of labeled instances (training subset). Note that the very first training subset can be randomly selected and labeled using the oracle. The learner then samples or queries one or a group of unlabeled instances from the input space or distribution using an appropriate method. Three querying approaches, which include membership query synthesis, stream-based selective sampling, and pool-based sampling, are discussed later in the paper. Depending on their degree of informativeness as measured by the acquisition function (e.g. uncertainty sampling), unlabeled instances are sent to the oracle for labeling, after which those labeled instances are added to the training subset. The learner's model is then updated using the most current training subset. Lastly, a stopping criterion is checked to determine whether to terminate the procedure or continue learning by repeating the previously outlined steps. In other words, if the stopping criterion is satisfied, the learner (i.e., training subset and model) is provided as output, otherwise the next set of unlabeled instances is sampled from the input distribution. From an infrastructure risk assessment perspective, the training subset that is provided as output from the active learning

framework represents the set of hazard scenarios that is used as the basis for the system performance evaluation.

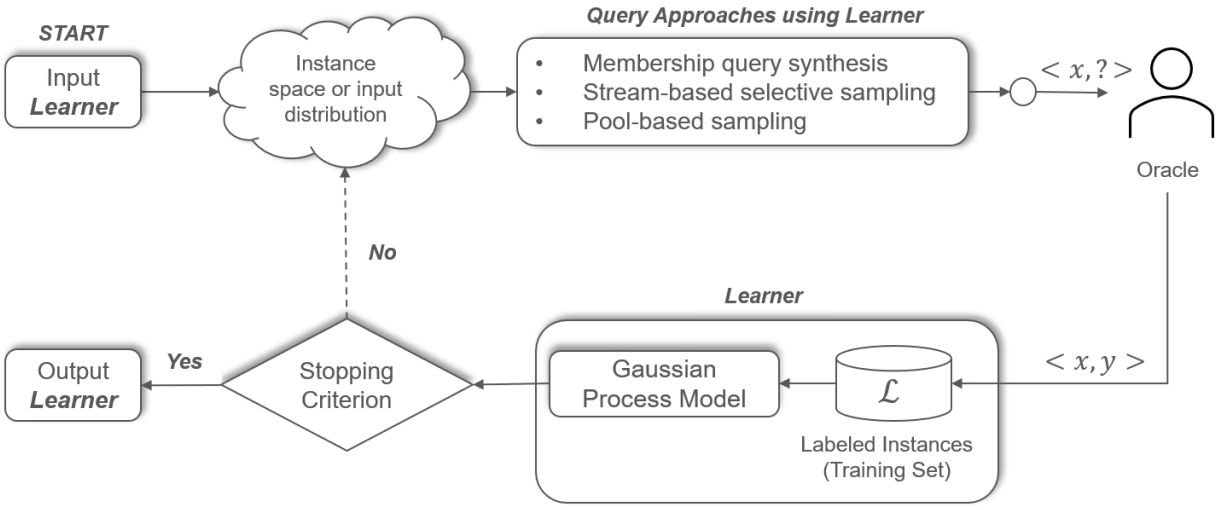


Figure 4-1 Active learning framework

4.3 Gaussian Process

4.3.1 Overview

A Gaussian Process ($\mathcal{GP}$) is a stochastic process (a collection random variables indexed by time or space) with a finite collection of random variables that follow a multivariate normal distribution. $\mathcal{GP}$ is a non-parametric model that enables predictions about a dataset by incorporating prior knowledge. Many real-world problems require an assessment of unknown functions that map inputs to outputs. Such functions may not have an analytical solution or other requirements (e.g. design costs) that complicate the process of information acquisition. In these cases, Gaussian Process Regression ($\mathcal{GPR}$) can serve as a useful tool for performing passive and active inferencing [132]. Whereas passive inferencing seeks to produce the best possible description of a given dataset and make predictions about future instances of that data, the goal of active inferencing is to quickly learn the function while iteratively choosing input points or produce the maximum possible outputs or both. $\mathcal{GPR}$ is a non-parametric Bayesian approach to regression problems that captures the mapping between inputs and outputs by utilizing a theoretically infinite number of parameters and letting the data determine the level of model complexity [133].

Given a training set $\mathcal{D}$ of n observations, $\mathcal{D} = \{(x_i,\ y_i) \mid i = 1, 2, \ldots, n\}$, and letting f denote a mapping (unknown) from inputs $\boldsymbol{x}$ to outputs y $(f: X \rightarrow \boldsymbol{y})$, $\mathcal{GPR}$ can be used to (i) *model* f using x and y, which enables inferencing on the full distribution of y, (ii) *explore* f i.e., actively choose the input points $x_{sub} \in x$ for which the outputs $y_{sub} = f(x_{sub})$ accurately model the function f, and (iii) perform *exploration-exploitation* of f to find inputs that produce the maximum (or minimum) outputs (similar to Bayesian optimization). One interpretation of $\mathcal{GPR}$ is based on a Bayesian analogue to a simple linear regression model where the output is a linear combination of the inputs. This interpretation is described as the *weight-space view* because the inferencing takes place in the space of weights (or predictor coefficients) that are modeled as random variables. Alternatively, a $\mathcal{GPR}$ model can be represented as a distribution over a set of possible functions. This interpretation is described as the *function-space view* because the inferencing takes place directly in the space of functions. The weight-space view and function-space view are further discussed in the next two sub-sections.

4.3.2 Weight-space view

Before presenting the details of the weight-space view modeling function, the Bayesian analysis of a standard linear regression model with Gaussian noise is reviewed as presented in Equation 4.1:

$$f(\boldsymbol{x}) = \boldsymbol{x}^T \boldsymbol{w}, \qquad y = f(\boldsymbol{x}) + \varepsilon \tag{4.1}$$

where $\boldsymbol{x}$ is the input vector and $\boldsymbol{w}$ is a vector of weights. It is assumed that the outputs (y) are a linear function of the inputs plus a noise term (ε) that follows an independent, identical Gaussian distribution with zero mean and variance σ_ε^2 i.e. $\varepsilon \sim \mathcal{N}(0, \sigma_\varepsilon^2)$. Given this assumption, the likelihood function for a set of observations given the parameters can be taken as the product of the probability distributions corresponding to the individual data points, which can be represented as presented:

$$p(\boldsymbol{y}|X, \boldsymbol{w}) \sim \mathcal{N}(X^T \boldsymbol{w}, \sigma_\varepsilon^2 \mathrm{I}) \tag{4.2}$$

In Bayesian formalism, a *prior* is specified, which expresses the beliefs about the parameters before encountering the observations. By specifying a zero mean Gaussian prior with covariance

matrix Σ for the weights i.e. $\boldsymbol{w} \sim \mathcal{N}(\mathbf{0}, \Sigma)$, inferences from the Bayesian linear model are based on the posterior distribution of the weights, which is computed using Baye's rule

$$p(\boldsymbol{w}|\boldsymbol{y}, X) = \frac{p(\boldsymbol{y}|X, \boldsymbol{w})p(\boldsymbol{w})}{p(\boldsymbol{y}|X)} \tag{4.3}$$

where $p(\boldsymbol{w}|\boldsymbol{y}, X)$ and $p(\boldsymbol{w})$ are the posterior and prior distribution of the weights, respectively. The normalizing constant or marginal likelihood is independent of the weights and is expressed as

$$p(\boldsymbol{y}|X) = \int p(\boldsymbol{y}|X, \boldsymbol{w})p(\boldsymbol{w})d\boldsymbol{w} \tag{4.4}$$

The posterior in Equation 4.3 can be written as

$$p(\boldsymbol{w}|\boldsymbol{y}, X) = \mathcal{N}(\widetilde{\boldsymbol{w}} = \frac{1}{\sigma_\varepsilon^2} A^{-1} X\boldsymbol{y},\ A^{-1}) \tag{4.5}$$

where $A^{-1} = \Sigma^{-1} + \sigma_\varepsilon^2 XX^T$. Readers are referred to Williams and Rasmussen (2006) for the complete derivation of Equation 4.5.

The mean of the posterior distribution $p(\boldsymbol{w}|\boldsymbol{y}, X)$ is also described as the maximum a posteriori (MAP) estimate of $\boldsymbol{w}$**.** Since the inferencing is performed using the weights, this formulation is described as "*the weight space view of regression*". To make predictions for a test case (x_*) represented by $f_* = f(x_*) = y_* - \varepsilon_*$, the average of all possible parameter values (which averages out the error term and focuses on the expected value) weighted by their posterior probability is computed. Thus, the predictive distribution for f_* is obtained by averaging the output of all possible linear models with respect to the Gaussian posterior

$$p(f_*|x_*, X, \boldsymbol{y}) = \int p(f_*|x_*, \boldsymbol{w})p(\boldsymbol{w}|X, \boldsymbol{y})d\boldsymbol{w} \tag{4.6a}$$

$$p(f_*|x_*, X, \boldsymbol{y}) = \mathcal{N}(\frac{1}{\sigma_\varepsilon^2} x_*^T A^{-1} X\boldsymbol{y},\ x_*^T A^{-1} x_*) \tag{4.6b}$$

The predictive distribution is also Gaussian with a mean given by the posterior mean of the weights from Equation 4.5 multiplied by the test input. The predictive variance is a quadratic form of the test input with the posterior covariance matrix.

Because of the linearity assumption, Bayesian linear models suffer from limited expressiveness. Nonlinear dependence can be incorporated by mapping the inputs $\boldsymbol{x}$ into some high dimensional space using a set of *basis* functions and then applying the Bayesian linear model to the transformed input. For example, a scalar input x can be projected into the space of powers of x: $\phi(x) = (1, x, x^2, x^3, \ldots)^T$ to implement polynomial regression. Mapping input variables into a feature space offers higher flexibility and allows one to model functions of any shape. However, this flexibility also acts as a drawback since the number of possible mappings is infinite and one must be chosen either a priori or by model comparison within a set of possible mappings. Gaussian process formalism offers a principled solution of choosing the *basis* function implicitly, effectively letting "the data decide" on the complexity of the function.

4.3.3 Function-space view

In Section 4.3.2, the probability distribution of the mapping weights is determined. Since each set of weights represents a particular function, the probability distribution of the weights also represents a distribution of the functions. Therefore, alternatively, $\mathcal{GPR}$ can focus directly on the distribution of the functions instead of the distribution over weights. A $\mathcal{GP}$ defines a distribution over functions such that, if any two or more points in a function are selected, the outputs at these points follow a joint (multivariate) Gaussian distribution. This feature is also described as the marginalization property. More formally, a Gaussian process is defined as a collection of random variables, any finite number of which have a joint (multivariate) Gaussian distribution. A $\mathcal{GP}$ that is denoted by $f(\boldsymbol{x})$ is completely specified by its mean function $m(\boldsymbol{x})$ and co-variance function $k(\boldsymbol{x}, \boldsymbol{x}')$ i.e. $f(\boldsymbol{x}) \sim \mathcal{GP}(m(\boldsymbol{x}), k(\boldsymbol{x}, \boldsymbol{x}'))$. The mean function $m(\boldsymbol{x})$ reflects the expected function value at $\boldsymbol{x}$ i.e. $m(\boldsymbol{x}) = \mathbb{E}[f(\boldsymbol{x})]$ and the covariance function $k(\boldsymbol{x}, \boldsymbol{x}')$ models the dependence between the function values at different input points $\boldsymbol{x}$ and $\boldsymbol{x}'$.

$$k(\boldsymbol{x}, \boldsymbol{x}') = \mathbb{E}[(f(\boldsymbol{x}) - m(\boldsymbol{x}))(f(\boldsymbol{x}') - m(\boldsymbol{x}'))] \tag{4.7}$$

The prior mean function is often set to $m(\boldsymbol{x}) = \mathbf{0}$ to avoid expensive posterior computations and the inferencing is only performed via the covariance function. Empirically, setting the prior to $\mathbf{0}$ is often achieved by subtracting the (prior) mean from all observations. The function $k(\blacksquare)$ is also known as the *kernel* of the Gaussian process. The choice of an appropriate kernel is based on assumptions of smoothness and anticipated patterns within the data. It can be reasonably assumed

that the correlation between two points decays with the distance between the points. A commonly used kernel that is consistent with this assumption is the *radial basis function* (RBF) kernel, which is defined as:

$$k(\boldsymbol{x}, \boldsymbol{x}') = \sigma_f^2 \exp\left(-\frac{\|x-x'\|^2}{2\lambda^2}\right) \tag{4.8}$$

The RBF kernel, also known as the "squared exponential kernel", is an example of a stationary (invariant to translations in the input space) covariance function since it is a function of $|\boldsymbol{x} - \boldsymbol{x}'|$. The two hyper parameters λ (length-scale) and σ_f^2 (signal variance) are tuned to match the priori correlation between input points and consequently the variability of the resulting function. Additional details about different kernels and strategies for choosing the most appropriate one are discussed in Duvenaud (2014) and Williams and Rasmussen (2006). After specifying the mean and covariance functions, the $\mathcal{GP}$ is used to determine priori and posterior function values conditioned on the specified observations.

4.3.3.1 Sampling functions from a $\mathcal{GP}$

Sampling a function from a $\mathcal{GP}$ is generally achieved by computing the function values at a selected set of input points. Theoretically, a continuous function can be represented as a vector of infinite size. However, since only a finite number of predictions are needed, the outputs for these points are obtained from a multivariate normal distribution with a covariance matrix generated by the kernel (i.e., exploiting the marginalization property). Given the chosen set of input points, X_* comprised of $\boldsymbol{x}_i^*, i = 1, \dots, n$, the first step in sampling a function is to compute the covariances between all inputs in X_*, which is expressed as:

$$K(X_*, X_*) = \begin{bmatrix} k(\boldsymbol{x}_1^*, \boldsymbol{x}_1^*) & \cdots & \boldsymbol{k}(\boldsymbol{x}_1^*, \boldsymbol{x}_n^*) \\ \vdots & \ddots & \vdots \\ k(\boldsymbol{x}_n^*, \boldsymbol{x}_1^*) & \cdots & k(\boldsymbol{x}_n^*, \boldsymbol{x}_n^*) \end{bmatrix} \tag{4.9}$$

Then a random Gaussian vector is generated with the above covariance matrix by sampling from a multivariate normal distribution, $f_* \sim \mathcal{N}(\mathbf{0}, K(X_*, X_*))$, where $f_* = [f(\boldsymbol{x}_1^*), \dots, \boldsymbol{x}_n^*]^T$ is a sample of function values. To sample observations of f_*, an additional and independent sample of the noise term ε is added.

4.3.3.2 Posterior predictions from a $\mathcal{GP}$

In the context of active learning, the goal is not to sample random functions from the prior but to incorporate the knowledge that a given set of observations provide about the function. Recalling the set of observations $\mathcal{D} = \{(x_i,\ y_i) \mid i = 1, 2, \ldots, n\}$, which can be labeled as the training set $\mathcal{D}_t = \{X_t, \boldsymbol{y}_t\}$ and with the goal of predicting new inputs X_* or testing observations, f_* is sampled from the posterior distribution $p(f|\mathcal{D}_t)$. As implied by the $\mathcal{GP}$, the training outputs $\boldsymbol{y}_t$ and function values f_* follow a joint (multivariate) normal distribution which can be written as:

$$\begin{bmatrix}\boldsymbol{y}_t \\ f_*\end{bmatrix} \sim \mathcal{N}\left(\mathbf{0}, \begin{pmatrix} K(X_t, X_t) + \sigma_\varepsilon^2 \mathrm{I} & K(X_t, X_*) \\ K(X_*, X_t) & K(X_*, X_*)\end{pmatrix}\right) \tag{4.10}$$

where $K(X_t, X_t)$ is the covariance matrix between all collected observations, $K(X_*, X_*)$ is the covariance matrix for the new points, $K(X_*, X_t)$ is the covariance matrix between the training and testing points, I is the identity matrix and σ_ε^2 is the assumed noise level of the observations. The predictive distribution $p(f_*|X_t, \boldsymbol{y}_t, X_*)$ follows a multivariate normal distribution represented as $f_*|X_t, \boldsymbol{y}_t, X_* \sim \mathcal{N}(\bar{f}_*, cov(f_*))$, where

$$\bar{f}_* \triangleq \mathbb{E}[f_*|X_t, \boldsymbol{y}_t, X_*] = K(X_*, X_t)[K(X_t, X_t) + \sigma_\varepsilon^2 \mathrm{I}]^{-1} \boldsymbol{y}_t \tag{4.11a}$$

$$cov(f_*) = K(X_*, X_*) - K(X_*, X_t)[K(X_t, X_t) + \sigma_\varepsilon^2 \mathrm{I}]^{-1} K(X_t, X_*) \tag{4.11b}$$

The predictive distribution is a $\mathcal{GP}$ with mean function $\bar{f}_*$ and covariance function $cov(f_*)$, which can be used to predict f_*. Readers are referred to Williams and Rasmussen (2006) for complete derivation.

4.3.3.3 Relationship between function-space and weight-space view

To understand the correspondence between function space view and weight space view of $\mathcal{GPR}$, we can switch from function space view to weight space view as follows:

The mean function $\bar{f}_*$ in Equation 4.11a can be rewritten as

$$\bar{f}_*(x) = \sum_{i=1}^{t} w_i k(x_i, x) \tag{4.12}$$

where each x_i is a point in the set of training observations and the weights are computed as $\boldsymbol{w} = [K(X_t, X_t) + \sigma_\varepsilon^2 \mathrm{I}]^{-1} \boldsymbol{y}_t$. Equation 4.12 shows that the $\mathcal{GPR}$ is equivalent to a linear regression

model that uses the basis function $k(\blacksquare, \blacksquare)$ to project the inputs into a high-dimensional feature space. To make new predictions, every output y_t is weighted based on its similarity with the associated input x_t as measured by the kernel. Therefore, the final predictions or new set of observations is simply the weighted sum of the inputs. In other words, a conceptually infinite parameter space is reduced to a finite sum, which is also referred to as the "*kernel trick*". Theoretically, $\mathcal{GPR}$ has as many parameters $\boldsymbol{w}$ as there are observations but making predictions involves only a finite sum over all past observations, which makes $\mathcal{GPR}$ a non-parametric technique.

4.3.3.4 Optimizing hyper-parameters

The covariance function of a Gaussian process typically contains hyper-parameters. For example, λ (length-scale) and σ_f^2 (signal variance) from the radial basis function as well as the noise variance, σ_ε^2, are unknown parameters that are inferred from the data. Since obtaining their posterior distribution is intractable, full Bayesian inferencing of the hyper-parameters is not typically used in practice. Instead, point estimates of the hyper-parameters are obtained by maximizing the marginal (log) likelihood. This is similar to parameter estimation by maximum likelihood and is also referred to as type-II maximum likelihood (ML-II) [132]. Given the data $\mathcal{D} = \{X, \boldsymbol{y}\}$ and hyper-parameters $\boldsymbol{\theta}$ (e.g., $\boldsymbol{\theta} = (\lambda, \sigma_f^2, \sigma_\varepsilon^2)$), the marginal log likelihood is computed as

$$\log p(\boldsymbol{y}|X, \boldsymbol{\theta}) = -\frac{1}{2}\boldsymbol{y}^T K_y^{-1}\boldsymbol{y} - \frac{1}{2}\log|K_y| - \frac{n}{2}log2\pi \tag{4.13}$$

where $K_y = K(X, X) + \sigma_\varepsilon^2 \mathrm{I}$ is the covariance of the noisy output values $\boldsymbol{y}$. The marginal log-likelihood in Equation 4.13 can be viewed as a penalized fit, where $-\frac{1}{2}\boldsymbol{y}^T K_y^{-1}\boldsymbol{y}$ measures the data-fit, $-\frac{1}{2}\log|K_y|$ is the complexity penalty term that only depends on the covariance function and the inputs and $-\frac{n}{2}log2\pi$ is a normalization constant.

4.3.3.5 Alternatives to the RBF kernel

In most real-world applications, the RBF kernel is chosen and its length-scale is optimized to account for potential mismatches between prior smoothness assumptions and the observed data. There are alternatives to the RBF kernel that offer flexibility in terms of smoothness, periodicity,

symmetry, and interaction between variables. In general, more complex kernels can be created by combining simpler ones using operations such as addition or multiplication [132]. For instance, kernels for $\mathcal{GP}$ models can be determined by defining an open-ended space and adding and multiplying kernels from a fixed set [134,135].

