8-2023

Application of Business Analytics Approaches to Address Climate-Change-Related Challenges

Donald J. Jenkins

APPLICATION OF BUSINESS ANALYTICS APPROACHES TO ADDRESS CLIMATE-CHANGE-RELATED CHALLENGES

A Dissertation Presented

by

DONALD J. JENKINS

Submitted to the Office of Graduate Studies,
University of Massachusetts Boston,
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

August 2023

Business Administration Program

APPLICATION OF BUSINESS ANALYTICS APPROACHES TO ADDRESS CLIMATE-CHANGE-RELATED CHALLENGES

A Dissertation Presented

by

DONALD J. JENKINS

Approved as to style and content by:

Foad Mahdavi Pajouh, Associate Professor
Stevens Institute of Technology
Chairperson of Committee

Jeffrey M. Keisler, Professor
Member

Josephine Mazzi Namayanja, Assistant Professor
Member

Paul Kirshen, Professor
Member

Ehsan Elahi, Program Director
Management Science and Information Systems Program

Peng Xu, Chair
Management Science and Information Systems Department

ABSTRACT

APPLICATION OF BUSINESS ANALYTICS APPROACHES TO ADDRESS CLIMATE-CHANGE-RELATED CHALLENGES

August 2023

Donald J. Jenkins, A.A.S., Charter Oak Community College, Connecticut
B.S./B.A., University of San Diego
M.S., Naval Postgraduate School, Monterey
Ph.D., University of Massachusetts Boston

Directed by Associate Professor Foad Mahdavi Pajouh

Climate change is an existential threat facing humanity, civilization, and the natural world. It poses many multi-layered challenges that call for enhanced data-driven decision support methods to help inform society of ways to address the deep uncertainty and incomplete knowledge on climate change issues. This research primarily aims to apply management, decision, information, and data science theories and techniques to propose, build, and evaluate novel data-driven methodologies to improve understanding of climate-change-related challenges. Given that we pursue this work in the College of Management, each essay applies one or more of the three distinct business analytics approaches (i.e., descriptive, prescriptive, and predictive analysis) to aid in developing decision support capabilities. Given the rapid growth in data availability, we evaluate important data characteristics for each analysis, focusing on the data source, granularity, volume, structure, and quality. The final analysis consideration is the methods used on the data output to help coalesce the various model outputs into understandable visualizations, tables, and takeaways. We pursue three distinct business analytics challenges. First, we start with a natural language processing analysis to gain insights into the evolving climate change adaptation discussion in the scientific literature. We then create a stochastic network optimization model with recourse to provide coastal decision-makers with a cost-benefit anal-

ysis tool to simultaneously assess risks and costs to protect their community against rising seas. Finally, we create a decision support tool for helping organizations reduce greenhouse gas emissions through strategic sustainable energy purchasing. Although the three essays vary on their specific business analysis approaches, they all have a common theme of applying business analytics techniques to analyze, evaluate, visualize, and understand different facets of the climate change threat.

ACKNOWLEDGMENTS

I am deeply grateful to many who helped me complete this dissertation. My advisor, Dr. Pajouh, supported me throughout and provided helpful insights to guide me on my journey through the Ph.D. program. I am thankful to my committee members who shared their particular expertise on each of the three essays, specifically:

- Dr. Namayanja on the analysis in Chapter 2, especially for her work in helping me develop the validation steps of the topic analysis.
- Dr. Pajouh on the analysis in Chapter 3, especially for his contributions on helping me develop the proof of complexity for the Flood Risk Mitigation model.
- Dr. Keisler on the analysis in Chapter 4, especially for his contributions in creating and defining the general multi-attribute utility and strategy table Analytica modules used in Chapter 4.
- Dr. Kirshen on Chapters 2 and 3, but more broadly for sharing his deep experience and assisting me in developing more profound insights into the challenges of climate change and the need for data-driven approaches to help us solve these challenges.

In addition to my committee, I would like to thank the following people and organizations for their assistance in particular areas:

- Dr. David Levy for contributing as a subject matter expert on the topic naming and validation analysis completed in Chapter 2.
- Dr. Mahyar Eftekhar of Arizona State University for his mentoring and help in rewriting portions of Chapter 3 as we targeted this work for future publication.

- Carly Foster from Arcadia for assisting in understanding the GIS tax-based data associated with the City of Boston for the work done in Chapter 3.

- The Resource Computing Lab of UMass Boston for providing the computing resources and training to use those systems. These capabilities enabled faster computation for the work done in Chapter 3 during the early phases of data analysis and model runs. I especially want to thank Jeff Dusenberry for having the patience to help me learn how to use the resources available to the UMass Boston research community.

- Dr. Max Henrion and the Lumina Analytica team for providing the licenses necessary to unleash the full capability of Analytica for the work done in Chapter 4.

- My Dissertation Coach, Dr. Rebecca Schuman, for keeping me on track when procrastination was holding me back.

- My editing assistant Nermina Ahmić for applying her extensive Latex skills and helping me properly structure and format the final version of this dissertation.

Finally, I could not have completed this work without the support of my wife, Lorrie. She put up with the many late nights and the lost opportunities to enjoy a sunny weekend while I worked to complete all the work associated with this dissertation. She is the reason why I could keep moving forward whenever I hit a roadblock. I owe her a deep debt of gratitude for being there and supporting me during the six-year journey.

TABLE OF CONTENTS

CHAPTER Page

CHAPTER Page

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

1.1 Background

Climate change is real, and its effects are worsening (United Nations Treaty Collection 2015). Much climate-change-related research highlights the associated risks and the deep uncertainty of the timing and severity of climate change impacts (Hallegatte et al. 2012; Haasnoot et al. 2013). Given the complexity of climate change, numerous fields of study are actively pursuing research into climate change causes and how to adapt to or mitigate its worst effects. Because of this, there is substantial research applying various analytical approaches to manage the associated risks. Just a small sampling of the topics covered include predicting rising sea levels (Garner et al. 2018; Sweet et al. 2017; Neumann et al. 2015), mitigating greenhouse gas creation (Fernandez and Watterson 2012; Vuuren et al. 2011; Google 2013), adapting natural and built environments to climate changes (Rosenzweig et al. 2011; Aerts et al. 2014; Ibáñez-Forés, Bovea, and Pérez-Belis 2014; Toimil et al. 2020; Reiblich et al. 2019; Sutton-Grier, Wowk, and Bamford 2015), addressing impacts on vulnerable communities (Hallegatte et al. 2013; Mikulewicz 2017; Reed et al. 2013), and managing the significant effects of climate-change-related catastrophes (Neumann et al. 2015; Hsiang et al. 2017; Nordhaus 2017). Researchers are applying analytical means to better understand climate change risks.

The complexity of many climate change challenges calls for interdisciplinary approaches to attacking climate change issues head-on. With the rapid increases in data availability, computing resources, and data science capabilities (Pollard, Spencer, and Jude 2018; Hassani, Huang,

and Silva 2019), there is a growing focus on leveraging computer and information sciences to apply unique analytical and algorithmic techniques to tackling climate-change-related issues (Faghmous and Kumar 2014). This increasing focus is most visible in how researchers are forming communities around computational sustainability (Gomes et al. 2019) and climate informatics (Monteleoni et al. 2016), as well as information and decision sciences conferences holding dedicated tracks for sustainability-related issues (INFORMS 2020). This momentum of applying advanced analytics presents a real opportunity to help in the fight against climate change. In furthering the cause of using business analytics to better inform the issues associated with climate change, this dissertation will combine the capabilities of management, decision, information, and data sciences to explore three distinct issues regarding climate change challenges. In applying business analytics, there are several considerations we will discuss in the following paragraphs.

Before applying business analytics to a problem, the analyst needs to understand the purpose of the analysis. The primary goal of business analytics is to help decision-makers understand and solve problems and to make decisions (Evans 2016). In order to make data-driven decisions, the analyst applies a problem-solving approach to define, structure, and analyze the problem (Phillips-Wren, Daly, and Burstein 2021). Upon completion of the analysis, the analyst presents results that will inform relevant stakeholders of the problem while potentially providing possible courses of action and associated challenges of taking those courses of action. In short, sometimes, the analysis may only be undertaken to inform a particular situation, enabling stakeholders to ground their understanding of a problem in a data-driven manner. Other times, the situation may be well understood, and the analysis is applied to aid in the stakeholders' decision-making. Understanding when an analysis' purpose is meant to inform versus aid in decision-making is critical context to know before starting an analysis.

The next consideration is the analytics approach; specifically, the approaches are descriptive, predictive, and prescriptive analyses (Evans 2016). In many business analyses, one or

more of these approaches are used depending on the objective of the analysis. Descriptive analysis is commonly used to understand a business or field's past and current performance. A descriptive analysis allows an interested party to understand the state of a problem and find areas for opportunity. Predictive analytics attempts to predict the future by examining historical data to detect patterns or relationships that can help inform possible futures. Predictive analytics allows one to ask "what if?" questions based on information already captured or understood from the past. Finally, prescriptive analytics attempts to identify the "best" alternatives to choose based on a combination of choices or alternatives that would typically be too challenging for a person to determine effectively. Businesses often apply prescriptive analytics to maximize financial outcomes or minimize overall costs. Although ample literature highlights each analytics approach, in reality, many analytical challenges often apply two or three approaches to address the needs of a particular problem (Evans 2016).

The next essential consideration is the data used in an analysis. In recent years, there has been a significant increase in the amount of data generated and made available for numerous purposes, whether in business, research, or other areas (Desjardins 2019). According to IBM in 2017, "Every day, we create 2.5 quintillion bytes of data. To put that into perspective, 90% of the data in the world today has been created in the last two years alone." (Watson Marketing 2017). This continuing growth in data is due to many factors, including the growth of the internet, the widespread use of digital technologies, and the increasing availability of computing power and storage (Desjardins 2019). The increased data availability has led to myriad new opportunities for researchers in many fields associated with climate change (Monteleoni et al. 2016). However, with this increased data availability, there are new challenges in managing, analyzing, and interpreting larger datasets. As such, researchers must be well-equipped with the necessary skills and tools to navigate this new data landscape effectively (Sebestyén, Czvetkó, and Abonyi 2021).

Before starting an analysis, the analyst needs to identify key data sources to use in the analysis and understand the characteristics of the data. Many sources of research in data usage exist that highlight elements one needs to consider when capturing or using data (Madnick et al. 2009). For our purposes, Table 1 captures the key data characteristics we considered when applying a business analytics approach to the climate-change problems discussed in this research. When conducting analysis, it is essential to use data fit for purpose (Devillers et al. 2007). More often than not, the data is not fit for purpose "as-is", but there may be transformations the analyst can apply to enable using that data for the analysis. Alternative data sources must be identified and obtained if available data cannot be made fit for purpose. Sometimes data may be too costly or unavailable to conduct the desired analysis, so simulated or randomized data may be needed to complete the analysis. Even after identifying the data source, the analyst must spend substantial time assessing the data's usefulness. In today's business world, analysts spend 70-80% of their time developing an understanding of their data and wrangling it to be used within their analysis (Rehman et al. 2019). One should expect the same effort when applying business analytics to climate change challenges. We essentially encountered that same level of effort in conducting the research in the following three chapters.

The final consideration when applying business analytics is the analysis output. Some factors to consider are the presentation format, key findings, visualization methods, model limitations, and actionable recommendations (Evans 2016). Given that we focus the work in our research from a management information sciences perspective, we will seek novel ways to identify the key findings and actionable recommendations through visualizations. Visualization is a powerful tool to help analysts connect to their audience in ways that the written word or numerical tables cannot (Tufte 2006). Often the volume of numerical information precludes effectively communicating patterns or results in a tabular format, but using summary visualizations enables effective display and discussion of the results. There are many visualization techniques to assist in displaying information, with the method used being dependent on the

Table 1: Common Data Characteristics to consider when applying business analytics.

Characteristic	Definition	Examples
Source	The origin of the data	Published reports, sensor data, Geographic Information System (GIS) data, open-source government reports, online text repositories
Granularity	The level of detail or specificity of the data	In business, fine-grained data includes individual transaction details, while coarse-grained data may be monthly or quarterly sales figures
Volume	The size or amount of data that is collected	Big data can contain billions or even trillions of data points, whereas much smaller datasets may contain tens to hundreds of data points
Structure	Data may or may not be organized in a structured manner	Structured data is organized into a specific format, such as a spreadsheet, database, or tables of numbers, while unstructured data is not, such as free text, video, or audio files
Quality	The accuracy, completeness, and consistency of the data (often context dependent)	High-quality data is accurate, complete, and consistent, while poor quality data is prone to errors and inaccuracies

characteristics of the information to be displayed, the purpose of the analysis, and the audience interpreting the results (Keim 2002). We believe that applying business analytics techniques discussed in the preceding provides additional interdisciplinary capability in attacking the challenges of climate change.

1.2 Problem statement

Climate change is a grave threat to both humanity and the natural world. We can better understand climate change's multifaceted challenges by applying data-driven theories and techniques commonly used in management and business settings. It is crucial to develop effective decision-making strategies relying on data analyses that assist society in addressing knowledge gaps and understanding complex uncertainties related to climate change. This study aims to assess the effectiveness of the business analytics approach in addressing climate change issues through applications such as understanding adaptation discussions, aiding businesses in energy purchasing decisions that contribute to mitigation efforts, and equipping coastal decision-

makers with decision support tools to assess risks and costs associated with safeguarding their communities.

1.3 Research purpose

This research aims to apply decision, information, and data science theories and techniques to propose, build, and evaluate novel data-driven methodologies to improve understanding climate-change-related challenges. Additionally, given that we undertake this work in the College of Management, we will bring a management focus to the discussion. We will apply the business analytics considerations discussed above, considering the analysis purpose, the approach used, characteristic data assessment, and effective output methodologies. Each essay will incorporate one or more of the three distinct business analytics approaches (i.e., descriptive, prescriptive, and predictive analysis) (Evans 2016) to aid in developing decision support capabilities. Regardless of the analytics approaches used in each paper, all analyses ultimately have a common thread of using various analytical techniques to evaluate, visualize, and assess applying data science to understand or address associated aspects of the climate change threat.

1.4 Research essays

To commence our exploration, we will start by laying out the flow of the chapters. We begin in Chapter 2 by delving into existing research on climate change. We aim to gain insights from the existing research to enable us to apply business analytics to address climate-change-related challenges. Within Chapter 2, we discover existing topics where a business analytics mindset can be applied. Of particular significance to our overall investigation is that, among nearly 15K papers encompassing climate change adaptation, finance-related issues never emerge as the majority focus in any paper. Building upon this revelation, we approach climate change-

related challenges from two finance-related perspectives. In the first perspective, discussed in Chapter 3, we present work that explores the effects of climate change, specifically focusing on investing to safeguard an urban coastal region against sea-level rise. Chapter 3 demonstrates that the costs associated with flooding due to sea-level rise can be substantial and highly uncertain. These costs lead us to explore the second finance-related perspective in Chapter 4. Instead of passively enduring and managing the costly and uncertain repercussions of climate change, we propose a decision-support model that empowers global companies to curtail their contributions to greenhouse gas emissions while considering the value of going green. Specifically, we incorporate financial, sustainability, and brand attributes to help decision-makers consider the overarching value of their sustainable energy purchasing needs. Ultimately, the three core chapters of this dissertation each contain an essay dedicated to understanding or addressing climate-change-related challenges. Knowing the reasoning behind the chapter flow, we will next discuss in more detail the purpose, approach, data characteristics, and outputs of each essay and summarize the discussions in Tables 2, 3, and 4.

We begin in Chapter 2 by evaluating the discourse occurring within published climate change adaptation literature using the natural language processing technique of Latent Dirichlet Allocation (LDA) (Blei, Ng, and Jordan 2003). We use the LDA model results, conducting a descriptive analysis of climate change adaptation topics discussed over the last two decades, their interrelation, and their prevalence over time. Even though the final output is primarily a descriptive analysis, the underlying LDA model used is predictive, transforming unstructured research article abstracts into a quantitative matrix for analysis. The data source comprises published literature abstracts identified and downloaded from the Web of Science's extensive library (Clarivate Analytics 2020). The dataset contains nearly 15K abstracts, much larger than that used for a typical literature review but relatively modest for our LDA analysis. One would expect high-quality data from this source. However, given that we are looking across many fields of study, we had substantial data-wrangling considerations. For instance, we under-

Table 2: Relevant analysis considerations for Chapter 2 (*Text-mining climate change adaptation research for thematic patterns emerging in the 21st century.*)

Consideration	Details
Purpose	Inform
Approach	Descriptive approach with underlying LDA predictive text-mining application
Data Source	Published articles from Web of Science research database and subject matter expertise from climate change experts
Data Granularity	Rough, the abstracts of papers are used rather than the full articles
Data Volume	Small, 15K abstracts comprised of hundreds of words
Data Structure	Mostly unstructured text data, with additional metadata fields in a table format (e.g., publication date, journal, field of study)
Data Quality	Expected high quality, but some cleansing required (e.g., inconsistent acronym usage, English language differences)
Output	Predictive text model output evaluated by subject matter experts then used to evaluate patterns in the data using time series, correlation, and heatmap visualizations supported with summary table outputs

took data cleaning efforts to account for acronyms used inconsistently across fields, different English word spellings, and bi-lingual entries for some abstracts. The LDA model output is relatively abstract and complicated. Therefore, we used Subject Matter Experts (SMEs) to review and validate the output of relevant words in order to name the topics meaningfully. This multi-tiered analysis enabled us to conduct subsequent descriptive analyses to plot time-series graphs, heatmaps, correlation matrices, and summary tables showing the most pertinent information describing the relationships within the dataset. Ultimately, this descriptive analysis exposed the depth and breadth of topics covered within climate change adaptation research over the last two decades and helped set the stage for enabling the analyses done in Chapters 3 and 4.

In Chapter 3, we conduct an analysis that applies all three of the business analytics approaches. At its core we have the creation of a network optimization model, which is prescriptive in nature. However, as seen in Table 3, because of the model's complexity and the inclusion of multiple datasets, we also require a liberal application of both predictive and de-

scriptive analytics to solve the model and understand its inputs and outputs. In this chapter, we develop a cost-benefit analysis model for evaluating coastal protection along a city coastline. This model focuses on adapting to climate change over time using a stochastic approach incorporating future sea level rise. Our model is novel in that we use a network graph-based approach to assess a non-protected coastal area's risk while simultaneously evaluating the costs of implementing a flood protection system. Although the core model is a prescriptive network optimization, the ultimate decision support analysis also applies a predictive approach in the use of probabilisitic scenarios and simulation to assess the model outputs on real-world data. In evaluating the model's generalizability, we create and assess random network data using the same techniques. With voluminous input and output data, the summary discussions also require descriptive analysis in the form of geographic map visualizations, cost-benefit analysis charts and distribution boxplots aggregating billions of data points for millions of model simulation runs. To enable coastal decision-makers to potentially use our modeling approach elsewhere, we demonstrate our case study by using real-world data from only open-source data repositories, including geographical elevation data, city tax data, tidal gauge data, and published sea level forecasts. The open-source data required significant transformation to convert into usable data to create a network model. The characteristics of the input and output datasets cut across the entire spectrum, ranging in size from large GIS elevation datasets to simple data tables with a handful of tidal values. We believe the novelty of this model, combined with the ability to execute it with fully open-source datasets, presents an opportunity for coastal decision-makers to evaluate the risks in their geographic area, assess the potential costs to adapt to rising sea levels, and ultimately mitigate its effects.

In seeing the potentially astronomical costs for mitigating damages due to rising sea levels in Chapter 3, we go upstream of the problem and aim to reduce the underlying greenhouse gas emissions in Chapter 4's research. The overarching analytical framework is prescriptive in its focus, incorporating multi-attribute utility theory (Keeney, Raiffa, and Meyer 1993) and

Table 3: Relevant analysis considerations for Chapter 3 (*Which is more rewarding in managing sea level rise and hurricane storm surge flooding: mitigation or response?)*

Consideration	Details
Purpose	Decision Support
Approach	Prescriptive approach creating network optimization model, with liberal application of predictive simulation and descriptive analysis
Data Source	Varied, Open source using published research for sea level rise, sensor data from tide level gauges, GIS land elevation and tax data, tax assessor data tables, and simulated network data for testing model's generalizability
Data Granularity	Varied, some data very precise and granular (e.g., land elevation heights to one square meter), some rough estimations based on past literature (e.g., costs to build levees), and some data simulated from random generation (e.g., random network generation)
Data Volume	Varied, inputs Big (e.g., GIS land elevation), to Small (e.g., discount rate single values), to outputs being Big (billions of network output data points)
Data Structure	Varied, very structured datasets (e.g., GIS land elevation) to semi-structured (e.g., tax file metadata), to probability distributions (e.g., storm and rising sea levels)
Data Quality	Varied, high quality (e.g., GIS land elevation), estimated (e.g., cost assumptions for building levees), and moderate (e.g., incomplete tax assessor data)
Output	Multiple, geographic network (e.g., defining relevant area for analysis), chloropleth displays (e.g., grid flood damages and investments), aggregated stacked bar charts (e.g., cost benefit graphs), boxplot, and panel graphs (e.g., parameter sensitivity analysis)

strategy tables (McNamee and Celona 2008) to strategically assess a multinational company's sustainable energy purchases over an extended period as shown in Table 4. This analysis also incorporates a predictive approach in using the underlying Monte Carlo simulation (Mooney 1997). In this study, our model is novel in that it goes beyond just a cost-benefit model and incorporates three distinct attributes of energy cost, sustainability requirements, and brand prestige to determine the option with the best utility for the decision-maker. In this case study, while assessing tools, we needed to create generalizable multi-attribute utility theory and strategy table modules for use in the decision support software Lumina Analytica. This final analysis focuses on helping mitigate the worst of climate change effects by helping a company potentially reduce greenhouse gas emissions. The case study's data source includes both anonymized real-world energy usage data for a factory based in Mexico and estimated values for other market dynamics based on input from SMEs who worked directly with the customer. Bringing in distributions for several key parameters, we run simulations to provide the decision maker with output distributions of potential costs and utility values. We capture the analysis outputs with descriptive analysis to describe the results using expected value tables, probability distribution charts, and a parameter sensitivity analysis. These outputs will enable a decision-maker to understand the potential outcomes for each strategic scenario presented. Utilizing a business analytics method on this problem enables decision support for decision-makers to reduce greenhouse gas emissions from their energy usage and mitigate the worst-case scenarios associated with climate change.

It is clear from the variety of elements found in Tables 2, 3, and 4 that business analytics can be applied to help fight climate change. The diversity of analytical approaches in these chapters shows that climate change is a complex issue requiring an interdisciplinary approach. Business analytics can provide insights into the complex interactions between human activities and the environment, helping organizations to develop effective strategies to mitigate and adapt to the impacts of climate change. The upcoming chapters will evaluate potential analytical opportu-

Table 4: Relevant analysis considerations for Chapter 4 (*A decision-analytic tool for corporate strategic sustainable energy purchases.*)

Consideration	Details
Purpose	Decision Support
Approach	Prescriptive approach in overarching framework, predictive approach involved in using Monte-Carlo simulation, and descriptive approach for summaries
Data Source	Anonymized company energy usage data and simulated values for remaining inputs
Data Granularity	Low granularity, with real-world facility energy usage broken down across monthly billing periods and estimates for other parameters based on conversations with subject matter experts
Data Volume	Small with thousands of data points for both input and output
Data Structure	Real-world energy usage data simple time series tables across a handful of dimensions, with remaining datasets based on simple tables of probability distributions or lookup tables.
Data Quality	Energy usage data is of high quality, while the estimations and assumptions from subject matter experts of potentially lower quality.
Output	Cost tables and probability distributions of expected costs across various strategies, plus individual sensitivity analysis visualization

nities, define problems, conduct detailed analyses, and present outcomes of unique analytical situations, all aimed at informing stakeholders and providing decision-support frameworks to those on the front line of climate change. Business analytics can help organizations transition to a more sustainable and resilient future.

CHAPTER 2

TEXT-MINING CLIMATE CHANGE ADAPTATION RESEARCH FOR THEMATIC PATTERNS EMERGING IN THE 21ST CENTURY

2.1 Introduction

The rapid growth of ever-changing climate change knowledge presents society with meaningful challenges and opportunities (Wuebbles, Fahey, and Hibbard 2018). This changing knowledge, combined with deep uncertainty of climate change's future impacts, makes effective communication about potential mitigation and adaptation opportunities challenging. Such challenges impede society's ability to develop and implement appropriate climate change adaptation (CCA) strategies. In just the last few decades, the volume of research being conducted and published regarding CCA has exploded across numerous fields (Siders 2019). CCA research was a relatively new and emerging concept at the turn of the century (Smit et al. 1999) primarily concentrated in the environmental and natural sciences (Biesbroek et al. 2018; Haunschild, Bornmann, and Marx 2016). Since then, CCA research has systematically grown to encompass many research fields. Potential challenges associated with this rapid growth are the introduction of new frameworks, divergent word meanings, and fractured approaches to researching CCA. However, this substantial increase in the volume of published articles and expansion to include a broader range of research fields also opens up the possibility for emerging interdisciplinary research opportunities.

To adapt successfully to face climate change head-on, in a timely manner, the scientific community must effectively evaluate and understand the discourse within the research litera-

ture (Berrang-Ford, Pearce, and Ford 2015). Do the ideas and issues discussed in this growing CCA research provide thematic patterns within and across various research fields to better inform society and policymakers? This question exposes an opportunity to identify and describe themes that open new opportunities for interdisciplinary CCA research. However, given this rapid growth, traditional literature review techniques are challenged to efficiently and effectively digest this burgeoning body of published research (Asmussen and Møller 2019). Given the anticipated ongoing growth in the published research, there is an opportunity to apply text-mining techniques to evaluate this growing body of research (Debortoli et al. 2016). Such an analysis could shed light on the coherence of ideas, relevancy of issues, and approaches used between research fields over time (Lesnikowski et al. 2019). Using natural language processing, unsupervised learning, and data mining techniques, we can expose the topics and themes underlying the conversation in the CCA scientific literature during this period of rapid research growth. CCA subject matter expertise can then be applied to interpret the text-mining analysis results and tie the findings back via an informed understanding of CCA research.

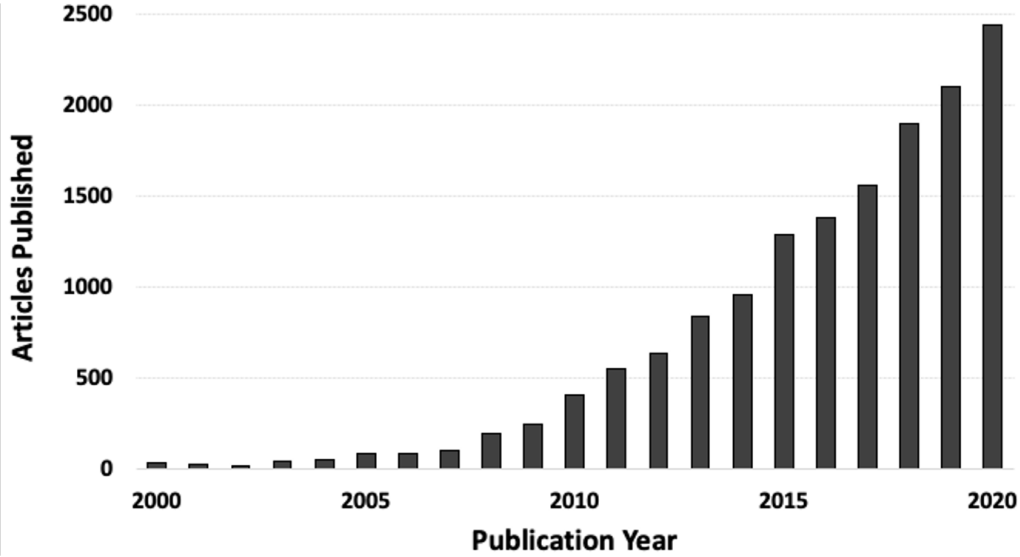
Given the insights above, we now discuss our research in the following sections. We start with a literature review covering the state of peer-reviewed CCA literature, research classification approaches, and available techniques for unsupervised topic modeling, along with our problem statement and expected contributions in Section 2.2. We then discuss in Section 2.3 the model development and methods applied to identify topics in the CCA literature. Section 2.4 covers the interpretation of our model output along with additional analysis of the topics. Finally, in Section 2.5, we cover the contributions and limitations of this study, along with the potential for future research opportunities, before wrapping up with our conclusions.

2.2 Literature review

2.2.1 Climate change adaptation research

There is a growing awareness of climate change, and its associated societal challenges (Nerlich, Koteyko, and Brown 2010). Increased risks and costs are associated with rising sea levels, increasing storm severity, worsening drought, increasing temperatures, and advancing species extinctions, all potentially leading to slow-moving disasters (Neumann et al. 2015). With growing awareness of this broad range of issues, the research on CCA has substantially expanded across many research fields in the 21st century (Haunschild, Bornmann, and Marx 2016). As shown in Figure 1, our search for CCA literature in the Web of Science (Clarivate Analytics 2020) results in more than 15,000 research articles published in peer-reviewed journals from the year 2000 to the year 2020. Only 41 articles from this search were published in 2000, predominantly from the environmental sciences. Contrast that with the nearly 2,500 published articles from more than 100 research fields in 2020. This substantial increase demonstrates the considerable growth in the CCA research published during that period.

Figure 1: Published articles by year in climate change adaptation research since 2000.



The interest in CCA research has ballooned to include research areas as diverse as Economics (Linnenluecke, Smith, and McKnight 2016; Jawid and Khadjavi 2019), Urban Studies (Chen 2015; Carter 2018; Douglas, Reardon, and Täger 2018), Law (Marjanac and Patton 2018; Hecht 2008), Energy (Solaun and Cerdá 2017; Metz, Darch, and Workman 2016; Chandramowli et al. 2016), Operations Research (Truong, Trück, and Mathew 2018; Chesney, Lasserre, and Troja 2017), Public Administration (Samaddar et al. 2015; Schlager and Heikkila 2011), Sociology (Smith, Anderson, and Moore 2012; Helicke 2019), Architecture (Pedersen Zari 2014; Solera Jimenez 2017), and Art (Inwood and Kennedy 2020). With such rapid growth across so many disciplines, researchers must be challenged to stay abreast of advances in CCA. To gain a sense of the breadth of research areas, Table 5 shows a breakdown of the top 25 research areas captured in CCA research found during our search of the Web of Science (Clarivate Analytics 2020). Environment-related science areas still dominate the coverage (e.g., Environmental Sciences, Environmental Studies, Meteorology Atmospheric Sciences), but one may also notice non-environmental areas with more than 100 publications in this timeframe (e.g., Economics, Public Administration, Energy & Fuels). As shown in Table 5, we can see the Web of Science categories for each article. However, we do not get a sense of the topics being discussed, nor do we see the potential interdisciplinary nature of the research dialogue.

2.2.2 Methods to categorize research content

Researchers have been attempting to categorize and synthesize research topics in published literature for many years (Krippendorff 2004). Methods used to identify themes and topics found in the literature vary by research area but include terms such as comprehensive literature review, formal systematic reviews, systematic mappings, narrative reviews, and research text coding (Berrang-Ford, Pearce, and Ford 2015; Okoli 2015). In line with these approaches, one methodology currently used in capturing advances in CCA is the systematic literature review

Table 5: Top 25 Web of Science categories in the climate change adaptation literature since 2000. *Note:* Some papers are tagged with more than one category.

Web of Science Category	**Articles**
Environmental Sciences	3750
Environmental Studies	2480
Meteorology & Atmospheric Sciences	1411
Ecology	1032
Water Resources	923
Green & Sustainable Science & Technology	630
Geosciences, Multidisciplinary	591
Geography	583
Development Studies	438
Economics	424
Forestry	411
Multidisciplinary Sciences	381
Biodiversity Conservation	374
Agronomy	323
Regional Urban Planning	316
Engineering, Civil	295
Agriculture, Multidisciplinary	280
Urban Studies	275
Engineering, Environmental	245
Public, Environmental & Occupational Health	237
Plant Sciences	232
Geography, Physical	220
Public Administration	163
Marine & Freshwater Biology	153
Energy & Fuels	141

(Berrang-Ford, Pearce, and Ford 2015). Systematic literature reviews are a useful way for researchers to summarize existing evidence, identify research gaps, and provide frameworks for positioning new research (Brereton et al. 2007). Similar to systematic literature reviews

are systematic mapping studies, with the critical difference being that in a mapping study, the research questions are more general and aim to classify research in a particular domain, rather than address particular research questions (Kitchenham, Budgen, and Brereton 2011). In either case, there is a heavy reliance on manual coding of texts, whether using a top-down, bottom-up, or dictionary approach to the classification (Debortoli et al. 2016). Regardless of the terminology used for synthesizing research topics covered in the literature, these manual categorization approaches face challenges. Some of these challenges include limited research population size, time-intensive manual reviews, significant time requirements from subject matter experts, and natural bias of the experts conducting the review (Asmussen and Møller 2019; Berrang-Ford, Pearce, and Ford 2015; Okoli 2015; Brereton et al. 2007). In essence, these manual approaches are supervised methods in classifying, synthesizing, or categorizing the existing literature on a given topic or research area.

With advances over the last few decades in information retrieval, computer processing speeds, parallel computing, and natural language processing algorithms, improved text analysis methods have made it possible to quantitatively analyze extensive collections of text (Asmussen and Møller 2019). Applying a data-mining framework, such as business standard CRISP-DM (Cross-Industry Standard Process for Data Mining) (Wirth and Hipp 2000) and pairing it with supervised or unsupervised text-mining methods, social scientists can evaluate substantial collections of text to identify patterns or topics (Ignatow and Mihalcea 2017). The social sciences are learning and applying these new tools that convert words to analyzable quantitative values using techniques such as Latent Semantic Analysis (LSA) (Deerwester et al. 1990), Latent Dirichlet Allocation (LDA) (Blei, Ng, and Jordan 2003), and Correlated Topic Modelling (CTM) (Blei and Lafferty 2006). This list is not exhaustive, as there are frequent text-mining developments, including the creation of new techniques and the introduction of enhancements to existing techniques. Each type of text-mining analysis has its advantages and disadvantages, so researchers must evaluate the research they will be conducting before choosing any particu-

lar method (Lee, Song, and Kim 2015). Researchers have employed these relatively new tools to convert extensive text into a quantitative representation they can then manipulate to develop meaningful insights such as themes, topics, and trends over time. Applying text-mining techniques to the expanding literature base mentioned in Section 2.2.1 has the potential to advance our understanding of CCA research (Sarma 2017).

2.2.3 Latent Dirichlet Allocation (LDA)

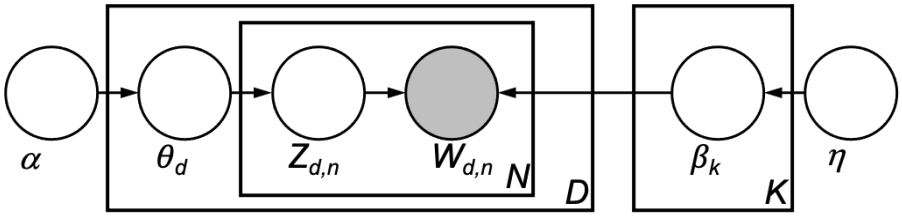
As highlighted in Section 2.2.2, there are various methods we could use for an unsupervised topic model analysis of the CCA literature. Given the exploratory nature of this analysis, we will be using the LDA methodology (Blei, Ng, and Jordan 2003) to explore and identify topics within the research literature. LDA has seen significant growth as a topic modeling method within academic research, business analytics, machine learning, and general online text analysis (Asmussen and Møller 2019; Jelodar et al. 2019). LDA modeling and analyses on academic literature has been done in research areas as diverse as Information Sciences (He et al. 2013), Accounting (Fang et al. 2018), Computational Linguistics (Hall, Jurafsky, and Manning 2008), Biology (Zheng, McLean, and Lu 2006), and Business & Economics (Piepenbrink and Nurmammadov 2015). In most studies of this type, the work was done in an exploratory manner in search of the underlying topics being discussed. However, there are also LDA-related studies that have been conducted for side-by-side comparisons with pre-existing manual coding schemes (Nelson et al. 2018) or conducted as the input to a separate manual categorization exercise using the LDA model topics (Grimmer and Stewart 2013). LDA is widely used and accepted, as noted by the substantial citation count (more than 43,000 on Google Scholar as of July 2022) associated with the original paper that introduced the LDA methodology (Blei, Ng, and Jordan 2003).

Given the extensive use of LDA in past studies, we now describe the methodology and underlying model structure. LDA ignores the order of occurrence of words and the syntactic

information, only treating documents as a collection of words, or more commonly known as a "bag of words" (Steyvers and Griffiths 2007). The basic premise underlying LDA is that each document is a mixture of topics and each topic is a mixture of words (Blei 2012). By finding the words in a topic, we can then use those words to identify the topics in a document. Breaking down LDA into its component words (Sharma 2020), the assumption is that there is a "Latent", or hidden, distribution of topics within a corpus of documents. These hidden distributions are composed of "Dirichlet" distributions for both the distribution of words to topics and topics to documents. Finally, the "Allocation" signifies that the topics are allocated across the documents, meaning each document is likely to have multiple topics covered within it. The general graphical structure of LDA is shown in Figure 2 (Blei 2012). Documents are treated as if they are generated from a random mixture of topics, and topics are seen as a probability distribution over the words. Each box (plate) in Figure 2 represents the sampling steps needed to achieve the number of samples in the lower right corner of the box (D = documents, N = words in a document, and K = topics). The term α can be thought of as the prior observation count (latent count) for the number of times a topic is sampled in a document. The term η can be thought of as the prior observation count (latent count) of the number of times words are sampled before any word from the corpus is observed (Steyvers and Griffiths 2007). More specifically, LDA assumes that documents are generated from the probabilistic processes as follows (Blei 2012).

1. For each topic k, draw from topic distribution $\beta_k \sim Dirichlet(\eta), k \in \{1, \ldots, K\}$
2. For each document d, draw topic proportions from $\theta_d \sim Dirichlet(\alpha)$
3. For each word n in each document d,
    - Draw a topic assignment from the topic proportions $z_{d,n}|\theta_d \sim Multinomial(\theta_d)$
    - Draw the word from the corresponding topic, $w_{d,n}|z_{d,n}, \beta_{1:K} \sim Multinomial(\beta_{z_{d,n}})$

Figure 2: Probabilistic graph model of Latent Dirichlet Allocation.
Source: Image from Blei 2012.



