UNIVERSITY OF OKLAHOMA

GRADUATE COLLEGE

BAYESIAN KERNEL METHODS FOR THE RISK ANALYSIS AND RESILIENCE MODELING OF CRITICAL INFRASTRUCTURE SYSTEMS

A DISSERTATION

SUBMITTED TO THE GRADUATE FACULTY

in partial fulfillment of the requirements for the

Degree of

DOCTOR OF PHILOSOPHY

By

HIBA BAROUD
Norman, Oklahoma
2015

BAYESIAN KERNEL METHODS FOR THE RISK ANALYSIS AND RESILIENCE MODELING OF CRITICAL INFRASTRUCTURE SYSTEMS

A DISSERTATION APPROVED FOR THE
SCHOOL OF INDUSTRIAL AND SYSTEMS ENGINEERING

BY

_____
Dr. Kash Barker, Chair

_____
Dr. Pakize Simin Pulat

_____
Dr. Naiyu Wang

_____
Dr. Charles Nicholson

_____
Dr. Jose Ramirez-Marquez

To Wassim Tabet
And to my family:
My father Sami, my mother Nada, and my brother Charbel

Acknowledgements

First, I would like to give thanks to God the father, Jesus, the Holy Spirit, the Virgin Mary, my Guardian Angel, and all the Saints. I am grateful for the graces and blessings I receive every day that helped me accomplish the work done in this dissertation and throughout my studies.

I am grateful for my family, my father Sami, my mother Nada, and my brother Charbel. They have supported me, prayed for me, and encouraged me in all my pursuits.

My advisor, Dr. Kash Barker, has been a great educator and mentor. I would not have been the person I am today nor would I have accomplished this great milestone without his guidance. He encouraged me to go after every opportunity I had to gain experience and branch out. He truly went above and beyond to help me succeed.

I want to thank my committee members, Dr. Pakize Simin Pulat, Dr. Naiyu Wang, Dr. Charles Nicholson, and Dr. Jose Ramirez-Marquez. I want to also thank Dr. Theodore Trafalis. I greatly appreciate your help and feedback on my work.

To my colleagues at the University of Oklahoma and to my friends in Lebanon, Canada, and the United States, thank you for believing in me and encouraging me especially in challenging times.

I appreciate the help and support of the faculty and staff of the School of Industrial and Systems Engineering, especially the director, Dr. Randa Shehab, thank you for all the insightful conversations.

To the love of my life and my support system, Wassim Tabet, this achievement is not possible without your love, patience, and encouragement. I am grateful for your presence in my life. I love you.

Table of Contents

List of Tables

List of Figures

Abstract

The protection of critical infrastructures has recently garnered attention with an emphasis on analyzing the risk and improving the resilience of such systems. With the abundance of data, risk managers should be able to better inform preparedness and recovery decision making under uncertainty. It is important, however, to develop and utilize the necessary methodologies that bridge between data and decisions. The goal of this dissertation is to (i) predict the likelihood of risk, (ii) assess the consequences of a disruption, and (iii) inform preparedness and recovery decision making.

This research presents a data-driven analysis of the risk and resilience of critical infrastructure systems. First, a new Bayesian kernel model is developed to predict the frequency of failures and a Beta Bayesian kernel model is deployed to model resilience-based importance measures. Bayesian kernel models were developed for Gaussian distributions and later extended to other continuous probability distributions. This research develops a Poisson Bayesian kernel model to accommodate count data. Second, interdependency models are integrated with decision analysis and resilience quantification techniques to assess the multi-industry economic impact of critical infrastructure resilience and inform preparedness and recovery decision making under uncertainty.

Examples of critical infrastructure systems are inland waterways, which are critical elements in the nation's civil infrastructure and the world's supply chain. They allow for a cost-effective flow of approximately $150 billion worth of commodities annually across industries and geographic locations, which is why they are called "inland marine highways." Aging components (i.e., locks and dams) combined with

adverse weather conditions, affect the reliability and resilience of inland waterways. Frequent disruptions and lengthy recovery times threaten regional commodity flows, and more broadly, multiple industries that rely on those commodities. While policymakers understand the increasing need for inland waterway rehabilitation and preparedness investment, resources are limited and select projects are funded each year to improve only certain components of the network. As a result, a number of research questions arise.

What is the impact of infrastructure systems disruptions, and how to predict them? What metrics should be used to identify critical components and determine the system's resilience? What are the best risk management strategies in terms of preparedness investment and recovery prioritization?

A Poisson Bayesian kernel model is developed and deployed to predict the frequency of locks and dams closures. Economic dynamic interdependency models along with stochastic inoperability multiobjective decision trees and resilience metrics are used to assess the broader impact of a disruption resulting in the closure of a port or a link of the river and impacting multiple interdependent industries. Stochastic resilience-based measures are analyzed to determine the critical waterway components, more specifically locks and dams, that contribute to the overall waterway system resilience. A data-driven case study illustrates these methods to describe commodity flows along the various components of the U.S. Mississippi River Navigation System and employs them to motivate preparedness and recovery strategies.

Chapter 1

Introduction

A number of infrastructure systems, both in the U.S. and globally, have been identified as *critical* due to their ubiquitous influence on society's way of life. Among these critical infrastructures are energy, healthcare, and transportation sectors (DHS, 2009). Their criticality is due to their interconnectedness with other infrastructure systems, as well as industries and workforces which rely upon them. A disruption to such a critical infrastructure, whether the result of a terrorist attack, a natural disaster, an accident, or common failure, could incur widespread losses of functionality that affect not only the infrastructure itself but all the industries depending on it. Therefore, to effectively plan for the protection of these infrastructures from failure, and more importantly the response to and recovery from such failures when they inevitably occur, an important challenge exists in (i) estimating the likelihood of a disruptive event, (ii) portraying the widespread economic losses resulting from a disruptive event, as well as (iii) measuring the efficacy of risk management to determine the appropriate investment to enable preparedness and recovery.

Of interest in this work are transportation systems, specifically inland waterway navigation systems, though the methodologies developed in this dissertation have broader applicability to preparedness in other infrastructure systems. As truck traffic continues to increase, congestion on highway networks will worsen and become a major issue for commodity flows (USDOT, 2009, 2011; NCFRP, 2010; GAO, 2011). The highway and railway networks in many cities near coastal ports are already experiencing bottlenecks. A viable alternative to these modes for freight transport is the

inland waterway navigation system. Vital to commodity flows in the U.S., almost 80% of all U.S. international trade is transported through coastal ports, with 40% of these shipments moving inside the U.S. through inland ports before reaching their final destination (Haveman & Shatz, 2006). However, commodity flows through inland waterways are a distant third behind highway and rail traffic (USACE, 2010), giving them an opportunity for expanded use.

Infrastructure networks in today's global landscape can been characterized as exhibiting the characteristics of many complex and large-scale systems: (i) a large number of interacting components and subsystems, (ii) a large number of decision and state variables, (iii) complicated, complex, and often nonlinear functional relationships, (iv) uncertainty and variability, (v) hierarchical and/or networked interdependencies, (vi) multiple and often conflicting performance objectives, (vii) multiple decision makers, and (viii) dynamic changes, among others (Haimes, 2009b). As such, disruptive events, whether malevolent attacks, natural disasters, manmade accidents, or common cause failures, can have devastating, widespread, and often unpredictable, results.

Consider, for example, the August 2003 US blackout which "contributed to at least 11 deaths and cost an estimated $6 billion" (Minkel, 2008), or the largest blackout in history experienced by India during August 2012, affecting over 600 million people. Against the effects of these events, most research efforts have been devoted to developing traditional measures of protection (hardening) (Ramirez-Marquez, Rocco, & Levitin, 2011; Levitin & Hausken, 2010; Bier, Haphuriwat, Menoyo, Zimmerman, & Culpen, 2008) and policies that can be expensive, degrade typical performance, and are non-reactive. Recent attention has been placed on preparedness, response, and recovery

from these events (e.g., in the large-scale homeland security preparedness domain (DHS, 2009)). A perspective that has recently been collectively referred to in the systems engineering community as *designing for resilience*, is considered an essential component in the design of systems and enterprises (Comfort, Boin, & Demchak, 2010).

The goal of this dissertation is to develop and utilize data-driven methods and statistical tools as well as stochastic simulation to address the preparedness for, response to, and recovery from critical infrastructure systems disruptions. The main objectives of the research are to (i) predict the likelihood of a disruption, (ii) assess the consequences, and (iii) provide recovery insights.

General Literature Review

While a more specific literature review will be presented in each chapter in the dissertation, the following is a general overview of the risk analysis, resilience modeling, and inland waterways research literature. Significant methodological and theoretical contributions that founded this research will be presented in subsequent chapters.

Risk Analysis

According to Lowrance (1976), risk is a "measure of the probability and severity of adverse effects." The analysis of risk is composed of the (i) risk assessment and (ii) risk management. The risk assessment would generally consider three questions (Kaplan & Garrick, 1981):

(i) *What can go wrong?*

(ii) *What are the chances of something going wrong?*

(iii) *What are the consequences if the undesirable event occurs?*

Haimes (2009b) complements these questions with another set of three questions aiming at risk management:

(i) *What can be done?*

(ii) *What options are available and what are the associated tradeoffs in terms of cost, risks, and benefits?*

(iii) *What are the impacts of the current management decisions on future options?*

The work in this dissertation addresses the second and third question of the risk assessment as well as the three questions of the risk management. Statistical methods are used to estimate the frequency of disruptions occurring in a critical infrastructure system. Interdependency modeling is used to assess the consequences of disruptions, and integrated with decision analysis techniques, it is used to identify risk management strategies and assess their efficacy. Finally, statistical tools and stochastic simulation techniques are deployed to assess current decisions on the system's resilience in the future.

Probabilistic Risk Assessment (PRA) is a common tool used to analyze the risk as it has been applied in cases where there is little to no data available such as the case with the estimation of accidents in nuclear power plants (Lewis et al., 1979). The method relies on the functional form and the physical structure of the system coupled with other techniques (e.g., fault tree, event tree, among others) that are used to extract the analysis from a component-level to a system-level. Ezell, Farr, & Wiese (2000) deploy PRA techniques for the risk analysis of infrastructure systems.

Guikema (2009) highlights the need to consider statistical data-driven methods to address the risk assessment of large-scale critical infrastructure systems facing natural disasters. The structure of such systems is too complex for PRA methods and they have an abundance of data for which statistical tools would be more suitable such as the Generalized Linear Models (GLM). The use of statistical techniques to provide data-driven analysis of the risk of infrastructure systems has recently emerged with applications in the prediction of power outages (Liu, Davidson, Rosowsky, & Stedinger, 2005), traffic accidents (Lord, Washington, & Ivan, 2005), and water distribution systems reliability (Yamijala, Guikema, & Brumbelow, 2009), among others. The key factor of utilizing such tools is the selection of the best model based on the type of data available and the outcome variable.

The first part of this dissertation develops a new statistical tool to analyze the risk in critical infrastructures. Such systems are large and complex in nature that PRA would be computationally expensive to implement. However, the data available for some of these systems is not always abundant enough to provide robust estimation using traditional statistical methods such as GLM. The proposed model has the potential of addressing risk using data-driven analysis when data is scarce.

Resilience Modeling

The importance of having robust and resilient infrastructure systems has gained the attention of decision makers and government officials in the past decade. Critical infrastructure systems, such as power grids and transportation systems have been vulnerable to numerous disruptive events including natural disasters, willful attacks, and accidents. DHS announced a set of grant programs targeting different areas prone to

willful attacks or natural disasters (DHS, 2012), aiming to provide resources helpful in supporting the National Preparedness Goal (NPG) in succeeding in its mission of insuring "a secure and *resilient* Nation with the capabilities required across the whole community to prevent, protect against, mitigate, respond to, and *recover* from the threats and hazards that pose the greatest risk" (DHS, 2011). Further motivation comes from Boin, Comfort, & Demchak (2009):

> *"If we accept that dominant trends such as globalization, increasing interdependence and complexity, the spread of potentially dangerous technologies, new forms of terrorism, and climate change create new and unimaginable threats to modern societies, it is only a small step to recognizing and accepting the inherent shortcomings of contemporary approaches to prevention and preparation. If we cannot predict or foresee the urgent threats we face, prevention and preparation become difficult. The concept of resilience holds the promise of an answer*."

Resilience is often thought of as the ability exhibited by a system to "bounce back" following a disturbance. In material science, a modulus of resilience is also defined to represent the energy absorbed per unit volume of material when stressed to the proportional limit (Ugural, 2003). In socio-ecological systems, resilience is defined as the magnitude of disturbance that can be absorbed before the system changes its structure by changing the processes that control behavior (Holling, 1973; Carpenter, Walker, Anderies, & Abel, 2001). Regarding enterprise systems, Jackson (2007, 2009) defines resilience as the ability of organizational, hardware, and software systems to mitigate the severity and likelihood of failures or losses, to adapt to changing

conditions, and to respond appropriately. In business terms, resilience has been defined as the ability of an organization to sustain the impact of a business interruption and recover and resume operations to continue to provide minimum services (Hoffman, 2007). With respect to critical infrastructure, the Infrastructure Security Partnership (2011) noted that a resilient infrastructure sector would "prepare for, prevent, protect against, respond or mitigate any anticipated or unexpected significant threat or event" and "rapidly recover and reconstitute critical assets, operations, and services with minimum damage and disruption." In an engineering context, Hollnagel, Woods, and Leveson (2007) promote *resilience engineering* as a new paradigm for safety engineering. And many others have defined resilience in various ways (Fiksel, 2003; Wreathall, 2006; Vogus & Sutcliffe, 2007; Rose, 2009).

This research utilizes the resilience modeling framework developed by Henry and Ramirez-Marquez (2012) that quantifies resilience based on the system's service function. The framework is the basis for the modeling approach developed to analyze the resilience of critical infrastructure systems and its impact on decision making under uncertainty.

Inland Waterways Research

> "*The current system of inland waterways lacks resilience. Waterway usage is increasing, but facilities are aging and many are well past their design life of 50 years. Recovery from any event of significance would be negatively impacted by the age and deteriorating condition of the system, posing a direct threat to the American economy*." 2009 Report Card for America's Infrastructure, American Society for Civil Engineers (ASCE) (2009).

The multi-modal transportation system plays a vital role in maintaining commodity flows across multiple industries and multiple regions. As a result of their critical role, the effects of large-scale disruptive events could result in the closure of key transportation links and nodes. These critical components in a transportation network (e.g., inland waterways) are particularly susceptible to disruptions in commodity flows (Lee, Park, & Lee, 2003; Sacone & Siri, 2009; Lee & Kim, 2010). The recovery of transportation networks from disruptions has been given some recent attention (Cadarso, Marín, & Maróti, 2013; Chen & Miller-Hooks, 2012; Zeng, Durach, & Fang, 2012; Zhang & Peeta, 2011).

Although inland ports face many of the same risks as coastal ports, relatively few studies have developed risk assessments of inland ports (Folga et al. 2009; Pant, Barker, Grant, & Landers, 2011; MacKenzie, Barker, & Grant, 2012a). Some studies have focused on forecasting commodity flows in inland waterway networks (Babcock & Lu, 2002; Beuthe, Jourquin, Geerts, & a Ndjang'Ha, 2001), however, such models do not capture the effect of uncertain disruptive events and their impacts on the commodity flows. Pant et al. (2011) provide a simulation model of inland port activities to parameterize a port disruption within a multi-regional interdependency model, while MacKenzie et al. (2012a) focus on the multi-regional impacts in functionality and in economic losses of decision making for shipping alternatives following an inland port disruption.

During recent ASCE testimony to the U.S. Senate, it was stated that the costs attributed to delays in U.S. inland waterways were $33 billion in 2010 (rising to $49 billion by 2020), with interdependent impacts cascading to economic sectors that

require inland waterway transport (e.g., petroleum, coal) (ASCE, 2013a). As such, the study of the risk and resilience of inland waterway networks is an important area of focus.

The case study for inland waterways is analyzed for each of the methods developed. In each chapter, an aspect of the risk or resilience of the network is analyzed such as the prediction of the frequency of disruptions (Chapter 2), the assessment of interdependent economic impacts of a port closure and the corresponding risk management efficacy analysis (Chapter 3), the impact of the resilience of the waterway network on the interdependent economic impacts of a disruption (Chapter 4), the prioritization of recovery activities sets after a disruption occurred (Chapter 5), and finally, the identification of resilience-based critical components of the inland waterway (Chapter 5).

Research Contributions

Each chapter in the dissertation outlines the methodology and modeling approach for the risk and resilience analysis with an application to the inland waterway network. Thus, each chapter will address a specific literature review of the tools discussed.

There are three levels of contributions in this research: theory, methodology, and application. A contribution is made to the theory of statistics and machine learning through a new model, the Poisson Bayesian kernel model, which expands on the class of Bayesian kernel methods to accommodate count data (Chapter 2). The work has appeared in Floyd, Baroud, & Barker (2014). A methodological contribution is made to the field of risk analysis through the deployment of the Poisson Bayesian kernel model

in the risk analysis of critical infrastructure systems, in particular, inland waterways (Baroud, Barker, & Lurvey, 2013a). Other methodological contributions include the integration of the economic interdependency model (i) with stochastic decision trees to analyze the impact of port closures and assess risk management strategies (Chapter 3), the work has appeared in Baroud, Barker, & Grant (2014a), and (ii) with resilience metrics to quantify the impact of the resilience of the disrupted infrastructure system on the multi-industry economic impacts (Chapter 4), the work appears in Baroud, Barker, Ramirez-Marquez, & Rocco (2013b, 2014c). Finally, the third type of contribution is at the application level where Bayesian kernel methods are used to model the resilience of critical infrastructures. In particular, the Beta Bayesian kernel model is used to analyze resilience-based importance measures to identify critical components of the inland waterway network. Data-driven and statistical tools have not been previously deployed to analyze the resilience of critical infrastructure systems. This work has appeared in Baroud and Barker (2014).

Chapter 2

Bayesian Kernel Methods for the Prediction of Risk

In many situations, the likelihood of an event is found using the average rate at which the event occurs. And often that rate is a function of characteristics surrounding the event. To integrate the impacts of both the component characteristics and any prior failure information, a Bayesian kernel model is proposed as an approach to a more accurate estimation of the rate of occurrence of an event. More specifically, an extended version of this method is developed, the Poisson Bayesian kernel model to accommodate count data and estimate the rate of occurrence. This chapter includes an extensive literature review of Bayesian kernel methods and count data models. The Poisson Bayesian kernel model is introduced as a new approach to predict the likelihood of occurrence of an event. An empirical analysis of this model using different types of data sets and measures of goodness of fit and prediction accuracy is used to validate the model in comparison to classical approaches. Finally, the Poisson Bayesian kernel model is deployed in a case study to analyze the rate of disruptions along an inland waterway network, the Mississippi River Navigation System.

Literature Review

Kernel methods, first introduced in a pattern recognition setting several decades ago (Aizerman, Braverman, & Rozonoer, 1964), have found popularity across a number of data mining domains, including bioinformatics (Schölkopf, Guyon, & Weston, 2003; Ben-Hur & Noble, 2005), sensing (Arias, Randall, & Sapiro, 2007; Camps-Valls, Rojo-Alvarez, & Martinez-Ramon, 2006), and financial risk management and forecasting (Wang & Zhu, 2010; Mitschele, Chalup, Schlottmann, & Seese, 2006), among many

others. Kernel functions are used to map input data, for which no pattern can be recognized, to a higher dimensional space, where patterns are more readily detected. Such functions enable algorithms designed to detect relationships among data in the higher dimensional space, including least squares regression and support vector machines (SVM) classification (Cherkassky & Mulier, 1998; Cristianini & Shawe-Taylor, 2000; Hastie, Tibshirani, & Friedman, 2001). Integrating Bayesian methods with kernel methods has recently garnered attention (Seeger, 2000; Bishop & Tipping, 2003, Mallick, Ghosh, & Ghosh, 2005; Zhang, Dai, & Jordan, 2011), as Bayesian methods make use of historical data to estimate posterior probability distributions of the parameter of interest given that it follows a specific prior distribution.

The integration of Bayesian and kernel methods enables a classification algorithm which provides probabilistic outcomes as opposed to deterministic outcomes (i.e., such as those resulting from SVM classification). That is, rather than assigning a class to a data point, Bayesian kernel methods assign a probability that the data point belongs to a particular class. Several extensions to Bayesian kernel models have appeared, including (i) the relevance vector machine (RVM) which assumes a Gaussian distribution for the probability to be estimated (Tipping, 2001; Schölkopf & Smola, 2002), and (ii) non-Gaussian distributions for binary problems (Montesano & Lopes, 2009; Mason & Lopes, 2011; MacKenzie, Trafalis, & Barker, 2014b). However, there has been no Bayesian kernel model developed for count data.

Bayesian Methods

The classic Bayes rule assumes that a prior probability for an event of interest, A, is given as $P(A)$, and a likelihood of event B conditioned on the occurrence of A is

given as $P(B|A)$. With these probabilities, along with $P(B)$, one can calculate the posterior distribution for the event of interest given knowledge of B, or $P(A|B)$, shown in Eq. (2-1) (Bayes & Price, 1763).

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \tag{2-1}$$

This manifests itself, for example, when one wants to develop a posterior distribution for a parameter of interest, t, from (i) the prior distribution for that parameter, $P(t)$, and (ii) the data describing that parameter in the form of a likelihood function, $P(x|t)$, which is a conditional likelihood of obtaining the data given the parameter. In such a case, the denominator does not depend on the parameter of interest and can be excluded from the Bayes rule equation when maximum likelihood calculations are performed. More specifically, the posterior probability distribution for this parameter t can be estimated as being proportional to its prior distribution multiplied by the likelihood function, as depicted in Eq. (2-2).

$$P(t|x) \propto P(t)P(x|t) \tag{2-2}$$

An important concept used in the Bayesian framework is the notion of conjugate priors, which assumes that the posterior, $P(t \mid x)$, and the prior, $P(t)$, distributions are from the same family of distributions. Having the prior and posterior follow the same distribution insures that the overall data properties are kept while modifying the details of the distribution such as the parameters to better explain the trends. In addition, conjugate priors provide analytical solutions for the posterior distribution, allowing for a much faster computation time than other Bayesian tools that require simulation or optimization techniques. MacKenzie et al. (2014b) use the Beta-Bernoulli conjugate prior to build non-Gaussian Bayesian kernel models for binary classification. The

research in this dissertation uses the Gamma conjugate prior to construct a Poisson Bayesian kernel method to model count data.

Kernel Methods

Algorithms for detecting non-linear relationships have started to emerge with heuristic tools such as decision trees (Breiman, Friedman, Olshen, & Stone, 1984; Quinlan, 1983, 1986) and artificial neural networks (ANN) (Yegnanarayana, 2009) with applications across many disciplines such as healthcare (Baxt, 1995; Fonarow et al., 2005), geotechnical engineering (Shahin, Jaksa, & Maier, 2001), and atmospheric sciences (Gardner & Dorling, 1998), among many others. These techniques lack theoretical foundation and robustness, and they have a tendency to overfit the data. In addition, ANN can be computationally expensive (Broussard, Kennell, Ives, & Rakvic, 2008).

Kernel methods introduced a new class of machine learning for non-linear classification that was more flexible, capable of accommodating different types of data and detecting different types of relations (Vapnik, 2013; Vapnik & Vapnik, 1998). As mentioned earlier, a kernel function maps the data to a higher dimensional space, called the feature space, and detects linear classifiers in that space. Such technique provides simple and easy detection to nonlinear and complex relations among the data. For example, Figure 2-1 is a graphical representation of a simple binary classification problem. The kernel function is used to map the input data in R^2 (plot on the left hand side) into the feature space R^3 (plot on the right hand side) in which the inner product is computed then integrated into the classification algorithm to find a linear classifier in the feature space that would be equivalent to an ellipse in the input space.

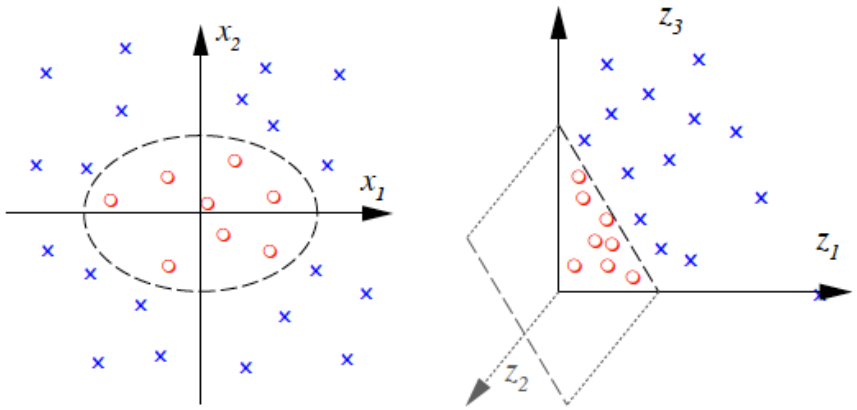


Figure 2-1: Binary classification example using a polynomial kernel function (Schölkopf & Smola, 2002)

The kernel function used to map and compute the inner product in the figure above, is the polynomial kernel function of degree $d = 2$, Eq. (2-3), where ϕ is the mapping function that maps a two-dimensional data point, (x_1, x_2), to a three-dimensional point in the feature space, $\phi(x) = \left(x_1^2, x_2^2, \sqrt{2x_1x_2}\right)$, then the kernel function computes the dot product of the mapped data points.

$$\begin{aligned} K(x,y) &= <\phi(x), \phi(y)> \\ &= x_1^2y_1^2 + 2x_1x_2y_1y_2 + x_2^2y_2^2 \\ &= (x_1y_1 + x_2y_2)^2 \\ &= (<x,y>)^2 \end{aligned} \tag{2-3}$$

Note in the equation above, the kernel is able to implicitly perform both the mapping and the dot product computation in the feature space in a single function. This is one of the most attractive properties of kernel methods, called the kernel trick, which saves computation time and complexity. The kernel matrix, **K,** resulting from such a

computation is symmetric and positive definite where each entry is a similarity measure corresponding to the kernel between two data points.

$$\mathbf{K} = \begin{bmatrix} K(x_1, x_1) & \cdots & K(x_1, x_m) \\ \vdots & \ddots & \vdots \\ K(x_m, x_1) & \cdots & K(x_m, x_m) \end{bmatrix} \tag{2-4}$$

Algorithms that use such technique include but are not limited to Parzen's Windows (Schölkopf & Smola, 2002), Support Vector Machines (SVM) (Cristianini & Shawe-Taylor, 2000; Shawe-Taylor & Cristianini, 2004), Ridge Regression (Saunders, Gammerman, & Vovk, 1998), Fisher Linear Discriminant Analysis (LDA) (Schölkopft & Mullert, 1999), and Principal Component Analysis (PCA) (Schölkopf, Smola, & Müller, 1997).

While kernel methods have a wide range of uses in different applications, this research employs kernel functions as a similarity measure to improve the prediction accuracy of Bayesian methods for count data and introduce probabilistic predictions to traditional classification tools. In particular, the kernel matrix produced by the implicit mapping and the computation of the inner product is used to integrate information from the covariates in the input data into the Bayesian computation of the parameters of the posterior distribution.

Bayesian Kernel Models

Bayesian kernel methods have recently been introduced to the machine learning literature providing probabilistic solutions as opposed to deterministic solutions. Most Bayesian kernel methods are developed with Gaussian prior distributions (Figueiredo, 2001; Tipping, 2001; Zhang et al., 2011). For an $m \times d$ data matrix $\mathbf{X}$ with rows corresponding to m data points each with d attributes, the function $\mathbf{t}(\mathbf{X})$ is considered to

be a random vector of length m mapping input data $\mathbf{X}$ into a class. Gaussian Bayesian kernel models assume the vector-valued function $\mathbf{t}$ follows a multivariate normal distribution with mean $E[\mathbf{t}(\mathbf{X})] = \mathbf{0}$ and covariance matrix $\mathrm{Cov}(\mathbf{t}(\mathbf{X})) = \mathbf{K}$, where matrix $\mathbf{K}$ is positive definite and matrix element K_{ij} is the kernel function $k(x_i, x_j)$, between the i^{th} and j^{th} data points. The multivariate normal distribution for the realization of $\mathbf{t}$ is found in Eq. (2-5).

$$P(\mathbf{t}) = \frac{1}{\sqrt{(2\pi)^m}}(det\,\mathbf{K})^{-1/2} exp\left(-\frac{1}{2}\mathbf{t}^T\mathbf{K}^{-1}\mathbf{t}\right) \tag{2-5}$$

As computing the inverse of the kernel matrix in the probability density function of $\mathbf{t}$ can be cumbersome, Schölkopf and Smola (2002) introduce a new vector-valued variable of length m, $\boldsymbol{\omega}$, such that $t(\mathbf{x}_i) = \mathbf{k}(\mathbf{x}_i, \mathbf{X})\boldsymbol{\omega}$, whose prior is also a multivariate normal distribution, Eq. (2-6).

$$P(\boldsymbol{\omega}) = \frac{1}{\sqrt{(2\pi)^m}}(det\,\mathbf{K})^{-1/2} exp\left(-\frac{1}{2}\boldsymbol{\omega}^T\mathbf{K}\boldsymbol{\omega}\right) \tag{2-6}$$

Since the first term in the probability density function, $\frac{1}{\sqrt{(2\pi)^m}}(\det\mathbf{K})^{-1/2}$, does not depend on the parameter $\boldsymbol{\omega}$, the prior distribution can be reduced to Eq. (2-7).

$$P(\boldsymbol{\omega}) = exp\left(-\frac{1}{2}\boldsymbol{\omega}^T\mathbf{K}\boldsymbol{\omega}\right) \tag{2-7}$$

In the case of a binary classification, an appropriate likelihood function would be the logit function shown in Eqs. (2-8) and (2-9).

$$P\big(y = 1\big|t(\mathbf{x})\big) = \frac{1}{1 + exp\big(-t(\mathbf{x})\big)} \tag{2-8}$$

$$P\big(y=-1\big|t(\mathbf{x})\big)=\frac{1}{1+exp\big(t(\mathbf{x})\big)} \tag{2-9}$$

The posterior distribution is then the product of the likelihood function and the prior distribution for a data set of m data points, found in Eq. (2-10) .

$$\begin{aligned} P(\boldsymbol{\omega}|\boldsymbol{y})\ \alpha & \prod_{i=1}^{m}\left(\frac{1}{1+exp(-\mathbf{k}(\mathbf{x}_i,\mathbf{X})\boldsymbol{\omega})}\right)^{0.5+0.5y_i} \\ & \times\prod_{i=1}^{m}\left(\frac{1}{1+exp(\mathbf{k}(\mathbf{x}_i,\mathbf{X})\boldsymbol{\omega})}\right)^{0.5-0.5y_i} \\ & \times exp\left(-\frac{1}{2}\boldsymbol{\omega}^T\mathbf{K}\boldsymbol{\omega}\right) \end{aligned} \tag{2-10}$$

To estimate the parameter of interest, $\boldsymbol{\omega}$, Eq. (2-10) is maximized (or its negative log is minimized) using any of several optimization algorithms (e.g., the Newton-Raphson method).

An important extension to the basic Bayesian kernel model is the non-Gaussian Bayesian kernel model (Montesano & Lopes, 2009; Mason & Lopes, 2011; MacKenzie et al., 2014b), which can improve the predictive accuracy for certain problems where a Gaussian distribution for model parameters should not realistically be assumed. MacKenzie et al. (2014b) highlight some of the drawbacks of using the Gaussian distribution for binary classification problems, use a Beta conjugate prior, and offer an alternative likelihood function to the logit. The research expands on previous work done on Non-Gaussian kernel models (Montesano & Lopes, 2009; Mason & Lopes, 2011) by introducing a more generalized model based on the Beta conjugate prior which provides much faster computation time than traditional Bayesian kernel methods that rely on

either optimization or simulation to find the solutions. Beta Bayesian kernel methods will be revisited in details in Chapter 5 to model the resilience of infrastructure systems.

Bayesian Kernel Model for Count Data

This dissertation proposes a new Bayesian kernel method that can accommodate count data. The purpose of such model is to integrate prior information in the form of experts' knowledge with historical data and attribute information to produce a probability distribution of the rate of occurrence of a particular event. Using this distribution, the frequency of events such as disruptions can be predicted. This section first reviews count data modeling approaches in the literature then discusses the structure of the Poisson Bayesian kernel model. The model is empirically tested using sample data and deployed in a case study of a critical infrastructure system disruptions prediction.

Count Data Modeling

One of the classical approaches used to analyze count data are Generalized Linear Models (GLM) (Agresti, 2002; Cameron & Trivedi, 2013; Nedler & Wedderburn, 1972). The Poisson GLM is most commonly used to model count data. The method assumes that the rate to be estimated has an exponential relationship with a set of covariates representing coefficients for the different attributes, shown in Eq. (2-11).

$$\hat{\lambda} = e^{\beta_i X} \tag{2-11}$$

Under the Poisson GLM, the response follows a Poisson distribution, Eq. (2-12), and the log function is the link function that relates the set of covariates and coefficients to the response variable.

$$P(y) = \frac{\lambda^y e^{-\lambda}}{y!} \tag{2-12}$$

Another type of GLM for modeling count data is the Negative Binomial GLM which relaxes the constraints of homoscedasticity imposed by the Poisson GLM (Cox, 1983; Lawless, 1987). The Negative Binomial GLM assumes that the marginal distribution of the response follows a Negative Binomial distribution, Eq. (2-13), where k is the overdispersion parameter and λ is assumed to follow a Gamma distribution.

$$P(y) = \frac{\Gamma\left(y + \frac{1}{k}\right)}{\Gamma(y+1)\Gamma\left(\frac{1}{k}\right)} \left(\frac{k\lambda}{1+k\lambda}\right)^y \left(\frac{1}{1+k\lambda}\right)^{\frac{1}{k}} \tag{2-13}$$

The Negative Binomial GLM also assumes a log function for the link function and as a result, the response variable has an exponential relationship with the covariates. While the Negative Binomial GLM is a formal way of handling overdispersion in count data, other approaches developed extensions to the classical Poisson regression (Breslow, 1984; Paul & Placket, 1978; Johnson, Kotz, & Kemp, 2005; Yip, 1991). A well-known approach is to add a dispersion parameter such as the case with the Quasi-Poisson regression. Many other count data regression models have been developed to introduce further flexibility and complexity into the modeling approach.

One example of added complexity to the Poisson GLM is the Zero-Inflated Poisson (ZIP) model which assumes a form of mixture modeling to account for a specific type of data expressing a large number of occurrences that are equal to zero. Such an approach has shown improved goodness of fit compared to Poisson and Negative Binomial GLM, but it is a much more complicated approach, it is difficult to interpret, and it is only applicable to particular cases where the majority of the response

variable observations are equal to zero. Examples of such applications are insurance claims (Mouatassim & Ezzahid, 2012), dental caries prevention methods (Hall & Shen, 2010), and defects in manufacturing (Lambert, 1992), among others. Also, since the likelihood is constructed for a mixture of models, there is no analytical solution for the estimation of parameters, instead, algorithms such as Newton-Raphson or EM (Fong & Yip, 1993) are used to maximize the log-likelihood function and compute estimates of the coefficients.

An example of added flexibility to count data modeling is the Conway-Maxwell Poisson GLM (Conway & Maxwell, 1962). The model can address both overdispersion and underdispersion in one framework and has been applied in the risk and reliability context using a Bayesian framework to account for the uncertainty in the regression parameters and improve on their accurate estimation (Guikema & Goffelt, 2008). Since such Bayesian techniques employ simulation, the higher accuracy comes at the cost of a longer computation time.

Other approaches of analyzing count data using a Bayesian framework are conjugate priors. These methods are quite attractive as they offer the benefit of uncertainty modeling using Bayesian techniques without adding any computational cost. Given a specific prior distribution and a specific likelihood function, the posterior distribution will have the same form as the prior distribution but with updated posterior parameters. Different forms of conjugate priors are discussed in this dissertation, one of which is the Gamma conjugate prior used to model count data. The method assumes that the rate of occurrence follows a Gamma prior and updates the distribution using information represented by a Poisson likelihood. The Gamma conjugate prior is the

foundation of the Poisson Bayesian kernel model proposed in this research and will be further discussed in the following section. Sophisticated extensions to this conjugate prior include the analysis of the parameters of the gamma prior distribution (Winkelmann, 2008). Other extensions to Bayesian Poisson methods consider hierarchical models (Tunaru, 2002). The model is based on the multivariate Poisson-log normal distribution with a hierarchical Bayesian application. This multivariate distribution is used to model discrete multiple count data and is shown in Eq. (2-14).

$$\begin{aligned} Y_{ki}|\lambda_{ki} &\sim^{\text{ind}} \text{Pois}(\lambda_{ki}) \\ (\log(\lambda_{ki}))_{i=1,\dots,4}|\mu,T &\sim^{\text{iid}} N_4(\mu,T) \\ \mu_i &\sim^{\text{iid}} N(0,0.0001) \\ T &\sim \text{Wishart}(R,4) \end{aligned} \tag{2-14}$$

$N_M(\mu,T)$ is the M-dimensional multivariate normal distribution. The mean vector is represented by μ, and T is the inverse of the covariance matrix. The hyperprior parameters R and $\pi = M$ are known. The model is advantageous in that it can model joint responses and can detect relationships among the categories of count variables. However, Markov Chain Monte Carlo methods are utilized to make inferences about the model parameters, for which computation can oftentimes be complex and lengthy.