4.3.4 Advantages and disadvantages of $\mathcal{GP}$

Gaussian processes are a powerful way to model functions and allows estimation of unknown functions based on exploration and exploitation strategies. $\mathcal{GP}$ allows analytical inferencing i.e., given a kernel function and some observations, the predictive posterior distribution can be computed exactly in close form. The flexibility of choosing a covariance function for $\mathcal{GP}$ allows a wide range of modeling assumptions. The fact that a $\mathcal{GP}$ posterior, given a fixed kernel, enables exact integration over a wide range of hypotheses space means that overfitting is less of an issue relative to other comparable model classes (example, neural networks) and complex optimization and regularization schemes are not needed. Additionally, exact integration over all hypotheses provides a principled way of comparing different models through marginal likelihood of the data given a model. The predictive posterior of a $\mathcal{GP}$ at a set of test points follows a closed-form distribution (multivariate Gaussian distribution).

The are some issues with the $\mathcal{GP}$ that make it difficult to use in specific cases. The main drawback is that computing the matrix inverse in Equation 6b takes $\mathcal{O}(N^3)$ time, which slows the inferencing when there are more than a few thousand points. However, recent approximate inference schemes address this challenge [136,137]. Another issue with $\mathcal{GP}$ is that choosing the kernel for a given problem requires human experts. Readers are referred to Görtler et al. (2019) for a user-interactive exploration of Gaussian processes and the mathematical intuition behind them. The next section on active learning discusses the use of Gaussian process regression to model the underlying function in a dataset and estimate the utility of available queries (candidates to sample for the next set of input points).

4.4 Active Learning

4.4.1 Query Approaches in Active learning

As discussed earlier, one of the first main steps in the active learning framework is selecting or querying the input dataset or distribution. Settles (2009) discusses three main approaches to querying in an active learning context: (i) membership query synthesis, (ii) stream-based selective sampling, and (iii) pool-based sampling. An illustrative comparison of these three querying approaches is presented in Figure 4-2. In the membership query synthesis approach, the learner may request "label membership" (i.e. assignment of a label) for any instance of unlabeled data under the assumption that the definition of the input space is known [140]. In other words, the instance to be labeled may be chosen from the input space or generated by the learner. This approach requires the model to capture the data distribution well enough to generate reasonable instances that would have a clear label. Query synthesis is used in several types of regression tasks such as learning to predict the absolute location of a robot hand given the joint angles of its mechanical arm as inputs [141] and executing autonomous biological experiments to discover metabolic pathways in yeast [142].

In the query synthesis approach, the data-generating distribution is not necessarily considered (and may not even be known). Under such conditions, the active learner might select arbitrary instances that are not useful to the task at hand. For example, in the case of classifying images, if the learner generates an image of pure noise, the oracle won't be able to label it. This issue is addressed in the stream-based selective sampling approach, which assumes that obtaining an unlabeled instance is free (or inexpensive), so it can first be sampled from the input distribution, and then the learner can decide whether or not to request its label [143]. Researchers have used utility metrics [144], region of uncertainty [143] and version space [145] to decide whether to query (label) or discard an instance. The stream-based selective sampling approach has been implemented in real-world tasks such as part-of-speech tagging [144], sensor scheduling [146], and learning to rank functions for information retrieval [147].

Pool based sampling is motivated by the rise of data gathering techniques that enable the collection of large unlabeled datasets [148]. Unlike the stream-based approach which selects individual instances from a dataset, pool-based sampling selects a subset (or pool) of data and the

learner decides which of them to query. Consider a dataset with a large number of unlabeled instances (U) and a small labeled subset (L). In pool-based sampling, queries are selected from U in a greedy manner based on a utility measure that is evaluated on all the unlabeled data. Pool based sampling has been implemented in several real-world machine learning problems including text classification [148–150], information extraction [151,152], image classification and retrieval [153,154], video classification and retrieval [155,156], speech recognition [157] and cancer diagnosis [158]. Pool-based sampling is the most popular scenario for applied research in active learning, whereas query synthesis and stream-based selective sampling are more common in the theoretical literature. Pool-based sampling is adopted in this study as it is more relevant to cases where gathering data is simple, but the labeling process is computationally expensive.

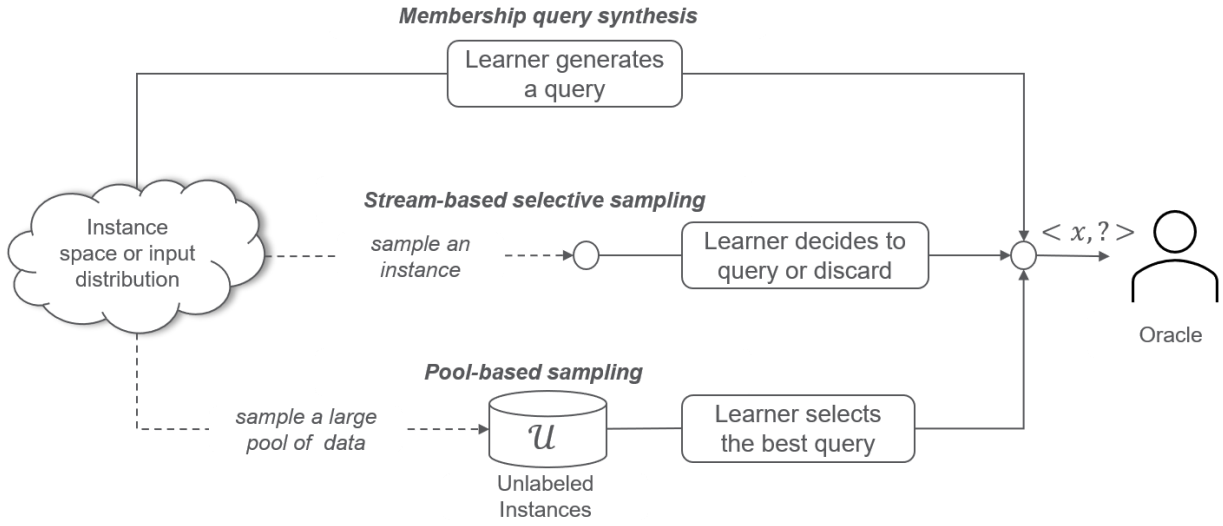


Figure 4-2 Query approaches in active learning

4.4.2 Evaluating the informativeness of queries via an acquisition function

Regardless of the querying approach, the amount of information provided by an unlabeled instance must be evaluated. This is achieved through the use of an acquisition function, V_t, which estimates the usefulness (or utility) of candidate input points based on how much it enables function learning (exploration) and generating the best possible output (exploitation). In a Bayesian setting, learning about a function can be represented as reducing the dispersion of the posterior distribution over all possible functions. A useful measure of the uncertainty about a random variable Y with probability distribution p is the differential entropy $H(\blacksquare)$ represented as

$$H(Y) = -\int p(y) log p(y) dy = \mathbb{E}[\log p(Y)] \tag{4.14}$$

The information gain $\boldsymbol{I}$ provided by input x about the random variable is the reduction in entropy when an input and the corresponding output are observed

$$\boldsymbol{I}(Y; x) = H(Y) - H(Y|x) \tag{4.15a}$$

$$\boldsymbol{I}(Y; x) = -\int (p(y) log\ p(y) + p(y, x) log p(y, x)) dy \tag{4.15b}$$

If Y follows a d-variate Gaussian distribution with mean μ and covariance Σ, then the entropy is

$$H(Y) = \frac{1}{2} \log(2\pi e)^d\ |\Sigma| \tag{4.16}$$

For a $\mathcal{GPR}$ model, the information gain can be written as

$$\boldsymbol{I}(f; \boldsymbol{y}) = \frac{1}{2} \log |\mathrm{I} + \sigma^{-2} \boldsymbol{K}| \tag{4.17}$$

where $\boldsymbol{K} = [k(x, x')]$ and I represents the identity matrix [159].

Even though finding the overall information gain maximizer is NP-hard, it can be approximated by an efficient greedy algorithm based on Gaussian process regression. If $F(A) = \boldsymbol{I}(f; \boldsymbol{y}_A)$ is the information about the function f after having observed a set of points A, the greedy querying algorithm selects the point with the most uncertain predicted output i.e. $x_t = \arg\max F(A_{t-1} \cup \{\boldsymbol{x}\})$. Here, uncertainty is measured by the variance of f at input $\boldsymbol{x}$.

$$V_t(\boldsymbol{x}) = k_{t-1}(\boldsymbol{x}, \boldsymbol{x}') \tag{4.18}$$

The algorithm begins with a Gaussian process prior for f and at each time $t = 1, \ldots, T$, the input points where the current posterior predictive distribution $p(f|\mathcal{D}_{t-1})$ evaluated at $\boldsymbol{x}$ shows the highest variance (the highest predictive uncertainty) are sequentially sampled. This is a "greedy algorithm" in the sense that it focuses on minimizing the current uncertainty, rather than looking further ahead in the future. Also described as "uncertainty sampling" [148,160], which algorithm is one of the most commonly used acquisition functions. The basic premise of uncertainty sampling is that the learner avoids querying instances of high confidence and focuses its attention on the least certain unlabeled instances. Figure 4-3 shows the algorithm for a generic pool-based active

learning scenario that uses uncertainty sampling to select the most uncertain instance x^* from U based on the model θ. This selection is performed iteratively until the stopping criterion is reached and the algorithm returns the trained model θ^*.

Krause et al. (2008) showed that uncertainty sampling can obtain at least a constant fraction of maximum information gain using at most T samples.

$$F(A_T) \geq \left(1 - \frac{1}{e}\right) \max_{|A| \subseteq T} F(A) \tag{4.19}$$

where $F(A_T)$ measures the information about f at time point t within the set A and e represents Euler's number. Equation 4.19 is based on two properties of the acquisition function: submodularity and monotonicity. Submodularity refers to a natural diminishing returns property of an acquisition function, whereby newly sampled points will add less and less information about the underlying functions. Monotonicity implies that there is no downside to additional information i.e., it is always helpful to observe more points.

Uncertainty sampling has been implemented in learning probabilistic models for binary [160] and multi-class classification [162], statistical sequence models in information extraction tasks [152,163] and support vector machines [149]. Uncertainty sampling is also applicable in regression problems where the learner queries the unlabeled instance for which the model has the highest output variance in its prediction. Closed-form approximations of output variances can be computed for several models, including Gaussian random fields [164] and neural-networks [165].

Algorithm 1: Generic pool-based uncertainty sampling algorithm

Result: Trained model θ^* using Active Learning
U = a pool of unlabeled instances $\left\{x^{(u)}\right\}_{u=1}^{U}$
L = set of initial labeled instances $\left\{\langle x, y\rangle^{(l)}\right\}_{l=1}^{L}$
while *stopping criterion* **do**
  $\theta = train(L)$
  select $x^* \in U$, the most uncertain instance according to model θ
  query the oracle to obtain label y^*
  add $\langle x^*, y^*\rangle$ to L
  remove x^* from U
end

Figure 4-3 Generic pool-based uncertainty sampling algorithm

4.4.3 Stopping Criterion

An important element of the active learning algorithm is knowing when to stop learning, or at least to stop posing queries. The stopping should be done at a point where the cost of acquiring new training data is greater than the cost of the errors made by the current system or the accuracy of the learner has reached a plateau. In the current study, the selected uncertainty (SC_{SU}) based on the overall uncertainty method proposed by Zhu et al. (2010) is adopted for the stopping criterion. SC_{SU} works on the principle that, in an iterative learning setting, the variance on the top-m selected samples (i.e., m most uncertain cases) at each learning cycle will be a good signal to indicate the confidence of the current model on the remaining unlabeled instances. SC_{SU} can be defined by Equation 4.20.

$$SC_{SU}(C) = \begin{cases} 1, & \frac{\sum_{x \in C} Var(x)}{m} < \delta_{SU} \\ 0, & otherwise \end{cases} \tag{4.20}$$

where δ_{SU} is the user-predefined variance threshold, C is the set of top-m selected unlabeled samples and the function $Var(x)$ evaluates the variance on set C based on the current model.

4.5 Case Study

The active learning framework is demonstrated through a risk-based assessment of the city of Napa's water distribution system. More specifically, active learning is used to select the subset of ground motion maps that are most critical to the water distribution system seismic risk. This particular problem is chosen because the state-of-the-art methodology involves performing a large number of end-to-end computational simulations that incorporate stochastic catalogue hazard characterization, damage assessment, inspection and repair processes and system-level hydraulic simulations. Gaussian process-based exploration is used to learn the full probability distribution of system-level impacts as quickly and accurately as possible using active learning. For ease of illustration, we will focus on a function f that takes a one-dimensional and discretized input x (related to the ground motion hazard) and maps to an output y (probability distribution of system-level performance).

4.5.1 Napa water network

Readers are referred to Section 2.2.1 for a detailed discussion on the Napa water network.

4.5.2 Risk assessment of the Napa water network using an end-to-end simulation framework

In this study, post-earthquake functional loss and restoration assessment is performed using an end-to-end simulation framework. Hazard characterization is the first step, which culminates with the generation of a stochastic catalogue with a set of ground motion fields corresponding to each event. Fragility functions for the components that comprise the system (pipes, pumping stations and tanks) are then coupled with the ground motion maps in a Monte Carlo framework to produce realizations of damage. Post-earthquake functional restoration of the system is represented using discrete event simulation (DES) and a system-level hydraulic performance model. The network performance is probabilistically quantified using metrics that are extracted from the recovery trajectories. Readers are referred to Section 3.2 for a detailed discussion risk-based assessment framework.

In the active learning context, the dataset $\mathcal{D}$ includes 2343 inputs, x (median ground motion intensity (PGV) for each scenario) and the corresponding output y (network performance metric). Figure 4-4 plots the $T_{100\%SSI}$ histogram representing the total time to fully recover after the earthquake considering all events. The x-axis bins are the $T_{100\%SSI}$ values and the y-axis represents the fraction of ground motion inputs in each bin. Similarly, Figure 4-5 plots the histogram of SSI_{CL} which is the area above the restoration curve normalized by the total area. The x-axis bins the SSI_{CL} values and the y-axis represents the fraction of ground motion maps in each bin. The higher SSI_{CL} values represent greater disruption in functionality disruption for a typical damage map. A 100% value corresponds to a complete loss of service over the considered time interval and 0% means no loss.

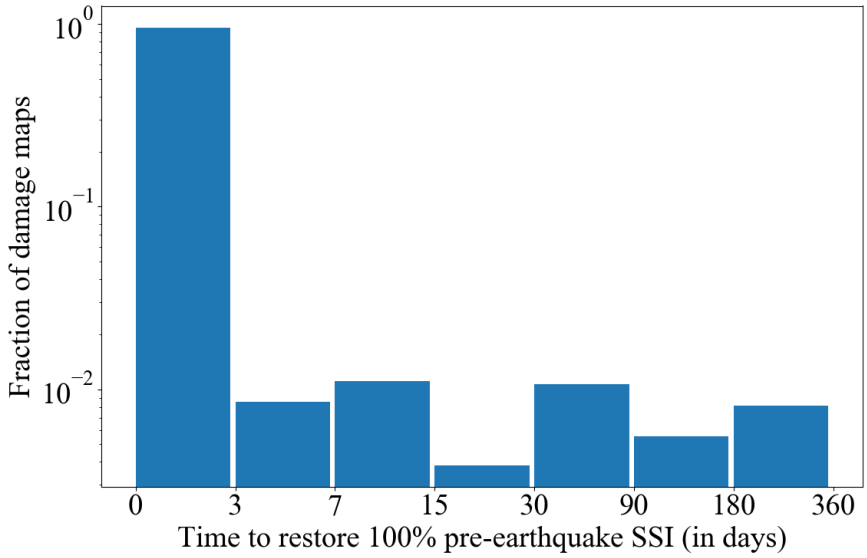


Figure 4-4 Distribution of time to restore 100% pre-earthquake SSI

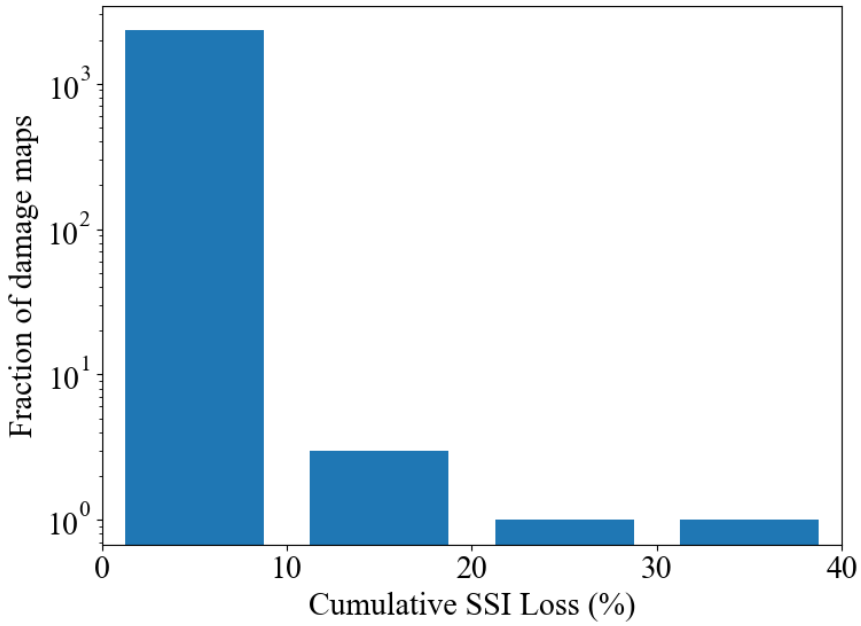


Figure 4-5 Distribution of cumulative loss in SSI

4.5.3 Map selection algorithm using active learning

This section discusses the development of the active learning-based map selection algorithm that is illustrated in Figure 4-6. To demonstrate the algorithm, it is assumed that the input x is

known but the output (the system-level performance distribution obtained from the end-to-end simulation framework) is not. Based on the results from a preliminary evaluation where several parameters were considered, the median PGV for all maps associated with a particular scenario was chosen as the input x. Therefore, initially, there are a set of unlabeled instances $U = \{x^1, x^2, \dots, x^{2343}\}$ containing the median PGV for all scenarios and the labeled instances $L = \{\emptyset\}$ is an empty set.

The algorithm starts by randomly sampling 10 ground motion maps and executing the end-to-end simulation framework to observe (evaluate) the network performance metric. At this point, the randomly sampled maps can be referred as the training set $x_{training}$ and $y_{training}$ is the output that corresponds to the network performance metric. The set of labeled instances are added to $L^1 = [{x_{training}}^T, {y_{training}}^T]$, which contains the labeled training samples. The remaining set of unlabeled instances are added to $U^1 = \{U \backslash x_{training}\}$.

A Gaussian process regression model is fitted to $\mathcal{D}_{training} = [{x_{training}}^T, {y_{training}}^T]$ with the mean function set to $m(x) = \mathbf{0}$ and an RBF kernel is used as the covariance function. The Scikit learn module [167] is used for this purpose. The hyper-parameters for the RBF kernel are optimized using the marginal log-likelihood (ML-II) procedure described in Section 4.3.3.4. After obtaining the functional form of the Gaussian process regressor that is fitted to $\mathcal{D}_{training}$, the variance on $x_{testing} = \{U \backslash x_{training}\}$ is calculated.

For checking the stopping criterion, the 10 testing points with the highest variance (i.e., points having highest dispersion in the fitted $\mathcal{GPR}$ model) are selected and tested for $\delta_{SU} = 0.6$. If the criterion is satisfied, then the algorithm stops and the resulting fitted $\mathcal{GPR}$ model represents a surrogate for the end-to-end simulation framework with ${x_{training}}^T$ as the selected ground motion maps. Otherwise, if the stopping criterion is not satisfied, the next sample x^* is selected from U^1 based on uncertainty sampling (i.e., the sample with the highest variance) and the corresponding network performance metric is calculated using the end-to-end simulation framework. This sample is added to $\mathcal{D}_{training}$ and the process is repeated starting with the construction of a new $\mathcal{GPR}$ model. The algorithm sequentially repeats this overall process until the stopping criterion is satisfied and the results from the selected subset of maps closely represent the performance metric

distribution for the entire stochastic event set. The Python-based modular active learning framework, modAL [168], is used to code the map selection framework.