2.2.4 Application of LDA topic modeling to climate change adaptation research

The 2019 review "Frontiers in Data Analytics for Adaptation Research" highlights potential opportunities to apply topic modeling to the burgeoning corpus of CCA policy-making documentation (Lesnikowski et al. 2019). The authors demonstrate LDA application by evaluating speeches given at the United Nations Framework Convention on Climate Change and evaluating adaption efforts in Canadian city meeting minutes and staff reports. In line with the authors' call to action, LDA has also found traction in studying aspects of climate change in other document collections. Some examples include papers published on the identification of climate change bias in newspapers (Bohr 2020), finding the signals of climate change doubt in climate change denier organizations (Boussalis and Coan 2016), evaluating climate change communities and topic discussions on blogs (Elgesem, Steskal, and Diakopoulos 2014), studying the climate change discourse in industrial ecology literature (Dayeen, Sharma, and Derrible 2020), and assessing climate change technology transfer from US patents (Kulkarni 2020). In (Lesnikowski et al. 2019), the authors identify four critical areas where topic modeling could be applied to help inform the CCA conversation, specifically:

1. Analyzing framing and issue salience in adaptation discourse
2. Shedding light on the coherence of ideas, issues, and approaches to adaptation
3. Connecting adaptation's thematic patterns by applying topic models

4. Testing relationships between the content of texts and variables.

We believe the call to action in items 1-4 is just as applicable to assessing the corpus of CCA research given its explosive growth in the published literature over the last few decades (Siders 2019). Potential challenges associated with this rapid growth in research and publication can be various workstreams introducing new frameworks, fracturing word meanings, and causing divergent paths to research CCA. In line with the recommendations from (Lesnikowski et al. 2019), this rapid growth in the CCA literature provides us an opportunity to apply text-mining techniques to evaluate the whole body of research. Such an analysis could shed light on the coherence of ideas, relevancy of issues, and approaches used over time and between research fields.

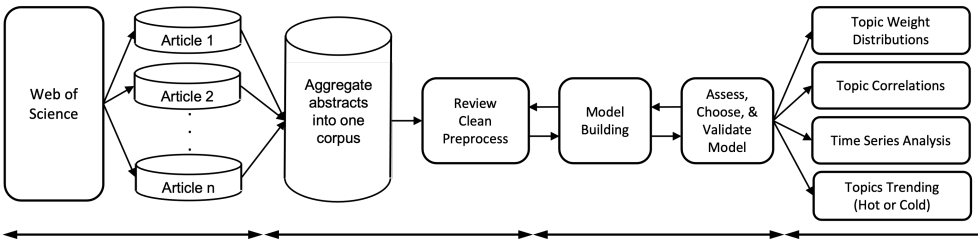
As an exploratory, descriptive analysis, this study uses the call to action in item 3, above, to gain insights into items 1 and 2 while laying the groundwork for future analysis that could test relationships as called out in item 4. We apply the LDA topic modeling approach to explore and evaluate a corpus built using research abstracts found on the Web of Science (Clarivate Analytics 2020; Li, Rollins, and Yan 2017). In doing so, we describe and evaluate the associated topics in the peer-reviewed literature published over the last two decades.

2.3 Model development and methods

We conduct our descriptive analytics effort with the four-phased approach adapted from the business standard CRISP-DM (Wirth and Hipp 2000), with relevant sections shown in Figure 3. We discuss these four stages in the following subsections. The first step is to conduct a comprehensive literature search to capture the relevant population of published research that forms our corpus (Section 2.3.1). The next stage entails data preparation to ready the entire corpus of texts for the LDA transformation and text-mining phase (Section 2.3.2). The third stage is composed of model development, where we explore the corpus and identify topics

using LDA analysis (Blei, Ng, and Jordan 2003) (Section 2.3.3), immediately followed by reviews from climate change experts to help name and validate the topics (Section 2.4.1). The final analysis stage explores the identified topics, including additional analyses looking at topic interactions and changes over time (Section 2.4.2.2).

Figure 3: Workflow for text-mining analysis adapted from CRISP-DM.
Source: CRISP-DM model from Wirth and Hipp 2000.



Study Analysis Phase	Data Collection	Data Preparation	Model Development	Analysis
CRISP-DM Analogue	Business and Data Understanding	Data Preparation	Modelling and Evaluation	Deployment

2.3.1 Data collection

The first stage of this study requires identifying the population of research papers to be included in our corpus. Even though we are not doing a systematic literature review, we conduct the initial phase of our data collection according to the best practices identified in (Okoli 2015). The first step is identifying and refining the appropriate search criteria to evaluate across as many disciplines as possible. As shown in Figure 1, there has been explosive growth in the volume and breadth of CCA literature since 2000; therefore, we choose 2000 to 2020 as the search coverage period. To collect the data, we searched the Web of Science (WoS) online search engine (Clarivate Analytics 2020). WoS is a highly effective database leveraged by thousands of large-scale data-intensive studies, typically involving activities cutting across diverse domains (Li, Rollins, and Yan 2017). WoS provides functionality to download up to 500 results at a time, which enables downloading all of the 67 available fields as shown in Table 7.

Table 6 shows the search criteria we apply during two distinct search phases. The search is limited to English language publications to minimize interpretability challenges. Our initial search focuses on CCA-related research by looking for synonymous words to "adaptation" located within three words of the term "climate change". As highlighted in Table 6, we utilize a wildcard with roots of words associated with adaptation (i.e., adapt*, resilien*, prepared*), which will capture terms such as resilience, preparedness, and adaptation. We limit our boolean search to the abstract, title, and keywords as the abstract comprises the text we include in our analysis, and the title and keywords should reflect the authors' perspective on the discussed topics. After reviewing the data from our first search with a subject matter expert, he postulated that in recent times researchers might be dropping the word "change" in their published research to avoid politically charged push-back from climate change deniers (Nerlich, Koteyko, and Brown 2010). Using that feedback, we conduct a second phase of the search using the keyword search shown in Table 6.

Table 6: Search criteria used to build corpus.

Criteria	**Description**
Database Source	Web of Science
Period Covered	Research published between 2000 - 2020
Text Searched	Title, abstract, and keywords associated with the article
Boolean keyword search (Phase I)	"climate change" NEAR3 (adapt* OR prepared* OR resilien*)
Boolean keyword search (Phase II)	"climate" NEAR3 (adapt* OR prepared* OR resilien*)

We conduct our search using the WoS interface to identify articles meeting the search term criteria. This download includes each full abstract needed to complete the LDA analysis. Additionally, as shown in Table 7, the WoS database contains other fields for each article available for the data prep and subsequent analysis, such as authors, title, keywords, journal, publication date, and research areas. Our search results include 10,525 documents in Phase I and

grows to 15,309 documents when adding the results of Phase II. We appreciate that conducting a comprehensive text-mining of each article's entirety has demonstrated better accuracy in some evaluations (Rezaeian, Montazeri, and Loonen 2017). However, there is also evidence that points to abstracts providing comparable coherence scores when the corpus and document sizes are sufficiently large (Syed and Spruit 2017). In Syed and Spruit 2017, the authors saw a significant difference in coherence scores when comparing LDA model runs for 4,417 full articles and again for their abstracts. When looking at LDA model runs for 15,004 articles, they found a negligible difference in coherence scores and human-ranked scores between models built with either the full article or the abstract. Therefore, given the expected volume and accessibility of abstracts, we choose the research abstract as a representative subset of text for our corpus development for all articles of interest.

2.3.2 Data preparation

The workflow we use for the data preparation stage is shown in Figure 4. As with any data analysis, we first need to evaluate the abstracts for any consistency, quality, or other issues that could be problematic (Chu et al. 2016). We initially profile the data, searching for blank abstracts, duplicate titles, and matching doi numbers. This initial review removes several hundred articles that show up as two entries due to being early-access and later published articles. We next visually inspect the abstracts for oddities, including leading or trailing information associated with the publication, such as "©2018 The Research Center for Eco-Environmental Sciences, Chinese Academy of Sciences. Published by Elsevier B.V." Other abstract clean-up includes removing citations, forced structure words like "Purpose" or "Aim", and hyperlink information to avoid these types of entries from skewing the actual abstract language. Finally, one unanticipated cleaning activity involves removing sections of non-English words that were in the abstracts, caused mainly by abstracts with both an English and non-English translation

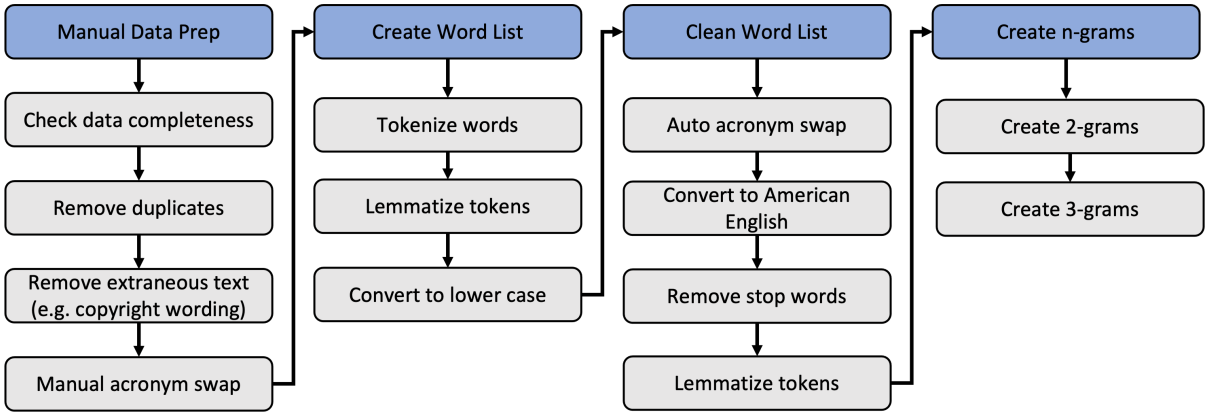
Table 7: Available fields from Web of Science. Fields of interest for current analysis highlighted and boldfaced.

Publication Type	Reprint Addresses	Volume
Authors	Email Addresses	Issue
Book Authors	Researcher Ids	Part Number
Book Editors	ORCIDs	Supplement
Book Group Authors	Funding Orgs	Special Issue
Author Full Names	Funding Text	Meeting Abstract
Book Author Full Names	Cited References	Start Page
Group Authors	Cited Reference Count	End Page
Article Title	Times Cited, WoS Core	Article Number
Source Title	Times Cited, All Databases	DOI
Book Series Title	180 Day Usage Count	Book DOI
Book Series Subtitle	Since 2013 Usage Count	Early Access Date
Language	Publisher	Number of Pages
Document Type	Publisher City	WoS Categories
Conference Title	Publisher Address	Research Areas
Conference Date	ISSN	IDS Number
Conference Location	eISSN	UT (Unique WOS ID)
Conference Sponsor	ISBN	Pubmed Id
Conference Host	Journal Abbreviation	Open Access Designations
Author Keywords	Journal ISO Abbreviation	Highly Cited Status
Keywords Plus	Publication Date	Hot Paper Status
Abstract	**Publication Year**	Date of Export
Addresses		

contained within the same abstract, such as that found in (Rondinini and Visconti 2015). There are 14,780 unique abstracts remaining after this data preparation stage.

While reviewing abstracts, it became apparent that researchers used acronyms inconsistently across the various research fields. For instance, when evaluating CCA literature, one might reasonably expect "CCA" to be an acronym for "Climate Change Adaptation", which is not always the case. Some uses of CCA are for word combinations like "Canonical Correspon-

Figure 4: Data preprocessing workflow.



dence Analysis" or "Crustose Coralline Algae". If left unaddressed, a topic model would treat CCA as having the same meaning within our corpus even though there are obvious differences across abstracts. We extract a list of all the acronyms in the corpus and then conduct a manual review of all acronyms that appear 25 or more times. During this review, we tag acronyms with all potential meanings. If an acronym has multiple meanings, but a majority with the same meaning, we manually convert the oddballs to their words within the abstract (e.g., we replaced "CA" with "California" where that was its meaning). In cases with no clear majority acronym, we evaluate and manually expand all instances of that acronym within the abstract (e.g., "ES" in Table 8). After this manual cleaning, acronyms with only one meaning are left to be expanded by the code in a later preprocessing stage. Finally, there are instances where changing the acronym is unnecessary because the acronym is consistently used. An example of this is the common usage of "EU" for the European Union. Table 8 shows some examples of each of these acronym instances.

The next few steps include preprocessing the corpus of text to be machine-readable for natural language processing (Ignatow and Mihalcea 2017). For this next preprocessing stage, we use "spaCy", a free, open-source library for advanced Natural Language Processing (NLP) in Python. The spaCy API is available online at https://spacy.io (Vasiliev 2020). First, we

tokenize, identifying words from the string of characters and returning a list of words without punctuation. We also lemmatize the tokens, identifying appropriate punctuation like contractions and abbreviations in the text. To execute this stage, we call the spaCy lemmatization function. First, identifying the part-of-speech of each word in an abstract and then tokenizing that word using its proper base form using lemmatization. We only keep nouns, verbs, adverbs, adjectives, or proper nouns. We keep the proper nouns in this phase due to the many acronyms that will be converted to their proper words in the next step. Finally, as part of conducting the lemmatization, all tokens are made lower-case to ensure the same words would not be treated as two different tokens due to different capitalizations. The output of this stage is a list of tokens representing each abstract.

The next few preprocessing steps are needed to clean up the text for consistency. It is at this point where we automatically swap out remaining acronyms that have only one meaning of their actual underlying words (e.g., "Greenhouse Gas" in Table 8). Next, we sweep through the list of words again, looking for words with British spelling and swapping them with their American spelling. Some examples of this include "behaviour" changed to "behavior" or "grey" changed to "gray". To facilitate these word swaps, we use the online dictionary of British to American translations located online at https://github.com/hyperreality/American-British-English-Translator (Hyperreality 2016). We then remove stopwords, which are high-frequency words such as pronouns (e.g., *we*, *them*), determiners (e.g., *an*, *the*), and prepositions (e.g, *in*, *on*, *of*) (Ignatow and Mihalcea 2017). We do this using the Natural Language Toolkit (NLTK) and its built-in stopword dictionary (Bird, Klein, and Loper 2009). We then clean up the lists of words for each abstract by removing any words with one or two characters. In most cases, these words do not actually add value to the actual topics being discussed (Resch, Usländer, and Havas 2017). Finally, we rerun the lemmatization one more time for two reasons. The first is to remove proper nouns based on their part-of-speech. The second is to account for

any remaining transformation of words to their bases (e.g., any plurals remaining are turned into the singular form).

Table 8: Example of acronyms in the corpus and methods used to address each example.

Acronym	Count	Instances	Meanings	Correction
CA	194	4	Conservation Agriculture	Auto/Manual
			Cellular Automata	
			California	
			Central Africa	
ES	81	4	Ecosystem Services	Manual
			Environmental Sustainability	
			Environmental Stewardship	
			Expected Shortfall	
CC	189	3	Climate Change	Auto/Manual
			Canopy Cover	
			Correlation Coefficient	
GHG	350	1	Greenhouse Gas	Auto
EU	221	1	European Union	No Change

The final data preprocessing step is to conduct collocation identification, which entails identifying sequences of words (n-grams) that have special meanings (Mikolov et al. 2013). In this research, some examples of likely n-grams are "greenhouse_gas_emission" or "decision_maker". To make n-grams, we start with 2-grams, then rerun the collocation algorithms with the tokens (including 2-grams) as the input to the 3-grams. This is where we introduce the Gensim Python package (Rehurek and Sojka 2011), which is commonly used for conducting LDA modeling. To do the collocation, we use the native functionality of Gensim, specifically the "gensim.models.phrases" functionality, with settings of 100 for the minimum number of occurrences to create an n-gram and a threshold setting of 100. The threshold setting is used by Gensim to determine the ease with which n-grams are allowed to be created. The higher the setting, the harder it is for n-grams to be created. We create 609 n-grams, with the top

10 (based on count in abstracts) shown in Table 9. These representative n-grams seem appropriately linked to CCA, except for potentially "take_account", which frequently occurs with phrases such as "taking into account" and "take account of" found within many abstracts. Upon completion of n-gram creation, we save two different lists of the corpus, one without n-grams and one with both 2-grams and 3-grams. We keep these different versions in order to later train LDA models on both versions while looking for results with the highest coherence scores and reader interpretability.

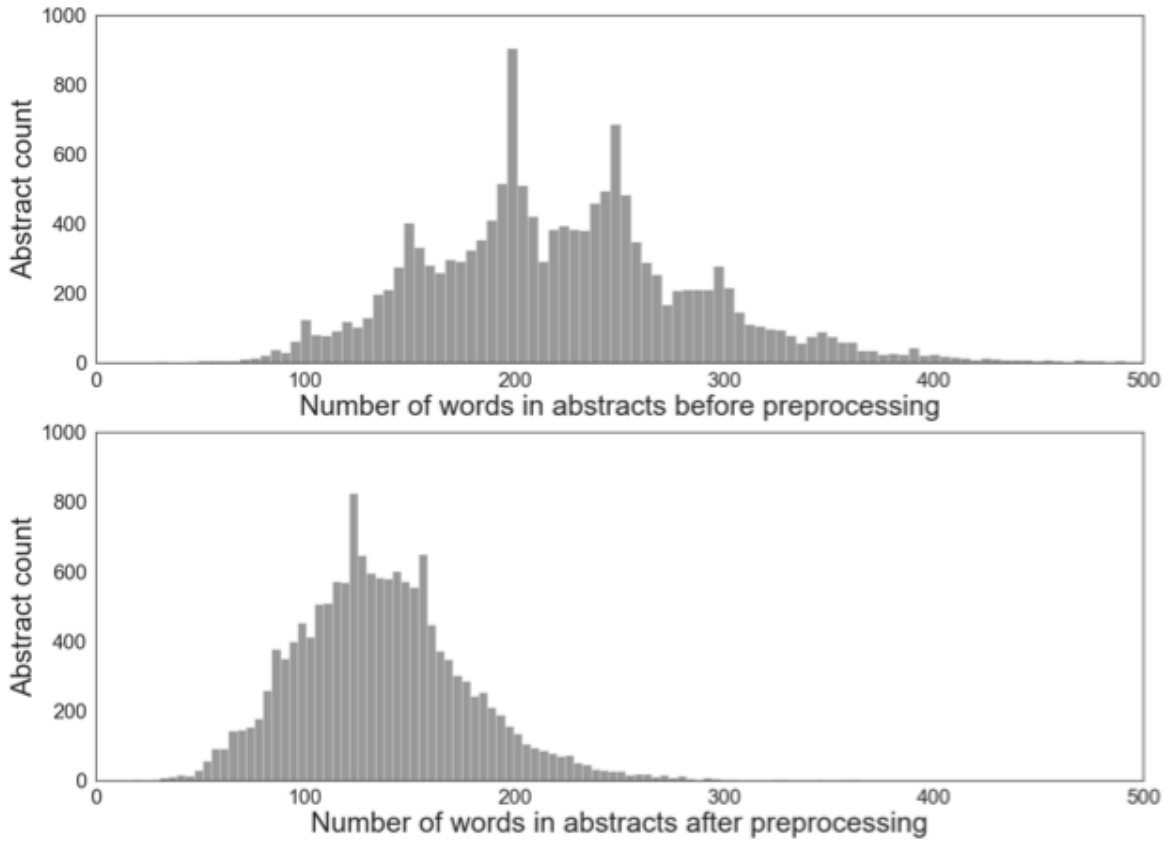
Table 9: Count of top 10 n-grams in corpus and the number of abstracts containing n-gram.

n-gram combination	**n-gram count**	**count of abstracts**
long_term	1908	1519
food_security	853	598
socio_economic	737	553
decision_maker	599	492
greenhouse_gas_emission	689	469
policy_maker	492	441
short_term	438	371
take_account	385	365
greenhouse_gas	469	322
representative_concentration_pathway	550	295

Figure 5 shows the original number of words in the abstracts compared to the number of words upon completion of the data preprocessing. One sidebar note of interest in Figure 5, the word counts of the abstracts before preprocessing show observable spikes at 100, 150, 200, 250, and 300. These spikes are likely due to authors meeting word count limits in their abstracts that align with these multiples of 50. At the completion of data preprocessing, we have an average reduction of 38.9% in the number of words per abstract. This sizeable reduction is expected for two reasons. First, we remove stop words that commonly represent 25% to 30% of words in English language documents (Schütze, Manning, and Raghavan 2008). Additionally,

the lemmatization process reduces the number of unique words by reducing the variations of specific words (e.g., plural words converted to singular). Upon completing the preprocessing phase, our corpus consists of 35,946 distinct tokens across the 14,780 abstracts.

Figure 5: Distribution of abstract word counts versus token counts available after pre-processing abstracts.



2.3.3 Model development

During this phase of the analysis, we conduct a topic modeling analysis using the LDA technique (Blei, Ng, and Jordan 2003) to identify topics, specifically using the Gensim (Rehurek and Sojka 2011) Python package (version 3.8.0. with Python version 3.8.5). LDA Analysis requires providing the model with a set number of topics to start the algorithm; however, we

do not know the number of topics apriori. Therefore, we explore a range of potential topics, iteratively running the LDA analysis on a potential range (e.g., 3 - 30) and assessing the best number of topics. Finding the best number of topics can be debated. However, there are qualitative methods and quantitative measures to assess LDA models to help identify a reasonable number of topics, such as coherence and perplexity (Greene, O'Callaghan, and Cunningham 2014; Röder, Both, and Hinneburg 2015). This section on model development walks through how we evaluate potential models before arriving at our best model.

2.3.3.1 Parameter tuning

As discussed in Section 2.2.3, the LDA model requires selection of settings and hyperparameters as shown in Table 10. For the first three hyperparameters in Table 10, we use the recommendation from Hoffman, Bach, and Blei (2010) for the best settings of offset ($\tau_0 = 64$), decay ($\kappa = 0.5$), and batch size ($S \geq 256$). For α and η described in Section 2.2.3, Gensim has three potential settings: symmetric, asymmetric, and auto. One additional alternative is to provide the settings in a vector. However, providing a vector requires knowing expected values for the α and η settings apriori, something we do not know. Therefore, we use the "auto" setting, which allows the Gensim model to automatically seek out and converge on favorable settings as it runs through multiple iterations. The downside of using "auto" is that it takes longer for the model to run than the symmetric or asymmetric settings.

Additional settings include the number of "passes" and "iterations" Gensim applies when creating the model (Hoffman, Bach, and Blei 2010). The "iterations" setting represents a max limit of how many times the algorithm may repeat each document's probability distribution assignments, thereby imposing an upper cap on how long the process will run if all documents have not achieved convergence. The setting "passes" is the number of times the algorithm trains the model on the entire corpus. To determine appropriate settings for passes and iterations, we ran multiple versions of the model process while recording every pass (i.e., "eval_every" = 1)

to determine how quickly the model was converging. After conducting multiple logging runs on topic models ranging from 7 to 30 topics, we chose values of 40 passes and 100 iterations based on where the convergence, coherence, and perplexity scores leveled out. In Figure 6, we show an example of the per pass convergence of coherence scores when creating models with 8 and 9 topics, each with lower limits on word presence in 3 or 5 abstracts. Finally, to enable comparison between runs, we choose a random seed (i.e., "random" = 42) to set a constant start point for model runs.

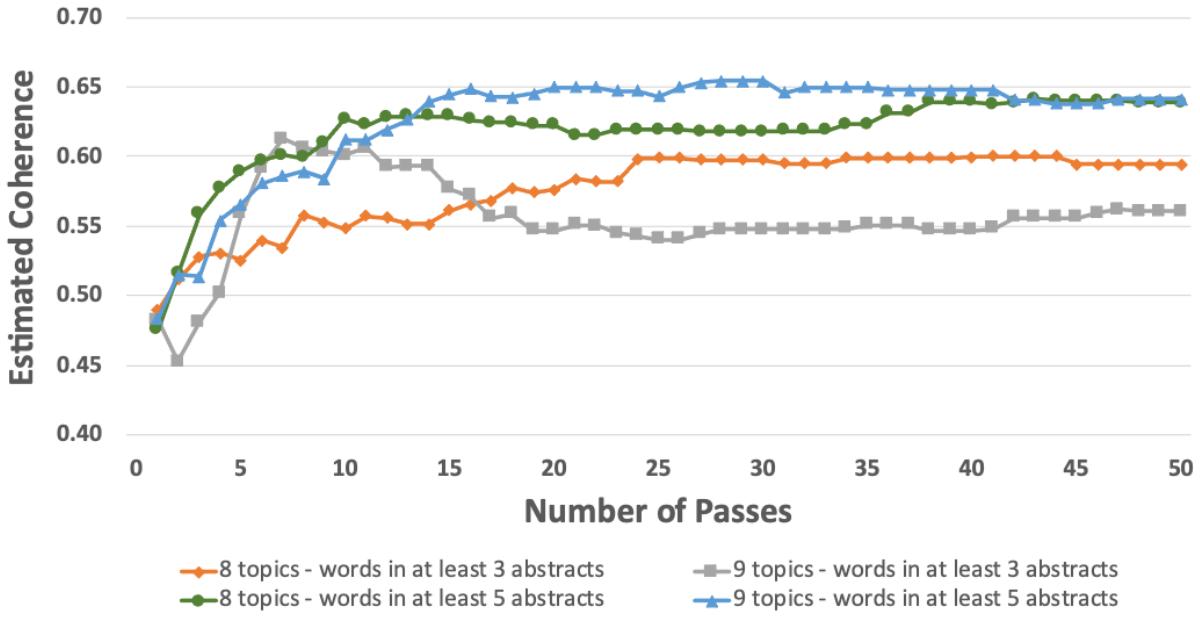
Table 10: LDA model settings and hyperparameters used when training LDA models with Gensim.

Parameter	**Setting**	**Details**
κ	0.5	decay recommended from (Hoffman, Bach, and Blei 2010)
τ_0	64	offset recommended from (Hoffman, Bach, and Blei 2010)
S	256	batch size recommended from (Hoffman, Bach, and Blei 2010)
α	auto	allow algorithm to determine value
η	auto	allow algorithm to determine value
eval_every	1	used for assessment of model runs
passes	40	selected based on model convergence in logging runs
iterations	100	selected based on model convergence in logging runs
random	42	random seed used to allow repeatability

Highlighted by bidirectional arrows in Figure 3, developing the model requires an iterative process between data preparation, model building, and model assessment. With each iterative run, we learn more about the data and the potential topics in the corpus. For instance, we create our first LDA models before addressing the acronyms sprinkled throughout the text. Upon evaluating these first models, it was clear that acronyms are not being treated accordingly with the underlying words that defined the acronyms. The next round of model building helps us determine a tighter range for the number of potential topics to explore, reducing our initial range from 3-30 down to 7-21. We make this assessment based on models with ≥ 22 topics seeming to have quite a few nonsense or non-interpretable topics and models with ≤ 6 topics

lacking sufficient topic granularity. That leads us to the next section, where we conduct many model runs with changes to a handful of key parameters to assess the most interpretable models.

Figure 6: Estimated coherence from convergence logging runs of initial test models when varying topic count and infrequent word settings.



2.3.3.2 Creating full range of models

To identify candidate models, we conduct Gensim model runs while varying the parameters shown in Table 11. The first parameter we vary is the number of topics ranging from 7 to 21. Next, we choose the corpus from either the simple unigram or the combined n-grams (specifically both bi- and tri-grams). We use both because we do not know in advance if one will be better than the other. The following two parameters we change are related to the pruning of frequently and infrequently used words (Grimmer and Stewart 2013; Maier et al. 2018). We provide the model with three cut-points for the frequently used words, removing words that appear in more than 30%, 20%, and 15% of the abstracts. Frequently used or ubiquitous words within the corpus will not help distinguish clear lines between the topics. We also use

cut-points for words that show up in no more than 5, 10, or 15 abstracts for the infrequently used words. Infrequently used words appearing within only a tiny subset of abstracts will not help identify the discussion within the corpus. Additionally, removing infrequent words also reduces the size and sparsity of the document-term matrix, thereby reducing overall model runtimes. When looking at all possible combinations of these parameters, there are 270 unique models we build with Gensim. Each model takes approximately 20-30 minutes to train and build with our established parameters in Table 10. To expedite the creation of the 270 models, we use a multicore instance of an Amazon Web Services (AWS 2021) server to create our models in parallel, saving all results for later analysis and assessment.

Table 11: Parameters for running multiple models for assessment.

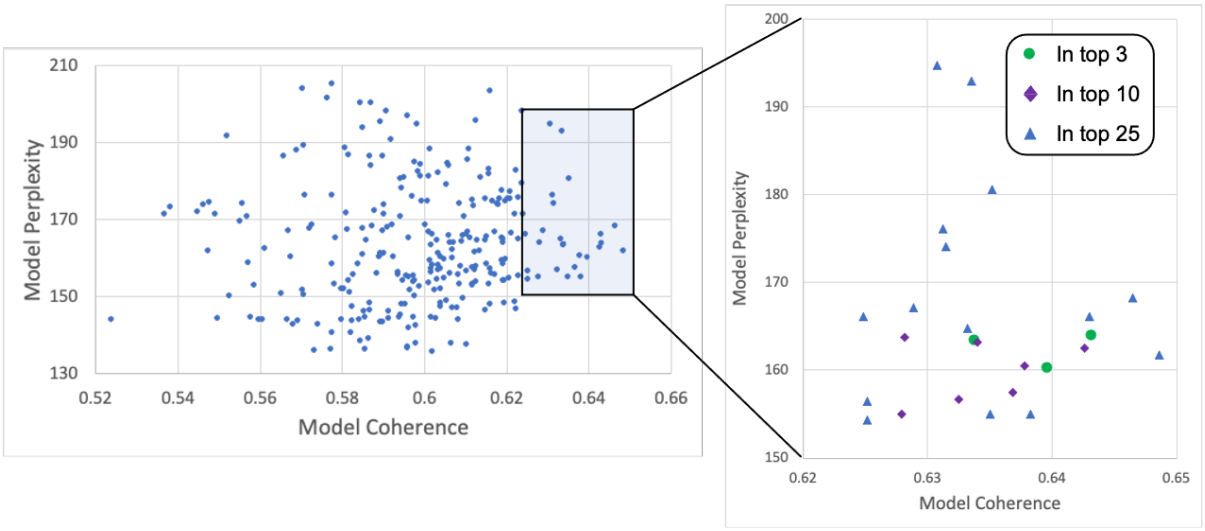
Parameter	**Values**	**Combinations**
Topics	7 - 21	15
Corpus	unigram vs. n-grams (n = 2 and 3)	2
Frequent words	words in no more than 15%, 20%, or 30% of abstracts	3
Infrequent words	words in at least 5, 10, or 15 abstracts	3
	Total combinations $(15 \cdot 2 \cdot 3 \cdot 3)$	270

2.3.3.3 Assessing full range of models and topics

After completing the 270 model runs, we assess the models based on coherence and perplexity scores to look for models with interpretable topics based on these quantitative measures. To determine the coherence for each model, we use the C_v measure developed and described in (Röder, Both, and Hinneburg 2015) and provided for in the Gensim package. The C_v coherence provides values in the range of 0.0 to 1.0, with larger coherence numbers ideally relating to more interpretable topics. Likewise, lower values for perplexity typically represent better generalization performance (Hoffman, Bach, and Blei 2010). Figure 7 shows a scatter plot

of all 270 models' coherence and perplexity scores. To screen for representative models, we initially limit our review to models with the top 25 coherence scores as shown in Figure 7. We review each of these top 25 models using the pyLDAVis tool discussed in Sievert and Shirley (2014), with an example shown in Figure 8. The pyLDAVis interactive tool provides a concise capability to visualize the most relevant words within each topic, their distribution across topics, and the relative distance of topics when reduced to a two-dimensional (2D) space. To assess the models, we visually inspect each of the top 25 models looking at multiple elements of pyLDAVis. Specifically, we look for a good breakdown of topics without any single topic thoroughly dominating the space (i.e., the area of the topic bubble) and a distributed spread of topics throughout the 2D space (topics separated and filling the entire space). We also evaluate if the top 10 tokens associated with each topic were semantically similar. Finally, where topics overlapped, we evaluate if the overlaps made sense semantically.

Figure 7: Comparison of model coherence and perplexity scores for the 270 LDA models, with inset showing breakdown of screening results to get to top 3 models.



Following our initial screening of the 25 models, we narrow the field down to 10 models for deeper inspection. We expand this review to inspect the 30 most relevant tokens in the topic,

with relevancy determined by setting the pyLDAVis λ value to 0.6 as recommended in Sievert and Shirley (2014). Of note, as shown by the green circles in the inset graph in Figure 7, the top 3 models do not have the highest coherence or the lowest perplexity scores, but they are still strong contenders based on the pyLDAVis review. By assessing multiple quantitative measures (coherence and perplexity) and visualizing the model with pyLDAVis, our assessment methodology mitigates the limitations of choosing the most representative model solely by quantitative scores. As shown in Table 12, we narrow the model selection to three viable candidates to be assessed by two CCA subject matter experts in more detail in Section 2.4.1. Figure 8 shows the pyLDAVis visualization for one of the top three candidate models, specifically the model that contained 16 topics.

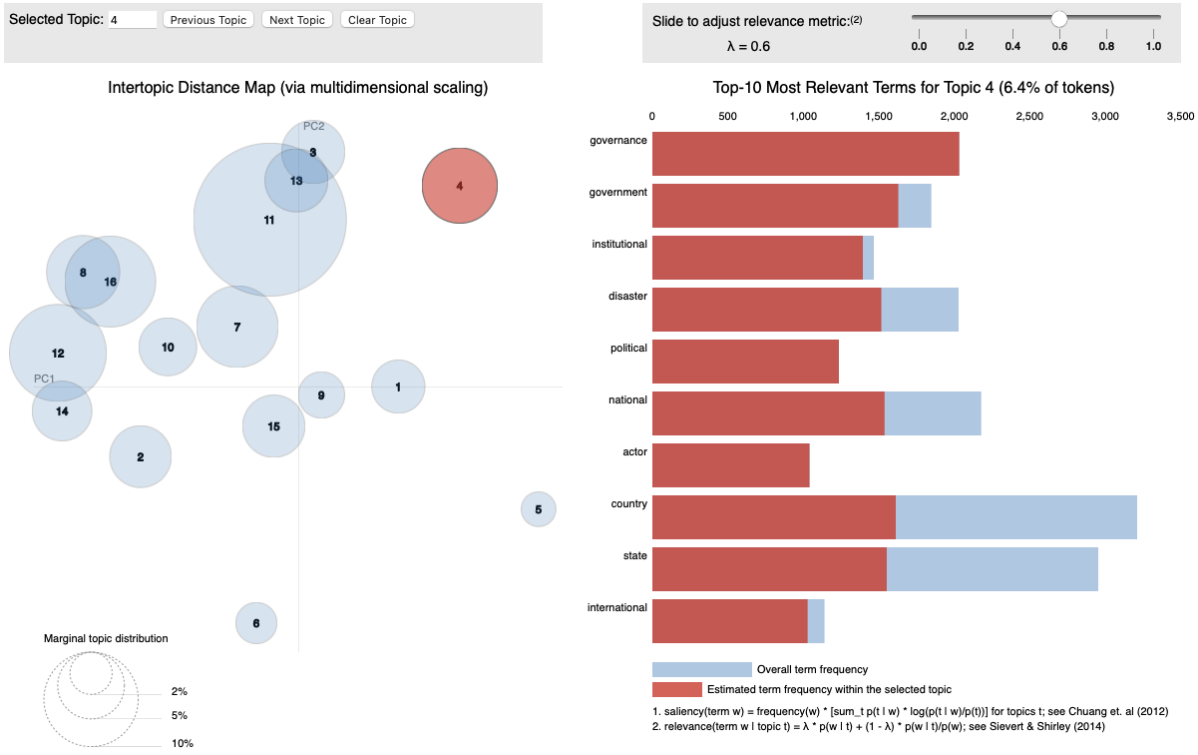
Table 12: Parameters and scores for the three candidate models.

Parameter	**Model 1**	**Model 2**	**Model 3**
Number of Topics	12	15	16
Corpus	unigrams	n-grams	unigrams
Frequent Words Cutoff	>15%	>15%	>15%
Infrequent Words Cutoff	10 abstracts	10 abstracts	10 abstracts
Perplexity	164.0	160.2	163.5
Coherence	0.643	0.640	0.634

2.4 Results

With the successful identification of three viable topic models, we begin working with two subject matter experts (SMEs) in climate change. Their role is to review the candidate models, label the topics, and assess each model's interpretability. Below we discuss this SME model review process, topic labeling, and semantic validation. Finally, after selecting the final model, we then explore its topics, topic correlation, and topic weighting changes over time.

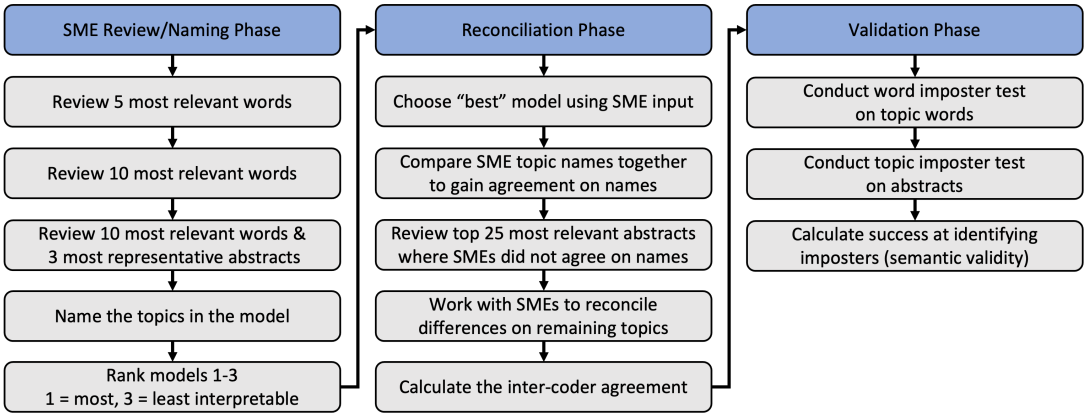
Figure 8: Visual representation of 16 topic LDA model in PyLDAViz with one of the topics highlighted. *Source:* Tool from Sievert and Shirley 2014.