The Poisson Bayesian kernel model developed in this research is simple enough to avoid expensive computations but detailed enough to overcome issues in basic Bayesian modeling approaches, such as the Gamma conjugate prior, and in count data regression models, such as the GLM.

Poisson Bayesian Kernel Model

Poisson Bayesian kernel methods estimate the rate of occurrence of the event rather than estimating a deterministic value for the number of times the event is estimated to occur. A common distribution to model count data within a Bayesian framework is the Gamma-Poisson conjugate prior. The development of the Poisson Bayesian kernel method discussed can be found in Baroud et al. (2013a) and Floyd et al. (2014). The approach uses the Gamma conjugate prior as the basis of the model.

It is assumed that the parameter to be estimated is the rate of occurrence, $\lambda > 0$, which follows a Gamma prior distribution with parameters $\alpha > 0$ and $\beta > 0$, as shown in Eq. (2-15).

$$P(\lambda) = \frac{\beta^{\alpha}}{\Gamma(\alpha)} \lambda^{\alpha-1} \mathrm{e}^{(-\beta\lambda)} \tag{2-15}$$

For the likelihood function, the product of the Poisson density function, shown in Eq. (2-16), is used, since this is a Gamma-Poisson conjugate prior approach.

$$L = \prod_{i=1}^{m} P(y_i) = \prod_{i=1}^{m} \frac{(\lambda_i^{y_i} e^{-\lambda_i})}{y_i!} = \frac{\lambda_i^{\sum_{i=1}^{m} y_i} e^{-m\lambda_i}}{\prod_{i=1}^{m} y_i!} \tag{2-16}$$

Thus, the posterior distribution is the product of Eqs. (2-15) and (2-16). Rearranging the product of the likelihood function and the prior distribution function results in a Gamma posterior distribution where $\alpha^* = \sum_{i=1}^{m} x_i + \alpha$ and $\beta^* = m + \beta$.

$$\begin{aligned} P(\lambda|x) &= \left(\frac{\beta^{\alpha}}{\Gamma(\alpha)} \lambda^{\alpha-1} e^{-\beta\lambda}\right)\left(\lambda^{\sum_{i=1}^{m} y_i} e^{-m\lambda}\right) \\ &= \frac{\lambda^{\left(\sum_{i=1}^{m} y_i + \alpha - 1\right)}\, e^{-\lambda(m+\beta)} (n+\beta)^{\sum_{i=1}^{m} y_i + \alpha}}{\Gamma\left(\sum_{i=1}^{m} y_i + \alpha\right)} \\ &= \text{Gamma}\left(\alpha^*, \beta^*\right) \end{aligned} \tag{2-17}$$

This result is the basic Gamma conjugate prior approach used in Bayesian analysis. This approach assumes the notion of exchangeability meaning that for different sets of training and testing data, the resulting posterior parameter will be similar since they are a function of the prior parameter, the size of the dataset, and the summation of all the data points. The characteristics of each outcome are not taken into consideration in this case, but rather the overall property of the dataset (Mackenzie et al., 2014b).

The Poisson Bayesian kernel approach extends the notion of the conjugate prior such that the posterior parameters computation not only depends on the prior parameters and the historical data but also on the attributes through the kernel matrix. The parameters for the Bayesian kernel model for counts are expressed in Eqs. (2-18) and (2-19). **K** is the $m \times m$ kernel matrix, **Y** is an $m \times 1$ vector containing the output data associated with the m observations of **X**, and **V** is an $m \times 1$ vector containing ones. Each entry in the kernel matrix represents the similarity measure between the attributes of the testing set and the training set. As such, the new data point is compared with the training set and according to the similarities of the attributes, new values for the parameter of the posterior distribution are computed. Note that in this case, the training and testing sets are assumed to have the same size, m. However, when the model is deployed, the sets can be of different sizes, and in some cases, the testing set could include only one data point such as in a leave-one-out analysis that will be illustrated in the case study.

$$\alpha^* = \mathbf{KY} + \alpha \tag{2-18}$$

$$\beta^* = \mathbf{KV} + \beta \tag{2-19}$$

As with other statistical and mathematical models, there are a few assumptions underlying the deployment of such modeling approach. Even though the form of the prior distribution is known from the conjugate prior, the model user would still need to identify the values of the prior parameters. While there are formal ways to determine the prior parameters (Kass & Wasserman, 1996), the selection of such parameters might not always be considered (Montesano & Lopes, 2009; Mason & Lopes, 2011). Oftentimes, the priors are either assumed to be known or are assigned such that the prior distribution is non informative. In other cases, these parameters are estimated using data and prior knowledge by matching the sample mean and variance to those of the prior distribution (MacKenzie et. al, 2014b; Carlin & Louis, 2008). Further discussion on the choice and impact of prior parameters is provided in the case study of this chapter. Another assumption to consider is the choice of the kernel function which depends on the application and the model user. This research uses the most popular kernel function, the radial basis function (RBF) in Eq. (2-20), where $k(\mathbf{x}_i, \mathbf{x}_j)$ is one entry in the matrix **K** representing the kernel function between the attributes of the i^{th} and j^{th} data points.

$$k(\mathbf{x}_i, \mathbf{x}_j) = \exp\left(-\frac{\|\mathbf{x}_i - \mathbf{x}_j\|^2}{2\sigma^2}\right) \tag{2-20}$$

In addition to being commonly used in kernel methods, RBF has nice properties. The function has only one parameter, σ, to be tuned to an optimal value. This reduces computation efforts significantly in comparison to other kernel functions with two or more parameters requiring a grid search to estimate them. Also, the structure of the function is based on the Euclidean distance, whereby similar data points are closer to

each other in the feature space. Finally, the kernel matrix of the RBF has full rank and the entries fall between zero and one resulting in kernel functions of the data points acting as weights in the computation of the posterior parameters (Schölkopf & Smola, 2002). More discussion on the tuning process of the RBF parameter, σ, will follow in the empirical analysis of this chapter.

The rate for the new data point follows then a Gamma distribution with parameters α^* and β^*. As a point estimate for this parameter, the expected value of the posterior distribution is considered, shown in Eq. (2-21) as the ratio of the Gamma distribution parameters α^* and β^*.

$$\hat{\lambda} = \frac{\alpha^*}{\beta^*} \tag{2-21}$$

Note that a different point estimate for the rate can be used such as the median, the mode, or the variance, depending on the type of problem and the model users.

Goodness of Fit Measures

In order to assess the performance of the model, goodness of fit measures are analyzed to identify the capability of the model to capture data patterns. The empirical analysis and the case study compare the Poisson Bayesian kernel (PBK) model to other classical methods for modeling count data, the Poisson generalized linear model (GLM) and the Negative Binomial GLM (Cameron & Trivedi, 1986, 2013). The Poisson and Negative Binomial GLM, presented above, assume that the rate to be estimated has an exponential relationship with a set of covariates representing coefficients for the different attributes, $\hat{\lambda}_{PGLM} = e^{\beta_i X}$, while the predicted rate for the PBK is equal to the expected value of the posterior probability distribution, $\hat{\lambda}_{PBK} = \frac{\mathbf{KY}+\alpha}{\mathbf{KV}+\beta}$.

The functional values of two metrics are used to compare how well the models fit the data and are able to explain the variance. The first metric is the *deviance*, which computes the difference in the log-likelihood function between the fitted model and the saturated model, Eq. (2-22), where y_i is the true value of the data point and $\hat{\lambda}$ is the estimated rate for the particular data point.

$$D = 2 \times \left(l(\boldsymbol{y}|\boldsymbol{y}) - l\left(\hat{\boldsymbol{\lambda}}|\boldsymbol{y}\right)\right) \tag{2-22}$$

The deviance is the generalized form of the sum of squared errors used in the linear regression model, it is a metric that analyzes the discrepancy between the observed and estimated values and it is the most commonly used goodness of fit measure by GLM users (McCullagh & Nelder, 1989). The deviance for a Poisson regression model is represented in Eq. (2-23), where $y_i \log\left(y_i/\hat{\lambda}_i\right) = 0$ when $y_i = 0$. D_P is used to assess how well the fitted values are representing the observed rate of occurrences in the Poisson Bayesian kernel model in comparison to the Poisson and Negative Binomial GLM.

$$D_P = 2 \times \sum_{i=1}^{m} y_i \log\left(\frac{y_i}{\hat{\lambda}_i}\right) - \left(y_i - \hat{\lambda}_i\right) \tag{2-23}$$

The second metric used is the functional value of the *log-likelihood*, shown in Eq. (2-24), which is to be maximized. The log-likelihood function represents the joint probability of the observed data as a function of the parameter of interest which is $\hat{\lambda}$ in this case. The larger the value of this function, the better the model is able to capture the data patterns using the estimated parameters.

$$l\left(\hat{\boldsymbol{\lambda}}|\boldsymbol{y}\right) = \sum_{i=1}^{m} [y_i \ln(\hat{\lambda}_i) - \hat{\lambda}_i - \ln(y_i!)] \tag{2-24}$$

Prediction Accuracy

The ultimate objective of building the Poisson Bayesian kernel model is to deploy it in risk analysis problems, such as predicting the frequency of disruptions in a particular network system. While the goodness of fit is important to assess whether the model is capturing the pattern and variability in the data, is it equally important to analyze the prediction power of a statistical model if it is going to be used for forecasting purposes. Prediction accuracy is assessed by the out-of-sample error, which accounts for the discrepancy between the estimated parameter and the actual observation of data points that were not in the set used to train the model. Figure 2-2 is a representation of the error in the training sample (in-sample error) and the test sample (out-of-sample error) as a function of the model complexity.

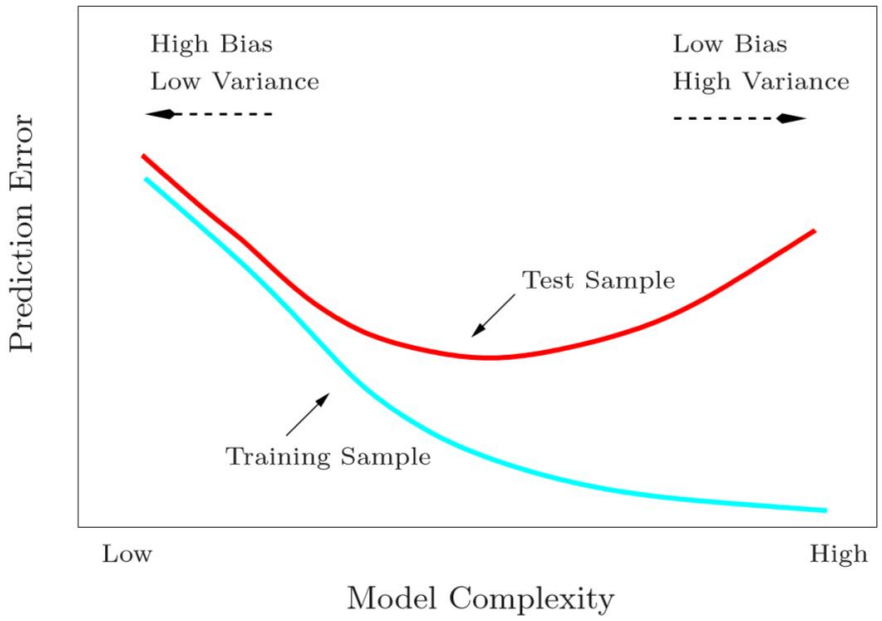


Figure 2-2: Test and training error as a function of the model complexity (Hastie et al., 2001)

Adding complexity to the model will decrease the training error but may cause overfitting at some point resulting in a poor prediction accuracy when the model is applied to an independent data set. Depending on the ultimate application of the method, a model can be selected based on either the test or the training error. The model developed in this dissertation is aimed at ultimately predicting the frequency of disruptions to inform preparedness and recovery decision making strategies and investment. An accurate prediction represented by a small test error is desired.

In order to validate the prediction power of the models, several metrics are evaluated to assess the out-of-sample error, and they are summarized in Table 2-1.

Table 2-1: Prediction error measurement metrics

Prediction accuracy metrics	**Formula**
Root Mean Square Error	$RMSE = \frac{1}{n}\sqrt{\sum_{i=1}^{n}\left(Y_i - \hat{\lambda}_i\right)^2}$
Normalized Root Mean Square Error	$NRMSED = \frac{\frac{1}{n}\sqrt{\sum_{i=1}^{n}\left(Y_i - \hat{\lambda}_i\right)^2}}{sd(Y_i)}$
	$NRMSEM = \frac{\frac{1}{n}\sqrt{\sum_{i=1}^{n}\left(Y_i - \hat{\lambda}_i\right)^2}}{Y_{maximum} - Y_{minimum}}$
Mean Absolute Error	$MAE = \frac{1}{n}\sum_{i=1}^{n}\left\|Y_i - \hat{\lambda}_i\right\|$

While RMSE and MAE are the most commonly used measurements of error, the normalized RMSE (NRMSE) is also considered to account for the variability across the different data sets evaluated in the empirical study. NRMSE can either be normalized based on the standard deviation of the observed values, $sd(Y_i)$, or the range of values in the testing set, $Y_{maximum} - Y_{minimum}$, and both cases are considered as the different

data sets exhibit different variability patterns in the observed values of the outcome variable.

Empirical Analysis

The PBK is applied to several data sets (Agresti & Finlay, 2008; Kutner, Nachtsheim, Neter, & Li, 2005; Lovett & Flowerdew, 1989; Roberts & Foppa, 2006; Sepkoski & Rex, 1974), and its performance is compared to the Poisson and Negative Binomial GLM using the goodness of fit and prediction accuracy metrics discussed in the previous section. A brief description of the data sets is found in Table 2-2. Most of the data sets are similar in terms of the number of predictors and the size of the data. One of the sets has a larger number of predictors for a small data set, and another is a large data set with a small number of predictors. Note that the number of predictors in all the models is held constant across the data sets to ensure consistency in the comparison. Also, the parameters of the prior distribution for all data sets are assumed to be $\alpha = \beta = 1$, also in order to maintain consistency in the evaluation of the performance of all models.

Table 2-2. Description of data sets in the Poisson Bayesian kernel model validation study

Data set	Number of attributes	Data set size	Dependent variable	Predictors
Crime	4	50	Crime rate	Race, percentage of high school graduates, percentage below poverty level, percentage with a single parent
Murder	4	51	Murder rate	Race, percentage of high school graduates, percentage below poverty level, percentage with a single parent
Mussels	8	45	Number of species of mussels	Area, number of stepping stones (intermediate rivers) to 4 major species-source river systems, concentration of nitrate, solid residue, concentration of hydronium
Customer	5	110	Number of customers visiting a store from a particular region	Number of housing units in the region, average household income in the region, average housing unit age in the region, distance to the nearest competitor, distance to the store
West Nile virus in birds	4	46	Numbers of cases West Nile virus in birds	Numbers of farms, area, population, human density
West Nile virus in equines	4	46	Numbers of cases West Nile virus in equines	Numbers of farms, area, population, human density
Apprentices Migration to Edinburgh	4	33	number of apprentices migrating	Distance, population, degree of urbanization, direction from Edinburgh

A holdout analysis is performed where each data set is randomly split into training and testing sets for 100 trials. Traditional holdout analyses would train the model on a portion of the data and deploy it on the testing set to make predictions and compare them to the actual observations. With the PBK, an intermediate step in the training process is added to tune the unknown parameter, σ, in the kernel function. This parameter is optimized based on the minimum mean square error. As a result, 30% of the data was used for testing the model, with 50% of the data used as a training set and 20% as a tuning set. The training and the tuning sets were then combined into one training set to perform the testing. For each of the three models, the estimated rate of occurrence is computed for the testing test and used to evaluate the deviance, the log-likelihood functional value, and the four out-of-sample error measurements given the observed values. This process is repeated 100 times where, at each iteration, random samples of training, tuning, and testing sets are chosen. Table 2-3 contains a summary of the analysis. The performance metrics values presented in the table below are the average values of the performance measures evaluated over 100 trials. PBK refers to the Poisson Bayesian kernel model and PGLM refers to the Poisson GLM, and NBGLM refers to the Negative Binomial GLM. Recall that the model with a smaller deviance and errors and a larger log-likelihood functional value is a better model.

Table 2-3: Performance metrics results for the empirical analysis

Data	Metrics	PBK	PGLM	NBGLM
Crime	LL	-276.25	-256.56	**-155.61**
	DEV	352.87	**313.49**	343.14
	RMSE	**26.47**	33.15	37.69
	NRMSEM	**0.28**	0.35	0.39
	NRMSED	**0.89**	1.13	1.29
	MAE	**21.26**	21.97	23.19
Murder	LL	-120.96	**-77.79**	**-77.79**
	DEV	107.69	**21.38**	**21.38**
	RMSE	9.81	**3.85**	**3.86**
	NRMSEM	0.28	**0.17**	**0.17**
	NRMSED	0.98	**0.58**	**0.58**
	MAE	4.68	**2.59**	**2.59**
Mussels	LL	-97.33	**-78.91**	**-78.72**
	DEV	66.55	29.71	**26.43**
	RMSE	**5.60**	5.84	5.83
	NRMSEM	**0.27**	0.31	0.31
	NRMSED	**0.96**	1.08	1.07
	MAE	**4.00**	4.32	4.31
Customer	LL	-230.02	**-194.46**	**-194.46**
	DEV	149.15	**78.04**	**77.69**
	RMSE	5.13	**3.58**	**3.58**
	NRMSEM	0.18	**0.13**	**0.13**
	NRMSED	0.77	**0.55**	**0.55**
	MAE	3.78	**2.75**	**2.75**
West Nile virus in birds	LL	-135.35	-103.94	**-80.76**
	DEV	181.74	118.93	**36.14**
	RMSE	**7.78**	8.44	9.14
	NRMSEM	**28.85**	33.47	36.44
	NRMSED	**98.22**	113.65	123.84
	MAE	**4.91**	5.09	5.23
West Nile virus in equines	LL	-40.12	-40.47	**-39.42**
	DEV	43.42	44.12	**32.57**
	RMSE	**1.75**	2.08	2.05
	NRMSEM	**0.30**	0.41	0.41
	NRMSED	**0.95**	1.24	1.25
	MAE	**1.17**	1.33	1.29
Apprentices Migration to Edinburgh	LL	-127.98	-106.03	**-64.46**
	DEV	442.9	146.62	**25.78**
	RMSE	**31.23**	32.29	32.45
	NRMSEM	**0.43**	0.53	0.51
	NRMSED	**1.32**	1.61	1.55
	MAE	15.98	**14.71**	15.86

Overall, there are five out of seven data sets for which the Poisson Bayesian kernel model outperforms the Poisson and Negative Binomial GLM in terms of the predictive accuracy. In particular, those five cases are all among the six small data sets. The RMSE, NRMSEM, NRMSED, and MAE all behave similarly for all the datasets and lead to the same conclusion of the model performance, except for a minor difference in the *Apprentices Migration to Edinburgh* where the PBK performs similarly to the NBGLM and slightly worse than the Poisson GLM in terms of MAE values. With respect to goodness of fit measures, the GLMs perform better than the PBK. Overall, the Negative Binomial fits the best. PGLM and NBGLM perform similarly in the two data sets for which the GLM outperforms the PBK in the predictive accuracy, *Customer* and *Murder*. The Poisson Bayesian kernel model appears to be a good model for prediction purposes when the data set is small with a small number of predictors, a situation known to cause issues with regression modeling (Cameron & Trivedi, 1986, 2013).

Case Study: Prediction of Inland Waterways Disruptions

With over 200 lock chambers and more than $150 million worth of goods flowing yearly (US Army Corps of Engineers, 2011), the inland waterway system plays an important role in the nation's economy. Unfortunately, the system's reliability is declining due to the aging components of the network (Grier, 2009). According to the American Society of Civil Engineers' most recent report card on America's infrastructure, inland waterways received a grade of "D^-" while dams received a grade of "D". Among the most common causes for the degrading status of inland waterways are aging components. On average, dams in the United States are 52 years old, and by

the year 2020, 70% of the dams will be over 50 years old (ASCE report card, 2013b). As a result, locks and dams are frequently closed for unscheduled or scheduled maintenance which causes delays in the flow of commodities and incurs large economic losses across the nation. In 2009, 90% of locks and dams in the US experienced service interruption resulting in an average of 52 delays a day.

The Poisson Bayesian kernel model is applied to analyze the frequency of lock closure due to disruptive events on the Mississippi River transportation network. The network has 29 locks acting as key connectors between different ports nationwide. The navigation system reflects 9,000 miles of navigable waterway with 70.5% of the U.S. inland waterway commodity flowing through the network (Clark, Henrickson, Thoma, 2005).

The data, retrieved from the database collected by the U.S. Army Corps of Engineers (2011), contains detailed information on each lock's characteristics including the river mile, the total number of vessels passing by the lock, the total tonnage, and the frequency and average delay for the vessels and tows experiencing delay time due to the lock's closure. In addition to that, data is available on the yearly frequency of closure for each lock which is considered in this case the outcome to be estimated. A sample of the data is represented in Table 2-4.

Table 2-4: Sample of the inland waterway disruption data

	Y	X_1	X_2	X_3	X_4	…
Lock & Dam	**Closure Frequency**	**River Mile**	**Vessels**	**Tonnage**	**Lockages**	. . .
L&D 3	0	797	9,397	6,747	4,406	. . .
L&D 13	6	523	2,810	14,545	3,155	. . .
L&D 2	0	815	4,478	6,735	2,893	. . .
L&D 20	23	343	2,508	20,828	3,582	. . .
L&D 22	40	301	2,280	22,476	3,486	. . .
L&D 8	6	679	4,333	10,277	2,620	. . .
⋮	⋮	⋮	⋮	⋮	⋮	. . .

Empirical Analysis of the Inland Waterway Data

The goal of deploying the Bayesian kernel model is to obtain an accurate prediction of the frequency of disruptions to inform preparedness strategies and investment decision making. Using the Poisson Bayesian kernel model, decision makers are able to produce a probability distribution of the number of times a particular lock and dam will close each year. The distribution can be used to improve risk management along the inland waterways and make them a more reliable transportation system.

As a first step, the prediction accuracy of the PBK model is tested in comparison with the PGLM and the NBGLM. Similarly to the analysis done in the empirical study, a holdout analysis is performed to assess the goodness of fit and prediction accuracy of PBK for the inland waterway, and the results are summarized in Table 2-5.

Table 2-5: Performance metrics results for the inland waterway data analysis

	Metrics	PBK	PGLM	NBGLM
Full model	LL	-285.06	-148.06	**-75.09**
	DEV	486.03	211.65	**23.89**
	RMSE	**32.82**	63.94	131.03
	NRMSEM	**0.34**	0.66	1.24
	NRMSED	**0.94**	1.88	3.45
	MAE	**21.53**	33.23	57.15
Best model - PGLM	LL	-252.25	**-146.01**	
	DEV	420.53	**208.06**	
	RMSE	**32.60**	42.37	
	NRMSEM	**0.34**	0.46	
	NRMSED	**0.95**	1.34	
	MAE	**20.76**	25.08	
Best model - NBGLM	LL	-238.07		**-78.13**
	DEV	391.33		**24.15**
	RMSE	**28.46**		46.74
	NRMSEM	**0.30**		0.54
	NRMSED	**0.87**		1.56
	MAE	**18.00**		26.01

According to the values of the average out-of-sample error expressed in the four metrics, RMSE, NRMSEM, NRMSED, and MAE, PBK does a better job at making accurate predictions of the average frequency of lock and dam closures even though based on the values of the log-likelihood and deviance, GLM, more specifically the Negative Binomial GLM is better at fitting the data. One of the reasons a GLM might not be providing good prediction errors is overfitting. In order to check whether the results obtained, after fitting a full model that includes all covariates, are due to the GLM overfitting the data, the analysis is performed given the best version of the GLMs. The selection of covariates for each of the Poisson and Negative Binomial GLM is based on Akaike's Information Criterion (AIC) (Akaike, 1970) that penalizes additional parameters contributing to the model complexity. In terms of goodness of fit measures, both reduced GLMs did not express any change in the values of the log-likelihood and

deviance from the full model. However, the prediction accuracy improved significantly for the reduced model with about 60% decrease in the values of RMSE, NRMSEM, NRMSED, and MAE for NBGLM and about 30% decrease for PGLM error measurement values. The covariates selected for the best models were also included in the PBK evaluation for consistency. The PBK still performed better than the best version of both GLMs and maintained better predictive error measures even though the reduced versions of GLMs significantly improved their prediction accuracy.

Prior Distribution Implications

One of the advantages of using Bayesian methods in risk analysis is the flexibility of the approach in (i) establishing assumptions, and (ii) interpreting the results. Any prior belief about the risk measure to be estimated can be embedded in the prior distribution. Determining the prior parameters can be challenging and can result in significant implications on the posterior parameters' estimation. Not much work has been devoted to analyzing priors, but discussions on selecting priors can be found in MacKenzie et al. (2014b), Carlin & Louis (2008), and Guikema (2007). So far, the analysis considered the same prior distribution with prior parameters $\alpha = \beta = 1$ to insure consistency in the empirical study across the different data sets and models. This section examines the implications of changing the priors on the posterior parameters.

In risk analysis problems, experts in the field can help in assessing any prior knowledge about the parameter to be estimated. Ideally, risk managers are interviewed, and using probability elicitation techniques (Spetzler & Stael von Holstein, 1975), a prior probability distribution is defined. Three levels of knowledge are considered in this case that influence the estimation of the priors. For each case scenario, the posterior

frequency of disruptions is computed and compared to results from fitting a PGLM and a NBGLM. The distribution of the RMSE across the three models under each case scenario is used to assess the impact of the priors.

The first approach assumes the experts have a perfect knowledge about the frequency of disruptions and the prior parameters are estimated from the data using the method of moments, Eq. (2-25) , where $\bar{Y}$ and s^2 are respectively the mean and variance of the historical data.

$$\alpha = \frac{\bar{Y}^2}{s^2}$$
$$\beta = \frac{\bar{Y}}{s^2} \qquad (2\text{-}25)$$

The plot in Figure 2-3 shows that the distribution of RMSE values is skewed towards the smaller values (around 25), while the PGLM and NBGLM distributions of RMSE values are spread across larger values with thicker tails. The dashed lines correspond to the mean RMSE showing that PBK performs the best in terms of prediction accuracy.

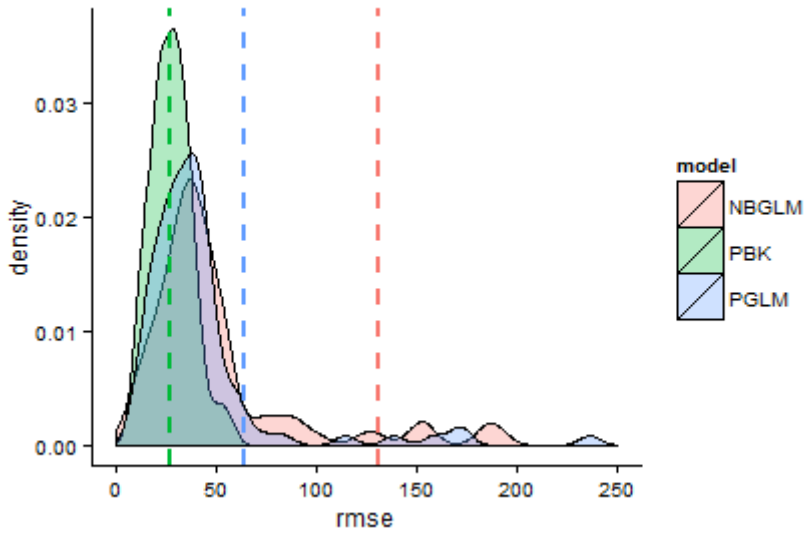


Figure 2-3: RMSE distribution with perfect prior knowledge

The second case scenario assumes that the risk managers have some prior knowledge but it is not perfect like in the first approach. The bias introduced by the risk managers is modeled with a random noise and the distribution of RMSE for the three models is depicted in Figure 2-4. The expected value of RMSE and the overall distribution of the values are both quite similar to the case where the knowledge is assumed to be perfect. Note that the added noise in this case is not very significant. If the risk manager expressed a stronger bias, more noise would be added which could impact the prior parameters estimation and ultimately the posterior distribution and predicted values of the frequency of disruption.

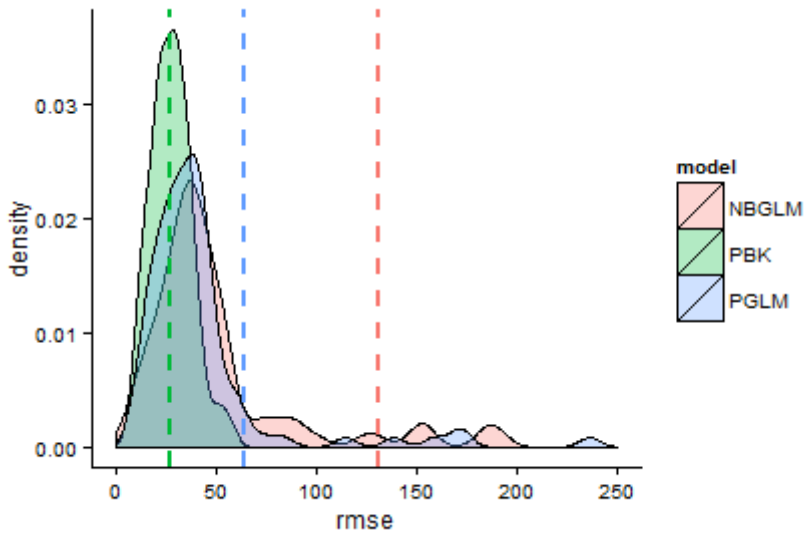


Figure 2-4: RMSE distribution with imperfect prior knowledge

For the third approach, it is assumed that the risk managers have no prior knowledge, or there is no access to a reliable source of information to estimate informed prior parameters. Therefore, the priors are arbitrarily determined and the distribution of the RMSE values is plotted in Figure 2-5. The smaller values of RMSE still have the

highest frequency in the distribution of RMSE for PBK; however this peak is now centered around values equal to 50 as opposed to 25 in the first two approaches. The overall distribution shifted to the right, towards larger values of the RMSE, and the distribution is overlapping with PGLM and NBGLM RMSE distributions. In addition, the expected value of RMSE for PBK increased and is approaching the RMSE expected value of the PGLM, although it is still significantly smaller.

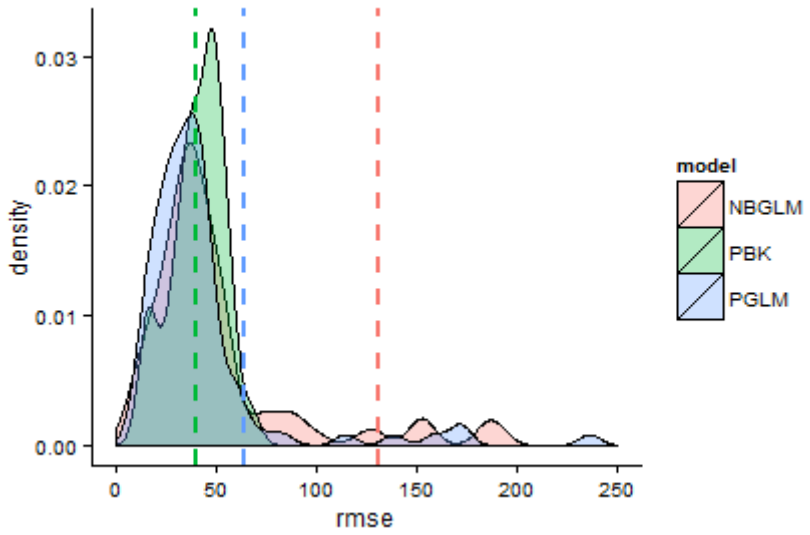


Figure 2-5: RMSE distribution with no prior knowledge

The selection of prior parameters has an implication on the form of the posterior distribution. A poorly formulated prior distribution can impact the performance of the PBK in predicting the frequency of lock and dam closures. On the other hand, a perfect prior distribution relying solely on the historical data is unrealistic. Therefore, the model user must carefully formulate the prior distribution and prior parameters to ensure accurate posterior inferences.

Research Outcome

The model developed in this chapter can impact the decision making process for the protection and rehabilitation of the U.S. inland waterways. As mentioned earlier, this critical infrastructure system is suffering from aging components resulting in frequent disruptions of the flow of commodities across the nation. The Department of Homeland Security announced a set of grant programs to protect and rehabilitate critical infrastructure systems. These grants are normally assigned based on the priority of the rehabilitation project due to the limited availability of resources. Using the hold-one-out analysis approach, the PBK is used to produce a rank of the locks and dams of the inland waterway network. Such a ranking of an infrastructure system's components is one way to implement data-driven risk analysis into real world decision making. Table 2-6 contains the top five locks & dams with the highest predicted frequency of closures per year.

Table 2-6: Locks/dams with highest frequency of closures

Ranking	**Lock/Dam ID**
1	L&D 27
2	Mel Price L&D
3	L&D 19
4	L&D 21
5	L&D 22

Frequent disruptions might be one indication of a component's reliability and urgent need for rehabilitation. As such, the ranking produced in the table above can either be used to allocate grants accordingly or it can be integrated into a multiobjective decision tool that incorporates other factors into the assessment of a rehabilitation

project. In fact, the methodologies developed throughout this dissertation yield different ways to prioritize the inland waterway recovery and repair strategies. Other factors are considered in the following chapters such as the economic interdependent impacts of a disruption and cost of preparedness strategies (Chapter 3), the resilience interdependent effects and cost of repair (Chapter 4), and the resilience importance measures of each component (Chapter 5).

Concluding Remarks

Bayesian kernel methods are powerful tools in forecasting data. These models make use of the Bayesian property by relying on historical data and experts' knowledge, but they also add more specificity to the model by using the kernel function. Gaussian Bayesian kernel models became very popular recently and were extended and applied to a number of classification problems. An important extension to those models is the non-Gaussian model which gives more flexibility in applying this methodology to all types of data set, however, there has been no Bayesian kernel model in the literature that addresses count data.

This chapter introduced count data modeling to the class of Bayesian kernel methods. Using the notion of the conjugate prior, the rate of occurrence is assumed to follow a Gamma prior and posterior distribution using the Poisson likelihood function. The parameters of the posterior distribution are constructed using results from the classical Bayesian Gamma conjugate prior and the exchangeability argument.

The Poisson Bayesian kernel model presented in this chapter is empirically tested and compared with the classical Poisson and Negative Binomial GLM. The three models were used to fit several datasets having similar characteristics in terms of the

size of the data and the number of predictors. The evaluation of the performance of each model is based on the values of metrics corresponding to the goodness of fit and prediction accuracy. Based on the results obtained, the Poisson Bayesian kernel model outperforms the Poisson and Negative Binomial GLM in the majority of the sets for most of the performance metrics representing the out-of-sample error. Also, the Poisson Bayesian kernel model is potentially a better model for small-sized data sets having few predictors. Such a result can be very useful in risk analysis applications to estimate the rate of occurrence of a certain disruption in transportation systems or power grids. In such cases, data can be limited due to the lack of occurrence of the event and the possible factors that might cause a disruption. The need for a more accurate estimation of the rate of disruption can help save lives and lead to more efficient preparedness and recovery investment and allocation.

The Poisson Bayesian kernel model is illustrated using waterway transportation network data of the frequency of lock closure along the Mississippi river, and compared to the classical Poisson and Negative Binomial GLM for the six metrics used in the empirical study. While GLMs exhibit a better fit of the data, the Bayesian kernel model produces a smaller out-of-sample error suggesting a better prediction power.

Accurate predictions of the frequency of disruptions are used to rank the locks and dams and allocate rehabilitation resources accordingly. Realistically, the rank would be one of many criteria used in the decision making process. This chapter addresses the prediction of risk of infrastructure disruptions, the second step would be to understand and quantify the interdependent economic impacts of a disruptive event and how they influence preparedness decision making accordingly.

Chapter 3

Managing the Risk of Interdependent Impacts of Infrastructure Disruptions

Decision making for managing risks to critical infrastructure systems requires accounting for (i) the uncertain behavior of disruptive events which was addressed in Chapter 2, and (ii) the interdependent nature of such systems that lead to large-scale inoperability which will be addressed in this chapter. This work integrates a dynamic risk-based interdependency model, the Dynamic Inoperability Input-Output Model, with a multiobjective decision tree to analyze preparedness decisions. The use of a dynamic model allows for resilience and recovery decisions to be incorporated in the decision making framework, and uncertainty is accounted for using probability distributions. The multiobjective inoperability decision tree is applied to the study of transportation infrastructure disruptions, namely closures of an inland waterway port.

Research focusing on economic impact analyses of disruptions to transportation systems has primarily been devoted to highway and railway transport systems (Sohn, Hewings, Kim, Lee, & Jang, 2004; Gordon et al. 2004; Ham, Kim, & Boyce, 2005a,b) and coastal ports (Rosoff & von Winterfeldt, 2007; Park, 2008; Jung, Santos, & Haimes, 2009). Little work is done to understand the impacts of inland waterway port and network closures: Pant et al. (2011) provide a simulation model of inland port activities to parameterize a port disruption within a multi-regional interdependency model, while MacKenzie et al. (2012a) focus on the multi-regional impacts in

functionality and in economic losses of decision making for shipping alternatives following an inland port disruption.