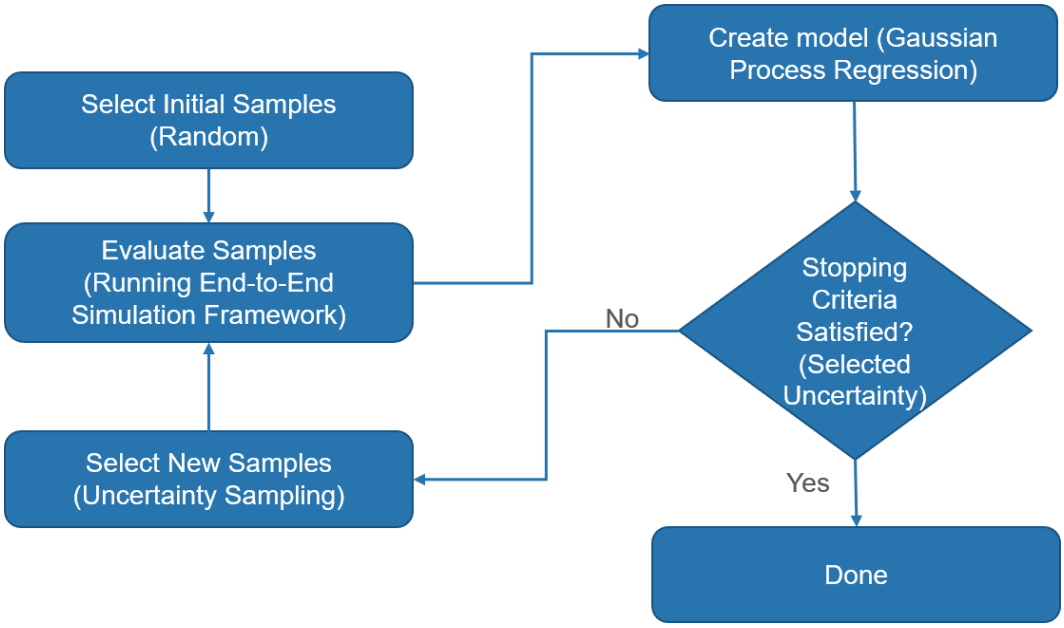


Figure 4-6 Map selection algorithm using active learning

When defining the stopping criterion, the adopted algorithm should be able to deal with challenges associated with the selection of initial random points. Recall that at the start of the algorithm, 10 points are sampled randomly, a $\mathcal{GPR}$ is fitted to this initial training set and then the stopping criterion is checked. At this stage, zero variance may be obtained for some testing points (for example, testing points that are spaced farther apart than the training points) and the stopping criterion will be satisfied. To avoid this problem, after the initial random selection, the stopping criteria is not checked until at least 10% of all the input points are selected. This modification allows the stopping criterion to gain stability and increases its robustness during the early stages of selecting random points. Additionally, the negative network performance predictions obtained from the fitted $\mathcal{GPR}$ models are taken as 0.

4.6 Results and Discussion

This section presents the results from the active learning-based map selection algorithm for two network performance metrics: $T_{100\%SSI}$ and SSI_{CL}. For $T_{100\%SSI}$, the active learning framework performed 919 queries before reaching the stopping point. The $\mathcal{GPR}$ model fitted to

919 queried points and the corresponding outputs from the end-to-end simulation framework are used to predict the full $T_{100\%SSI}$ distribution. The empirical and predicted $T_{100\%SSI}$ histograms shown in Figure 4-7 indicate that the fitted Gaussian process model closely estimates the $T_{100\%SSI}$ distribution. For SSI_{CL}, the active learning algorithm needed 234 queries to learn the entire distribution. Figure 4-8 shows the empirical and predicted SSI_{CL}. It is observed that the $\mathcal{GPR}$ model almost exactly captures the skewed SSI_{CL} distribution.

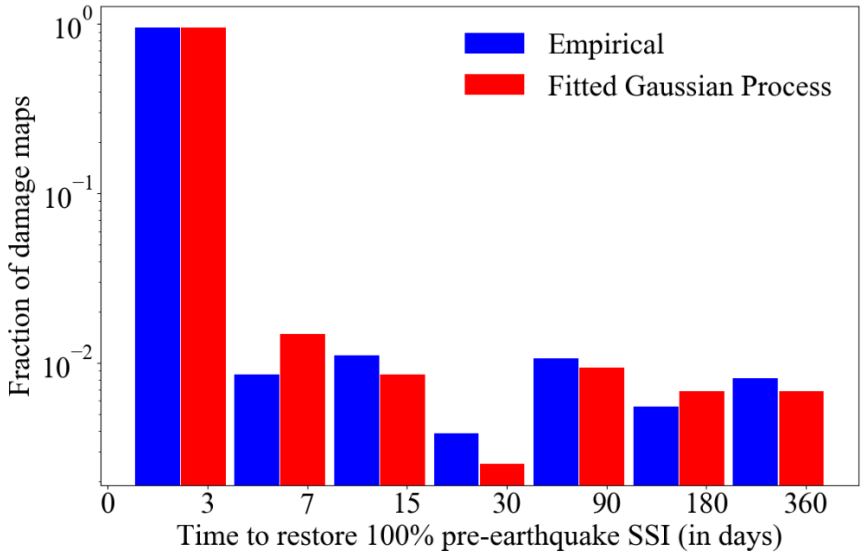


Figure 4-7 Histogram of empirical and predicted $\boldsymbol{T_{100\%SSI}}$ distribution

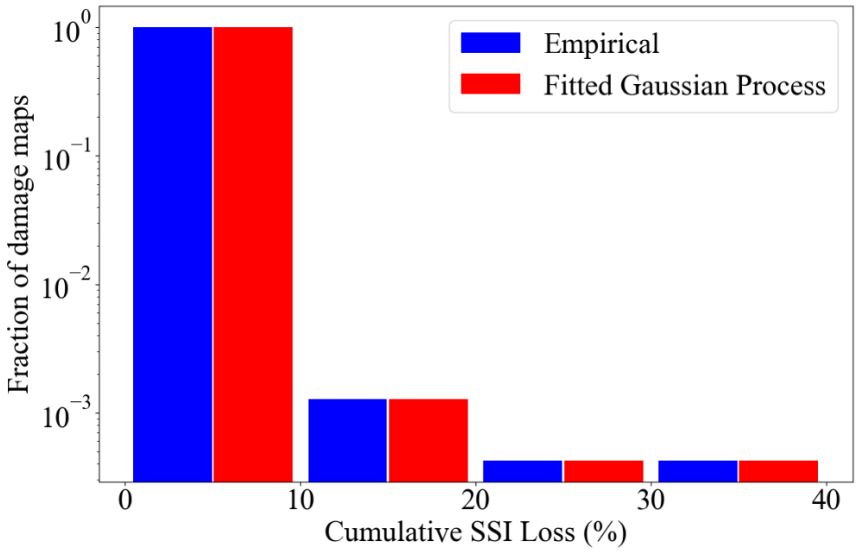


Figure 4-8 Histogram of empirical and predicted $\boldsymbol{SSI_{CL}}$ distribution

The network performance distribution along with the annual exceedance rates of the associated scenarios can be used to calculate the annual exceedance rate of each performance metric using Equation 4.23.

$$\lambda(\hat{p}) = \sum_{j \in M} w_j \mathrm{I}[p_j \geq \hat{p}] \tag{4.23}$$

where $\lambda(\hat{p})$ is the annual exceedance rate for a specified network performance level $\hat{p}$, M is the set of ground motion maps, w_j is the weight (annual occurrence rate) of ground motion map j, p_j is the network performance resulting from ground motion map j, and I[■] represents an indicator function that is equal to 1 when the network performance resulting from ground motion map j exceeds the specified level, and 0 otherwise.

Figure 4-9 shows the annual exceedance curve for $T_{100\%SSI}$ calculated using the complete set (2,343) and selected subset (919) of ground motion maps. The selected subset of maps closely estimates the annual exceedance rates for $T_{100\%SSI}$ with a R^2 score of 0.92. Similarly, Figure 4-10 shows the annual exceedance curve for SSI_{CL} calculated from the complete set and selected subset (234) of ground motion maps. Again, the selected subset of maps closely estimates the annual exceedance rates for SSI_{CL} with a R^2 score of 0.86.

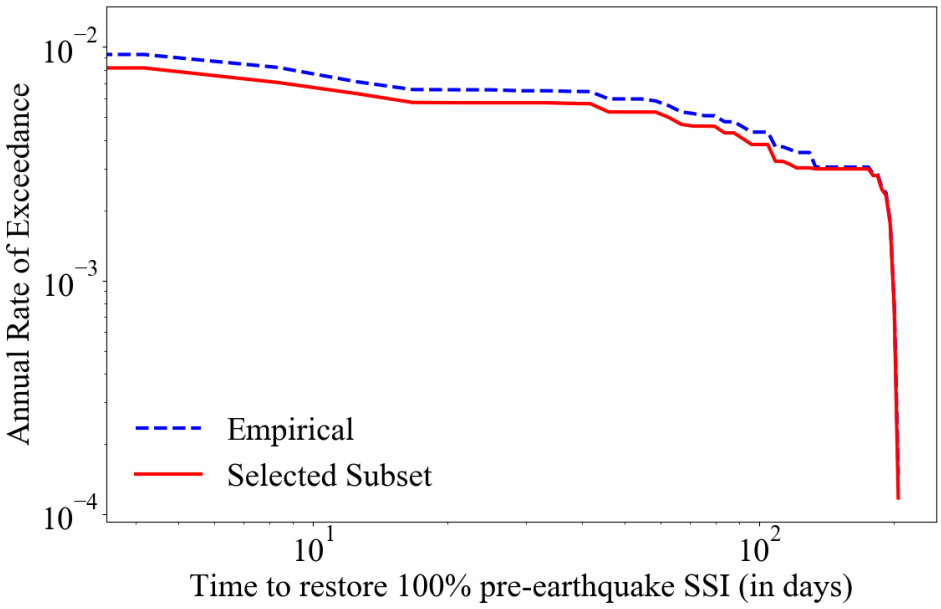


Figure 4-9 $\boldsymbol{T_{100\%SSI}}$ annual exceedance curve for the empirical and selected subset of ground motion maps

Figure 4-10 $\boldsymbol{SSI_{CL}}$ annual exceedance curve for the empirical and selected subset of ground motion maps

To understand the implications of the initial random sampling of maps and stability of the stopping criterion, the active learning framework is repeated 100 times for the $T_{100\%SSI}$ metric. In other words, the initial set of random samples is the only thing that distinguishes the 100 realizations. Figure 4-11 compares the $T_{100\%SSI}$ annual exceedance curve for the empirical (all scenarios) and selected subset of ground motion maps for all 100 realizations. The number of queries (selected map subset) ranges from 848 to 931 maps and each simulation follows a similar pattern of convergence after performing approximately 500 queries. The R^2 score for the predicted $T_{100\%SSI}$ annual rate of exceedance curves in Figure 4-11 ranges from 0.68 to 0.99 with a median value of 0.98. It is also observed that the dispersion in $T_{100\%SSI}$ annual rate of exceedance for the 100 realizations decreases as the value of $T_{100\%SSI}$ increases. This can be attributed to the relative fraction of scenarios in each bin. More specifically, the dispersion is higher in the bins with a greater fraction of scenarios. Note that, if needed, the dispersion in $T_{100\%SSI}$ annual exceedance curves can be reduced by choosing a lower δ_{SU} value. Figure 4-12 shows the time series of the stopping criterion (SC_{SU}) for every realization corresponding to $\delta_{SU} = 0.6$. It shows the instability of the stopping criterion in the initial stages of training. Recall that this is addressed by querying at least 10% of all maps before beginning the stopping criterion check.

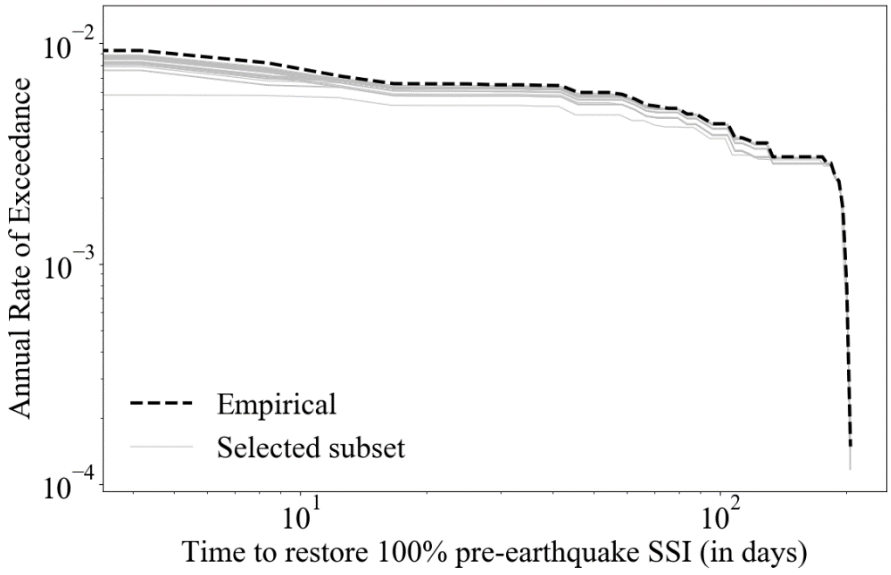


Figure 4-11 $\boldsymbol{T_{100\%SSI}}$ annual exceedance curve for the empirical and selected subset of ground motion maps in the MCS

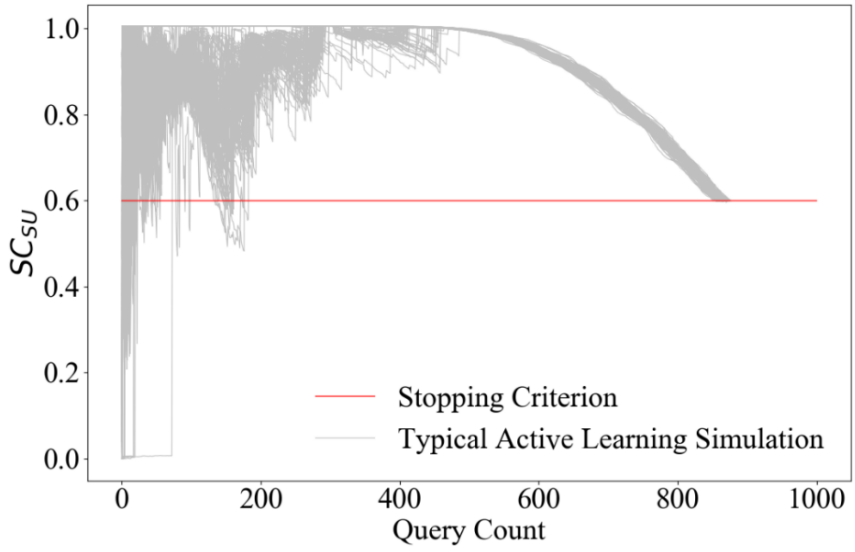


Figure 4-12 Time series of the stopping criterion ($\boldsymbol{SC_{SU}}$) for every case of MCS with $\boldsymbol{\delta_{SU}}$

4.7 Summary and Conclusion

Risk assessment of distributed infrastructure system entails consideration of all possible hazard scenarios, corresponding component-level damage assessment and network-level performance quantification. The computational expense of performing stochastic event set assessment can be reduced by utilizing a subset of the possible scenarios for hazard characterization. This study presented a computationally efficient method for selecting such a subset using active learning (AL). The framework begins by defining a learner (in this case, a gaussian process regression model), which is used to select a subgroup of unlabeled (risk assessment has not yet been performed) samples (scenarios) from the input distribution (the entire stochastic event catalogue). Based on their level of informativeness, which is gleaned from uncertainty sampling (the acquisition function), some of those samples are sent to an oracle (end-to-end simulation-based performance assessment) for labeling (calculating the corresponding performance metric). These labeled instances are added to the training subset, which is used to update the learner's model. The selected uncertainty stopping criterion is used to determine whether to repeat this process or terminate the learning. The training set that is obtained after the stopping criterion is satisfied represents the selected subset of events that are used as the basis for the risk assessment.

Gaussian process regression ($\mathcal{GPR}$), which is used as the learner in the AL framework, is introduced as a general-purpose technique for modeling and exploring unknown functions. The *weight-space view* and *function-space view* are discussed to gain insight into the sampling procedure, posterior prediction and hyper-parameter optimization. The radial basis function is used as the kernel for the $\mathcal{GPR}$ models. However, other kernels and kernel combinations are also discussed.

Three alternative querying approaches for active learning are explained, and pool-based sampling is adopted for the proposed risk assessment framework. The uncertainty sampling procedure is used as the activation function, which evaluates the informativeness of unlabeled instances. The selected uncertainty SC_{SU} stopping criterion is used to determine whether to terminate or continue the learning process based on sample variances.

A case study involving a risk-based assessment of the city of Napa water distribution system is used to demonstrate how the active learning framework could be used to select a subset of events from a larger set of candidates. Two metrics of system performance are considered, both of which are derived from explicit quantification of water service recovery as measured by the system serviceability index (SSI). The first is the time to restore 100% of the pre-event SSI ($T_{100\%SSI}$) and the other is the cumulative loss of service (SSI_{CL}), which is taken as the area above the SSI restoration curve normalized by the cumulative area. The results show that the AL-based event subset provides an accurate assessment of the full probability distribution of these two-performance metrics. For example, the $T_{100\%SSI}$ annual exceedance rate curve based on 40% of the complete catalogue provided a match with an R^2 score of 0.92 when measured with the one obtained from all events. Additionally, using only 10% of the original event set, a similarly high R^2 score (0.86) was obtained for SSI_{CL}. These results demonstrate the significant computational gains that AL can provide in infrastructure risk assessments.

The proposed framework can be extended to analyze risk from other types of natural hazards (e.g. floods, hurricanes) and infrastructure systems (e.g. water, power, building portfolios). The current study utilized only a single parameter (median peak ground velocity) the input. Future work can investigate the use of multiple dimensions in the input vector. For instance, multiple ground motion intensity measures, spatial correlation network summary metrics (e.g. degree of centrality) can be explored as potential inputs. However, it is important to recognize that, as the dimension of the input vector increases, advanced sampling methods such as query by committee, expected error reduction, hierarchical sampling may be required (instead of uncertainty sampling). More research is also needed on selecting kernels for the $\mathcal{GPR}$ model that incorporate domain expertise and prior knowledge. The stopping criterion for the active learning framework is another area that is ripe for exploration where more generalizable adaptive strategies (e.g. learner-decided threshold) may prove useful.

CHAPTER 5 Dynamic Updating of Post-Earthquake Damage and Functional Restoration Forecasts of Water Distribution Systems using Bayesian Inferencing

---

This chapter is adopted from the following study:

Tomar, A., Burton, H. V. and Mosleh, A. (2020). "Dynamic Updating of Post-Earthquake Damage and Functional Restoration Forecasts of Water Distribution Systems using Bayesian Inferencing," *Earthquake Spectra* (under review).

5.1 Introduction

The United States Department of Homeland Security National Infrastructure Protection Plan identifies water and wastewater systems as one of several critical infrastructure systems [10]. The plan also describes the sourcing, distribution and recycling of water as being essential to the operation of most critical infrastructure sectors. The impact of earthquakes on water systems begins with the physical damage and functional disruption of individual components, which then degrades the overall operability of the network. Many prior studies on the impact of seismic events on water systems have assessed the physical damage to individual components of the system [169–173]. The focus here is on modeling the spatial distribution of shaking, which is then coupled with component-level fragility functions to simulate damage. Other studies have sought to link component-level damage to the disruption of service and overall system performance immediately following the event [31,32,82,174–176]. Static (i.e. not time dependent) metrics such as the hydraulic power capacity [24,25], resilience index [26], connectivity loss [177,178], and serviceability ratio [21] have been used to quantify overall system performance. Hydraulic power capacity is the probability that there exists a feasible flow of hydraulic power in the network. Computing this performance metric requires assumptions regarding the minimum loss of network power and the subsequent redistribution of that loss to the nodes. The aforementioned resilience index is the ratio between the surplus of internal power in the network to the maximum power that could be dissipated internally, after satisfying the nodal demand and head constraints. Recognizing that this metric cannot be used in a network with multiple sources, a modified resilience index [27]

was proposed to overcome this drawback by considering the ratio between the available surplus and minimum required power at the demand nodes. It can be used to compare the performance of one network relative to another, which is useful for new design and rehabilitation problems. Connectivity loss measures the average reduction in the ability of sinks to receive flow from a source. Serviceability ratio captures the number of distribution nodes in the network that remain accessible from at least one supply facility following the earthquake. Some prior studies have also evaluated the direct and indirect economic implications of physical damage and/or reduced performance [52,179].