2.4.1 Incorporation of domain expertise

Topic models do not explicitly label topics; instead, they probabilistically weight lists of words belonging to each topic. In an exploratory topic modeling analysis, best practices highlight the need for domain expertise in evaluating model output (DiMaggio, Nag, and Blei 2013). We do so in this research with the assistance of two CCA SMEs who identify the underlying theme for each topic and then assign topic names based on domain knowledge. They use the most relevant words per topic and the most representative abstracts to assign names to the topics. This model validation exercise takes place in three stages, as shown in Figure 9, with each stage discussed in the following sections.

Figure 9: Workflow for Subject Matter Expert topic review, naming, reconciliation, and validation.



2.4.1.1 SME review/naming phase

In the first phase of the SME review, we introduce the SMEs to the three candidate topic models described in Section 2.3.3. The SMEs are provided with a breakdown of each model that includes the ten most relevant words and the three most representative abstracts for that topic. Word relevance as defined in (Sievert and Shirley 2014) is a weighted score of the term's probability and its lift. In line with the findings of (Sievert and Shirley 2014), we use a weighting value of $\lambda = 0.6$ to determine word relevance. Given that no single topic makes up 100% of any abstract, we use the three abstracts in the corpus that have the highest weighting of that topic to find the most representative abstracts. In most cases, these weights are in the 0.70-0.90 range, meaning the preponderance of topic distribution within the abstract is for the topic of interest. In some cases, some topics do not represent a majority of the topics in any one abstract; instead, the topic has the largest percentage weight for that abstract (e.g., topic weight is 0.45).

As shown in Figure 9, we ease the SMEs into the topic naming work in three steps. The first step is to look at the top five most relevant words and assign a topic name. Next, they are given the following five most relevant words and tasked with naming the topic based on the

ten most relevant words. Finally, the SMEs look at the three most representative abstracts for each topic and write down their final name for each topic. Upon completing this exercise, the SMEs then tag each topic within a given model with a ranking of 1 through 3, with 1 = easy to interpret, 2 = moderate to interpret, and 3 = hard to interpret. The SMEs conduct these steps for all three models in our candidate pool. After the SMEs finish, we evaluate two components of their topic rankings for each model. The first component is the SMEs' average topic interpretability rankings for each model. This score indicates the relative, subjective interpretability for each model. The second component is a per-topic difference of SME interpretability rankings between the two SMEs. This score indicates SME consistency of ease of interpretation. The scores provided by the SMEs results in Model 2 being seen as the least interpretable, and Model 3 as the most interpretable.

2.4.1.2 Reconciliation phase

The next step is to reconcile the SMEs' Model 3 results with our candidate model in hand. As expected, when working in isolation, the SMEs come up with similar yet somewhat different topic names or difficulty of topic interpretability. The name differences are often semantically minor, but joint agreement on the topic label is still necessary. For example, one SME named a topic "Carbon Storage in Soils" while the other called it "Land Carbon Storage". After further review, the SMEs agreed with "Soil Carbon Storage" as the topic name in this instance. Several other examples include "Climate Attitudes and Behaviors" versus "Behavior and Adaptation", which became "Attitudes and Behaviors", and "Genetics and Climate Adaptation" versus "Genetic Impacts", which became "Climate-related Genetics". After this initial review round, the SMEs agree on 12 of 16 topic names for an initial naming agreement of 75%. The rows tagged as Round 1 in Table 13 include the 12 resolved at this stage.

In instances where we do not gain initial agreement on the remaining four topics, we expand our review to the 25 most representative abstracts and look for language that might help close

the gap between the two SMEs. At this stage, the goal of expanding the review to the 25 abstracts is to look for commonalities on the topic that may not be readily discernable with just the top three abstracts. An example of one of these topics had one SME initially naming it "Climate Adaptation in African Agriculture" while the other SME named it "Improving Farmer Household Food Security". For further review of this topic, specific questions we have when looking at the expanded list of abstracts are: 1) Is this topic limited to African Agriculture, 2) Is there an element of Household Food Security throughout the abstracts, and 3) Is the topic more generally about Agriculture in the Developing World? This deeper review enables the SMEs to reconcile differences from their initial reviews, allowing us to gain inter-coder agreement on all 16 topics for agreement on 100% of topic names. Table 13 includes each topic's name, the ten most relevant words, which round of reconciliation it was finalized, and a short name used for ease of further analysis.

2.4.1.3 Validation phase

The final topic modeling step is a validation process that entails identifying word intrusion and topic intrusion tasks (Chang et al. 2009) conducted by the SMEs, researchers, and research assistants. The word intrusion task measures the topic coherence gauged by identifying an intruder word that does not belong within a given topic. More specifically, we present a list of six words for each topic, where five of the words are the five most relevant words for the topic plus one intruder word. We select the intruder words from a pool of words with low relevancy for the topic but high relevancy for one of the other topics. We select the intruder this way to avoid the word being rejected outright due to being a rare word (Ying, Montgomery, and Stewart 2019). We shuffle the six words such that the intruder shows up randomly in the list of six words. If the reviewer can identify the intruder word in the list, that implies good semantic coherence. If the reader cannot identify the intruder word, that suggests they had to guess due

to poor semantic coherence of the topic. The word lists we use for the word intrusion task are shown in Table 14.

The second intrusion task entails presenting reviewers with an abstract and four lists of the five most relevant words for four topics potentially in the abstract. Three topics are the highest probability-weighted topics for the abstract, while one topic is chosen randomly from the other low-probability topics for the abstract in the model. In this task, the reviewers need to identify the intruder topic with the expectation that by selecting the proper topics, the reviewers are verifying that the document-topic assignments are valid. Conversely, if the reviewers choose the wrong topic, then the assignment of the topics to the abstract may be poor matches. Figure 10 shows one topic intrusion example, with Topic 2 being the intrusion topic in this example.

For the word intrusion task, the seven reviewers are able to identify 101 of 112 intruder words, resulting in a precision of 90.2% (Chang et al. 2009). Only one of the seven reviewers identified the intruder word for Topic 12 ("Modelling Rainfall and Temperature Variability"), accounting for more than half of the incorrect guesses. In evaluating the topic intruders, the five reviewers are able to identify 87 of 100 topic intruders for precision of 87.0% (Chang et al. 2009). There were no readily identifiable patterns in the incorrect guesses for the topic intrusion task. Evaluating the inter-rater reliability, we find strong agreement with a Fleiss Kappa (Fleiss 1971) of 0.631 for the word intrusion task and 0.708 for the topic intrusion task. The strong inter-rater reliability and high detection percentage of intruders lead us to believe our model validly represents the topics contained within the CCA literature.

2.4.2 Analysis

2.4.2.1 Topic weights distributions

Our first analysis is to look at the distribution of topics. Figure 11 shows a histogram for the topic weights distributed within the corpus for those abstracts including the topic (i.e., topic

Table 13: Topics identified and named following Subject Matter Expert review of models.

Topic Name	Short Name for Analysis	Round	Ten Most Relevant Words in Topic
Forest Ecosystems	forest_ecosystems	1	forest, tree, ecosystem, habitat, site, fire, conservation, native, distribution, landscape
Climate-related Genetics	genetics	1	genetic, trait, thermal, gene, variation, selection, tolerance, cold, evolutionary, breed
Attitudes and Behaviors	attitudes_behaviors	1	health, perception, survey, perceive, school, behavior, education, respondent, public, participant
Soil Carbon Storage	soil_carbon_storage	1	soil, land, carbon, cover, surface, organic, vegetation, air, storage, measurement
Urban and Greenspace Environments	urban_and_greenspace	1	urban, city, green, infrastructure, building, build, design, urbanization, space, construction
Finance-related Issues	finance_issues	1	service, market, finance, insurance, financial, industry, private, investment, provision, business
Plant-related Impacts	plant_impacts	1	plant, leaf, treatment, root, growth, stress, fruit, content, flower, arctic
Government and Politics	govt_politics	1	governance, government, institutional, disaster, political, national, actor, country, state, international
Mitigation Economics and Markets	mitigation_econ_markets	1	energy, emission, greenhouse, gas, cost, mitigation, consumption, demand, carbon, country
Extreme Weather Events	extreme_wx_events	1	extreme, event, drought, flood, weather, heat, frequency, mortality, damage, disease
Coastal Effects	coastal_effects	1	coastal, sea, rise, island, ocean, marine, pacific, wetland, storm, coast
Freshwater Effects and Management	freshwater_effects_mgmt	1	river, basin, flow, groundwater, lake, hydrological, runoff, supply, catchment, stream
Frameworks, Planning, and Decision Support	frameworks_planning_dec_sup	2	framework, context, information, knowledge, assessment, tool, address, integrate, planning, exist
Modelling Rainfall and Temperature Variability	rainfall_temp_variability	2	precipitation, degree, period, rainfall, cultivar, trend, annual, season, wheat, variability
Farmers and Agriculture in Developing World	farmers_and_ag	2	crop, yield, farmer, agricultural, production, food, agriculture, farm, rice, maize
Rural Households and Livelihood Vulnerability	household_livelihood_vuln	2	vulnerability, household, livelihood, rural, vulnerable, indigenous, tourism, migration, gender, people

Table 14: Word intrusion list for validating semantic coherence. Intruder word is not disclosed to test subject until completion of task.

Topic	**Shuffled Word List for Topic**	**Intruder Word**
Topic1	city, green, building, cyclone, urban, infrastructure	cyclone
Topic2	group, soil, land, carbon, surface, cover	group
Topic3	perceive, survey, health, school, cereal, perception	cereal
Topic4	disaster, government, institutional, political, governance, vegetable	vegetable
Topic5	market, dam, service, financial, finance, insurance	dam
Topic6	net, basin, flow, river, groundwater, lake	net
Topic7	production, financial, yield, farmer, crop, agricultural	financial
Topic8	variation, gene, genetic, thermal, winter, trait	winter
Topic9	island, coastal, sea, ocean, rise, winter	winter
Topic10	event, flood, weather, extreme, dioxide, drought	dioxide
Topic11	downscale, assessment, framework, information, knowledge, context	downscale
Topic12	evaporative, degree, cultivar, rainfall, precipitation, period	evaporative
Topic13	household, vulnerable, vulnerability, river, rural, livelihood	river
Topic14	electricity, treatment, root, growth, plant, leaf	electricity
Topic15	gas, greenhouse, emission, cost, energy, enzyme	enzyme
Topic16	efficacy, ecosystem, forest, tree, habitat, site	efficacy

weight > 0 in a paper). Given that there are 16 topics, and most topics appear in only a subset of documents, we see that all topics skew towards zero weight. This visual makes sense as most papers generally have three or four meaningfully weighted topics, with most topics appearing in fewer than half the papers (i.e., topic weight = 0). All but three topics are present in at least 25% of abstracts.

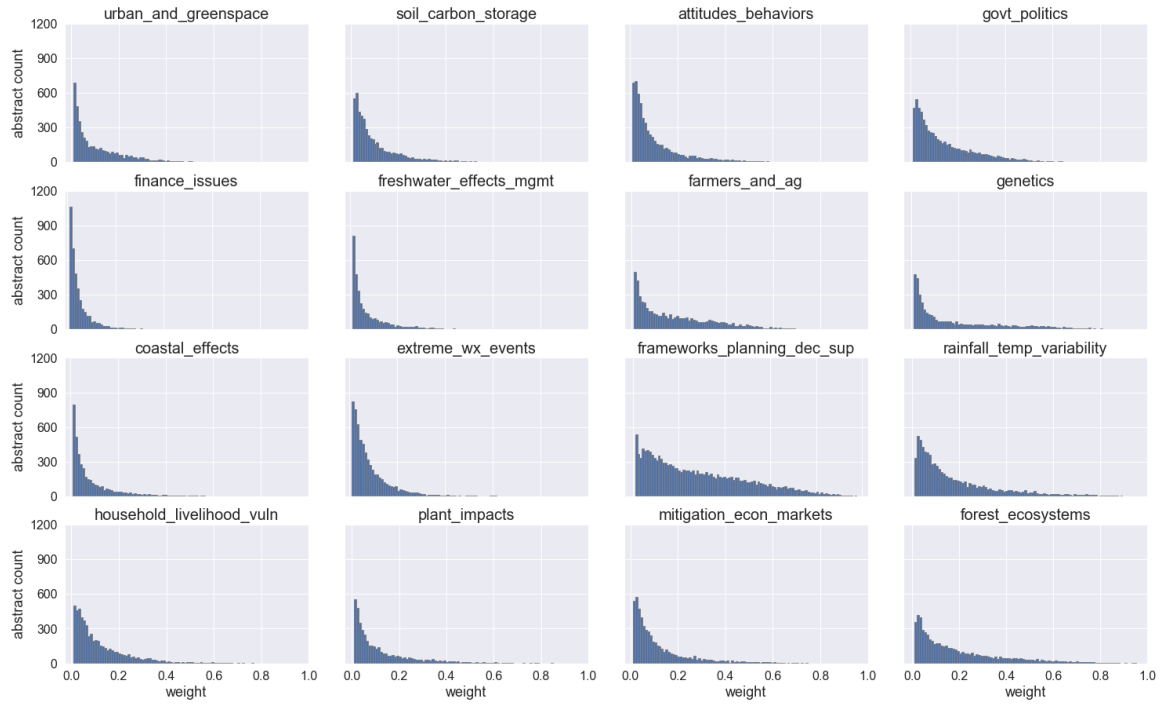
These histograms show that a single topic rarely dominates an article, given how few bars show up on the right-hand side of each histogram. One histogram that stands out is frameworks_planning_dec_sup. Because this topic appears in all but 163 abstracts, its histogram is denser and not skewed as heavily to the left. The preponderance of frameworks_planning_-

Figure 10: Example of topic intrusion task for validating semantic coherence.

Abstract #	Topic #1	['forest', 'tree', 'ecosystem', 'habitat', 'site']
8	Topic #2	['coastal', 'sea', 'rise', 'island', 'ocean']
	Topic #3	['precipitation', 'degree', 'period', 'rainfall', 'cultivar']
	Topic #4	['crop', 'yield', 'farmer', 'agricultural', 'production']
	Abstract	Changing climate is likely to impact on both tree species and agroforestry systems in a variety of ways. A multi-model ensemble approach based on ecological niche modelling was used to understand the impact of climate on distribution of agroforestry trees in Yunnan Province of China. Future changes in distribution of 10 agroforestry tree species were projected using an ensemble of climate projections derived from the results of 19 Earth System Models provided by the Coupled Model Inter-comparison Project-Phase 5. Our model explained suitable habitat, and identified potential locations for mixed agroforestry using selected species. The model suggested west and southwest Yunnan as important location for tea and alder-based agroforestry, while southern parts of Yunnan are better suited for tea and hog plum, and northern parts could support walnut-based agroforestry options. Agroforestry is an important adaptation option for climate change, which could benefiting farmers and enhancing environmental conservation and restoration of the landscape.

dec_sup points to the research community focusing on needed frameworks and decision support tools to address climate change challenges while feeding into potential efforts to improve adaptation efforts. The only other topic present in more than half the papers is rainfall_temp_variability, which is not surprising given the main effects of climate change will be increasing temperatures and uncertain variability in the climate systems. Alternatively, freshwater_effects_mgmt is present in fewer than 25% of the papers, possibly the result of freshwater management potentially studied as an issue wrapped up in more prominent issues such as rainfall variability or planning and decision support. Hence, maybe it is present in papers looking at the specific challenges of freshwater management in the face of climate change. Finally, of interest from a business perspective is that finance_issues is the only topic that is not a clear majority topic weight (i.e., topic weight > 0.5) in any single paper. Even with no clear majority, the finance_issues topic appears in more than 25% of articles, so it is finding its way into the discussions in the published research. This breakdown of topic weights leads us to the

Figure 11: Topic weight histogram distributions for abstracts containing given topic (excludes zeroes).



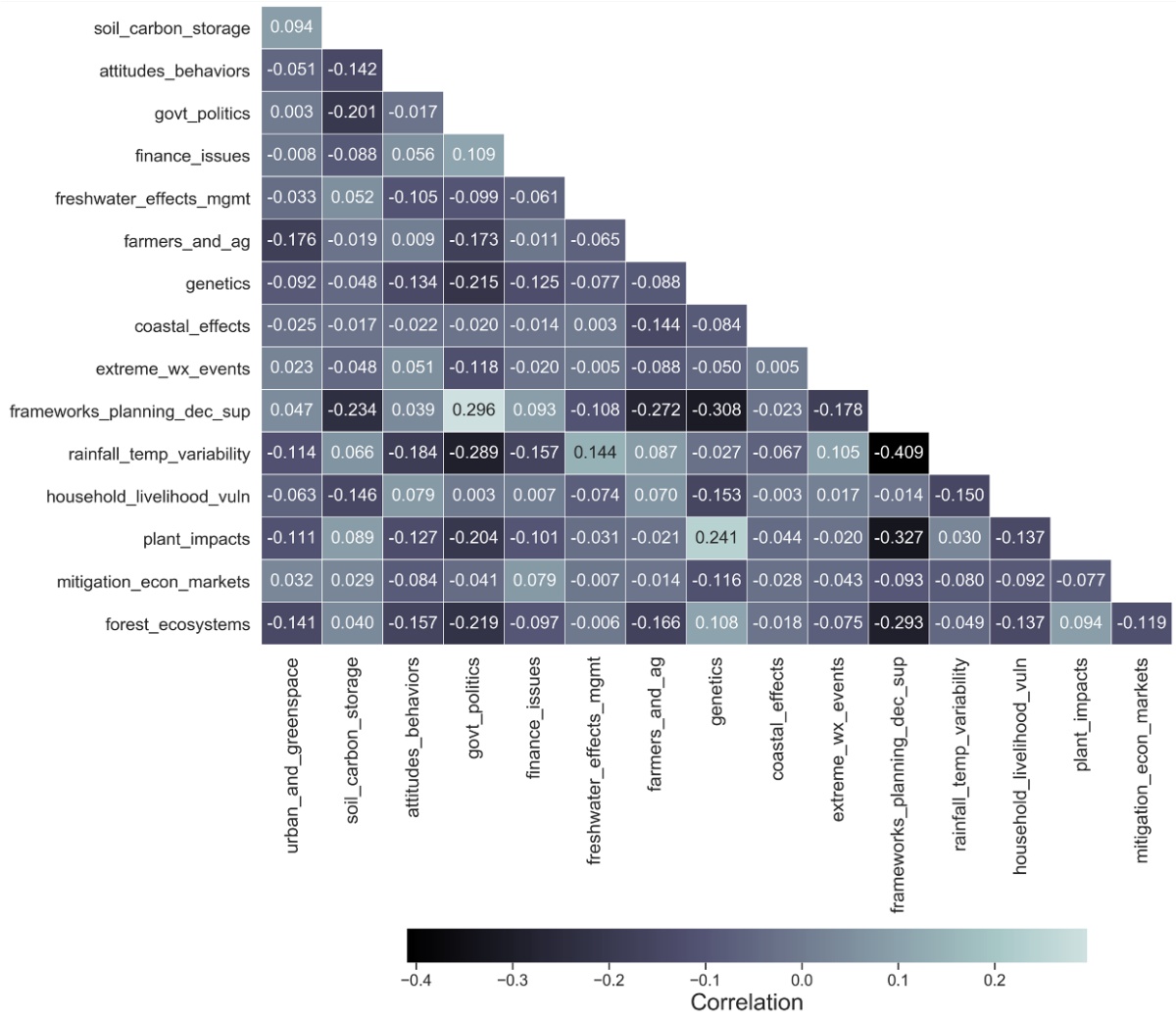
natural follow-on analysis, to wit, what are the interactions between individual topics within the documents in our corpus?

2.4.2.2 Topic correlations

In this next analysis, we investigate the correlations between topic weights within documents across the corpus. Given that each document contains multiple topics, the correlation of topics is essentially looking for which topics appear together in the article abstracts. Figure 12 shows a pairwise heatmap view of topic correlations. We use greyscale to highlight values further, with positively correlated topics being closer to white and negatively correlated topics being closer to black. Given that the total weight of all topics in a given document is equal to one, we expect the values to skew towards negative correlations, which shows in the scale

range at the bottom of Figure 12. Correlations that stand out in Figure 12 are the negative correlations associated with the frameworks_planning_dec_sup and govt_politics topics and other topics. These two topics also have the highest positive correlation, while tending to have higher negative correlations with the other topics.

Figure 12: Correlation of topic weights across all abstracts.



Delving deeper into the larger numbers, Table 15 shows the top five negative and top five positive correlations extracted from Figure 12. As noted from the heatmap, the top positive correlation is for frameworks_planning_dec_sup and govt_politics, which seems reasonable given

the work taking place in the public policy sphere focused on building frameworks for addressing climate change. Also noticeable in the positive correlations is freshwater_effects_mgmt pairing with rainfall_temp_variability, which also makes sense given that rainfall variability can certainly impact freshwater supplies. Additionally, genetics is positively correlated with both forest_ecosystems and plant_impacts, two natural systems genetically stressed by climate change. Alternatively, when looking at negative correlations, four of the top five are frameworks_planning_dec_sup to natural sciences studies associated with rainfall, genetics, plants, and forests. The fifth most negative correlation is govt_politics with rainfall_temp_variability. These negative correlations are somewhat surprising. One would expect the natural effects of climate change would entail the need for society to focus on developing frameworks to address changes in the nature-related sciences. This negatively correlated relationship is an area of interest potentially requiring additional focus in understanding if there is an insufficient cross-disciplinary focus on research within the natural sciences and policy-making spheres today.

Table 15: Top five negatively and positively correlated topics.

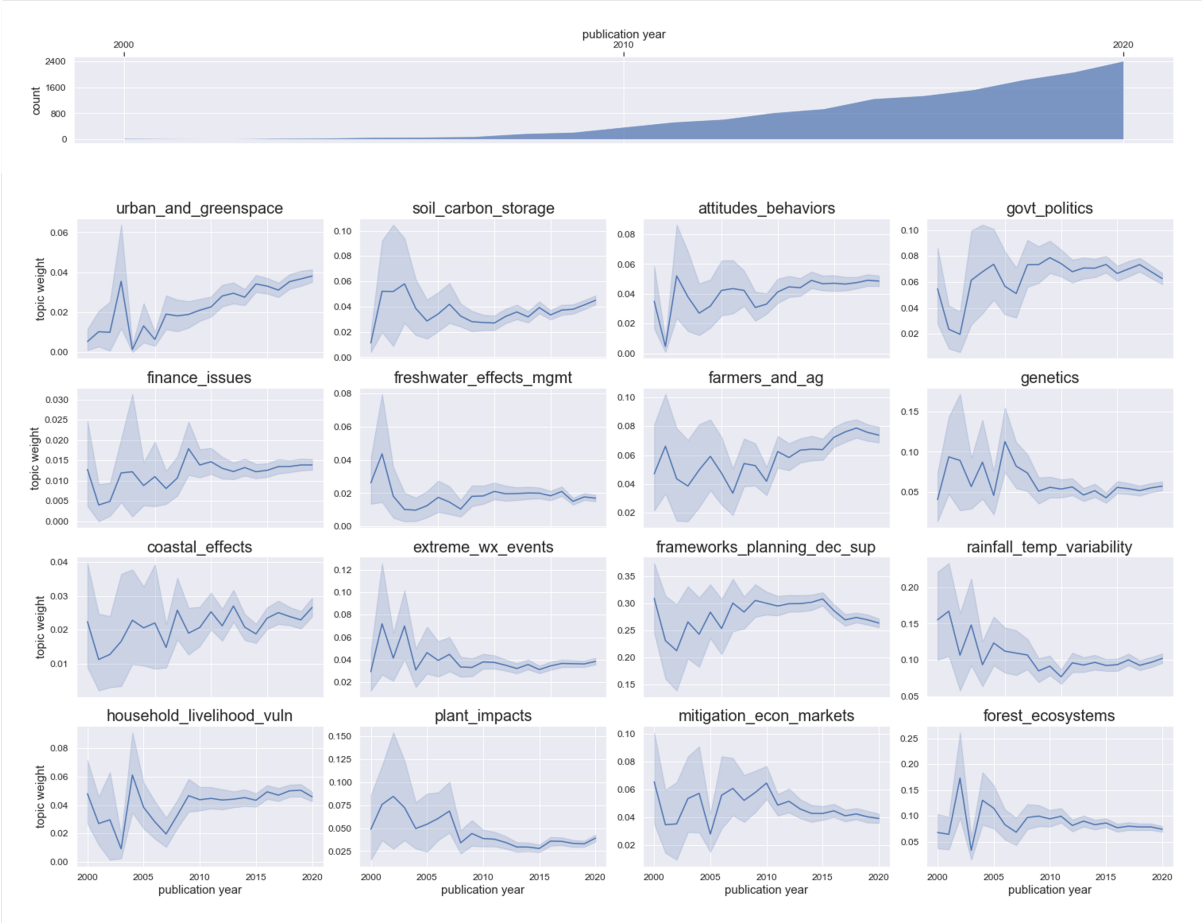
	First Topic	**Second Topic**	**Correlation**
Top Five Negative	frameworks_planning_dec_sup	rainfall_temp_variability	-0.409
	frameworks_planning_dec_sup	plant_impacts	-0.327
	frameworks_planning_dec_sup	genetics	-0.308
	frameworks_planning_dec_sup	forest_ecosystems	-0.293
	govt_politics	rainfall_temp_variability	-0.289
Top Five Positive	genetics	forest_ecosystems	0.108
	govt_politics	finance_issues	0.109
	freshwater_effects_mgmt	rainfall_temp_variability	0.144
	genetics	plant_impacts	0.241
	frameworks_planning_dec_sup	govt_politics	0.296

2.4.2.3 Time series analysis

One area of interest to explore with the topic model is the potential changes of topic weights within the documents over time. This next visualization can give us a sense of how the discourse around climate change is evolving. Figure 13 shows each of the 16 average topic weights over the 21 years of our corpus. At the top of Figure 13, the area plot is a reminder of the growth in the published CCA literature to provide perspective for the small volume in earlier periods. The line on each topic's time-series plot represents the mean topic weight across the corpus for each period, with the shaded region providing the 95% confidence interval around each period's mean. Given the much smaller count of abstracts in the earlier periods (fewer than 100 per year), the shaded confidence interval for data earlier than 2005 consists of much wider confidence intervals than data close to 2020, when there are thousands of abstracts for each year.

Of interest in evaluating the trend lines in Figure 13, several topic weights appear to have increased in recent years, most notably farmers_and_ag and urban_and_greenspace, with urban_and_greenspace consistently climbing since the mid-2000s. Conversely, a few topics decreased over the recent past, most notably mitigation_econ_markets and frameworks_planning_and_dec_sup. Given the substantial amount of abstracts containing some weighting of the frameworks topic, this seems to be pointing to frameworks_planning_dec_sup losing ground to other topics showing a recent increase in average topic weights. These plots provide a helpful visual to search for trends, but given the low volumes in the earlier periods, we have to be cautious in interpreting trends from the graphs. Instead, we will explore z-scores in the next section to identify and determine the significance of trends.

Figure 13: Overall growth over time in CCA literature and weighted topic means for each topic over time with shading representing the 95% confidence interval.



2.4.2.4 Topics trending hot and cold

Given the substantial increase in the number of CCA articles over time, looking for hot and cold topics based on a simple linear regression of the data shown in Figure 13 would be inappropriate. Instead, we will calculate z-scores from the weighted averages of the topic weights per document using a representative grouping of periods for the current period and the recent past. The time to publish new research can be quite extended, with some journals taking up to a year from initial paper submission until final acceptance and publication (Powell 2016).

This delay is for the papers that actually get published in a journal, so given that many papers get submitted to multiple journals in series as some journals reject the papers originally, it can take many months, even years, to get original research published. We choose the current period to include the most recent three-year period (2018-2020) to account for this publication delay. We use the five years preceding the current period for the historical comparison to represent the recent past historical period (2013-2017). Equation 2.1 shows how we calculate the z-scores, with the resulting values shown in Table 16.

$$z\text{-score} = \frac{\mu_{current} - \mu_{prior}}{\sigma_{prior}} \tag{2.1}$$

$\mu_{current}$ is the weighted topic mean for the most recent three year period

μ_{prior} is the weighted topic mean for the recent historical period

σ_{prior} is the weighted topic standard deviation for the recent historical period

Consistent with the trends in Figure 13, we can see that the z-scores show two of the three coldest topics are mitigation_econ_markets and frameworks_planning_and_dec_sup, with freshwater_effects_mgmt showing the most negative z-score. Even though the visual for freshwater_effects_mgmt does not show a consistent downward trend, the very low z-score seems to be an artifact of the last three years of this topic being lower than the preceding five years. The hottest topics by z-score appear to be urban_and_greenspace, soil_carbon_storage, and finance_issues, all of which are in line with the trends we see in Figure 13. Again, while farmers_and_ag has been visually trending positively for most of the last decade in Figure 13, the last two years have seen a drop-off causing the z-score for this topic to only show up fifth highest on the list.

Table 16: Topic z-scores for hot and cold topics using the latest 3 years trend (2018-2020) compared to the prior 5 years (2013-2017)

Topic	**z_score**
urban_and_greenspace	2.130
soil_carbon_storage	2.052
finance_issues	1.665
extreme_wx_events	1.242
farmers_and_ag	1.149
attitudes_behaviors	1.070
household_livelihood_vuln	1.015
plant_impacts	0.825
genetics	0.778
rainfall_temp_variability	0.611
coastal_effects	0.528
govt_politics	-1.028
forest_ecosystems	-1.148
frameworks_planning_dec_sup	-1.364
mitigation_econ_markets	-1.472
freshwater_effects_mgmt	-3.164

2.5 Discussion and conclusion

Although researchers use text-mining in many fields for identifying the underlying themes and topics addressed in published literature (He et al. 2013; Fang et al. 2018; Hall, Jurafsky, and Manning 2008; Zheng, McLean, and Lu 2006; Piepenbrink and Nurmammadov 2015), our analysis of the CCA literature using LDA is the first of its kind that we have seen. Using a well-documented analytical technique (LDA), we build a model that helps us identify 16 topics within the language used in the CCA literature. These 16 topics make sense when evaluated from a quantitative perspective scoring with coherence and perplexity measures. Better yet, with the help of two CCA SMEs, these topics also make sense from a qualitative perspective. We see a handful of themes when looking at these topics that loosely fall into themes that par-

allel effects of climate change (e.g., Extreme Weather Events; Coastal Effects; and Modeling Rainfall and Temperature Variability), effects on the natural environment (e.g., Forest Ecosystems; Plant-related Impacts; Climate-related Genetics; Freshwater Effects and Management), effects on people (e.g., Attitudes and Behaviors; Farmers and Agriculture; and Rural Households and Livelihood Vulnerability), and finally, individual or systemic efforts to address climate change (e.g., Soil Carbon Storage; Urban and Greenspace Environments; Finance-related Issues; Mitigation Economics and Markets; Government and Politics; and Framing, Planning, and Decision Support).

As shown in Figure 11, not all topics have the same presence throughout the literature. The Frameworks, Planning, and Decision Support topic is the dominant topic, showing up in all but 163 of the abstracts. It has an average weight of 0.281, a value that is three times larger than the next closest topic. This topic's ubiquitousness makes sense, given that we are looking at literature meant to address how civilization adapts to climate change. Researchers are actively working on ways to understand the effects of climate change and build out the adaptations necessary to address the worst that climate change will throw at us. These adaptations naturally will entail frameworks to understand the situation, discussion of planning for climate change, and the creation of decision support tools to enable decision-makers to make informed decisions that will combat uncertain climate change scenarios. That said, this topic tends to have meaningful negative correlations with those topics that have more of a natural environment theme. This apparent relationship opens up the logical question, should there be more interdisciplinary decision sciences focus for the natural environment studies associated with forests, plants, rainfall, and climate-related genetics? Is there an opportunity to leverage these research areas to enhance the conversation, including frameworks, planning tools, and decision support? We believe this is undoubtedly an outcome of our descriptive analysis worth investigating more deeply.

Notwithstanding, we are parsing the literature to look for areas ripe for improved cross-disciplinary research. By identifying the topics within the CCA research, we are able to do further analysis that could help us better understand them, how they are changing over time, and how they may interact. When looking at the topics growing during this period, we see three interesting topics in the most positive z-scores. Research on Urban and Greenspace Environments has seen the fastest growth in the recent past. It would be interesting to understand if this growth is due to the research ramping up in universities around the globe. Perhaps researchers are helping urban leaders learn how to help their cities adapt to climate change while leveraging improvements to the green space in their communities. The second fastest-growing topic is the topic of Soil Carbon Storage, possibly pointing to plant or agriculture-based methods progressing faster if researchers conduct more cross-disciplinary studies associated with carbon storage in the soil. The third highest z-score was for Finance Issues, yet interestingly Finance Issues did not have a single article where it was the clear majority topic. Is it possible that as CCA research has exploded, more researchers understand the significance of financial support? Going further, perhaps researchers are looking for appropriate finance mechanisms and funding available to take on what is arguably the biggest challenge facing humanity?

With our research grounded in the School of Management, we find Finance Issues' underrepresentation and recent growth within the corpus intriguing. This observation sets the stage for the two forthcoming chapters, where we delve into finance-related aspects of climate-change-related challenges. In Chapter 3, we will introduce a groundbreaking cost-benefit model to assess the overall expenses associated with investing in flood protection infrastructure in an urban coastal region to mitigate the damages caused by climate-change-induced sea level rise. In Chapter 4, we will develop a novel decision support model that considers the financial costs of adopting sustainable energy while incorporating the utility value of sustainability and branding attributes. These two studies will address existing research gaps in their respective areas while also approaching the challenges from a finance-focused perspective.

Any study attempting to distill themes out of a large set of literature has inherent limitations. We take a text-mining technique that assumes the words within the abstracts can be used to identify the themes discussed within the literature. The underlying assumptions of LDA that we can identify topics by finding the latent distribution of words and topics within the documents leaves us open to potentially finding topics that are nonsensical or that are not meaningful. To mitigate these limitations, we conduct our analysis by assessing hundreds of LDA model iterations, applying appropriate quantitative measures, evaluating with visualization tools, and incorporating the experience of CCA SMEs. Ideally, these efforts have produced sound results, but we must appreciate that this is not the only possible solution to such an analysis. We can only offer that the results are a reasonable representation of the topics being discussed in CCA literature and acknowledge that these results are a good foundation to delve deeper into the literature. The model output is also limited to the peer-reviewed published research available in one research database (Web of Science), meaning this is not the full scope of all CCA research that could be included. For instance, other ways to expand the corpus include exposing additional research databases (e.g., Proquest, JSTOR) or incorporating other types of published research such as conference proceedings, master's theses, and Ph.D. dissertations. Additionally, given that LDA is a method to turn words into tokens to create matrices for analysis, we focus solely on English language publications to avoid translation challenges. This language constraint means we may be missing out on potential topics captured in non-English research publications not found in our corpus.

As highlighted in Section 2.2.4, the LDA modeling done in this research can be used to potentially form hypotheses to test relationships between the topics we are seeing in the CCA literature and what we are seeing develop elsewhere. For instance, we may be interested in determining possible relationships between news articles about climate-related urban development and the topic of the Urban and Greenspace Environments. This research may include testing the hypothesis that the growth we see in the Urban and Greenspace Environments topic

is driven mainly due to activity discussed in news articles over the last decade. The output from our analysis enables the potential to test the relationship between published news articles and compare it to the timing of the growth seen in the period of interest. In Section 2.4.2, we only scratch the surface on the types of analysis that are possible using the dataset captured from Web of Science. For example, given the fields available from WoS, the relationship of specific topics could be analyzed by publication, publisher, author, or author location to determine concentrations of topics by any one of these parameters. Identifying relationships of topics to these fields could help bridge gaps in the research and capture opportunities for improved cross-disciplinary research. As noted in the limitations, this research could be extended by broadening the corpus to include other English language research (e.g., dissertations) or even to conduct the same text-mining exercise on a non-English language corpus to determine if the topics meaningfully change. Finally, there are additional text-mining approaches to dive deeper into the corpus that may enable us to segment further the topics found here in a hierarchical manner, which would enable a fuller understanding of particular topics and potential subsets of research captured therein.

As seen in Figure 1, there has been an explosion in CCA research over the last two decades, with nearly 100x growth in the WoS peer-reviewed literature during that time. Understanding the discourse in the literature has become more challenging because of this growth, stressing the traditional means of conducting comprehensive literature reviews or mapping exercises. As highlighted in (Lesnikowski et al. 2019), gaining insights into the CCA discourse within the peer-reviewed literature can be used to analyze framing and issue salience in the adaptation discourse. Our text-mining analysis included here applies the LDA method approach to published research in an attempt to do just that. Our analysis ties together best practices from past LDA text-mining based research to pull out meaningful topics present within the CCA literature since the beginning of the 21st century. Ideally, the topics we uncover in this analysis

will lead to better understanding this discourse and will feed into potentially new insights to expand opportunities for improved cross-disciplinary CCA research.

CHAPTER 3

WHICH IS MORE REWARDING IN MANAGING SEA LEVEL RISE AND HURRICANE STORM SURGE FLOODING: MITIGATION OR RESPONSE?

3.1 Introduction

Flooding accounts for nearly half of all natural disasters across the globe (Sodhi and Tang 2013). Economic losses caused by floods, only during 2009–2018, are estimated to be over $356T (EM-DAT 2020), making it one of the most devastating types of natural disasters (equal to earthquakes). Flooding, caused by the rapid melting of thick snow packs and ice jams, is Canada's most frequent and expensive natural disaster. For example, the 2013 flooding across Alberta displaced more than 100,000 people and imposed economic damages over $5 billion (Wang and Huang 2016). A more recent example is Hurricane Ida that unleashed a trail of destruction cutting from the Gulf Coast to the Northeast in the United States, causing an estimated $43B to $64B in total damages, much of it due to associated coastal flooding (Scism 2021). Projections indicate that flood hazard continues to accelerate. For example, by 2050, the city of Boston is expected to be annually hit by what is now a one in 10-year winter storm flood, under all emissions scenarios (Douglas and Kirshen 2022). Others report that, by 2100, the equivalent of today's one in 100-year flood event will probably become an annual disaster in Boston (Baranes et al. 2020; Thompson et al. 2019).