This work integrates a dynamic interdependency modeling methodology with a stochastic decision analysis methodology to assess investment strategies for infrastructure preparedness, guided by prior work by Santos, Barker, & Zelinke Iv (2008), who developed a static formulation for the interdependent effects of biofuel subsidies and the widespread adoption of biofuels. The dynamic interdependency model and the stochastic decision tree have both been separately used in risk analysis and decision making problems. However, the integration of both approaches with the addition of uncertainty analysis through probabilistic measures is a novel idea constituting the main contribution of this work. The sections of this chapter, in order, provide: the methodological background of the dynamic multiobjective inoperability decision tree (MOIDT); an extension of the MOIDT to address stochastic decision tree problems; a case study of a disruption of the Port of Catoosa, an inland waterway port on the Mississippi River Navigation System near Tulsa, Oklahoma; and concluding remarks. This research has appeared in Baroud et al. (2014a).

Literature Review

This section provides a discussion of several of the components that will make up the dynamic stochastic multiobjective inoperability decision tree deployed for infrastructure preparedness.

Stochastic Decision Tree

A common tool for aiding the decision making process through the graphical depiction of a sequence of decisions and uncertain events is the decision tree (Raiffa,

1968). A simple one-period decision tree is depicted in Figure 3-1, with notation from Santos et al. (2008) and Barker and Wilson (2012). Decisions are made at decision nodes, designated by squares in the decision tree, and branches extending from a decision node represent actions, options, or alternatives from which the decision maker has to choose. The notation in Figure 3-1 for the lth decision alternative for the kth time period in which that decision is made is alternative a_l^k. Chance nodes are designated with a circle, and branches emanating from a chance node represent the states of nature, the occurrences which occur with some known probability. The jth state of nature in the kth time period in which the chance event may occur is represented by s_j^k. The probability of occurrence of a state of nature is represented by $p(s_j^k)$, where $\sum_{j \in J_l^k} p(s_j^k) = 1$ and set J_l^k refers to the set of state of nature subscripts which follow the lth alternative in period *k* and which themselves occur in period *k*. Stochastic decision trees, like that depicted in Figure 3-1, can better incorporate uncertainty by using probability distributions for the likelihood of the occurrence of states of nature instead of probability point estimates (Hespos & Strassman, 1965). The use of probability distributions in stochastic decision trees, as opposed to the point estimates in the traditional decision trees, provides a more comprehensive way to model uncertainty and a wider selection of decision key parameters. For example, some decision makers might be interested in the mean of the distribution while more risk averse decision makers base their analysis on the 70th or higher percentile. The point estimate alone does not offer sufficient information in the selection of parameters representing the outcome function.

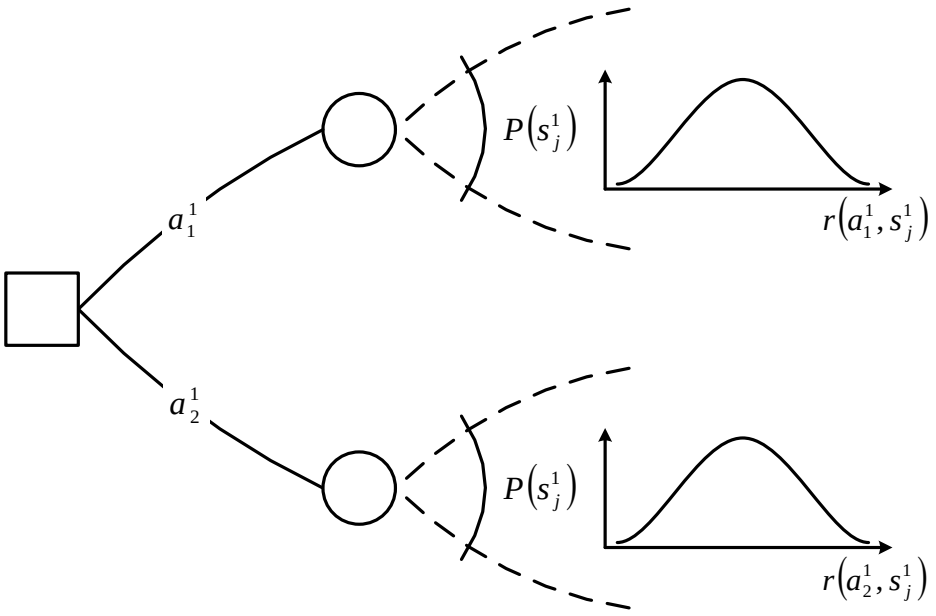


Figure 3-1: Depiction of a general stochastic decision tree

The outcome of a particular path of alternatives and states of nature is quantified with the function $r(\cdot)$. Alternatives are compared by the rollback method, which consists of computing the expected value of each outcome resulting from a specific path followed in the tree. For a *v*-period tree, the expected value for the *l*th alternative at the vth decision is $E[a_l^v] = \sum_{j \in J_l^v} p(s_j^v)\, r(a_l^v, s_j^v)$, and in prior periods, the expected value of the lth alternative at the kth decision is $E[a_l^k] = \sum_{j \in J_l^k} p(s_j^k)\, E[a_j^{k-1}]$. If outcome $r(\cdot)$ represents an adverse outcome (e.g., risk), the alternative which minimizes $E[a_l^k]$ would be chosen at time k. Likewise, if $r(\cdot)$ represents a beneficial outcome (e.g., profit), the alternative that maximizes $E[a_l^k]$ would be chosen. Folding back stochastic decision trees is done using Monte Carlo simulation, as the point estimates $p(s_j^k)$ are replaced with distributions (Dicdican & Haimes, 2005), and consequently $r(\cdot)$ functions

follow a distribution function. The decision is then based on the probability distribution of the expected value.

There are typically a finite number of alternatives and states of nature. The drawback of using a decision tree is that the number of branches emanating from the nodes must be small, otherwise computations can easily become cumbersome. In addition, it might not be possible to sufficiently specify all the alternatives necessary to represent the uncertainty of the event.

Many real-world decisions often require the trade-off of multiple objectives. Haimes, Li, and Tulsiani (1990) provide an approach for dealing with multiple objectives in sequential decision making with the multiobjective decision tree (MODT). The single outcome function $r(\cdot)$ corresponding to a sequential set of alternatives is replaced with a vector-valued outcome function or a vector of m outcomes, $(r_1(\cdot), \ldots, r_m(\cdot))$. The rollback procedure for MODTs is similar to that for decision trees with a single objective. Each sequential path of decisions and chances results in a vector-valued outcome function at the final period *v*, generally represented as $\left[r_1\left(a_i^v, s_j^v\right), \ldots, r_m\left(a_i^v, s_j^v\right)\right]$, noting again that the value of the vector-valued outcome function is found for a specific path of alternatives and states of nature from periods 1 to v. When each of the m outcomes in $(r_1(\cdot), \ldots, r_m(\cdot))$ represents an adverse outcome, a minimum to the vector-valued function is sought at each decision node. However, there may exist no single optimal alternative to roll back to the previous period but more likely a set of noninferior, or Pareto-optimal, solutions. A noninferior solution is defined as a solution to a multiobjective problem where any improvement in one objective comes only at the expense of another objective (Chankong & Haimes, 2008).

Trade-offs among the m competing objectives can be calculated between each string of alternatives.

Interdependency Model

A widely accepted model for describing the interconnected relationships among infrastructure systems and industry sectors is the Nobel Prize-winning economic input-output model (Leontief, 1966), as shown in Eq. (3-1). For a set of n infrastructure and industry sectors, $n \times 1$ vector $\mathbf{x}$ quantifies production outputs in each sector, $n \times n$ matrix $\mathbf{A}$ represents the proportional interdependence among sectors (that is, $\mathbf{Ax}$ represents intermediate demand resulting from the production of $\mathbf{x}$), and $n \times 1$ vector $\boldsymbol{c}$ provides final exogenous consumer demand. As such, Eq. (3-1) describes how changes in consumer demand lead to widespread changes in sector production.

$$\mathbf{x} = \mathbf{Ax} + \mathbf{c} \Longrightarrow \mathbf{x} = [\mathbf{I} - \mathbf{A}]^{-1}\mathbf{c} \qquad (3\text{-}1)$$

The extensive usage of input-output models is due, in part, to the availability of economic interdependency data describing the interconnected nature of infrastructures and industries in a number of countries (OECD, 2011), including an extensive data collection effort by the US Bureau of Economic Analysis (BEA), which maintains input-output tables at different levels of aggregations (BEA 2010). The input-output framework has also been used in modeling interdependent systems that are not economic in nature (e.g., by Setola and De Porcellinis (2008)).

The input-output model was extended to describe the propagation of inoperability, or the proportional extent to which sectors are not performing in an as-planned manner (e.g., reduced production capability), through several interdependent

infrastructure and industry sectors (Santos & Haimes, 2004b). This model, the Inoperability Input-Output Model (IIM), is expressed in Eq. (3-2).

$$\mathbf{q} = \mathbf{A}^{\star}\mathbf{q} + \mathbf{c}^{\star} \Longrightarrow \mathbf{q} = [\mathbf{I} - \mathbf{A}^{\star}]^{-1}\mathbf{c}^{\star} \tag{3-2}$$

Vector $\mathbf{q}$ is a vector of infrastructure and industry inoperabilities, proportional reductions in production, describing the extent to which ideal functionality is not realized following a disruptive event. Inoperability for sector i is defined in Eq. (3-3), where as-planned total output is represented with $\hat{x}_i$ and degraded total output resulting from a disruption is represented with $\tilde{x}_i$.

$$q_i = (\hat{x}_i - \tilde{x}_i)/\hat{x}_i \iff \mathbf{q} = [\text{diag}(\hat{\mathbf{x}})]^{-1}(\hat{\mathbf{x}} - \tilde{\mathbf{x}}) \tag{3-3}$$

An inoperability of 0 suggests that an industry is operating at normal production levels, while an inoperability of 1 means that the industry is not producing at all. Normalized interdependency matrix $\mathbf{A}^{*}$ is a modified version of the original $\mathbf{A}$ matrix describing the extent of economic interdependence among a set of infrastructure and industry sectors. Shown in Eq. (3-4), $\mathbf{A}^{*}$ is an interdependency matrix in which every entry represents how much inoperability is contributed by the column industry to the corresponding row industry.

$$a_{ij}^{\star} = a_{ij}\left(\hat{x}_j/\hat{x}_i\right) \iff \mathbf{A}^{\star} = [\text{diag}(\hat{\mathbf{x}})]^{-1}\mathbf{A}[\text{diag}(\hat{\mathbf{x}})] \tag{3-4}$$

Such a matrix would translate the relationship between industries and the impact caused to an industry by the disruption of another industry sector. If, for example, a disruption affects petroleum production, electric power and transportation industries would also be adversely impacted by the disruption. Note that industry sectors are impacted differently by the disruption of one particular industry, and in some cases, the impact is infinitesimal or does not exist, as a result, the matrix entry would be null. Eq.

(3-5) calculates $\mathbf{c}^{\star}$, a vector of normalized demand reduction, that drives inoperability in the demand-reduction IIM (Santos, 2006).

$$c_i^{\star} = (\hat{c}_i - \tilde{c}_i)/\hat{x}_i \iff \mathbf{c}^{\star} = [\text{diag}(\hat{\mathbf{x}})]^{-1}(\hat{\mathbf{c}} - \tilde{\mathbf{c}}) \qquad (3\text{-}5)$$

The elements of $\mathbf{c}^{\star}$ represent the difference in as-planned demand $\hat{c}_i$ and perturbed demand $\tilde{c}_i$ divided by as-planned production, quantifying the reduced final demand for sector i as a proportion of total as-planned output. The effects of inoperability across multiple sectors can be expressed with total economic losses, $Q = \mathbf{x}^T\mathbf{q}$, or the amount of production made inoperable due to a disruption.

A dynamic version of this model, the Dynamic Inoperability Input-Output Model, or DIIM, calculates the inoperability at any point in time using the recursive formula in Eq. (3-6), quantifying the temporal nature of how inoperability propagates across sectors then dissipates with recovery (Lian & Haimes, 2006). The inoperability vector, $\mathbf{q}(t)$, as well as the vector of demand perturbation, $\mathbf{c}^{\star}(t)$, change in time. An $n \times n$ resilience matrix, $\mathbf{K}$, represents the capability of a certain sector to recover from the disruptive event and reach a desired performance state.

$$\mathbf{q}(t+1) = (\mathbf{I} - \mathbf{K})\mathbf{q}(t) + \mathbf{K}[\mathbf{A}^{\star}\mathbf{q}(t) + \mathbf{c}^{\star}(t)] \qquad (3\text{-}6)$$

One way to estimate the entries in matrix $\mathbf{K}$ is in Eq. (3-7), a result of the dynamic version of Eq. (3-6) with no temporal demand perturbations (Lian & Haimes, 2006). Value $q_i(0)$ is the initial inoperability experienced in sector i following a disruptive event, $q_i(T_i)$ is the desired inoperability state after recovery (assumed to be small but nonzero), which requires T_i time periods to achieve, and a_{ii}^{*} is the diagonal entry in the interdependency matrix.

$$k_i = \frac{\ln\left(\frac{q_i(0)}{q_i(T_i)}\right)}{T_i(1 - a_{ii}^*)} \tag{3-7}$$

The IIM and its extensions have been used in a number of risk-based applications, including inventory decision making (MacKenzie, Santos, & Barker, 2012b; Barker & Santos, 2010a), workforce disruptions (Barker & Santos, 2010b; Orsi & Santos, 2010), and electric power outages (MacKenzie & Barker, 2012; Anderson, Santos, & Haimes, 2007), among others.

Distribution Assumptions

Two sources of uncertainty enter the analysis of infrastructure preparedness as described here: (i) the uncertainty associated with the states of nature that may occur, as depicted in the decision tree, and (ii) the uncertainty associated with how sequences of decisions and states of nature will impact demand perturbations in the interdependency model (thereby driving the calculation of inoperability across sectors). These two representations of uncertainty are described subsequently.

Discussed in more detail in the methodological development section, the states of nature describe the magnitude of a disruptive event at the inland port, where such an event may be rare in nature with large impacts. Originally introduced through the idea of the scale invariance (Richardson, 1948), which is an inverse power scaling between independent and dependent variables, the power-law distribution has been used to model different social, political, and financial patterns in history. Studies aimed at determining the probability distribution of the severity of a terrorist attack agree that, according to empirical data on worldwide terrorist events from 1968 to 2008, the

probability of a terrorist event claiming x deaths follows a power-law distribution (Clauset & Young, 2005; Johnson et al., 2006; Clauset, Young, & Gleditsch, 2007).

$$f(d) = Cd^{-\lambda} \tag{3-8}$$

This model, which is provided in Eq. (3-8), for scale parameter λ and normalizing constant C, has been extended to a more general form to describe other covariates such as the size and experience of the terrorist organization and the type of weapons used in the attack (Clauset & Gleditsch, 2012; Clauset & Wiegel, 2009). Power-law relationships are also used to model the severity of natural disasters such as tornadoes (Malamud & Turcotte, 2012), other severe weather conditions (Dessai & Walter, 2000), and large earthquakes (Mega et al., 2003), among others. Statistical methods such as the least square method, the maximum likelihood function, or an extended combination of these with other statistical tests are used to estimate the parameters of this distribution (Clauset, Shalizi, & Newman, 2009): , C, and minimum value of the random variable d_{min}.

The manner in which the disruptive event affects changes in demand is itself represented with a probability distribution. Like Santos et al. (2008) and Santos (2008), the beta distribution is used here to describe uncertainty in $\mathbf{c}^{\star}$. An advantage of using this distribution, whose probability density function is shown in Eq. (3-9), is that its support is on [0,1]. This is beneficial in describing a parameter such as $c_i^{\star}$, the minimum and maximum are controllable, and the range of values of its parameters determine several shapes for the probability distribution.

$$f(x) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1} \tag{3-9}$$

The mean and variance of the beta distribution are shown in Eqs. (3-10) and (3-11), respectively.

$$\mu = \frac{\alpha}{\alpha + \beta} \tag{3-10}$$

$$\sigma^2 = \frac{\alpha\beta}{(\alpha + \beta)^2(\alpha + \beta + 1)} \tag{3-11}$$

The parameters of this distribution are estimated using the method of moments. Given the values of the sample mean, $\bar{X}$, and sample variance, s^2, the parameters are estimated with Eqs. (3-12) and (3-13).

$$\hat{\alpha} = \bar{X}\left(\frac{\bar{X}(1-\bar{X})}{s} - 1\right) \tag{3-12}$$

$$\hat{\beta} = (1-\bar{X})\left(\frac{\bar{X}(1-\bar{X})}{s} - 1\right) \tag{3-13}$$

Methodology: Dynamic Multiobjective Inoperability Decision Tree

Extending a static version of the multiobjective inoperability decision tree (MOIDT) (Santos et al., 2008), this section describes a dynamic and stochastic MOIDT to model the (i) investment in infrastructure preparedness, and (ii) the interdependent benefits of such investments (or adverse impacts of a lack of investment).

The use of the DIIM considers the cumulative effect of the total economic loss, a risk measure of interest to decision makers. Also, the problem is solved here by taking into account the uncertainty in the parameters of the DIIM as well as in the states of nature in the decision tree. As a consequence, the parameters of the beta distribution for

each entry in the demand perturbation vector, $\mathbf{c}^{\star}$, depend on the state of nature as well as on the point estimates of each entry.

Infrastructure Preparedness Decision Problem

An influence diagram is a useful graphical tool that helps to visualize the decision problem in terms of the decision, uncertainties, objectives, and the influential relationship each has on the others. For the infrastructure preparedness decision problem in Figure 3-2, the decision to make, represented by a rectangle, is a certain amount of dollars invested in preparedness activities. This decision will directly affect one of the objective functions, represented by a diamond shape, which is the cost. It will also indirectly affect the other objective function which is the expected total economic loss as it will impact two uncertainties, represented by a circular shape: (i) the probability of a disruptive event occurring (in case of a manmade attack), and (ii) the severity of the disruptive event. Hence, the problem presented is a multiobjective decision problem in which the decision maker is interested in minimizing both the cost of preparedness and the expected total economic loss in case of a disruptive event. Both objectives are competing (as increased investment in preparedness would expectedly lead to fewer losses). And while both are measured in dollar terms, the objectives are noncommensurable in that investment budgets would likely come primarily from government sources, while losses would be experienced across a wide number of infrastructure and industry sectors.

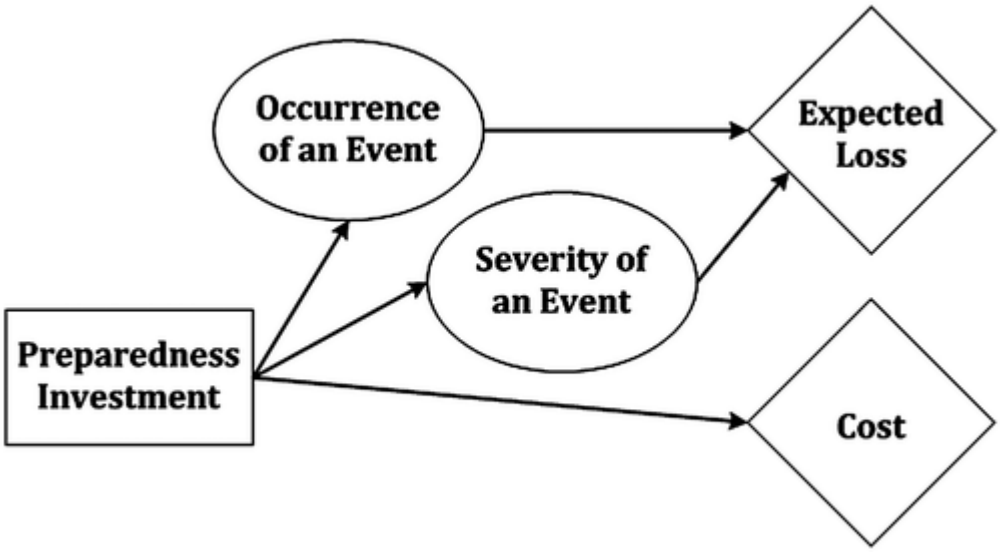


Figure 3-2: Influence diagram describing the infrastructure preparedness investment decision-making process

From Figure 3-2, the preparedness investment decision directly affects the cost, and this investment has an effect on the likelihood and severity of a disruptive event. The way in which the investment affects these others quantities is through a factor of influence, θ_I, linearly related to the investment amount, I, and the maximum amount to be invested, I_{max}, as shown in Eq. (3-14).

$$\theta_I = \frac{I}{I_{max}} \tag{3-14}$$

The factor of influence is used as a translation of the amount invested from dollars to fraction terms expressing how large is the investment compared to other possible investments. In particular, the amount invested is compared to the maximum investment that can be made and is then expressed as a value between 0 and 1, providing the factor of influence. The impact of the investment on the likelihood and severity of the disruption can then be modeled using this factor. A simple linear relationship between the investment and the factor of influence is more than enough to

express how large the investment is in proportion terms. A more sophisticated relationship would make unnecessary complications to the overall model. Note that the impact of investment through that factor of influence may not be as simple as a linear relationship and is thoroughly discussed in the following section.

Decision Tree Construction and Parameters

A dynamic MOIDT is constructed in Figure 3-3 to address the investment in infrastructure preparedness, where the decision addresses the amount of investment in period k, I_j^k. Two chance nodes are assumed: the first concerns the occurrence of a disruptive event, and the second concerns the severity of the event if it occurs. The occurrence of the event is a deterministic chance node where the likelihood of occurrence is known to be p. The second chance node representing the severity has infinitely many states of nature to represent the severity of the disruption (thus, a stochastic chance node), and consequently the demand perturbation and ultimate total economic losses across all sectors. The stochastic nature of the severity of the disruption is modeled with the power law distribution, and the demand perturbation is modeled with the beta distribution. While multiple planning periods could be explored with this problem, a one-period tree is addressed here.

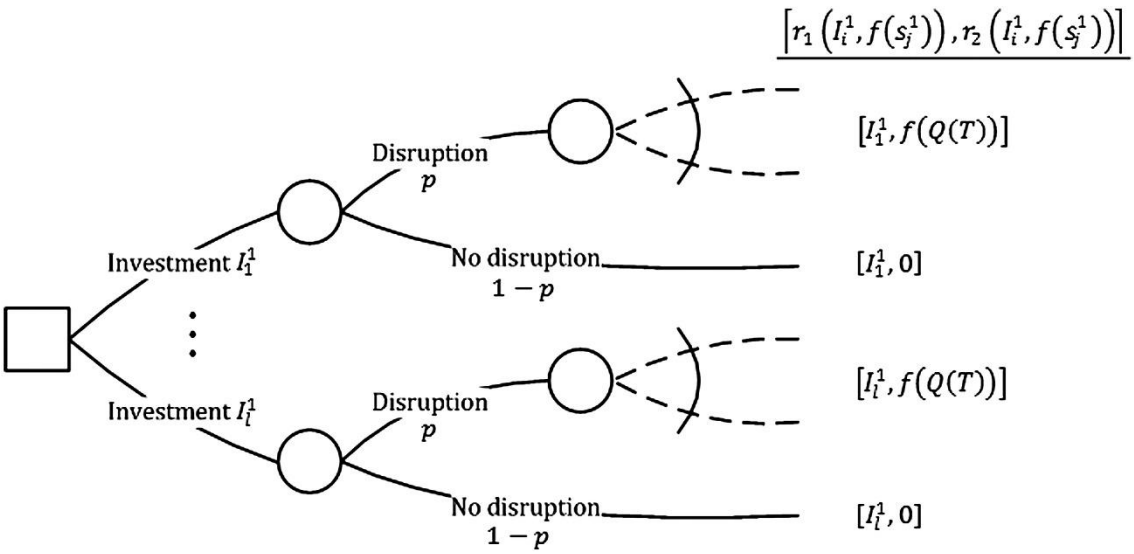


Figure 3-3: Multiobjective stochastic inoperability decision tree for infrastructure preparedness

The outcome of each sequential decision and state of nature is a (i) cost of investment and (ii) total economic loss across all sectors resulting from the disruptive event corresponding to the chance node. The relationships below describe how the investment decision, through the factor of influence, θ_I, from Eq. (3-14), alters the different parameters in the model. The probability of a disruptive event occurring is decreased with increasing investment, as governed by Eq. (3-15).

$$p_I = e^{-\theta_I} p \tag{3-15}$$

That is, depending on the nature of the investment, some amount of protection is assumed in preventing the event (or reducing its likelihood of occurrence). The probability of the occurrence of the event, p, is reduced to p_I.

Dynamic Recovery Model

It is assumed that a disruptive event befalling the infrastructure system impacts demand for commodities whose flow or production is enabled by the infrastructure (e.g., commodities that flow through a port would be impacted by the port's closure).

Such a demand perturbation for some commodities could be the result of a supply reduction in other commodities. Given the magnitude of disruptive events in infrastructure systems and their large economic impacts, industries and consumers overall tend to become more conservative in demand activity due to the instability of the economy after the disruption. In addition, supply or production could also be impacted by the disruption. In fact, the inoperability vector in the DIIM is representative of the supply reduction following disruptive events. To account for the uncertainty in the demand reduction, the values of the entries in the $\mathbf{c}^{\star}$ vector are drawn from the beta distribution whose parameters depend on the severity of the disruptive event. The severity is generated from the stochastic chance node in the decision tree following a power-law distribution.

The severity of the event and the amount invested in preparedness activities are two factors having opposite impacts on several distributional parameters as well as other parameters in the model. As noted earlier, the shape and scale parameters of the beta distribution for $\mathbf{c}^{\star}$ are estimated using the method of moments according to Eqs. (3-12) and (3-13). This method considers the sample mean and variance to be known. The sample mean of the demand reduction, $\bar{X}$, is directly related to the maximum demand perturbation $\mathbf{c}^{\star}_{\max}$, with the latter representing the worst case scenario that could occur. For example, this could describe a one-year closure of a transportation facility (e.g., inland waterway port) or any type of prolonged dysfunction of an infrastructure affecting the economy. Such scenarios are generally assessed by means of expert solicitation.

The sample mean of the demand perturbation is in fact a proportion of the maximum perturbation, and this proportion, γ, depends on the severity of the event. For example, if severity is measured in terms of the number of months a certain facility is closed, and the worst case scenario assumes a full year of closure, then $\gamma = 0.75$ for a nine-month closure. Further, the sample mean has a negative exponential relationship with the factor of influence, θ_I, computed in Eq. (3-14). This relationship ensures that as the amount of investment is increased, the reduction in demand after a disruptive event is exponentially decreased. These relationships are represented in vector form in Eq. (3-16), where $\bar{\mathbf{X}}$ increases as the proportion of the maximum demand perturbation increases, depending on the severity of the event, and it decreases exponentially as the factor influence (or equivalently the amount invested) increases.

$$\bar{\mathbf{X}} = \gamma e^{-\theta_I} \mathbf{c}^{\star}_{\max} \tag{3-16}$$

Likewise, standard deviation $\mathbf{s}$ is computed by assuming that the maximum demand perturbation lies three standard deviations above the mean, shown in vector form in Eq. (3-17). No further restrictions or assumptions are applied to the sample variance, which results in larger variations when the sample mean is smaller.

$$\mathbf{s} = \frac{1}{3}(\mathbf{c}^{\star}_{\max} - \bar{\boldsymbol{X}}) \tag{3-17}$$

Similar to the sample mean, the initial inoperability is a function of both the investment strategy, θ_I, and the severity of the event, γ, as shown in Eq. (3-18), where $\mathbf{q}_{\max}(0)$ represents the vector of initial inoperability under the most extreme circumstances. The values of $\mathbf{q}_{\max}(0)$ could either be determined by experts or estimated depending on the system considered in the application.

$$\mathbf{q}(0) = \gamma e^{-\theta_I}\mathbf{q}_{\max}(0) \tag{3-18}$$

The concept of resilience will be revisited in much more details and in-depth analysis in Chapter 4 and Chapter 5. For the purpose of this analysis, resilience is defined in two aspects: (i) reducing the impact of a disruptive event, and (ii) improving the speed with which recovery occurs (Henry & Ramirez-Marquez, 2012; Zobel, 2011). The effect of an investment in improving the first aspect is found in Eqs. (3-16) and (3-18), where the initial inoperability and the initial demand perturbation experienced after a disruptive event is lessened with an investment in infrastructure preparedness. The second aspect is addressed in Eq. (3-19), where preparedness investments can also decrease the time to full recovery of industry i, T_i. Like the previous relationships, recovery time is reduced according to $e^{-\theta_I}T_i$, an exponentially decreasing function of the factor of influence, θ_I. As recovery commences, inoperability decreases over time, consequently impacting total economic losses across all sectors over time, $Q(t) = \mathbf{x}(t)^T\boldsymbol{q}(t)$.

$$k_i = \frac{\ln\left(\frac{q_i(0)}{q_i(e^{-\theta_I}T_i)}\right)}{(e^{-\theta_I}T_i)(1-a_{ii}^{\star})} \tag{3-19}$$

The amount invested in port security and system hardening, in general having an exponential impact on the effect of a disruption, is a more realistic representation of the real life behavior of improvements in the risk management of infrastructure systems. The idea is deemed useful in such as way that is not too simplistic, such as the case of a linear relationship, and not too complicated in a manner to make it problem-specific. Similar ideas have been implemented in other scenarios, such as resource allocation problems (MacKenzie, Baroud, & Barker, 2014a).

Note the double effect on the input parameter from the severity of an event leading to more perturbation, more inoperability, and less resilience; while risk management helps in decreasing the initial impact and increasing the recovery rate by decreasing the time required to full recovery.

It is also worth noting that the demand perturbation, which is a function of time, starts to decrease as recovery commences and is updated at discrete time points from $t = 1$ to $t = T = \max(T_1, T_2, \dots, T_n)$. Updating the demand reduction during the recovery process depends on the case study and will be further developed in the case study section. The decision is then based on the aggregate economic loss computed at the time, T, when all industries are fully recovered, and since the decision tree is a one-period, more precisely one-year decision tree, there is one investment to be made.

Simulation of the Decision Tree

Alluded to previously, stochastic decision trees require the use of Monte Carlo simulation in the folding back process. For each alternative emanating from the first decision node, a fixed cost representing the amount invested is assumed, and θ_I for an investment of I can be computed. The following steps guide the simulation for a sufficiently large number of iterations.

1. Generate a random variable from the power law distribution within the bounds of the minimum and maximum severity, guided by Eq. (3-8).
2. Given the severity, factor of influence, θ_I, and the elicited values $\mathbf{c}^{\star}_{\max}$ and $\mathbf{q}_{\max}(0)$, calculate the sample mean, $\overline{\mathbf{X}}$**,** and standard deviation, **s**, of the demand perturbation from Eqs. (3-16) and (3-17), the parameters of the

DIIM from Eqs. (3-18) and (3-19), and the probability, p_I, of a disruptive event occurring from Eq. (3-15).

3. Using the method of moments, compute estimates for the parameters of the beta distribution from Eqs. (3-12) and (3-13) and draw random variables from this distribution for each industry element in the demand perturbation vector. It is assumed that the number n' of initially perturbed industries can represent a subset of all industries, $n' \leq n$.
4. Compute the inoperability vector at each point in time from Eq. (3-6).
5. Compute the total economic loss, $Q = Q(T) = \mathbf{x}^T \sum_{t=1}^{\tau} \mathbf{q}(t)$.
6. Compute the expected value of the total economic loss as the product of the loss and the probability p_I of a disruptive event, $EQ = EQ(T) = Q(T)p_I$

Repeating the above steps for a number of iterations, N, for one particular investment cost results in a distribution for *EQ* under this specific investment. The mean values of the *EQ* distributions for each investment could be compared to determine the appropriate investment level. Further, the conditional expected value (upper-tail value) provides an idea of how each investment performs in extreme conditions (Asbeck & Haimes, 1984).

Case Study: Inland Waterways Preparedness Strategies Assessment

The methodology developed is illustrated with a decision problem regarding inland waterway port security and resilience. This data-driven case study involves the Port of Catoosa in Tulsa, Oklahoma, the largest inland port in the U.S. in terms of area. Located on McClellan-Kerr Arkansas River, the port is part of the Mississippi River Navigation System and roughly two million tons of commodities flow annually through

the port. As such, preparing for disruptive events is crucial to the port itself as well as to the economic system depending on the trading activity at the port.

Motivation

The Department of Homeland Security announced a set of grant programs targeting different areas prone to willful attacks or natural disasters (DHS 2012). Among this set of grants is the Port Security Grant Program (PSGP) whose purpose is "to support increased port-wide risk management; enhanced domain awareness; training and exercises; expansion of port recovery and resiliency capabilities; and further capabilities to prevent, detect, respond to, and recover from attacks…and…assist ports in obtaining the resources required to support the NPG's associated mission areas and core capabilities" (DHS, 2012). According to the grant overview report (DHS 2012), the ports in the U.S. were categorized into three different groups, and DHS divided the grant of the PSGP among the three groups. However, the ports within each category were to compete for the funding available in their group. According to that report, the group in which the Tulsa Port of Catoosa is placed contains a total of 22 ports and received a grant of $4,875,000. The approach developed in this research could provide a helpful means to determine the appropriate investment amount allocated to each port based on the port's size, its mix of annual commodity flows, and its location, among other factors impacting the probability of a disruptive event occurring as well as its consequences in terms of economic losses.

Assumptions

The probability of a (rare) event causing a disruption of an extreme nature at the port is assumed to be similar to the probability of a terrorist attack since this type of

event would incur large impacts and rarely occurs. This assumption is merely for illustration purposes, a sensitivity analysis will follow to address scenarios with different probabilities. Bun (2006) developed a mathematical model to measure the probability of a nuclear terrorist attack, finding that, with a set of plausible parameters, the probability of a nuclear terrorist attack in the next 10 years is 0.29. Extending from this result, the probability associated with a terrorist event occurring in a one-year period is assumed to be 0.29/10 = 0.029 which corresponds to one period of a decision tree. This computation assumes independence between probabilities of occurrence of a terrorist attack over the years and generalizes the likelihood of a nuclear terrorist attack to any broadly-defined disruptive event. Therefore, a baseline probability for a disruptive event occurring in one year is assumed to be 0.02 for this particular application. Given that the result used is specific to a nuclear attack, it might seem unrealistic to model any type of disruptive event occurring at the port of Catoosa. However, a baseline probability is set to be slightly less than the given result and followed up by a sensitivity analysis considering other values for the probabilities. The choice of this value serves solely for illustration purposes, though motivated from previous work.

Generally extreme events that are modeled with a power-law distribution result in an estimated scale parameter ranging between 2 and 3, for further discussion on the distribution of disruptive events, please refer to the *Distribution Assumptions* section above. It is then assumed that $\lambda = 2.5$. In addition, it is assumed that the severity of the disruptive event is bounded with a minimum and a maximum number of days during which the port is closed, between 15 and 60 days.

The cost of investment, I, ranges from \$0 to \$1,000,000 and the proportion parameter of the severity of the event, γ, is the number of days the port is closed divided by 365, meaning that the maximum perturbation and initial inoperability represent a one year closure of the port. Since this case study is motivated by the amount of grant from PSGP allocated to the group of ports to which the Port of Catoosa belongs, the maximum amount of investment is determined accordingly. It is also assumed that all the industries require the same time, $T = T_i$, to fully recovery, where T depends on the duration of the closure of the port. In this particular example it is assumed that $\kappa = 0.5$, and that if the port was closed for d days, the industries relying on the port as well as the interdependent industries will all recover in $T = \kappa d$ days. This assumes that while the port was closed, the products that were supposed to be exported were held at the port and those that were supposed to be imported were held at their original port, and after the port reopens, the port needs T days to ship all the products held and hence the industries would be fully recovered. This would mostly affect the manner in which the demand perturbation is decreasing in the recovery period, and instead of adding $\mathbf{c}^{\star}(t)$, in Eq. (3-6), the change in the demand perturbation, $\Delta_{\mathbf{c}^{\star}}$, Eq. (3-20), is added, which considers that once the recovery starts, the commodity flows through the port at a rate of $1/\kappa$. In this case the flow is twice as much as it usually is to make up for the loss and the time during which the commodity had to wait at the port. During the recovery, the daily demand perturbation is subtracted from the initial demand perturbation and updated accordingly every day until $\Delta_{\mathbf{c}^{\star}}(T) = 0$. Note that while the port was closed for d days, the recovery process does not start until the port reopens. This suggests that the port and associated industries are not functioning for d

days, after which the recovery process commences and spans $T = \kappa d$ days before the port and industries are fully recovered. However, since the DIIM models the recovery process, the model is implemented for $T = \kappa d$ days and not the full period of time required to fully recover, $d + \kappa d$.

$$\Delta_{\mathbf{c}^\star}(1) = \mathbf{c}^\star(0) - \frac{1}{e^{-\theta_I T}}\mathbf{c}^\star(0)$$
$$\Delta_{\mathbf{c}^\star}(t) = \Delta_{\mathbf{c}^\star}(t-1) - \frac{1}{e^{-\theta_I T}}\mathbf{c}^\star(0) \tag{3-20}$$

Finally, the demand perturbation comes from the loss in exports and is computed as the ratio of exports of industry i to the total production output of that industry at a specific time, t, as shown in Eq. (3-21).

$$c_i^\star(t) = \frac{e_i(t)}{x_i(t)} \tag{3-21}$$

Similarly, the inoperability is the result of imports not reaching the port and causing a shortage in the material needed to produce in the industries relying on such commodity, it is then the ratio of imports to the total production output of industry i, provided in Eq.(3-22) (MacKenzie et al., 2012a; Pant et al., 2011).

$$\hat{q}_i(t) = \left(\frac{m_i(t)}{x_i(t)}\right) \tag{3-22}$$

If both the numerator and the denominator in Eqs. (3-21) and (3-22) are yearly estimates, the above metrics can serve as the maximum demand perturbation and the maximum inoperability since the worst case scenario is considered to be a full year of port closure and they are used to compute the initial demand perturbation and inoperability. The elements of the sample mean and initial inoperability vectors in Eqs. (3-16) and (3-18) are then expressed in Eqs. (3-23) and (3-24), where $e_i(t)$, $m_i(t)$, and

$x_i(t)$ respectively represent the exports, imports and total production output for industry i at time t of one year.

$$\bar{X}_i = \gamma e^{-\theta_I}\left(\frac{e_i(t)}{x_i(t)}\right) \tag{3-23}$$

$$q_i(0) = \gamma e^{-\theta_I}\left(\frac{m_i(t)}{x_i(t)}\right) \tag{3-24}$$

The input data described above relies on either (i) parameter assumptions that could be altered depending on the application, or (ii) port-specific data. Several data sources from the Tulsa Port of Catoosa, the US Army Corps of Engineers, and the Bureau of Transportation Statistics were used by MacKenzie et al. (2012a) to derive estimates of the dollar amount of commodities flowing through the port of Catoosa. These estimates are inputs to Eqs. (3-21) and (3-22). The interdependencies between the critical infrastructure and the rest of industries are expressed by two matrices in the DIIM model: (i) the $\mathbf{A}^*$ matrix describes how inoperability propagates among industries when the infrastructure is disrupted, and (ii) the **K** matrix governs the recovery of infrastructure and industry sectors. The entries in the $\mathbf{A}^*$ matrix are computed using data from the Bureau of Economic Analysis, and the formula for the entries of the **K** matrix is discussed the *Interdependency Model* background section, Eq. (3-7). The idea of expressing initial inoperability and demand perturbation as a fraction of the worst case scenario has been used in a different study on port disruptions (Pant et al., 2011). The length of closure of the port is chosen to be anywhere between two weeks and two months, a reasonable assumption in the case of any disaster (Pant et al., 2011). And the time to full recovery controlled by κ, is randomly chosen. These parameters can be

altered according to each case and largely depend on the decision makers and the type of risk they seek to mitigate.