With the emergence of seismic resilience as a core issue related to the impact of earthquakes on built systems, models have been developed to simulate repair activities and functional recovery of the water network [37,45,51,180–182]. The earliest attempt to model post-earthquake functional restoration of water systems used a series of analytical equations with variables that capture the number, efficiency and scheduling of repair workers [51]. Chang et al. 2002 adopted a similar "resource-constrained" analytical modeling approach while also facilitating repair prioritization. Network models have also been used to simulate post-earthquake performance of water systems and optimize repair and recovery processes [54]. More recently, the application of discrete event simulation (DES) [72] as the primary engine for modeling water system restoration has gained popularity (e.g. [37,55,56,73,180,181,183,184]). DES models use a set of event-driven interacting entities (e.g. pipe, tank, pump) to represent system processes. These entities have attributes (e.g. pipe damaged, pump functional) and the associated processes (e.g. repair pipe) rely on a set of limited resources (e.g. repair materials, repair workers). A common theme among all the aforementioned studies is that they rely on expert judgment and limited empirical data to calibrate the corresponding damage, functional disruption and/or restoration model. This introduces large uncertainties in predicting post-event impacts. This is especially true for those models that attempt to capture restorative activities and the temporal variation in the system serviceability and overall performance. One strategy for reducing this uncertainty is to use information acquired during the recovery process to update forecasts of component-level physical damage and system-level restoration trajectories.

Probabilistic models based on directed acyclic graphs (DAGs) have a long and rich tradition, which began with the geneticist Sewall Wright (1921) [185]. Motivated by the need to model the

top-down (semantic) and bottom-up (perceptual) combination of evidence in reading, Bayesian networks (BNs), a variant of DAGs, was initially developed in the late 1970s. The capability for bi-directional inferences, combined with a rigorous probabilistic foundation, led to the rapid emergence of BNs as the method of choice for uncertain reasoning in artificial intelligence and expert systems, replacing earlier, ad-hoc rule-based schemes [186–190]. Since that time, BNs have been extensively applied in domains such as statistical modeling, language processing, image recognition, machine learning and database systems [191–195]. In the past 25 years, BNs have received considerable interest in risk modeling of engineered systems [196–207]. Dynamic Bayesian networks (DBNs), are BNs that capture the evolution of variable relationships over time [208]. DBNs are common in robotics [209,210] and have shown potential for a wide range of data mining applications [211,212].

Bayesian updating provides a mathematical framework for combining new information into an existing model [213–216]. One of the earliest applications in the broad areas of seismic risk, reliability and resilience, was performed by [217], who developed a Bayesian statistical analysis method for updating fragility functions when new data became available. The methodology was used to update the ground motion-damage relationships for reinforced concrete structures. This initial study was followed by several others that implemented Bayesian updating for fragility analysis of structures [218–220] and lifeline components [221] and reliability-based assessment of individual [222–228] and distributed infrastructure systems [229–232].

This study presents a novel approach to dynamically updating post-earthquake damage and restoration forecasts in water distribution networks. Central to the proposed framework is the integration of three types of models: a BN model of the Napa water piping system, a process-based discrete-event simulation (PBDES) model of the post-earthquake functional recovery and Bayes' theorem for updating the pipe damage and temporal recovery parameters (e.g. pipe repair time). The BN model takes the pipe damage and repair time probability distributions and the number of repair crews as the inputs and provides a probability distribution of the total recovery time as output. The PBDES-based post-earthquake recovery model takes the pipe damage distribution, inspection and repair crew, pipe inspection and repair time probability distributions and the crew reduction schedule as the inputs and provides a simulated pipe recovery curve. The priori distribution and observed values of key input variables (e.g. damaged state of individual pipes,

pipe repair time) are dynamically updated using Bayes' theorem. The proposed framework can be used by water management agencies to identify the critical aspects of the restoration process and act on them in real time. The methodology can also be used to plan and train crew members by creating a suite of earthquake scenarios and the corresponding recovery response.

5.2 Overview of the Proposed Methodology

A schematic overview of the proposed methodology for dynamic updating of post-earthquake damage and functional restoration forecasts of the water pipe network is shown in Figure 5-1. Only the pipe network is considered in the current study because there is an available dataset of documented pipe damage and recovery trajectory from a real earthquake (discussed later) that is used as a benchmark. However, the framework is general enough such that it can be adapted to consider other components in the water system (e.g. tanks, pumps). The horizontal axis in Figure 5-1 represents time and two points are highlighted (Day 0 or the day of the earthquake and Day t: the t^{th} day after the earthquake). The vertical axis represents the variables (input and output) related to earthquake-induced pipe damage and recovery whose values are changed during the recovery process. The input variables are *pipe damage, pipe repair time distribution, inspection crew and repair crew* and the *recovery time* and *recovery trajectory* are the outputs. Each variable has an observed and simulated value at each time point. Starting at $t = 0$ (represented as Day 0), an earthquake occurs and causes damage to the pipe network. As the exact distribution of pipe damage is unknown immediately following the earthquake, a simulation model is used to provide an initial estimate. Assuming that the magnitude and location of the earthquake is available within a few hours of its occurrence, correlated ground motion maps can be generated using the appropriate models ([91,92,119–122]), which can then be coupled with pipe fragility functions to produce an initial (Day 0) estimate of (priori) the damage distribution (Figure 5-1).

Given the priori pipe damage distribution and an appropriate model, a Day 0 forecast of the pipe recovery time or trajectory can be obtained. Note the distinction between the recovery time and trajectory. The latter provides an empirical function of the state of the pipe network (in terms of the percentage of repaired pipes) as a function of time (a recovery curve) while the former is an estimate of the time to complete the repairs (the time to 100% recovery). Both are considered in this study. The event duration distribution of pipe inspection and repair time at Day 0, which are needed as inputs into the recovery model, can be taken from empirical data from past events. Using

the priori pipe damage distribution, pipe repair time distribution parameters and available number of repair crews at Day 0, an estimate (mean, standard deviation) of the pipe recovery duration (in days) is obtained from the BN model of the Napa water piping system (blue arrows in Figure 5-1). Additionally (or alternatively), the PBDES pipe recovery model is used to obtain a pipe recovery trajectory (orange arrows in Figure 5-1). Note that these initial estimates of pipe recovery time and trajectory are based on the assumed Day 0 parameters.

The proposed framework seeks to dynamically update the estimates of the post-earthquake restoration forecast (pipe recovery time and trajectory). At Day t, the values of variables between Day 0 and Day t have been observed. More specifically, as the restoration proceeds in real time, the individual occurrences of pipe damage throughout the network can be updated. In other words, some pipes that have been simulated as "damaged" might be changed to "intact" following their inspection. Similarly, the time needed to repair various types of pipes and changes in the crew size (increase or decrease in crew members) are observed and recorded. Based on the observations made between Day 0 and Day t, the pipe damage distribution and the number of available inspection and repair crews are updated. Additionally, the priori individual pipe repair time distribution can be updated using Bayesian theorem (green arrows in Figure 5-1). Following these variable updates, a new estimate of the total pipe recovery duration is obtained from the BN model and a new pipe recovery trajectory obtained from the PBDES model. The updating process between Day 0 to Day t is typical and continues at specific time-intervals until the restoration is complete.

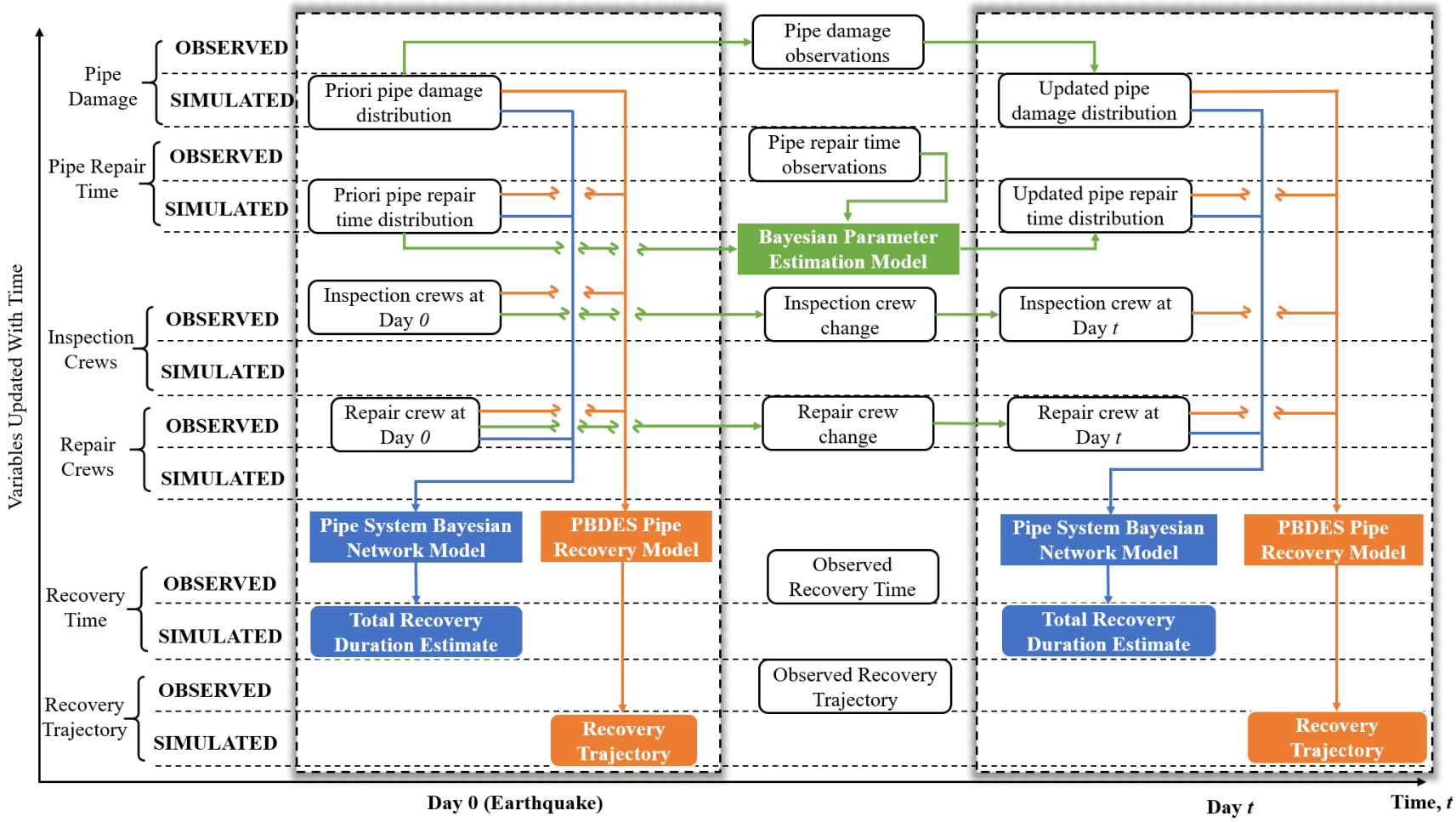


Figure 5-1 Overview of the proposed methodology

5.3 Description of the Water Network

Readers are referred to Section 2.2 for a detailed discussion on the Napa water network and the damage and recovery following the 2014 South Napa earthquake.

5.4 Modeling Post-Earthquake Pipe Repairs using Discrete Event Simulation

Readers are referred to Section 2.4 for a detailed discussion on modeling pipe repairs using PBDES model and replicating the recovery following 2014 earthquake.

There are several sources of uncertainty (see Figure 5-2) that are inherent in a "blind" (no data available for the event under consideration) post-earthquake functional restoration simulation. When the network component damage is simulated, there are uncertainties associated with the spatial distribution of shaking intensity for a given event (addressed by generating multiple ground motion maps) and the component-level damage conditioned on a specified shaking intensity (considered in fragility relationships). Within the PBDES functional restoration model, there are uncertainties in the temporal input parameters (i.e. inspection and repair times) and the availability of repair crews. The framework developed in the current study seeks to dynamically update the functional restoration forecasts over time (i.e. recovery trajectories generated at multiple time points) while using the data acquired from the actual event (as time progresses) to reduce these inherent uncertainties.

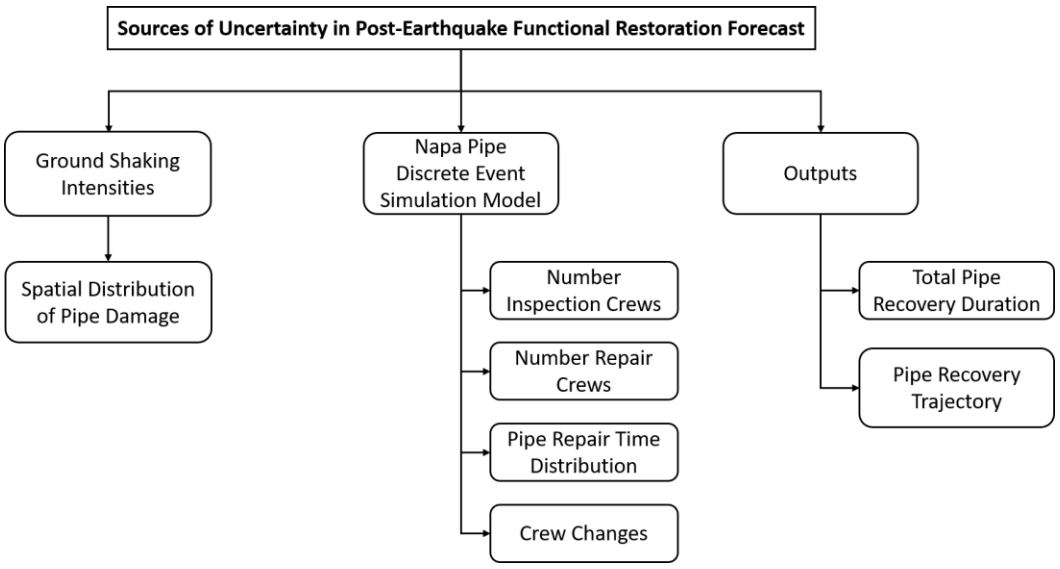


Figure 5-2 Sources of uncertainty in post-earthquake functional restoration forecast

5.5 Probabilistic Graphical Models (Bayesian Networks)

5.5.1 Background

Most daily tasks require reasoning either by a human or an automated system, which uses available information to make decisions about how to act. This reasoning process can be mathematically represented by exploiting the concept of declarative representation to construct a model of the system. This model encodes our knowledge of the system and answers specific questions through the manipulation of reasoning algorithms. Due to the inherent limitations in the ability of humans to observe and represent the innate nondeterminism of the world, these models have aleatory and epistemic uncertainty. Aleatory uncertainty (variability) is the natural randomness in a process whereas epistemic uncertainty is the scientific uncertainty in the model of the process [233]. The latter is due to limited data and knowledge and is characterized by alternative models. Given the presence of these two types of uncertainty, the possible as well as probable states of the system are considered.

Probabilistic graphical models (PGM) follow a declarative representation and use a graph-based approach to compactly encode the complexities of a system. Bayesian networks (BN) (also called belief networks, Bayesian belief networks, causal probabilistic networks, or causal networks) are a family of probabilistic graphical models that use a directed acyclic graph (where edges have a source and a target) to model the system [186]. PGM supports representation (compact encoding), inferencing (answering queries using our model of the world) and learning (data-driven approach to making changes in our model of the world) [234].

There is an excellent body of literature that defines all the key properties of graphs [234–236]. This section focuses on the ones used in the current study, which are specific to directed graphs. A graph is a data structure, G, consisting of a set of nodes ($\Lambda: X_1, X_2, \ldots, X_n$) and edges (E). In G, nodes can be connected through a directed edge ($X_i \rightarrow X_j$) or an undirected edge ($X_i - X_j$). A graph is directed when all edges are directed edges and acyclic when the graph does not contain any cycles. An induced subgraph, $G[\Lambda']$, where $\Lambda' \subset \Lambda$, is defined as (Λ', E'), where $\mathrm{E}'$are all the edges $X \leftrightarrows Y \in \mathrm{E}$ such that $X, Y \in \Lambda$. In other words, an induced subgraph is formed from a subset of the nodes (and all the edges connecting pairs of those nodes) in another graph. Using the basic notion of edges, longer-range connections in a graph can be defined. For instance, ($\Lambda_k: X_1, X_2, \ldots, X_k$) forms a path in G if, for every $i = 1, \ldots, k-1$, $X_i \rightarrow X_{i+1}$. For a directed edge $X_i \rightarrow X_j$ in G, X_j is the child of X_i in G and X_i is the parent of X_j in

G. Also, X is described as being an ancestor of Y in G and Y is a descendant of X, if there exists a directed path $X_1, X_2, \ldots, X_k$ with $X_1 = X$ and $X_k = Y$. The set of nodes which are descendants, ancestors and non-descendants of X are denoted as $Descendants_X$, $Ancestors_X$ and $NonDescendants_X$, respectively.

5.5.2 Bayesian Networks

A Bayesian network is a directed acyclic graph, G, whose nodes represent random variables $X_1, X_2, \ldots, X_n$. Given that $Pa_{X_i}^G$ denotes the parents of X_i in G and $NonDescendants_{X_i}$ denotes the variables in the graph that are not descendants of X_i, G encodes a set of conditional independence assumptions (also known as local independencies and denoted by $I_l(G)$). It then follows that for each $X_i\left(X_i \perp NonDescendants_{X_i} | Pa_{X_i}^G\right)$, the joint distribution P of all nodes in G can be expressed as a product:

$$P(X_1, X_2, \ldots, X_n) = \prod_{i=1}^{n} P(X_i | Pa_{X_i}^G) \tag{5.1}$$

Equation 5.1 describes the chain rule for Bayesian networks and the individual factors $P(X_i | Pa_{X_i}^G)$ are conditional probability distributions. A Bayesian network is represented as $B = (G, P)$ where P factorizes over G, and P is specified as a set of conditional probability distributions associated with the nodes of G. As noted earlier, the graph data structure provides the skeleton for compactly representing a joint distribution in a factorized manner.

Figure 5-3 shows a Bayesian network for a simple "pipe repair time" problem. The network models the flow of dependence for the pipe repair time. The pipe type node can take two distinct values: distribution and trunk. The type and length are physical attributes of a pipe and are therefore represented as nodes without parents. The ground shaking intensity depends on the properties of the earthquake (e.g. magnitude, fault type) and soil and is independent of the pipe type and length. Since the ground shaking intensities are simulated using OpenSHA, which encapsulates all of the relevant dependencies, the associated node is not assigned a parent. Therefore, there are three nodes (pipe type, pipe length and ground shaking intensity) that don't have parents. Pipe damage is defined using a fragility function which only depends on the pipe length and ground shaking intensity. Therefore, the pipe damage node has pipe length and

ground shaking intensity as parent nodes. Pipe repair time is conditioned on pipe type and damage. Therefore, the node of the former is the parent of the nodes of the latter two.