Flooding conditions are significantly affected by higher groundwater elevations, especially in coastal areas as sea level rises (Douglas and Kirshen 2022). There are many low-lying islands and coastal regions around the world, housing millions of people, that face increased

flooding and potential inundation year-round due to rising sea levels (Geest and Berg 2021; Nicholls et al. 2007). Thus, it is not surprising that the total estimated value of potential flooding damages from sea level rise (SLR) is in the trillions of dollars (Abadie 2018). The recurring theme for SLR-related coastal flooding is the lack of existing infrastructure such as levees and seawalls to protect coastal areas that can significantly mitigate the hazard (Chakravarty 2018). Given the effectiveness of storm barricades in mitigating the risk of coastal flooding, we aim to develop a decision support system to optimally manage investments in building a flood protection system in the form of dikes and levees.

This study is inspired by a critical issue that the City of Boston is encountering. Boston is an example of a coastal city facing the potential risk of flooding due to its rapid SLR. Over the past 20 years, the city has experienced an average SLR rate of 5.4 mm/yr, much faster than the global mean and twice the rate of Boston's SLR over the last century (Douglas and Kirshen 2022). Projections of SLR in Boston harbor diverge as a function of future emissions. Yet, compared to a 2000 baseline, SLR in 2100 is 35–78 cm under the most optimistic scenario, and it might exceed two meters under the worst case scenario (Oppenheimer et al. 2019; Climate Change 2019; Douglas and Kirshen 2022), warning of a significant risk of coastal area flooding (Sweet et al. 2017). Even without anticipated sea level rise, Boston has more than 3,000 properties that face substantial damage from flooding, so it can expect to see flood costs in excess of $35M each year (Abel 2021).

As highlighted in Chapter 2, the decision frameworks topic is dominant, while finance-related issues were a growing topic within climate change adaptation literature. These two topics stand out in the work done by the City of Boston to evaluate the SLR threat highlighted above. The city's management focuses on eleven strategic initiatives to address the expanded effects of climate change (City of Boston 2016). Among these are five flood-related strategies: monitoring up-to-date climate change projections, creating a coastal flood protection system, updating zoning and building regulations, retrofitting existing buildings, and insuring buildings

against flood damage. Striving to keep momentum, as part of the recommended actions in City of Boston 2016, this coastal protection strategy called for the city to launch a harbor-wide feasibility study within two years. The subsequent 2018 Boston harbor barrier feasibility report recommended forgoing a barrier system while implementing incremental steps and continued monitoring to see how the SLR situation unfolds (Kirshen, Borrelli, et al. 2018). The findings recommend using other multi-layer adaptation strategies (i.e., protection, accommodation, and retreat), at least for the next few decades, while monitoring actual SLR to better understand the uncertainty of the city's risks.

It is a positive sign that cities like Boston are working to overcome this inertia of inaction, but the latest Boston report still takes a wait-and-see approach (see Kirshen, Borrelli, et al. 2018). Unfortunately, cities face challenges requiring unprecedented foresight, complex coordination, and heightened urgency. While facing these challenges, multiple stakeholders are clamoring for attention, such as state and federal agencies, developers, landowners and nonprofit organizations (Wissman-Weber and Levy 2021). In light of these challenges, there are opportunities to improve acknowledging the risks of SLR-related flooding and develop methods that evaluate differential benefits and costs of public and policymaker (in)action (Mechler et al. 2014; Wissman-Weber and Levy 2021). Policymakers taking a wait-and-see approach in the face of rising seas require new decision support tools to provide the flexibility to more frequently assess the potential flood risks and investment costs their communities face.

In this paper, we focus on the finance-related issues of one effect of climate change. Specifically, we propose a cost-benefit analysis model to optimize investment decisions at each period, to mitigate flood hazards. To conduct a more granular analysis, we focus on flooding caused by sea level rise and hurricane storm surge along the sea coast. The flood protection infrastructure sought throughout this paper is in the form of creating dikes and levees via coastal land elevations. We propose a generalized modeling approach to minimize a cost function composed of two components: (1) the estimated cost of constructing an infrastructure of dikes and levees,

and (2) the potential SLR-related flood cost. We develop a multi-stage stochastic program with recourse to determine the least cost option when weighing both permanent and temporary flood damages in conjunction with investment costs to create a flood protection system of dikes and levees. This combined approach provides enlightening managerial insights into related costs for coastal areas if decision-makers are proactive in preventing damages versus waiting for the damages to occur and cleaning up the disaster after the fact. We also present a network-based framework for modeling complex flood movement dynamics on land to identify regions at risk. Our model is more generalized than the existing cost-benefit analyses, all limited to starting from a pre-existing infrastructure and enhancing that infrastructure over time. Creating such a cost-benefit analysis that simultaneously evaluates flood damage costs and the costs of building a shore-based infrastructure while also incorporating changing SLR projections would enable policymakers to meet the periodic monitoring recommended by experts.

One additional goal of this study is to demonstrate that completing this complex cost analysis could be done using only open-source data (USGSA 2009), which enables our approach to be applied more broadly. To this end and to demonstrate the performance of our proposed model, we discuss a case study of East Boston, which is a coastal region facing a substantial risk of SLR-related flooding. We consider a grid network representing the neighborhood of East Boston using open-source land elevation, tax appraisals, tidal gauge data, and published sea level rise elevations for possible climate change scenarios. We supplement this dataset with Google street view visualization to fill in gaps in the open-source tax data.

Using a simulation-based approach, we illustrate that our proposed method results in a cost reduction of as much as 92.2%, with an average of 63.2%, when compared to a "do nothing" strategy. Moreover, in a scenario-based experiment, we illustrate that our model results in similar cost savings when compared to a "do nothing" strategy in four different scenarios. Specifically, in the optimistic scenario (best-case), we see a cost reduction of as much as 85.0% and an average of 60.2%. These same maximum and average cost reductions are 92.5% and

59.2% in the expected-low scenario, 96.5% and 78.3% in the expected-high, and 96.5% and 78.9% in the high (worst-case) scenario. In addition to the East Boston case study, we repeated the experiments using 50 random networks, and showed that the methodology and insights are generalizable beyond the East Boston case. Finally, across all experiments, we present an extensive parameter sensitivity analysis, offering decision-makers the ability to compare the outcomes by incorporating the latest financial data or economic values.

We identify a few key takeaways from a policymaker's perspective. The first is that a modest investment at a fraction of the cost of expected damages under the "do nothing" strategy results in a meaningful reduction in flood-related and overall costs. Doing nothing and hoping for the best will only set up a coastal area for much higher costs over the evaluated time horizon. Sea level rise is a real threat which can significantly increase the cost of coastal flooding. Even under a scenario with no sea level rise and only expected annual storm flooding, the model suggests making investments if the build costs of levees are in the low to moderate range. Our model is also a powerful tool which can provide meaningful estimations for the optimal investment amounts required per period. This enables city planners to make more reasonable budgeting decisions in their financial planning for disaster prevention. From our sensitivity analysis, we could identify key factors (namely the costs to build the levees, the minimum build height, and the discount rate) that have significant impact on the investment amounts and their timing. Planners must pay close attention to these parameters to not underestimate the costs or overestimate the risk. Underestimating costs may result in building infrastructure that costs more than the damages it is meant to mitigate. Conversely, overbuilding for the potential sea level rise may limit the areas that can be protected due to quickly eating into any available budget. Finally, our experiments show that there will be areas that are not cost-effective to protect. No matter how much budget is available, the investment costs to protect these areas exceed the potential flood damage mitigation. This allows policymakers to assess areas under their control for a potential retreat rather than trying to protect them at any cost.

3.2 Literature review

The literature of disaster operations management has substantially grown in the recent past (Galindo and Batta 2013b; Besiou and Van Wassenhove 2020). Yet, much of this attention skews towards crisis response and logistics with little regarding mitigation policies (Akter and Wamba 2019; Galindo and Batta 2013b; Gupta et al. 2016; Besiou and Van Wassenhove 2020). The attention to response operations is justified by a variety of reasons, such as the high deprivation costs during response operations (Eftekhar, Song, and Webster 2022), media attention, and donors' sensitivity (Eftekhar et al. 2016).[1] Consequently, the majority of humanitarian assistance donations come with restrictions to focus on short-term relief operations, limiting opportunities for long-term investment to mitigate potential disasters (Oloruntoba and Gray 2006), and so mitigation strategies are often under-financed.

Looking at the special issues of 2014 Production and Operations Management, 2016 Journal of Operations Management, and 2018 European Journal of Operational Research, Besiou and Van Wassenhove 2020 found only one paper related to mitigation. Likewise, in their seminal review of papers published during 1957–2014 on disaster management, Gupta et al. 2016 identified 50 of 268 (18.7%) papers as being in the administrative function of prevention/mitigation, collectively referring to activities aimed at reducing the severity of a disaster's impact or ensuring that a man-made/natural event does not result in disaster. Of these, the majority concentrate on terrorism prevention policies following 9/11, with papers such as allocation of resources for airport screening (Bagchi and Paul 2014), response planning to bioterror attacks in airport terminals (Berman, Gavious, and Menezes 2012), and strategic terrorism deterrence in two-country frameworks (Roy and Paul 2013). Within these papers, there is a preponderance of papers not focused on specific disaster types (i.e., they treat disasters as a

1. For example, studies show that on average, it takes 38,920 deaths for a "food shortage crisis" to receive media coverage, while major U.S. networks cover news of an earthquake if it leads to two deaths (Eisensee and Strömberg 2007).

general problem). These studies tend to evaluate overarching methods or frameworks to apply generally to disasters, with some examples including evaluating disaster severity assessments (Rodríguez et al. 2011), representing perceived trade-offs between disaster impact and time to recovery to define disaster resilience (Zobel 2011), and developing a general methodology for inductive rule-building for NGOs involved in responding to natural disasters (Rodríguez, Vitoriano, and Montero 2012). Consequently, Gupta et al. 2016 emphasizes the need for more research in prevention/mitigation.

In digging deeper into the papers labeled as prevention/mitigation, there appear to be scant references centered on planning for mitigation of some disaster types such as flooding, epidemics, and wildfires. In the case of disaster-related research, this includes modeling with specific disaster characteristics to help practitioners develop adequate frameworks for the prevention and mitigation of disasters (Kovacs and Moshtari 2019). For example, although hurricane disaster management has received significant attention (see e.g., Uichanco 2022, Galindo and Batta 2013a, Campbell and Jones 2011, Lodree and Taskin 2008 and Davis et al. 2013), almost all of these studies focus on the response phases of crisis management. Gupta et al. 2016 found only seven papers related to floods, with only two of those focusing on prevention/mitigation. Of the two papers with some focus on prevention/mitigation, one was modeling disruption to freight transportation networks (Miller-Hooks, Zhang, and Faturechi 2012) and the other was covering optimal deployment for search and rescue operations during disasters (Chen and Miller-Hooks 2012).

Constructing a storm barricade system of levees and dikes is an effective technique for mitigating the risk of coastal flooding. Jonkman et al. 2009 employ an economic optimization approach for risk-based design of levee systems for the New Orleans metropolitan area. Suddeth, Mount, and Lund 2010 present an economic decision analysis approach for levee upgrade and repair investments in 34 major islands in California's Sacramento-San Joaquin Delta. Issues related to construction and maintenance of locally operated levees, and an overview of

federal programs addressing them are discussed in Keegan et al. 2011. Eijgenraam et al. 2017 discuss improving the dike infrastructure in the Netherlands to protect more than 55% of the land area that is below sea level. Perhaps the closest to the current paper is Chakravarty 2018, which proposes an optimization model integrating multiple decisions pre- and post-disaster to determine how investment in constructing levees can be leveraged in procuring relief items during preparation and response phases. Chakravarty 2018 considers a setting where a levee capacity decision is made by a governmental agency at the beginning of a multi-year planning horizon, while humanitarian relief agencies make procurement decisions every year. With this integrated model, Chakravarty 2018 analytically shows how increasing the levee capacity creates more social surplus over time. Nevertheless, depending on the severity of storms, the levees can be damaged or completely destroyed. An example is the 2005 Hurricane Katrina, which shattered the protective barriers and caused the disaster in New Orleans. Sills et al. 2008 investigate the Southeast Louisiana Flood and Hurricane Protection System that was in place at the time of Hurricane Katrina to further highlight the deficiency of knowledge in the technology and expertise needed for development of levee systems. Given the importance for reliability of a levee system, revamping these systems is very common (Chakravarty 2018), reflecting that flood hazard mitigation requires continuous investment from city management.

Upon evaluating the literature, there is a surprisingly small number of disaster-related papers focused on what many consider to be a slow-motion disaster in the making, coastal flooding caused by rising sea levels (IPCC 2014), and there is scant coverage for addressing the sea level rise in areas where infrastructure is non-existent today. In light of these limitations, our research aims to contribute to the literature by building a model that supports an adaptive strategic approach to mitigate potential disasters caused by flooding of a coastal city. This article puts a spotlight on the need for local and national government investment in infrastructure to lessen the impending risk of climate-change-induced flooding. To our knowledge, this is the first study that uses network-based modeling and linear algebra logic to represent complex wa-

ter movement dynamics on land for detecting regions at risk of flooding. Moreover, our model is more general than the existing methods because it incorporates both permanent and temporary flood damages along with investment costs, can be used in regions without any preexisting infrastructure, and can be built by using only open-source data.

3.3 Model, complexity, and solution

We model the Flood Risk Mitigation (FRM) problem as a multistage stochastic program with recourse. To present the model, we will define some sets, parameters and variables in this section. A list summarizing all these defined sets, parameters and variables is presented in Appendix A. The proposed model incorporates the risk of flooding over time, with t_{max} periods within the planning horizon $t \in \mathcal{T} = \{1, \ldots, t_{max}\}$, on a network of interconnected land grids. We first address the input parameters (i.e., associated with the sea level rise, grid partitioning, investment costs, flood damages, and available budget) and the necessary assumptions in section 3.3.1, and then present the full model in section 3.3.2. Then, we discuss the FRM problem's computational complexity and solution in section 3.3.3.

3.3.1 Input parameters and assumptions

3.3.1.1 Sea level states and their probabilities

Given that we are only focusing on flooding caused by SLR and hurricane storm surge along the sea coast, we model the state of the sea level during a period (denoted by $\mathcal{S}$) using two components. The first component, denoted by s, represents the sea level during a period solely due to the climate change effects. The second one, denoted by $\hat{s}$, indicates the sea level during a period due to both climate change effects and hurricane storm surge factors. Notice that we assume climate change effects and hurricane storm surge factors are independent of

each other. We also assume that the change in s and $\hat{s}$ happens at the start of a given period, and these two components stay unchanged during the period. These two components together shape the sea level state during a period (i.e., $\mathcal{S} = (s, \hat{s})$), and the set containing all possible sea level states during a period t is denoted by Ξ_t. At time zero, we assume that both components of the sea level state are zero (i.e., ($s = 0$, $\hat{s} = 0$)), and define the set containing this sea level state as $\Xi_0 = \{(0,0)\}$. Since the hurricane storm surges increase the sea level temporarily within a period, they pose even higher sea levels during the period, i.e., $s \leq \hat{s}$ for all $t \in \mathcal{T}$ and $\mathcal{S} \in \Xi_t$. Given $t \in \{0, \ldots, t_{max} - 1\}$, $\mathcal{S} \in \Xi_t$ and $\mathcal{S}' \in \Xi_{t+1}$, let $p_t^{\mathcal{S}\mathcal{S}'}$ denote the probability that the sea level state during period t is $\mathcal{S}$ and during period $t+1$ is $\mathcal{S}'$. We assume that probabilities $p_t^{\mathcal{S}\mathcal{S}'}, \forall t \in \{0, \ldots, t_{max} - 1\}, \mathcal{S} \in \Xi_t, \mathcal{S}' \in \Xi_{t+1}$, are known.

3.3.1.2 Grid-based partitioning

To model the SLR and hurricane storm flooding system as a network, we use a grid partitioning that segments a coastal region into hexagonal grids. More precisely, let us denote the coastal area in which we have control to create dikes and levees by elevating the land and we are also responsible for the cost of land elevation and flooding as the "region of interest." We only concern ourselves with areas within the region of interest that might get flooded during the planning horizon. Some parts of the region of interest might be adjacent to the sea at the start of the planning horizon (referred to as "time zero") and might get flooded directly from the sea. Depending on the land formations, other areas of the region of interest that are not adjacent to the sea may also get flooded due to water passing through the surrounding region during future time periods. To capture the possibility of the sea approaching a nonadjacent part of the region of interest during a future period, we also need to take into account parts of the region surrounding the region of interest that might get flooded at some point in future.

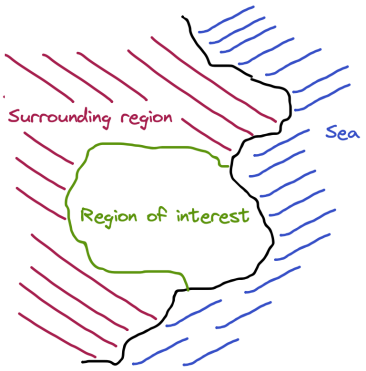
To identify parts of the region of interest and the surrounding region that are at risk of flooding during the planning horizon, we first partition the land in these two regions into hexagonal

grids. We assume that the elevation of the land on the surface of a grid is uniform and constant, and is equal to the average elevation of all points across the surface of that grid. We then consider the highest sea level across all sea level states (i.e., $\hat{s}_{max} = \max\{\hat{s} : (s,\hat{s}) \in \cup_{t\in\mathcal{T}}\Xi_t\}$), and identify land grids within the region of interest and the surrounding region at time zero that will get flooded under this sea level (i.e., the land grid elevation is below $\hat{s}_{max}$). A flooded land grid in the surrounding region will be considered in our model if it has a water path to a flooded land grid in the region of interest without going through the sea under the highest sea level $\hat{s}_{max}$. Let us refer to the flooded land grids in the region of interest under sea level $\hat{s}_{max}$ as the "area of interest" and denote the set containing these grids as Φ. We also refer to the flooded land grids in the surrounding region with a water path to some flooded land grid in the region of interest without going through the sea under sea level $\hat{s}_{max}$ as the "area of relevance." The set of land grids in the area of relevance is denoted by Ψ. Figure 14 shows the process of identifying the areas of interest and relevance in a simple example.

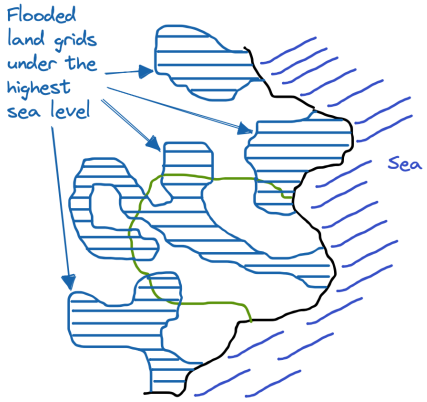
Land grids $i \in \Phi$ start with an initial elevation denoted by h_i. In our model, building levees and dikes within the area of interest is synonymous with raising the elevations of some land grids in set Φ incrementally over time throughout the planning horizon to prevent flooding within the area of interest. However, the elevations of land grids $i \in \Psi$ (also denoted by h_i) are going to stay unchanged throughout the planning horizon as we do not have control over these grids and we are not responsible for their flooding. The only reason grids in set Ψ are incorporated in our model is that these grids might create pathways for the sea to approach the area of interest. Notice that for some $i \in \Phi \cup \Psi$, we might have $h_i < 0$, which indicates that the elevation of the land grid i at time zero is below the sea level state at time zero (i.e., $(s = 0, \hat{s} = 0)$). We also similarly partition the sea at time zero into hexagonal grids designated as set O. These sea-based grids start with an elevation of zero, and rise accordingly with sea level changes over time.

Figure 14: Illustration of the process of identifying the areas of interest and relevance.

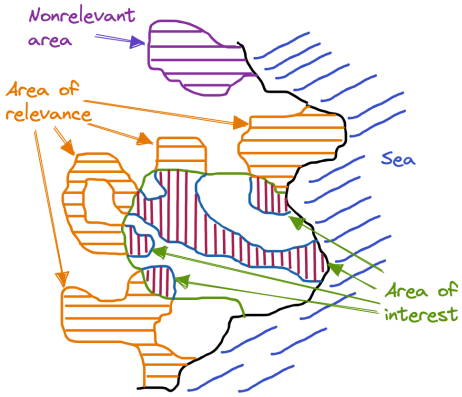
(a) An illustration of the sea, region of interest and its surrounding region at the beginning of the planning horizon.



(b) An illustration of the flooded land grids under the highest sea level across all sea level states (i.e., $\hat{s}_{max} = \max\{\hat{s} : (s, \hat{s}) \in \cup_{t \in \mathcal{T}} \Xi_t\}$).



(c) An illustration of the areas of interest and relevance identified under the highest sea level.



Note: All land grids in the area of relevance are flooded under the highest sea level and have a water path to some flooded land grid in the region of interest without going through the sea. The area marked as "Nonrelevant area" is also flooded under the highest sea level but any water path from this area to the region of interest passes through the sea.

To focus our modeling approach on the land grids subject to flooding during a given period t and under a given sea level state $\mathcal{S} \in \Xi_t$, we further segment the grids in sets Φ and Ψ into those grids at risk of temporary (hurricane storm surge related) flooding or permanent inundation flooding versus those grids that are not at risk of any flooding during period t and under sea level state $\mathcal{S}$. Assuming water can only flow between grids that share a physical border, we define the land grids at risk during a period t and under sea level state $\mathcal{S} \in \Xi_t$ as follows:

Definition 3.1. *Let $R_t^{\mathcal{S}}$ denote the subset of land grids in Φ at risk of permanent inundation flooding during period t and under sea level state $\mathcal{S} \in \Xi_t$. Similarly, let $Q_t^{\mathcal{S}}$ denote the subset of land grids in Ψ at risk of permanent inundation flooding during period t and under sea level state $\mathcal{S} \in \Xi_t$. Given $t \in \mathcal{T}$ and $\mathcal{S} = (s, \hat{s}) \in \Xi_t$, a land grid $i \in \Phi$ is in $R_t^{\mathcal{S}}$ if and only if it has an initial elevation h_i below the permanent sea level s during period t, and at least one of its neighbors is in set $O \cup R_t^{\mathcal{S}} \cup Q_t^{\mathcal{S}}$. Similarly, a land grid $i \in \Psi$ is in $Q_t^{\mathcal{S}}$ if and only if it has an elevation h_i below the permanent sea level s during period t, and at least one of its neighbors is in set $O \cup R_t^{\mathcal{S}} \cup Q_t^{\mathcal{S}}$. Notice that land grids in $R_t^{\mathcal{S}} \cup Q_t^{\mathcal{S}}$ are also logically at risk of temporary flooding.*

Definition 3.2. *Let $\hat{R}_t^{\mathcal{S}}$ denote the subset of land grids in Φ only at risk of temporary flooding during period t and under sea level state $\mathcal{S} \in \Xi_t$. Similarly, let $\hat{Q}_t^{\mathcal{S}}$ denote the subset of land grids in Ψ only at risk of temporary flooding during period t and under sea level state $\mathcal{S} \in \Xi_t$. Given $t \in \mathcal{T}$ and $\mathcal{S} = (s, \hat{s}) \in \Xi_t$, a land grid $i \in \Phi$ is in $\hat{R}_t^{\mathcal{S}}$ if and only if one of the following mutually exclusive cases happens:*

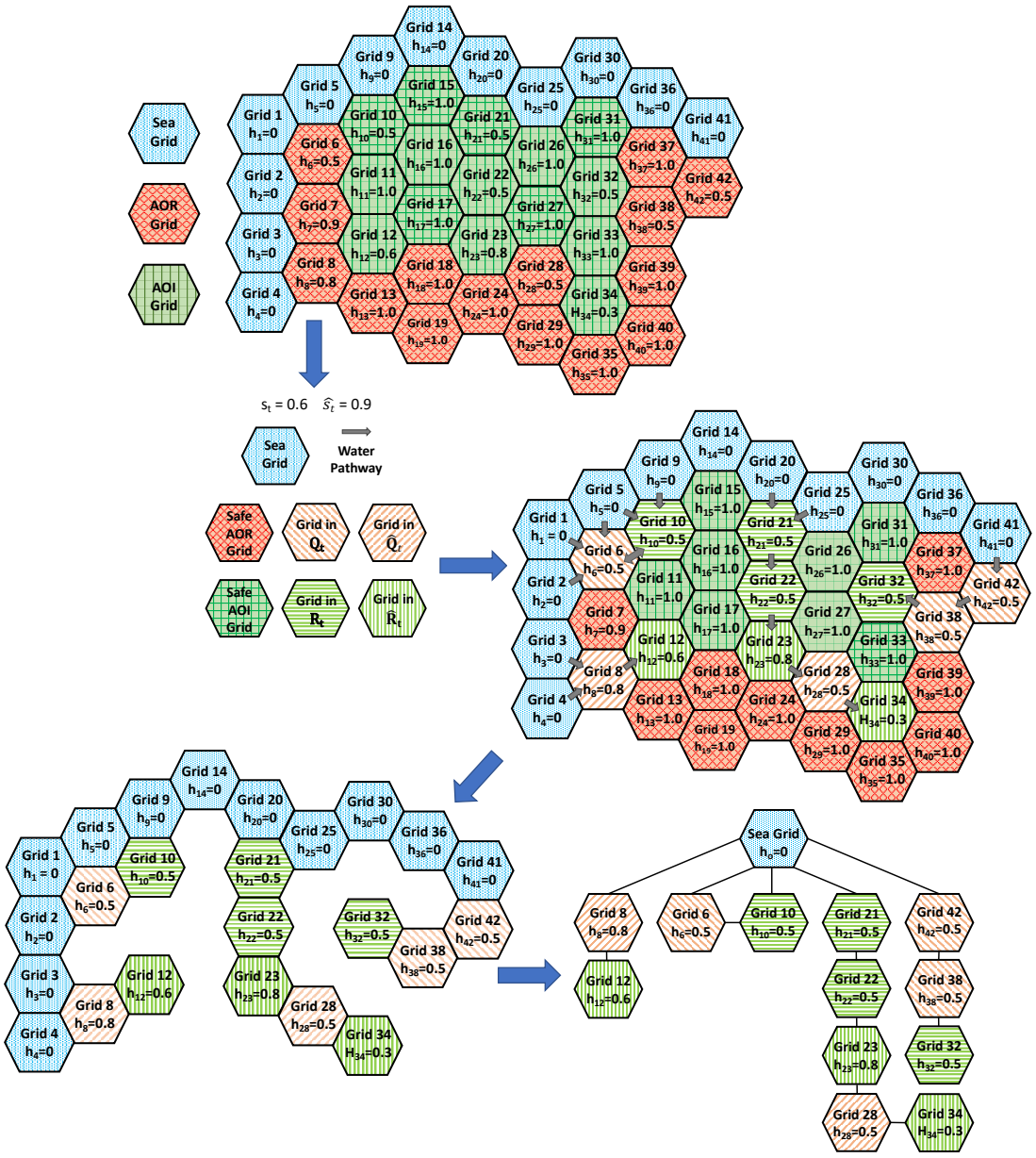
- *It has an initial elevation h_i below the temporary sea level $\hat{s}$ and above (or equal to) the permanent sea level s during period t, and at least one of its neighbors belongs to set $O \cup R_t^{\mathcal{S}} \cup Q_t^{\mathcal{S}} \cup \hat{R}_t^{\mathcal{S}} \cup \hat{Q}_t^{\mathcal{S}}$.*

- *It has an initial elevation h_i below the permanent sea level s during period t, none of its neighbors belongs to set $O \cup R_t^{\mathcal{S}} \cup Q_t^{\mathcal{S}}$, and at least one of its neighbors belongs to set $\hat{R}_t^{\mathcal{S}} \cup \hat{Q}_t^{\mathcal{S}}$.*

Similarly, a land grid $i \in \Psi$ is in $\hat{Q}_t^{\mathcal{S}}$ if and only if one of the above mentioned mutually exclusive cases happens for land grid i.

Given a period t and a sea level state $\mathcal{S} \in \Xi_t$, the land grids in set $R_t^{\mathcal{S}} \cup \hat{R}_t^{\mathcal{S}} \cup Q_t^{\mathcal{S}} \cup \hat{Q}_t^{\mathcal{S}}$ represent the vertices in a network (referred to as the "at-risk network during period t and under sea level state $\mathcal{S}$"), and the grids that share borders are made adjacent via edges within the network. Figure 15 provides an example for transformation of a full grid layout of the areas of interest and relevance along with the sea into an at-risk network during a period t and under a sea level state $\mathcal{S} = (0.6, 0.9) \in \Xi_t$ in meters. The starting elevations are labeled in each land grid and the sea-based grids are highlighted in blue. Highlighting the grids at risk during this time period and under this sea level state results in the middle image with sets $R_t^{\mathcal{S}}$, $\hat{R}_t^{\mathcal{S}}$, $Q_t^{\mathcal{S}}$ and $\hat{Q}_t^{\mathcal{S}}$ identified. To create the associated at-risk network using the identified at-risk land grids, we collapse all sea-based grids into one vertex (denoted by vertex "o"), and build the network with the at-risk grids and edges based on shared borders. The at-risk network basically highlights which grids will flood during the given time step and under the given sea level state, if no elevation increase is made for flood protection throughout the planning horizon. As shown in the At-Risk Network in Figure 15, the set of vertices includes a risk vertex for each risk grid in set $R_t^{\mathcal{S}} \cup \hat{R}_t^{\mathcal{S}} \cup Q_t^{\mathcal{S}} \cup \hat{Q}_t^{\mathcal{S}}$ and a single sea vertex. This means that we will have a total number of $| R_t^{\mathcal{S}} \cup \hat{R}_t^{\mathcal{S}} \cup Q_t^{\mathcal{S}} \cup \hat{Q}_t^{\mathcal{S}} | + 1$ vertices. Note that if a risk grid shares a border with the sea, we also add an edge between its corresponding vertex and the sea vertex (i.e., vertex o). Given the at-risk network during a period t and under a sea level state $\mathcal{S} \in \Xi_t$, the set containing vertices that are adjacent to a vertex $i \in R_t^{\mathcal{S}} \cup \hat{R}_t^{\mathcal{S}} \cup Q_t^{\mathcal{S}} \cup \hat{Q}_t^{\mathcal{S}}$ is referred to as the neighbors of i within the at-risk network, and is denoted by $N_t^{\mathcal{S}}(i)$.

Figure 15: Example transformation of a full grid layout of the areas of interest and relevance along with the sea into an at-risk network for a given period t and a given sea level state $(s = 0.6m, \hat{s} = 0.9m) \in \Xi_t$.



3.3.1.3 Grid costs for investment and flood damage

There are four inputs required to model the costs associated with the FRM problem. The first parameter is related to the investment cost to elevate grids in set Φ by building dikes and levees on them. The cost c is what it takes to elevate a grid in Φ by one meter at the start of a given period. We assume that investment costs are uniform across grids, independent of the grid's surface structure, and do not vary much over the planning horizon. The units of c are in terms of dollars per meter of elevation raise. Similar to the case of the sea level rise, we assume that the increase in grids' elevations happens at the start of a given period before realization of the sea level state at the start of that period, and the grids' elevations stay unchanged during the period.

The next two inputs provide the information needed to determine flood-related damages. The first parameter is the cost g_i of losing a grid $i \in \Phi$ due to inundation if the grid is in $R_t^{\mathcal{S}}$ during a period t and under a sea level state $\mathcal{S} \in \Xi_t$, and is permanently flooded. We assume the inundation cost is a constant value representing the full grid loss during a given period. We also assume that once a grid is inundated (permanently flooded) during a period, it is possible to raise the elevation of that grid at the start of the next period and pull it out of the inundation (full loss) state. This is specifically possible when the investment budget in early periods is limited. Notice that during a period t and under a sea level state $\mathcal{S} \in \Xi_t$, a grid may be in $R_t^{\mathcal{S}}$ but not be permanently flooded because that grid has been elevated at the start of period t before realization of $\mathcal{S}$ or during the previous periods, or there is no path to it from the sea due to other grids being elevated. We assume g_i is strictly positive, and its units are in dollars per time period for a given grid $i \in \Phi$. The second flood-cost related parameter is the cost for one meter of hurricane storm surge flood damage f_i to a grid i in $R_t^{\mathcal{S}} \cup \hat{R}_t^{\mathcal{S}}$ during a period t and under a sea level state $\mathcal{S} \in \Xi_t$ when the grid is temporarily flooded. Notice that we assume a grid will not experience hurricane storm surge related costs if it is permanently inundated

during a given period. This is due to the fact that when a grid i is permanently inundated during a period, the total value of the grid (i.e., g_i) is lost during that period, but the storm surge only causes partial grid value loss (e.g., only first floors of buildings being damaged) during a period. Therefore, during a period t and under a sea level state $\mathcal{S} \in \Xi_t$, a grid might be in $R_t^{\mathcal{S}} \cup \hat{R}_t^{\mathcal{S}}$, but not be temporarily flooded because it is permanently flooded, or it has been elevated at the start of period t or during previous periods, or there is no path to it from the sea due to other grids being elevated. Hurricane storm surge flood cost in a given grid is based on a linear depth damage curve for that grid. Similarly, we assume f_i is strictly positive, and its units are in dollars per time period per meter of sea level elevation above a grid i elevation.

The final input to incorporate realistic costs over time is to apply a discount rate per period, denoted by λ. The discount rate systematically adjusts the value of costs and benefits during future periods. Notice that to get λ, the standard annual discount rate (i.e., d) will need to be adjusted to match the time frame used for a period in our model.

3.3.1.4 Budget

We consider a fixed construction and maintenance budget for each period t (denoted by b_t) that does not carry over into other periods, where b_t is given for $t \in \mathcal{T}$. Coastal cities' resources are limited, and so this budget imposes a constraint on the amount of investment and construction in a given period. Inclusion of this budget constraint in the FRM problem makes this problem more realistic, but imposes significant computational challenge for its solution. We further discuss the FRM problem's complexity and solution in Section 3.3.3.

3.3.2 Model development

3.3.2.1 Decision variable

The main decision variables for the FRM model are the heights of each grid $i \in \Phi$ during each period $t \in \mathcal{T}$. As mentioned before, we assume that the decisions on the heights of the grids in Φ during a period t are made at the start of period t before disclosure of the sea level change at the start of this period. We also assume that the decisions on the heights of the grids in Φ at the start of any period $t \in \{1, \ldots, t_{max}\}$ only depend on the sea level state during period $t-1$, which is known to the decision maker at the time of decision making. This means that the heights of the grids during the first period (first-stage decisions) are decided while the only piece of information available is the sea level state at time zero (i.e., $(s = 0, \hat{s} = 0)$). So, given a period $t \in \mathcal{T}$ and a sea level state $\mathcal{S} \in \Xi_{t-1}$, we use notation $x_{it\mathcal{S}}$ to represent the decision variable associated with the height of a grid $i \in \Phi$ during period t if the sea level state during period $t-1$ is $\mathcal{S}$.

It is important to note that elevating grids by building dikes and levees on them cannot be done in small increments across the years. In practice, if the decision maker decides to elevate a grid in Φ during a time period, the elevation increase should be done up to a minimum threshold to justify the initial setup cost. Therefore, we incorporate a parameter m in our model that represents the minimum threshold of elevation increase in any grid in Φ during a period. Moreover, to model the FRM problem, we need to find a valid upper bound on the elevation increase in any grid in Φ during a period (denoted by M). One such valid upper bound is $M = \max\{\hat{s}_{max} - \min\{h_i : i \in \Phi \cup \Psi\}, m\}$, which is used in our model.

3.3.2.2 Objective function

This model's main objective is to minimize the total cost, with two primary components to address. The first is the expected investment cost for building dikes and levees by increasing

the elevation of grids to protect themselves and possibly other grids in the network. The second is the expected flood cost when grids are affected by either permanent inundation or hurricane storm surge flooding.

The cost for investment is evaluated on a per grid basis, and the overall cost for each grid is determined by looking at a grid's height change throughout the planning horizon. Equation (3.1) below captures the expected investment cost (denoted by *EIC*) for all at-risk grids during the planning horizon. To account for discounted future periods, we incorporate the adjusted discount rate λ in this equation.

$$EIC = c\sum_{i\in\Phi}(x_{i1(0,0)} - h_i) + c\sum_{i\in\Phi}\sum_{t=1}^{t_{max}-1}\sum_{\mathcal{S}\in\Xi_{t-1}}\sum_{\mathcal{S}'\in\Xi_t} p_{(t-1)}^{\mathcal{S}\mathcal{S}'}\frac{(x_{i(t+1)\mathcal{S}'} - x_{it\mathcal{S}})}{\lambda^t} \tag{3.1}$$

A grid $i \in \Phi\cup\Psi$ faces two mutually exclusive possibilities of flooding during a period and under a sea level state: (1) permanent inundation where the grid is deemed underwater (at least during daily high tides) for the full period t, and (2) temporary hurricane storm surge flooding where the grid only faces damage due to short duration flooding within the period. During a period $t \in \mathcal{T}$ and under sea level states $\mathcal{S} \in \Xi_{t-1}$ and $\mathcal{S}' \in \Xi_t$ for which $p_{t-1}^{\mathcal{S}\mathcal{S}'} > 0$, to capture if a grid $i \in R_t^{\mathcal{S}'} \cup Q_t^{\mathcal{S}'}$ is inundated, we designate a binary variable $w_{it\mathcal{S}\mathcal{S}'}$, where $w_{it\mathcal{S}\mathcal{S}'} = 0$ if the grid is not inundated, and $w_{it\mathcal{S}\mathcal{S}'} = 1$ otherwise. If a grid $i \in R_t^{\mathcal{S}'} \cup Q_t^{\mathcal{S}'} \cup \hat{R}_t^{\mathcal{S}'} \cup \hat{Q}_t^{\mathcal{S}'}$ is not inundated but faces hurricane storm surge related flooding during period t and under sea level states $\mathcal{S} \in \Xi_{t-1}$ and $\mathcal{S}' \in \Xi_t$ for which $p_{t-1}^{\mathcal{S}\mathcal{S}'} > 0$, we designate the water depth used to calculate the flood cost by a continuous variable $z_{it\mathcal{S}\mathcal{S}'}$. As mentioned before, if a grid $i \in R_t^{\mathcal{S}'} \cup Q_t^{\mathcal{S}'}$ is inundated, then it is assumed to be only subject to the permanent flooding cost, and not the hurricane storm surge cost. This means that if $w_{it\mathcal{S}\mathcal{S}'} = 1$ for some $t \in \mathcal{T}$, $\mathcal{S} \in \Xi_{t-1}$, $\mathcal{S}' \in \Xi_t$, and $i \in R_t^{\mathcal{S}'} \cup Q_t^{\mathcal{S}'}$, then $z_{it\mathcal{S}\mathcal{S}'}$ is assumed to be zero.