Decision Tree Solution

A simulation of 100,000 iterations is performed, and several metrics related to inoperability and economic loss are computed for each investment amount using the aggregate inoperability throughout the recovery period.

The distribution of total economic loss without preparedness is depicted in Figure 3-4. The property of the power-law distribution can be easily seen, as disruptive events with larger impacts have a smaller chance of occurring while less impactful events are more common. An important feature of this distribution is that it does not rule out extreme events as outliers but rather considers them as events with an infinitesimal likelihood of occurrence, important for a risk averse decision maker who is interested in minimizing extreme risks.

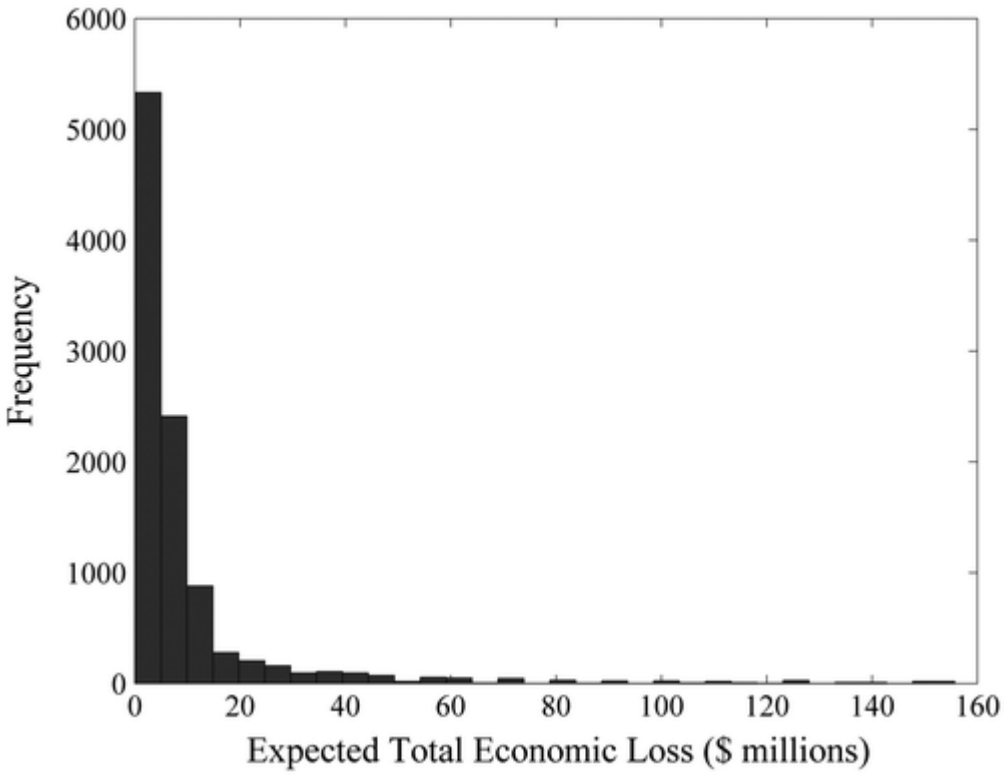


Figure 3-4: Frequency distribution of the expected total economic losses across all regional industries

Recall that this MOIDT is constructed for two minimization objectives: (i) expected total economic loss across all industries, and (ii) investment amount. While both of these objectives are measured in dollars, they are not necessarily commensurate as the preparedness investment would likely come from a port authority or DHS funding program while economic losses would be shared across multiple industries. The resulting Pareto-optimal frontier plotting expected total economic loss, $EQ(T)$, computed once the industries are fully recovered versus investment is shown in Figure 3-5. Note that, as Figure 3-4 suggests, there exists a probability distribution for expected total economic loss for each realization of investment. Shown in Figure 3-5 are the mean, M, of these distributions, as well as conditional means $CM_{0.01}$ and $CM_{0.05}$.

Although the Pareto-optimal frontiers in Figure 3-5 are not completely smooth, it is concluded that none of the investment strategies is dominant due to the overall shape of the curve. Additional iterations would likely result in smoother curves.

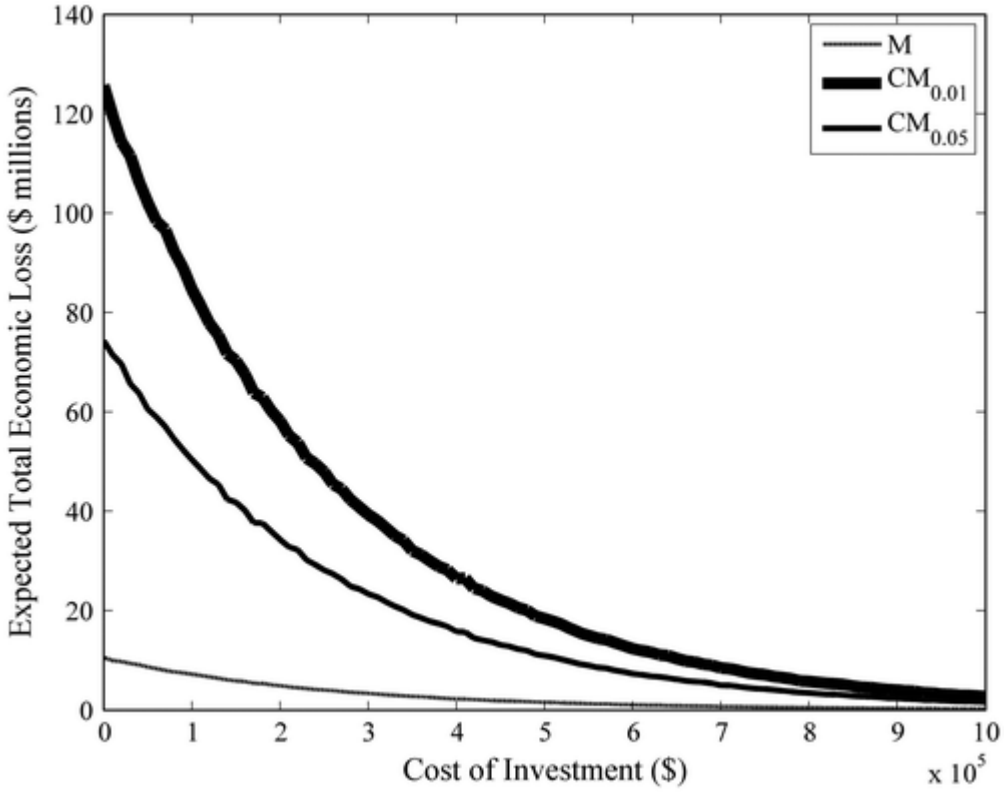


Figure 3-5: Pareto frontier for expected total losses versus the amount invested toward preparedness activities

The conditional expected total economic loss values are upper-tail or extreme values associated with higher consequence, but lower probability, events (Asbeck & Haimes, 1984). The conditional expected value is found with Eq. (3-25) and indicator function in Eq. (3-26).

$$CM_r = E[EQ|\beta < EQ < \infty] = \frac{\sum_{i=1}^{N} \delta(i) EQ(i)}{\sum_{i=1}^{N} \delta(i)} \qquad (3\text{-}25)$$

$$\delta(i) = \begin{cases} 1 & \text{if } EQ(i) > \beta \\ 0 & \text{otherwise} \end{cases} \quad (3\text{-}26)$$

The use of an upper-tail conditional expected economic loss calculation is particularly telling for the power-law distribution, which naturally has high consequence, low probability values in its upper tail. The indicator function is used to select the observations falling under the upper-tail of the probability distribution and include them in the computation of the conditional expectation, a means to determine the extent to which the extremity of the event is considered. Depending on the decision maker, more or less extreme case scenarios could be analyzed to determine the amount of preparedness investment needed (e.g., risk averse decision makers tend to be prepared for more extreme events). Hence, using the indicator function, it is possible to determine the degree of risk aversion of the decision maker (Santos & Haimes, 2004b). The formula in Eq. (3-27) calculates the significance level which is the probability corresponding to the upper-tail portion of the distribution of *EQ*.

$$r = \frac{1}{N}\sum_{i=1}^{N} \delta(i) \quad (3\text{-}27)$$

The larger the r, the less risk averse is the decision maker since the calculation is incorporating more points into the expectation and the risk aversion is moving towards a risk neutral decision making that uses the mean. According to Figure 3-5, all the metrics decrease as the investment cost increases, suggesting that the more the risk manager invests, the less impactful is the disruptive event in terms of economic losses. If the investment is higher than $800,000, both the mean and conditional means provide very similar estimated values for the expected total economic loss. $CM_{0.01}$ and $CM_{0.05}$ are two conditional expectation for different levels of risk aversion. Risk averse

decision makers would have to invest much more to get the same expected total economic loss as a risk neutral decision maker. For example, a risk neutral decision maker investing $100,000 would see an expected total economic loss of about $10 million, while a risk averse decision maker aiming to reach the same level of expected economic loss would need to invest $500,000 to $700,000 depending on their extent of risk aversion.

Recall that no particular restriction was applied to the variance of the beta distribution except that $\mathbf{c}^{\star}_{\max} = \bar{X} + 3s$. Figure 3-6 is a plot of the standard deviation of the total expected economic loss as the investment changes. Note that the variation in the expected total economic loss decreases as the investment increases, suggesting that no further restriction is required for the variance of the beta distribution as the model accounts for the behavior of the variance of the expected total economic loss, ensuring that it decreases with larger investments. Noted previously, additional iterations would result in smoother curves and hence the small jumps along the curve should be considered insignificant.

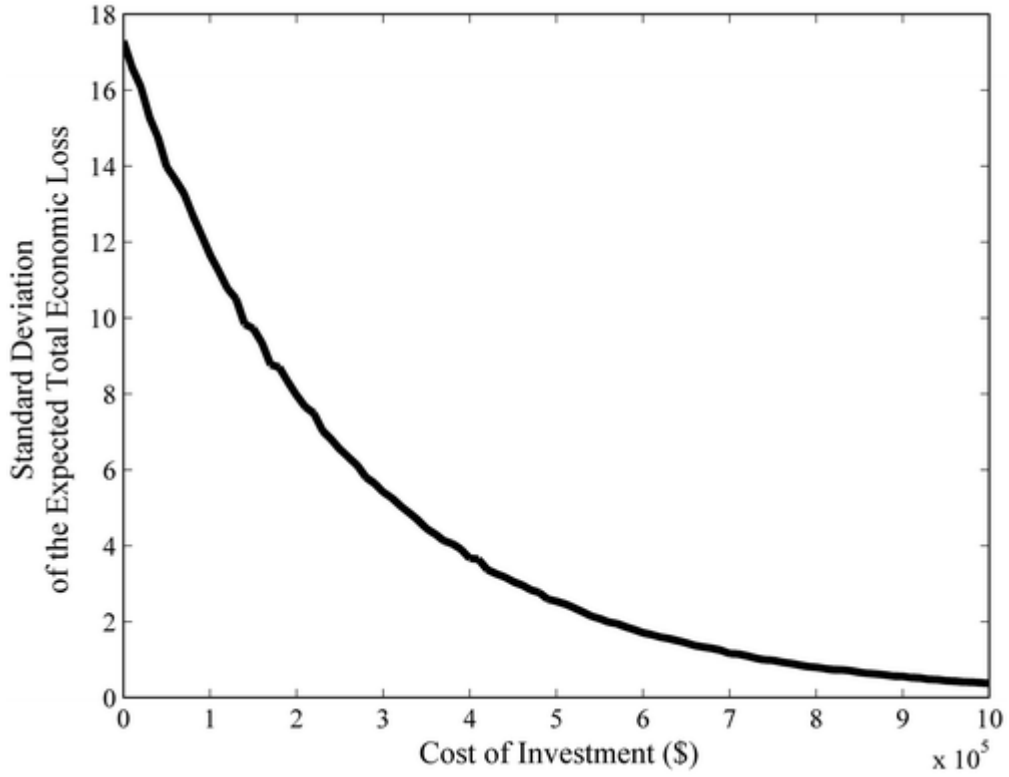


Figure 3-6: Behavior of standard deviation of the expected total economic loss as investment increases

Trade-off Analysis

A trade-off analysis is necessary to study the additional effect on the total economic loss of investing an extra dollar in port security. The trade-off analysis is done using the generalized trade-off function for discrete strategies used in Santos et al. (2008). The trade-off of a risk function f_m with respect to another risk function f_n for two different risk management strategies S_i and S_j is computed in Eq. (3-28).

$$\lambda_{(f_m, f_n|S_i S_j)} = -\frac{f_{n|S_i} - f_{n|S_j}}{f_{m|S_i} - f_{m|S_j}} \quad \text{for } i \neq j \tag{3-28}$$

The trade-off analysis gives insights on the efficacy of additional risk management investments. It is true that the more the decision maker invests, the less the expected economic loss would be, however, how efficient is this additional investment? Consider

the numerical example in Eq. (3-29), investing $100,000 more would decrease the expected total economic loss by $5 million. A trade-off of 50 means that the total expected loss decreases by $50 for each $1 invested in preparedness.

$$\lambda = -\frac{30 - 35}{0.2 - 0.1} = 50 \tag{3-29}$$

As long as the trade-off is positive, neither of the two strategies considered dominates the other. However, the value of that trade-off indicates how efficient the investment is. Note in Figure 3-7 that the trade-off decreases as the amount invested increases while it remains positive. This means that although the investment is helping to decrease the expected total economic loss, the efficiency of that risk management procedure is decreasing as the investment increases. The same pattern is observed in Figure 3-8, in which the trade-off is based on the values of $CM_{0.01}$ from Figure 3-5. Although it has a similar pattern of decreasing the efficiency of the investment as more money is invested, however the value of the trade-off is much larger than the one computed using the mean value of the expected total economic loss. That is due to the decision maker being risk averse, any additional investment would highly impact the risk function the decision maker is trying to minimize.

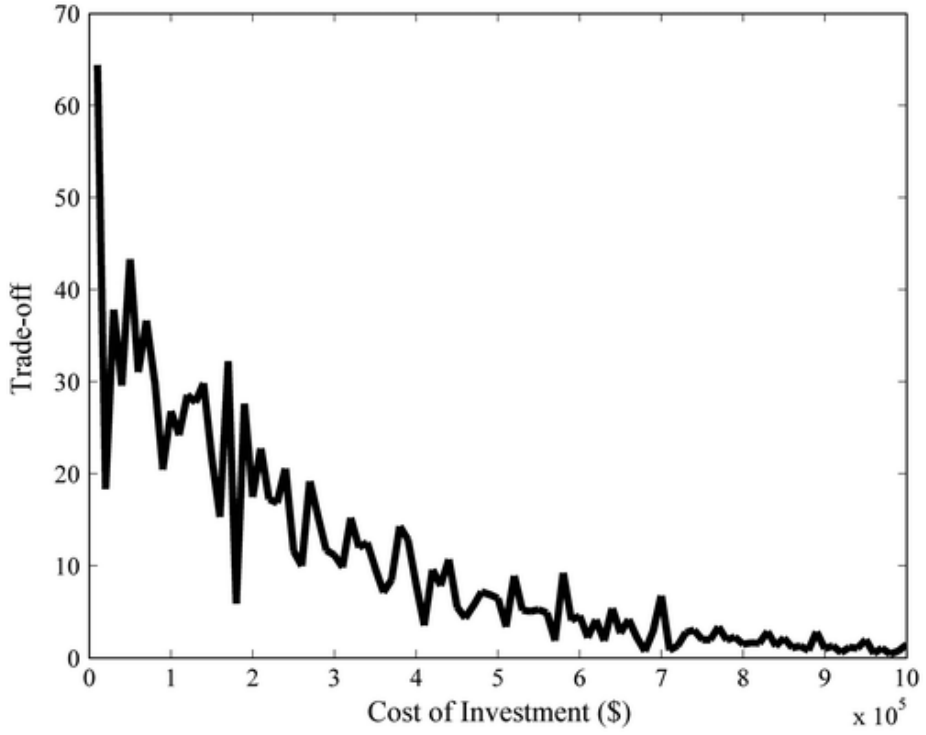


Figure 3-7: Trade-off as a function of the cost of investment for a risk-neutral decision maker

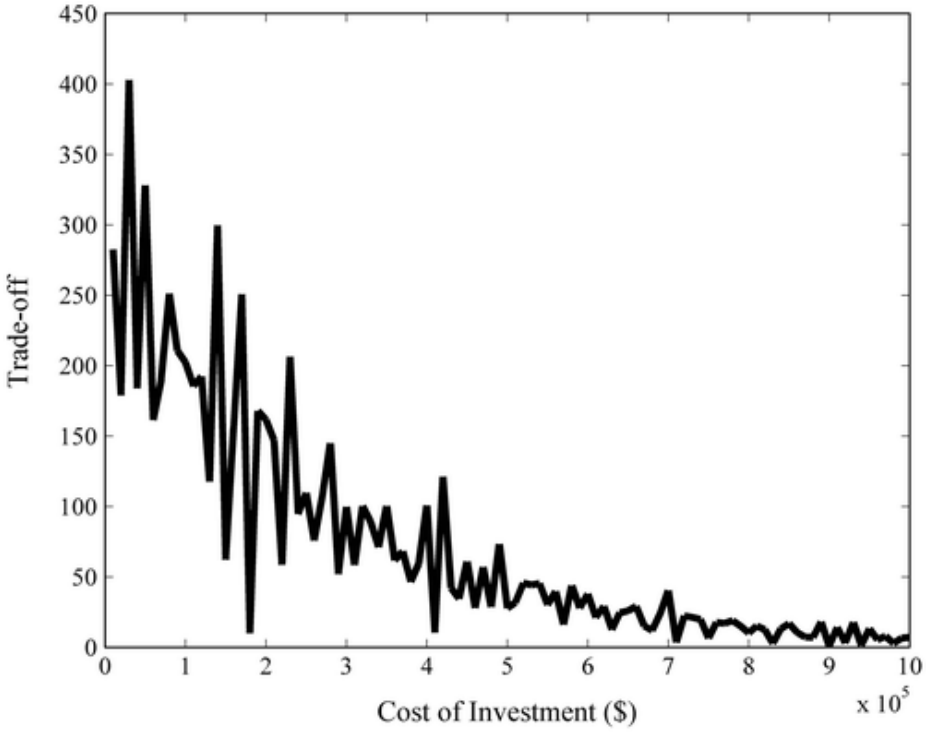


Figure 3-8: Trade-off as a function of the cost of investment for a risk-averse decision maker, $\boldsymbol{r = 0.01}$

Sensitivity Analysis

A number of input parameters needed to be estimated in this methodology, therefore a sensitivity analysis is useful in determining the impact of these parameters on the output of the model. One of the key parameters in this paper is the probability p of a disruptive event occurring, an important issue addressed in the probabilistic risk analysis applications. Determining the probability of an accident, a willful attack or a natural disaster occurring is challenging and has been a subject of research in the past decade.

A sensitivity analysis is performed to determine the effect of the probability of a disruptive event on the expected economic loss as a function of different preparedness investment strategies. Figure 3-9 depicts the mean of the distribution of the expected total economic loss, M, as a function of different preparedness strategies for several possible probabilities of a disruptive event occurring ranging from 0.01 to 0.05. Naturally, the higher the probability, the higher is the expected economic loss under any preparedness strategy. However, the more the decision maker chooses to invest, the less is the effect of the probability on the economic loss. More precisely, for preparedness investments of $600,000 or higher, the average expected total economic loss is almost the same regardless of the estimated value used to express the probability of a disruptive event.

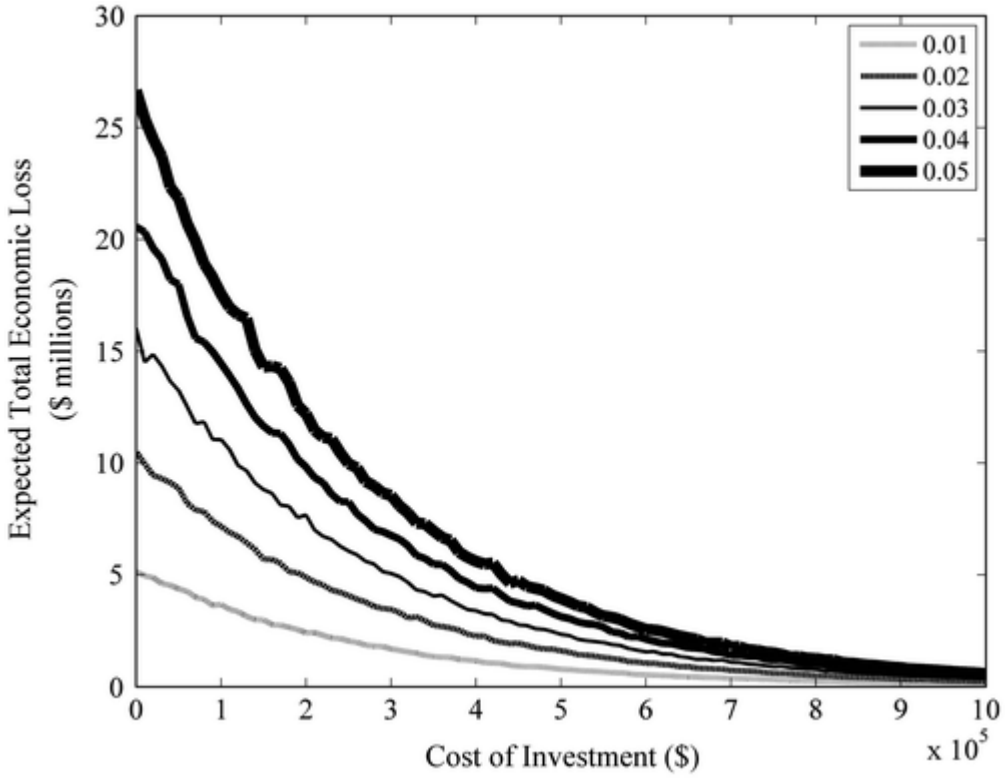


Figure 3-9: Sensitivity analysis on the probability of a disruptive event

Concluding Remarks

Preparedness decision making is an important issue that critical infrastructure systems have been dealing with for the past decade given their vulnerability to disruptive events and their interdependence with other industries. Therefore, it is crucial to consider accurate measures of the risk functions and use the appropriate methodologies to solve such decision problems. This research provides a new framework for analyzing such infrastructure preparedness problems, by integrating two well-known methodologies in risk analysis and decision making, a stochastic decision tree and a dynamic interdependency model to capture the uncertain and widespread economic impacts of disruptive events. With this framework, more comprehensive means are contributed to quantify risk and measure the efficacy of risk management

while accounting for uncertainties in the parameters of the models used in this integrated approach. Particular distributions, namely the power law and beta distributions, are used here, though any appropriate distributions could describe the parameters of the decision problem.

The framework is applied to an inland port preparedness investment problem. Results suggest that while an increase in the investment is providing better protection to the port in terms of the average and conditional expectation of the upper tail of the total economic loss distribution as well as the likelihood of a disruptive event occurring, the decision maker should be aware of the value this increased investment is adding. Sometimes investing an additional dollar in port security might not greatly improve the port security, depending on the risk preference of the decision maker.

The methodologies presented in this chapter aimed at introducing the concept of interdependent impacts of disruptions and their effect on decision making. Interdependent economic losses resulting from disruptive events play a key role in risk management and preparedness strategies. But how do these interdependencies impact the recovery process? And what are the best ways to account for the economic interdependent impacts given the resilience of the disrupted critical infrastructures? These questions are addressed in the following chapter.

Chapter 4

Interdependent Impacts of Infrastructure Systems Resilience Modeling

Recent studies in system resilience have proposed metrics to understand the ability of systems to recover from a disruptive event, often offering a qualitative treatment of resilience. This work provides a quantitative treatment of resilience and focuses specifically on measuring resilience in infrastructure networks. Inherent cost metrics are introduced: *Loss of Service Cost* and *Total Network Restoration Cost*. Further, "costs" of network resilience are often shared across multiple infrastructures and industries that rely upon those networks, particularly when such networks become inoperable in the face of disruptive events. As such, this work integrates the quantitative resilience approach with a model describing the regional, multi-industry impacts of a disruptive event to measure the *Interdependent Impacts of Network Resilience*. The approaches discussed are deployed in a case study of an inland waterway transportation network, the Mississippi River Navigation System.

Several qualitative schema have recently been offered for describing resilience (Haimes, Crowther, & Horowitz, 2008; Haimes, 2009a; Woods, 2006; Bruneau et al., 2003), with a quantitative treatment by Henry and Ramirez-Marquez (2012) describing how system performance is affected by the change of the state of the system in the presence of a disruption and throughout the recovery process. It is this resilience paradigm, developed for deterministic (Henry & Ramirez-Marquez, 2012) and stochastic (Pant, Barker, & Ramirez-Marquez, & Rocco, 2014) analyses that drives the resilience cost metrics in this research.

Figure 4-1 and Figure 4-2 illustrate a general approach for visualizing system performance over time when faced with a disruptive event e^j. The states of the system are depicted across the bottom of Figure 4-1: the original system state S_0 transitions to disrupted state S_d following event e^j, and then to recovered state S_f following a recovery effort. The performance of the system is quantified with the system's service function $\varphi(t)$ (e.g., $\varphi(t)$ could represent commodity flows along a waterway network). In Figure 4-1, larger values of $\varphi(t)$ are preferred (therefore e^j leads to reduced $\varphi(t)$), with the opposite (smaller values of $\varphi(t)$ preferred) depicted in Figure 4-2. The *reliability*, *vulnerability*, *survivability*, and *recoverability* system descriptors, among other details of the state transition, can be found in Baroud, Ramirez-Marquez, Barker, & Rocco, (2014d). Note that a recovered state of the system need not necessarily be the same as the initial state prior to the disruptive event. For instance, the state of infrastructure following the 2010 earthquake in Haiti may be improved over pre-disruption levels, as the recovery activities aiming at helping the infrastructure system regain its functionality might at the same time be helping in improving the system. Further, system performance could fluctuate over time even when no disruption occurs.

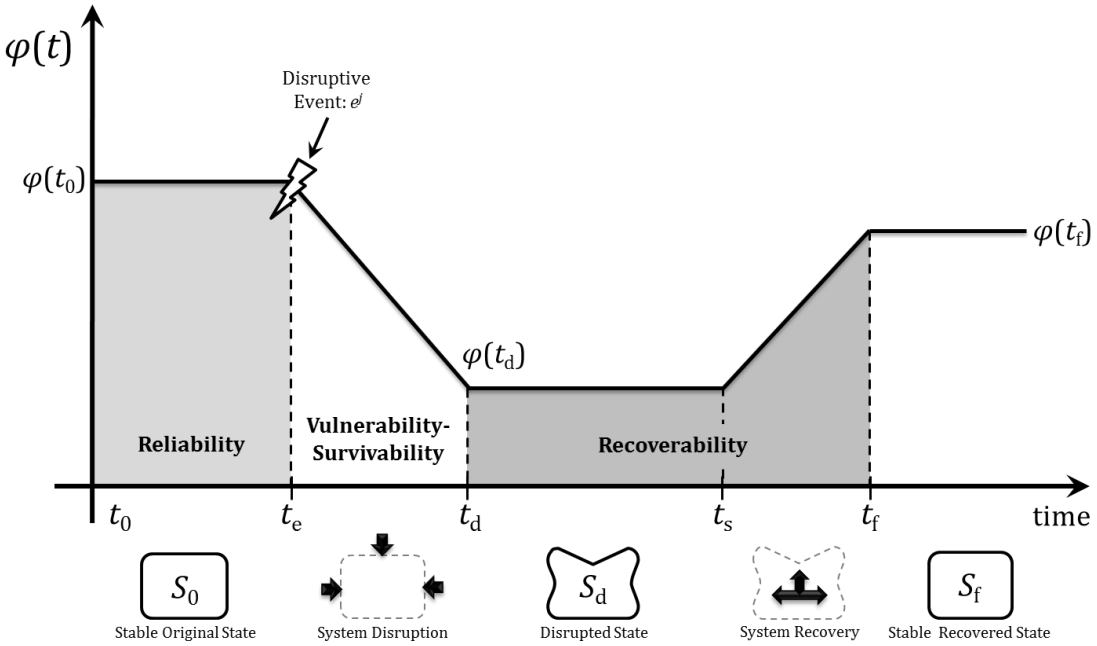


Figure 4-1: Graphical depiction of state transitions over time with respect to an increasing system service function, $\varphi(t)$

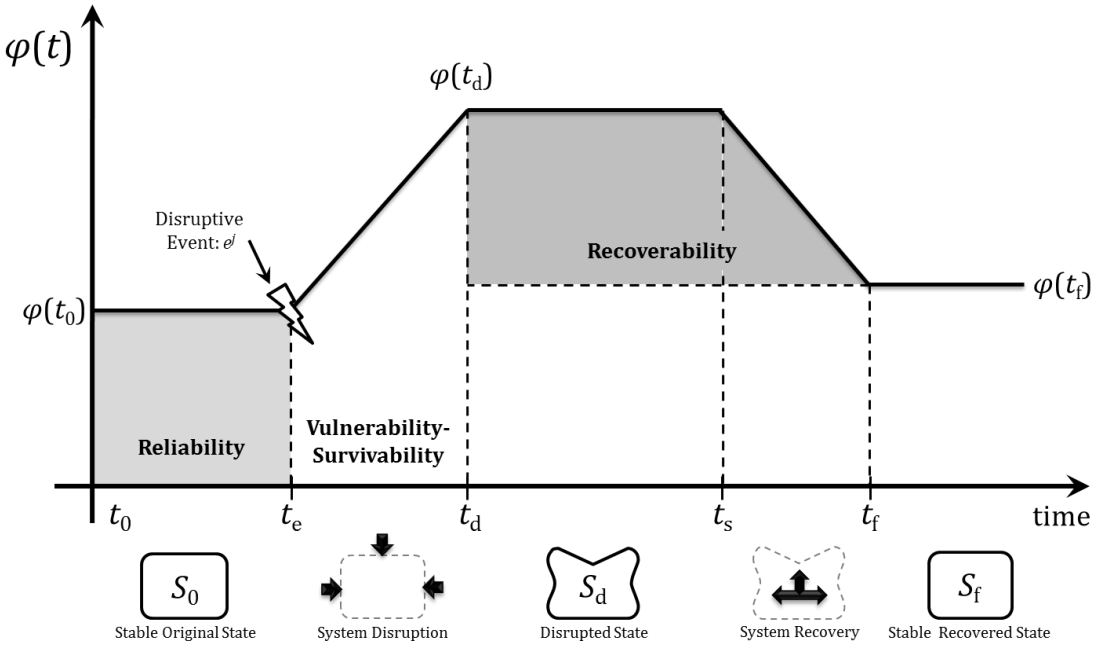


Figure 4-2: Alternative depiction of state transitions describing a decreasing system service function, $\varphi(t)$

Resilience is considered as a time-dependent proportional measure of how the system is performing relative to an as-planned performance level $\varphi(t_0)$, namely in how different the disrupted performance level $\varphi(t_\mathrm{d})$ is from $\varphi(t_0)$. Given that R is historically reserved for quantifying reliability, resilience is given the notation of $Я$ and computed at time t as the ratio of the network performance recovered by time t over the loss of the performance after the disruption occurred, Eq. (4-1). As such, resilience is a function of the extent of loss experienced at time t_d (or *vulnerability*) and the speed at which the system recovers (or *recoverability*) (Henry & Ramirez-Marquez, 2012; Pant et al., 2014; Baroud et al., 2014d).

$$Я(t) = \mathrm{Recovery}(t)/\mathrm{Loss}(t_d) \tag{4-1}$$

A thorough analysis of the recovery dynamics in a network are useful to accurately assess the time (Baroud et al. 2014d) and cost needed for a system to regain its functionality and improve the risk-informed decision making in regards to allocating resources after the disruption. Such costs resulting from a disrupted system are due to inherent costs such as (i) the loss of service, (ii) the cost of restoring the system back to a functional state, and other losses such as (iii) the interdependent effects on industries relying on the disrupted system (Baroud et al., 2013b). For example, if the power grid was disrupted in a certain region, (i) the loss of service could be quantified in terms of energy not supplied (megawatts-hours) to a number of customers, (ii) the cost of restoration involves the work crews and equipment needed for repair, and (iii) one of the interdependent impacts could be the loss of production in industries relying on the power generation in that area (and broader to the interdependent relationships outside of the directly impacted area). Other examples include a disruption in certain river links of

a waterway transportation network impacting the commodity flow throughout the entire system and in particular, disrupting the functionality of neighboring ports and industries relying on the import and export of commodities through the port.

This work, which appeared in Baroud et al. (2013b) and Baroud et al. (2014c), deals with cost and impact metrics aimed at assisting risk managers in accurately identifying and quantifying the multiple costs of a disruptive event in the context of building resilience with particular emphasis given to decision making under uncertainty. The primary contribution of this research is the integration of (i) the resilience modeling paradigm in Figure 4-1 and Figure 4-2 for a disrupted network with (ii) an economic interdependency model to more accurately quantify the resilience trajectory of indirectly impacted sectors and more effectively make network recovery decisions. A literature review section provides the methodological background with respect to stochastic resilience metrics. Inherent cost metrics are then introduced in the methodology section followed by a case study of the application to the Mississippi River Navigation System.

Literature Review

This section details the resilience metrics derived from the depiction of resilience in Figure 4-1 and Figure 4-2. For a review on the interdependency model, please refer to the Literature Review section in Chapter 3. The background on these methodologies will motivate the development of the inherent cost and interdependent impact metrics proposed thereafter. The discussion of resilience that follows will be centered on networks as opposed to general systems.

Assume that disruptive event e^j affects the original network state S_0 at time t_e. The effect on the network is assessed by quantifying the damage to the network service function $\varphi(\cdot)$. For example, if the network under study is an inland waterway network, $\varphi(\cdot)$ could measure commodity flows, noting that lower service function values are considered undesirable. After a period of degradation of length $(t_\mathrm{d} - t_\mathrm{e})$ the network service function is damaged from its original state, S_0 (with corresponding $\varphi(t_0)$), to a disrupted state, S_d (with corresponding $\varphi(t_\mathrm{d})$). That is, the disruptive effect of such an event is quantified via the analysis of a function $\varphi(t)$ describing the behavior of the network as a function of time. After a disrupted state of length $(t_\mathrm{s} - t_\mathrm{d})$, the network restoration commences until it reaches a stable system state, S_f, with corresponding $\varphi(t_\mathrm{f})$. Eq. (4-2) provides a more specific quantification of the value of resilience $Я_\varphi(t_r|e^j)$ evaluated at time $t_r \in (t_\mathrm{d}, t_\mathrm{f})$ (Henry & Ramirez-Marquez, 2012). Set $\mathcal{D}$ is the set of possible disruptive events.

$$Я_\varphi(t_r|e^j) = \frac{[\varphi(t_r|e^j) - \varphi(t_\mathrm{d}|e^j)]}{[\varphi(t_0) - \varphi(t_\mathrm{d}|e^j)]} \quad \forall e^j \in \mathcal{D} \tag{4-2}$$

This model enables quantifying and tracking the changes in the network state as a function of time and accurately observing the network response to the recovery strategies employed. Using this metric, decision makers can dynamically assess their resilience-building decisions during the aftermath of a disruption. It could also be used as a preparedness decision tool, whereby risk managers decide on investments in vulnerability reduction and/or increased recoverability.

The resilience from Eq. (4-2), $Я_{\varphi}(t_r|e^j)$, is operationalized with a set of three metrics describing the time required to achieve different resilience and restoration goals (Ramirez-Marquez & Rocco, 2012; Pant et al., 2014; Baroud et. al 2014d).