Based on the chain rule for Bayesian networks, the joint probability distribution of all nodes can be represented as shown in Equation 5.2.

$$\boldsymbol{P(PT, PL, GSI, PD, PRT) = P(PT) * P(PL|PT) * P(GSI|PL, PT)} \\ * P(PD|PL, PT, GSI) \ * P(PRT|PL, PT, GSI, PD) \tag{5.2}$$

where the random variables PT, PL, GSI, PD, PRT correspond to pipe type, pipe length, ground shaking intensity, pipe damage and pipe repair time respectively. The independence relationship among nodes in a Bayesian network are called local independencies and are formally represented as follows: (1) the pipe repair time only depends on pipe type and damage and is independent of pipe length and ground shaking intensity given pipe damage ($PRT \perp PL, GSI \mid PD$), (2) pipe type is independent of pipe length, damage and ground shaking intensity ($PT \perp PL, PD, GSI$), (3) pipe damage is independent of pipe type ($PD \perp PT$). Taking the conditional independence relationships into consideration, the joint probability distribution of all nodes represented in Equation 5.2 can be modified as shown in Equation 5.3. It can be observed that the conditional independence relationships enable the joint distribution to be represented more compactly. In general, if we had n binary nodes, the full joint distribution will be represented in $O(2^n)$ space but the factored form is represented in $O(n.2^k)$ space, where k is the maximum number of incoming edges of a node.

$$P(PT, PL, GSI, PD, PRT) \ = \ P(PT) * P(PL) * P(GSI) * P(PD|PL, GSI) * P(PRT|PT, PD) \tag{5.3}$$

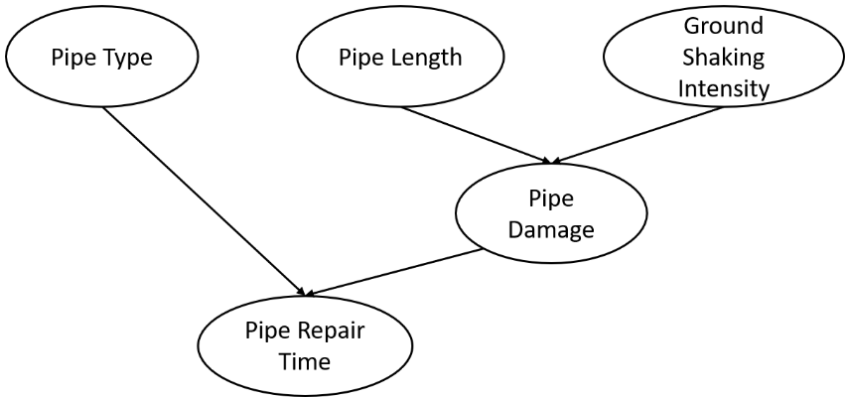


Figure 5-3 Bayesian network for a simple “pipe repair time” problem

5.6 Bayesian Inferencing

5.6.1 Background

Probability theory provides a formal foundation of semantics and rules for dealing with uncertain events. Combining the definition of conditional probability with the product and sum rule yields Bayes' rule [213]. It provides the basis for combining various sources of information and data and enables the treatment of the uncertainties in the parameters of the parent distribution of a random variable X. The unknown unobserved parameter Θ is regarded as a random variable which encodes model parameters. Given a new set of observations, x, the distribution of Θ can be updated by applying Bayes' rule (also illustrated in Figure 5-4 using pipe repair time as an example) as shown in Equation 5.4:

$$f_{(\Theta|X)}(\theta|x) = \frac{f_{(X|\Theta)}(x|\theta).f_{\Theta}(\theta)}{\int f_{(X|\Theta)}(x|\theta).f_{\Theta}(\theta).d\theta} \tag{5.4}$$

where $x = \{x_1, x_2, \ldots\ x_n\}$ is the set of n independent observations of the random variable X that has the underlying conditional probability distribution $f_{(X|\Theta)}(x|\theta)$ (also known as the likelihood function), $f_{\Theta}(\theta)$ is the prior distribution of Θ, which represents the state of knowledge about that parameter before the most recent set of observations of X, $f_{(\Theta|X)}(\theta|x)$ is the 'posterior distribution' of Θ, which represents the knowledge about that parameter after the most recent set of observations of X. The denominator is the integration of the likelihood function over all possible parameter assignments.

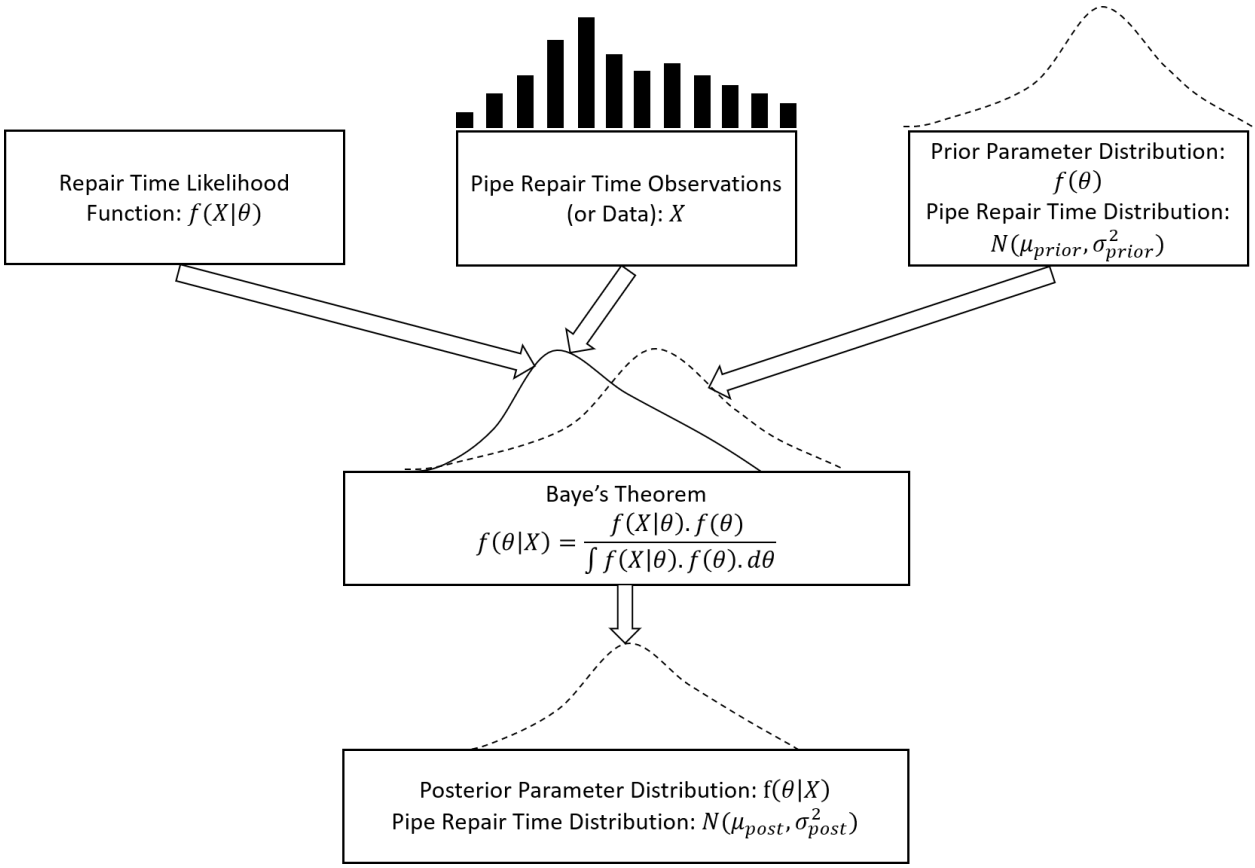


Figure 5-4 Schematic illustration of the Bayesian updating framework for pipe repair duration

In computational complexity theory, nondeterministic polynomial time hardness (NP-hardness) is defined as the property of a class of problems that are informally at least as hard as the hardest problem in NP. Several problems associated with networks belong to this class. For instance, the exact computation of the network reliability where any link has a discrete probability of being in two states (in service/out of service) belongs to this class [237]. Bayesian (or belief) updating is also a computationally complex problem. In the worst case, belief updating algorithms are NP-hard [238].

One of the main challenges in applying Bayes' theorem is the selection of an appropriate distribution for the parameter(s) Θ based on the available information. The variability captured by the distribution $f_{\Theta}(\theta)$ reflects the epistemic uncertainty which can be reduced with sufficient data. After specifying the prior distribution, estimating the posterior distribution can become complex and may not have a closed form solution. In those cases, Markov Chain Monte Carlo (MCMC) methods are used to estimate the posterior distribution by generating a sequence of points through a Markov chain. Markov chains can be constructed using various

algorithms such as Metropolis-Hastings, Gibbs sampler, slice sampling, and perfect sampling [239]. In special cases, the posterior and prior distribution belong to the same family for a particular likelihood function. These families of distributions are called "conjugate distributions" for the corresponding likelihood function. Using conjugates to model the parameter distributions reduces the mathematical complexity and guarantees a closed form solution.

5.6.2 Application of Bayesian Parameter Estimation to Update PBDES Event Durations

In the current study, Bayesian parameter estimation is used to update the event duration distributions (pipe repair) used by the PBDES model. Prior studies have used the triangular [55] and normal distribution [86] to probabilistically model the duration associated with repairing the components of a water network. Since the normal distribution is its own conjugate prior for a normal likelihood function with known variance, the current study assumes normal distributions for all event durations. Table 5-1 shows the priori mean and standard deviation of the repair duration for damaged pipes.

Table 5-1 Pipe repair duration mean and standard deviation (STD) [55]

Event		Mean (in hours)	STD (in hours)
Repairs	Distribution Leak	4.3	0.6
	Distribution Break	7.4	1.7
	Trunk Leak	112.0	11.3
	Trunk Break	191.9	19.8

By utilizing a normal prior sprobability density function (pdf) $f(\theta) \sim N(\mu_{prior}, \sigma_{prior}^2)$ and given a single measurement $x \sim N(\theta, \sigma^2)$, it is assumed that $\sigma^2 = \sigma_{prior}^2$. The posterior pdf after considering the observations is denoted as $f(\theta|x) \sim N(\mu_{post}, \sigma_{post}^2)$

where $a = \frac{1}{\sigma_{prior}^2}$, $b = \frac{1}{\sigma^2}$, $\mu_{post} = \frac{a\mu_{prior}+bx}{a+b}$, $\sigma_{post}^2 = \frac{1}{a+b}$. Note that μ_{post} is simply the weighted average of μ_{prior} and the data x.

5.7 Bayesian Networks for Computing the Total Recovery Time for the Napa Water Pipe System

A declarative representation of the post-earthquake total recovery time for the Napa water pipe system is created using Bayesian networks. This representation consists of two main

components: a pipe recovery plate model and a total recovery time node. Plate models are used as a language for encoding mathematical representations with repeated structure and shared parameters [240,241]. The pipe repair process is represented through a "pipe recovery plate model", which embeds a Bayesian network consisting of the following conditional probability nodes: *pipe type, pipe damage, repair time, damage identified indicator*. The following equation nodes (i.e. a node that has an equation as its attribute) are also included in the Bayesian network: recovery time and damage identified recovery time.

A schematic representation of the Bayesian network that is embedded in the pipe recovery plate model is shown in Figure 5-5 which is an extension of the one shown in Figure 5-3. The *pipe type* node has no associated parent node and two states: distribution and trunk. It has a discrete prior probability distribution (right after the earthquake), which is shown in Figure 5-3. The probability of being in each state (distribution or trunk) is defined by the ratio of the number of pipe types associated with that state to the total number of pipes in the network. It should be noted that, while the type of pipe at each location is known (deterministic), the type of pipe that will be damaged during earthquake shaking is uncertain, hence the discrete probability distribution. The *pipe damage node* has two parents: *pipe length* and *ground shaking* intensity as shown in Figure 5-3. The *pipe length* and *ground shaking intensity* are used to calculate the state of the *pipe damage node* using pipe fragility functions during the process of generating the priori pipe damage distribution. Therefore, in Figure 5-5, the *pipe damage node* has no associated parents and is used to represent the damaged condition of the pipe and has three possible states: no damage, leak, and break. The probability distribution for *pipe damage* is also discrete and describes the prior probability of being in each damage state, which is calculated using fragility curves [86]. The *repair time* node, which has the *pipe type* and *pipe damage* as parent nodes, is used to represent the repair duration for a damaged pipe. As noted earlier, the pipe repair time is represented using a normal distribution and the prior parameters (mean and standard deviation) are adopted from Tabucchi 2007. The *damage identified indicator* ($DamIdeInd$) has no parent nodes and two states: damage identified, and damage not identified. Before a pipe is inspected, it is in the damage not identified state i.e. prior probability of 1.0. If a pipe is identified as damaged during the inspection process, the state changes to "damage identified". The prior state of this node (immediately following the earthquake) is determined from the simulated pipe damage distribution. D*amage identified recovery time* ($DamIdeRecTime$) is an equation node that uses the *damage identified indicator* and *repair time* parent nodes to calculate its value. Note that nodes related to the

inspection are not considered in the pipe recovery Bayesian network because the *damage identified indicator* acts as a proxy for the inspection process.

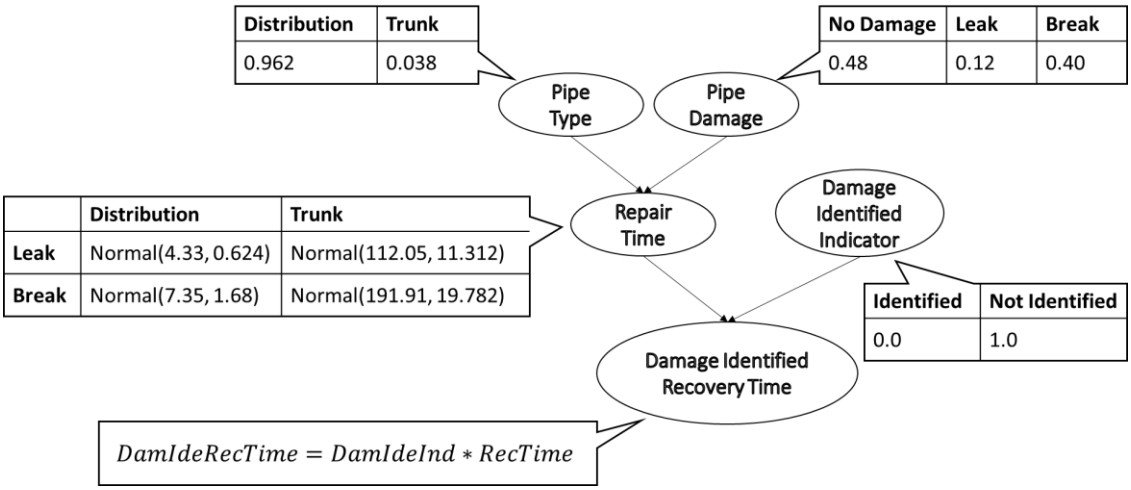


Figure 5-5 Schematic illustration of the Bayesian network embedded in pipe recovery plate model

Figure 5-6 shows a schematic illustration of the Bayesian network for the entire Napa water piping system. It consists of N (number of pipes in the water network) pipe recovery plate models, a node that accounts for the number of available repair crews on a given day and the total recovery time for the entire pipe system. All *pipe recovery plate models* are assumed to be independent. In other words, the recovery of any pipe is not affected by that of the other pipes in the network. The equation node representing the total recovery time is connected to all pipes through the *damage identified recovery time* with a dynamic edge that represents the behavior of *pipe damage identification* and is updated over time. In the post-earthquake environment, the number of pipes identified as being damage varies with time (through the inspection process, new damaged pipes are identified and pipes that were simulated as damaged were found to be intact) and becomes constant after all pipes have been inspected. For the pipe recovery plate model, the conditional decomposition based on the chain rule of Bayesian networks results in the following joint probability distribution as shown in Equation 5.5:

$$\begin{aligned} &P(PipeType, PipeDamage, RepairTime, DamIdeInd, DamIdeRecTime) \\ &= P(PipeType).P(PipeDamage).P(RepairTime|PipeType, PipeDamage). \\ &\qquad P(DamIdeInd).P(DamIdeRecTime|RepTime, DamIdeInd) \end{aligned} \tag{5.5}$$

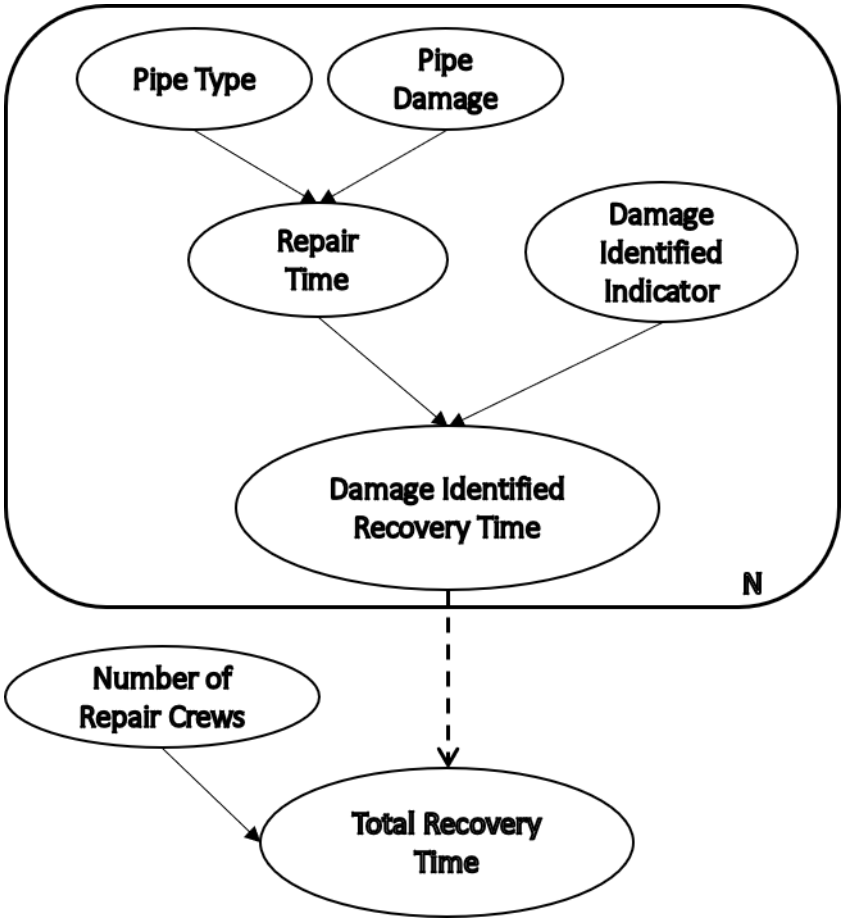


Figure 5-6 Bayesian network for the entire Napa water pipe system

The pipe recovery Bayesian network is modeled using the Structural Modeling, Inference, and Learning Engine (SMILE), which is a library of C++ classes [242] that was developed to support creating, editing, saving and loading graphical models. The Bayesian updating is also performed using SMILE. Since the PBDES model is written in Python, the SMILE wrapper for python (PySMILE) is used to integrate the functionalities of the two modules (functional restoration and Bayesian updating) and create an end-to-end simulation system.

5.8 Dynamic Updating of the Post-Earthquake Total Recovery Time and Trajectory

The primary objective of this study is to create a dynamic prediction system for estimating the total post-earthquake recovery time and trajectory for the entire pipe network of the water distribution system. Two alternative approaches are formulated. The first, which is described as the recovery duration dynamic prediction (RD-DP) model, only provides updated estimates of the time to repair all pipes in the network at regular time intervals (e.g. daily, weekly). The recovery trajectory dynamic prediction (RT-DP) model provides updated forecasts of the future

repair curves (similar to Figure 2-7) at regular time intervals. Both the RD-DP and RT-DP models use information collected during inspections (e.g. distribution of pipe damage) and repairs (e.g. repair duration) from the time of the earthquake to the prediction point, to update recovery time and trajectory estimates, respectively. To demonstrate the framework, the "actual" pipe damage and repair durations are taken from the 2014 South Napa earthquake dataset.

For the RD-DP model, the prior probability distribution for each node of the Bayesian network is assigned based on the results of the damage simulation immediately after the earthquake (Day 0). As time proceeds (e.g. at Day 10), evidence is collected on the time to repair a subset of pipes. More specifically, some pipes will be inspected and possibly repaired between Day 0 and Day 9. Some pipes that were simulated as damaged at Day 0 may or may not be actually damaged. Others that were simulated as undamaged at Day 0 may be encountered as actually damaged between Day 0 and Day 9. These new pieces of evidence will be used to update the states and conditional probability distributions, which will change the estimate of total recovery time. The updating procedure will continue until all pipes that were actually damaged during the earthquake are repaired.

Figure 5-7 shows a 3-dimensional plot of the predicted total recovery time distribution at 10 day intervals after the earthquake. One horizontal axis is used to represent the number of days following the earthquake. The second horizontal axis represents the total recovery time and the vertical axis represents the associated probability distribution function. The actual total recovery time observed during the 2014 South Napa earthquake is 156 days. At Day 0, the total recovery time prediction is normally distributed, with a mean of 48 days and a standard deviation of 12 days. The Day 0 prediction is based on the priori value of pipe damage (simulated by coupling ground shaking intensities with damage fragility functions). As the restoration progresses and actual pipe damage and repair time information is collected, the total repair time prediction is improved, which is evident by the convergence of the mean total repair time prediction to the actual value and a decrease in standard deviation.