If a grid $i \in \Phi$ is inundated during a period (permanent flooding), there is a fixed cost (i.e., g_i) for losing that grid during that period. If a grid $i \in \Phi$ is not inundated but is affected by hurricane storm surge level during a period (temporary flooding), then the cost is assumed to

be a linear depth damage curve (using parameter f_i) that depends on the depth of the flood in grid i during that period. Using variables $w_{it\mathcal{S}\mathcal{S}'}$ and $z_{it\mathcal{S}\mathcal{S}'}$, and incorporating the discount rate for each period, we have the expected flood cost (denoted by EFC) as shown in Equation (3.2).

$$EFC = \sum_{t=1}^{t_{max}} \sum_{\mathcal{S}\in\Xi_{t-1}} \sum_{\mathcal{S}'\in\Xi_t} \frac{p_{t-1}^{\mathcal{S}\mathcal{S}'}}{\lambda^t} \left(\sum_{i\in R_t^{\mathcal{S}'}\cup\hat{R}_t^{\mathcal{S}'}} f_i z_{it\mathcal{S}\mathcal{S}'} + \sum_{i\in R_t^{\mathcal{S}'}} g_i w_{it\mathcal{S}\mathcal{S}'} \right) \tag{3.2}$$

3.3.2.3 Associated constraints and the full model

Minimize $EIC + EFC$ (3.3)

Subject to:

$x_{i1(0,0)} \geq h_i + mv_{i1(0,0)}$ $\forall i \in \Phi$ (3.4)

$x_{i1(0,0)} \leq h_i + Mv_{i1(0,0)}$ $\forall i \in \Phi$ (3.5)

$x_{i(t+1)\mathcal{S}'} \geq x_{it\mathcal{S}} + mv_{i(t+1)\mathcal{S}\mathcal{S}'}$ $\forall i \in \Phi,\ \forall t \in \{1,\ldots,t_{max}-1\},\ \forall \mathcal{S}\in\Xi_{t-1}, \mathcal{S}'\in\Xi_t : p_{t-1}^{\mathcal{S}\mathcal{S}'} > 0$ (3.6)

$x_{i(t+1)\mathcal{S}'} \leq x_{it\mathcal{S}} + Mv_{i(t+1)\mathcal{S}\mathcal{S}'}$ $\forall i \in \Phi,\ \forall t \in \{1,\ldots,t_{max}-1\},\ \forall \mathcal{S}\in\Xi_{t-1}, \mathcal{S}'\in\Xi_t : p_{t-1}^{\mathcal{S}\mathcal{S}'} > 0$ (3.7)

$c\sum_{i\in\Phi}(x_{i1(0,0)} - h_i) \leq b_1$ (3.8)

$c\sum_{i\in\Phi}\frac{(x_{i(t+1)\mathcal{S}'} - x_{it\mathcal{S}})}{\lambda^t} \leq b_{(t+1)}$ $\forall t \in \{1,\ldots,t_{max}-1\},\ \forall \mathcal{S}\in\Xi_{t-1}, \mathcal{S}'\in\Xi_t : p_{t-1}^{\mathcal{S}\mathcal{S}'} > 0$ (3.9)

$$w_{it\mathcal{S}\mathcal{S}'} \geq \frac{(s' - x_{it\mathcal{S}})}{M} - (1 - y_{it\mathcal{S}\mathcal{S}'}) \quad \forall t \in \{1,\ldots,t_{max}\},\ \forall \mathcal{S}\in\Xi_{t-1},\ \forall \mathcal{S}' = (s',\hat{s}') \in \Xi_t : p_{t-1}^{\mathcal{S}\mathcal{S}'} > 0,\ \forall i \in R_t^{\mathcal{S}'} \tag{3.10}$$

$$w_{it\mathcal{S}\mathcal{S}'} \geq \frac{(s' - h_i)}{M} - (1 - y_{it\mathcal{S}\mathcal{S}'}) \quad \forall t \in \{1,\ldots,t_{max}\},\ \forall \mathcal{S}\in\Xi_{t-1},\ \forall \mathcal{S}' = (s',\hat{s}') \in \Xi_t : p_{t-1}^{\mathcal{S}\mathcal{S}'} > 0,\ \forall i \in Q_t^{\mathcal{S}'} \tag{3.11}$$

$$\sum_{i'\in N_t^{\mathcal{S}'}(i)\cap(R_t^{\mathcal{S}'}\cup Q_t^{\mathcal{S}'})} w_{i't\mathcal{S}\mathcal{S}'} \leq |N_t^{\mathcal{S}'}(i)|\, y_{it\mathcal{S}\mathcal{S}'} \quad \forall t \in \{1,\ldots,t_{max}\},\ \forall \mathcal{S}\in\Xi_{t-1},\ \forall \mathcal{S}'\in\Xi_t : p_{t-1}^{\mathcal{S}\mathcal{S}'} > 0,\ \forall i \in R_t^{\mathcal{S}'}\cup Q_t^{\mathcal{S}'} : o \notin N_t^{\mathcal{S}'}(i) \tag{3.12}$$

$$y_{it\mathcal{S}\mathcal{S}'} = 1 \qquad \forall t \in \{1,\ldots,t_{max}\},\ \forall \mathcal{S} \in \Xi_{t-1},\ \forall \mathcal{S}' \in \Xi_t : p_{t-1}^{\mathcal{S}\mathcal{S}'} > 0,\ \forall i \in R_t^{\mathcal{S}'} \cup Q_t^{\mathcal{S}'} : o \in N_t^{\mathcal{S}'}(i) \quad (3.13)$$

$$z_{it\mathcal{S}\mathcal{S}'} \geq (\hat{s}' - x_{it\mathcal{S}}) - M(1 - \hat{y}_{it\mathcal{S}\mathcal{S}'} + w_{it\mathcal{S}\mathcal{S}'}) \qquad \forall t \in \{1,\ldots,t_{max}\},\ \forall \mathcal{S} \in \Xi_{t-1},\ \forall \mathcal{S}' = (s',\hat{s}') \in \Xi_t : p_{t-1}^{\mathcal{S}\mathcal{S}'} > 0,\ \forall i \in R_t^{\mathcal{S}'} \quad (3.14)$$

$$z_{it\mathcal{S}\mathcal{S}'} \geq (\hat{s}' - h_i) - M(1 - \hat{y}_{it\mathcal{S}\mathcal{S}'} + w_{it\mathcal{S}\mathcal{S}'}) \qquad \forall t \in \{1,\ldots,t_{max}\},\ \forall \mathcal{S} \in \Xi_{t-1},\ \forall \mathcal{S}' = (s',\hat{s}') \in \Xi_t : p_{t-1}^{\mathcal{S}\mathcal{S}'} > 0,\ \forall i \in Q_t^{\mathcal{S}'} \quad (3.15)$$

$$z_{it\mathcal{S}\mathcal{S}'} \geq (\hat{s}' - x_{it\mathcal{S}}) - M(1 - \hat{y}_{it\mathcal{S}\mathcal{S}'}) \qquad \forall t \in \{1,\ldots,t_{max}\},\ \forall \mathcal{S} \in \Xi_{t-1},\ \forall \mathcal{S}' = (s',\hat{s}') \in \Xi_t : p_{t-1}^{\mathcal{S}\mathcal{S}'} > 0,\ \forall i \in \hat{R}_t^{\mathcal{S}'} \quad (3.16)$$

$$z_{it\mathcal{S}\mathcal{S}'} \geq (\hat{s}' - h_i) - M(1 - \hat{y}_{it\mathcal{S}\mathcal{S}'}) \qquad \forall t \in \{1,\ldots,t_{max}\},\ \forall \mathcal{S} \in \Xi_{t-1},\ \forall \mathcal{S}' = (s',\hat{s}') \in \Xi_t : p_{t-1}^{\mathcal{S}\mathcal{S}'} > 0,\ \forall i \in \hat{Q}_t^{\mathcal{S}'} \quad (3.17)$$

$$\sum_{i' \in N_t^{\mathcal{S}'}(i)} z_{i't\mathcal{S}\mathcal{S}'} + \sum_{i' \in N_t^{\mathcal{S}'}(i) \cap (R_t^{\mathcal{S}'} \cup Q_t^{\mathcal{S}'})} w_{i't\mathcal{S}\mathcal{S}'} \leq |N_t^{\mathcal{S}'}(i)|\,(M+1)\hat{y}_{it\mathcal{S}\mathcal{S}'} \qquad \forall t \in \{1,\ldots,t_{max}\},\ \forall \mathcal{S} \in \Xi_{t-1},\ \forall \mathcal{S}' \in \Xi_t : p_{t-1}^{\mathcal{S}\mathcal{S}'} > 0,\ \forall i \in R_t^{\mathcal{S}'} \cup Q_t^{\mathcal{S}'} \cup \hat{R}_t^{\mathcal{S}'} \cup \hat{Q}_t^{\mathcal{S}'} : o \notin N_t^{\mathcal{S}'}(i) \quad (3.18)$$

$$\hat{y}_{it\mathcal{S}\mathcal{S}'} = 1 \qquad \forall t \in \{1,\ldots,t_{max}\},\ \forall \mathcal{S} \in \Xi_{t-1},\ \forall \mathcal{S}' \in \Xi_t : p_{t-1}^{\mathcal{S}\mathcal{S}'} > 0,\ \forall i \in R_t^{\mathcal{S}'} \cup Q_t^{\mathcal{S}'} \cup \hat{R}_t^{\mathcal{S}'} \cup \hat{Q}_t^{\mathcal{S}'} : o \in N_t^{\mathcal{S}'}(i) \quad (3.19)$$

$$z_{it\mathcal{S}\mathcal{S}'} \geq 0 \qquad \forall t \in \{1,\ldots,t_{max}\},\ \forall \mathcal{S} \in \Xi_{t-1},\ \forall \mathcal{S}' \in \Xi_t : p_{t-1}^{\mathcal{S}\mathcal{S}'} > 0,\ \forall i \in R_t^{\mathcal{S}'} \cup Q_t^{\mathcal{S}'} \cup \hat{R}_t^{\mathcal{S}'} \cup \hat{Q}_t^{\mathcal{S}'} \quad (3.20)$$

$$w_{it\mathcal{S}\mathcal{S}'} \in \{0,1\} \qquad \forall t \in \{1,\ldots,t_{max}\},\ \forall \mathcal{S} \in \Xi_{t-1},\ \forall \mathcal{S}' \in \Xi_t : p_{t-1}^{\mathcal{S}\mathcal{S}'} > 0,\ \forall i \in R_t^{\mathcal{S}'} \cup Q_t^{\mathcal{S}'} \quad (3.21)$$

$$v_{i1(0,0)} \in \{0,1\} \qquad \forall i \in \Phi \quad (3.22)$$

$$v_{it\mathcal{S}\mathcal{S}'} \in \{0,1\} \qquad \forall t \in \{2,\ldots,t_{max}\},\ \forall \mathcal{S} \in \Xi_{t-2},\ \forall \mathcal{S}' \in \Xi_{t-1} : p_{t-2}^{\mathcal{S}\mathcal{S}'} > 0,\ \forall i \in \Phi \quad (3.23)$$

$$y_{it\mathcal{S}\mathcal{S}'} \in \{0,1\} \qquad \forall t \in \{1,\ldots,t_{max}\},\ \forall \mathcal{S} \in \Xi_{t-1},\ \forall \mathcal{S}' \in \Xi_t : p_{t-1}^{\mathcal{S}\mathcal{S}'} > 0,\ \forall i \in R_t^{\mathcal{S}'} \cup Q_t^{\mathcal{S}'} \quad (3.24)$$

$$\hat{y}_{it\mathcal{S}\mathcal{S}'} \in \{0,1\} \qquad \forall t \in \{1,\ldots,t_{max}\},\ \forall \mathcal{S} \in \Xi_{t-1},\ \forall \mathcal{S}' \in \Xi_t : p_{t-1}^{\mathcal{S}\mathcal{S}'} > 0,\ \forall i \in R_t^{\mathcal{S}'} \cup Q_t^{\mathcal{S}'} \cup \hat{R}_t^{\mathcal{S}'} \cup \hat{Q}_t^{\mathcal{S}'} \quad (3.25)$$

The elevation of a grid $i \in \Phi$ is assumed to stay constant or increase due to an investment in building dikes and levees on the grid. Using Inequalities (3.4)-(3.7), the model ensures that a grid $i \in \Phi$ cannot be lowered in elevation from its initial elevation h_i or any subsequent elevation it may be raised to in the planning horizon. These equations also ensure that if a grid $i \in \Phi$ is elevated at the start of a period, the increase in elevation is at least equal to the minimum required threshold m. Notice that Inequalities (3.6)-(3.7) are written for any possible transition of water level states from a period $t-1$ to a period t (with positive probability) as we do not need to enforce these requirements on impossible water level state transitions. This is the case for many of the constraints in our proposed model. Inequalities (3.8) and (3.9) limit the amount of money spent at the start of a given period t for raising the elevations of grids in Φ by a user-specified parameter b_t.

Given a period t and water level states $\mathcal{S} \in \Xi_{t-1}$ and $\mathcal{S}' = (s', \hat{s}') \in \Xi_t$ with positive transition probabilities, a grid $i \in R_t^{\mathcal{S}'} \cup Q_t^{\mathcal{S}'}$ is protected from inundation during period t, if its elevation (i.e., $x_{it\mathcal{S}}$ for $i \in R_t^{\mathcal{S}'}$, and h_i for $i \in Q_t^{\mathcal{S}'}$) is higher than permanent sea level s'. Grid i is also protected if it does not have a hydraulic connection to the sea via one or more paths through inundated grids in $R_t^{\mathcal{S}'} \cup Q_t^{\mathcal{S}'}$. This secondary protection is determined by checking if grid i is a neighbor of the sea grid or it has any adjacent grid $i' \in N_t^{\mathcal{S}'}(i) \cap (R_t^{\mathcal{S}'} \cup Q_t^{\mathcal{S}'})$ that is inundated, and is represented by a binary variable $y_{it\mathcal{S}\mathcal{S}'}$ in the model. If grid i is safe from inundation due to not being a neighbor of the sea grid and also due to the absence of a hydraulic connection to the ocean through inundated grids in $R_t^{\mathcal{S}'} \cup Q_t^{\mathcal{S}'}$, $y_{it\mathcal{S}\mathcal{S}'}$ is zero, and one otherwise. Using variables $y_{it\mathcal{S}\mathcal{S}'}$, Inequalities (3.10)-(3.13) along with the objective function guarantee that grid $i \in R_t^{\mathcal{S}'} \cup Q_t^{\mathcal{S}'}$ is inundated during period t (i.e., $w_{it\mathcal{S}\mathcal{S}'} = 1$) if and only if its elevation is below the sea level s' and it is a neighbor of the sea grid or has a hydraulic path through inundated grids to the sea (i.e., $y_{it\mathcal{S}\mathcal{S}'} = 1$).

With the temporary flooding during a period t and under water level states $\mathcal{S} \in \Xi_{t-1}$ and $\mathcal{S}' = (s', \hat{s}') \in \Xi_t$ with positive transition probabilities, a grid $i \in R_t^{\mathcal{S}'} \cup Q_t^{\mathcal{S}'} \cup \hat{R}_t^{\mathcal{S}'} \cup \hat{Q}_t^{\mathcal{S}'}$ incurs

hurricane storm surge flooding, if and only if its elevation (i.e., $x_{it\mathcal{S}}$ for $i \in R_t^{\mathcal{S}'} \cup \hat{R}_t^{\mathcal{S}'}$, and h_i for $i \in Q_t^{\mathcal{S}'} \cup \hat{Q}_t^{\mathcal{S}'}$) is below water level $\hat{s}'$, it is a neighbor of the sea grid or has a hydraulic connection to the sea via a path through flooded grids (permanent or temporary), and it is not inundated. A grid $i \in R_t^{\mathcal{S}'} \cup Q_t^{\mathcal{S}'} \cup \hat{R}_t^{\mathcal{S}'} \cup \hat{Q}_t^{\mathcal{S}'}$ being a neighbor of the sea grid or existence of a hydraulic path to the sea from this grid is captured by the binary variable $\hat{y}_{it\mathcal{S}\mathcal{S}'}$, which is equal to one if grid i is a neighbor of the sea or such a path exists, and zero otherwise. Inequalities (3.14)-(3.20) along with the objective function, use variables $\hat{y}_{it\mathcal{S}\mathcal{S}'}$ to assure that a grid $i \in R_t^{\mathcal{S}'} \cup Q_t^{\mathcal{S}'} \cup \hat{R}_t^{\mathcal{S}'} \cup \hat{Q}_t^{\mathcal{S}'}$ is temporarily flooded during period t (i.e., $z_{it\mathcal{S}\mathcal{S}'} > 0$) if and only if its elevation is below the sea level $\hat{s}'$, it is a neighbor of the sea grid or has a hydraulic path through flooded grids to the sea (i.e., $\hat{y}_{it\mathcal{S}\mathcal{S}'} = 1$), and it is not inundated (i.e., $w_{it\mathcal{S}\mathcal{S}'} = 0$ if $i \in R_t^{\mathcal{S}'} \cup Q_t^{\mathcal{S}'}$).

3.3.3 Computational complexity and solution approach

Having defined the FRM problem to this point in Section 3.3, we address its computational complexity in Theorem 1 below, and establish that the decision version of this problem is indeed NP-complete.

Theorem 1. *The decision version of the FRM problem is NP-complete.*

We present the proof of Theorem 1 by reducing the well-known Knapsack problem (Karp 1972) to a special case of the FRM decision version. Given a planning horizon $\mathcal{T} = \{1, \ldots, t_{max}\}$, set $\boldsymbol{\Xi} = [\Xi_t : t \in \mathcal{T}]$ containing the sets of sea level states during each period $t \in \mathcal{T}$, sea level state transition probabilities $\boldsymbol{p} = [p_t^{\mathcal{S}\mathcal{S}'} : t \in \{0, \ldots, t_{max}-1\}, \mathcal{S} \in \Xi_t, \mathcal{S}' \in \Xi_{t+1}]$, a coastal region with the associated set of land grids in the area of interest Φ and the area of relevance Ψ, initial elevations $\boldsymbol{h} = [h_i : i \in \Phi \cup \Psi]$, minimum elevation threshold m, sets $\boldsymbol{R} = [R_t^{\mathcal{S}} : t \in \mathcal{T}, \mathcal{S} \in \Xi_t]$, $\boldsymbol{Q} = [Q_t^{\mathcal{S}} : t \in \mathcal{T}, \mathcal{S} \in \Xi_t]$, $\hat{\boldsymbol{R}} = [\hat{R}_t^{\mathcal{S}} : t \in \mathcal{T}, \mathcal{S} \in \Xi_t]$, and $\hat{\boldsymbol{Q}} = [\hat{Q}_t^{\mathcal{S}} : t \in \mathcal{T}, \mathcal{S} \in \Xi_t]$ containing sets of land grids in Φ or Ψ at risk of permanent or temporary

flooding during each period $t \in \mathcal{T}$ and under each sea level state $\mathcal{S} \in \Xi_t$, a neighborhood set $\mathcal{N} = [N_t^{\mathcal{S}}(i) : t \in \mathcal{T}, \mathcal{S} \in \Xi_t, i \in R_t^{\mathcal{S}} \cup Q_t^{\mathcal{S}} \cup \hat{R}_t^{\mathcal{S}} \cup \hat{Q}_t^{\mathcal{S}}]$, cost c of elevating a grid in Φ by one meter at the start of a period, permanent flooding costs $\boldsymbol{g} = [g_i : i \in \Phi]$ and hurricane storm surge related flooding costs $\boldsymbol{f} = [f_i : i \in \Phi]$, a discount rate per-period λ, budgets per-period $\boldsymbol{b} = [b_t : t \in \mathcal{T}]$, minimum elevation threshold m, and a scalar $\mathcal{C}$, the decision version of the FRM problem, denoted by $\langle \mathcal{T}, \boldsymbol{\Xi}, \boldsymbol{p}, \Phi, \Psi, \boldsymbol{h}, \boldsymbol{R}, \boldsymbol{Q}, \hat{\boldsymbol{R}}, \hat{\boldsymbol{Q}}, \mathcal{N}, c, \boldsymbol{g}, \boldsymbol{f}, \lambda, \boldsymbol{b}, m, \mathcal{C} \rangle$, is defined as follows: "Is there an assignment to variables $\boldsymbol{x} = [x_{it\mathcal{S}} : i \in \Phi, t \in \mathcal{T}, \mathcal{S} \in \Xi_{t-1}]$ that satisfies Constraints (3.4)-(3.9) and (3.22)-(3.23) for which the value of Objective Function (3.3) (calculated with respect to Constraints (3.10)-(3.21) and (3.24)-(3.25)) is at most $\mathcal{C}$". Now we will employ a reduction from Knapsack problem to establish the intractability of the decision version of the FRM problem in Theorem 1.

Proof of Theorem 1. It can be easily verified that given a decision vector $\boldsymbol{x}^0$, verifying whether Constraints (3.4)-(3.9) and (3.22)-(3.23) are satisfied and the value of Objective Function (3.3) incorporating Constraints (3.10)-(3.21) and (3.24)-(3.25) is at most $\mathcal{C}$, can be done in polynomial-time. Therefore, the decision version of the FRM problem belongs to class NP.

Next, we prove that the decision version of the FRM problem is NP-hard by performing a polynomial-time reduction from the well-known NP-complete Knapsack problem (Karp 1972). The decision version of the Knapsack problem, denoted as $\langle \boldsymbol{\alpha} = [\alpha_i : i \in \{1, \ldots, n\}], \boldsymbol{\beta} = [\beta_i : i \in \{1, \ldots, n\}], \sigma, \eta \rangle$, is "Given a finite set of n items each one with the value of α_i and weight of β_i, a desired total value σ, and a knapsack with weight capacity of η, is there a subset $J \subseteq \{1, \ldots, n\}$ such that $\sum_{i \in J} \beta_i \leq \eta$ and $\sum_{i \in J} \alpha_i \geq \sigma$?"

Given an instance of the Knapsack decision problem $\langle \boldsymbol{\alpha}, \boldsymbol{\beta}, \sigma, \eta \rangle$, we transform this instance into a special case of the decision version of the FRM problem in polynomial-time as follows:

1. The length of the planning horizon is set to be equal to one (i.e., $\mathcal{T} = \{1\}$).

2. There is only one sea level state during period 1, which is

$$(s_1, \hat{s}_1) = (\max\{\beta_i : i \in \{1,\ldots,n\}\} + 1, \max\{\beta_i : i \in \{1,\ldots,n\}\} + 1).$$

Therefore, $\boldsymbol{\Xi} = [\Xi_1 = \{(s_1, \hat{s}_1)\}]$.

3. The transition probability $p_0^{(0,0)(s_1,\hat{s}_1)}$ is also set to be equal to one. Therefore, $\boldsymbol{p} = [p_0^{(0,0)(s_1,\hat{s}_1)} = 1]$.

4. The set of land grids in the area of interest is composed of n grids (i.e., $\Phi = \{1,\ldots,n\}$). The set of land grids in the area of relevance is empty (i.e., $\Psi = \emptyset$). Each land grid in Φ is assumed to be surrounded by the sea, and land grids in Φ themselves do not share any border.

5. The initial elevation of each land grid $i \in \Phi$ (i.e., h_i) is set to be equal to $\max\{\beta_i : i \in \{1,\ldots,n\}\} + 1 - \beta_i$. Therefore, $\boldsymbol{h} = [\max\{\beta_i : i \in \{1,\ldots,n\}\} + 1 - \beta_i : i \in \Phi]$.

6. The set of land grids in Φ at risk of permanent flooding during the single period considered and under the single sea level state given (i.e., $R_1^{(s_1,\hat{s}_1)}$) is set to be equal to $\{1,\ldots,n\}$. The set of land grids in Φ only at risk of temporary flooding during the single time period and under the single sea level state (i.e., $\hat{R}_1^{(s_1,\hat{s}_1)}$) is set to be empty. Therefore, $\boldsymbol{R} = [R_1^{(s_1,\hat{s}_1)} = \{1,\ldots,n\}]$ and $\hat{\boldsymbol{R}} = \boldsymbol{Q} = \hat{\boldsymbol{Q}} = [\emptyset]$.

7. The set containing neighbors of each land grid $i \in R_1^{(s_1,\hat{s}_1)}$ within the at-risk network corresponding to the given single period and single sea level state (i.e., $N_1^{(s_1,\hat{s}_1)}(i)$) is set to only contain the sea grid. Therefore, $\mathcal{N} = [\{o\} : i \in \{1,\ldots,n\}]$. The at-risk network for the given single time period and sea level state is illustrated in Figure 16 below.

8. The cost of raising a grid $i \in \Phi$ by one meter (i.e., c) is set to be equal to one dollar.

9. The cost of losing a grid $i \in \Phi$ due to inundation (i.e., g_i) is set to be equal to $\alpha_i + \beta_i$, and the cost of hurricane storm surge flood damage to a grid $i \in \Phi$ (i.e., f_i) is set to be equal to zero. Therefore, $\boldsymbol{g} = [\alpha_i + \beta_i : i \in \Phi]$ and $\boldsymbol{f} = [0 : i \in \Phi]$.

10. The discount rate per-period (i.e., λ) is set to be equal to one.

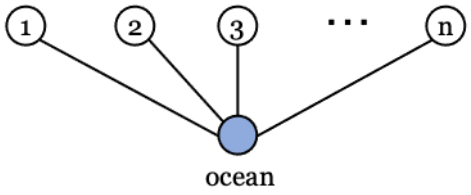
11. The budget for the single considered period (i.e., b_1) is set to be equal to η. Therefore, $\boldsymbol{b} = [\eta]$.

12. The minimum elevation threshold m is set to be equal to $\min\{\beta_i : i \in \Phi\}$.

13. Finally, the upper bound on the total cost in the FRM model (i.e., $\mathcal{C}$) is set to be equal to

$$\sum_{i=1}^{n} (\alpha_i + \beta_i) - \sigma.$$

Figure 16: Illustration of the at-risk network used in the proof of Thereom 1.



We now show that the answer to the Knapsack decision problem is a "yes", if and only if the answer to the constructed special case of the FRM decision problem is a "yes". Suppose, there exists a set $J \subseteq \{1, \ldots, n\}$ that satisfies the Knapsack capacity constraint (i.e., $\sum_{i \in J} \beta_i \leq \eta$) and the total value for items in set J is at least σ (i.e., $\sum_{i \in J} \alpha_i \geq \sigma$). Consider the solution corresponding to elevating the land grids in set J in the constructed FRM problem so that each grid in this set is raised up to the permanent (or storm-related) sea level, and is not permanently flooded. The nodes in set J each cost β_i when elevated, therefore, incurring a total investment cost of $\sum_{i \in J} \beta_i \leq \eta$, meeting the FRM budget limit. So, this solution satisfies Constraints (3.4)-(3.9) and (3.22)-(3.23). Additionally, as shown in Equation (3.26) below, the value of Objective

Function (3.3) (calculated with respect to Constraints (3.10)-(3.21) and (3.24)-(3.25)) is also at most $\mathcal{C}$.

$$EIC + EFC =$$

$$\sum_{i \in J} \beta_i + \sum_{i \in \{1,\ldots,n\} \setminus J} (\alpha_i + \beta_i) = \sum_{i=1}^{n} (\alpha_i + \beta_i) - \sum_{i \in J} \alpha_i \leq \sum_{i=1}^{n} (\alpha_i + \beta_i) - \sigma = \mathcal{C} \quad (3.26)$$

Conversely, suppose there exists a set of land grids that are built up to protect against permanent (and temporary) flooding in a way that the investment cost is bounded above by the FRM budget η (i.e., Constraints (3.4)-(3.9) and (3.22)-(3.23) are satisfied) and the value of Objective Function (3.3) (calculated with respect to Constraints (3.10)-(3.21) and (3.24)-(3.25) is at most $\mathcal{C} = \sum_{i=1}^{n} (\alpha_i + \beta_i) - \sigma$. Let $J \subseteq \{1, \ldots, n\}$ denote the set of land grids in the FRM model that are not permanently flooded. Consider a solution to the Knapsack problem is which we select the items in set J. Since the grids in J are not permanently flooded, then they should have been elevated to a level higher than or equal to the permanent (or storm-related) water level. Therefore,

$$\sum_{i \in J} \beta_i \leq EIC \leq \eta, \quad (3.27)$$

which means that the items in set J satisfy the Knapsack capacity constraint. Additionally, we know that

$$EIC + EFC \leq \mathcal{C} = \sum_{i=1}^{n} (\alpha_i + \beta_i) - \sigma.$$

Since

$$EFC = \sum_{i \in \{1,\ldots,n\} \setminus J} (\alpha_i + \beta_i),$$

then, by Inequality (3.27), we have

$$\sum_{i \in J} \beta_i + \sum_{i \in \{1,\ldots,n\} \setminus J} (\alpha_i + \beta_i) \leq \sum_{i=1}^{n} (\alpha_i + \beta_i) - \sigma.$$

This means

$$\sum_{i \in J} \alpha_i \geq \sigma,$$

which indicates that the total value for items in set J is at least σ.

Therefore, the decision version of the FRM problem is NP-hard. Combined with the earlier mentioned fact that this decision problem belongs in class NP, we also conclude that the decision version of the FRM problem is NP-complete. □

Given the intractability of the FRM problem and considering the extremely large number of variables and constraints in Formulation (3.3)-(3.25), solving this problem by classical branch-and-cut algorithms available via commercial solvers is impractical and computationally expensive. In our case study in Section 3.4, to solve the FRM problem within practical time limits and obtain managerial insights, we employed two different methods: a simulation-based approach and a scenario-based approach. The reader is referred to Sections 3.5.1 and 3.5.2 for the details of these proposed methods.

3.4 Case study

In this section, we employ our proposed model to develop a decision-support system for building levees to protect the neighborhood of East Boston, using only publicly available data. This research is the fifth paper of a series of articles on climate change adaptation in Boston. Douglas et al. 2012 identified major obstacles and incentives for adaptation based upon representative focus groups, Kuhl et al. 2014 examined in more details some of the challenges and implementation barriers for evacuation in an environmental-justice community, Kirshen, Ballestero, et al. 2018 addressed how to involve vulnerable exposed populations in urban adaptation strategy planning and the use of multi-stakeholder collaborative processes, and Zandvoort et al. 2019 studied how pathway thinking can be used to inform landscape architects to design sustainable and adaptive landscapes.

3.4.1 Data and experiment settings

The neighborhood of East Boston presents a large coastal front with a relatively dense population and a variety of building structures. This region has a nonuniform topography with several hilly areas that overlook the city and are not at risk of flooding because of their higher elevations. As shown in Figure 17(a), the neighborhood of East Boston marked within the solid boundary line is the region of interest in our case study. The surrounding region includes the neighboring towns of Winthrop and Revere, as well as Boston Logan airport owned by MassPort and not under control of the city (Aloisi 2017). By following the procedure outlined in Section 3.3.1.2, we create the network for our model by overlaying the grids as shown in Figure 17(b), and then Figure 17(c) shows the identified areas of interest and relevance in our case study. In creating the grid attributes (i.e., h_i, f_i, and g_i), we use open source tax appraisal data from the City of Boston (Boston 2020, 2016) and Light Detection and Radar (LIDAR) elevation data from the Massachusetts Commonwealth (MassGIS 2017). Full details of the data sources and transformations conducted to create the network and estimate grid parameters h_i, f_i, and g_i are found in Section 3.4.2.

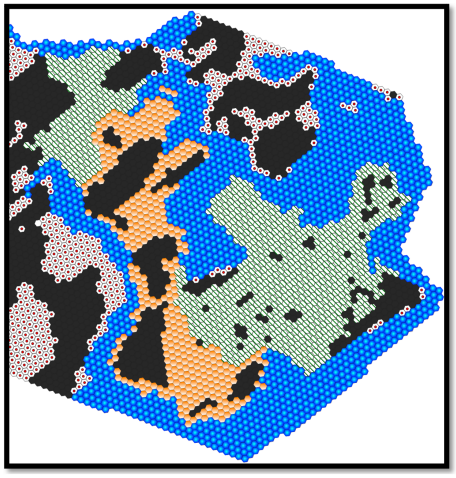
Given that the model's parameters are estimated based on available open-source data, there might be inaccuracies associated with the estimated values. Therefore, we used a range of possible values for some of the primary parameters as shown in Table 17, and conducted a sensitivity analysis in our experiments to show the behavior of the optimal objective as the values for these parameters change. We go through data used in the model in greater detail in Section 3.4.2 below, but justify the range of values chosen for each one of these primary parameters as follows. In evaluating flooding of coastal mega-cities, Aerts et al. 2014 provide multiple sources and a range of discount rates (d) applicable to studies evaluating flood protection investment, leading to our chosen values of 3, 5, and 7 percent. We chose minimum levee heights (m) of 1, 3, and 5 meters, which is also based on past studies showing a breakdown of

Figure 17: Map of East Boston and overlays to make the network.

(a) Overlay of Region (East Boston marked).

(b) Hex grids overlaid.

(c) Resulting Grid Network.



heights for levee projects around the globe as discussed in Jonkman et al. 2013. To determine the levee build cost (c), we initially started with a linear estimate of $450 per foot build-up per linear foot of wall (Hecht and Kirshen 2019) as a potential lower value. In order to provide for assessing sensitivity, we evaluated and chose values of $5M, $15M, and $25M per kilometer of wall built one meter in height based on historical and regional factors affecting these values as discussed in Jonkman et al. 2013. To conduct sensitivity analysis on the values used for storm surge flooding, we shift the slope of the estimated depth damage function curve ($\bar{f}_i$) by $\pm 25\%$ to attain f_i. Finally, we determined the budget values (b_t) by initially running the full range of simulations with an unlimited budget. We then looked for clear breakpoints for the initial period spends across all scenarios to use them as possible values for budgets.

Table 17: Parameter values used to conduct sensitivity analysis.

Parameter	**Values used**	**Units**
Discount rate (d)	3, 5, 7	%
Minimum elevation increase (m)	1, 3, 5	meters
Grid elevation cost (c)	5, 15, 25	$M/km per m
Storm flood damage curve (f_i)	$0.75\bar{f}_i$, $\bar{f}_i$, $1.25\bar{f}_i$	$M/m
Budget per period (b_t)	0, 25, 50, 75, 100, 150, 200, 400, 600	$M

3.4.2 Data Collection and Transformation for Case Study

Given that East Boston is the location of interest in our case study, we need to compile the necessary data to create the actual model parameter values for this region. As we aim to create a useful model for any coastal area, we only use available open-source datasets. Figure 18 shows the breakdown of the data references used for compiling the input parameters, with three main categories: sea level, network, and financial data. Given the model's inherent spatial nature, the network-related data are captured and transformed from Geographical Information System (GIS) open-source data. All elevations are referenced to the North American Vertical Datum of

1988 (NAVD88) (Vanicek 1991) to ensure consistency in the spatial data used. Figure 18 highlights the three main elements of the data preparation, including identifying the data sources, interactions, and transformations to model parameters. In the following subsections, we will detail the data sources, assumptions made to apply the data accordingly, and the transformation methods needed to produce the model parameters.

3.4.3 Sea level data

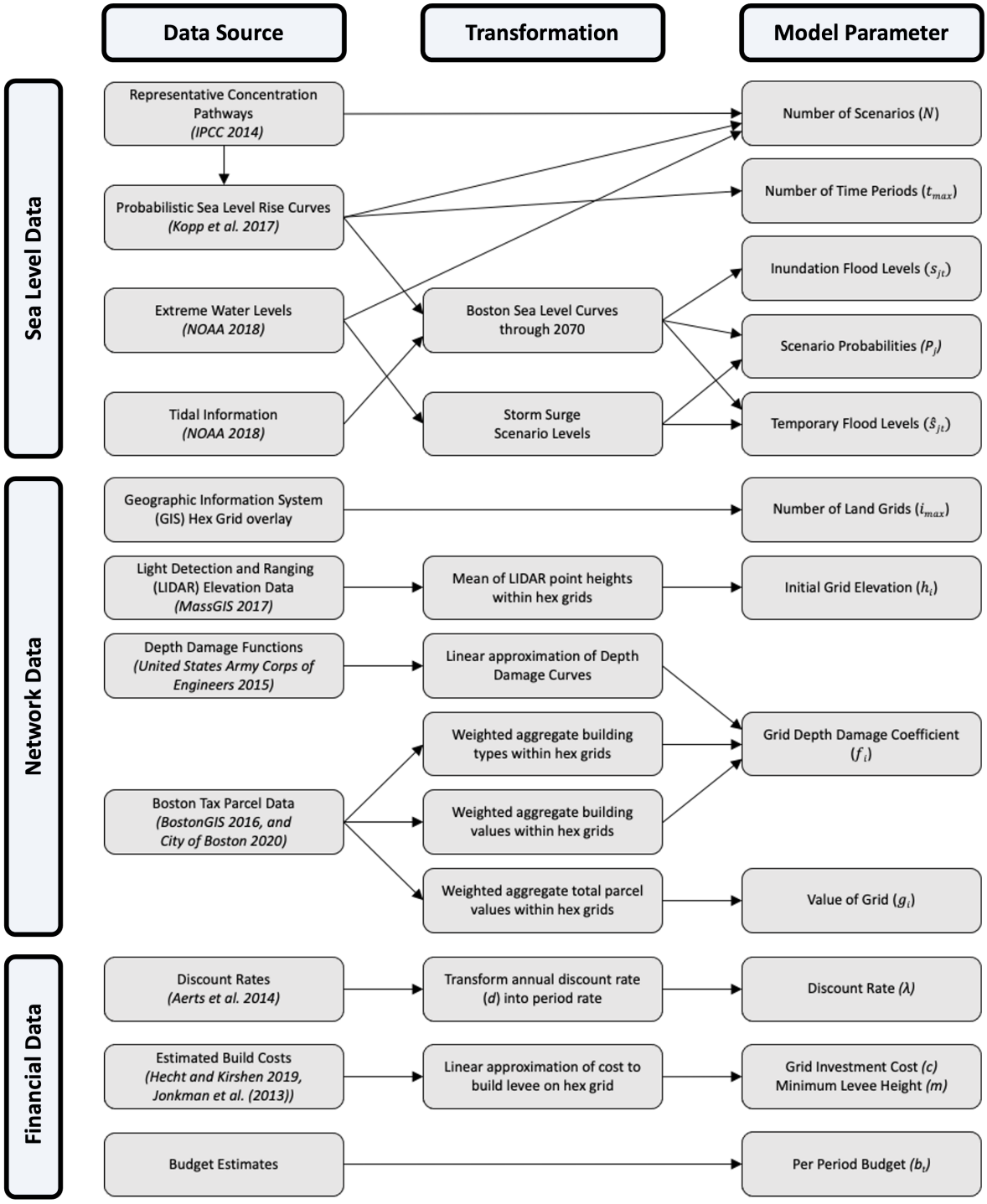
To capture the full range of sea level states (parameters s and $\hat{s}$) over time, we need to incorporate three important factors. These three factors include the climate-induced sea level rise heights and probabilities, tidal range captured by tidal data, and hurricane storm surge heights and probabilities. We will discuss each of these factors in the next four subsections.