First, the metric *Time to Total Network Restoration*, $T_T(e^j)$, records the total time spent from the point when recovery activities commence, at time t_s, up to the time when all recovery activities finalize, $T_T(e^j)$. Since these are stochastic metrics, one can calculate the probability that total system restoration is finished before mission time t as $P_R(t) = P(T_T(e^j) \leq t)$.

The second metric, *Time to Full Network Service Resilience*, $T_{\varphi(t_0)}(e^j)$, records the total time spent from the point when recovery activities commence, at time t_s, up to the exact time, t_f, when network service is completely restored. From $T_{\varphi(t_0)}$, one can define the probability that network service restoration is finished before mission time t_f as $P_F(t_f) = P(T_{\varphi(t_0)}(e^j) \leq t_f)$. Note that $T_T(e^j) \geq T_{\varphi(t_0)}(e^j)$, or the time at which the network is fully restored, is at least as lengthy as the time until a desired network resilience, say $Я_{\varphi}(t_r|e^j) = 1$ (though a different target, either better or worse than $\varphi(t_0)$, may be desired), is achieved. For example, flows along a network can occur with full capacity despite not all arcs being restored: full capacity would suggest full network service resilience without all network components being completely restored.

Finally, the metric *Time to α×100% Resilience*, $T_{\alpha}(e^j)$, records the total time spent from the point when recovery activities commence, at time t_s, up to the exact time, t_{α}, when the system service is restored to $\alpha\varphi(t_0)$. From T_{α}, one can define the probability that network service is restored by α×100%, or $\alpha\varphi(t_0)$, before mission time

t_α as $P_\alpha(t_\alpha) = P\left(T_\alpha\left(e^j\right) \leq t_\alpha\right)$. This metric provides a means to compare different recovery strategies, determining which strategy achieves α×100% resilience the quickest holding resilience constant. Similarly, $Я_\varphi\left(t_r|e^j\right)$ can be found for different strategies, holding t_r constant.

A comparison of the different distributions for the different resilience time metrics can help decision makers choose the best recovery strategy that would lead to the optimal time to full recovery. For a more detailed discussion on these metrics and a stochastic analysis of the time to full recovery in inland waterways, more specifically the Mississippi River Navigation System, please refer to Baroud et al. (2014d).

The work in this chapter utilizes the concept of resilience in Eq. (4-2) to model cost and economic impacts of resilience by developing resilience-based economic interdependent metrics of critical infrastructure disruptions. More specifically, the resilience paradigm of Eq. (4-2) is integrated with a risk-informed interdependency model to quantify the economic impact of a disruptive event on the interdependent industries that rely on the directly impacted and disrupted network.

This work is using the DIIM that has been presented and reviewed in Chapter 3. The interdependency model from Eq. (3-6) uses an $n \times n$ resilience matrix, **K**, representing the capability of a certain sector to recover from the disruptive event and reach a desired performance state. One means to estimate the entries in matrix **K** is to use the formula in Eq. (3-7) (Lian & Haimes, 2006) that assumes a constant recovery rate over time. In this case, **K** is a diagonal matrix with zero non-diagonal entries and k_{ii} as its diagonal entries. An alternative approach to estimate **K** that incorporates the resilience paradigm of Eq. (4-2) is proposed in this chapter.

Methodology: Inherent Cost Metrics

Alluded to previously, two primary inherent costs of network resilience are (i) the cost of lost service and (ii) the cost of restoring the system back to a desired state. Stochastic measures of these two costs are described here.

Loss of Service Cost

When a disruptive event occurs, a loss in the service of the disrupted network is expected, and the change in the service function quantifies the extent of loss. For example, in the case of a road transportation network, the performance measure could be the traffic flow. Different problems might use different performance measures for the same network, with relevant performance measures determined by the risk manager or the decision maker.

A disruptive event is assumed to impact a certain number of components in the network. For road networks, components would be bridges or roads; for inland waterway networks, river links, ports, or dams/locks. Impacted components are assumed to cease functioning for a certain period of time during which the network is in a disrupted state (Figure 4-1). It is assumed that the length of the disrupted state depends on the severity of the event. In Baroud et al. (2014a), the intensity of a disruptive event is assumed to follow a power-law distribution suggesting that more severe events resulting in longer disrupted state periods of time have a lower probability of occurring. For more details on the power-law distribution, please refer to the *Distribution Assumptions* section of Chapter 3.

The distribution used to model the severity of a disruptive event is presented in Eq. (3-8) where d is the number of days the impacted components are not functional

(i.e., the period of time during which the system is in a disrupted state). The loss of service cost is then computed as a function of the severity of the disruptive event. If the performance measure is the commodity flow, then the loss of service is the aggregate commodity that was supposed to flow across the disrupted components for the duration of the disruption. Eq. (4-3) is the performance measure of the network system at the disrupted state.

$$\varphi(t_d) = \varphi(t_0) - \sum_{i=1}^{m} \varphi_i(t_0) \times \frac{d}{365} \tag{4-3}$$

$\varphi_i(t_0)$ is the original performance measure of component i prior to disruption e^j. It is assumed that annual flow $\varphi(t_0)$ across the network is the sum of individual $\varphi_i(t_0)$ such that flows are not counted more than once in the sum. If daily commodity flow are available and if a disruption renders m components completely inoperable for d days, then the disrupted flow across the network, $\varphi(t_d)$, can be measured with Eq. (4-3). Since d follows a probability distribution, simulation techniques (e.g., Monte Carlo) can be used to construct the probability distribution of the loss of service.

Total Network Restoration Cost

Restoring a disrupted network is not only time consuming but costly as a consequence. Careful preparedness strategies and accurate resource allocation should be made to determine the right amount of resources invested at the right time.

To develop a probability distribution for the cost of network restoration, it is assumed that the cost of repairing one component is stochastic. More specifically, $C_i(e^j) = C_i^j$ is introduced to be the cost of repairing link i disrupted by event e^j. $\mathbf{C}(e^j)$ defines the vector of costs for all the links disrupted by the same event. The individual

component's restoration cost probability distribution is described in Eq. (4-4). Note that this distribution could also be a function of the severity of the event, the relationship between the component restoration cost distribution and the severity of the event would be case dependent. One particular example will be explored in the case study.

$$C_i^j = \left\{ C_i^j | P\left(c_s < C_i^j \leq c_r\right) = \int_{c_s}^{c_r} f\left(c_i^j\right) dc_i^j \right\} \tag{4-4}$$

The total network restoration cost would then be the sum of the individual components' restoration costs. However, note that taking the sum assumes that component repair is performed in series. To account for potential parallel recovery activities, a constant factor θ_i is introduced that would depend on the order in which components are repaired. This factor would have higher values in cases where more components are repaired in parallel and would be multiplied by the individual component's cost of restoration, shown in Eq. (4-5). This is similar to the idea of a weighted average, where more weight (in this case higher cost), is given to components repaired in parallel.

$$C_{\text{total}}\left(e^j\right) = \sum_i \theta_i C_i^j \tag{4-5}$$

Similar to the loss of service cost, the total network restoration cost's probability distribution is constructed by means of simulation, such as Monte Carlo simulation.

Ultimately, decision makers would be interested in a metric describing the overall aggregate cost of the entire disruptive event that covers both the loss of service and the cost of restoration. Such a metric can also be computed using simulation of each of the other metrics, provided that they are expressed in the same unit, either dollars, tons of commodity flow, or other units.

Interdependent Impacts of Network Resilience

A disruptive event impacting an infrastructure network does not only have impacts on the network itself but also on the surrounding regional infrastructure systems and industries related to and relying upon it. The costs discussed previously are considered to be inherent costs related directly to the disrupted network, and indirect impacts would be losses and costs incurred by infrastructures and industries related to the disrupted system that were not necessarily directly impacted by the disruptive event.

In order to quantify for those indirect losses, an integration of the resilience paradigm in Eq. (4-2) with the discrete time dynamic interdependency model in Eq. (3-6) is developed. The original interdependency model considered a resilience matrix **K** that is held constant throughout the recovery time. That matrix represented the capability of the economy to restore its functionality without taking into consideration the resilience of the underlying disrupted physical infrastructure. Hence, the **K** matrix in Eq. (3-6) assumed $Я(t) = 1$ for the disrupted sector, and not effectively accounting for the recovery of the underlying physical infrastructure that caused the economic perturbation in the first place.

The approach considers a dynamic version of the resilience matrix that governs the trajectory of interdependent recovery, introducing a matrix whose values are functions of time, $\mathbf{K}(t)$. Further, the resilience matrix is updated with information regarding the trajectory of resilience as a function of time, shown in Eq. (4-6). Matrix **K** is considered to be a baseline matrix of recovery trajectory whose entries are computed according to Eq. (3-7) which updates the resilience matrix at each point in time with the cumulative resilience of the physically disrupted system. It is also assumed that the

perturbation is expressed through a production inoperability, hence, $\mathbf{c}^{\star}(t) = 0, \forall\, t$. The new resilience-based dynamic interdependency model is expressed in Eq. (4-7), where $0 < Я_{\varphi}(t|e^{j}) \leq 1$.

$$\mathbf{K}(t) = \mathbf{K}\, e^{Я_{\varphi}(t|e^{j})} \tag{4-6}$$

$$\mathbf{q}(t+1) = (\mathbf{I} - \mathbf{K}(t))\mathbf{q}(t) + \mathbf{K}(t)[\mathbf{A}^{\star}\mathbf{q}(t)] \tag{4-7}$$

The disrupted system might recover before the rest of the economy does, for which case $Я(t) = 1$ for all $t \geq t_r$, where t_r is the time at which the physically disrupted infrastructure network is recovered. In some other instances, the economy's recovery does not start until the physically disrupted system is fully recovered, which is the case of a port closure for example. In other cases in which components of the system (nodes or links) are disrupted, the recovery of the economy would overlap with the recovery of the physically disrupted system, the relation in Eq. (4-6) takes into account the quantitative analysis of such overlap and how the recovery strategy of the physically disrupted system impacts the recovery trajectory of the entire economy.

Case Study: Interdependent Impacts of Inland Waterways Resilience

The methodology discussed is applied to the Mississippi River Navigation System. This type of transportation system has a special feature that differentiates it from other regular network systems. Generally, there is a single point to point access: there is no redundancy due to the nature of the links being a part of the river.

The nation's economy depends strongly on this waterway as it carries the equivalence of 51 truck trips of commodity circulating through the network each year (ASCE 2013b). The National Waterway Network (NWN) is composed of a large

number of links and nodes. A link represents either a shipping lane or simply a path in open water, and a node could be a facility such as a port, lock, dam, or perhaps another intermodal terminal. The case study analyzes the resilience of a known number of links that might become completely inoperable due to a disruptive event. The US Army Corps of Engineers database was used to construct the 3046 links of the Mississippi River Navigation System network shown in Figure 4-3.

Figure 4-3: Inland waterway network of the Mississippi River Navigation System

The Mississippi River is prone to different types of disruptive events, including periods of drought (Schwartz, 2012) and flooding (Jevis & Bath, 2012). Closing sections of the river impacts the nation's economy by incurring losses to a large number of industries relying on the shipments that are being delayed, with macro-level,

interdependent losses becoming quite significant (MacKenzie et al., 2012a; Pant et al., 2011). As such, resilience planning for inland waterway networks is of high importance.

Parameters and Assumptions

A few assumptions were made in this case study. There is one disruptive event impacting four specific links of the river, highlighted in red in Figure 4-3. Note that the four links represent in fact 10 segments of the river according to the data of the Army Corps of Engineers. Those segments were condensed into four links due to similar commodity flow capacities and proximity of their location. Hence, the event in this case is disrupting operations along a stretch of 141.6 miles of the river located in the area surrounding the port of Catoosa in Oklahoma resulting in delays in the flow of imports and exports to and from Oklahoma. The individual component restoration cost follows a uniform distribution, with a multiplicative relationship with the severity of the event, Eq. (4-8). The severity is expressed in terms of days of disrupted state, d, and follows a power-law distribution, Eq. (4-9).

$$C_i \sim \mathrm{UNI}(0,1) \times d \tag{4-8}$$

$$f(d) = (1.5 \times d_{\min}^{1.5}) d^{-2.5} \tag{4-9}$$

Generally, extreme events that are modeled with a power-law distribution result in an estimated scale parameter λ ranging between 2 and 3 (as discussed in Chapter 3). It is then also assumed that λ = 2.5 in this case. In addition, the severity of the disruptive event is bounded with a minimum and a maximum number of days during which the links are disrupted, between $d_{\min} = 5$ and $d_{\max} = 30$ days. The values chosen for the parameters can be thought of as starting values, and sensitivity analysis should be done

to observe the model's outcome over a specific range for the parameters. For example, λ controls the spread of the probability distribution, with larger values resulting in a larger spread of the likelihood of more severe events. Risk averse decision makers might consider larger values to design preparedness options that account for extreme events expressed by the upper tail of the distribution. Also, since the application pertains to a transportation network measured with the number of days transportation flow capability is reduced, there is a need to specify a threshold for the severity of the event to avoid unreasonable impacts that alter the preparedness decision and resource allocation without a significant tradeoff of protection.

In the interdependency model, industry inoperability is assumed to be the result of imports failing to reach the port and causing a shortage in the material needed to produce in the industries relying on that commodity. As such, inoperability in industry i is then the ratio of its imports to its total production output, as an industry can only be as productive as its most disrupted supplier (MacKenzie et al., 2012a; Pant et al., 2011). Eq. (4-10) illustrates this thought process as the maximum initial inoperability experienced in industry *i*. The yearly estimate of imports from industry *j* is m_j, the total production output for industry *j* is x_j, and the number of industries that typically circulate on the disrupted links is *h*. This maximum perturbation corresponds to a disruption lasting for a year, therefore for a disruption of duration *d* days, the initial inoperability is computed in Eq. (4-11). This approach has been used in previous interdependent impact analyses for disruption of inland waterway ports (MacKenzie et al., 2012a; Baroud et al., 2014a).

$$q_i^{\max} = \max\left(\frac{m_1}{x_1}, \dots, \frac{m_j}{x_j}, \dots, \frac{m_h}{x_h}\right) \tag{4-10}$$

$$q_i(0) = \left(\frac{d}{365}\right) q_i^{\max} \quad (4\text{-}11)$$

Commodity flow data for each link in the Mississippi River Navigation system is provided by the US Army Corps of Engineers, comprised of the yearly tonnage of commodity flow, with the commodity flows of five chosen links provided in Table 4-1. The daily commodity flow is assumed to be the annual flow divided by 365 days.

Table 4-1: Annual commodity flow (in tons) across the five links chosen from the Mississippi River Navigation System

Link ID	Total annual commodity flow
231500	5,007,904
231600	240,925
231709	4,766,979
231810	6,236,462

Three possible recovery activities sets are considered for the four disrupted links with IDs 500, 600, 709, and 810, the first three numbers of the ID are omitted to simplify the notation. The recovery sets are described in Table 4-2.

Table 4-2: Four recovery sets considered for the restoration of the four disrupted waterway links

Recovery set	Description
W1	Repair links in series in the order: 500 – 600 – 709 – 810
W2	Repair link 500 first, then 600 and 709 in parallel in the second order, and 810 in the third order
W3	Repair link 500 first, then links 600, 709, and 810 in parallel in the second order

Resilience and Restoration Time Results

One possible realization for the resilience trajectory over time is first observed. One observation is drawn from a triangular distribution with parameters randomly

selected for each link, and the resilience is computed at each point in time using Eq. (4-2) based on the three strategies in Table 4-2. Note the difference in the time required to achieve full network resilience, portrayed in Figure 4-4. W1 requires approximately 15 additional time units of recovery activities when compared with W2, and almost 35 additional time units when compared with W3. Also, W2 and W3 differ by 20 time units. Clearly, as more links are repaired in parallel, the trajectory towards a fully resilient network tends to be faster leading to a shorter recovery time. Note that the step function is used here to describe the resilience over time, while other types of linear and non-linear functions can be investigated (Zobel, 2014).

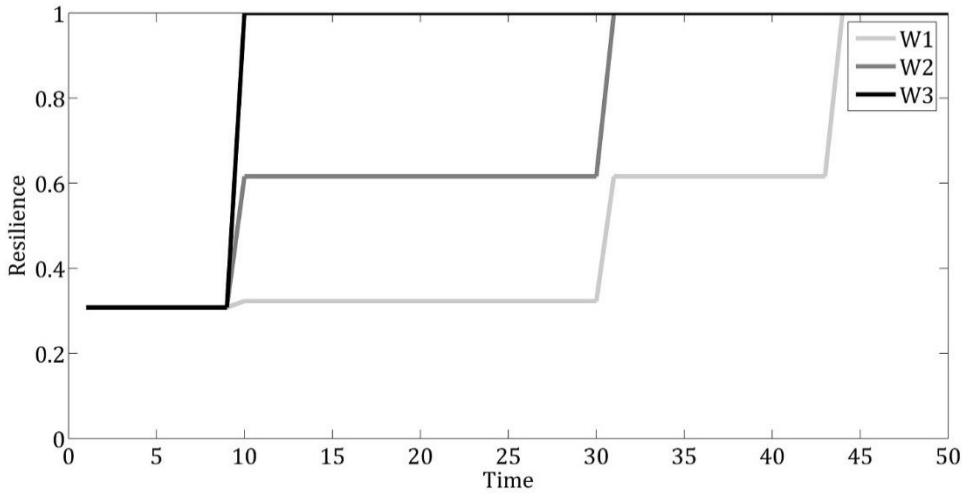


Figure 4-4: Resilience trajectory based on one realization of the distribution of recovery time

To more effectively represent the variability in the underlying model parameters, 2000 scenarios of possible disruptive events were simulated and the cumulative distribution function (cdf) of time to full restoration under the three strategies was constructed. The conclusions align with the observations from Figure 4-4, as more links are repaired in parallel, the overall time to full network restoration

decreases. Figure 4-5 suggests that W3 generally dominates the other two strategies for much of the length of the cdfs. For example, the probability that full network restoration occurs before 20 days is approximately 0.2 for W1, 0.60 for W2, and 0.80 for W3. Also, the average point estimate of the simulated times to full recovery suggests a similar outcome. There is, on average, a decrease of 9.5 time units between W1 and W2, 15.4 between W1 and W3, and 5.9 between W2 and W3. It is clear that going from strategy W1 to strategy W2 has a greater impact, almost double, than going from strategy W2 to strategy W3. To make a better decision regarding which strategy to choose, risk managers look at other objectives, such as costs and interdependent impacts, to determine the trade-off between the strategies.

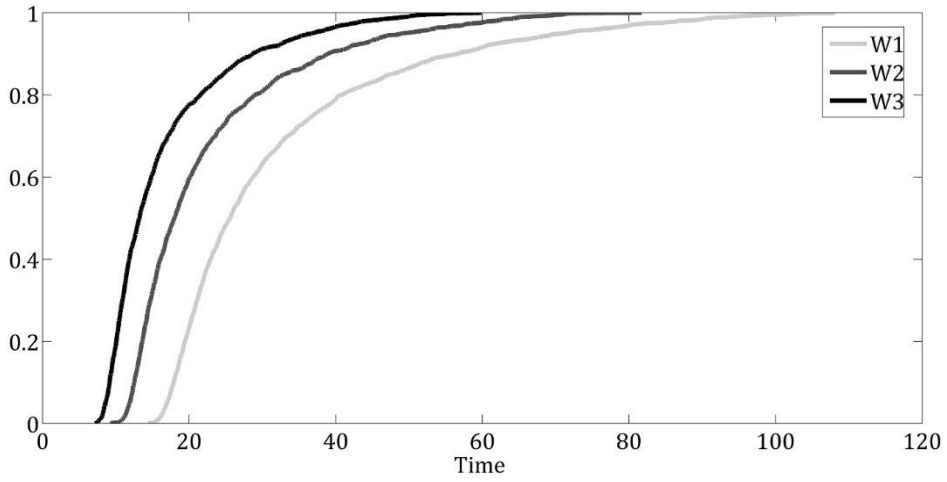


Figure 4-5: Cumulative distribution function for the time to full network restoration

Inherent Cost Results

Mentioned previously, the cost of a disruptive event has several dimensions: the loss of service cost, the cost of network restoration, and the cost incurred by interdependent impacts. Using the 2000 simulations of possible scenarios for disruptive events, the probability distribution function (pdf) and cumulative probability

distribution (cdf) are constructed for the two cost metrics: the loss of service cost in Figure 4-6a and Figure 4-6c, and the total restoration cost in Figure 4-6b and Figure 4-6d for recovery strategy W1. Given the nature of the power-law distribution, smaller values of cost are more likely to occur with extreme values becoming increasingly less likely. Also, note the difference in the units of the cost between the loss of service cost and the network restoration cost: the loss of service cost is measured in terms of tons of commodity flow and the restoration cost is measured in thousands of dollars. To commensurate costs, the commodity flow in tons would be converted into dollar amounts (if available) before adding the two random variables and generating a distribution for the total cost of the disruption.

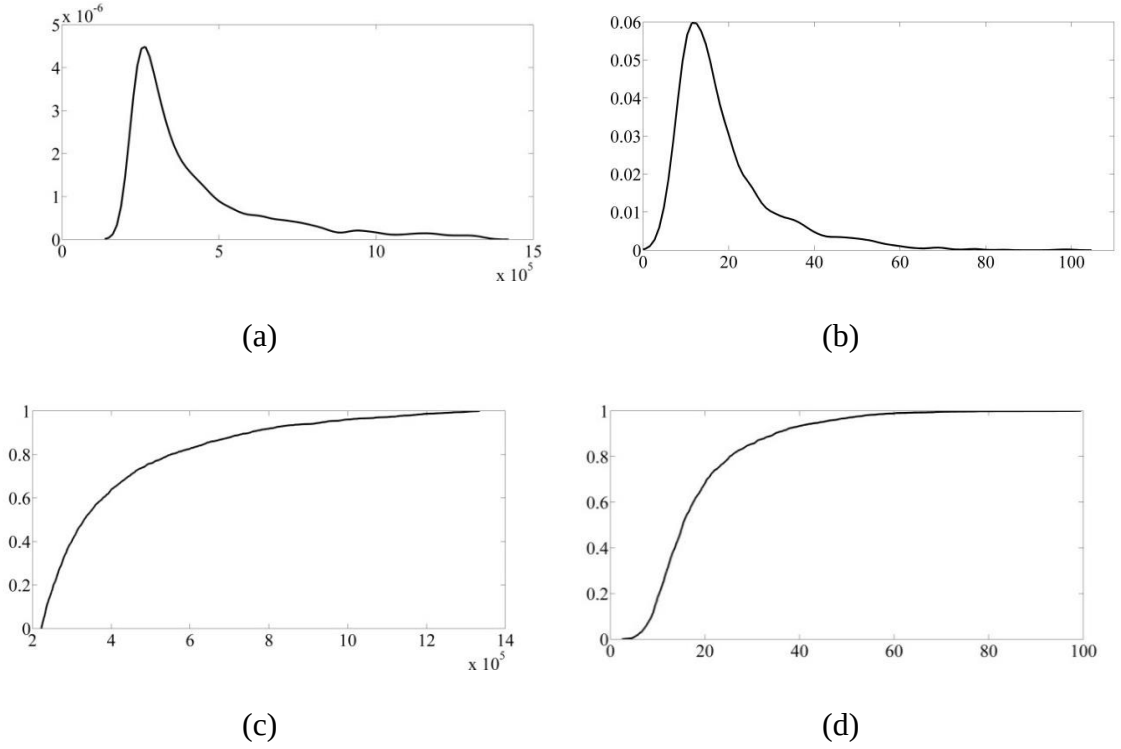


Figure 4-6: Approximate pdf results for 2000 simulations of (a) the loss of service cost and (b) the network restoration cost, along with their respective cdfs in (c) and (d).

The loss of service cost depends solely on the severity of the event and is not impacted by the difference in the recovery strategies. However, from Eq. (4-5), different strategies have different costs that increase as more links are repaired in parallel, which impacts the cost incurred by the recovery activities. The individual cost for repairing each link is multiplied by the number of links that are being repaired in parallel with this particular link. Given the manner in which this is modeled in Eq. (4-5), Figure 4-7 suggests that W3 incurs the highest cost with three out of the four disrupted links being repaired in parallel, while W1 has the lowest cost with all links being repaired in series. Also, note that going from W2 to W3 involves a larger investment than going from W1 to W2. In fact, on average, W2 cost more than W1 by 9.4 cost unit and by 19.2 for W3, this will help decision makers in choosing a recovery strategy by assessing the trade-off between the strategies.

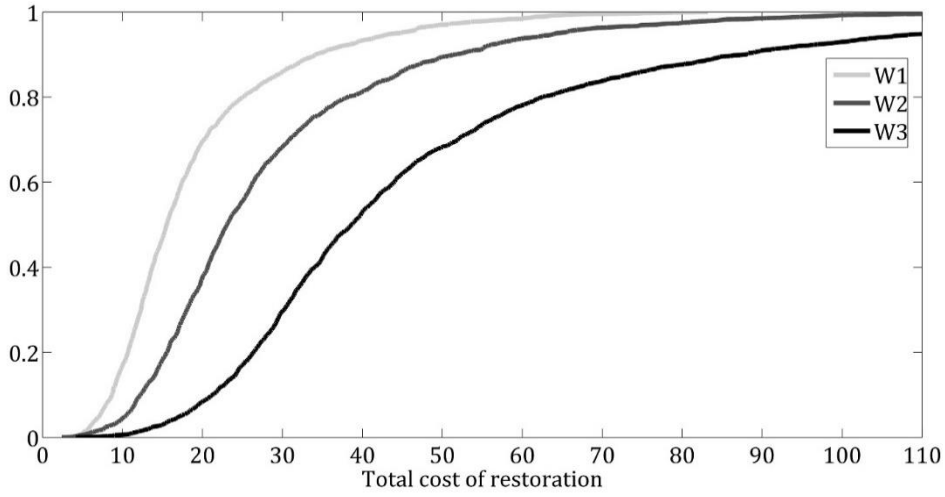


Figure 4-7: Cumulative distribution function for the network restoration cost

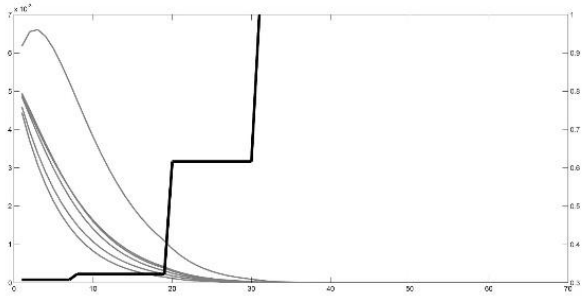
Interdependent Impacts Results

This section examines the interdependent impacts of a disruptive event in the waterway network on the industries relying on the commodities flowing on the network, in light of the resilience quantification and the three strategies considered above.

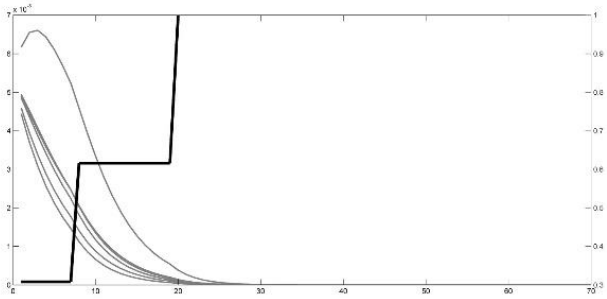
Figure 4-8 depicts the relationship among (i) network resilience, $Я_{\varphi}(t|e_j)$, represented as a dark line, and (ii) inoperability of a selected number of industries, $q_i(t)$, represented with lighter gray lines. The selected industries are those with imports flowing primarily on the disrupted links: food and beverage and tobacco products (FBT), petroleum and coal products (PC), chemical products (CH), nonmetallic mineral products (NMM), primary metal products (PM), and fabricated metal products (FM). Originally considered over 120 days in Figure 4-5, the number of days is reduced along the horizontal axis in Figure 4-8 to better illustrate (i) when $Я_{\varphi}(t|e_j)$ approaches 1, or full network resilience, and (ii) when $q_i(t)$ approaches 0 for the selected industries. Inoperability is depicted on the vertical axis on the left, while resilience, ranging from 0 to 1, appears on the vertical axis on the right.

Naturally, resilience as measured by $\varphi(t)$ is increasing with time, while inoperability is decreasing. For these particular recovery examples, the individual sectors recover faster with strategies W2 and W3 with steeper decreasing trajectories for the individual industries' inoperability. This is due to the parallel recovery activities in strategies W2 and W3 speeding the recovery of the disrupted system and hence resulting in a faster recovery for the rest of the economy. Also, note that for recovery strategies W1 and W2, the full recovery of the economic sectors almost aligns with the full recovery of the disrupted system, while for strategy W3, the disrupted system

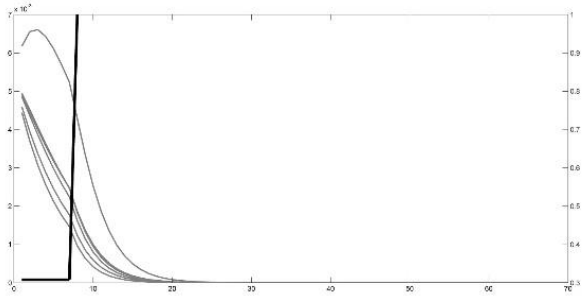
recovered very fast and the economy fully recovers shortly after that. Finally, for one of the sectors, the inoperability continues to increase at the beginning, this is mainly due to the relatively high initial inoperability resulting in an infinitesimal impact of the resilience matrix on the recovery process. Such observation might trigger a need for system hardening or extra resources if decision makers want this sector to start to recover sooner.



(a) W1



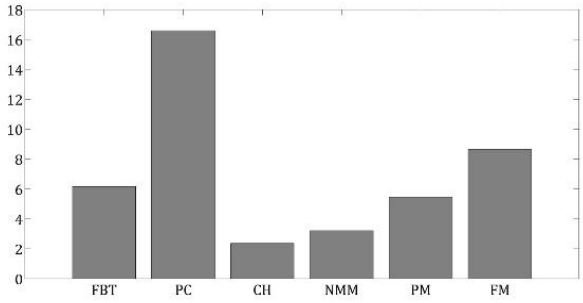
(b) W2



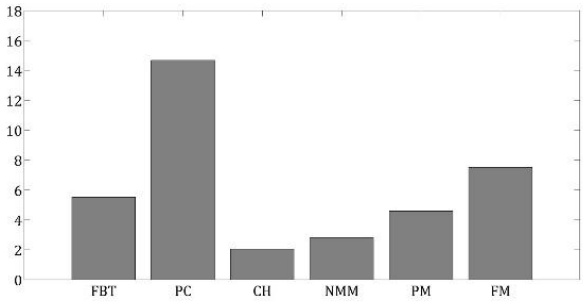
(c) W3

Figure 4-8: Resilience (black line, right vertical axis) and sector inoperabilities (gray lines, left vertical axis) over 50 time periods for the three recovery strategies

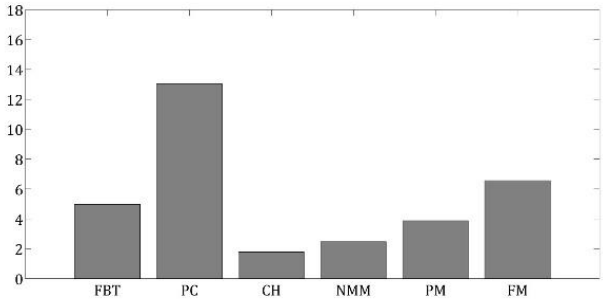
When economic losses are aggregated across time, that is the cumulative effect of inoperability multiplied by the as-planned output of each industry, the effect of recovery strategy can impact industries in different ways. The average economic losses from 2000 simulations are depicted in Figure 4-9 for the three strategies. The same pattern is seen across all three strategies: the Petroleum and Coal Products industry experiences the most economic losses by far, with Fabricated Metal Products next, and the Chemical Products industry affected the least in economic terms. However, the extent to which these industries are impacted does depend on the strategy, with W3 resulting in the fewest losses across industries. Such a breakdown can point a decision maker in the appropriate direction when patterns point to key industries.



(a) W1



(b) W2

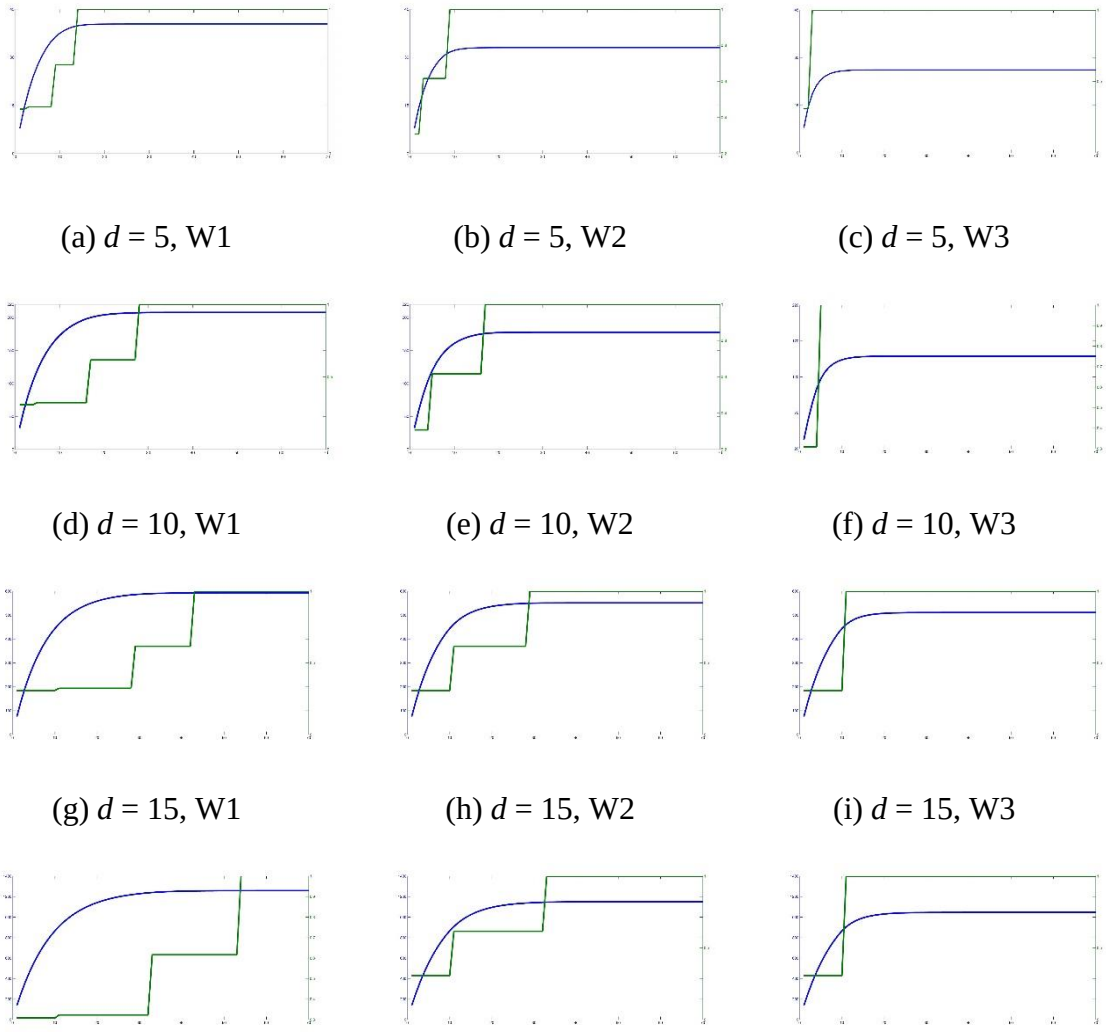


(c) W3

Figure 4-9: Average economic losses experienced in each of the six primary waterway industries for the three network recovery strategies

Figure 4-8 and Figure 4-9 provide average behavior when d, the number of days the waterway links are disrupted, is treated as a random variable. Figure 4-10 focuses on four particular disruption lengths: $d = 5$, $d = 10$, $d = 15$, and $d = 20$. Behavior of economic losses across all industries is depicted with the blue curve, corresponding to the left axis. The trajectory of resilience over time is depicted with the green curve and

the right axis ranging from 0 to 1. Note that resilience reaches 1 in a shorter time as the recovery strategies are switched from W1 to W3 for one particular disruptive scenario, and it takes longer for scenarios with larger impacts. Also, economic losses increase as the disruptive event's severity increases, but they decrease as recovery strategies become faster. Total economic losses range roughly from $45 to $1,200 million for disruptions lasting from 5 to 20 days under strategy W1, for example. The decrease in total economic losses for faster recovery strategies can be observed by examining plots of the same disrupted scenario as the *y*-axis of the total economic loss is the same across different recovery strategies. Hence, observations in the plots below align with the conclusions from Figure 4-8 and Figure 4-9. A faster recovery strategy means that at a certain point in time the resilience of the disrupted system is larger than one of a slower recovery strategy, and this results in more resilient industries being able to recover faster which is portrayed by larger entries in the resilience matrix, according to Eq. (4-6).



(a) d = 5, W1 (b) d = 5, W2 (c) d = 5, W3

(d) d = 10, W1 (e) d = 10, W2 (f) d = 10, W3

(g) d = 15, W1 (h) d = 15, W2 (i) d = 15, W3

(j) d = 20, W1 (k) d = 20, W2 (l) d = 20, W3

Figure 4-10. Economic losses across the six primary waterway industries (blue curve, left vertical axis) and network resilience (green curve, right vertical axis) over time for the three recovery strategies and four disruptive scenarios. Note that the economic loss axis is held constant for the same disruptive scenario

In order to compare the three recovery strategies W1, W2, and W3, interdependent impacts for a particular disruption scenario (d = 10) are further examined. Consider first the total economic loss, an aggregation of the losses incurred across all sectors at each point in time from the disruption through recovery. This function is cumulative, therefore naturally exhibits an increasing pattern, illustrated in

Figure 4-11. Note that the total economic loss is larger under strategy W1, and W3 results in the lowest estimate for the total economic loss. Also, note that the different recovery strategies are impacting the interdependency model through (i) the resilience trajectory and (ii) the time to full recovery. A faster recovery of the disrupted system leads to less total economic losses incurred as the economy can recover faster and more effectively. However, the faster the strategy is, the more costly it is.