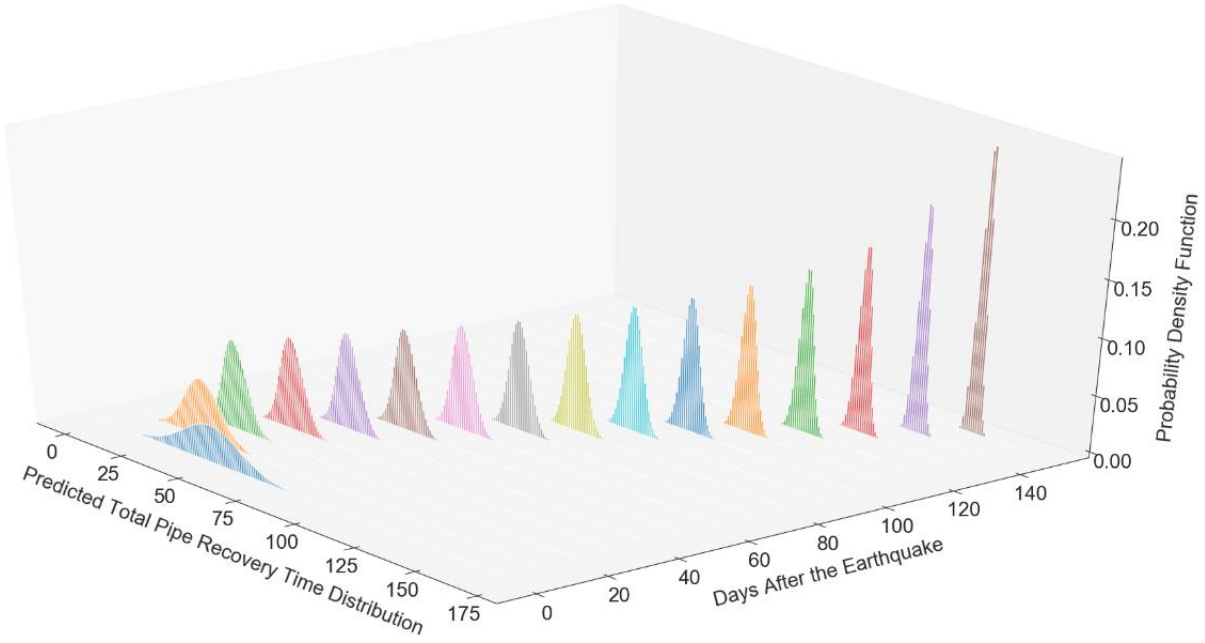


Figure 5-7 Predicted total pipe recovery time distribution at different time-points after the earthquake

As noted earlier, the RD-DP model provides temporally updated estimates of the time to repair all damaged pipes in the network using the knowledge acquired for the distribution of pipe damage and recovery progress. However, the RD-DP model cannot generate a complete recovery curve (like the one shown in Figure 2-7). For the RT-DP model, PBDES is used to simulate the functional restoration process at regular time intervals using the updated input parameters. The probability distribution of the event durations is updated based on the information collected between the time of the earthquake (Day 0) and the time of the new prediction (e.g. Day 10). Naturally, the dynamic recovery curve that is generated by the RT-DP model is more informative than just the total recovery time estimate because the relevant authorities can use it to assess their resource availability and protocols for the remainder of the restoration time horizon. However, it also requires more human and computational resources to construct, update and re-run the PBDES model. Figure 5-8 shows the output of the RT-DP model where the black line represents the actual recovery curve, and each greyscale (darkens over time) curve represent a single trajectory generated at regular one day intervals. Note that Monte Carlo simulation (1000 realizations) is used to generate the PBDES-based trajectories. In other words, for a single realization, the inspection and repair duration values are sampled from their respective distributions and a single trajectory is simulated. The grey-scale curves in Figure 5-8 represent the median trajectory. The Day 0 predicted recovery curve is observed

to be very different from the observed trajectory. This difference can be quantified by taking the residual sum of squares (RSS) between the predicted and observed trajectories. A plot of this RSS versus the number of days following the earthquake is shown in Figure 5-9. The overall decrease in RSS over time indicates the increase in the accuracy of the predicted trajectory. The temporary drop (followed by an increase) in the RSS at approximately day 12 happens because of "corrections" in the priori damage simulation. In other words, as pipes simulated as damaged are found to be intact (and vice versa) are encountered, the forecasted trajectory bifurcates from the observed one. Because most of the damaged pipes are inspected and repaired during the first 25 days following the earthquake, there is a significant improvement in the predicted trajectory moving forward, which is evidenced by the significant drop in the RSS at Day 24. From that point forward, which also corresponds to a reduction in the rate of pipe repairs, the predicted and observed recovery curves are very much aligned.

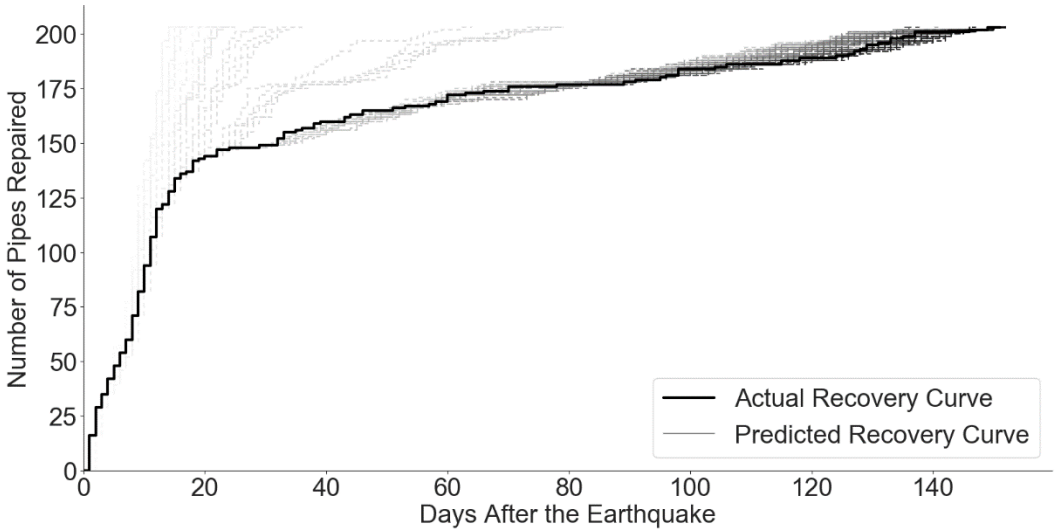


Figure 5-8 Comparing the actual pipe repair trajectory with dynamic predictions generated using the RT-DP model

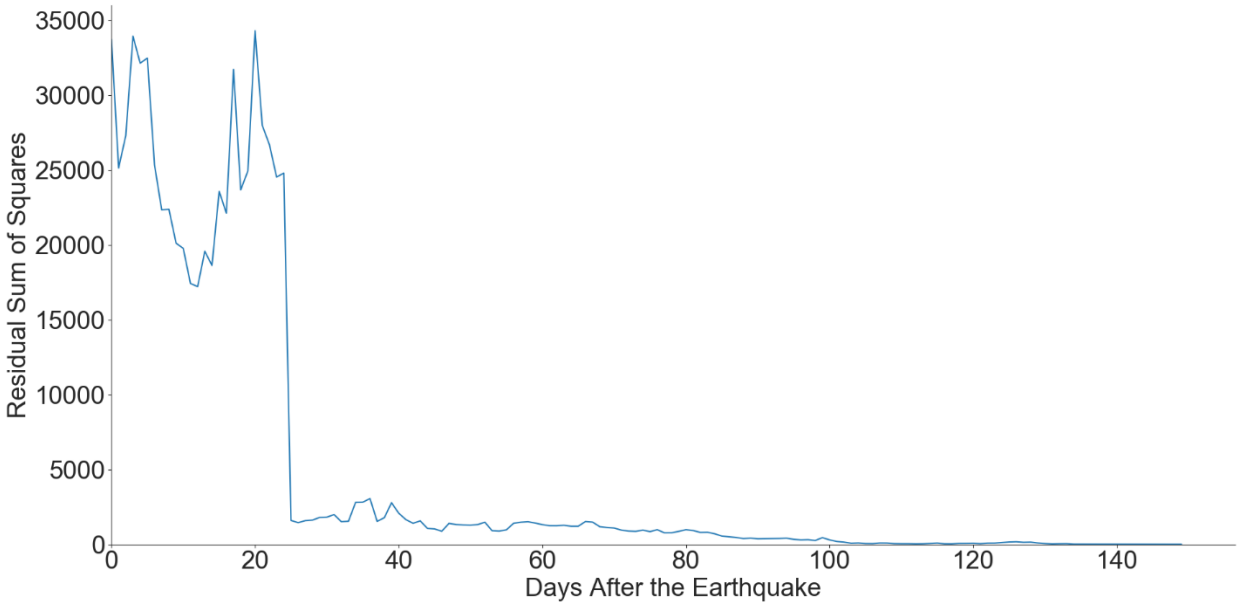


Figure 5-9 Residual sum of squares between predicted and actual pipe repair curves versus the number of days following the earthquake

5.9 Summary and Conclusion

Several sources of uncertainty are inherent in post-earthquake infrastructure functional recovery models such as the distribution of damage caused by the event and the adopted temporal parameters (e.g. time to inspect and repair various components). To reduce these uncertainties over time, a framework is developed to support dynamic updating of post-earthquake functional restoration forecasts for a water distribution system. Central to the proposed methodology is the ability to update forecasts of the total recovery time and/or recovery trajectory as information becomes available during the period following the event. A case study is presented where the framework is applied to the pipe network of the water distribution system in the City of Napa. Only the pipe network is considered because a dataset of damage and repair times are available from the 2014 South Napa earthquake, which was used to validate the proposed methodology.

The proposed framework has three core elements. Bayes' theorem is applied to the pipe damage and repair time parameters that are embedded in the recovery model. These parameter-updates are embedded in a Bayesian Network (BN) of the piping system, which provides dynamic estimates of the total recovery time for the entire piping system. A key limitation of this approach, which is described as a recovery duration dynamic prediction (RD-DP) model, is that the total restoration time is the only output i.e. no recovery curve is provided.

Alternatively, a recovery trajectory dynamic prediction (RT-DP) model provides estimates of the entire restoration curve. This approach relies on time-dependent post-earthquake functionality estimates generated by a process-based discrete-event simulation (PBDES) model whose temporal and damage input parameters are updated over time (also using Bayes' theorem). The disadvantage of the RT-DP model is that it requires more computational resources than the BN-based RD-DP model.

The proposed framework was validated by comparing the dynamic estimates of the recovery trajectory (from the RT-DP model) and total duration (from the RD-DP model) to observations following the 2014 South Napa earthquake. Improved estimates of the probabilistic total recovery duration from the RD-DP model was evidenced by convergence of the mean predicted and observed values over time as well as a reduction in the dispersion of the predicted value. Similarly, by quantifying the residual sum of squares for the predicted and observed trajectories, the RT-DP forecasts was shown to improve over time. In fact, near-perfect convergence between the observed and predicted trajectories was obtained within 20 days of the earthquake.

The proposed framework will enable decision-makers to determine the level and timing of resource allocation to different restorative activities at different periods following a disruptive event, by utilizing the most recent data and knowledge. It can also be used for planning and to train crew members by creating a suite of earthquake scenarios and the corresponding recovery response. A key limitation in the current study is that the methodology was only applied to the pipe system of the water network. Future efforts should focus on extending dynamic recovery estimates to other components of the network (e.g. tanks, pumps, treatment plants) while considering serviceability-related metrics (e.g. hydraulic performance of the system).

CHAPTER 6 Conclusion, Limitations and Future Work

6.1 Overview

The primary objective of the study is to present a comprehensive framework for quantifying the seismic risk and resilience of water distribution systems integrating probabilistic assessment of seismic hazard, damage, restoration, and decision-making. It focused on four main objectives: (1) developing a DES model to hindcast post-earthquake functional loss and restoration of water distribution systems including performance measures based on hydraulic analysis; (2) capturing uncertainties in the seismic hazard, vulnerability and impact of water distribution systems by developing a stochastic event-based framework along with dispersion disaggregation by sources of uncertainty; (3) developing a computationally efficient framework for selecting a subset of damage maps, corresponding ground motion maps, and associated occurrence rates for a probabilistic distributed infrastructure system risk assessment using active learning; and (4) developing a framework to capture relevant uncertainties in the outcomes of the DES model and dynamically reduce them using real-time data. More specifically, the issues addressed in this dissertation can be summarized as follows.

6.2 Findings

6.2.1 Chapter 2: Hindcasting the Functional Loss and Restoration of the Napa Water System Following the 2014 Earthquake using Discrete Event Simulation

Chapter 2 focused on models for simulating post-earthquake functional loss and restoration of water distribution systems supporting the development and implementation of more informed resilience-enhancing strategies. The Discrete Event Simulation (DES) model along with pipe damage and repair data is used to hindcast the water system damage and functional loss, and restoration following the 2014 South Napa earthquake and extended to simulate the post-earthquake disruption and restoration of the City of Napa's water supply for a magnitude 6.7 event on the West Napa fault. By propagating the uncertainties in the damage to the network components (e.g., pipe) conditioned on the shaking intensity and the temporal (e.g., repair time) and resource-related (e.g., the number of inspection crews) variables, probabilistic descriptions of network component repair and water service restoration are generated. The probabilistic outcomes were described using restoration "clouds", which show single-realization and central

tendency trajectories. The results from a two-sample Kolmogorov-Smirnov test showed that the normal distribution is appropriate for probabilistically describing the time needed to achieve specific levels of cumulative pipe repair and SSI. The developed model can be used to inform decision-making in the pre- and post-earthquake environment.

6.2.2 Chapter 3: Risk-Based Assessment of the Post-Earthquake Functional Disruption and Restoration of a Water Distribution System

Event-based methods have been used to assess the risk of infrastructure systems by modeling the hazard, corresponding damage, and, in some cases, functional restoration. Chapter 3 discussed a stochastic event-set-based framework for simulating the post-earthquake functional loss and restoration of water distribution systems and was used to conduct a risk-based assessment of the city of Napa's water distribution system. A stochastic event set catalogue comprised of 2,343 scenarios was assembled using the UCERF2 seismic source model. For each scenario, 50 spatially correlated ground motion maps were generated to obtain shaking intensities at the locations of the individual components that make up the water network. Using Monte Carlo simulation, a set of 1000 damage maps is produced for each scenario using the randomly sampled (from the set of 50) ground motion maps. The damaged state of each component (pipes, tanks and pump stations) in the network is represented by each damage map. Only 135 scenarios with magnitudes ranging from 6.25 to 8.0 caused damage to the Napa water network.

A pressure-driven hydraulic simulation was performed at regular time intervals over the entire duration of the restoration process to quantify the functional recovery in terms of the system serviceability index (SSI). This formed the basis of the network performance assessment using metrics related to the initial loss of SSI (SSI_{IL}), the time to restore the pre-event SSI ($T_{100\%SSI}$) and the cumulative SSI loss (SSI_{CL}). The results indicated that approximately 2.5% of the damage maps from the entire stochastic catalogue resulted in post-earthquake water service disruption. Additionally, 95% of the damage maps achieved full SSI restoration within 3 days i.e. $T_{100\%SSI} \leq 3$ days.

Dispersion disaggregation was used to isolate the individual contributions of multiple random variables to the uncertainty bounds in the restoration trajectory. Disaggregation of the dispersion in the network performance outcomes was used compare the relative contribution of different sources of uncertainty including ground shaking intensity, component fragility, crew size and event duration. When all sources of uncertainty were considered, the dispersion

associated with recovery time metrics (e.g. $T_{100\%SSI}$) was found to be higher than that of SSI_{IL} and SSI_{CL}. The results from the disaggregation showed that the ground shaking intensity had the highest contribution to the dispersion in the restoration time at lower SSI levels (e.g. time to restore 25% and 50% of pre-event level). Whereas at higher SSI levels (75% and 100% of pre-event level), the number of crew members dominated the dispersion in the restoration time. In general, the dispersion in restoration time increases with the SSI level.

The study also discussed an extension of the Miller and Baker (2012) optimization-based map selection algorithm that includes multiple network performance metrics in the objective function. Through a case study of Napa water distribution system, I demonstrated how to use the optimization formulation to select a subset of damage maps (with corresponding ground-motion intensity maps and occurrence rates) from a larger set of candidate maps. We have also shown that the results from the subset are a good estimate of the results from an extensively sampled baseline set of maps. The proposed framework can be used as a resilience-based risk-informed decision-making tool to provide much-needed information to stakeholders and decision-makers.

6.2.3 Chapter 4: Active Learning Framework for Risk Assessment of Distributed Infrastructure Systems

This study presented an active learning framework for selecting a subset of hazard scenarios for infrastructure risk assessment. Uncertainty sampling was used to select a fraction of events from a stochastic catalogue that reasonably estimated the full probability distribution of the system performance. Active learning enables the efficient training of a gaussian process predictive model by choosing the data from which it learns. The framework was illustrated with a case study of the Napa water distribution system where a risk-based assessment of the post-earthquake functional loss and recovery is performed. A subset of earthquake scenarios is sequentially selected using a variance reduction stopping criterion. The full probability distribution and annual exceedance curves of the network performance metrics are shown to be reasonably estimated.

Through, the case study of the Napa water distribution system, I demonstrated how to use the active learning framework to select a subset of ground motion maps from a larger set of candidate maps and model the distribution of performance measures. It was shown that the results from the subset are a good estimate of the results from an extensively sampled baseline set of maps. For example, the annual exceedance rates for $T_{100\%SSI}$ from the selected subset of

919 maps using active learning matched the empirical set of 2343 maps with an R^2 score of 0.92. The estimation brings significant computational gain in the risk assessment process as some network performance metrics are computationally expensive to calculate. Previous research on map selection methods use the baseline network performance results to select the subset of maps which defeats the purpose of map selection as one would need to have the baseline for map the selection to work. Active learning addresses this limitation by making selections based on observed points rather than the baseline results which makes it a more generalized method of map selection. Using these map subsets, policy makers can closely estimate the exceedance rates of a target performance measure and efficiently the analyze seismic risk. The subset of maps can also be used to conduct a "what-if" scenario approach using an event-based probabilistic loss estimation model for assessing risk efficiently and identifying critical aspects of restoration. This framework can also be easily extended to analyze risk from other natural hazards impacting the infrastructure system.

6.2.4 Chapter 5: Dynamic Updating of Post-Earthquake Damage and Functional Restoration Forecasts of Water Distribution Systems using Bayesian Inferencing

Chapter 5 discussed a framework for dynamically updating post-earthquake functional recovery forecasts using Bayesian inferencing to reduce the inherent uncertainties in the outcomes of DES model. The framework comprised of two models: (i) a Bayesian Network (BN) model which is used to provide estimates of the total recovery time and (ii) a process-based discrete event simulation (PBDES) model which is used to generate forecasts of the complete recovery trajectory. Both models rely on component damage and temporal input parameters that are dynamically updated using Bayes theorem, as information becomes available throughout the recovery process. The proposed framework was validated by comparing the dynamic estimates of the recovery trajectory (from the RT-DP model) and total duration (from the RD-DP model) to observations following the 2014 South Napa earthquake. Improved estimates of the probabilistic total recovery duration from the RD-DP model was evidenced by convergence of the mean predicted and observed values over time as well as a reduction in the dispersion of the predicted value. Similarly, by quantifying the residual sum of squares for the predicted and observed trajectories, the RT-DP forecasts was shown to improve over time. In fact, near-perfect convergence between the observed and predicted trajectories was obtained within 20 days of the earthquake. Also, despite a crude initial estimate, the median trajectory generated by the PBDES model provides a reasonable approximation of the observed recovery within thirty days following the earthquake. The

proposed framework will enable decision-makers to determine the level and timing of resource allocation to different restorative activities at different periods following a disruptive event, by utilizing the most recent data and knowledge. It can also be used for planning and to train crew members by creating a suite of earthquake scenarios and the corresponding recovery response.