3.4.3.1 Climate-induced sea level rise

Before discussing the actual data source for the potential sea level rise scenarios, it is important to briefly mention the Representative Concentration Pathway (RCP) carbon emission trajectories adopted by the Intergovernmental Panel on Climate Change (IPCC 2014). The pathways describe different possible climate futures. All of them are possible and depend on the volume of greenhouse gases emitted now and in the future. Most sea level rise analyses align with these RCP trajectories; therefore, we will segment our data into the four original RCP trajectories of RCP2.6, RCP4.5, RCP6.0, and RCP8.5. Given the non-deterministic future of greenhouse gas emissions reductions, we will treat each of the four RCPs as independent and equally likely.

There are many sources of potential sea level rise within the literature. Looking back over the last few decades (Garner et al. 2018), the one common theme in all the predictions is substantial uncertainty in the best to worst-case sea level rise predictions. One can follow the flow of sea level rise predictions from both plausible or probabilistic perspectives (Ruckert,

Figure 18: Outline of data sources and their incorporation into model's parameters.

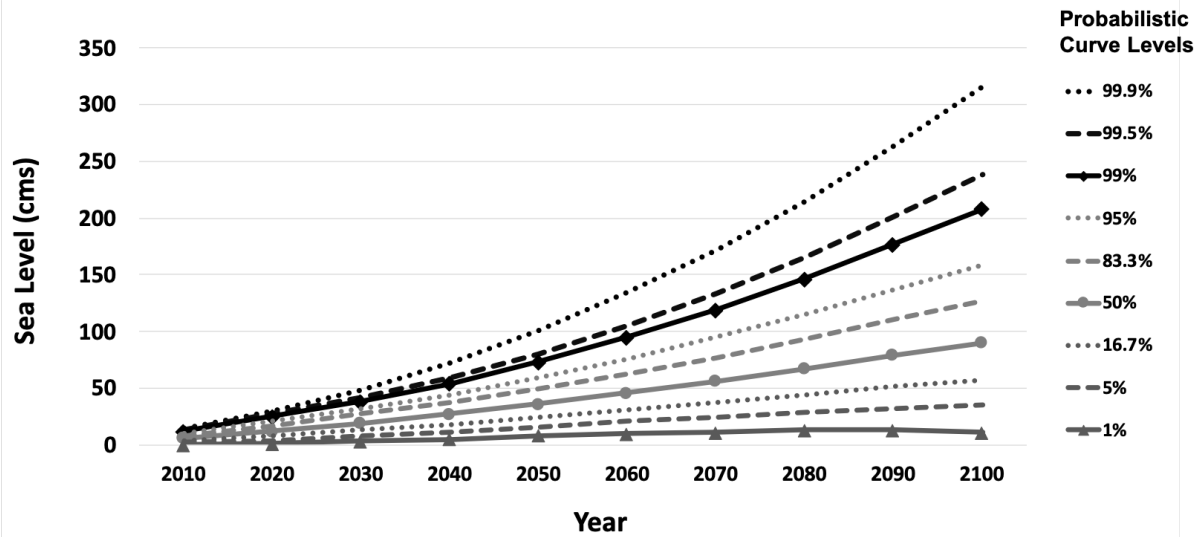


Srikrishnan, and Keller 2019). Given these ever-evolving sea level rise predictions, the model we propose leaves open the ability to incorporate a range of sea level rise predictions. This ability will allow us to be flexible in capturing the expected costs of the inundation and hurricane storm surge flood damages as sea level rise understanding evolves. For parameters s in the sea level states of our model, we utilize the sea level rise predictions captured in Kopp et al. 2017 as the authors provide nine sets of probabilistic sea level curves for each of the four RCPs. These curves predict sea level rise each decade through the year 2100, with further 50-year predictions out to the year 2300. Here we will only be using the period data up to 2070 starting from 2020, which results in five decade-long periods in our model. The data from Kopp et al. 2017 incorporate applicable sea level rise causes from multiple sources, including ice sheet melt, glacier and ice cap melt, land water storage, oceanographic processes, glacial isostatic adjustments, tectonics, and other non-climate local effects. The authors made their model calculations publicly available for many coastal areas associated with local tidal gauges, including the Boston tidal gauge.

In Kopp et al. 2017, there are nine probabilistic curves associated with each RCP, with the example for RCP8.5 shown in Figure 19. For explanation, the curves in Figure 19 represent the probability that the sea level rise will be less than or equal to that sea level associated with that curve at that point in time. For instance, in the year 2100, one can see that the sea level for the 50% curve will be 90 cm. That is interpreted to mean that the sea level is 50% likely to be at or below that level in the year 2100. Similarly, when looking at the 99.9% curve in the year 2100, we would expect the sea level is 99.9% likely to be at or below 316 cm in the year 2100. This data provides us with the required sea level curves, the periods t we will need, and a basis for assigning their probabilities, as discussed in Section 3.4.3.4 below.

Figure 19: Boston sea level rise curves for Representative Concentration Pathway 8.5 (RCP8.5). *Source:* Graph from Kopp et al. 2017

.



3.4.3.2 Tidal range

To obtain the tidal value used in the model, we reference the National Oceanic and Atmospheric Administration (NOAA) tide gauge based in Boston harbor (NOAA 2018). In using the tidal data, we assume the sea level rise data will account for any changes over time discussed in Section 3.4.3.1. The data we use is relative to North American Vertical Datum of 1988 (NAVD88) (Vanicek 1991), providing us with the Mean Higher High Water level of 1.52 meters as the highest elevation caused by high tides in our sea level data. We add this high tide value to the values reported in Kopp et al. 2017 to obtain the parameters s used in the sea level states of our model.

3.4.3.3 Hurricane storm surge levels

To determine our storm surge levels, we use the hurricane storm surge data from the National Oceanic and Atmospheric Administration (NOAA) tide gauge based in Boston harbor (NOAA 2018). We assume that storm surge heights over time will remain constant based on

the hurricane storm surges and any change in absolute height are captured by factors discussed in Section 3.4.3.1. Table 18 shows the surge levels and associated exceedance probability levels for the four hurricane storm surge levels we use in our model. The exceedance storm surge levels represent the likelihood of a storm surge height above the Mean Higher High Water high tide level captured in Section 3.4.3.2. For instance, on average, the 1% level will be exceeded in only one year per century, while the 10% level will be exceeded in ten years per century. As shown in Table 18, we use the median value for the four exceedance probabilities for the Boston tidal gauge published by NOAA (NOAA 2018) to simplify the model development and to reduce the risk of overly long model runtimes. We add these values to the parameters s, as determined above, to obtain parameters $\hat{s}$ in the sea level states of our model.

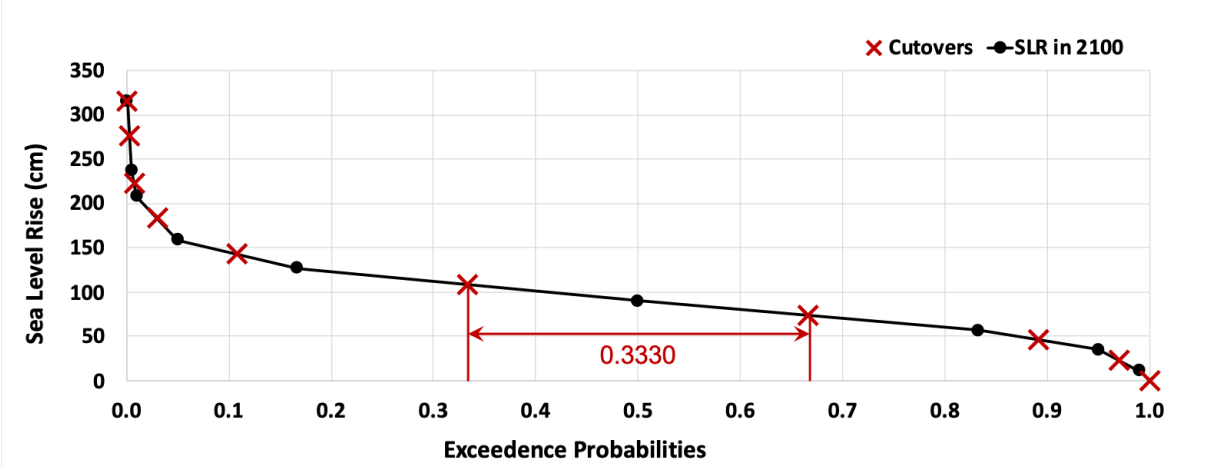
Table 18: National Oceanic and Atmospheric Administration storm surge levels.

Storm Surge	Exceedance Probability	Water Level above Mean Higher High Water
100-year	0.01	1.41 m
10-year	0.10	1.07 m
2-year	0.50	0.81 m
1-year	0.99	0.56 m

3.4.3.4 Sea level states paths and their probabilities

To calculate the associated probabilities for each curve under a given RCP, we first transform the RCP's probabilistic curves into time-dependent exceedance curves. Figure 20 shows an example of an exceedance curve for RCP8.5 in the year 2100. Using a linear interpolation methodology (Kirshen et al. 2012), we calculate an estimated probability for each probabilistic curve for a given RCP at a given time. For explanation, considering the exceedance curve in Figure 20, we estimate the probability for the 50%-probabilistic curve by taking the halfway points between the 50%- and 83.3%- curves and the 50%- and the 16.7%- curves, and then measure the distance between those two midway points as shown in Figure 20. It is important

Figure 20: Representative Concentration Pathway 8.5 (RCP8.5) sea level rise exceedance curves for year 2100 adapted for Boston. *Source:* Graph from Kopp et al. 2017.



to note that for a given p%-curve in a given RCPα, the actual sea level increases over time, while the estimated probability for the curve (denoted by $P_{curve}^{\alpha,p}$) does not change as the distance between the two midway points remains constant over time. We show the estimated probabilities for each of the nine p%-curves in Table 19. We apply the same technique to each β-year storm surge exceedance curve to determine estimated probabilities P_{surge}^{β} as shown in Table 20. As mentioned, the probability distribution for the hurricane storm surge levels is assumed to be fixed over time.

Now that we have estimates for the probabilities of each RCP's probabilistic curves and the probabilities of hurricane storm surge instances, we can determine the complete set of possible sea level states paths and their probabilities by bringing this information together. Given an RCP α, to calculate the probability for each of the associated sea level states paths, we treat the probabilities for sea level rise curves $P_{curve}^{\alpha,p}$ and hurricane storm surge levels P_{surge}^{β} as independent. By doing so, we can calculate the estimated probability for a path composed of a p%-curve sea level rise in a given RCP α and a β-year storm surge (denoted by $P_{\alpha,p,\beta}$) as shown in Equation (3.28). When combining the four RCPs with equal probabilities, we

Table 19: Estimated probabilities for sea level rise curves for a given Representative Concentration Pathway α.

SLR Curve (p)	$P_{curve}^{\alpha,p}$
99.9%	0.0030
99.5%	0.0045
99.0%	0.0225
95.0%	0.0785
83.3%	0.2250
50.0%	0.3330
16.7%	0.2250
5.0%	0.0785
1.0%	0.0300

Table 20: Estimated probabilities for four storm surge curves.

Storm	Surge curve	P_{surge}^{β}
100 year	1%	0.055
10 year	10%	0.245
2 year	50%	0.445
1 year	99%	0.255

produce a total of $N = 144$ unique sea level states paths, each with the probability of $0.25 * P_{\alpha,p,\beta}$. These will be the paths we use in our simulation-based and scenario-based solution approaches discussed in Sections 3.5.1, 3.5.2, and 3.5.3. In Table 31 in Appendix B, we show the probabilities for each of the 144 sea level states path and the associated values of s and $\hat{s}$ for each path in each period.

$$P_{\alpha,p,\beta} = P_{curve}^{\alpha,p} * P_{surge}^{\beta} \tag{3.28}$$

3.4.4 Network data

To create our network, we start by overlaying a grid in the geographic region of interest, which is East Boston and the surrounding region in our case study. We chose to create our grid overlay using hexagonal grids. Although many flood studies use square grids, we use hex grids for two main reasons. First, hex grids provide a tessellated grid structure with equidistant grids between their centers. Second, hex grids have distinct boundary lines with limited ambiguity. The reduced boundary ambiguity overcomes the challenge of determining water flow with square cells as to whether it is just across the four shared sides (four neighbors) or also across the corners (eight neighbors). Using hex grids, the network nodes' neighbors are just the six neighbors sharing same-length borders, which are very straightforward to define and account for flow between nodes. These are the two main reasons we use hex grids to develop our network, but the interested reader can find more background on the pros and cons of hex grids in geospatial analysis applications (Birch, Oom, and Beecham 2007; De Sousa et al. 2017).

We conduct all geographical data inputs and transformations for modeling purposes using the Quantum Geographic Information Software (QGIS) application, specifically version 3.10. To create our grid overlays, we use the MMQGIS Python library that allows creating a grid overlay by providing information covering the geographic region of interest and individual grid size. In determining the hex grids' size, we have to trade off between hex grid area size and expected model runtime. On the one hand, the smaller we make the hex grids, the larger the number of overall grids there will be, which will result in a more complex model and longer runtimes to reach suitable solutions. Conversely, the grids should be sufficiently small to keep the model relevant for protecting discrete areas that can be built up over time to provide potential protection for the network. In evaluating the model, we evaluated various size grid overlays to get a sense of model runtime compared to realistic grid sizes for planning. In the end, we have used a grid overlay with hex grids with a side-to-side length of 100 meters and

an area of 8660 m^2, resulting in 404 at-risk land grids in the region of interest. We will create our network within this grid structure and conduct the necessary transformations to create grid attributes. In the following sections, we will discuss the assumptions and transformations needed to calculate the grid attributes for elevation (h_i), depth damage coefficient (f_i), and value (g_i) used in the model.

3.4.4.1 Grid elevation data

To determine each grid's elevation, we start with available Light Detection and Ranging (LIDAR) data for Massachusetts. This dataset is open-source and available online from the MassGIS repository (MassGIS 2017). MassGIS terrain data comes in varying degrees of granularity. The raster file data used in this analysis had elevation data by one square-meter of land coverage. The surveys that captured this data were conducted in 2013-2014 to assist in evaluating hurricane storm damage and erosion of the local environment as part of the United States Geologic Survey (USGS) response to Hurricane Sandy. For the bathtub-type flood analysis used in this study, LIDAR data with this level of accuracy are considered to be of sufficient quality (Gesch 2018). To capture each hex grid's elevation, using the QGIS software, we sampled the elevation from the LIDAR raster values within a given hex grid and then assigned the mean as the elevation of the hex grid. Given the overall exploratory nature of the model, we must accept some limitations to the ability to capture the grid elevations perfectly.

3.4.4.2 Tax appraisal data

We use the open-source Geographic Information System (GIS) tax appraisal data available from the City of Boston to obtain the value and types of buildings located within the study area. Specifically, we use the tax appraisal data from the 2016 tax year because it was the fullest dataset available at the beginning of this study. The tax appraisal dataset is available from the City of Boston at their GIS data repository (BostonGIS 2016). Supporting information

and metadata were also needed and available from the Analyze Boston website (Boston 2020), including the tax parcel data key and associated property occupancy codes.

It was necessary to profile the data to ensure its quality and applicability due to it being an open-source dataset. There were several elements of the data that required subsequent validation. For example, to properly determine the building structure, we needed to know how many floors were in a given building. Unfortunately, for multi-unit buildings with multiple taxpayers, the dataset's records only had the number of floors associated with a taxable unit (e.g., condo or apartment). To correctly classify the number of floors in a building, we used secondary validation to visually inspect the buildings using Google Streetview (Anguelov et al. 2010) to assess the total number of floors for multi-tenant buildings. The number of floors is essential in determining the classification to assign each building for assigning the appropriate depth damage function. We also used Google Streetview for additional validation when a building in the tax database no longer existed or a building did exist but was not in the tax database. Finally, a subset of buildings and lots in the tax database was categorized as "exempt" for taxation purposes, and had zero appraised value. Examples of such buildings included churches, government buildings, hospitals, and state transportation nodes. Because these tax-exempt parcels have a value of zero, exempt parcels in the model understate the potential total damages. That said, in the overall dataset, only 2% of the tax parcels were exempt, so we leave their value as zero for purposes of this model evaluation.

3.4.4.3 Depth damage curve estimation

To determine the flood damage associated with temporary related flooding from hurricane storm surges, we need to use the tax parcel data to capture the type of building within each tax parcel. We also need a method of determining flood damage based on the level of flooding. Depth damage functions are typically defined by interpolating flooding depth and damage data through systematic procedures that analyze historical flood events or insurance claims data, or

even from synthetic damage data from simulation models (Armal et al. 2020). The resulting function can be used to estimate how much damage a building may experience based on a given depth of flooding.

For purposes of this study, we adapt the hurricane storm surge-related depth damage curves captured following Hurricane Sandy-related flooding in 2012. In the aftermath of Hurricane Sandy, the US Army Corps of Engineers worked with a group of experts to elicit depth damage functions for the building types common in the area affected by Hurricane Sandy. The result was a listing of 14 building types and the associated depth damage function for these buildings when experiencing inundation, erosion, and wave impacts (USACE 2015). For our model with the bathtub-type flooding methodology, we only use the inundation-related damage curves from (USACE 2015) to develop our associated depth damage functions. More specifically, of the 14 building types identified within the study, there are five that we apply to the buildings in our model. Specifically, we used the depth damage curves for one-story commercial, one-story residential, two-story residential, three-story apartment, and high-rise buildings. Not every building in East Boston fits perfectly into these categories, so we had to apply a judgment call in some instances. For instance, there are many three- and four-story residential buildings in East Boston with very similar characteristics. Given these similarities, we put all of these buildings into the three-story apartment category. Table 21 shows the breakdown of buildings and their type classification. Due to our model's linear nature, we applied a linear approximation of the depth damage curves in (USACE 2015), with all costs starting at zero. This linearization method introduces potential errors into the model, so we check its sensitivity in the model by running three potential values at 125%, 100%, and 75% of the estimated values captured using our method. Using the linear approximation for each curve, we arrive at the coefficient of the depth damage function for each building in a tax parcel as shown in Table 21.

Table 21: Breakdown of building types in East Boston and associated depth damage coefficients.

Building Type	Count	Depth Damage Coefficient per meter
Three Story Apartment	3290	0.203
Two Story Residential	2040	0.203
One Story Commercial	206	0.233
One Story Residential	155	0.246
High Rise	6	0.108

3.4.4.4 Calculating flood loss values f_i and g_i

With each building's values and types now captured in the tax parcel dataset, it is necessary to transform the applicable tax parcel data within each hex grid into the flood loss coefficients that are the attributes for each hex grid. We next derive the flood cost parameters f_i and g_i by weighting and aggregating the previously described tax appraisal data. Below, we describe the steps necessary to calculate these two grid monetary-related attributes as discussed in (Molenaar 1998).

Within the tax dataset, tax entities represent units that are taxable within a tax parcel. There can be more than one tax entity in a tax parcel, but a tax entity can only reside in one and only one tax parcel. We represent the tax entity by index k, where there are K = 7,979 tax entities in the dataset (i.e., $k \in \{1, \ldots, K\}$). Tax parcels are the geographic areas within the data that contain one or more tax entities. Tax parcels will be represented by index l, where there are L = 6,467 tax parcels in the dataset (i.e., $l \in \{1, \ldots, L\}$).

Tax entities have three appraised value fields in the tax dataset: *building value*, *lot value*, and *total value* consisting of the sum of building and lot values. In this analysis, for a given tax entity k, we use the appraised building value (denoted by BV_k) to determine the flood damage costs from temporary flooding, and use the appraised total value (denoted by TV_k) to determine

the loss caused due to inundation. We used the appraised values available via the Boston 2016 tax parcel data (Boston 2020, 2016).

Using the information we have for each tax entity, the tax entity values can then be aggregated into the applicable geographic-based tax parcels as described next.

- Appraised Building Value for Tax Parcel l is

  $$PBV_l = \sum_{k \in l} BV_k \quad \forall l \in \{1,..,L\}$$

- Appraised Total Value for Tax Parcel l is

  $$PTV_l = \sum_{k \in l} TV_k \quad \forall l \in \{1,..,L\}$$

Each tax parcel represents an area of land on the map where the tax entities are located. To allow for aggregating portions of the tax parcels that intersect with the hex grids, we need to determine a unit measure of value within each tax parcel. In this case, we can determine that unit measure by dividing the tax parcel's appraised values by the tax parcel area. Area of the Tax Parcel l is designated as A_l as measured in square meters and is determined by using QGIS software. Using this tax parcel area, we can determine the unit values for both building and total appraised values per square meter within each tax parcel as follows.

- Unit Building Value in parcel l is

  $$pbv_l = \frac{PBV_l}{A_l} \quad \forall l \in \{1,..,L\}$$

- Unit Total Value in parcel l is

  $$ptv_l = \frac{PTV_l}{A_l} \quad \forall l \in \{1,..,L\}$$

The aggregation's next step is to determine the depth damage value for a square meter of tax parcel l. We can do this by using the information in the data set that indicates the property type along with the applicable depth damage coefficient from Table 21. The depth damage coefficient is a linear approximation used to determine the dollars of property damage as a

percentage of value per depth of one meter flood. We designate this by D_l, then multiply it with the unit building value pbv_l to attain the per meter squared depth damage per meter of flood level. In other words, the damage value in a meter squared of tax parcel $l \in \{1,\ldots,L\}$ per meter of flooding is $pdv_l = D_l\ pbv_l$

Now that we have our values for parcel depth damage and total value, we need to aggregate those values into the hex grids i to develop the per grid estimated depth damage values $\bar{f}_i$ and the inundation flood loss value g_i. To this aim we will merge the two map layers and evaluate the overlapping areas of the tax parcels captured within each hex grid. The overlapping area between a hex grid $i \in \{1,\ldots,i_{max}\}$ and a tax parcel $l \in \{1,\ldots,L\}$ is denoted by OA_{il}, and is calculated using the QGIS software. Using these overlapping areas, we can determine the aggregated depth damage and total values associated with each hex grid as described next.

- Hex grid i depth damage value (dollars/meter) is

$$\bar{f}_i = \sum_{\{l\in\{1,\ldots,L\}:i\cap l\neq\emptyset\}} pdv_l OA_{il} \quad \forall i \in \{1,..,i_{max}\}$$

- Hex grid i total value (dollars) is

$$g_i = \sum_{\{l\in\{1,\ldots,L\}:i\cap l\neq\emptyset\}} ptv_l OA_{il} \quad \forall i \in \{1,..,i_{max}\}$$

3.4.5 Financial data

As discussed in Section 3.4.1, there are three financial components that we need to incorporate into our model. The first is the cost of investment to elevate a grid (i.e., parameter c). We treat the investment cost as the cost of building a levee on the grid. Typically a levee would have a range of values per meter of height build-up based on the levee height. To determine the cost in our model, we started with a linear estimate of $450 per foot build-up per linear foot of wall (Hecht and Kirshen 2019) (or $4,841.28 per meter per linear meter). Given the

expected importance of this parameter to the model solutions, and based on values from previous projects (Jonkman et al. 2013), we check the model's cost sensitivity using three values of \$5M, \$15M, and \$25M per kilometer of length for one meter of elevation (\$M/km per m). Converting these costs for the model to use per grid, the values we used were \$0.866M, \$2.598M, and \$4.33M per meter of elevation change for one grid. We do this by using the linear distance measurement of half the perimeter of our hex grids for the length to build.

The next financial parameter we include is a discount rate (d). Wide ranges of discount rates can be used for this type of cost-benefit analysis (Aerts et al. 2014). The value used often depends upon the decision-maker's financial risk tolerance. To account for a range of sensitivity, we incorporate three annual discount rate values of 3%, 5%, and 7%. Given that the time duration of each period in our model is ten years, we end up with an adjusted discount rate parameter in the model of $\lambda = (1+d)^{10}$.

The final financial parameter needed for our model is the budget spent to build up grid elevations for each period (i.e., b_t for all $t \in \{1,\ldots,t_{max}\}$). Given this study's exploratory nature, we use this financial value primarily as a per-period constraint to assess the model under various budget limitations. In reality, if running this model for a government agency, we would need to work with the agency to understand their current budget plans, constraints, and limits on what they would spend for the area under investigation. A simple way to arrive at an acceptable budget range is to look at the city's current budget for its infrastructure development and use that as a constant value throughout the study period. In the end, we use a band of budgets determined by initially running the model without the budget constraint and picking representative values based on the breakpoints for the initial period spends across all scenarios. This results in the case study budgets ranging from \$0 to \$600M per period as shown in Table 17 for the case study of East Boston. These values span a wide range, with the lower to mid-ranges being within a reasonable range of what we would expect the city to spend to avoid flood damages. The final factor we need to address in the financial data is that the model

periods are a decade long, while the flood damages are based on annual factors. Thus, we increase the expected damages by a factor of 10 in calculating all expected flood damages that the model determines.

3.5 Results

As discussed in Section 3.3.3, given the intractability and impracticality of solving the FRM model, we employ two different approaches to solving the case study, namely a simulation-based approach (Section 3.5.1) and a scenario-based approach (Section 3.5.2) to handle this challenge. To check the generalizability of the model, we subsequently use the same approaches on 50 randomly generated networks (Section 3.5.3). In all experiments, we built the related models using Python 3.8.5 and used Gurobi version 9.5.2 as the commercial solver. The experiments were conducted on Amazon Web Services EC2 c5n.4xlarge instances (AWS 2023). Given more than 100K individual optimization runs for the case study, we set the termination condition for each optimization run as either one percent optimality gap or one hour running time limit, whichever is observed first. Most (98.5%) of the runs completed by reaching the one percent optimality gap. In the remainder of this article, we refer to the best solution found before reaching the termination condition in each optimization run as the "optimal" solution. We discuss the details of our experiments next.

3.5.1 Simulation-based approach

The first approach is a simulation-based approach in which we solve the FRM Formulation (3.3)-(3.25) for each possible combination of chosen values for the model's parameters on simulated sea level states sample paths each composed of a collection of possible sea level states over the next five decades. In this approach, each run of the model is done on a single simulated sea level states path (i.e., we assume $\Xi_t = \{\mathcal{S}_t\}$, where $\mathcal{S}_t$ is the sea level state on

the considered path during period $t \in \{1, \ldots, 5\}$), which makes this solution approach computationally practical. Then, incorporating the probability associated with each simulated sea level state path, we compute and analyze the expected optimal objective of the FRM Formulation (3.3)-(3.25) across all simulation runs.

As discussed in Section 3.4.3, we sample sea level states paths over the next five decades in our simulation, using a collection of four RCP greenhouse gas emission pathways (IPCC 2014), nine probabilistic sea level rise curves per RCP (Kopp et al. 2017), and four potential hurricane storm surge levels (NOAA 2018). Using this sampling approach discussed in Section 3.4.3, we can derive the actual distribution for all possible simulated five-period sea level states paths. This distribution contains a total of 144 five-period sea level states paths and their probabilities of occurrence as shown in Table 31 in Appendix B. We use Ω to denote the set containing these 144 paths, and $p(\boldsymbol{S})$ to represent the probability associated with a path $\boldsymbol{S} \in \Omega$.

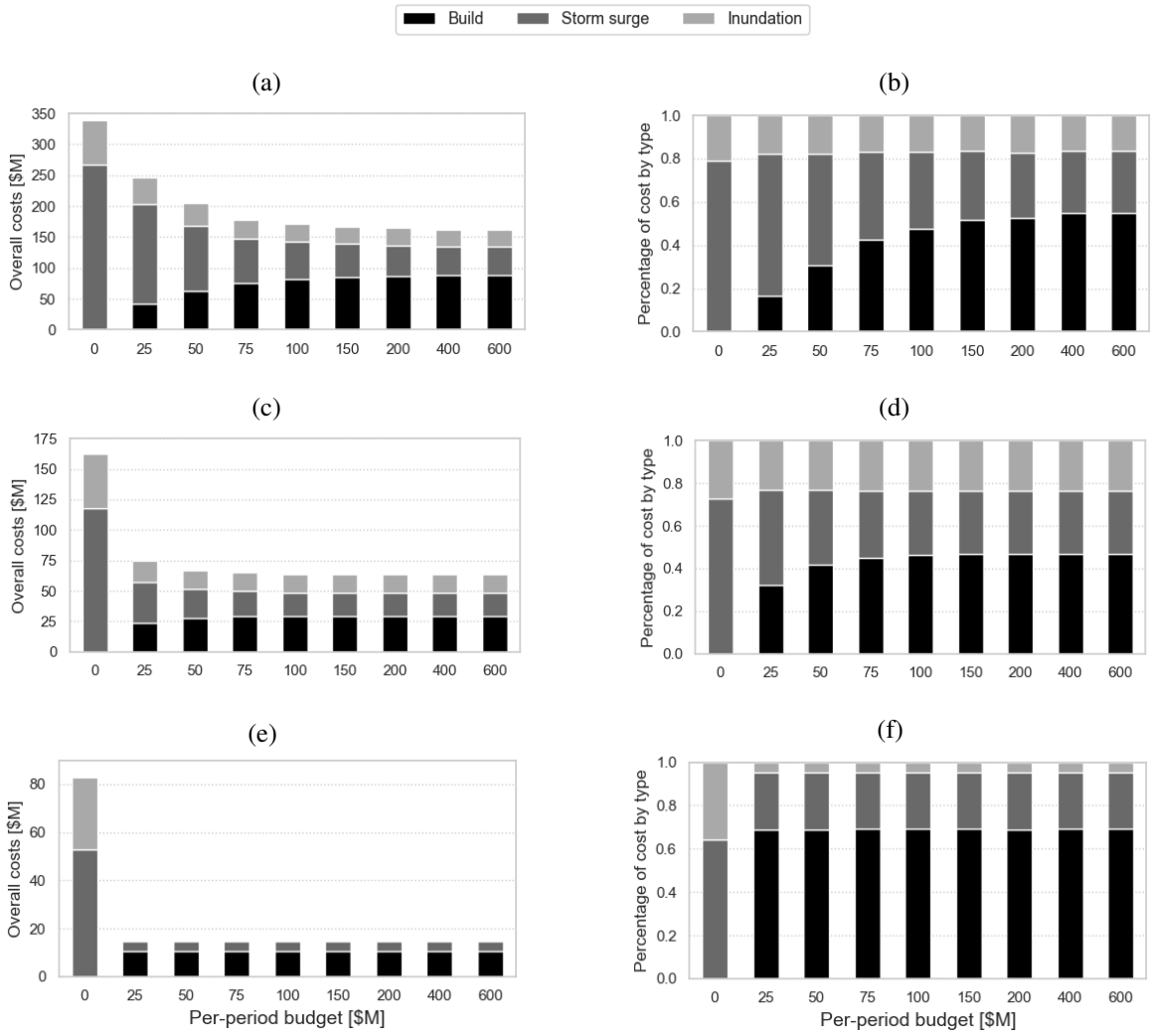
Assuming that the distribution of all possible simulated sea level states paths is given by Table 31, we solve the FRM Formulation (3.3)-(3.25) for a possible combination of chosen values for the model's parameters 144 times, each time on a distinct five-period simulated sea level states path, and compute the expected optimal objective for the chosen parameters' combination in our simulation experiment as $\sum_{\boldsymbol{S} \in \Omega} p(\boldsymbol{S}) z^*_{\boldsymbol{S}}$, where $z^*_{\boldsymbol{S}}$ is the optimal objective of the FRM Formulation (3.3)-(3.25) for the chosen parameters' combination using path $\boldsymbol{S} \in \Omega$. We discuss key takeaways from our simulation-based approach below, and refer the reader to Appendix C for the full set of simulation results.

In Figure 21, we present three simulation results with expected optimal costs reported for two extreme parameter settings and the mid-point of these settings as shown in the inset table in this figure. In Figure 21, charts (a) and (b), the parameters selection results in the worst-case combination (i.e., highest overall costs). The "do nothing" (zero budget) expected costs nearly reach \$340M. The optimal expected investment is \$88.7M with a per-period budget of at least \$400M. Notice that \$88.7M is an expected value and there might be a sea level states path under

which the build cost could be substantially more in a given period, but due to the low probability for such sea level states path, the expected value is much lower. This is the reason that with budgets less than $400M, the overall expected costs are higher. It is also important to note that even with budget values of more than $400M, there are combined expected storm surge and inundation costs tallying more than $70M. This is due to the fact that given the highest build costs (i.e., c) and the largest increments in levee heights (i.e., m) in this worst-case parameters setting, more grids are sacrificed to flooding and inundation over the planning horizon. This simulation's discounted flooding costs are higher due to the low discount rate causing future flooding to be more expensive at today's rates. Figure 21(b) provides a percentage by type breakdown of costs. Storm-surge flooding under lower budgets makes up the bulk of costs. Not until the model reaches a $400M per-period budget does the expected investment cost stabilize at an optimal point where an $88.7M expected investment reduces total expected costs from "do nothing" by 52.3% while also reducing total expected flood-related damages by 78.5%. On the other extreme parameters setting, Figure 21 charts (e) and (f) show that with lower minimum levee heights (i.e., m) and building costs (i.e., c), expected costs are the same under all budget values. Only in the "do nothing" case does the model reach total expected costs of $82.8M, with inundated grids making up nearly 40% of those costs. With just a $10.6M expected investment over the planning horizon, the total expected costs are reduced by 81.0%, while total expected flood-related damages are reduced by 94.3%. Of note, even at these investment costs, the model still sacrifices some grids to storm-surge flooding and inundation, meaning they are not protected even with the extra funding available to build. The behaviour of the expected optimal costs is in between the two extreme cases for the mid-point parameters setting as shown in Figure 21 charts (c) and (d). As mentioned before, Appendix C shows the full breakdown of expected costs and percentages observed for all possible parameter combinations. The best overall expected cost reduction compared to a "do nothing" policy is 92.2% across all parameter combinations and budget values (attained when d = 3%, m = 1m,

$c =$ \$5M/km per m, $f_i = 1.25\bar{f}_i$, and $b_t \geq \$25M$), while the average cost reduction is 63.2%. Across the board, investment shows a meaningful reduction in flood damages, but only until further investment is no longer cost-beneficial.

Figure 21: Overall expected optimal costs and their percentage breakdown by per-period budget for worst-case (charts (a) & (b)), mid-case (charts (c) & (d)), and best-case (charts (e) & (f)) parameter settings across full 144 simulated sea level states paths.



In Figure 22, we show overall expected costs across per-period budgets with combinations of minimum levee heights (m) and levee costs (c) while holding the discount rate (d =3%)

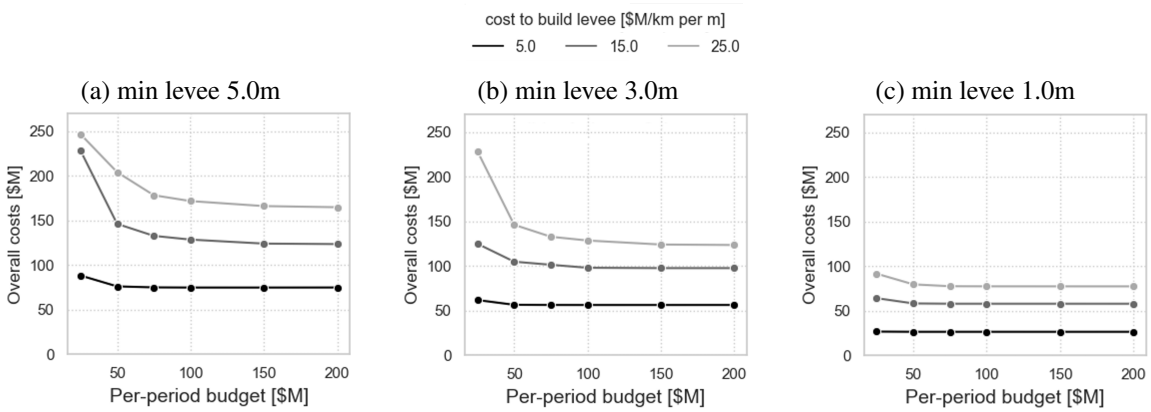
Table 22: Parameter values used in each chart of Figure 21.

Parameter values used in each chart	**Chart**		
Parameter	(a) & (b)	(c) & (d)	(e) & (f)
Discount rate (d) [%]	3	5	7
Minimum elevation increase (m) [meters]	5	3	1
Grid elevation cost (c) [\$M/km per m]	25	15	5
Storm flood damage curve (f_i) [\$M/m]	$1.25\bar{f}_i$	$\bar{f}_i$	$0.75\bar{f}_i$

and the depth damage function ($f_i = 1.25\bar{f}_i$) constant. In Figure 22(c), with $m =$ 1m and $c =$ \$5M/km per m, the lowest curve is constant regardless of budget. As the value of c increases, the other curves begin increasing when the budget falls below \$75M. In Figure 22(b), with $m =$ 3m, the $c =$ \$5M/km per m line remains constant until the budget falls to \$25M. With higher values of c, the overall expected costs increase as the per-period budgets fall. At the lowest budget of \$25M, there is a significant uptick in overall costs due to the inability of the model to build up enough grids to protect the network overall. Finally, in Figure 22(a), the expected costs are rising again, especially at the lower budgets. At $m =$5m, both the \$15M/km per m and \$25M/km per m values of c show increased overall costs when falling below the \$200M per-period budget. At these higher values of m and c, the model hinders building levees on enough grids to provide adequate protection within the network overall. It uses constrained funding to protect the most valuable grids and initially requires higher funding to protect the network more broadly.