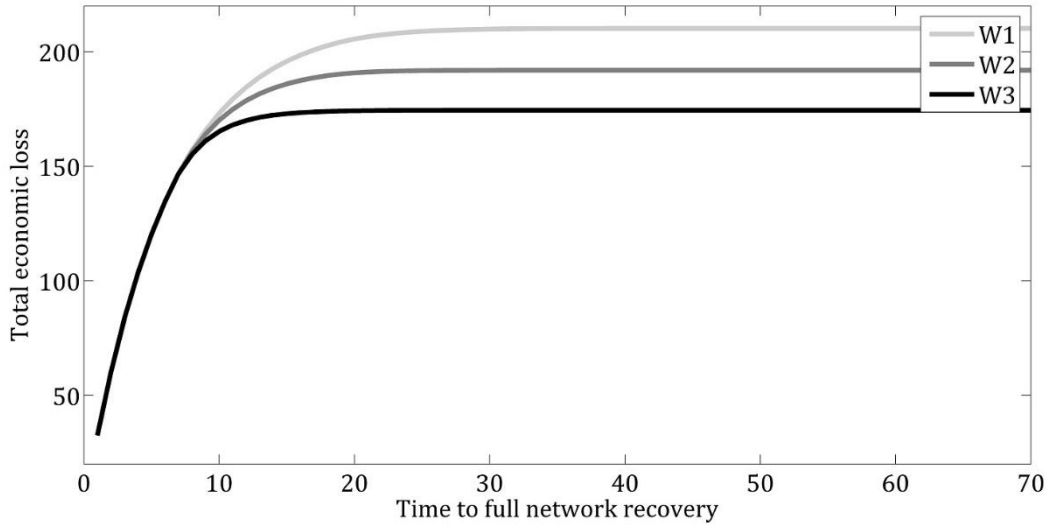


Figure 4-11. Total economic loss computed under three recovery activities Strategies

Rather than a multi-industry economic loss perspective, focus could be given to a particular industry. In particular, the industries whose commodities flow along the disrupted waterway links are considered: food and beverage and tobacco products (FBT), petroleum and coal products (PC), chemical products (CH), nonmetallic mineral products (NMM), primary metal products (PM), and fabricated metal products (FM). Figure 4-12 and Figure 4-13 are plots of the difference in the inoperability of the

individual sectors when considering strategies W1 and W2, and strategies W2 and W3, respectively.

In both cases, the sector of primary metal products is the most impacted by the change of strategies while the rest of the sectors have comparable and a smaller change in the estimated inoperability. Note that all observe the same pattern in the difference in inoperability over the time to full network recovery. This pattern suggests that opting for a faster recovery strategy is not always beneficial. After time $t = 15$, the magnitude of the tradeoff of switching to W2 from W1 starts to decrease, suggesting that this might be a good time for decision makers to switch back to a cheaper strategy. A similar conclusion can be drawn from the comparison between W3 and W2 with an overall much smaller difference in the inoperability between the two strategies.

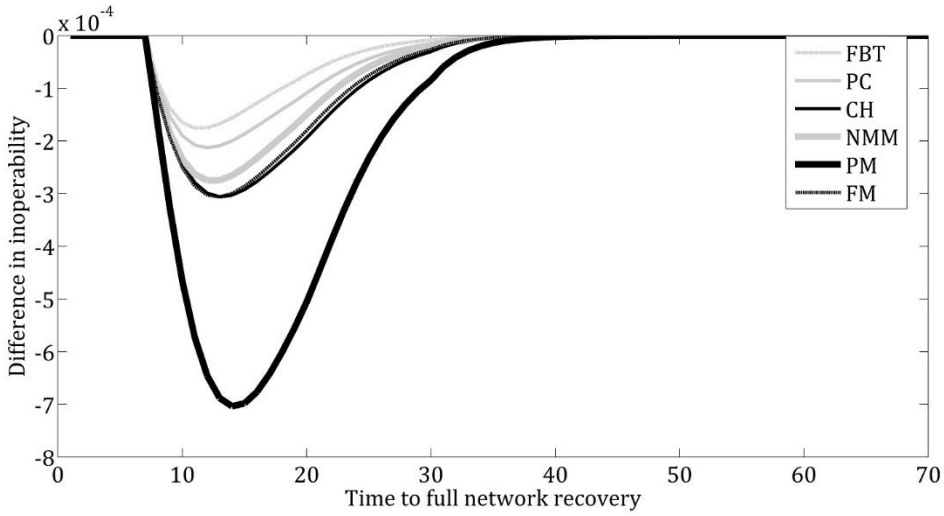


Figure 4-12. Impact on individual sector inoperability for adopting strategy W2 as opposed to W1

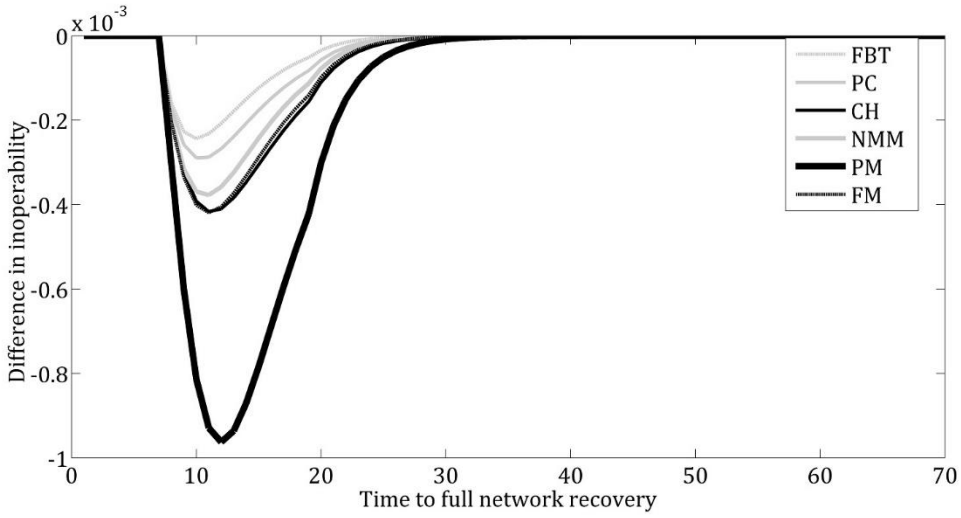


Figure 4-13. Impact on individual sector inoperability for adopting strategy W3 as opposed to W2

Figure 4-14 shows the inoperability trajectory for the sector of primary metal products under the three different recovery strategies. The results shown comply with Figure 4-11, total economic loss under the three recovery strategies. Adopting strategy W1 results in an estimation of a largest inoperability, while W3 results in a much steeper decreasing trajectory for the inoperability of the Primary Metal Products sector.

With such an extensive analysis, decision makers have a range of metrics to consider in order to choose the best recovery strategy that balances cost and risk management. Using the metrics above, decision makers can examine the possible strategies that minimize the cost (W1), or minimize the total economic losses (W3), or a combination of the two based on how a strategy is significantly better than the other over time.

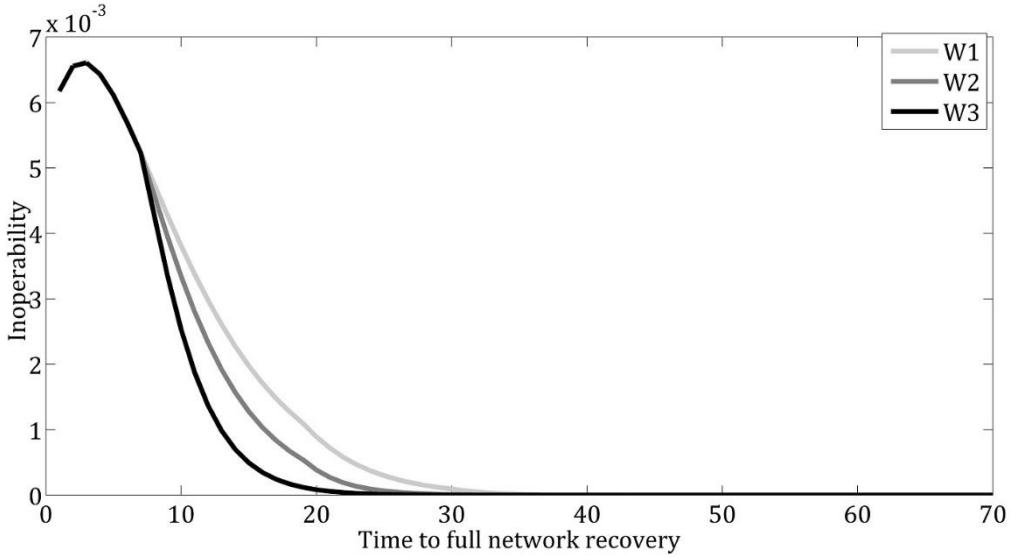


Figure 4-14. Sector inoperability over time for the Primary Metal Products industry for the three recovery strategies

Concluding Remarks

Risk managers preparing for all such sources of disruptive events must plan for the interconnected relationship of infrastructure networks with the industries that rely upon them. Most work in infrastructure networks addresses esoteric graph theoretic measures of topology (e.g., centrality, betweenness) that may provide little insight for decision making (Hines et al., 2010). However, the ultimate usefulness of understanding interdependent effects for a sustained decision making is not just a descriptor of physical damage, but of economic interruption (the result of a lack of functionality) (Tierney, 1997; Webb, Tierney, & Dahlhamer, 2000). That is, the benefit of physical models of interdependence is lost unless they ultimately translate into (i) dollars of losses incurred, and (ii) extent and duration of system inoperability.

This work presents a stochastic approach to compute three metrics of the resilience of an infrastructure network following a disruption: (i) the loss of service

cost, (ii) the total network restoration cost, and (iii) the cost of interdependent impacts. These three metrics extend from prior work in stochastic network resilience (Ramirez-Marquez & Rocco, 2012; Baroud et. al 2014d). The first two metrics are modeled using simulation of probability distributions. The economic impacts of a disruptive event is a well-studied topic (Rose, 2009; Hallegatte, 2008, 2014; Okuyama, 2004; Jiang & Haimes, 2004), as a result, the third metric developed here represents a first step in measuring the broader multi-industry impacts of resilience in infrastructure networks, integrating a network resilience model and an economic interdependency model.

A case study involving the disruption of links on a waterway infrastructure network, the Mississippi River Navigation System, illustrates these concepts, and results demonstrate the importance of considering such measures in risk-informed decision making problems. Incorporating resilience in the interdependency model is helpful to accurately assess the patterns in the cost metrics over time as the system is recovering. Strategies differ in their cost of implementation and interdependent impacts, allowing a decision maker to understand tradeoffs among different objectives. In this particular example, the Petroleum and Coal Products industry was most impacted on average by the disruption, measured by a stochastic duration of commodity flow stoppage, and the interdependent inoperability in the Primary Metal Products industry was most affected by the change in recovery strategy.

Resilience in this chapter has been modeled using simulation techniques and integrated into data-driven techniques, the interdependency model. In addition, the decisions to be made using the analysis above would be based on the overall cost and economic impact of the disruption. As a result, two questions arise. Is it possible to use

data to model resilience metrics using data-driven tools and statistical methods as opposed to simulation of probability distributions? Also, with respect to preparedness and recovery decision making, how can risk managers identify critical components based on their resilience as opposed to analyzing the overall resilience of the disrupted system discussed in this chapter? The following chapter addresses both of these questions to complement the resilience analysis framework of infrastructure systems.

Chapter 5

Bayesian Kernel Methods for Resilience Importance Measures Prediction

When planning for transportation networks such as inland waterways, it is important to understand which components (e.g., locks, dams, waterway links) are most influential on the performance of the entire network and are most influenced by other components in the network. This is a well-studied topic in reliability engineering, with component importance measures (CIMs) quantifying the influence of particular components on the overall structural performance or reliability of the system (Leemis 2009; Kuo & Zhu, 2012). Other explorations of CIMs in a network context include those by Murray-Tuite and Mahmassani (2004), who determine transportation link importance based on the disruption of an optimal traffic assignment network, Jenelius, Petersen, & Mattsson (2006), who provide several vulnerability-based importance measures for transportation networks, and Nagurney and Qiang (2007) and Qiang and Nagurney (2008), who develop a more general flow efficiency metric with which to rank the importance of transportation network components. Natvig, Huseby, and Reistadbakk (2011) suggest that importance measures are helpful in (i) determining which components merit resources to improve overall system performance, and (ii) preparing an efficient component repair checklist in the event of system failure. This work addresses these two items in the context of waterway transportation resilience.

As the quantification of resilience has become a vital component of infrastructure risk analysis, stochastic simulation and the Beta Bayesian kernel model are used to estimate resilience metrics to analyze the recovery process of disrupted

critical infrastructure systems. More specifically, stochastic resilience based component importance measures are assessed using the component's characteristics and disruption data. Such estimates would help risk managers determine the overall best recovery strategy of an infrastructure system in case of a disruption impacting multiple components. The model is deployed in an application to an inland waterway transportation network, the Mississippi River Navigation system, for which the recovery activities sets are analyzed based on the components' resilience importance measures. These measures are estimated using either stochastic simulation techniques or statistical tools such as Bayesian kernel models.

The contributions of this research lie in (i) deploying and validating resilience-based importance measures (Barker, Ramirez-Marquez, & Rocco, 2013) to study the important links of inland waterway networks and (ii) improving the prediction of resilience-based importance measures using data-driven and statistical methods. The importance measures, extended from stochastic measures of the time required for a network to achieve full resilience after a disruption (Pant et al., 2014; Baroud et al., 2014d), along with the cost of recovery activities, are aimed at determining the best recovery set to restore the service in the disrupted links.

The first section of this chapter presents a stochastic analysis of the resilience importance measures that highlights a decision making approach to prioritize recovery for disrupted links of the river (Baroud, Barker, & Ramirez-Marquez, 2014b). The second section provides a data-driven modeling approach to estimate and predict resilience-based importance measures of locks and dams which constitute nodes in the waterway network (Baroud & Barker, 2014). The case study of the Mississippi River

Navigation System is used in each section to illustrate the two approaches in identifying critical components of the waterway network.

Stochastic Analysis of Resilience Importance Measures

Literature Review

Common reliability-based CIMs include (Vasseur & Llory, 1999): (i) Birnbaum importance, or $\partial R_S/\partial R_i$ where R_S and R_i are system and component i reliability, respectively, which describes the probability that component i is critical to the functioning of the system, (ii) reliability achievement worth (RAW), or the maximum proportion increase in system reliability generated by a given component, (iii) risk reduction worth (RRW), an index that quantifies the potential damage to a system caused by a particular component, and (iv) Fussell-Vesely, or an index quantifying the maximum decrement in system reliability caused by a particular component. Several other discussions of importance measures include those by Ramirez-Marquez and Coit (2005), Zio, Marella, and Podofillini (2007), and Rocco and Ramirez-Marquez (2012), among others, and they generally calculate CIMs as a ratio of the measure of component contribution to system reliability and a measure of system reliability itself.

When considering component importance in a resilience setting, one may want to understand the effect that both the disruption magnitude and the recovery speed of components will have on the time to full network service resilience, $T_{\varphi(\mathbf{x}(t_0))}(e^j)$, introduced in the literature review of Chapter 4.

This first resilience-based CIM is illustrated in Eq. (5-1), where V_i^j refers to a component's vulnerability, or its ability to maintain performance after the disruptive event e^j, and $\varphi(t)$ refers to the system's performance measure (for more discussion on

the performance measure, please refer to the literature review in Chapter 4). For example, when $x_i(t)$ describes traffic flow on the ith highway link, $V_i^j = 0.3$ suggests a 30% reduction in flow at the onset of the event. Decreasing performance $\varphi(t)$ occurs until t_d when the new disrupted state is reached, $x_i(t_\mathrm{d}) = \left(1 - V_i^j\right)x_i(t_0)$. A complete reduction in the functionality of the link occurs when $V_i^j = 1$, and $V_i^j = 0$ when the event does not impact the functionality of link i. The numerator in Eq. (5-1) quantifies the network service loss due to the disruption effect on link i, while the denominator describes the maximum loss among all the links. This ratio is then multiplied by the time required to restore the system service to its original state, providing the proportion of restoration time attributed to each link.

$$\mathrm{CI}\text{Я}_{\varphi,i}\left(t_\mathrm{r}|e^j\right) = \frac{\varphi\left(\mathbf{x}(t_0)\right) - \varphi\left(\left(\mathbf{x}(t_0), x_i\left(t_d|V_i^j\right)\right)\right)}{\max_i\left\{\varphi\left(\mathbf{x}(t_0)\right) - \varphi\left(\left(\mathbf{x}(t_0), x_i\left(t_d|V_i^j\right)\right)\right)\right\}} T_{\varphi\left(\mathbf{x}(t_0)|V_i^j\right)} \tag{5-1}$$

As V_i^j and $T_{\varphi\left(\mathbf{x}(t_0)|V_i^j\right)}$ are stochastic terms, a probability distribution can be generated for $\mathrm{CI}\text{Я}_{\varphi,i}\left(t_\mathrm{r}|e^j\right)$ for $t_\mathrm{r} \in [t_\mathrm{s}, t_\mathrm{f}]$. This CIM is comparable to the risk reduction worth (RRW) importance measure (Ramirez-Marquez, Rocco, Gebre, Coit, & Tortorella, 2006)

The second resilience-based CIM addresses the perspective of reliability achievement worth (RAW). That is, Eq. (5-2) defines the "resilience worth" of link , $\mathrm{W}\text{Я}_{\varphi,i}\left(t_\mathrm{r}|e^j\right)$, or an index that quantifies how the time to total network service resilience is improved for event e^j if link *i* is invulnerable (or $V_i^j = 0$).

$$\text{WЯ}_{\varphi,i}\left(t_{\text{r}}|e^{j}\right)=\frac{T_{\varphi\left(\boldsymbol{x}(t_0)|V_i^j\right)}-T_{\varphi\left(\boldsymbol{x}(t_0)|V_i^j=0\right)}}{T_{\varphi\left(\boldsymbol{x}(t_0)|V_i^j\right)}} \tag{5-2}$$

These two resilience-based CIMs are illustrated using general networks in Barker et al. (2013). In this work, these metrics are (i) integrated with a decision making approach, (ii) used in a Beta Bayesian kernel model, and (iii) applied to the study of inland waterway network resilience.

Methodology: Recovery Strategies Decision Process

A disruptive event could impact one or more links, and in the case of the disruption of multiple links, a decision should be made regarding the order in which the links should be repaired. That order is dependent on several factors such as the degree of importance of the links, the time required to achieve full recovery, and the resources available to perform recovery activities, among others. A heuristic approach is presented here to determine the optimal recovery strategy as a function of three metrics representing several factors impacting the overall recovery: (i) the time to network restoration importance, $\text{CIЯ}_{\varphi,i}\left(t_{\text{r}}|e^{j}\right)$, (ii) the resilience worth, $\text{WЯ}_{\varphi,i}\left(t_{\text{r}}|e^{j}\right)$, and (iii) the total cost of recovery.

Recall, $C_i\left(e^{j}\right)=C_i^j$ as the cost of repairing link i disrupted by event e^{j}. $\mathbf{C}(e^{j})$ is then the vector of costs for all the links disrupted by the same event.

For a disruptive event e^{j}, the recoverability strategy, $\mathbf{s}^{j}\left(e^{j}\right)=\left(s_1^j,s_2^j,\ldots,s_m^j\right)$, is a vector of link recovery activities to restore the performance of the system following disruptive event e^{j}. Each element of the recovery activity vector $\mathbf{s}^{j}\left(e^{j}\right)$ is described by (i) the order in which recovery is performed, and (ii) the cost required for recovery to occur. This is represented with a duple, as shown in Eq. (5-3).

$$\begin{aligned}\mathbf{s}(e^j) &= \left(\mathbf{o}(e^j), \mathbf{C}(e^j)\right) \\ &= \left(\left(o_1^j, C_1^j\right), \ldots, \left(o_i^j, C_i^j\right), \ldots, \left(o_m^j, C_m^j\right)\right)\end{aligned} \tag{5-3}$$

To count link i among those links that are disrupted by e^j, $\tilde{V}_i^j$ is introduced in Eq. (5-4).

$$\tilde{V}_i^j = \begin{cases} 1 & \text{if } V_i^j > 0 \\ 0 & \text{otherwise} \end{cases} \tag{5-4}$$

As such, Eqs. (5-5) and (5-6) describe the $\left(o_i^j, C_i^j\right)$ duple in more detail, respectively. The order in which the recovery activity for link i is accomplished is represented by o_i^j and the cost required to complete this activity is C_i^j, and both are respective to the disruptive event, e^j. For example, if all disrupted links are repaired at the same time, $o_i^j = 1, \forall\, i$. Also, C_i^j is a random variable described by its probability density function (pdf), $f\left(c_i^j\left(e_i^j\right)\right)$.

$$o_i^j(e^j) = \left\{ o_i^j | o_i^j = h, h \in Z^+, \sum_i o_i^j = \sum_i \tilde{V}_i^j \right\} \tag{5-5}$$

$$C_i^j(e^j) = \left\{ C_i^j | P\left(C_s < C_i(e^j) \leq C_r\right) = \int_{C_s}^{C_r} f\left(c_i^j\right) dc_i^j \right\} \tag{5-6}$$

If the recovery orders are known and the probability distributions for the components recoveries are given, then a schedule for recovery can be devised. The set $A_h = \left\{ s_i^j | o_i^j = h, \forall i \right\}$ is the collection of all those components having the same order of recovery planning. The recovery planning activity schedule is thus given in Eq. (5-7). Each element in set A_h shows those activities which are planned in parallel (i.e., all

occurring at order h), while the different sets show the series planning of the overall recovery activities, $W^p(e^j)$.

$$W^p(e^j) = \left\{A_1, A_2, \dots, A_l, l \leq \sum_i \tilde{V}_i^{\,j}\right\} \text{ where } p = 1, \dots, P_L \qquad (5\text{-}7)$$

Special cases include scenarios where all the recovery activities are in series, $l = \tilde{V}_i^{\,j}$, or when they are all in parallel, $l = 1$. The number, P_L, of possible recovery sets, W^p, is governed by the different combinations of recovery activities the sets contain. If the number of element sets, A_h, is fixed to L, then the number of possible recovery sets can be represented in Eq. (5-8).

$$P_L = \sum_{n_r=0}^{n-n_1-\cdots-n_L} \dots \sum_{n_2=0}^{n-n_1} \sum_{n_1=0}^{n} \prod_{i=1}^{L} \binom{n - n_1 - \cdots n_i}{n_i} \qquad (5\text{-}8)$$

One way of solving this optimization problem is to consider all the possible, P_L, combinations of recovery sets and choosing the best set with respect to either time to full network resilience (Baroud et. al, 2014d) or with respect to the total cost of recovery activities (Baroud et. al, 2014c). Another option would be to consider a bi-objective optimization problem examining the tradeoffs between the time to full network resilience and recovery cost. The approach considered here is different in a way that it does not need to check on all the possible recovery set combinations. The approach consists of first identifying recovery sets satisfying the priority ranking of the disrupted links according to the resilience-based $\text{CI}\text{Я}_{\varphi,i}(t_\text{r}|e^j)$ and $\text{W}\text{Я}_{\varphi,i}(t_\text{r}|e^j)$, of each disrupted link. Once the recovery sets are identified, the best recovery set is chosen as a function of the total cost required for each set to achieve full network resilience.

Such an optimization problem is of a stochastic order in nature (Dentcheva & Ruszczyński, 2003, 2004), hence a heuristic stochastic ranking approach based on the Copeland Score (CS) method is used to rank the different recovery sets. The CS is a technique used to rank objects characterized by a set of attributes (Al-Sharrah, 2010). The technique assumes that the ranking of the objects could be defined without considering the decision maker's preferences and it is considered a nonparametric approach. The CS is computed based on pairwise comparisons between objects in a set and is defined as the difference between the number of times an object a is better (with respect to attribute q_k) than the other objects and the number of times that object a is worse (with respect to the same attribute q_k) to the other objects. $C_k(a,b)$ provides a value based on a comparison between object a and object b for attribute q_k, $k = 1, \dots, \Omega$, performed according to the rule in Eq. (5-9). Note that a minimum to the objective is desired in this case. Before the first attribute, q_1, $C_0(a,b)$ is initialized at zero, and Eq. (5-9) iterates through all Ω attributes.

$$C_k(a,b) = \begin{cases} C_{k-1}(a,b) - 1 & q_k(a) < q_k(b) \\ C_{k-1}(a,b) + 1 & q_k(a) > q_k(b) \\ C_{k-1}(a,b) & q_k(a) = q_k(b) \end{cases} \quad (5\text{-}9)$$

The method by Al-Sharrah (2010) dictates that the CS of object a is obtained by adding $C_i(a,b)$ over all b, each representing the other objects, as shown in Eq. (5-10). The object with the largest CS value is assumed to stochastically dominate all other objects with respect to the set of attributes. Note that CS assumes that all attributes have the same importance. If this assumption is not valid for a decision maker, then other approaches could be considered such as the ordered weighted averaging (Yager, 1988).

$$\text{CS}(a) = \sum_{b \neq a} C_{\Omega}(a, b) \tag{5-10}$$

Mathematically, the algorithm is performed according to the conditions outlined in the steps in Eq. (5-11).

$$\begin{aligned}
&\text{Step 1} \\
&W^1 = \left\{A_1, A_2, \dots, A_l, l \leq \textstyle\sum_i \tilde{V}_i^j \;\middle|\; CS_{\text{CIЯ}_1} < CS_{\text{CIЯ}_2} < \cdots < CS_{\text{CIЯ}_l}\right\} \\
&\text{Step 2} \\
&W^2 = \left\{A_1, A_2, \dots, A_l, l \leq \textstyle\sum_i \tilde{V}_i^j \;\middle|\; CS_{\text{WЯ}_1} < CS_{\text{WЯ}_2} < \cdots < CS_{\text{WЯ}_l}\right\} \\
&\text{Step 3} \\
&W^* = \left\{A_1, A_2, \dots, A_l, l \leq \textstyle\sum_i \tilde{V}_i^j \;\middle|\; P\left(\underset{\{W^*\}}{TC(e^j)} < c\right) \geq P\left(\underset{\{W^1, W^2\}}{TC(e^j)} < c\right)\right\}
\end{aligned} \tag{5-11}$$

In Eq. (5-11), $CS_{\text{CIЯ}_h}$ represents the Copeland Score for links that have an order h of recovery activity computed based on the $\text{CIЯ}_{\varphi,i}(t_{\text{r}}|e^j)$ of those links. Similarly $CS_{\text{WЯ}_h}$ is the Copeland Score computed according to $\text{WЯ}_{\varphi,i}(t_{\text{r}}|e^j)$. Step 1 selects one recovery activity set satisfying the Copeland Score for $\text{CIЯ}_{\varphi,i}(t_{\text{r}}|e^j)$, meaning that the strategy chosen in this case repairs network components that are most impactful according to the importance of the component from a resilience perspective. In a second step, another recovery set is selected based on the Copeland Score for $\text{WЯ}_{\varphi,i}(t_{\text{r}}|e^j)$. Note that more than one recovery set could be selected in each of steps 1 and 2. Step 3 chooses between the recovery sets selected in the first step and second step, the selection in this step is done with respect to the cost of implementing each recovery set.

Case Study: Stochastic Analysis of Locks and Dams Resilience Importance Measures

This section applies (i) the stochastic time-to-resilience metric, (ii) the resilience-based component importance measures, and (iii) the resulting stochastic

ranking approach to a case study focusing on the Mississippi River Navigation System. More information on this waterway network and its important contribution to the U.S. economy is highlighted in the case studies of Chapter 3 and Chapter 4. The case study of this chapter analyzes the resilience of a known number of links that might go completely or partially inoperable due to a disruptive event. Note that the links in this illustration are randomly selected and are different than the links considered in Chapter 4 which were specifically selected to be near the port of Catoosa in order to analyze regional interdependent impacts.

Each link is considered to be subject to a disruptive event. The impact on each link is dictated by the vulnerability parameter V_i^j which is a random variable following a uniform distribution, $V_i^j \sim$ UNI(0,1). In addition to that, the time each link requires to recover after the disruption also follows a uniform distribution with values going from 0 to 1 time unit, $U_i^j\left(V_i^j\left(e^j\right)\right) \sim$ UNI(0,1). The network performance measure $\varphi(t)$ is the total weight of commodities flowing throughout the Mississippi River Navigation System in a certain period of time (or the sum of the commodity flow for each link over all 3046 links comprising the network). Commodity flow data for each link in the Mississippi River Navigation System is provided by the US Army Corps of Engineers, comprised of the yearly tonnage of commodity flow, with the commodity flow of five chosen links provided in Table 5-1. The IDs of the links are 210, 310, 500, 708, and 800 (the first three numbers are omitted in this case to simplify the notation). The daily commodity flow is assumed to be the annual flow divided by 365. A disruption resulting in a vulnerability of 0.25 for a certain link would decrease the commodity flow for that link by 25%.

Table 5-1: Annual commodity flow (in tons) across the five links chosen from the Mississippi River Navigation System

Link ID	Total annual commodity flow
200210	1,626,645
200310	20,402,014
200500	15,240,458
200708	62,011,198

Recall the linear behavior of $\varphi(t_f)$ between t_e and t_d in Figure 4-1 and Figure 4-2. Another functional form could be considered (e.g., exponentially decreasing), though this case considers the linear form.

One means to identify the importance of each link is to investigate how the vulnerability of each link affects the vulnerability of the entire system. According to the plot in Figure 5-1, links 708 and 800 are the most impactful for the whole system in the event they are disrupted. Link 210 is the least important in terms of adverse effects on the system in a disruption.

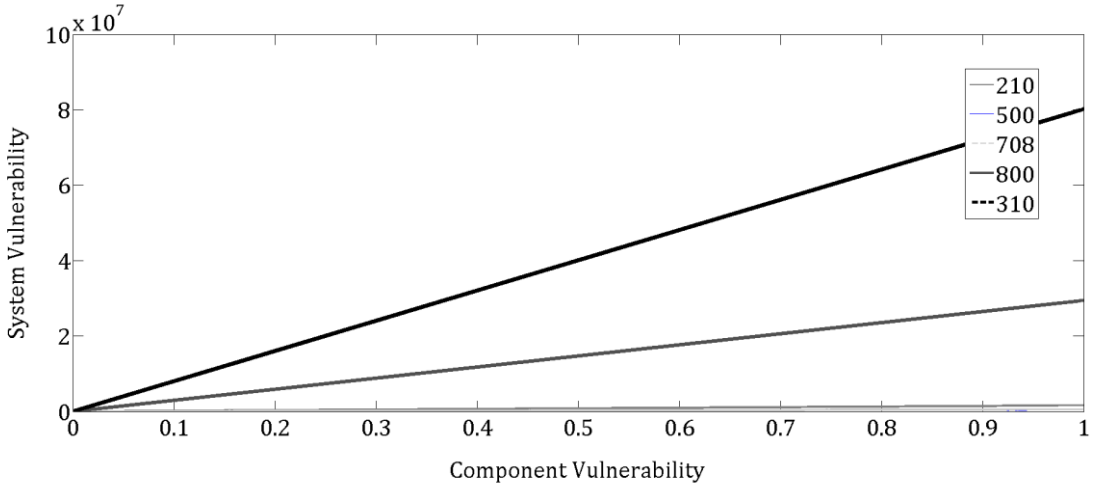


Figure 5-1: Total network-wide vulnerability (in tons) as a function of individual component vulnerability

The first importance measure, $\mathrm{CI}\text{Я}_{\varphi,i}(t_\mathrm{r}|e^j)$, is applied to the five links of the waterway network. Figure 5-2 depicts the cumulative probability distribution of $\mathrm{CI}\text{Я}_{\varphi,i}(t_\mathrm{r}|e^j)$ for each link constructed using 2000 simulations of possible disruptive events. Note that links 310 and 800 are overall dominated by the rest of the links. Links 210, 500, and 708 have a $\mathrm{CI}\text{Я}_{\varphi,i}(t_\mathrm{r}|e^j)$ that is always less than roughly 0.1, while links 310 and 800 could reach much higher measures of resilience. This means that links 310 and 800 are the most impactful to the overall system's resilience when a disruptive event occurs. In this specific example, the cumulative distributions are easily distinguishable, and an initial recovery set can be identified in this case to be $W^1 = \{A_1, A_2, A_3, A_4, A_5\}$ such that $A_1 = \{s_{310}^1\}, A_2 = \{s_{800}^1\}, A_3 = \{s_{210}^1\}, A_4 = \{s_{708}^1\}, A_5 = \{s_{500}^1\}$. This recovery set performs recovery activities with the sequence 310-800-210-708-500.

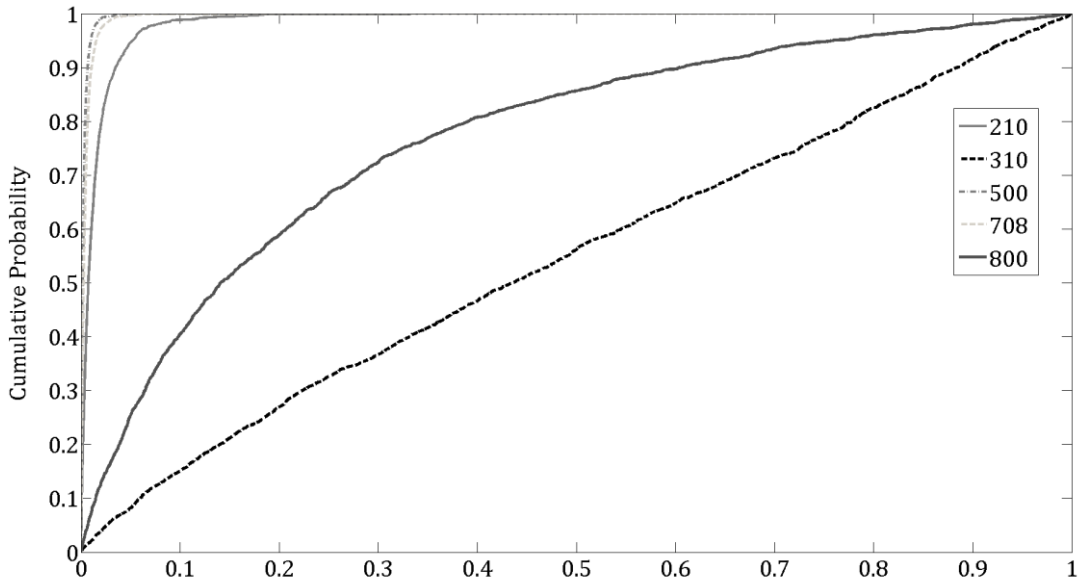


Figure 5-2: Cumulative probability distribution for the resilience-based component importance measure, $\mathbf{CI}\text{Я}_{\boldsymbol{\varphi},\boldsymbol{i}}(\boldsymbol{t}_\mathbf{r}|\boldsymbol{e}^j)$

Although a Copeland Score analysis is not necessarily needed in that case, the Copeland scores are computed and shown for those links as an illustration and validation of the method. According to the histogram in Figure 5-3, the same recovery set is suggested based on the value of the Copeland score for each link, suggesting that link 310 is the most impactful, followed by links 800, 210, 708, and 500. The choice of order priorities in this case is solely based on the value of the Copeland Score, meaning that the link with the highest score is first repaired. However, different decision makers might interpret the scores differently: (i) as 310 and 800 stand out from the rest, they could be repaired in parallel first, then the remaining three each repaired in series, or (ii) 310 and 800 could be repaired in parallel, then 210, 708, and 500 could be repaired in parallel. These two recovery activities sets are defined below as $W^{1\prime}$ and $W^{1\prime\prime}$.

(i) $W^{1\prime} = \{A_1, A_2, A_3, A_4\}$ such that $A_1 = \{s^1_{310,}, s^1_{800}\}, A_2 = \{s^1_{210}\}, A_3 = \{s^1_{708}\}, A_4 = \{s^1_{500}\}, A_4 = \{\emptyset\}$, this recovery set performs recovery activities with the sequence 310-210-708-500 in series with 800 in parallel.

(ii) $W^{1\prime\prime} = \{A_1, A_2, A_3, A_4, A_5\}$ such that $A_1 = \{s^1_{310,}, s^1_{800}\}, A_2 = \{s^1_{210}, s^1_{708}, s^1_{500}\}, A_3 = \{\emptyset\}, A_4 = \{\emptyset\}, A_5 = \{\emptyset\}$, this recovery set repairs links 310 and 800 in parallel at first then repairs links 210, 708, and 500 in parallel.