6.3 Limitations and future work

- One of the main focusses of the current study was to formulate a comprehensive framework for hindcasting post-earthquake damage and functional recovery including performance measures based on restoration hydraulic analysis. It is important to note that the key assumptions, and model parameters were established based on discussions with the Napa Water Division and therefore the findings of the present study in terms of potential seismic resilience and recovery are limited to Napa water distribution system. More studies with a variety of water distribution systems, ground motions and sites with different regional characteristics are needed to quantify the impact of earthquakes on post-earthquake recovery performance more broadly.
- The models used in the current study were manually tuned using a dataset from a single (moderate) earthquake and characteristic region (small city). To develop more generalized models, further studies are needed where variations in the regional context and size and impacts from the earthquake are incorporated (e.g., urban vs rural regions, different scales of ground shaking and damage etc.).
- Future work can extend the framework to incorporate multiple hazards, operational adaptations, other lifeline systems including their interaction and interdependence, other network performance indicators (like centrality measures) and economic loss assessment.
- The ctive learning framework was demonstrated on a 1-dimensional input vector using uncertainty sampling as the acquisition function and radial-basis function as the kernel of the gaussian process regression model. Future work can extend the framework to incorporate multiple dimensions in the input vector (i.e., considering multiple vectors in input like ground motion intensity (PGA, PGV), spatial correlation, network summary metrics (like centrality)). As the dimension of the input vector increases, advanced sampling methods (like query by committee, expected error reduction, hierarchical sampling) in place of uncertainty sampling

should be explored. More research in selecting kernels for the $\mathcal{GPR}$ model is needed which can incorporate domain expertise and prior knowledge. The stopping criterion for the active learning framework is another area for exploration where adaptive criteria (learner-decided threshold) could prove to be more generalizable.

- The Bayesian updating framework was only applied to the pipe system of the Napa water network. Future efforts should focus on extending dynamic recovery estimates to other components of the network (e.g., tanks, pumps, treatment plants) while considering serviceability-related metrics (e.g., hydraulic performance of the system).

References

[1] Sarma H, Russell R. Second day of India's electricity outage hits 620 million. USA Today Http//Usatoday30 Usatoday Com/News/World/Story/2012-07-31/Indiapower-Outage/56600520/1 2012.

[2] Rose A, Lim D. Business interruption losses from natural hazards: conceptual and methodological issues in the case of the Northridge earthquake. Glob Environ Chang Part B Environ Hazards 2002;4:1–14.

[3] Rose A, Liao S-Y. Modeling regional economic resilience to disasters: A computable general equilibrium analysis of water service disruptions. J Reg Sci 2005;45:75–112.

[4] Song J, Ok S-Y. Multi-scale system reliability analysis of lifeline networks under earthquake hazards. Earthq Eng Struct Dyn 2010;39:259–79.

[5] Kaplan S, Garrick BJ. On the quantitative definition of risk. Risk Anal 1981;1:11–27.

[6] Kanamori H. Real-time seismology and earthquake damage mitigation. Annu Rev Earth Planet Sci 2005;33:195–214.

[7] Manuel J. The long road to recovery: environmental health impacts of Hurricane Sandy 2013.

[8] Smith CM, Graffeo CS. Regional impact of Hurricane Isabel on emergency departments in coastal southeastern Virginia. Acad Emerg Med 2005;12:1201–5.

[9] MacKenzie CA, Santos JR, Barker K. Measuring changes in international production from a disruption: Case study of the Japanese earthquake and tsunami. Int J Prod Econ 2012;138:293–302.

[10] DHS. National Infrastructure Preparedness Plan, NIPP 2013. 2013.

[11] Lewis TG. Critical infrastructure protection in homeland security: defending a networked nation. John Wiley & Sons; 2019.

[12] National Academy of Sciences. Disaster Resilience: A National Imperative. Washington, DC: The National Academies Press; 2012. doi:10.17226/13457.

[13] Hosseini S, Barker K, Ramirez-Marquez JE. A review of definitions and measures of system resilience. Reliab Eng Syst Saf 2016;145:47–61.

[14] Vogus TJ, Sutcliffe KM. Organizational resilience: towards a theory and research agenda. 2007 IEEE Int. Conf. Syst. Man Cybern., 2007, p. 3418–22.

[15] Keck M, Sakdapolrak P. What is social resilience? Lessons learned and ways forward. Erdkunde 2013:5–19.

[16] Youn BD, Hu C, Wang P. Resilience-driven system design of complex engineered systems. J Mech Des 2011;133.

[17] Rokneddin K, Ghosh J, Dueñas-Osorio L, Padgett JE. Bridge retrofit prioritisation for ageing transportation networks subject to seismic hazards. Struct Infrastruct Eng 2013;9:1050–66.

[18] Basöz N, Kiremidjian AS. A bridge prioritization method based on transportation system performance using GIS. Proc. 6th US-Japan Work. Earthq. Disaster Prev. Lifeline Syst. Tsukuba Sci. City, Japan, 1995, p. 437–49.

[19] Lee Y-J, Song J, Gardoni P, Lim H-W. Post-hazard flow capacity of bridge transportation network considering structural deterioration of bridges. Struct Infrastruct Eng 2011;7:509–21.

[20] Dueñas-Osorio L, Craig JI, Goodno BJ. Seismic response of critical interdependent networks. Earthq Eng Struct Dyn 2007;36:285–306.

[21] Adachi T, Ellingwood BR. Serviceability of earthquake-damaged water systems: Effects of electrical power availability and power backup systems on system vulnerability. Reliab Eng Syst Saf 2008;93:78–88.

[22] Winkler J, Duenas-Osorio L, Stein R, Subramanian D. Performance assessment of topologically diverse power systems subjected to hurricane events. Reliab Eng Syst Saf 2010;95:323–36.

[23] Shinozuka M, Dong X, Chen TC, Jin X. Seismic performance of electric transmission network under component failures. Earthq Eng Struct Dyn 2007;36:227–44.

[24] Park JI, Lambert JH, Haimes YY. Hydraulic power capacity of water distribution

networks in uncertain conditions of deterioration. Water Resour Res 1998;34:3605–14.

[25] Prasad TD, Park N-S. Multiobjective genetic algorithms for design of water distribution networks. J Water Resour Plan Manag 2004;130:73–82.

[26] Todini E. Looped water distribution networks design using a resilience index based heuristic approach. Urban Water 2000;2:115–22.

[27] Jayaram N, Srinivasan K. Performance-based optimal design and rehabilitation of water distribution networks using life cycle costing. Water Resour Res 2008;44.

[28] Fragiadakis M, Christodoulou SE. Seismic reliability assessment of urban water networks. Earthq Eng Struct Dyn 2014;43:357–74.

[29] Kessler A, Ormsbee L, Shamir U. A methodology for least-cost design of invulnerable water distribution networks. Civ Eng Syst 1990;7:20–8.

[30] Goulter IAN. Analytical and simulation models for reliability analysis in water distribution systems. Improv. Effic. Reliab. water Distrib. Syst., Springer; 1995, p. 235–66.

[31] Romero N, O'Rourke TD, Nozick LK, Davis CA. Seismic hazards and water supply performance. J Earthq Eng 2010;14:1022–43.

[32] Yoo DG, Jung D, Kang D, Kim JH, Lansey K. Seismic hazard assessment model for urban water supply networks. J Water Resour Plan Manag 2015;142:4015055.

[33] Wu J, Dueñas-Osorio L. Calibration and validation of a seismic damage propagation model for interdependent infrastructure systems. Earthq Spectra 2013;29:1021–41.

[34] Li J, He J. A recursive decomposition algorithm for network seismic reliability evaluation. Earthq Eng Struct Dyn 2002;31:1525–39.

[35] Christodoulou SE, Fragiadakis M. Vulnerability assessment of water distribution networks considering performance data. J Infrastruct Syst 2015;21:4014040.

[36] Laucelli D, Giustolisi O. Vulnerability assessment of water distribution networks under seismic actions. J Water Resour Plan Manag 2015;141:4014082.

[37] Tabucchi T, Davidson R, Brink S. Simulation of post-earthquake water supply system

restoration. Civ Eng Environ Syst 2010;27:263–79. doi:10.1080/10286600902862615.

[38] Fujiwara O, Li J. Reliability analysis of water distribution networks in consideration of equity, redistribution, and pressure-dependent demand. Water Resour Res 1998;34:1843–50. doi:10.1029/98WR00908.

[39] Shi P, O'Rourke TD. Seismic response modeling of water supply systems. Mceer-08-0016, 352 2008.

[40] Pathirana A. EPANET2 desktop application for pressure driven demand modeling. Water Distrib. Syst. Anal. 2010, 2010, p. 65–74.

[41] Muranho J, Ferreira A, Sousa J, Gomes A, Marques AS. WaterNetGen: an EPANET extension for automatic water distribution network models generation and pipe sizing. Water Sci Technol Water Supply 2012;12:117–23.

[42] Klise KA, Bynum M, Moriarty D, Murray R. A software framework for assessing the resilience of drinking water systems to disasters with an example earthquake case study. Environ Model Softw 2017;95:420–31. doi:10.1016/j.envsoft.2017.06.022.

[43] González AD, Dueñas-Osorio L, Sánchez-Silva M, Medaglia AL. The interdependent network design problem for optimal infrastructure system restoration. Comput Civ Infrastruct Eng 2016;31:334–50.

[44] Han S-R, Guikema SD, Quiring SM, Lee K-H, Rosowsky D, Davidson RA. Estimating the spatial distribution of power outages during hurricanes in the Gulf coast region. Reliab Eng Syst Saf 2009;94:199–210.

[45] Davis CA. Water system service categories, postearthquake interaction, and restoration strategies. Earthq Spectra 2014;30:1487–509. doi:10.1193/022912EQS058M.

[46] O'Rourke T, Toprak S. Using GIS to assess water supply damage from the Northeidge earthquake 1997. http://mceer.buffalo.edu/publications/bulletin/97/11-03/jul97n3.html (accessed January 25, 2018).

[47] TCLEE. Fire Following Earthquake. Scawthorn, C, J M Eidinger, AJ Schiff (Editors), Tech Counc Lifeline Earthq Eng Monogr No 26 2005:345pp.

[48] Porter KA, Sherrill R. Utility performance panels in the ShakeOut Scenario. Earthq

Spectra 2011;27:443–58.

[49] Scawthorn C, Khater M, Rojahn C, Cluff LS. Seismic Vulnerability and Impact of Disruption of Lifelines in the Conterminous United States. Applied Technology Council; 1991.

[50] Porter KA. “A new model of water-network resilience, with application to the HayWired scenario.” in The HayWired earthquake scenario—Engineering Implications: chap. N edited by S.T. Detweiler, and Wein, A.M.: U.S. Geological Survey Scientific Investigations Report 2. 2018.

[51] Isumi M, Nomura N, Shibuya T. Simulation of post-earthquake restoration of lifeline systems. Int J Mass Emerg Disasters 1985;3:87–105.

[52] Chang SE, Svekla WD, Shinozuka M. Linking infrastructure and urban economy: simulation of water-disruption impacts in earthquakes. Environ Plan B Plan Des 2002;29:281–301.

[53] Kozin F, Zhou H. System study of urban response and reconstruction due to earthquake. J Eng Mech 1990;116:1959–72.

[54] Nojima N, Kameda H. Optimal strategy by use of tree structure for post-earthquake restoration of lifeline network systems. Proc. 10th World Conf. Earthq. Eng., 1992, p. 5541–6.

[55] Tabucchi THP. Modeling Post-earthquake Restoration of the Los Angeles Water Supply System. 2007 2007.

[56] Luna R, Balakrishnan N, Dagli CH. Postearthquake Recovery of a Water Distribution System: Discrete Event Simulation Using Colored Petri Nets. J Infrastruct Syst 2011. doi:10.1061/(ASCE)IS.1943-555X.0000039.

[57] Guidotti R, Chmielewski H, Unnikrishnan V, Gardoni P, McAllister T, van de Lindt J. Modeling the resilience of critical infrastructure: The role of network dependencies. Sustain Resilient Infrastruct 2016;1:153–68.

[58] O’Rourke TD. Critical infrastructure, interdependencies, and resilience. Bridg Acad Eng 2007;37:22.

[59] Bruneau M, Eeri M, Chang SE, Eeri M, Ronald T, Eeri M, et al. A Framework to Quantitatively Assess and Enhance the Seismic Resilience of Communities 2003;19:733–52. doi:10.1193/1.1623497.

[60] Franchin P. A computational framework for systemic seismic risk analysis of civil infrastructural systems. SYNER-G Syst. Seism. vulnerability risk Assess. complex urban, Util. lifeline Syst. Crit. Facil., Springer; 2014, p. 23–56.

[61] Ellingwood BR, Cutler H, Gardoni P, Peacock WG, van de Lindt JW, Wang N. The centerville virtual community: A fully integrated decision model of interacting physical and social infrastructure systems. Sustain Resilient Infrastruct 2016;1:95–107.

[62] U.S. Geological Survey (and California - California Geological Survey). Quaternary fault and fold database for the United States 2018. https://earthquake.usgs.gov/hazards/qfaults/ (accessed April 26, 2019).

[63] Napa Water Division. Urban Water Management Plan: 2015 Update. 2017.

[64] Johnson LA, Mahin SA. The Mw 6.0 South Napa Earthquake of August 24, 2014 2016.

[65] CESMD. Center for engineering strong motion data 2012.

[66] Worden CB, Thompson EM, Baker JW, Bradley BA, Luco N, Wald DJ. Spatial and Spectral Interpolation of Ground-Motion Intensity Measure Observations. Bull Seismol Soc Am 2018;108:866–75.

[67] Eidinger J. South Napa M 6.0 Earthquake of August 24, 2014. G&E Eng Syst March 2015;17.

[68] SPA Risk LLC. 24 August 2014 South Napa Mw 6 Earthquake Reconnaissance Report. 2014.

[69] SPA Risk LLC. 24 August 2014 South Napa Mw 6 Earthquake Reconnaissance Report. 2014.

[70] Varga A. Discrete event simulation system. Proc. Eur. Simul. Multiconference, 2001.

[71] Fishman GS. Principles of Discrete Event Simulation 1978.

[72] Fishman GS. Discrete-Event Simulation: Modeling, Programming, and Applications

2001.

[73] Cagnan Z. Post-earthquake restoration modeling for critical lifeline systems. 2005.

[74] Müller K, Vignaux T. SimPy: Simulating Systems in Python. ONLampCom Python Devcenter 2003.

[75] Eguchi RT. Seismic vulnerability models for underground pipes. Proc., Int. Symp. Earthq. Behav. Saf. Oil Gas Storage Facil. Buried Pipelines Equip., 1983, p. 368–73.

[76] O'Rourke TD. An overview of geotechnical and lifeline earthquake engineering. Geotech. Earthq. Eng. soil Dyn. III, 1998, p. 1392–426.

[77] Toprak S. Earthquake effects on buried lifeline systems. 1998.

[78] Toprak S, Taskin F. Estimation of earthquake damage to buried pipelines caused by ground shaking. Nat Hazards 2007;40:1–24.

[79] Nyman DJ, others. Guidelines for the seismic design of oil and gas pipeline systems. Comm Gas Liq Fuel Lifelines Am Soc Civ Eng New York, 473p 1984.

[80] Fragiadakis M, Christodoulou SE, Vamvatsikos D. Reliability Assessment of Urban Water Distribution Networks Under Seismic Loads. Water Resour Manag 2013;27:3739–64. doi:10.1007/s11269-013-0378-0.

[81] Adachi T, Ellingwood BR. Serviceability assessment of a municipal water system under spatially correlated seismic intensities. Comput Civ Infrastruct Eng 2009;24:237–48.

[82] Wang Y, O'Rourke TD. Seismic performance evaluation of water supply systems 2006.

[83] U S Nuclear Regulatory Commission. Reactor Safety Study. Wash 1400. Washington, DC 1975.

[84] American Lifelines Alliance. Seismic Fragility Formulation for Water Systems: Guideline. Am Lifelines Alliance 2001. doi:10.1007/s13398-014-0173-7.2.

[85] Pitilakis K, Crowley H, Kaynia AM. SYNER-G: Typology Definition and Fragility Functions for Physical Elements at Seismic Risk: Buildings, Lifelines, Transportation Networks and Critical Facilities. Geotech Geol Earthq Eng 2014. doi:10.1007/978-94-007-7872-6.

[86] FEMA FEMA. Hazus-MH 2.1 Technical Manual: Earthquake Model 2012.

[87] Wald DJ, Worden BC, Quitoriano V, Pankow KL. ShakeMap manual: technical manual, user's guide, and software guide. 2005.

[88] Field EH, Jordan TH, Cornell CA. OpenSHA: A developing community-modeling environment for seismic hazard analysis. Seismol Res Lett 2003;74:406–19. doi:10.1785/gssrl.74.4.406.

[89] Hanks TC, Bakun WH. M-log A observations for recent large earthquakes. Bull Seismol Soc Am 2008;98:490–4. doi:10.1785/0120070174.

[90] Wells DL, Coppersmith KJ. New Empirical Relationships among Magnitude, Rupture Length, Rupture Width, Rupture Area, and Surface Displacement. Bull Seismol Soc Am 1994;84:974–1002. doi: .

[91] Campbell KW, Bozorgnia Y. NGA-West2 ground motion model for the average horizontal components of PGA, PGV, and 5% damped linear acceleration response spectra. Earthq Spectra 2014;30:1087–115.

[92] Jayaram N, Baker JW. Correlation model for spatially distributed ground-motion intensities. Earthq Eng Struct Dyn 2009;38:1687–708.

[93] Massey Jr FJ. The Kolmogorov-Smirnov test for goodness of fit. J Am Stat Assoc 1951;46:68–78.

[94] Guidotti R, Gardoni P, Rosenheim N. Integration of physical infrastructure and social systems in communities' reliability and resilience analysis. Reliab Eng Syst Saf 2019;185:476–92.

[95] Wagner JM, Shamir U, Marks DH. Water distribution reliability: simulation methods. J Water Resour Plan Manag 1988;114:276–94. doi:10.1061/(ASCE)0733-9496(1988)114:3(276).

[96] Muranho J, Ferreira A, Sousa J, Gomes A, Marques AS. Pressure-dependent demand and leakage modelling with an EPANET extension--WaterNetGen. Procedia Eng 2014;89:632–9.

[97] Wu ZY, Wang RH, Walski TM, Yang SY, Bowdler D, Baggett CC. Efficient pressure

dependent demand model for large water distribution system analysis. Water Distrib. Syst. Anal. Symp. 2006, 2008, p. 1–15.

[98] Rossman L. Epanet 2 users manual 2000.

[99] Trifunovic N. Pattern Recognition for Reliability Assessment of Water Distribution Networks: UNESCO-IHE PhD Thesis. CRC Press; 2012.

[100] Laucelli D, Berardi L, Giustolisi O. Assessing climate change and asset deterioration impacts on water distribution networks: Demand-driven or pressure-driven network modeling? Environ Model Softw 2012;37:206–16.

[101] McAllister TP. Community resilience planning guide for buildings and infrastructure systems, volume i. 2015.

[102] Freeman LC. A set of measures of centrality based on betweenness. Sociometry 1977:35–41.

[103] Tomar A, Burton H V., Mosleh A, Lee JY. Hindcasting the Functional Loss and Restoration of the Napa Water System Following the 2014 Earthquake using Discrete Event Simulation. ASCE J Infrastruct Syst 2020. doi:10.1061/(ASCE)IS.1943-555X.0000574.

[104] Masoomi H, Burton H, Tomar A, Mosleh A. Simulation-Based Assessment of Postearthquake Functionality of Buildings with Disruptions to Cross-Dependent Utility Networks. J Struct Eng (United States) 2020;146. doi:10.1061/(ASCE)ST.1943-541X.0002555.

[105] Wu J. End-to-end seismic risk analysis framework for the identification of infrastructure network retrofits. Ph. D. dissertation, Dept. of Civil and Environmental Engineering, Stanford Univ, 2017.

[106] Eugster S, Rüttener E, Liechti D. The risk premium distribution (annual average loss) with respect to earthquake magnitude. Proc XXIV Eur Geophys Soc Gen Assem 1999.

[107] Liechti D, Rüttener E, Zbinden A. Disaggregation of average annual loss. XXV Eur Geophys Soc Gen Assem 2000.

[108] Tantala MW, Nordenson GJP, Deodatis G, Jacob K. Earthquake loss estimation for the

New York City metropolitan region. Soil Dyn Earthq Eng 2008;28:812–35.

[109] Windeler D, Morrow G, Williams CR, Rahnama M, Molas G, Peña A, et al. Earthquake risk estimates for residential construction in the western United States. Proc. 13th World Conf. Earthq. Eng., 2004, p. 1–6.

[110] Crowley H, Bommer JJ. Modelling seismic hazard in earthquake loss models with spatially distributed exposure. Bull Earthq Eng 2006;4:249–73.