Figure 23 shows expected overall, build, storm surge, and inundation cost distributions for the three values of each parameter d, c, m, and f_i when holding the budget constant at \$50M. The top row shows overall expected cost distributions for the simulations when varying each parameter. As anticipated, increasing discount rate reduces the expected overall costs, while increasing the other parameters causes higher overall expected costs. The discount rate has the largest overall effect across model runs. In contrast, changes to the depth damage function (f_i)

Figure 22: Effects of varying minimum levee heights (m) and levee costs (c) on expected optimal costs for varying per-period budgets ranging from \$25M to \$200M (discount rate (d) at 3% and storm depth damage function slope (f_i) at $1.25\bar{f}_i$).



have the smallest effect, which makes sense given that it primarily affects storm surge flooding costs. Looking at individual cost charts is where we see some interesting effects. For instance, in Figure 22, we saw that changes to m and c result in significantly more flood-related damage at lower budgets. Figures 23 (f) and (g) show substantially lower build costs for levee costs of $c =$ \$5M/km per m and minimum levee height of $m =$ 1m than for the higher values of each parameter. Additionally, the median build cost for $m =$ 3m is higher than when $m =$ 5m. The combination of c and m values are critical factors in evaluating building a levee system that can protect as much of the network as possible. At the lower values, adequate funding exists to leverage the network effect and build levees that protect as many grids as possible. However, at higher levee costs with higher minimum levee heights, the model shifts to protecting the most valuable grids in the network because the investment to protect more of the network is too costly.

When evaluating discount rate sensitivity in Figures 23 (e) and (m), there are notable observations in the individual build and inundation costs. At the 3% discount rate, the build costs' boxplot stretches upward, with its lower quartile higher than the upper quartile of the 5% box-

plot. Similarly, at 3%, the inundation costs cover a wider range of values, while at 5%, the inundation cost distribution is much smaller, with a handful of outliers. The smaller discount rates cause future storm surge and flood costs to be higher in discounted terms. So, we interpret these observations to mean that the model is inclined to protect against inundation when those costs can be higher due to lower discount rates. The model invests earlier to protect those grids from future damages. However, with the higher discount rate of 5%, the model invests less because the discounted inundation costs are lower in future periods. Essentially, the lower discount rate tends to cause earlier investments due to flooding and inundation costs in future periods having a more considerable effect on costs in discounted terms.

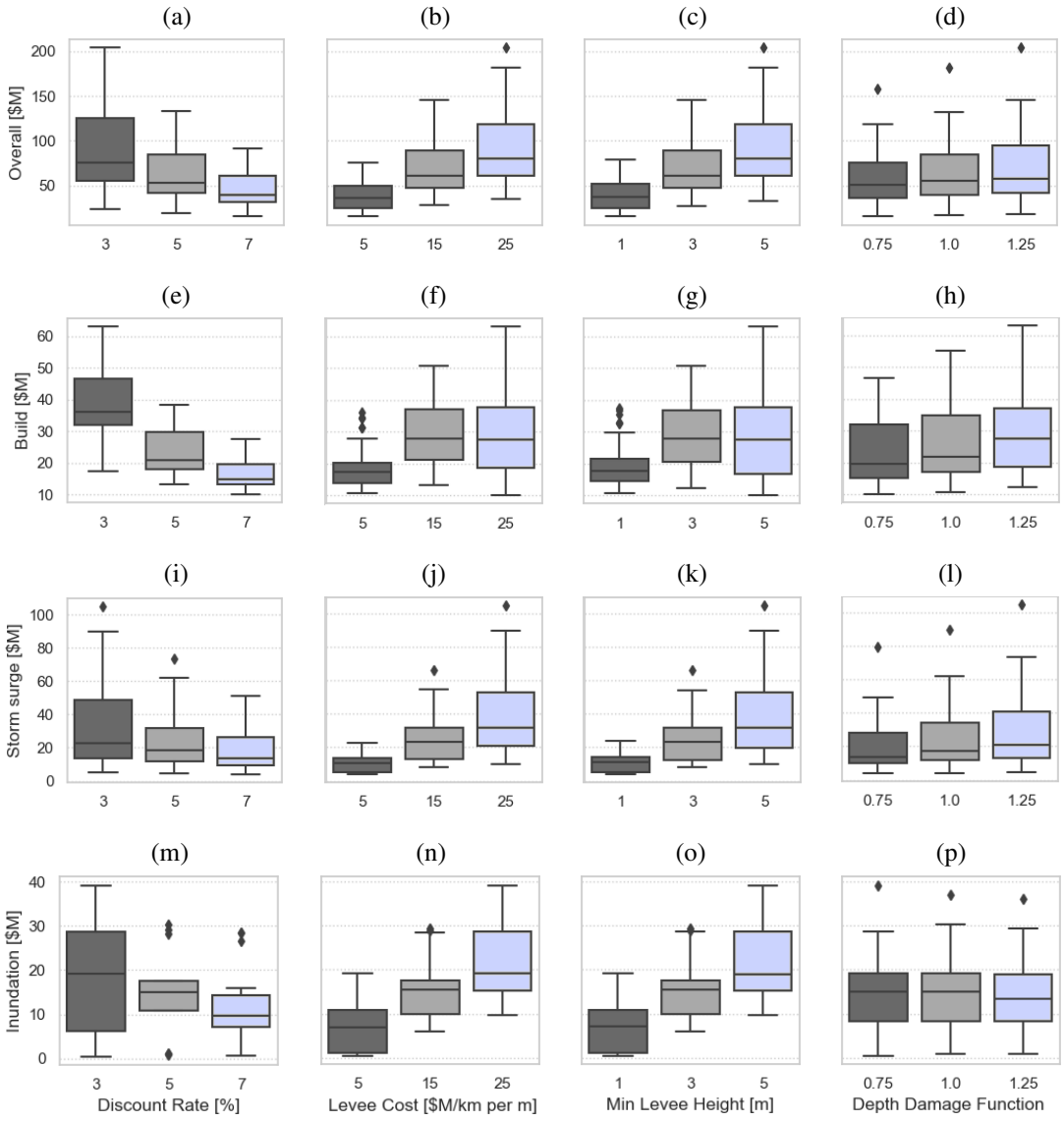
Some key takeaways from the simulation experiment and its sensitivity analysis are:

1. From Figure 21, potential overall expected costs can be significantly reduced by investing only a small fraction of the "do nothing" flood-related costs independent of parameters' values. This quantifiably proves the effectiveness of a mitigation policy in dollar values and shows the extent of the loss for following a response-type strategy.

2. From Figure 21, some grids appear too expensive to protect through the network effect (or individually) and incur storm-surge flooding and inundation costs even when a surplus budget is available. Identifying such grids for planning purposes is not a trivial task and our proposed model can be an effective tool for this purpose.

3. From Figures 21 and 22, our model can be used to find optimal budget per period values that yield the minimum expected overall costs for a given combination of input parameters. We also observed that the levee cost (c) and minimum levee height (m) have the biggest effect on the amount of budget needed to reach the minimum expected overall cost before no further spending occurs. Our proposed model is a powerful tool for determining such meaningful budget values and can be used in financial planning for development of a levee system.

Figure 23: Boxplots showing effects of changing parameters on expected costs for all $50M budget runs.



4. From Figure 23, the discount rate (d) has the largest effect on overall expected costs while the changes to the storm depth damage function (f_i) has the smallest effect. This is an important capability of our model as it can be used to identify parameters that require more accurate estimations because of their significant effects on levee construction planning and timing.

3.5.2 Scenario-based approach

Given the uncertainty associated with the expected sea level rise used in the model, policymakers might be interested in adopting a scenario-based approach by investigating individual scenarios ranging from the best- to the worst-case sea level rise predictions. For example, a policymaker might want to highlight the range of values for investment and flood cost across four different scenarios, namely optimistic, expected-low, expected-high, and high sea level rise scenarios for the next five periods (50 years). This scenario-based approach provides policymakers with meaningful insights to make decisions based on their judgment on anticipated future sea levels. To this aim, in this section, we focus on solving the FRM Formulation (3.3)-(3.25) on four scenarios (i.e., optimistic, expected-low, expected-high, and high sea level rise scenarios) chosen from the 144 simulated sea level states paths mentioned in Section 3.5.1. The chosen scenarios are paths numbered 114 (high), 130 (expected-high), 85 (expected-low) and 64 (optimistic) in Table 31 of Appendix B, respectively. The optimistic and high scenario values represent points near the extremes of the 144 simulated sea level states paths, while the expected-high and expected-low scenarios represent points near the middle. Similar to the case of the simulation-based method, we use the same ranges of values for the model's parameters to conduct a sensitivity analysis for each of the four scenarios considered. Figure 24 shows the results for these four chosen scenarios using the same worst-, mid-, and best-case parameter settings as in Section 3.5.1.

Figure 24: Overall optimal costs by per-period budget for worst-case, mid-case, and best-case parameter settings for no SLR with annual flooding, optimistic, expected-low, expected-high, and high sea level rise scenarios.

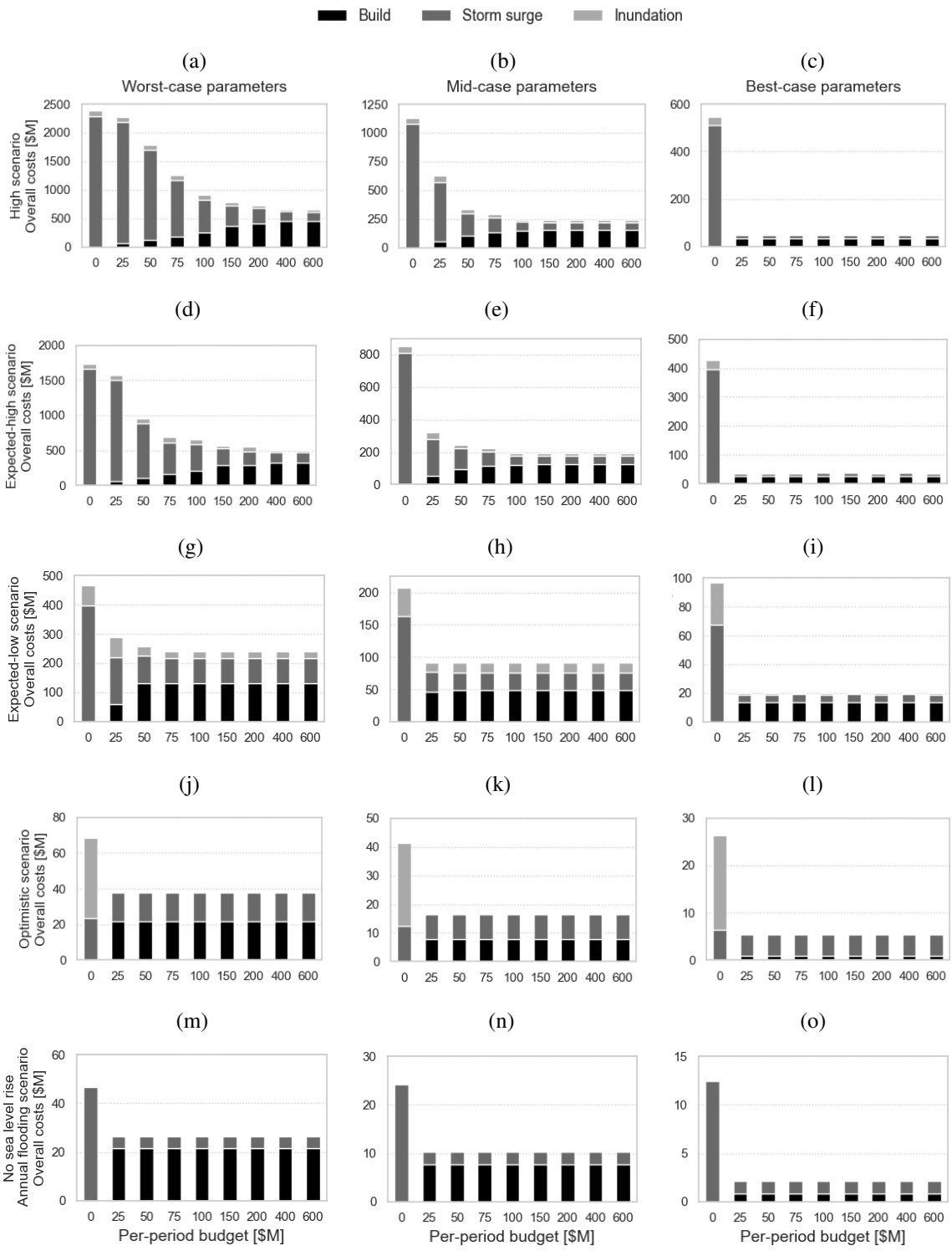


Table 23: Parameter values used in each chart of Figure 24.

Parameter	Worst-case	Mid-case	Best-case
Discount rate (d) [%]	3	5	7
Minimum elevation increase (m) [meters]	5	3	1
Grid elevation cost (c) [$M/km per m]	25	15	5
Storm flood damage curve (f_i) [$M/m]	$1.25\bar{f}_i$	$\bar{f}_i$	$0.75\bar{f}_i$

We see similar patterns in Figure 24 when compared to Figure 21. In the best-case parameters column, all four scenarios show minimal change in total costs across all non-zero budgets. Therefore, mitigating the "do nothing" damages can be met with funding available in the $25M per-period budget. The investment required in the four scenarios under best-case parameters is 3.5% to 14.1% of the "do nothing" damages, resulting in a 79.2% to 91.3% reduction in overall costs. Contrasting that with the worst-case parameters column, we observe that the overall costs increase significantly due to 1) higher investment required due to higher and more expensive levees and 2) more costly damages due to a lower discount rate and steeper depth damage function. In the high scenario and under the worst-case parameters, the overall costs of a "do nothing" policy nearly reaches $2.4B over 50 years. A substantial investment of $454.8M is only 19.1% of the "do nothing" total costs, but reduces overall costs by 72.4%. Appendix C shows the full breakdown of costs and percentages observed for all possible parameter combinations for each of the four sea level rise scenarios. In all four scenarios, investing in flood protection infrastructure at a fraction of the potential flood-related damages results in a meaningful reduction in overall costs.

While evaluating the costs under the four scenarios, we observed consistent sensitivity analysis behavior to that seen previously in Figure 23. The main difference in the scenario-based analysis is the extensive range of investment costs and flood damages. This wide range of sea level states causes substantial variation across the four scenarios, with the "do nothing" overall costs in the high scenario being 27.3 times larger than in the optimistic scenario when aver-

aged across the different parameter settings. Investigating this range for an individual scenario and across different parameter settings also reveals interesting facts. In the optimistic scenario, when given the best-case parameters, the investment required is \$0.9M to handle the addressable risk, while in the worst-case parameters, the investment required is \$21.7M. This presents a reasonably manageable range for a policymaker trying to address the sensitivity of the optimal investment costs to the parameters' estimation accuracy while protecting East Boston from the optimistic flooding. Contrast that with the challenge posed under the high scenario when these numbers go to \$33.0M and \$454.8M, respectively. This presents an extremely risk-averse policymaker with potentially hard trade-offs, and the policymaker must ensure sufficient diligence in estimation of parameters to defend their coastal areas adequately.

In addition to the four sea level rise scenarios discussed above, we also include a no sea level rise (only hurricane storm) flooding scenario in Figures 24 (m)-(o). Comparing this scenario with the other four potential sea level rise scenarios further emphasizes the magnitude of additional flooding costs caused by sea level rise, and calls for more attention to this potential threat. We see investments made in Figure 24 where the costs are balanced in the case of only-storm flooding. However, there are several combinations of parameters for which the model forgoes any investment regardless of the budget amount (see Figure 42 in Appendix C). This happens when investment costs are high due to higher minimum levee heights and construction costs and future flood costs are low due to higher discount rates. For a policymaker believing that sea levels do not rise, given the proper cost structure and levee scope, there is still considerable financial benefit to building such protection infrastructure. Figure 42 in Appendix C shows the full breakdown of costs observed for all possible parameter combinations for no sea level rise (only-storm) flooding scenario.

Table 24 shows the per-period spend for the scenario and parameter combination shown in Figure 24(a). In this table, we include per-period budget data only up to \$400M, because both the \$600M and unlimited budget runs had the same optimal solutions as the \$400M case.

The investment costs per decade shown in the $400M row are essentially the actual amounts required in each period to reach the optimal solution, because even with the higher per-period budgets ($600M and unlimited cases), the model will only spend up to these levels, and then spends no more. As the per-period budget decreases, however, the budget constraints start enforcing limits on per-period spend. We see that with reduced spending in 2030, the model shifts development costs into future periods to mitigate as much damage as possible. This effect results in total investment decreasing as the per-period budget decreases, while the total flood-related damages proliferate with each reduction in per-period budget. Of note, if possible, investment costs are pushed to future periods at the discounted rate. For instance, in both the $150M and $200M budgets, one can see reduced spending in 2060 compared to 2070. One key takeaway from Table 24 when looking at the $400M row is that an initial influx of cash in the first period can reduce future cash needs while significantly reducing overall total costs experienced throughout the planning horizon.

Table 24: Investment costs per period for the high sea level rise scenario and worst-case parameter combination given for each per-period budget value.

Per-period	Discounted Investment Cost [$M]					Total	Total Flood
Budget [$M]	2030	2040	2050	2060	2070	Investment [$M]	Related Costs [$M]
0.0	0.0	0.0	0.0	0.0	0.0	0.0	2,379.0
25.0	21.7	16.1	12.0	8.9	6.6	65.3	2,205.0
50.0	43.3	32.2	24.0	17.8	13.3	130.6	1,657.1
75.0	65.0	48.3	36.0	26.8	13.3	189.3	1,064.1
100.0	86.6	64.4	47.9	35.7	26.5	261.2	649.6
150.0	129.9	96.7	71.9	26.8	39.8	365.1	424.8
200.0	173.2	145.0	24.0	26.8	39.8	408.7	319.2
400.0	281.5	80.5	24.0	35.7	33.2	454.8	202.9

In summary, the following are the key takeaways from our scenario-based experiment.

1. From Figure 24, similar to the simulation results, all four scenarios show opportunities to significantly reduce potential overall costs with levels of investment that are a fraction of "do nothing" flood-related costs. This again demonstrates how rewarding a mitigation approach could be compared to a wait-and-see response-type policy.

2. From Figure 24, the wide range of potential investment and flood costs shows the importance of adequately assessing the potential risks and estimating the relevant parameters for making an investment decision. The more risk-averse the decision maker is, the more accurate their estimation of the model's parameters need to be.

3. From Figure 24, sea level rise threat is real, and can potentially increase the storm-only flood damages by several orders of magnitude. Even if policymakers do not believe sea levels are rising, there is still value to invest in protecting against annual storm flooding if the anticipated cost structures and discount rate support building a levee.

4. From Table 24, policymakers get a view into the actual funding required per decade to mitigate flood-related damages, allocating only as much money as needed to address risks over time. This again proves the value of our model when used for budgeting and financial planning purposes.

3.5.3 Random networks experiment

In order to evaluate the generalizability of the takeaways from our East Boston case study to other coastal areas with different at-risk network structures, we conducted the same simulation-based and scenario-based experiments on 50 randomly generated networks. We created the random networks consisting of 402 hexagonal nodes with random selections of approximately 100 sea nodes, 200 surrounding region nodes, and 100 contiguous region of interest nodes that have at least one border grid with the sea. Values for h_i, f_i, and g_i for each grid were selected

by randomly sampling the attribute data from East Boston. These random network experiments resulted in more than 3M optimization runs across the full range of parameters and sea level states paths. Full details of random network creation and experiment outputs are available in Appendix D.

After comparing the results from the 50 random network experiments with the East Boston case study, we conclude that the key takeaways highlighted in Sections 3.5.1 and 3.5.2 are generalizable to any other coastal area as discussed below.

- We still observe meaningful cost reductions with investments that are a small portion of the "do nothing" flood-related damages.
- Some grids still incur storm-surge flooding and inundation costs even when a surplus budget is available. However, the East Boston case study incurs more inundation damage than most random networks. The features of East Boston contributing to this effect are its extensive border with the sea where sea level incursion can occur, and multiple areas within East Boston connected through groupings of low-lying grids.
- Levee build cost (c) and minimum levee heights (m) still meaningfully affect how much budget is needed to reach minimum overall expected costs before no further spending occurs.
- Overall discounted costs behave consistently with the East Boston case, and we see similar effects when delving into the specific costs as discussed in Figure 23. We still observe that the discount rate (d) has the largest effect on overall expected costs, and changes to the storm depth damage function (f_i) has the smallest effect.
- When evaluating the costs from optimistic to high scenarios for sea level states paths, there is still a substantial range of differences in investment required and total overall costs. The same patterns emerge across parameter sensitivity and sea level states paths.

Similar to the Boston case, the overall costs are more sensitive to errors in parameter estimations as the decision-maker becomes more risk-averse.

3.6 Discussion and conclusion

In this study, we employ networks to model the movement of temporary (storm-related) and permanent sea level rise floods on land and propose a multi-stage stochastic program with recourse for cost-benefit analysis of creating dikes and levees in a coastal city to mitigate climate-change-induced flood damages. In reviewing the experiments in Section 3.5, our model enables an improved understanding of the costs associated with protecting an urban coastal neighborhood from rising sea levels. We find that a "do nothing" strategy of zero flood protection infrastructure investment incurs significant flooding costs. When evaluating the full range of scenarios, modest investment in creation of dikes and levees enhances protection significantly, thereby causing a precipitous drop in overall long-term costs. With a relatively limited investment in mitigation, the worst-case effects will be reduced, and the response needed during a flooding catastrophe is likely not as acute or urgent as would be the case if there were no investment in mitigation.

By including a constrained budget, our model provides planners with a powerful tool for budgeting and financial planning. It also enables a what-if analysis to evaluate potential flood-related damages if they constrain available funding for investment. Even if policy-makers are confronted with uncertainty and competing stakeholders, our model provides the ability to show that mitigation still yields real value in preventing flooding risks and mitigates the potential severity of future flooding disasters for that coastal area. Moreover, our model is very effective in identifying critical parameters whose estimation requires high accuracy in order to avoid large costs due to miscalculations.

The multi-stage character of our model and its recourse feature allow prospective community leaders to revise their adaptation measures as the sea level rise situation unfolds, as recommended in Kirshen, Borrelli, et al. 2018. The model also enables quickly incorporating the latest thinking in sea level rise probabilities, thereby interpreting and applying potential probabilities for a broader range of sea level rise as found in studies by Kopp et al. 2017 and Sweet et al. 2017. The agile nature of our proposed method also enables quick solution adaptation when facing infrastructure changes in built-up areas, new risks in specific locations, and changes in flood protection design and costs.

In pursuing this model, we aimed to be able to build and run the model using readily available open-source data. We believe planners can apply this methodology in any coastal area with the same data availability. For example, low-lying cities like Miami and New Orleans have publicly available elevation data (through the NOAA Data Access Viewer (NOAA 2013)) and appraised tax data (through local municipal Open Data Hubs (USGSA 2009)). In practice, these cities could similarly use this methodology to evaluate their city's changing situation.

As potential directions for future studies, researchers may focus on potential impacts of hurricane storm and sea level rise flooding on other kinds of infrastructure such as roads and transportation networks. Moreover, future research may try to mitigate the non-financial impacts of flooding associated with disrupted communities, lost lives, and displacement of people particularly those from socially or economically marginalized communities. Although we do not currently capture these non-monetary parameters, we believe there is an opportunity to incorporate these considerations in the future development of our proposed model.

CHAPTER 4

A DECISION ANALYTIC TOOL FOR CORPORATE STRATEGIC SUSTAINABLE ENERGY PURCHASES

4.1 Introduction

Within the industry sector, energy consumption accounts for nearly 40% of global energy usage (International Energy Agency 2021). Decision frameworks rooted in finance-related considerations are crucial to transition this sector away from greenhouse-gas-emitting energy sources. As discussed in Chapter 2, decision frameworks were prevalent across various studies, yet finance-related issues never emerged as the predominant focus. Nevertheless, energy procurement for global companies continues to grow more complex due to the rise of corporate sustainability initiatives, increased greenhouse gas compliance reporting, declining costs of renewable energy, and ongoing regulatory changes in electricity markets. Moreover, companies strive to manage costs amidst intensifying international competition (Google 2013). This complexity poses significant challenges for company executives, sustainability leaders, procurement professionals, and energy managers. Confronted with numerous objective decision-making scenarios, it demands substantial effort to arrive at optimal sustainable energy purchasing decisions (Esty and Winston 2008).

Many companies find they do not have adequate decision-making frameworks for proper tactical decision-making at the facility or regional level, let alone ensuring that company strategic initiatives and objectives are met more broadly. This apparent need in industry can seemingly be met with the ample literature that exists in research publications that spell out chal-

lenges of integrating renewable energy and sustainability initiatives on a regional or country-level (Wang et al. 2009; Brede and Vries 2013; Strantzali and Aravossis 2016; Kumar et al. 2017; Martínez-García et al. 2018). There are also many vendors with point solutions for optimizing energy usage in combination with energy storage that can be integrated to varying degrees with on-site generation and electricity markets more broadly. Among these competing products, there are real-time energy usage optimization products, including, but not limited to Athena from STEM, DER Optimization from Enel X, AEROS control suite from NEC Energy Solutions, IQ from Fluence, and GEMS OS from Wartsila (Baxter 2021). However, there is little research covering these challenges from the perspective of a multinational company looking to gain a competitive advantage while optimizing energy purchasing strategies through a combination of achieving energy cost savings, managing volatile energy market risks, and meeting greenhouse gas compliance reporting requirements. Companies must balance these needs while aiming to optimally run their organization and improve their brand recognition.

In this chapter, we create general decision-support modules in the Lumina Decision Systems Analytica 5.0 decision-support modeling software (Lumina Decision Systems 2020) for both multi-attribute utility theory decision-making and strategy development methodologies. These general modules being made available within Analytica is a meaningful contribution for decision sciences researchers to apply within not only energy purchasing decisions, but in other complex decision scenarios with multiple attributes and substantial complexity to possible strategies. We then apply these modules to further build a decision model that helps corporate decision-makers pick their optimal sustainable energy purchasing strategy for a given location. The model inputs are based on anticipated energy usage combined with the decision-makers' applied utility functions for energy cost, sustainability requirements, and brand prestige. We then apply the model in a use case from a multinational corporation to demonstrate the model's capabilities. Throughout the model design and case study demonstration, we incorporate best practices from similar efforts in decision analysis and energy-related journals to

introduce a decision-support framework, enabling a more systematic approach to sustainable energy procurement management for global companies.

4.2 Literature review

A search for published research on multi-criteria decision-making (MCDM) in sustainable energy yields many hundreds of articles; therefore, to begin our search, we first identified several holistic reviews of articles published on MCDM in the energy decision space. The first holistic review paper evaluated conducts a thorough review of more than 100 studies involving multi-criteria decision analysis aids in sustainable energy decision-making (Wang et al. 2009). This paper performs a breakdown of criteria for what constitutes sustainable energy, methods for criteria selection, weighting methods, multiple MCDM analysis methods, and aggregation methods. A subsequent holistic review paper also conducts a similarly thorough review of MCDM application in sustainable and renewable energy development (Kumar et al. 2017). The authors in this second instance evaluate sustainable and renewable energy for electrification of rural areas along five dimensions of sustainability indicators: Economic, Technical, Environment, Social/Ethical, and Organizational/Institutional. The authors also conduct a complete breakdown of the differing MCDM methods applied from eight possible MCDM methodologies, including the one we will use in our framework – multi-attribute utility theory. The third comprehensive review paper in this space ventures further into sustainability, evaluating literature applicable to assessing and selecting optimal technology alternatives from a sustainability perspective (Ibáñez-Forés, Bovea, and Pérez-Belis 2014). That assessment explores the literature and buckets selection criteria along five dimensions: Economic, Technical, Environmental, Social, and Political. When taken together, these three review articles evaluated more than 300 papers, conducting thorough reviews of the application of decision-making methods for choosing sustainable material or energy from multiple criteria.

Evaluating these comprehensive sustainability-related MCDM reviews reveals a gap in the research, highlighted by the growing need for companies to capture value from a demonstrated commitment to sustainability. There is an increasing importance of consumer awareness on sustainability as consumers actively decide among competitors based on their focus on sustainability (Galbreth and Ghosh 2013). In the United States, some forward-leaning multinational companies pay for commercials in prime sports television spots to paint their brands as sustainable (The Climate Group 2019). In a recent study conducted from a company perspective, the authors evaluate how branding products as "made with renewable energy" will positively affect Australian consumers' impression of a company (Mydock III et al. 2018). In a drive to increase consumer awareness, consumers are being influenced through environmentally friendly initiatives to learn more about how their purchasing behavior can contribute to more sustainable activities by the companies where they purchase goods and services (Frostell et al. 2015).

In light of this trend, few articles focus on energy purchasing strategies that balance the competing priorities of short- and long-term energy costs with sustainability reporting requirements while also enhancing brand prestige as a sustainability-focused company. Some papers approach the challenge from a strategic perspective of defining corporate energy policy and strategy to achieve carbon emissions reduction targets (Finnerty et al. 2018). In contrast, other papers get to a more tactical decision-support framework to prioritize energy efficiency projects in an industrial organization (Contreras et al. 2017), but none tie all of the strategic needs together. Still, companies worldwide are increasingly sensitive to climate change challenges and the need to reduce greenhouse gas emissions globally (Esty and Winston 2008) and are thus making sustainable energy choices every day.

The situation highlighted above created the context for conducting the effort in this analysis. In conducting this analysis, we had access to a US-based Energy Services company that helps guide customers through the complicated process of developing and executing sustainable energy purchasing strategies. Through this company, we had access to the facilitator of

a strategic energy purchasing workshop held for a global company operating in Mexico. The facilitator provided an interview about the workshop in general and also provided workshop information, including data and participant input about concerns and drivers of their decisions. We took the general framework discussed, used it to evaluate the company's needs, and created a more structured and formal method to concisely capture, analyze, and present the possible strategic options and outcomes for the decision-maker. We populated the model with (here, disguised) energy usage data from the global company. This analysis provided a better understanding of the customer's priorities and how the Energy Advisory team from the US company could help guide the customer to the "best" decision based on customer-provided criteria. In the following sections, we will walk through the general framework and model formulation before applying it to this particular case. We also discuss the results from applying the model specifically to the case study, but the framework here is generalizable to many situations where companies make similar sustainable energy purchasing decisions. To the best of our knowledge, this is the first study to incorporate brand prestige into the strategic sustainable energy purchasing decision process for a large multinational company.

4.3 Model development and methods

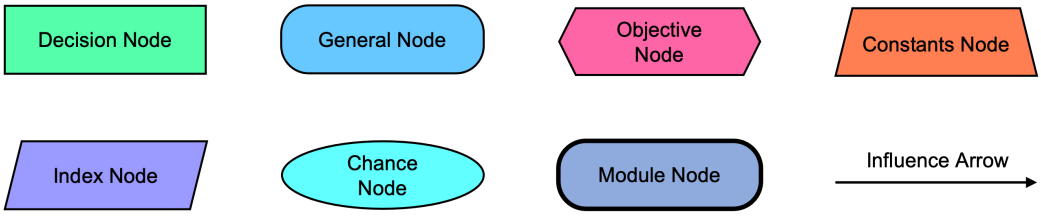
After evaluating the case study mentioned above, we synthesize the key elements discussed in the workshop and break them down to create a generalizable framework for decision support. We apply several decision-support methodologies to balancing short/long term energy costs with the need to meet green reporting requirements and the strategic value of leveraging the purchase of sustainable energy to build a better brand image. To facilitate the modeling, we use the Lumina Decision Systems Analytica 5.0 decision-support modeling software to create, explore, and share the quantitative decision model presented in the following sections (Lumina Decision Systems 2020). Analytica provides a rich quantitative modeling environment using a

graphical interface based on influence diagrams. It was initially developed for decision analysis use and, over time, added many general quantitative modeling and analysis capabilities (Morgan, Henrion, and Small 1990). It has been used in a wide range of complex challenges in related areas, including energy (Stadler et al. 2009; Tylock et al. 2012), climate change (Dowlatabadi 1998; Senbel, McDaniels, and Dowlatabadi 2003), emissions policy (McKinley et al. 2005), and transportation (Maizlish et al. 2013). Analytica has an active, though selective, user community, and it supports libraries of reusable components from both public and organization-specific models.

Before discussing model development, it is important to understand the objects used within the Analytica model. In the discussions that follow, our model will use the Analytica nodes as described below and shown in Figure 25. More detailed background on Analytica objects is available on the Lumina website (Lumina Decision Systems 2020).

- Decision Node – Variable that the decision-maker can control directly
- General Node – Variable typically representing a deterministic input or a functional dependency
- Objective Node – Variable evaluating overall value or desirability of outcomes
- Constants Node – Variable representing constant values
- Index Node – Class of variable defined as a list or sequence
- Chance Node – Variable with uncertain values, usually contains probability distribution
- Module Node – A collection of related objects and nodes
- Influence Arrow – An arrow from one node to another showing influence between nodes

Figure 25: Node types used in Analytica model.



Next, we will discuss the decision analysis methods included in this analysis. We first introduce and discuss the creation of general modules to apply both Strategy Table and Multi-attribute Utility Modules and how to implement them in Analytica (Section 4.3.1 and 4.3.2). Then we discuss integrating these general modules into a particular implementation of an overall model in a case study for purchasing sustainable energy. We then follow up with a description of the entire model implementation (Section 4.3.3), the data used in the model for the case study (Section 4.3.4), and the results from running the model (Section 4.4). We finish the paper with a discussion of our findings, which is followed by conclusions (Section 4.5).

4.3.1 Strategy tables

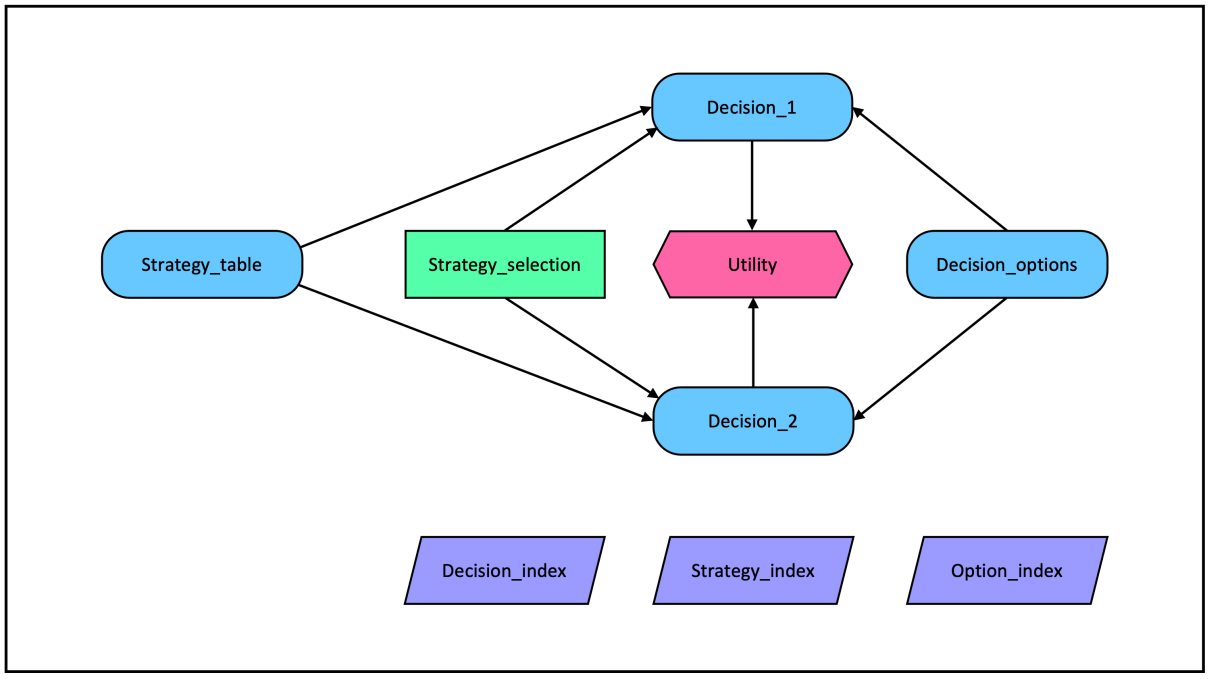
4.3.1.1 Strategy table module development

The first decision analysis structure we will use in creating our model is a strategy table (McNamee and Celona 2008). A strategy table is used for generating strategic alternatives when there are many components of a decision to consider. A strategy table is typically constructed by identifying a set of decision variables and creating a table with one column for each decision variable. Next, a set of options is identified for each decision variable entered in rows of the corresponding columns. Finally, a new column is added to the left of the decision variables, and this column contains a set of strategic alternatives. Each strategic alternative is associated with a left-to-right path through the table, selecting one option for each decision variable. Each path should be feasible, internally coherent, and consistent with the particular

theme of its associated strategy (Jeffrey M. Keisler 2003). These strategic alternatives then form the decision-maker's overall choices to consider by evaluating the strategies in the usual way. The strategy with the highest expected utility can then be identified.

We note that in many ways, the Analytica modeling environment has powerful capabilities for decision-analytic computations, but it does not explicitly incorporate strategy tables. Thus, we developed a module that should integrate smoothly with Analytica across many applications (not just for the energy purchasing model). The structure of the module in Analytica is shown in Figure 26 and is outlined further below:

Figure 26: Strategy table module design in Analytica (Jeffrey M Keisler 2020b)



1. The decision variables, represented by the columns of the strategy table, are indexed with the *Decision_index*.

2. The options, represented by the rows of the strategy table, are indexed with the *Option_index*.

3. The table *Decision_options* is indexed by *Option_index* and *Decision_index*. The cell in row i and column j contains the i^{th} option associated with the j^{th} decision variable. Thus the first column contains the options for the first decision variable, and the second column contains the options for the second decision variable, which then continues for all j columns. Cells may contain numeric values or other types of values. If the cells contain non-numeric values, decision variable values would also be non-numeric, and in order to be evaluated, any subsequent nodes (such as *Utility*) would require expressions that operate on these types of values in their predecessors.