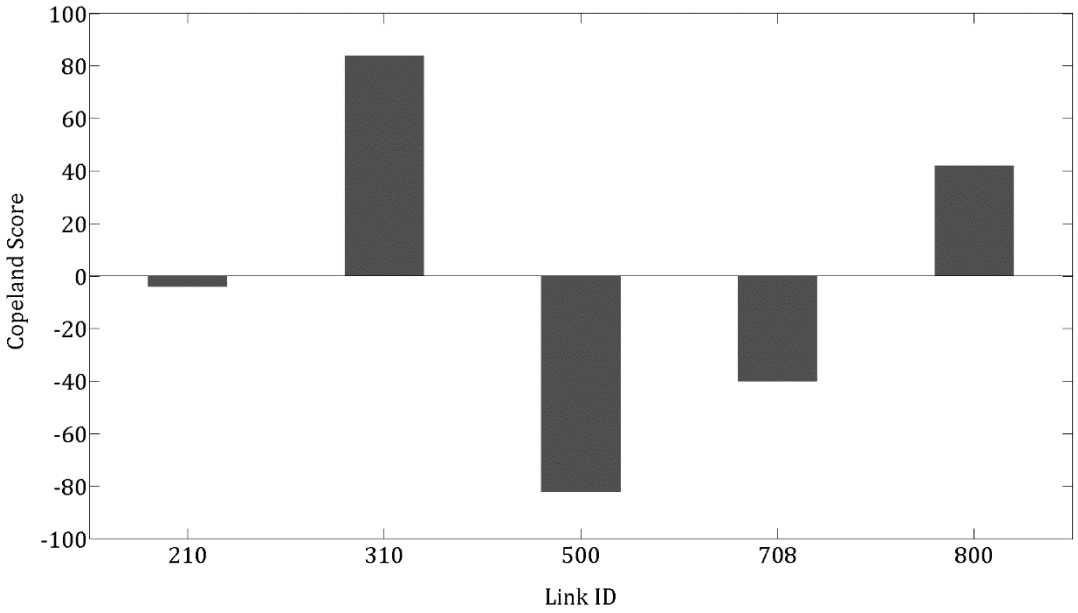


Figure 5-3: Copeland Score for each link computed based on the resilience-based component importance measure, $\mathbf{CI}\mathcal{Я}_{\boldsymbol{\varphi},\boldsymbol{i}}(\boldsymbol{t}_{\mathrm{r}}|\boldsymbol{e}^{\boldsymbol{j}})$

The second importance measure, $\mathrm{W}\mathcal{Я}_{\varphi,i}(t_{\mathrm{r}}|e^{j})$, quantifies how time to full network resilience improves when a particular link is not affected by the disruptive event. Given that the time to full network resilience is stochastic, it is possible to construct the corresponding cumulative probability distribution of the resilience worth of a link using simulation techniques. Figure 5-4 illustrates the cumulative probability distribution for $\mathrm{W}\mathcal{Я}_{\varphi,i}(t_{\mathrm{r}}|e^{j})$ for all five links, expressing the probability that the resilience worth of a link is smaller than a target value. The distributions are constructed using the same simulation technique in Figure 5-3. For example, consider the target value $\mathrm{W}\mathcal{Я}_{\varphi,i}(t_{\mathrm{r}}|e^{j}) = 0.5$ on the horizontal axis. Note that links 708 and 800 have a smaller probability of having a resilience worth less than 0.5, while the rest of the links have a higher probability for this target value suggesting that links 708 and 800 are more impactful in terms of time to full recovery. That is, the time to full network

resilience is much shorter when either link 708 or 800 is not disrupted. Also note that the probability distributions are not easily ranked as in the $\mathrm{CI}\text{Я}_{\varphi,i}(t_\mathrm{r}|e^j)$ case. For example, comparing the distributions of $\mathrm{WI}\text{Я}_{\varphi,i}(t_\mathrm{r}|e^j)$ for links 210 and 310 depends on the target value considered. For resilience worth values smaller than 0.5, link 310 is more impactful, while link 210 is more impactful in cases where the resilience worth is larger than 0.5. The Copeland Score method is deemed useful in distinguishing those differences. The plot for the cumulative probability distribution helps in identifying the first three links to be repaired which are links 800, 708, and 500. The ranking of the last two links will be determined by the corresponding Copeland Score of each link.

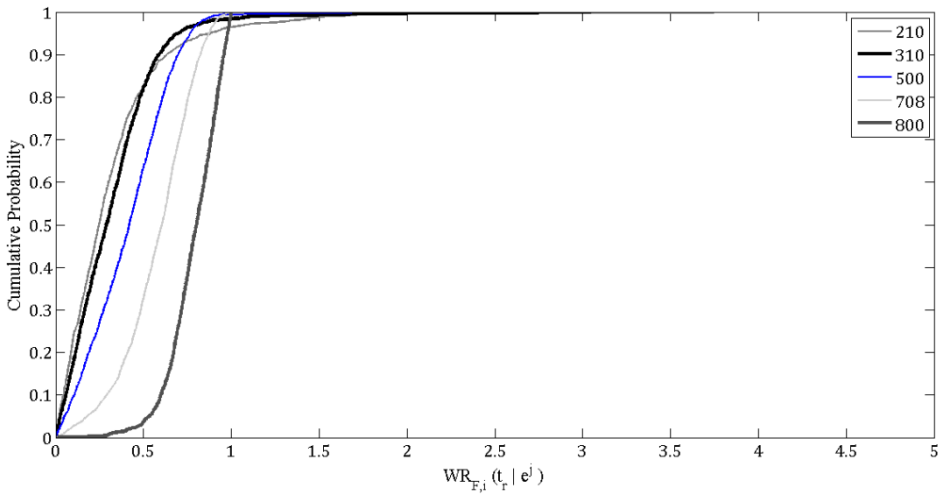


Figure 5-4: Cumulative probability distribution for the component resilience-worth, $\mathbf{WI}\text{Я}_{\boldsymbol{\varphi,i}}(\boldsymbol{t_\mathrm{r}}|\boldsymbol{e^j})$

Figure 5-5 is a histogram of the values of the Copeland scores for each link, as expected from the analysis of the cumulative probability distributions, the links with the highest scores are links 800, 708, and 500 in descending order. The comparison of links 210 and 310 is much clearer with the Copeland Score, where link 310 is deemed more impactful with a higher score than link 210. Another candidate for the optimal recovery

set based on $\mathrm{W}\text{Я}_{\varphi,i}(t_{\mathrm{r}}|e^{j})$ is $W^{2}=\{A_1,A_2,A_3,A_4,A_5\}$ such that $A_1=\{s^1_{800}\}, A_2=\{s^1_{708}\}, A_3=\{s^1_{500}\}, A_4=\{s^1_{310}\}, A_5=\{s^1_{210}\}$, performing series recovery activities with the sequence 800-708-500-310-210. Similarly to the analysis of the $\mathrm{CI}\text{Я}_{\varphi,i}(t_{\mathrm{r}}|e^{j})$, different decision makers might have different perspectives on how to interpret such results and might choose different recovery sets. For the purpose of this research, the value of the score is used to rank the links, and having links repaired in parallel would only be the case where the links have the same value of the Copeland Score.

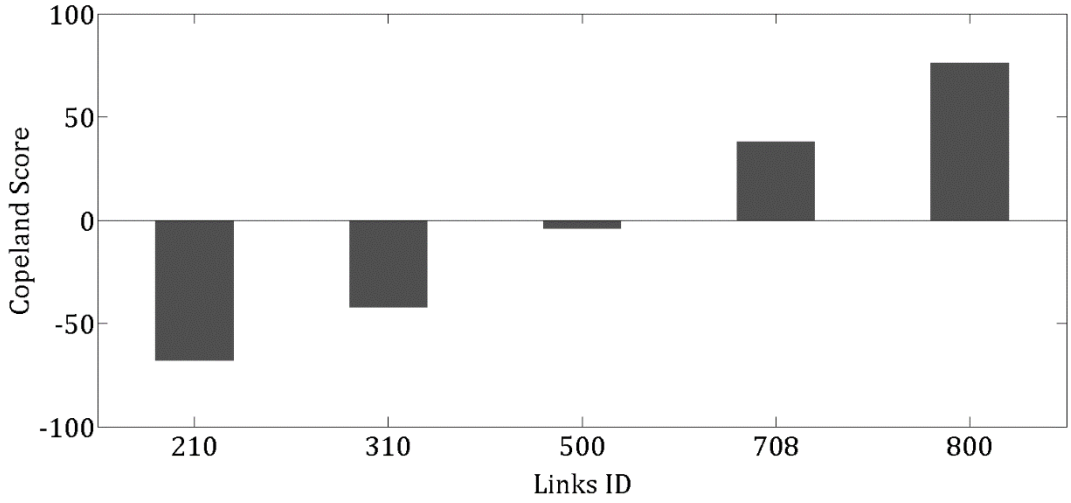


Figure 5-5: Copeland Score for each link computed based on the component resilience-worth, $\mathbf{W}\text{Я}_{\boldsymbol{\varphi},\boldsymbol{i}}(\boldsymbol{t}_{\mathrm{r}}|\boldsymbol{e}^{\boldsymbol{j}})$

Consider only recovery sets W^1 corresponding to 310-800-210-708-500 from the $\mathrm{CI}\text{Я}_{\varphi,i}(t_{\mathrm{r}}|e^{j})$ and W^2 corresponding to 800-708-500-310-210 from the $\mathrm{W}\text{Я}_{\varphi,i}(t_{\mathrm{r}}|e^{j})$. These recovery strategies are compared with the third step of Eq. (5-11) using the cost of implementation. This example is for illustration purposes, and the cost for repairing each link is drawn from a discrete uniform distribution with parameters randomly assigned to each link that range from 1 to 10 time units, larger values were

randomly assigned to lengthier links without having a specific relationship between the parameters of the distribution and the length of the link. Ideally, such a distribution and such parameters would either be determined using expert solicitation or in the case of the availability of historical data, they would be determined using statistical analysis and distribution fitting of the data. For instance, in this particular case, risk managers might want to fit a triangular distribution determining the minimum, maximum and most likely value for the cost of repairing a particular link. The parameter might highly depend on the available resources and workers who perform repair activities, it is then linked to the geographic location of the disrupted links, among other factors. By examining Figure 5-6, it is impossible to choose the best recovery set as the cumulative probability distributions overlap in some instances while in other cases one recovery set outperforms the other. The best recovery set is then determined using the Copeland Score method with a slight variation from Eq. (5-9). Since a smaller value of cost is preferred, a larger Copeland Score value represents the preferred option in Eq. (5-12).

$$C_k(a,b) = \begin{cases} C_{k-1}(a,b) + 1 & q_k(a) < q_k(b) \\ C_{k-1}(a,b) - 1 & q_k(a) > q_k(b) \\ C_{k-1}(a,b) & q_k(a) = q_k(b) \end{cases} \quad (5\text{-}12)$$

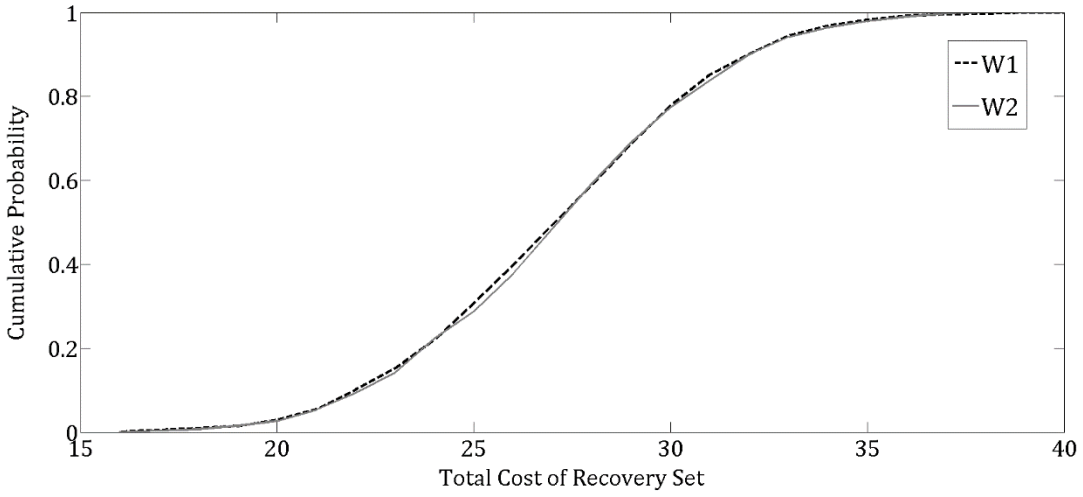


Figure 5-6: Cumulative probability distribution of the total cost (in thousands of dollars) of each recovery set

With a Copeland score equal to 4, recovery set W^1 is chosen to be the optimal recovery set in terms of cost. That is, the recovery set according to $\mathrm{CI}\mathcal{Я}_{\varphi,i}(t_\mathrm{r}|e^j)$ results in a lower cost to implement than $\mathrm{W}\mathcal{Я}_{\varphi,i}(t_\mathrm{r}|e^j)$. Thus, an appropriate balance between link importance and cost is met.

Decision makers might choose to include more options into the last analysis by taking more possibilities for the recovery sets. For example, if more than five links were considered, a mix of the results from the two importance rankings could help guide series and parallel repair to appropriately balance cost considerations. The analysis in this case study has been guided by a simulation of possible values of the resilience-based importance measures that rely on random generations of the time, vulnerability, and cost of the recovery process. A more accurate way of quantifying resilience importance measures is to use statistical methods to draw inference from data.

Data-Driven Prediction of Resilience-Based Importance Measures

There has been a particular emphasis on data-driven methods recently as more decision makers are incorporating statistical analysis of historical data into the decision making process to provide accurate quantification of risk and resilience metrics. Chapter 2 of this dissertation developed a new class of machine learning tools to predict critical infrastructure risk by estimating the frequency of disruptions using historical data, experts' knowledge, and characteristics of the system. The work here is concerned with using data-driven tools to improve the way resilience importance measures are computed. Resilience has often been evaluated qualitatively. In this dissertation, methods were developed to quantify resilience. More specifically, this section is reinforcing the quantification mechanism with data-driven methods and statistical modeling.

This work uses the Beta Bayesian kernel model to assess the resilience importance of critical infrastructure networks. More specifically, the model is used to predict the resilience worth, $\mathrm{W}\text{Я}_{F,i}\left(t_{\mathrm{r}}|e^{j}\right)$, that was deployed in a stochastic analysis in the preceding section. Most of the methodological background has been reviewed in earlier sections and chapters. The literature review of this section will present the Beta Bayesian kernel method followed by a discussion of how the model is used to assess the resilience worth of critical infrastructure systems. The model is deployed in a case study of inland waterways. More specifically, the application illustrates how the model can help analyze the resilience of locks and dams along the Mississippi River Navigation System.

Literature Review

Recall that kernel functions are used to map input data, for which no pattern can be recognized to model their behavior, to a higher dimensional space, where patterns are more readily detected. Such functions enable algorithms designed to detect relationships among data in the higher dimensional space. On the other hand Bayesian methods make use of previous data to estimate posterior probability distributions of the parameter of interest that follows a specific prior distribution. As a result, the integration of Bayesian and kernel methods allows for a classification algorithm which provides probabilistic outcomes (i.e., probability of a data point belonging to a particular class) as opposed to deterministic outcomes (i.e., the mere classification of a data point to a particular class). A detailed overview of Bayesian methods, kernel functions, and Bayesian kernel models is presented in Chapter 2.

The basic form of Bayesian kernel methods assumes a Gaussian distribution. Several extensions have been applied to Bayesian kernel models which assume both Gaussian and non-Gaussian distributions for this classification probability to be estimated. In particular, for the non-Gaussian case, models were developed with a Beta conjugate prior to model binary classification by estimating the probability of a data point belonging to one class (MacKenzie et al. 2014b), while another used a Poisson Bayesian kernel model based on the Gamma conjugate prior to estimate the frequency of disruptive events (Baroud et al., 2013a; Floyd et al., 2014). The Poisson Bayesian kernel model has been developed, tested, and applied to a case study in Chapter 2 of this dissertation.

In this chapter, a Beta Bayesian kernel model is presented and deployed to estimate the resilience of components of a critical infrastructure system. The Beta Bayesian kernel model developed by MacKenzie et. al (2014b) is reviewed here. The other methodological background used in this section has been reviewed earlier in the chapter such as the concept of resilience worth, $\mathrm{W}\text{Я}_{F,i}\left(t_{\mathrm{r}}|e^{j}\right)$ and the Copeland score method to rank probability distributions.

Since the Beta Bayesian kernel model is developed based on a conjugate prior, both the prior and the posterior distributions of the parameter of interest, θ_i, are a Beta distribution with parameters (α, β) and (α^*, β^*), respectively. Eq. (5-13) shows the relationship between prior and posterior parameters.

$$\alpha^* = \alpha + \frac{m_-}{m} \sum_{\{j|y_j=1\}} k(x_i, x_j)$$

$$\beta^* = \beta + \frac{m_+}{m} \sum_{\{j|y_j=-1\}} k(x_i, x_j) \tag{5-13}$$

The kernel function is $k(x_i, x_j)$ and m_+ is the number of positive labels while m_- is the number of negative labels in the training set of size m. The ratios representing the proportions of each class insure an unbiased estimation of the posterior parameters in the presence of imbalanced data sets. Finally, a radial basis kernel function is used, Eq. (2-20), to compute the kernel matrix.

Methodology: Resilience Worth Data-Driven Prediction

The work presented so far in this research computes the resilience worth by assuming that the time to full network resilience is stochastic and follows a particular probability distribution. $\mathrm{W}\text{Я}_{F,i}\left(t_{\mathrm{r}}|e^{j}\right)$ is then computed by means of simulation. This

section provides a similar approach to the non-Gaussian Bayesian kernel models discussed above and applies it to model the resilience worth of the components of a network. The outcome of the model, $\mathrm{W}\text{Я}_{F,i}(t_{\mathrm{r}}|e^{j})$, is a number between 0 and 1, where 0 represents a non-impactful component and 1 represents a highly impactful component. Therefore, a suitable conjugate prior in this case is the Beta distribution for which the range of the random variable is [0,1]. Eq. (5-14) is a representation of the Beta probability distribution with parameters $\alpha > 0$ and $\beta > 0$, where WЯ is the resilience worth described in Eq. (5-2) and $\mathrm{B}(\alpha, \beta)$ is the beta function. Bayesian kernel methods provide a more accurate estimate of the resilience worth as the posterior probability distribution relies on information pertaining to experts' knowledge, the component's characteristics, and historical data of disruptions.

$$P(\mathrm{W}\text{Я}) = \frac{\mathrm{W}\text{Я}^{\alpha-1}(1-W\text{Я})^{\beta-1}}{\mathrm{B}(\alpha,\beta)} \tag{5-14}$$

There are two ways to analyze the outcome of the resilience worth from the Beta Bayesian kernel model. One possibility is to analyze components using a point estimate (e.g., the expected value or condition expected value of the posterior distribution) and examine the resilience worth of a component based on this estimate. The larger the estimate is, the more impactful the component is. Another possibility is to analyze components using the entire probability distribution instead of only the point estimate. This has a great benefit in assessing the resilience worth of components with similar characteristics as information about the entire probability distribution is being used in the analysis. The former way was used in Chapter 2 to analyze the outcome of the Poisson Bayesian kernel model and estimate the frequency of disruption. The predicted frequency in that case was equal to the mean of the posterior distribution, although

other possibilities could have been considered such as the conditional mean in the case of a risk averse decision maker. Conditional expectations were used in Chapter 3 to inform preparedness decision making given the interdependent economic impacts of a critical infrastructure disruption.

The analysis in this chapter will look into the entire probability distribution instead of considering one point estimate (e.g. expected value or conditional expected value). In order to compare probability distributions of the resilience worth the Copeland Score (CS) method is used, which is, in general, a multiattribute ranking technique (Al-Sharrah, 2010). As applied here, the CS method will be used to compare different components according to the distribution of their WЯ importance measure, where the attributes studied here are different percentiles of the distribution. Since the most critical components (largest WЯ) needs to be identified, a maximum $C_k(a,b)$ is desired according to Eq. (5-9). For more discussion on how the Copeland Score is computed, please refer to the literature review of the previous section in this chapter.

In case of a disruptive event impacting several components in the system, determining the components' resilience worth helps decision makers in identifying the best strategy to recover the disrupted critical infrastructure by ordering the components' repairs according to their resilience worth.

Case Study: Bayesian Kernel Modeling of Locks and Dams Resilience Importance Measures

The framework discussed above is applied to analyze the resilience worth of locks and dams on the Mississippi River Navigation System network. The river has 29 locks acting as key connectors between different ports nationwide. The data, retrieved

from the database collected by the U.S. Army Corps of Engineers (2011) contains detailed information on each lock's characteristics including the river mile, the total number of vessels passing by the lock, the total tonnage, the frequency and average delay for the vessels and tows experiencing delay time due to the lock's closure, and the yearly frequency of closure for each lock. A sample of the data is presented in Table 2-4. No prior data are available for the resilience worth, but it is assumed that such data can be elicited from risk managers or government officials. Given the characteristics of each lock and dam, an individual can be asked to classify each lock and dam as either impactful or non-impactful.

Using the Beta Bayesian kernel model and a uniform Beta distribution for the prior, the posterior distribution parameters α^* and β^* are computed and the distribution of the expected value is presented in Figure 5-7. Note that the distribution is dispersed around a range of values going from approximately 0.25 to 0.4. Variability is mainly due to the data set being small. Also, the median of the distribution reflects the actual number of positive classification originally in the data. With such information, risk managers can identify the degree to which the lock and dam is impactful with the probabilistic outcome rather than a simple classification of 0 or 1. This helps in a more accurate allocation of recovery resources.

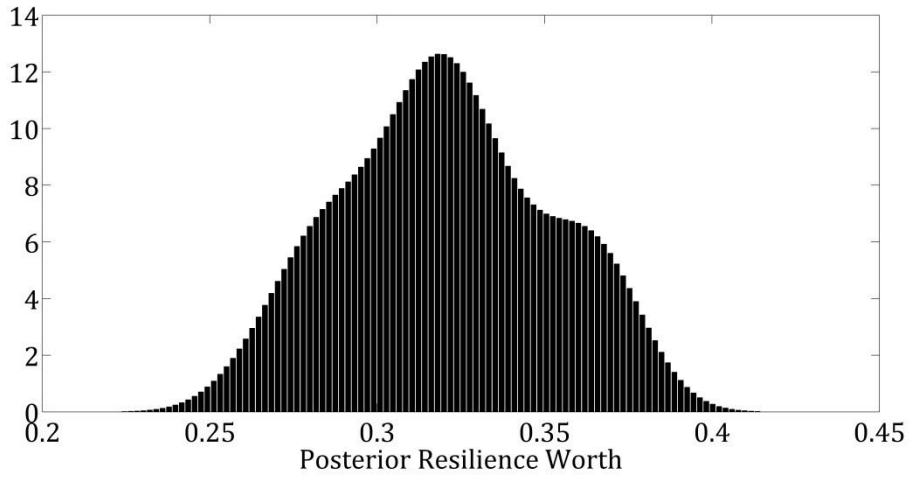


Figure 5-7: Distribution of the posterior expected value of the resilience worth

The five most impactful locks/dams are considered judging by the expected value of resilience worth. Based on the cumulative probability distributions of the resilience worth of these five most impactful locks and dams in Figure 5-8, it is difficult to distinguish their actual ranking of importance, which is the case for locks and dams with similar characteristics. Such cases arise when disruptions occur in a particular region and result in the closure of a number of similar locks and dams. Since it is possible to construct the posterior probability distribution with the Beta Bayesian kernel model, the locks and dams can be ranked according to their Copeland score with approximated percentiles of the resilience worth as attributes (the top five appear in Figure 5-9). Table 5-2 shows the ranking of the locks and dams based on (i) the posterior expected value and (ii) the posterior Copeland Score. Note that each method results in a different ranking, the reason for which is that the Copeland Score represents the entire distribution (lower and upper tails) while the expected value is only a point estimate of the average resilience worth.

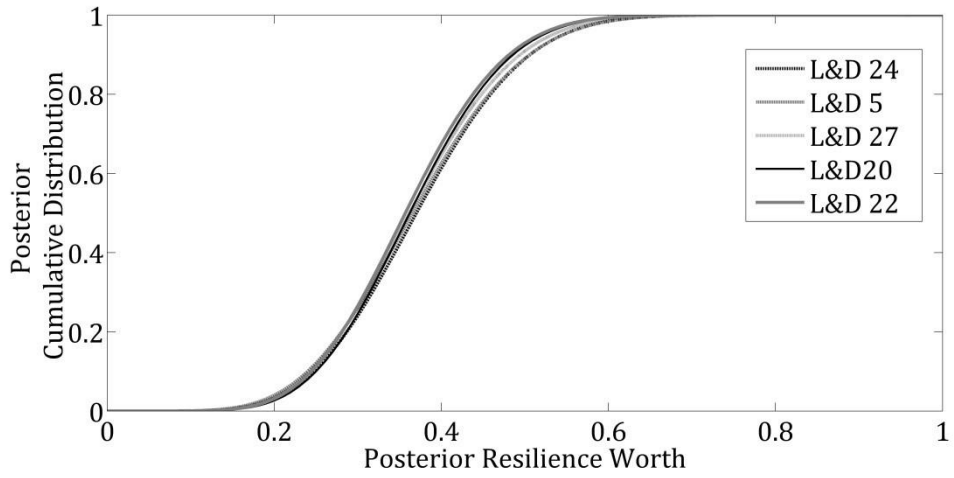


Figure 5-8: Posterior cumulative probability distribution of the five most impactful locks and dams of the navigation system

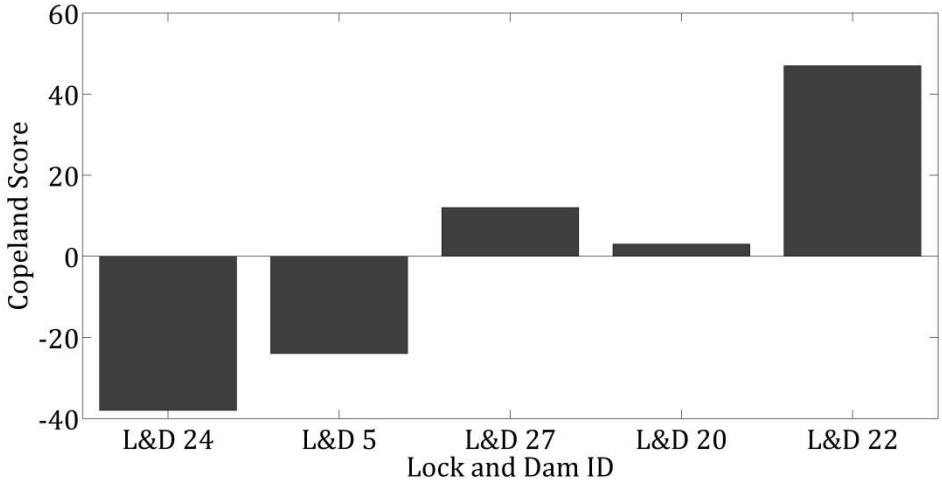


Figure 5-9: Copeland score of the five most impactful locks and dams of the navigation system

Table 5-2: Lock and dam repair order based on the resilience worth values

***WЯ* ranking**	**Posterior expected value**	**Posterior Copeland score**
1	L&D 24	L&D 22
2	L&D 5	L&D 27
3	L&D 27	L&D 20
4	L&D 20	L&D 5

Recall the ranking provided by the Poisson Bayesian kernel method in Table 2-6, it was based on locks and dams that have the highest frequency of closures and could have also been used to inform recovery decision making. The ranking in Table 5-2 is based on the contribution these locks and dams have on the recovery of the overall waterway network in the case of a disruptive event. Note that some locks and dams are common between the two rankings such as L&D 27. There is not one single correct answer to which component is the most critical or needs to be prioritized in resource allocation. Instead, decision makers should account for multiple factors when determining the importance of a link or a node in an infrastructure system. This dissertation has provided several means to approach this problem by considering different criteria to produce a ranking of the most critical components in the inland waterway network.

Concluding Remarks

The ability of a network to "bounce back" from seemingly inevitable disruptive events is a vital consideration. Hence, the ability to quantify the vulnerability and recoverability of inland waterways, as it pertains to the inclusive measure of network resilience, is addressed in this chapter.

A means to quantify vulnerability, or the initial impact experienced in a network following a disruptive event, and recoverability, or the ability of a network to recover functionality in a timely manner is presented. As such the work in this chapter contributes a new way to implement two approaches to measure the importance of network components from the perspective of component contribution to network resilience as a function of stochastic vulnerability and recoverability terms.

The first resilience-based component importance measure, $\mathrm{CI}\text{Я}_{\varphi,i}(t_r|e^j)$ in Eq. (5-1), quantifies the potential adverse impact on system resilience at time t_r when disruption e^j affects link i. Analogous to the risk reduction worth CIM common in the reliability engineering literature, it measures the proportional contribution of link i to the time required to achieve full network service resilience. The second resilience-based component importance measure, $\mathrm{WI}\text{Я}_{\varphi,i}(t_r|e^j)$ in Eq. (5-2), quantifies the potential positive impact on network resilience when vulnerability-strengthening measures are put into place such that link i cannot be disrupted.

The two importance measures discussed in this chapter can serve as a guide to prioritize resilience improvement activities. Using the two measures, the links are ranked with respect to their importance to the overall system resilience, the rank is then used to construct candidates for the optimal recovery strategy that would be later determined based on the cost of implementation of the strategy. Due to the stochastic nature of the elements comprising $\mathrm{CI}\text{Я}_{\varphi,i}(t_r|e^j)$, $\mathrm{WI}\text{Я}_{\varphi,i}(t_r|e^j)$, and the cost, ordering the components according to such measures requires a stochastic ranking technique and the ranking of links and recovery sets is done using the Copeland Score method.

The case study illustrates a particular case of systems, which is the waterway transportation system. Such system is not redundant, and hence, component importance analysis has a large impact on the decision making under uncertainty. As illustrated by the results of the analysis of disruptive events along the Mississippi River Navigation System, some links are deemed more important than other and might require special consideration from risk managers. While the case study in this chapter considers the disruption of five links, the methodology can be used to solve more complex problems

due to the inexpensive computation effort of the Copeland Score and the simplification strategy of the feasible region in the optimization.

Another methodology is presented in this chapter to quantify resilience importance metrics using data-driven and statistical methods. A Beta Bayesian kernel model is applied to analyze the resilience of critical infrastructure networks by estimating the resilience worth of each component in the network using prior information as well as the component's characteristics and historical data. The methodology is applied to an inland waterway transportation network, and the resilience worth of locks and dams is estimated to rank components depending on how impactful they are to the rest of the network. Results show that while the expected value can be used as an estimator, a more accurate metric is the Copeland Score which considers the entire posterior distribution and accounts for more uncertainty in all the possible disruption scenarios. Such analysis assists risk managers and decision makers in allocating resources and determining the ranking order of the repair activities in case of an event resulting in multiple disrupted components.

Based on the ranking of the locks and dams and the analysis of the best recovery set, improvement activities can be suggested to address the prioritized components or recovery sets. Such activities can be in the form of vulnerability reduction policies (i.e., protecting or hardening components) or in the form of accelerating the speed of recovery activities.

Chapter 6

Conclusions

The contributions of this research are focused on three main areas: (i) risk analysis, (ii) interdependent impacts, and (iii) resilience modeling. The tools developed and deployed in this research are used to predict risk, analyze its consequences, and suggest recovery strategies.

This chapter includes an overview of the key findings in this dissertation and a discussion of the future direction of this research.

Insights and Lessons

Statistical Modeling for Risk Analysis

The theoretical contribution of this research is made to the statistics and machine learning literature through the Poisson Bayesian kernel model. It is the first Bayesian kernel method that can be applied to count data. The model has been tested with sample count data and found to outperform traditional count data modeling approaches such as the Poisson and the Negative Binomial GLM in the prediction accuracy. The majority of the data sets (five out of seven sets) in the empirical study had better predictive accuracy for the proposed model across the four different metrics considered. The Poisson Bayesian kernel model appears to be a good model for prediction purposes when the data set is small with a small number of predictors, which is common among risk analysis problems.

The Poisson Bayesian kernel model has been applied to predict the frequency of disruptions in critical infrastructure systems, which is considered to be a methodological contribution to the risk analysis field in which data-driven tools have received little

attention due to the limited availability of data. A case study to an inland waterway network, the Mississippi River Navigation System, illustrates the advantages of using the Poisson Bayesian kernel model as opposed to GLM. The model proposed in this research provides more accurate predictions for the number of closures of locks and dams along the waterway. In addition, the Poisson Bayesian kernel is a flexible tool where decision makers are able to input their knowledge and expertise into the prior distribution to guide the inference process. Also, the decision makers' risk preference can be taken into account in the analysis. The case study assumed a risk neutral decision maker and considered the expected value of the posterior distribution as a point estimate of the frequency of closures. However, risk averse decision makers will be interested in either a different point in the upper tail of the posterior distribution expressed through a conditional expected value, or the variability in the prediction.

Overall, the Poisson Bayesian kernel model is a tool that integrates several sources of information (such as the decision maker's expertise, prior information, historical data, and predictors) to provide a more comprehensive quantification and analysis of risk.

Economic Impacts of Disruptions

The first part of this dissertation answered questions related to the likelihood of a disruptive event: *What can go wrong? What are the chances of something going wrong?* The second part is concerned with the impact of a disruptive event: *What are the consequences if the undesirable event occurs?* (Kaplan & Garrick, 1981).

Analyzing the impacts of a disruption in critical infrastructure systems goes beyond assessing the losses pertaining to disrupted system. As discussed in the

introduction and case studies throughout this dissertation, critical infrastructure systems are highly interdependent. As a result, accounting for the impact of a disruption is concerned with a number of factors at different levels. Chapter 3 and Chapter 4 developed tools that integrate the inoperability interdependency model with decision analysis and resilience modeling techniques aimed at (i) quantifying the interdependent economic impacts of a disruption, (ii) assessing the efficacy of risk management strategies, and (iii) analyzing the impact of infrastructure resilience on the economic losses.

The deployment of these methods in the case study of the Mississippi River Navigation System revealed insights in inland waterway preparedness strategies and interdependent impacts modeling. Results from the Stochastic Multiobjective Inoperability Decision Tree suggest that while an increase in the investment is providing better protection to the port in terms of the average and conditional expectation of the total economic losses distribution as well as the likelihood of a disruptive event occurring, the decision maker should be aware of the value this increased investment is adding. Sometimes investing an additional dollar in port security might not greatly improve the port security, depending on the risk preference of the decision maker. In addition, the interdependent analysis provided commodity specific resilience insights. In the particular case of the Mississippi River Navigation System, the Petroleum and Coal Products industry was most impacted on average by disruptions, measured by a stochastic duration of commodity flow stoppage, and the interdependent inoperability in the Primary Metal Products industry was most affected by the change in recovery strategy.

The tools developed in Chapter 3 and Chapter 4 allow for a general analysis of the interdependent impacts and risk management strategies assessment of the overall inland waterway system, while also considering the impact of industry-specific resilience of the disrupted and interdependent systems. Such methods can inform decision making at a higher systems level as well as provide specific insights about the different components of the system and any other interdependent systems for a more targeted resource allocation effort.

Resilience Modeling

The third part of this dissertation is concerned with resilience modeling. Resilience is identified by four dimensions: reliability, vulnerability, survivability, and recoverability. It has been modeled from different perspectives such as analyzing the time to full network recovery, assessing the system's performance measure, and modeling the cost of recovery activities, among others.

This research deploys resilience-based importance measures developed by Barker et al. (2013) to assess the resilience importance of inland waterways. More specifically, the stochastic analysis gives insights on the system's resilience contributed by certain links of the network. The links that are deemed more important than others would require special consideration from risk managers.

In order to analyze the resilience-based importance of locks and dams which are considered to be nodes in the waterway network, a data-driven approach is used to rank the components based on their resilience worth. The method utilizes prior belief in the form of experts' knowledge integrated with characteristics of the network component to produce a probability distribution expressing the magnitude of the resilience worth.

Risk managers can utilize the outcome of this research to inform their decisions by either relying on a particular ranking of components or recovery sets, or by integrating the different results for a more comprehensive analysis of the overall system resilience.

Future Research

A number of research extensions can be investigated to expand on the ideas developed in this dissertation. This section provides a discussion on the future direction of the research in each of the three parts of the dissertation.

Bayesian Kernel Methods

Modeling extensions can be incorporated into the Poisson Bayesian kernel method to make it more flexible and generalized.

Recall the discussion on the prior parameters estimation in the case study of Chapter 2. Results show that the selection of the priors has an impact on the posterior parameters estimation and prediction accuracy. One way to overcome the misspecification of the priors is to assign a hyperprior distribution to account for the uncertainty in estimating the prior parameters. This class of methods is called Hierarchical Bayesian models which have been used in different applications but were never integrated with Bayesian kernel models.

The analysis performed for the Poisson Bayesian kernel model did not account for the impact of covariates on the performance metrics. To maintain consistency all covariates were included in the models for the empirical analysis. However, for the case study, reduced models were considered to compare the Poisson Bayesian kernel model to the best version of the GLM. A next step would be to explore model selection criteria

and techniques for Bayesian kernel methods to investigate the possibility of improving the model's performance as a function of the predictors.

Interdependency Modeling

The uncertainty modeling of the stochastic multiobjective decision tree developed in Chapter 3 considered two sources of uncertainty: (i) the uncertainty associated with the likelihood of a disruption occurring, and (ii) the uncertainty associated with the severity of a disruption and its impact on the demand perturbations. The two uncertainties considered are not related to the interconnectedness of the economy represented by the $\mathbf{A}^{\star}$ matrix. Therefore, although the interdependent impact between industries involves uncertainties, this is not analyzed in the study. An integration of the suggested approach with models that consider not only the uncertainty of the inoperability but also the uncertainty of the $\mathbf{A}^{\star}$ matrix (Barker & Rocco, 2011; Oliva, Panzieri, & Setola, 2011) constitutes the subject of future research.