[111] Miller M, Baker J. Ground-motion intensity and damage map selection for probabilistic infrastructure network risk assessment using optimization. Earthq Eng Struct Dyn 2015;44:1139–56.

[112] Vanzi I. Seismic reliability of electric power networks: methodology and application. Struct Saf 1996;18:311–27.

[113] Cristian F. Synchronous and asynchronous. Commun ACM 1996;39:88–97.

[114] Ostfeld A, Kogan D, Shamir U. Reliability simulation of water distribution systems--single and multiquality. Urban Water 2002;4:53–61.

[115] Masoomi H, Burton H, Tomar A, Mosleh A. Simulation-Based Assessment of Postearthquake Functionality of Buildings with Disruptions to Cross-Dependent Utility Networks. J Struct Eng 2020;146:4020070.

[116] Field EH, Dawson TE, Felzer KR, Frankel AD, Gupta V, Jordan TH, et al. Uniform California earthquake rupture forecast, version 2 (UCERF 2). Bull Seismol Soc Am 2009;99:2053–107.

[117] Field EH, Arrowsmith RJ, Biasi GP, Bird P, Dawson TE, Felzer KR, et al. Uniform California earthquake rupture forecast, version 3 (UCERF3)—The time-independent model. Bull Seismol Soc Am 2014;104:1122–80.

[118] Graves R, Jordan TH, Callaghan S, Deelman E, Field E, Juve G, et al. CyberShake: A Physics-Based Seismic Hazard Model for Southern California. Pure Appl Geophys 2011.

[119] Idriss IM. An NGA-West2 empirical model for estimating the horizontal spectral values generated by shallow crustal earthquakes. Earthq Spectra 2014;30:1155–77.

[120] Boore DM, Stewart JP, Seyhan E, Atkinson GM. NGA-West2 equations for predicting PGA, PGV, and 5% damped PSA for shallow crustal earthquakes. Earthq Spectra 2014;30:1057–85.

[121] Chiou BS-J, Youngs RR. Update of the Chiou and Youngs NGA model for the average horizontal component of peak ground motion and response spectra. Earthq Spectra 2014;30:1117–53.

[122] Abrahamson NA, Silva WJ, Kamai R. Summary of the ASK14 ground motion relation for active crustal regions. Earthq Spectra 2014;30:1025–55.

[123] Rezaeian S, Petersen MD, Moschetti MP, Powers P, Harmsen SC, Frankel AD. Implementation of NGA-West2 ground motion models in the 2014 U.S. National Seismic Hazard Maps. Earthq Spectra 2014;30:1319–33. doi:10.1193/062913EQS177M.

[124] Chang SE, Shinozuka M. Measuring improvements in the disaster resilience of communities. Earthq Spectra 2004;20:739–55.

[125] Cauffman SA. Community resilience planning guide for buildings and infrastructure systems. 2015.

[126] Poland C. The resilient city: Defining what San Francisco needs from its seismic mitigation polices. San Fr Plan Urban Res Assoc Report, San Fr CA, USA 2009.

[127] ATC. Critical Assesment of Lifeline System Performance: Understanding Societal Needs in Disaster Recovery. US Department of Commerce, National Institute of Standards and Technology; 2016.

[128] Boyd S, Boyd SP, Vandenberghe L. Convex optimization. Cambridge university press; 2004.

[129] Diamond S, Boyd S. CVXPY: A Python-embedded modeling language for convex optimization. J Mach Learn Res 2016;17:2909–13.

[130] Joshi S, Boyd S. Sensor selection via convex optimization. IEEE Trans Signal Process 2008;57:451–62.

[131] Peterson WW. Addressing for random-access storage. IBM J Res Dev 1957;1:130–46.

[132] Williams CKI, Rasmussen CE. Gaussian processes for machine learning. vol. 2. MIT press Cambridge, MA; 2006.

[133] Gershman SJ, Blei DM. A tutorial on Bayesian nonparametric models. J Math Psychol 2012;56:1–12.

[134] Duvenaud D. Automatic model construction with Gaussian processes. University of Cambridge, 2014.

[135] Lloyd JR, Duvenaud D, Grosse R, Tenenbaum J, Ghahramani Z. Automatic construction and natural-language description of nonparametric regression models. Twenty-eighth AAAI Conf. Artif. Intell., 2014.

[136] Hensman J, Fusi N, Lawrence ND. Gaussian processes for big data. ArXiv Prepr ArXiv13096835 2013.

[137] Quiñonero-Candela J, Rasmussen CE. A unifying view of sparse approximate Gaussian process regression. J Mach Learn Res 2005;6:1939–59.

[138] Görtler J, Kehlbeck R, Deussen O. A Visual Exploration of Gaussian Processes. Distill 2019;4:e17.

[139] Settles B. Active learning literature survey. 2009.

[140] Angluin D. Queries and concept learning. Mach Learn 1988;2:319–42.

[141] Cohn DA, Ghahramani Z, Jordan MI. Active learning with statistical models. J Artif Intell Res 1996;4:129–45.

[142] King RD, Whelan KE, Jones FM, Reiser PGK, Bryant CH, Muggleton SH, et al. Functional genomic hypothesis generation and experimentation by a robot scientist. Nature 2004;427:247–52.

[143] Cohn D, Atlas L, Ladner R. Improving generalization with active learning. Mach Learn 1994;15:201–21.

[144] Dagan I, Engelson SP. Committee-based sampling for training probabilistic classifiers. Mach. Learn. Proc. 1995, Elsevier; 1995, p. 150–7.

[145] Mitchell TM. Generalization as search. Artif Intell 1982;18:203–26.

[146] Krishnamurthy V. Algorithms for optimal scheduling and management of hidden Markov model sensors. IEEE Trans Signal Process 2002;50:1382–97.

[147] Yu H. SVM selective sampling for ranking with application to data retrieval. Proc. Elev. ACM SIGKDD Int. Conf. Knowl. Discov. data Min., 2005, p. 354–63.

[148] Lewis DD, Gale WA. A sequential algorithm for training text classifiers. SIGIR'94, 1994, p. 3–12.

[149] Tong S, Koller D. Support vector machine active learning with applications to text classification. J Mach Learn Res 2001;2:45–66.

[150] Hoi SCH, Jin R, Lyu MR. Large-scale text categorization by batch mode active learning. Proc. 15th Int. Conf. World Wide Web, 2006, p. 633–42.

[151] Thompson CA, Califf ME, Mooney RJ. Active learning for natural language parsing and information extraction. ICML, 1999, p. 406–14.

[152] Settles B, Craven M. An analysis of active learning strategies for sequence labeling tasks. Proc. 2008 Conf. Empir. Methods Nat. Lang. Process., 2008, p. 1070–9.

[153] Tong S, Chang E. Support vector machine active learning for image retrieval. Proc. ninth ACM Int. Conf. Multimed., 2001, p. 107–18.

[154] Zhang C, Chen T. An active learning framework for content-based information retrieval. IEEE Trans Multimed 2002;4:260–8.

[155] Yang J, others. Automatically labeling video data using multi-class active learning. Proc. Ninth IEEE Int. Conf. Comput. Vis., 2003, p. 516–23.

[156] Hauptmann AG, Lin W-H, Yan R, Yang J, Chen M-Y. Extreme video retrieval: joint maximization of human and computer performance. Proc. 14th ACM Int. Conf. Multimed., 2006, p. 385–94.

[157] Tur G, Hakkani-Tür D, Schapire RE. Combining active and semi-supervised learning for spoken language understanding. Speech Commun 2005;45:171–86.

[158] Liu Y. Active learning with support vector machine applied to gene expression data for cancer classification. J Chem Inf Comput Sci 2004;44:1936–41.

[159] Seeger M. Bayesian Gaussian process models: PAC-Bayesian generalisation error bounds and sparse approximations. 2003.

[160] Lewis DD, Catlett J. Heterogeneous uncertainty sampling for supervised learning. Mach. Learn. Proc. 1994, Elsevier; 1994, p. 148–56.

[161] Krause A, Singh A, Guestrin C. Near-optimal sensor placements in Gaussian processes: Theory, efficient algorithms and empirical studies. J Mach Learn Res 2008;9:235–84.

[162] Scheffer T, Decomain C, Wrobel S. Active hidden markov models for information extraction. Int. Symp. Intell. Data Anal., 2001, p. 309–18.

[163] Culotta A, McCallum A. Reducing labeling effort for structured prediction tasks. AAAI, vol. 5, 2005, p. 746–51.

[164] Cressie NAC. Statistics for spatial data John Wiley and Sons New York 900 1991.

[165] MacKay DJC. Information-based objective functions for active data selection. Neural Comput 1992;4:590–604.

[166] Zhu J, Wang H, Hovy E, Ma M. Confidence-based stopping criteria for active learning for data annotation. ACM Trans Speech Lang Process 2010;6:1–24.

[167] Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: Machine learning in Python. J Mach Learn Res 2011;12:2825–30.

[168] Danka T, Horvath P. modAL: A modular active learning framework for Python. ArXiv Prepr ArXiv180500979 2018.

[169] Hoshiya M, others. Seismic Damage Restoration of Underground Water Pipe Lines. Proc. Rev. Meet. US-Japan Coop. Res. Seism. Risk Anal. Its Appl. to Reliab. Des. Lifeline Syst. Honolulu, 1981, p. 229.

[170] O'Rourke M, Ayala G. Pipeline Damage Due to Wave Propagation. J Geotech Eng 1993;119:1490–8. doi:10.1061/(ASCE)0733-9410(1993)119:9(1490).

[171] Clough DP. Experimental evaluation of seismic design methods for broad cylindrical tanks 1977.

[172] O'Rourke MJ, So P. Seismic fragility curves for on-grade steel tanks. Earthq Spectra

2000;16:801–15.

[173] Malhotra PK, Wenk T, Wieland M. Simple procedure for seismic analysis of liquid-storage tanks. Struct Eng Int 2000;10:197–201.

[174] Wagner BJM, Shamir U, Marks DH. WATER DISTRIBUTION RELIABILITY: ANALYTIC METHODS. J Water Resour Plan Manag 1988;114:253–75. doi:10.1061/(ASCE)0733-9496(1988)114:3(276).

[175] Hwang HHM, Lin H, Shinozuka M. Seismic performance assessment of water delivery systems. J Infrastruct Syst 1998;4:118–25.

[176] Markov I, Grigoriu M, O'Rourke TD. An evaluation of seismic serviceability of water supply networks with application to the San Francisco auxiliary water supply system 1994.

[177] Poljanšek K, Bono F, Gutiérrez E. Seismic risk assessment of interdependent critical infrastructure systems: The case of European gas and electricity networks. Earthq Eng Struct Dyn 2012;41:61–79.

[178] Albert R, Albert I, Nakarado GL. Structural vulnerability of the North American power grid. Phys Rev E 2004;69:25103.

[179] Rose A. Defining and measuring economic resilience to disasters. Disaster Prev Manag An Int J 2004;13:307–14.

[180] Kang D, Lansey K. Post-earthquake restoration of water supply infrastructure. World Environ. Water Resour. Congr. 2013 Showcasing Futur., 2013, p. 913–22.

[181] Davidson RA, Cagnan Z. Discrete event simulation of the post-earthquake restoration process for electric power systems. Int J Risk Assess Manag 2007;7:1138–56. doi:10.1504/IJRAM.2007.015298.

[182] Chang SE, Shinozuka M, Svekla W. Modeling post-disaster urban lifeline restoration. Optim. Post-Earthquake Lifeline Syst. Reliab., 1999, p. 602–11.

[183] Çağnan Z, Davidson RA, Guikema SD. Post-earthquake restoration planning for Los Angeles electric power. Earthq Spectra 2006;22:589–608. doi:10.1193/1.2222400.

[184] Lee JY, Tomar A, Burton H V. A framework for water distribution system exposed to

seismic events and evolving conditions 2018.

[185] Wright S. Correlation and causation. J Agric Res 1921;20:557–85.

[186] Pearl J. Probabilistic reasoning in intelligent systems: networks of plausible inference. Elsevier; 2014.

[187] Schachter RD, Heckerman D. Thinking backward for knowledge acquisition. AI Mag 1987;8:55.

[188] Pearl J. Graphs, causality, and structural equation models. Sociol Methods Res 1998;27:226–84.

[189] Shafer G, Pearl J. Readings in uncertain reasoning. Morgan Kaufmann Publishers Inc.; 1990.

[190] Heckerman D, Geiger D. Learning Bayesian networks: a unification for discrete and Gaussian domains. Proc. Elev. Conf. Uncertain. Artif. Intell., 1995, p. 274–84.

[191] Mittal A. Bayesian network technologies: applications and graphical models: applications and graphical models. IGI Global; 2007.

[192] Koller D, Friedman N, Džeroski S, Sutton C, McCallum A, Pfeffer A, et al. Introduction to statistical relational learning. MIT press; 2007.

[193] Haug PJ, Koehler SB, Christensen LM, Gundersen ML, Van Bree RE. Probabilistic system for natural language processing 2003.

[194] Russell SJ, Norvig P. Artificial intelligence: a modern approach. Malaysia; Pearson Education Limited,; 2016.

[195] Gámez JA, Moral S, Cerdan AS. Advances in Bayesian networks. vol. 146. Springer; 2013.

[196] Friis-Hansen A, Friis-Hansen P, Pedersen PT. Bayesian networks as a decision support tool in marine applications. Technical University of Denmark (DTU); 2001.

[197] Friis-Hansen P. Structuring of complex systems using Bayesian network. Proc. Two-Part Work. DTU, 2004, p. 23–5.

[198] Weber P, Medina-Oliva G, Simon C, Iung B. Overview on Bayesian networks

applications for dependability, risk analysis and maintenance areas. Eng Appl Artif Intell 2012;25:671–82.

[199] Yazdani A, Shahidzadeh M-S, Takada T. Bayesian networks for disaggregation of structural reliability. Struct Saf 2020;82:101892.

[200] Havbro Faber M, Kroon IB, Kragh E, Bayly D, Decosemaeker P. Risk assessment of decommissioning options using Bayesian networks. J Offshore Mech Arct Eng 2002;124:231–8.

[201] Mahadevan S, Rebba R. Validation of reliability computational models using Bayes networks. Reliab Eng Syst Saf 2005;87:223–32.

[202] Grêt-Regamey A, Straub D. Spatially explicit avalanche risk assessment linking Bayesian networks to a GIS 2006.

[203] Nishijima K, Maes MA, Goyet J, Faber MH. Constrained optimization of component reliabilities in complex systems. Struct Saf 2009;31:168–78.

[204] Bensi MT, Straub D, Friis-Hansen P, Der Kiureghian A. Modeling infrastructure system performance using BN. Proc. ICOSSAR 2009, 2009.

[205] Bobbio A, Portinale L, Minichino M, Ciancamerla E. Improving the analysis of dependable systems by mapping fault trees into Bayesian networks. Reliab Eng Syst Saf 2001;71:249–60.

[206] Langseth H, Portinale L. Bayesian networks in reliability. Reliab Eng Syst Saf 2007;92:92–108.

[207] Neil M, Tailor M, Marquez D, Fenton N, Hearty P. Modelling dependable systems using hybrid Bayesian networks. Reliab Eng Syst Saf 2008;93:933–9.

[208] Dagum P, Galper A, Horvitz E. Dynamic network models for forecasting. Proc. eighth Int. Conf. Uncertain. Artif. Intell., 1992, p. 41–8.

[209] Tahboub KA. Intelligent human-machine interaction based on dynamic bayesian networks probabilistic intention recognition. J Intell Robot Syst 2006;45:31–52.

[210] Grimes DB, Chalodhorn R, Rao RPN. Dynamic Imitation in a Humanoid Robot through Nonparametric Probabilistic Inference. Robot. Sci. Syst., 2006, p. 199–206.

[211] Aggarwal CC, Zhai C. Mining text data. Springer Science & Business Media; 2012.

[212] Chapelle O, Zhang Y. A dynamic bayesian network click model for web search ranking. Proc. 18th Int. Conf. World wide web, 2009, p. 1–10.

[213] Bayes T. An essay towards solving a problem in the doctrine of chances. 1763. MD Comput Comput Med Pract 1991;8:157.

[214] Schlaifer R, Raiffa H. Applied statistical decision theory. 1961.

[215] Benjamin JR, Cornell CA. Probability, statistics, and decision for civil engineers. Courier Corporation; 2014.

[216] Ang AH-S, Tang WH. Probability concepts in engineering planning and design, vol. 2: Decision, risk, and reliability. JOHN WILEY SONS, INC, 605 THIRD AVE, NEW YORK, NY 10158, USA, 1984, 608 1984.

[217] Singhal A, Kiremidjian AS. Bayesian updating of fragilities with application to RC frames. J Struct Eng 1998;124:922–9.

[218] Choe D-E, Gardoni P, Rosowsky D. Closed-form fragility estimates, parameter sensitivity, and Bayesian updating for RC columns. J Eng Mech 2007;133:833–43.

[219] Huang Q, Gardoni P, Hurlebaus S. Probabilistic seismic demand models and fragility estimates for reinforced concrete highway bridges with one single-column bent. J Eng Mech 2010;136:1340–53.

[220] Der Kiureghian A. A Bayesian framework for fragility assessment. Proc., ICASP8 Conf, 1999, p. 1003–10.

[221] Der Kiureghian A. Bayesian methods for seismic fragility assessment of lifeline components. Accept Risk Process Nat Hazards 2002:61–77.

[222] Engelund S, Sørensen JD. A probabilistic model for chloride-ingress and initiation of corrosion in reinforced concrete structures. Struct Saf 1998;20:69–89.

[223] Enright MP, Frangopol DM. Condition prediction of deteriorating concrete bridges using Bayesian updating. J Struct Eng 1999;125:1118–25.

[224] Geyskens P, Kiureghian A Der, Monteiro P. Bayesian prediction of elastic modulus of

concrete. J Struct Eng 1998;124:89–95.

[225] Sørensen JD. Reliability-based inspection planning for structural systems. Reliab. Optim. Struct. Syst., 1993, p. 31–46.

[226] Zheng R, Ellingwood BR. Role of non-destructive evaluation in time-dependent reliability analysis. Struct Saf 1998;20:325–39.

[227] Zhang R, Mahadevan S. Model uncertainty and Bayesian updating in reliability-based inspection. Struct Saf 2000;22:145–60.

[228] Fan W, Ang AH-S, Li Z. Reliability assessment of deteriorating structures using Bayesian updated probability density evolution method (PDEM). Struct Saf 2017;65:60–73.

[229] Tien I, Der Kiureghian A. Algorithms for Bayesian network modeling and reliability assessment of infrastructure systems. Reliab Eng Syst Saf 2016;156:134–47.

[230] Tien I, Der Kiureghian A. Reliability assessment of critical infrastructure using Bayesian networks. J Infrastruct Syst 2017;23:4017025.

[231] Bensi MT, Der Kiureghian A, Straub D. A Bayesian network methodology for infrastructure seismic risk assessment and decision support 2011.

[232] Hosseini S, Barker K. Modeling infrastructure resilience using Bayesian networks: A case study of inland waterway ports. Comput Ind Eng 2016;93:252–66.

[233] Abrahamson NA. Seismic hazard assessment: problems with current practice and future developments. First Eur. Conf. Earthq. Eng. Seismol., 2006, p. 3–8.

[234] Koller D, Friedman N, Bach F. Probabilistic graphical models: principles and techniques. MIT press; 2009.

[235] Diestel R. Graph Theory (Graduate Texts in Mathematics). 2000.

[236] Newman M. Networks: An Introduction. 2010. doi:10.1093/acprof:oso/9780199206650.001.0001.

[237] Ball MO. Computational complexity of network reliability analysis: An overview. IEEE Trans Reliab 1986;35:230–9.

[238] Cooper GF. The computational complexity of probabilistic inference using Bayesian belief networks. Artif Intell 1990;42:393–405.

[239] Robert C, Casella G. Monte Carlo statistical methods. Springer Science & Business Media; 2013.

[240] Gilks WR, Thomas A, Spiegelhalter DJ. A language and program for complex Bayesian modelling. J R Stat Soc Ser D (The Stat 1994;43:169–77.

[241] Buntine WL. Operations for learning with graphical models. J Artif Intell Res 1994;2:159–225.

[242] Druzdzel MJ. SMILE: Structural Modeling, Inference, and Learning Engine and GeNIe: a development environment for graphical decision-theoretic models. Aaai/Iaai, 1999, p. 902–3.