Another extension to this research would be to consider multi-period decision trees and incorporate resource allocation in addition to preparedness decisions. Also, using a multi-regional dynamic interdependency model is useful in indicating the regional expected total economic losses.

The case study in Chapter 4 analyzed three possible recovery strategies. A more exhaustive set of strategies could be considered and compared with a stochastic ordering technique (Copeland score or optimization). Future work would include the resilience-based analysis of a more extensive set of disruptive scenarios and recovery strategies, as well as the exploration of network examples where disruptive events are not as localized as inland waterways.

Resilience Modeling

Note that the approach considered in the analysis of resilience importance measures in Chapter 5 does not take into account the cascading effect of a disruption in the system, this might be a concern in applications that are not related to waterway network systems, such as in power grids or other transportation networks. In that case, the methodology described here could be extended to a dynamic version whereby the measures are updated with newly disrupted links due to the cascading effect at each point in time.

While the resilience importance measures are key factors in determining the recovery strategy, it is equally important to account for the overall cost and time of implementing the strategy. Future research is involved in determining the optimal recovery strategy by taking into account the tradeoff of the Bayesian kernel estimates of the component importance and the time and cost of recovery.

Bibliography

Agresti, A. (2002). *Categorical Data Analysis*, 2nd edition. Hoboken, NJ: Wiley-Interscience.

Agresti, A. & Finlay, B. (2008). *Statistical Methods for the Social Sciences*, 4th edition. Prentice Hall.

Aizerman, M., Braverman, E., & Rozonoer, L. (1964). Theoretical foundations of the potential function method in pattern recognition learning. *Automation and Remote Control, 25*, 821–837.

Akaike, H. (1970). Statistical predictor identification. *Annals of the Institute of Statistical Mathematics*, *22*(1), 203-217.

Al-Sharrah, G. (2010). Ranking using the Copeland Score: a comparison with the Hasse diagram. *Journal of chemical information and modeling*, *50*(5), 785-791.

American Society of Civil Engineers. (2009). Report Card on America's Infrastructure 2009.

American Society of Civil Engineers. (2013a). Testimony of Andrew H. Cairns on Behalf of the American Society of Civil Engineers before the Environment and Public Works Committee, United States Senate, on the Water Resources Development Act.

American Society of Civil Engineers. (2013b). Report Card for America's Infrastructure 2013.

Anderson, C. W., Santos, J. R., & Haimes, Y. Y. (2007). A risk-based input–output methodology for measuring the effects of the August 2003 northeast blackout. *Economic Systems Research*, *19*(2), 183-204.

Arias, P., Randall, G., & Sapiro, G. (2007, June). Connecting the out-of-sample and pre-image problems in kernel methods. *In IEEE Conference on Computer Vision and Pattern Recognition, 2007. CVPR'07.* (pp. 1-8). IEEE.

Asbeck, E. L., & Haimes, Y. Y. (1984). The partitioned multiobjective risk method (PMRM). *Large Scale Systems in Information and Decision Technologies*, *6*(1), 13-38.

Babcock, M. W., & Lu, X. (2002). Forecasting inland waterway grain traffic. *Transportation Research Part E: Logistics and Transportation Review*, *38*(1), 65-74.

Barker, K., Ramirez-Marquez, J. E., & Rocco, C. M. (2013). Resilience-based network component importance measures. *Reliability Engineering & System Safety*, *117*, 89-97.

Barker, K., & Rocco S, C. M. (2011). Evaluating uncertainty in risk-based interdependency modeling with interval arithmetic. *Economic Systems Research*, *23*(2), 213-232.

Barker, K., & Santos, J. R. (2010a). Measuring the efficacy of inventory with a dynamic input–output model. *International Journal of Production Economics*, *126*(1), 130-143.

Barker, K., & Santos, J. R. (2010b). A risk-based approach for identifying key economic and infrastructure systems. *Risk Analysis*, *30*(6), 962-974.

Barker, K., & Wilson, K. J. (2012). Decision trees with single and multiple interval-valued objectives. *Decision Analysis*, *9*(4), 348-358.

Baroud, H., & Barker, K. (2014). Bayesian kernel methods for critical infrastructure resilience modeling. In *Second International Conference on Vulnerability and Risk Analysis and Management (ICVRAM) and the Sixth International Symposium on Uncertainty, Modeling, and Analysis (ISUMA)*.

Baroud, H., Barker, K., & Grant, F. H. (2014a). Multiobjective stochastic inoperability decision tree for infrastructure preparedness. *Journal of Infrastructure Systems*, *20*(2), 04013012.

Baroud, H., Barker, K., & Lurvey, R. (2013, January). Bayesian kernel model for disruptive event data. *In Proceedings of IIE Annual Conference*. (p. 1777). Institute of Industrial Engineers-Publisher.

Baroud, H., Barker, K., & Ramirez-Marquez, J. E. (2014b). Importance measures for inland waterway network resilience. *Transportation Research Part E: Logistics and Transportation Review*, *62*, 55-67.

Baroud, H., Barker, K., Ramirez-Marquez, J. E., & Rocco, C. M. (2013b, June). Modeling resilience in infrastructure networks: Inherent costs and interdependent impacts. In *Proceedings of the 11th International Conference on Structural Safety and Reliability, New York City*.

Baroud, H., Barker, K., Ramirez-Marquez, J. E., & Rocco, C. M. (2014c). Inherent costs and interdependent impacts of infrastructure network resilience. *Risk Analysis*, *35*(4), 642-662.

Baroud, H., Ramirez-Marquez, J. E., Barker, K., & Rocco, C. M. (2014d). Stochastic measures of network resilience: Applications to waterway commodity Flows. *Risk Analysis*, *34*(7), 1317-1335.

Baxt, W. G. (1995). Application of artificial neural networks to clinical medicine. *The Lancet*, *346*(8983), 1135-1138.

Mr. Bayes, & Price, M. (1763). An essay towards solving a problem in the doctrine of chances. By the late Rev. Mr. Bayes, FRS communicated by Mr. Price, in a letter to John Canton, AMFRS. *Philosophical Transactions* (1683-1775), 370-418.

Ben-Hur, A. & Noble, W.S. (2005). Kernel methods for predicting protein–protein interactions. *Bioinformatics, 21*(Suppl. 1), 38–46.

Beuthe, M., Jourquin, B., Geerts, J. F., & a Ndjang'Ha, C. K. (2001). Freight transportation demand elasticities: a geographic multimodal transportation network analysis. *Transportation Research Part E: Logistics and Transportation Review*, *37*(4), 253-266.

Bier, V. M., Haphuriwat, N., Menoyo, J., Zimmerman, R., & Culpen, A. M. (2008). Optimal resource allocation for defense of targets based on differing measures of attractiveness. *Risk Analysis*, *28*(3), 763-770.

Bishop, C.M. & Tipping, M.E. (2003). Bayesian regression and classification. In Suykens, J. A. K., Horváth, G., Basu, S., Micchelli, C., and Vandewalle, J. (Eds), *Advances in Learning Theory: Methods, Models and Applications* (pp. 267-288). Amsterdam, The Netherlands: IOS press.

Boin, A., Comfort, L. K., Demchak, C. C. (2009). *The Rise of Resilience*. In Comfort LK, Boin A, Demchak CC (eds). Pp. 1-12 in Design Resilience: Preparing for Extreme Events. Pittsburgh, PA: Pittsburgh University Press.

Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J. (1984). *Classification and Regression Trees*, Wadsworth International Group.

Breslow, N. E. (1984). Extra-Poisson variation in log-linear models. *Applied Statistics*, 38-44.

Broussard, R. P., Kennell, L. R., Ives, R. W., & Rakvic, R. N. (2008, February). An artificial neural network based matching metric for iris identification. In *Electronic Imaging 2008* (pp. 68120S-68120S). International Society for Optics and Photonics.

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

Bunn, M. (2006). A mathematical model of the risk of nuclear terrorism. *The ANNALS of the American Academy of Political and Social Science*, *607*(1), 103-120.

Bureau of Economic Analysis. (2010). Interactive access to input-output accounts data. http://www.bea.gov.in.

Cadarso, L., Marín, Á., & Maróti, G. (2013). Recovery of disruptions in rapid transit networks. *Transportation Research Part E: Logistics and Transportation Review, 53*, 15-33.

Cameron, A. C., & Trivedi, P. K. (1986). Econometric models based on count data. Comparisons and applications of some estimators and tests. *Journal of Applied Econometrics*, *1*(1), 29-53.

Cameron, A. C., & Trivedi, P. K. (2013). *Regression Analysis of Count Data* (Vol. 53). Cambridge university press.

Camps-Valls, G., Rojo-Alvarez, J.L., & Martinez-Ramon, M. (2006). *Kernel Methods in Bioengineering, Signal and Image Processing*. Hershey, PA: IGI Global.

Carlin, B. P., & Louis, T. A. (2008). *Bayesian Methods for Data Analysis*. CRC Press.

Carpenter, S., Walker, B., Anderies, J. M., & Abel, N. (2001). From metaphor to measurement: resilience of what to what?. *Ecosystems, 4*(8), 765-781.

Chankong, V. and Haimes, Y. Y. (2008). *Multiobjective Decision Making: Theory and Methodology*. Mineola, NY: Dover Publications, Inc.

Chen, L., & Miller-Hooks, E. (2012). Resilience: an indicator of recovery capability in intermodal freight transport. *Transportation Science, 46*(1), 109-123.

Cherkassky, V. & Mulier, F. (1998). *Learning from Data: Concepts, Theory, and Methods*. Hoboken, NJ: Wiley.

Clark, C., Henrickson, K. E., Thoma, P. (2005). *An Overview of the U.S. Inland Waterway System.* Institute for Water Resources, U.S. Army Corps of Engineers.

Clauset, A., & Gleditsch, K. S. (2012). The developmental dynamics of terrorist organizations. *PLoS one, 7*(11), e48633.

Clauset, A., Shalizi, C. R., & Newman, M. E. (2009). Power-law distributions in empirical data. *SIAM review*, *51*(4), 661-703.

Clauset, A., & Wiegel, F. W. (2009). A generalized aggregation-disintegration model for the frequency of severe terrorist attacks. *Journal of Conflict Resolution*, *54*(1), 179-197.

Clauset, A., & Young, M. (2005). Scale invariance in global terrorism. *arXiv preprint physics/0502014*.

Clauset, A., Young, M., & Gleditsch, K. S. (2007). On the frequency of severe terrorist events. *Journal of Conflict Resolution*, *51*(1), 58-87.

Comfort, L. K., Boin, A., & Demchak, C. C. (Eds.). (2010). *Designing Resilience: Preparing for Extreme Events*. University of Pittsburgh Pre.

Conway, R. W., & Maxwell, W. L. (1962). A queuing model with state dependent service rates. *Journal of Industrial Engineering*, *12*(2), 132-136.

Cox, D. R. (1983). Some remarks on overdispersion. *Biometrika*, *70*(1), 269-274.

Cristianini, N., & Shawe-Taylor, J. (2000). *An Introduction to Support Vector Machines and Other Kernel based Learning Methods.* Cambridge, UK: Cambridge University Press.

Dentcheva, D., & Ruszczyński, A. (2003). Optimization with stochastic dominance constraints. *SIAM Journal on Optimization*, *14*(2), 548-566.

Dentcheva, D., & Ruszczyński, A. (2004). Semi-infinite probabilistic optimization: first-order stochastic dominance constrain. *Optimization*, 53(5-6), 583-601.

Department of Homeland Security. (2012). *FY 2012 Preparedness Grant Program Overview*, Grant Program Directorate.

Department of Homeland Security. (2009). *National Infrastructure Protection Plan: Partnering to Enhance Protection and Resiliency*. Technical report.

Dessai, S., & Walter, M. E. (2000). Self-organised criticality and the atmospheric sciences: selected review, new findings and future directions. *XE Extreme Events: Developing a Research Agenda for the 21st Century*, 34-44.

Dicdican, R. Y., & Haimes, Y. Y. (2005). Relating multiobjective decision trees to the multiobjective risk impact analysis method. *Systems Engineering*, *8*(2), 95-108.

Ezell, B. C., Farr, J. V., & Wiese, I. (2000). Infrastructure risk analysis model. *Journal of Infrastructure Systems*, *6*(3), 114-117.

Figueiredo, M. (2001). Adaptive sparseness using Jeffreys prior. In *Advances in Neural Information Processing Systems* (pp. 697-704).

Fiksel, J. (2003). Designing resilient, sustainable systems. *Environmental Science & Technology, 37*(23), 5330-5339.

Floyd, M. S., Baroud, H., & Barker, K. (2014). Empirical analysis of Bayesian kernel methods for modeling count data. In *Systems and Information Engineering Design Symposium (SIEDS), 2014* (pp. 328-333). IEEE.

Folga, S., Allison, T., Seda-Sanabria, Y., Matheu, E., Milam, T., Ryan, R., & Peerenboom, J. (2009). A systems-level methodology for the analysis of inland waterway infrastructure disruptions. *Journal of Transportation Security,* 2(4), 121-136.

Fonarow, G. C., Adams, K. F., Abraham, W. T., Yancy, C. W., Boscardin, W. J., & ADHERE Scientific Advisory Committee. (2005). Risk stratification for in-hospital mortality in acutely decompensated heart failure: classification and regression tree analysis. *Jama*, *293*(5), 572-580.

Fong, D. Y., & Yip, P. (1993). An EM algorithm for a mixture model of count data. *Statistics & Probability Letters*, *17*(1), 53-60.

Gardner, M. W., & Dorling, S. R. (1998). Artificial neural networks (the multilayer perceptron)—a review of applications in the atmospheric sciences. *Atmospheric Environment*, *32*(14), 2627-2636.

Gordon, P., Moore II, J. E., Richardson, H. W., Shinozuka, M., An, D., & Cho, S. (2004). Earthquake disaster mitigation for urban transportation systems: An integrated methodology that builds on the Kobe and Northridge experiences. In *Modeling Spatial and Economic Impacts of Disasters* (pp. 205-232). Springer Berlin Heidelberg.

Government Accountability Office. (2011). *Surface Freight Transportation: A Comparison of the Costs of Road, Rail, and Waterways Freight Shipments That Are Not Passed on to Consumers*. GAO-11-134. Washington, DC: United States Government Accountability Office.

Grier, D. V. (2009). The declining reliability of the US inland waterway system. *Institute for Water Resources*, Army Corps of Engineers. Alexandria, VA.

Guikema, S. D. (2007). Formulating informative, data-based priors for failure probability estimation in reliability analysis. *Reliability Engineering & System Safety*, 92(4), 490-502.

Guikema, S. D. (2009). Natural disaster risk analysis for critical infrastructure systems: An approach based on statistical learning theory. *Reliability Engineering & System Safety, 94*(4), 855-860.

Guikema, S. D., & Goffelt, J. P. (2008). A flexible count data regression model for risk analysis. *Risk analysis, 28*(1), 213-223.

Haimes, Y. Y. (2009a). On the definition of resilience in systems. *Risk Analysis, 29*(4), 498-501.

Haimes, Y. Y. (2009b). *Risk Modeling, Assessment, and Management*. Hoboken, NJ: John Wiley & Sons.

Haimes, Y. Y., Crowther, K., & Horowitz, B. M. (2008). Homeland security preparedness: Balancing protection with resilience in emergent systems. *Systems Engineering, 11*(4), 287-308.

Haimes, Y. Y., Li, D., & Tulsiani, V. (1990). Multiobjective decision-tree analysis. *Risk Analysis, 10*(1), 111-129.

Hall, D. B., & Shen, J. (2010). Robust estimation for zero-inflated Poisson regression. *Scandinavian Journal of Statistics, 37*(2), 237-252.

Hallegatte, S. (2008). An adaptive regional Input-Output model and its application to the assessment of the economic cost of Katrina. *Risk Analysis, 28*(3), 779-799.

Hallegatte, S. (2014). Modeling the role of inventories and heterogeneity in the assessment of the economic costs of natural disasters. *Risk Analysis, 34*(1), 152-167.

Ham, H., Kim, T. J., & Boyce, D. (2005a). Assessment of economic impacts from unexpected events with an interregional commodity flow and multimodal transportation network model. *Transportation Research Part A: Policy and Practice, 39*(10), 849-860.

Ham, H., Kim, T. J., & Boyce, D. (2005b). Implementation and estimation of a combined model of interregional, multimodal commodity shipments and transportation network flows. *Transportation Research Part B: Methodological,39*(1), 65-79.

Hastie, T., Tibshirani, R., & Friedman, J. (2001). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. New York, NY: Springer.

Haveman, J. D., & Shatz, H. J. (2006). *Protecting the Nation's Seaports: Balancing Security and Cost*. San Francisco, Calif: Public Policy Institute of California.

Henry, D., & Ramirez-Marquez, J. E. (2012). Generic metrics and quantitative approaches for system resilience as a function of time. *Reliability Engineering & System Safety, 99*, 114-122.

Hespos, R. F., & Strassmann, P. A. (1965). Stochastic decision trees for the analysis of investment decisions. *Management Science, 11*(10), B-244.

Hines, P., Cotilla-Sanchez, E., & Blumsack, S. (2010). Do topological models provide good information about electricity infrastructure vulnerability? *Chaos: An Interdisciplinary Journal of Nonlinear Science, 20*(3), 033122.

Hoffman, E. (2007). Building a resilient business. *Raptor Networks Technology Inc*.

Holling, C. S. (1973). Resilience and stability of ecological systems. *Annual Review of Ecology and Systematics*, 1-23.

Hollnagel, E., Woods, D. D., & Leveson, N. (2007). *Resilience Engineering: Concepts and Precepts*. Ashgate Publishing, Ltd.

Jackson, S. (2007, June). System resilience: capabilities, culture and infrastructure. In *Proceedings of the 17th International INCOSE Symposium*, San Diego, CA.

Jackson, S. (2009). *Architecting Resilient Systems: Accident Avoidance and Survival and Recovery from Disruptions* (Vol. 66). John Wiley & Sons.

Jenelius, E., Petersen, T., & Mattsson, L. G. (2006). Importance and exposure in road network vulnerability analysis. *Transportation Research Part A: Policy and Practice*, 40(7), 537-560.

Jervis, R., & Bath, A. (2012). Gulf Coast stays on guard in Isaac's wake. *USA TODAY, August, 31*.

Jiang, P., & Haimes, Y. Y. (2004). Risk management for Leontief-based interdependent systems. *Risk Analysis, 24*(5), 1215-1229.

Johnson, N. F., Spagat M., Restrepo, J. A., Becerra, O., Bohorquez, J. C., Suraez, N., Restropo, E. M., Zarama, R. (2006). Universal patterns underlying ongoing wars and terrorism. *arXiv preprint physics/0605035*.

Johnson, N. L., Kemp, A. W., & Kotz, S. (2005). *Univariate discrete distributions* (Vol. 444). John Wiley & Sons.

Jung, J., Santos, J. R., & Haimes, Y. Y. (2009). International Trade Inoperability Input-Output Model (IT-IIM): Theory and application. *Risk analysis, 29*(1), 137-154.

Kaplan, S., & Garrick, B. J. (1981). On the quantitative definition of risk. *Risk analysis*, *1*(1), 11-27.

Kass, R. E., & Wasserman, L. (1996). The selection of prior distributions by formal rules. *Journal of the American Statistical Association*, *91*(435), 1343-1370.

Kuo, W., & Zhu, X. (2012). *Importance Measures in Reliability, Risk, and Optimization: Principles and Applications*. John Wiley & Sons.

Kutner, M. H., Nachtsheim, C., Neter, J., & Li, W. (2005). *Applied Linear Statistical Models*. 5th edition. New York: McGraw-Hill-Irwin.

Lambert, D. (1992). Zero-inflated Poisson regression, with an application to defects in manufacturing. *Technometrics*, *34*(1), 1-14.

Lawless, J. F. (1987). Negative binomial and mixed Poisson regression. *Canadian Journal of Statistics*, *15*(3), 209-225.

Lee, B. K., & Kim, K. H. (2010). Optimizing the block size in container yards. *Transportation Research Part E: Logistics and Transportation Review*, *46*(1), 120-135.

Lee, T. W., Park, N. K., & Lee, D. W. (2003). A simulation study for the logistics planning of a container terminal in view of SCM. *Maritime Policy & Management*, *30*(3), 243-254.

Leemis, L. M. (2009). *Reliability: Probabilistic Models and Statistical Methods*, 2nd edition. Williamsburg, VA: Lawrence Leemis.

Leontief, W. (1966). *Input-Output Economics*. New York: Oxford University Press.

Levitin, G., & Hausken, K. (2010). Separation in homogeneous systems with independent identical elements. *European Journal of Operational Research*, *203*(3), 625-634.

Lewis, H. W., Budnitz, R. J., Rowe, W. D., Kouts, H. J. C., Von Hippel, F., Loewenstein, W. B., & Zachariasen, F. (1979). Risk assessment review group report to the US Nuclear Regulatory Commission. *IEEE Transactions on Nuclear Science*, *26*(5), 4686-4690.

Lian, C., & Haimes, Y. Y. (2006). Managing the risk of terrorism to interdependent infrastructure systems through the dynamic inoperability input–output model. *Systems Engineering*, *9*(3), 241-258.

Liu, H., Davidson, R. A., Rosowsky, D. V., & Stedinger, J. R. (2005). Negative binomial regression of electric power outages in hurricanes. *Journal of Infrastructure Systems*, *11*(4), 258-267.

Lord, D., Washington, S. P., & Ivan, J. N. (2005). Poisson, Poisson-gamma and zero-inflated regression models of motor vehicle crashes: balancing statistical fit and theory. *Accident Analysis & Prevention*, *37*(1), 35-46.

Lovett, A., & Flowerdew, R. (1989). Analysis of count data using Poisson regression. *The Professional Geographer*, *41*(2), 190-198.

Lowrance, W. W. (1976). *Of Acceptable Risk: Science and the Determination of Safety*. Los Altos, Calif: W. Kaufmann.

MacKenzie, C. A., & Barker, K. (2012). Empirical data and regression analysis for estimation of infrastructure resilience with application to electric power outages. *Journal of Infrastructure Systems*, *19*(1), 25-35.

MacKenzie, C., Barker, K., & Grant, F. H. (2012a). Evaluating the consequences of an inland waterway port closure with a dynamic multiregional interdependence model. *Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions on*, *42*(2), 359-370.

MacKenzie, C. A., Baroud, H., & Barker, K. (2014a). Static and dynamic resource allocation models for recovery of interdependent systems: application to the Deepwater Horizon oil spill. *Annals of Operations Research*, 1-27.

MacKenzie, C. A., Santos, J. R., & Barker, K. (2012b). Measuring changes in international production from a disruption: Case study of the Japanese earthquake and tsunami. *International Journal of Production Economics*,*138*(2), 293-302.

MacKenzie, C. A., Trafalis, T. B., & Barker, K. (2014b). A Bayesian Beta kernel model for binary classification and online learning problems. *Statistical Analysis and Data Mining: The ASA Data Science Journal*, *7*(6), 434-449.

Malamud, B. D., & Turcotte, D. L. (2012). Statistics of severe tornadoes and severe tornado outbreaks. *Atmospheric Chemistry and Physics*, *12*(18), 8459-8473.

Mallick, B.K., Ghosh, D., & Ghosh, M. (2005). Bayesian classification of tumours by using gene expression data. *Journal of the Royal Statistical Society, Part B*, *67*(2), 219-234.

Mason, M., & Lopes, M. (2011, March). Robot self-initiative and personalization by learning through repeated interactions. *In Conference on Human-Robot Interaction (HRI), 2011 6th ACM/IEEE International* (pp. 433-440). IEEE.

McCullagh, P., & Nelder, J. A. (1989). *Generalized Linear Models* (Vol. 37). CRC press.

Mega, M. S., Allegrini, P., Grigolini, P., Latora, V., Palatella, L., Rapisarda, A., & Vinciguerra, S. (2003). Power-law time distribution of large earthquakes. *Physical Review Letters*, *90*(18), 188501.

Minkel, J. R. (2008). The 2003 Northeast Blackout--Five Years Later. *Scientific American*, 13.

Mitschele, A., Chalup, S., Schlottmann, F., & Seese, D. (2006). Applications of Kernel Methods in Financial Risk Management. *Computing in Economics and Finance, Society for Computational, 317*.

Montesano, L., & Lopes, M. (2009, June). Learning grasping affordances from local visual descriptors. In *Development and Learning, 2009. ICDL 2009. IEEE 8th International Conference on* (pp. 1-6). IEEE.

Mouatassim, Y., & Ezzahid, E. H. (2012). Poisson regression and zero-inflated Poisson regression: application to private health insurance data. *European Actuarial Journal*, *2*(2), 187-204.

Murray-Tuite, P., & Mahmassani, H. (2004). Methodology for determining vulnerable links in a transportation network. *Transportation Research Record: Journal of the Transportation Research Board*, (1882), 88-96.

Nagurney, A., & Qiang, Q. (2007, May). A transportation network efficiency measure that captures flows, behavior, and costs with applications to network component importance identification and vulnerability. *In Proceedings of the POMS 18th Annual Conference, Dallas, Texas, USA*, (Vol. 4).

National Cooperative Freight Research Program. (2010). *Research Results Digest, Subject Areas: IV Operations and Safety, VII Rail, VIII Freight Transportation, IX Marine Transportation.*

Natvig, B., Huseby, A. B., & Reistadbakk, M. O. (2011). Measures of component importance in repairable multistate systems—a numerical study. *Reliability Engineering & System Safety*, *96*(12), 1680-1690.

Nedler, J. A., & Wedderburn, R. W. M. (1972). Generalized linear models. *Journal of the Royal Statistical Society, Part A, 135(3)*, 370–384.

Okuyama, Y. (2004). Modeling spatial economic impacts of an earthquake: Input-output approaches. *Disaster Prevention and Management: An International Journal*, *13*(4), 297-306.

Oliva, G., Panzieri, S., & Setola, R. (2011). Fuzzy dynamic input–output inoperability model. *International Journal of Critical Infrastructure Protection*, *4*(3), 165-175.

Organization for Economic Co-Operation and Development. 2011. OECD.StatsExtrats. http://stats.oecd.org/Index.aspx.

Orsi, M. J., & Santos, J. R. (2010). Incorporating time-varying perturbations into the dynamic inoperability input–output model. *Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions*, 40(1), 100-106.

Pant, R., Barker, K., Grant, F. H., & Landers, T. L. (2011). Interdependent impacts of inoperability at multi-modal transportation container terminals. *Transportation Research Part E: Logistics and Transportation Review*, *47*(5), 722-737.

Pant, R., Barker, K., Ramirez-Marquez, J. E., & Rocco, C. M. (2014). Stochastic measures of resilience and their application to container terminals. *Computers & Industrial Engineering*, *70*, 183-194.

Park, J. (2008). The economic impacts of dirty bomb attacks on the Los Angeles and Long Beach Ports: Applying the supply-driven NIEMO (National Interstate Economic Model). *Journal of Homeland Security and Emergency Management*, *5*(1).

Paul, S. R., & Plackett, R. L. (1978). Inference sensitivity for Poisson mixtures. *Biometrika*. *65(3)*, 591-602.

Qiang, Q., & Nagurney, A. (2008). A unified network performance measure with importance identification and the ranking of network components. *Optimization Letters*, *2*(1), 127-142.

Quinlan, J. R. (1983). Learning efficient classification procedures and their application to chess eng games. In R. S. Michalski, J. G. Carbonell & T. M. Mitchell (eds), *Machine Learning: An Artificial Intelligence Approach*, Morgan Kaufmann, San Mateo, CA.

Quinlan, J. R. (1986). Induction of decision trees. *Machine Learning 1*(1):81-106.

Raiffa, H. (1968). *Decision analysis*. Addison-Wesley, Reading, MA.

Ramirez-Marquez, J. E., & Coit, D. W. (2005). Composite importance measures for multi-state systems with multi-state components. *IEEE Transactions on Reliability*, *54*(3), 517-529.

Ramirez-Marquez, J. E., & Rocco, C. M. (2012, June). Towards a unified framework for network resilience. In *Proceedings of the Third International Engineering Systems Symposium CESUN* (pp. 18-20).

Ramirez-Marquez, J. E., Rocco, C. M., Gebre, B. A., Coit, D. W., & Tortorella, M. (2006). New insights on multi-state component criticality and importance. *Reliability Engineering & System Safety, 91*(8), 894-904.

Ramirez-Marquez, J. E., Rocco, C. M., & Levitin, G. (2011). Optimal network protection against diverse interdictor strategies. *Reliability Engineering & System Safety, 96*(3), 374-382.

Richardson, L. F. (1948). Variation of the frequency of fatal quarrels with magnitude. *Journal of the American Statistical Association*, *43*(244), 523-546.

Roberts, R. S., & Foppa, I. M. (2006). Prediction of equine risk of West Nile virus infection based on dead bird surveillance. *Vector-Borne & Zoonotic Diseases*, *6*(1), 1-6.

Rocco, S. C. M., & Ramirez-Marquez, J. E. (2012). Innovative approaches for addressing old challenges in component importance measures. *Reliability Engineering and System Safety, 108,* 123-130.

Rose, A. (2009). Economic resilience to disasters: Community and Regional Resilience Institute (CARRI) research report 8. *Oakridge, TN: CARRI Institute*, 2009.

Rosoff, H., & Von Winterfeldt, D. (2007). A risk and economic analysis of dirty bomb attacks on the ports of Los Angeles and Long Beach. *Risk Analysis*,*27*(3), 533-546.

Sacone, S., & Siri, S. (2009). An integrated simulation-optimization framework for the operational planning of seaport container terminals. *Mathematical and Computer Modelling of Dynamical Systems*, *15*(3), 275-293.

Santos, J. R. (2006). Inoperability input-output modeling of disruptions to interdependent economic systems. *Systems Engineering*, *9*(1), 20-34.

Santos, J. R. (2008). Interdependency analysis with multiple probabilistic sector inputs. *Journal of Industrial and Management Optimization*, *4*(3), 489-510.

Santos, J. R., Barker, K., & Zelinke Iv, P. J. (2008). Sequential decision-making in interdependent sectors with multiobjective inoperability decision trees: application to biofuel subsidy analysis. *Economic Systems Research*, *20*(1), 29-56.

Santos, J. R., & Haimes, Y. Y. (2004a). Modeling the demand reduction Input-Output (I-O) inoperability due to terrorism of interconnected infrastructures. *Risk Analysis, 24,* 6, 1437-1451.

Santos, J. R., & Haimes, Y. Y. (2004b). Applying the partitioned multiobjective risk method (PMRM) to portfolio selection. *Risk analysis*, *24*(3), 697-713.

Saunders, C., Gammerman, A., & Vovk, V. (1998). Ridge regression learning algorithm in dual variables. In *(ICML-1998) Proceedings of the 15th International Conference on Machine Learning* (pp. 515-521). Morgan Kaufmann.

Schölkopf, B., Guyon, I., & Weston, J. (2003). Statistical learning and kernel methods in bioinformatics. In P. Frasconi und R. Shamir (Eds.), *Artificial intelligence and heuristic methods in bioinformatics*, (pp. 1-21). Amsterdam, The Netherlands: IOS Press.

Schölkopft, B., & Mullert, K. R. (1999). Fisher discriminant analysis with kernels. *Neural networks for signal processing IX*, *1*, 1.

Schölkopf, B. & Smola, A.J. (2002). *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*. Cambridge, MA: MIT Press.

Schölkopf, B., Smola, A., & Müller, K. R. (1997). Kernel principal component analysis. In *Artificial Neural Networks—ICANN'97* (pp. 583-588). Springer Berlin Heidelberg.

Schwartz, J. (2012). Drought clogs traffic on a shrunken Mississippi. *New York Times*, August, 20. http://www.nytimes.com/2012/08/21/us/drought-clogs-traffic-on-shrunken-mississippi-river.html.

Seeger, M. (2000). Bayesian model selection for support vector machines, Gaussian processes and other kernel classifiers. In Solla, S. A., Leen, T. K., and Müller, K. R. (Eds.), *Advances in Neural Information Processing Systems* (pp. 603-609). Cambridge: MIT Press.

Sepkoski, J. J., & Rex, M. A. (1974). Distribution of freshwater mussels: coastal rivers as biogeographic islands. *Systematic Biology*, 23(2), 165-188.

Setola, R., & De Porcellinis, S. (2008). A methodology to estimate input-output inoperability model parameters. In *Critical Information Infrastructures Security* (pp. 149-160). Springer Berlin Heidelberg.

Shahin, M. A., Jaksa, M. B., & Maier, H. R. (2001). Artificial neural network applications in geotechnical engineering. *Australian Geomechanics*, *36*(1), 49-62.

Shawe-Taylor, J., & Cristianini, N. (2004). *Kernel Methods for Pattern Analysis*. Cambridge university press.

Sohn, J., Hewings, G. J., Kim, T. J., Lee, J. S., & Jang, S. G. (2004). Analysis of economic impacts of an earthquake on transportation network. In *Modeling Spatial and Economic Impacts of Disasters* (pp. 233-256). Springer Berlin Heidelberg.

Spetzler, C. S., & Stael von Holstein, C. A. S. (1975). Exceptional paper-probability encoding in decision analysis. *Management Science*, *22*(3), 340-358.

Tierney, K. J. (1997). Impacts of recent disasters on businesses: The 1993 Midwest floods and the 1994 Northridge Earthquake (No. NCEER-SP-0001). *Buffalo: Multidisciplinary Center for Earthquake Engineering and Research, State University of New York at Buffalo.*

Tipping, M.E. (2001). Sparse Bayesian learning and the relevance vector machine. *Journal of Machine Learning Research*, *1*, 211-244.

Tunaru, R. (2002). Hierarchical Bayesian models for multiple count data. *Austrian Journal of statistics*, *31*(3), 221-229.

Ugural, A. (2003). *Mechanical design: an integrated approach*. McGraw-Hill Science/Engineering/Math.

U.S. Army Corps of Engineers. (2010). *Waterborne Commerce of the United States: Calendar Year 2009*. New Orleans, LA: U.S. Army Corps of Engineers Waterborne Commerce Statistics Center.

U.S. Army Corps of Engineers. 2011. Interactive access to website. http://www.ndc.iwr.usace.army.mil//lpms/lpms.htm.

U.S. Department of Transportation. (2009). *Freight Facts and Figures 2009*. Federal Highway Administration, Office of Freight Management and Operations.

U.S. Department of Transportation. (2011). *Maritime Administration America's Marine Highway*. Report to Congress.

Vapnik, V. (2013). *The Nature of Statistical Learning Theory*. Springer Science & Business Media.

Vapnik, V. N., & Vapnik, V. (1998). *Statistical Learning Theory* (Vol. 1). New York: Wiley.

Vasseur, D., & Llory, M. (1999). International survey on PSA figures of merit. *Reliability Engineering & System Safety*, *66*(3), 261-274.

Vogus, T. J., & Sutcliffe, K. M. (2007, October). Organizational resilience: towards a theory and research agenda. In *IEEE International Conference on Systems, Man and Cybernetics* (pp. 3418-3422). IEEE.

Wang, L. & Zhu, J. (2010). Financial market forecasting using a two-step kernel learning method for the support vector regression. *Annals of Operations Research*, *174*, 103-120.

Webb, G. R., Tierney, K. J., & Dahlhamer, J. M. (2000). Businesses and disasters: Empirical patterns and unanswered questions. *Natural Hazards Review*, *1*(2), 83-90.

Winkelmann, R. (2008). *Econometric Analysis of Count Data*. Springer Science & Business Media.

Woods, D. D. (2005). Creating foresight: Lessons for enhancing resilience from Columbia. *Organization at the Limit: Lessons from the Columbia Disaster*.

Wreathall, J. (2006). Properties of resilient organizations: an initial view. *Resilience Engineering Concepts and Precepts. Burlington, VT: Ashgate.*

Yager, R. R. (1988). On ordered weighted averaging aggregation operators in multicriteria decisionmaking. *Systems, Man and Cybernetics, IEEE Transactions on*, *18*(1), 183-190.

Yamijala, S., Guikema, S. D., & Brumbelow, K. (2009). Statistical estimation of water distribution system pipe reliability. *Reliability Engineering & System Safety*, *94*(2), 282-293.

Yegnanarayana, B. (2009). *Artificial neural networks*. PHI Learning Pvt. Ltd.

Yip, P. (1991). Conditional inference on a mixture model for the analysis of count data. *Communications in Statistics-Theory and Methods*, *20*(7), 2045-2057.

Zeng, A. Z., Durach, C. F., & Fang, Y. (2012). Collaboration decisions on disruption recovery service in urban public tram systems. *Transportation Research Part E: Logistics and Transportation Review*, *48*(3), 578-590.

Zhang, Z., Dai, G., & Jordan, M.I. (2011). Bayesian generalized kernel mixed models. *Journal of Machine Learning Research*, *12*, 111-139.

Zhang, P., & Peeta, S. (2011). A generalized modeling framework to analyze interdependencies among infrastructure systems. *Transportation Research Part B: Methodological*, *45*(3), 553-579.

Zio, E., Marella, M., & Podofillini, L. (2007). Importance measures-based prioritization for improving the performance of multi-state systems: application to the railway industry. *Reliability Engineering & System Safety, 92*(10), 1303-1314.

Zobel, C. W. (2011). Representing perceived tradeoffs in defining disaster resilience. *Decision Support Systems, 50*(2), 394-403.

Zobel, C. W. (2014). Quantitatively representing nonlinear disaster recovery. *Decision Sciences*, 45(6), 1053-1082.