4. The strategic alternatives are indexed with the *Strategy_index*.

5. The table *Strategy_table* is indexed by *Decision_index* and *Strategy_index*. The cell row k and column j of the strategy table contains the index value for the option that is to be selected for a j^{th} decision under the k^{th} strategy.

6. *Strategy_selection* is a decision variable, indicating the operative strategy (row) from the strategy table.

7. The variables *Decision_1* and *Decision_2* are for the values of the first and the second decision variables as determined by the strategy selection. The *Decision_1* expression is *Decision_options[Decision_index= 1 , Option_index=Strategy_table[Decision_index = 1]]*, which pulls out the appropriate option for decision 1 from the option table under each strategy, and a similar expression is used for *Decision_2*. One node will be required for each decision variable. The purpose of these nodes is to utilize the values of the specific decision variables associated with the strategy in later computations of, for example, profit or utility. Depending on the application, it may be more elegant to put the decision variables in a separate module.

8. The variable *Utility* is a stand-in for whatever later variables or modules will make use of the decision variables.

4.3.1.2 Strategy table model implementation

The Energy Purchasing Model specifies a strategy table that we incorporate into Analytica via the application of the above module in Section 4.3.1.1. Many decisions come into play when companies weigh the type of energy to purchase for one of their given locations. We need to apply a strategy table to enable reducing the decision set to suitable, feasible combinations of the decision points. Based on the previously mentioned case study in Section 4.2, we develop five critical decisions that the decision-maker will need to make to plan for their energy and sustainability needs. These decisions are shown in Table 25 and represent an essential subset of this customer's critical decision points. This table is representative but is not a complete set of all possible decisions that any specific firm may need to make. In Table 25, there are $2^3 \cdot 4^2 = 128$ possible combinations of decisions in just this relatively small table of choices. Not all of those options are feasible (e.g., "No Renewable Energy" should not be paired with "Own Green"), but there is still a non-trivial amount of complexity to the choices facing the decision-makers.

In evaluating the case study, we synthesize five possible strategy alternatives that are different enough to ensure the strategic options are distinct. We break down these five strategy alternatives as:

1. Install, use, and maintain their own green generation supply
2. Purchase cheapest retail energy (assumed it is not green)
3. Purchase green energy from a supplier on a short-term basis, without capital expense (CAPEX)

Table 25: Energy purchasing decision options.

Decision Points	**Decision Options**
Build or Buy	Build Energy Supply Buy Energy Supply
Location	On-site Off-site
Renewable Energy Strategy	No Renewable Energy Use Own Energy Buy Green Energy Buy Renewable Energy Credits
Purchase Strategy	Retail Green Own Green Power Purchase Agreement (PPA) Retail Any Energy
Equipment Type	Existing New

4. Purchase green energy from a supplier on a long-term basis, without CAPEX

5. Offset non-green retail energy purchases with Renewable Energy Credits (RECs)

Combining these strategic options with the decision choices in Table 25 and applying the appropriate options for each decision, we end up with the strategy table shown in Table 26. The five alternatives shown in Table 26 will form the foundation for the remainder of the model's implementation.

4.3.2 Multi-attribute utility

4.3.2.1 Multi-attribute utility module development

As noted previously in Section 4.2, there are many different methodologies to approach a multi-criteria decision analysis (MCDA). A recent review of literature focused on decision-

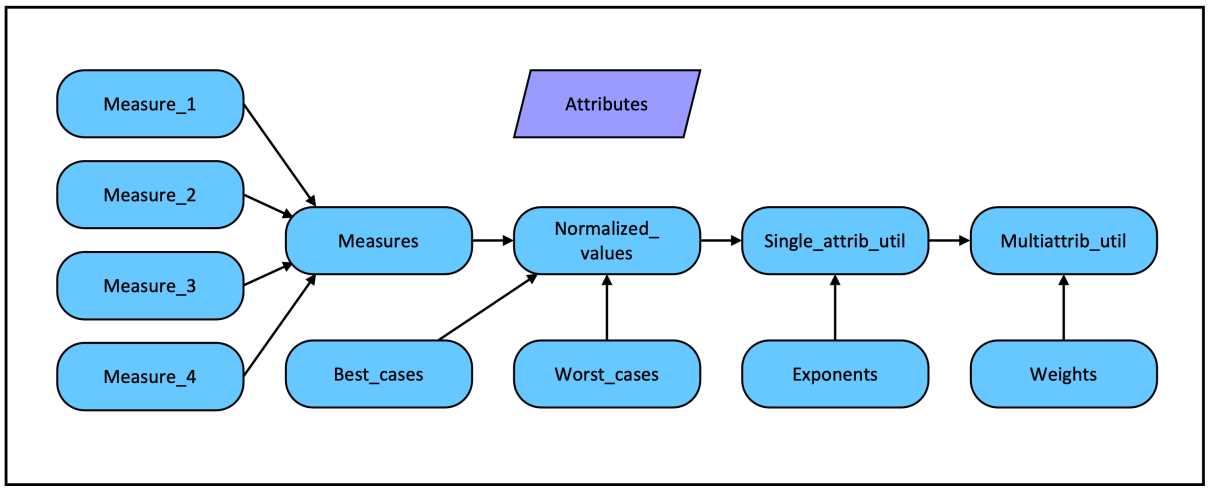
Table 26: Strategy table with strategy alternatives and associated decision options.

Strategy Alternatives	**Build or Buy**	**On or Off-Site**	**Renewable Energy Strategy**	**Purchase Strategy**	**Equipment Type**
Install own Green Gen	Build	On-Site	Use Own Energy	Own Green	New
Cheapest Retail Energy Cost	Buy	Off-Site	No Renewable Energy	Retail Any	Existing
No CAPEX w/ Green short-term	Buy	Off-Site	Buy Green Energy	Retail Green	Existing
No CAPEX w/ Green long-term	Buy	Off-Site	Buy Green Energy	PPA	Existing
Offset with RECs	Buy	Off-Site	Buy RECs	Retail Any	Existing

making for sustainable development shows a significant upward trend in the application of MCDA, with at least 343 papers published from 2010 – 2017 (Kandakoglu, Frini, and Ben Amor 2019). As highlighted in past studies, frequently, the main stakeholders involved in a decision-making process initially lack the expertise to thoroughly assess and apply the available techniques (Guikema and Milke 1999; Ewing and Baker 2009; Lerche et al. 2019). For our modeling, we choose the multi-attribute utility theory (MAUT) approach (Keeney, Raiffa, and Meyer 1993). Applying MAUT involves developing utility functions that define key metrics (attributes) representing the level of achievement of the decision-makers' objectives. Then an overall utility function is developed – often a linear-additive multi-attribute utility model, i.e., a weighted sum of single-attribute utility functions. When the utility function is adequately specified in terms of expressed preferences and tradeoffs over tractable reference cases, the course of action with the highest utility score can also be assumed to be the course of action that a rational decision-maker would ultimately prefer. Its simplicity is vital for explaining the inputs and results to decision-makers who may not have the background, time, or patience to understand the more mathematically complex approaches.

While Analytica software also supports the use of utility functions (in particular, using value nodes that are tracked in calculations and whose expectation models are intended to maximize), it does not come with specific capabilities for encoding multi-attribute utility functions. This module implements such a capability in a manner that should integrate smoothly across Analytica applications (not just for the Energy Purchasing Model). The structure of the module in Analytica is shown in Figure 27 and is outlined further below:

Figure 27: Multi-attribute utility theory implementation in Analytica (Jeffrey M Keisler 2020a).



1. The first set of nodes on the left (*Measure_1* through *Measure_4*, although any names and any number of measures may be used) are the measures for the attributes. Alternatively, the measures may be contained in a separate module and calculated in whatever way that makes sense for a particular application.

2. There is an index node named *Attribute* which lists the attributes.

3. The nodes *Measure_1*, . . . , *Measure_4* serve as inputs to a node named *Measures* which is indexed by *Attributes*.

4. *Worst_cases* contains the lowest allowable scores for each attribute, *Best_cases* contains the best scores for each. *Normalized_values* transforms the measure for each attribute to a 0-1 scale using the expression $(Measures - Worst\_cases)/(Best\_cases - Worst\_cases)$.

5. *Single_attribut_util* calculates single attribute utilities based on the normalized value and the exponent parameter for each attribute contained in the *Exponents* node.

6. *Multi_attribute_util* is not indexed by attributes. It calculates a multi-attribute utility score using the expression *sum(Single_attrib_util · Weights)*. If weights are between 0 and 1 and sum to 1, and if single attribute utilities range from 0 to 1, then the multi-attribute utility will also range from 0 to 1.

4.3.2.2 Multi-attribute utility model implementation

The Energy Purchasing Model specifies a multi-attribute utility function incorporated into Analytica via the above module in Section 4.3.2.1. This model highlights three key attributes that come into play when selecting a strategy: Energy Cost, Green Value, and (Brand) Prestige Value. The Energy Cost utility value comes straight from normalizing the calculated expected energy costs described below. For this model's purposes, we model the Green Value and Prestige Value on a five-point scale ranging from the worst case of "not important" being zero to the best case of "highest importance" being equal to 5. We use the 0 - 5 point scale as it was the method used by the energy advisors in this case study when discussing options with the decision-makers. The values were presented in the case study on a visual scale showing 0 = Not Important/None, 1 = Little Importance, 3 = Important, and 5 = Highly Important. Examples of the type of questions used in the case study to determine the green and prestige utility values include:

1. Are there corporate directives that call for environmental stewardship?

2. Can multiple renewable energy strategies be employed to lower or eliminate an individual plant's carbon footprint or applied to the entire organization?

3. Is there value in claiming renewable purchases for the corporation?

4. Can the company apply renewable attributes across geographies?

4.3.2.3 Energy cost

We use an Energy Cost calculation that is relatively straightforward based on expected energy usage, price rates for various usage parameters, and time covered. We use this simplified method as a starting point but acknowledge that electricity charges are often uniquely dependent upon the customer's market tariff, country, utility, and many other parameters related to the customer's energy usage. This model could get overly complicated quickly if we tried to replicate all possible charges for all possible scenarios (Capehart, Turner, and Kennedy 2006). Table 27 shows at a high level the energy cost factors we use in this model broken down into two columns; those factors used across all strategies versus elements unique to each strategy.

For all five electricity alternative strategies in Table 26, the data is ultimately summed up to evaluate the total cost over twenty years. By looking over twenty years, we can estimate the expected annual rate increases (e.g., Retail Elevator, Power Purchase Agreement (PPA) Elevator, Inflation Rate) for each scenario. Including these rates helps show the value of long-term contracts or the potential risk of not locking in rates in potentially volatile markets. Finally, for some types of electricity alternatives, in the model, there are unique costs we evaluate by applying a probability distribution to reflect each strategy's specific values (e.g., Capital Expense (CAPEX) and Operating Expense (OPEX) for Install Own Green Gen). Equation 4.1 below shows the shared calculation for Energy Cost across all types of electricity usage.

$$\text{EnergyCost}_j = \left(\sum_{i=1}^{I} (\text{Rate}_{ij} \cdot \text{Usage}_{ij}) + \text{Dem}_j \cdot \text{DemCharge}_j + \text{AdderRate}_j \cdot \sum_{i=1}^{I} \text{Usage}_{ij} \right) \cdot (1 + \text{VAT}) \quad \forall i \in I \qquad (4.1)$$

With each of the items in the equation defined as:

$i \in I$ = Types of Charges [Base, Intermediate, Peak periods]

$j \in J$ = Types of electricity [Retail Green, Own Green, Power Purchase Agreement, Retail Any]

Rate_{ij} = Charge for energy usage by usage type i and electricity type j ($ per kWh)

Usage_{ij} = Energy usage by usage type i and electricity type j (kWh)

Dem_j = Electricity demand during peak period in the past per tariff for electricity type j (kW)

DemCharge_j = Rate for charging Demand Charges for electricity type j ($ per kW)

AdderRate_j = Rate to apply for ancillary expenses for electricity type j ($ per kWh)

VAT = Value Added Tax (%)

Table 27: Cost factors applied to strategy to determine 20-year energy cost.

Strategy Alternatives	**Cost Factors Used in Cost Calculations for Individual Strategies**	**Cost Factors Applied to all Strategies**
Install own Green Gen	CAPEX and OPEX for installing and maintaining Green Energy Generation	Prior period peak demand Value-added tax (VAT) Inflation Rate Adder charge rates Energy usage & price rates (base, intermediate, peak)
Cheapest Retail Energy Cost	Retail Elevator and Possible Fines for failing to meet green regulations	
No CAPEX w/ Green short-term	Retail Elevator	
No CAPEX w/ Green long-term	PPA Elevator for contracted annual rate increases	
Offset with RECs	Cost of Renewable Energy Credits	

4.3.2.4 Green value

While most traditional energy-related business decisions are made concerning only the cost, due to the introduction of sustainability regulations, greenhouse gas reporting, and science-based goals, there is now more focus on being green within companies (Esty and Winston 2008). This value can be captured through economic means (e.g., selling renewable energy credits), utility value (e.g., greenhouse gas reporting requirements), or other numeric means (e.g., science-based targets for actual greenhouse gas reductions). For purposes of simplicity for this case study, we treat Green Value as a utility value on a scale of 0 to 5. Although 0 to 1 is typically used in decision analysis, the scale of 0 to 5 was more intuitive to the Energy Advisors and decision-makers in this case study.

4.3.2.5 Prestige value

Every company is always trying to get ahead of its competition (Porter 1998). Energy purchasing has not been at the forefront of improving the corporate brand as energy was simply a commodity that a company's procurement team worked to get as cheaply as possible. With the advent of increasing green energy opportunities, significant news coverage on climate change, and the changing regulatory environment, companies now have significant opportunities to leverage their energy purchasing strategy to improve the prestige of their brand (Winston, Favaloro, and Healy 2017). For this research, we treat Prestige Value as a utility value on a scale of 0 to 5, using similar reasoning as for the Green Value index.

4.3.3 Full model implementation

We use the Lumina Decision Systems software package, Analytica 5.0 (Lumina Decision Systems 2020), for this analysis. Analytica enables us to visually break down and map out the model's complexity, segmenting the entire model into five main modules. As seen in Fig-

ure 28, at the highest level of the model, we have one module each for the Strategy Table, the three attributes (Energy Cost, Green Value, and Prestige Value), and the MAUT Scores. The model's analytical output is also available at this level of the model, including Outputs and Sensitivity Analysis to conduct a deterministic sensitivity analysis for Energy Costs and the Multi-attribute Utility Scores. The Strategy Table and Multi-attribute Utility modules are implemented as described in Sections 4.3.1.1 and 4.3.2.1, respectively. There is one additional MAUT Score Importance node for an alternative sensitivity analysis using the variable importance functionality of Analytica. The Green Value, Prestige Value, and Energy Cost modules house all the inputs, variables, indices, and calculations necessary to provide outputs for each strategy alternative to act as the required inputs into the MAUT Score calculations. All data for monetary and electricity-related variables for the case study are easily ingested from a csv (comma separated values) file.

Figure 28: Analytica top level of decision model.

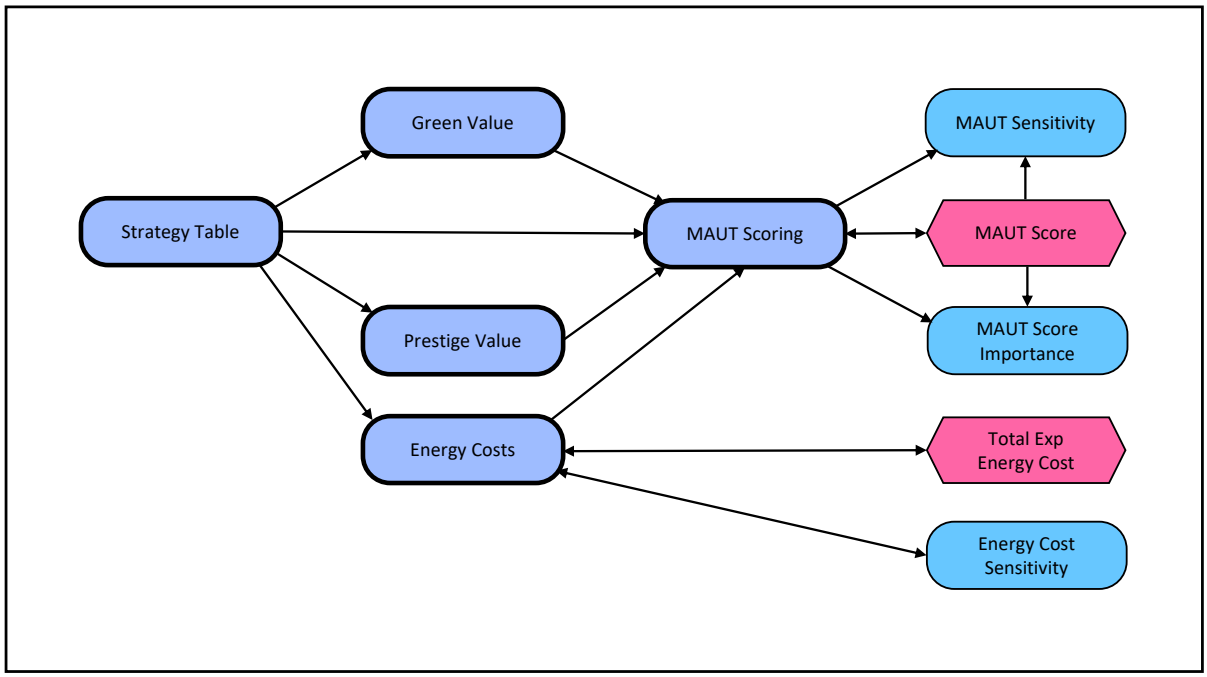


Figure 29 shows the structure for the implementation of our Strategy Table represented in Table 26. The Strategy Index node represents the rows of strategic alternatives in Table 26, and the Decision Index node represents the decisions to be made from the Table 26 columns. We replicate the table in an array in the Strategy Table node, and the strategic choices that flow throughout the model are based on the options in the green Strategy Selection decision node. The five variable nodes representing the five decisions create individual arrays used for further calculation triggered throughout the model.

Figure 29: Analytica model strategy table diagram.

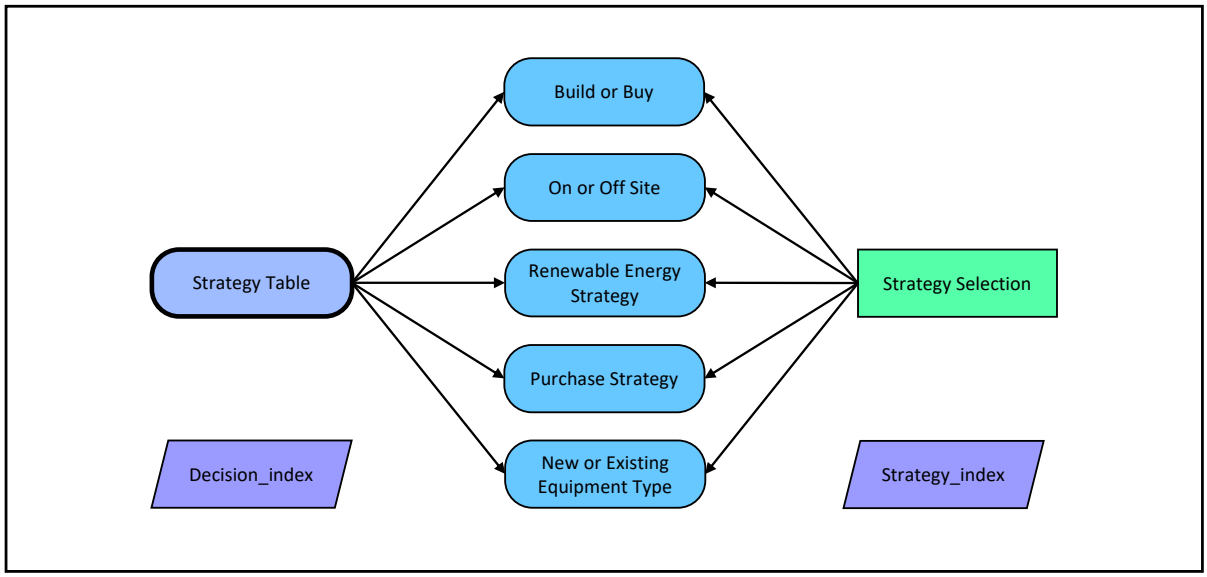


Figure 30 goes deeper into the Energy Cost module's model development, the most complex of the three attributes represented in this model. All strategies for Electricity Rates and Electricity Costs are defined therein. Because different strategic alternatives have additional costs, we also capture those unique costs in specific modules below. If a decision results in building and maintaining Green Generation equipment, those costs are modeled in the Cost to Build/Own module. If a strategy requires Renewable Energy Credits (RECs), those costs are modeled in the Cost for the RECs module. If the energy is from a non-green source, and there

is possible regulatory and penalty risk, those costs are modeled in the Cost for Fines module. All of the different modules' information is brought back together in the Combined Costs by applying applicable expenses to each of the strategic Energy Cost alternatives, which provides the final result in the Energy Cost Measure node. There is also a module that incorporates the discount rate to account for properly discounting future costs. Finally, we have two modules that help conduct a sensitivity analysis on just energy costs and total costs.

Figure 30: Energy cost module diagram.

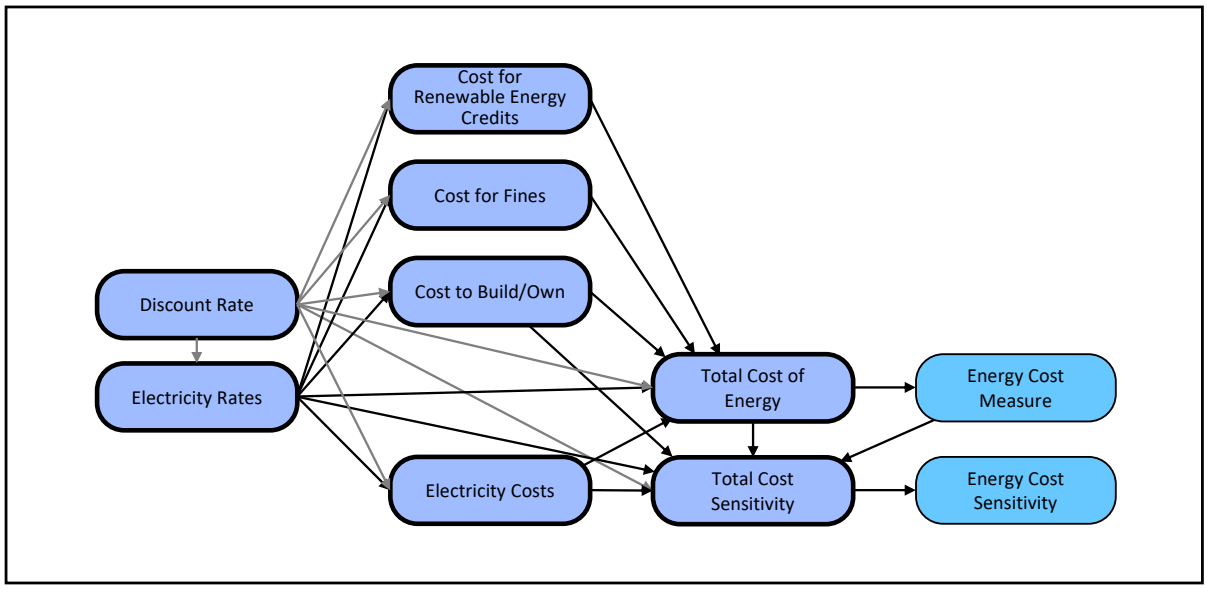


Figure 31 shows the inner flows for the Prestige Value and Green Value modules. For this analysis, we use a simple combination of utility values (Utility Raw variables) for both the additive (Good) and detrimental (Bad) strategies and shape them with probabilistic distributions (Utility Uncertainty variable nodes). These values are combined into the attribute values passed on to the MAUT calculations as Prestige Measure and Green Measure. We believe keeping these modules simple for this analysis is appropriate given the highly complex nature of these attributes for potential companies who would evaluate these utility functions differently based on their individual corporate needs.

Figure 31: Prestige and green value model diagrams.

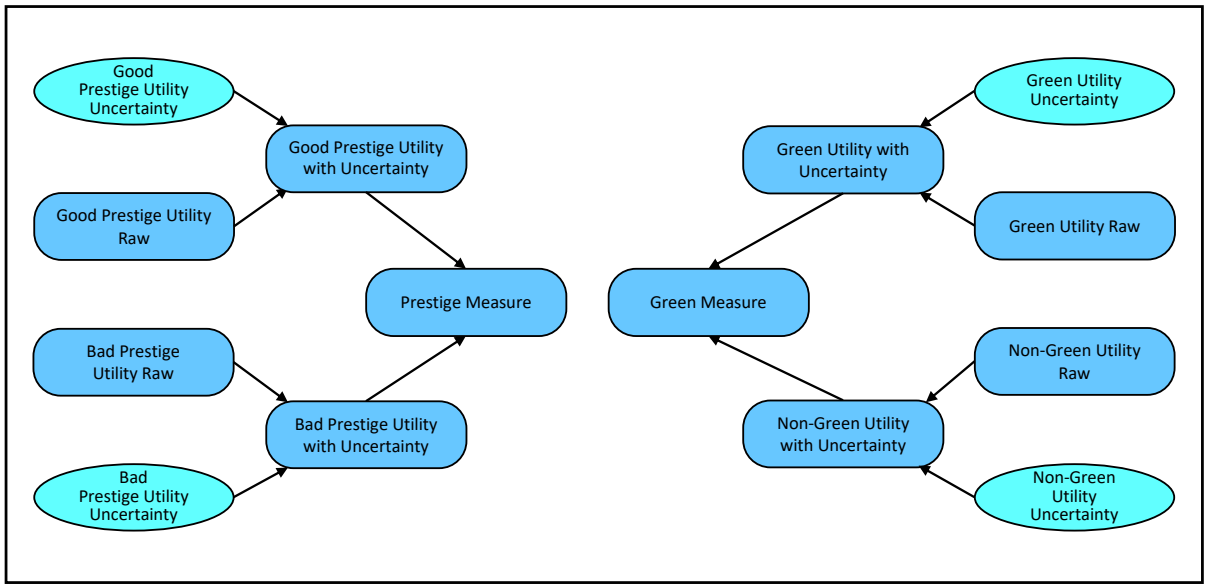
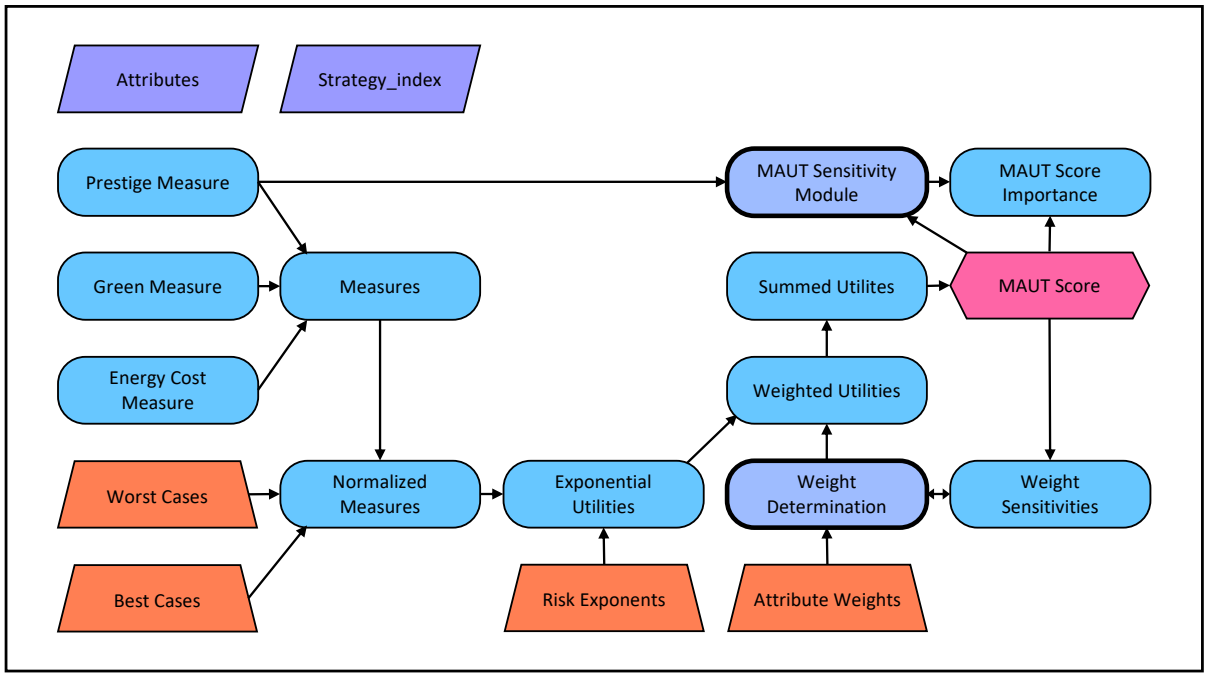


Figure 32 gets to the heart of the multi-attribute utility scoring methodology we apply in the model based on the previous discussion in Section 4.3.2.1. There is one main difference in our full model in that we actually use constant nodes for weights, risk exponents, and best/worst case measure values, given that we are providing the model with the values from the start. On the left are the raw score outputs from the Energy Cost, Green Value, and Prestige Value modules. Those raw scores are combined into a single table of values (Measures node) that are then normalized to a 0 to 1 scale (Normalized Measures node) by linearizing values from each measure from their worst to best case values. The next step converts the normalized measures into an exponential utility function (Exponential Utilities node) to account for the decision-maker's risk attitudes (Risk Exponents node). The scores are then weighted (Weighted Utilities node) based on their relative importance by multiplying each attribute's score by a weight (Weight Determination module) that is determined by the decision-maker or energy advisor. The final step is to sum up the attribute scores for all three attributes for each strategic alternative to give a resulting MAUT Score. There are a few remaining nodes within the MAUT

module, including the nodes used to calculate the Weights and MAUT Sensitivity Analysis outputs.

Figure 32: Analytica model multi-attribute utility score diagram.



4.3.4 Data summary

Next, we discuss the critical data fed into the model and used for analysis, focusing on the three main attributes previously discussed in Section 4.3. This discussion focuses on explaining the data we use in the case study, but the model framework is generalizable to numerous similar situations with different data formats and inputs. For the Cost of Energy, we use a combination of actual values from the case study and estimates where necessary. The energy-related data we use was based on prospective customer data, including historical energy usage and demand from the preceding 12 months, associated types of charges for that usage and demand, and Value Added Tax (VAT). This actual usage data feeds all the calculations for the different kinds of electricity, with the underlying assumption that the customer's electricity usage will

be consistent regardless of the electricity type procured. Using retail cost rates, we build rate estimates for the other types of electricity. We estimate the Adder Rate by determining the percentage of historical bill expenses due to non-core charges and then use that percentage as the baseline for the Adder Rate. We estimate the remaining inputs based on energy advisor input, including PPA Elevator, Retail Elevator, Inflation Rate, CAPEX, OPEX, cost of Renewable Energy Credits, and possible Fines for not mitigating greenhouse gases. Finally, we incorporate a discount rate to account for the time value of money, using a normal distribution for this value with a mean of 4% and standard deviation of 1.2%.

The data we use for both the Green Value and the Prestige Value are estimated utility values based on inputs from the case study, with the caveat that the utility values will change based on the customer, the decision-maker's risk attitude, and an individual company's strategy for energy procurement. We use a static number from 0 to 5 for the utility for each of the strategies, with 0 being the least utility value to 5 being the highest utility value. We make the utility values probabilistic by applying distributions skewed towards higher values if the strategy adds to the sustainability or skews towards a lower value if the strategy does not. To achieve what we believe to be representative skews, we use a Beta distribution for the Green Value and a Log-Normal distribution for the Prestige Value. Again, these distributions are representative of possible decision-makers' inputs and could be represented accordingly with other distributions based on the details of any given scenario.

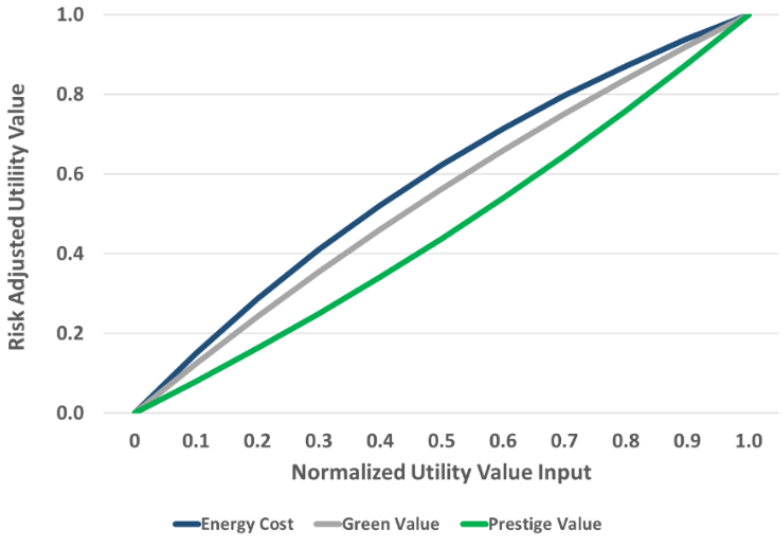
Finally, we introduce additional shaping variables to calculate the final MAUT score for the multi-attribute utility calculations. To start, we use upper and lower bounds to enable normalizing each attribute's measure to a value from 0 to 1. We use weights that allow each attribute to contribute to the overall score based on the importance level the decision-maker provides to the advisory team. The weights must sum to 1 to keep the overall utility score bounded between 0 and 1. Finally, we provide a mechanism to apply the decision-maker's risk attitude to each attribute through designated exponents shaping the utility value (McNamee and

Celona 2008). Table 28 shows the representative shaping values and weights for each of the attributes, and Figure 33 shows the risk-attitude-shaped curves for each attribute. For instance, the Energy Cost curve is adjusted to be convex, bowing upwards and to the left, which results in the decision-maker attitude being one where they have a lower tolerance to the risk associated with energy costs. In this instance, if the utility value of the model would be 0.5 for energy cost prior to accounting for the risk attitude, it will be closer to 0.6 after adjusting it for the risk attitude.

Table 28: Multi-attribute utility shaping values.

Attributes	Normalizing Lower Bound	Normalizing Upper Bound	Exponent (Risk Attitude)	Weights
Energy Cost	$1.01B	$2.81B	1	0.7
Green Value	0	5	0.5	0.2
Prestige Value	0	5	-0.5	0.1

Figure 33: Exponential utility functions for the three risk attitude attributes.



4.4 Results

In this section, we present and interpret the results of the Analytica model output as applied in our case study. Of interest is that we introduce several probabilistic variables in the model, which allows for a Median Latin Hypercube Sampling (McKay, Beckman, and Conover 1979) method to run many sample simulations. In this case, we use 1,000 simulations as a reasonable bound. The simulation functionality of Analytica provides a range of results that help model and analyze uncertainties such that we can evaluate more than just a static expected value result (Lumina Decision Systems 2020).

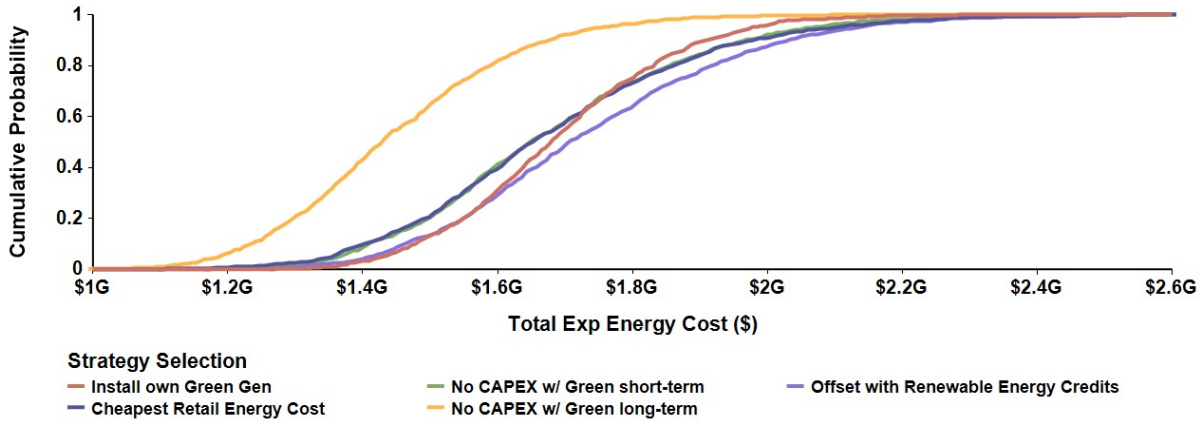
We first present the 20-year total cost determination results with expected cost and statistics by strategy in Table 29, followed by the Cumulative Distribution of those total costs in Figure 34. The results show that the strategy most likely to give us the lowest cost option over 20 years, at an expected discounted cost of $1.44B, is the "No CAPEX w/ Green long-term" strategy. This option entails working with a supplier to contract rates for a long-term Power Purchase Agreement (PPA) with known contracted annual increases. This option also is less volatile than most, with the second smallest standard deviation after "Install Own Green Gen" and is $55M smaller than that of the largest standard deviation for the "Cheapest Retail Energy Cost" strategy. The most expensive option is "Offset with Renewable Energy Credits", at an expected discounted cost of $1.72B over 20 years, which is $280M more than the PPA option expected value.

The following summary information below is the output of the multi-attribute utility score calculations, which takes the utility value of the cost information above and combines it with the utility values for Green Value and Prestige Value. Table 30 provides the expected utility value and summary statistics by strategy alternative. Once again, the best strategy score for utility is the "No CAPEX w/ Green long-term" strategy, with a utility value of 0.80, but this time around it is actually tied with the "Install Own Green Gen" strategy. When looking at

Table 29: Total discounted energy costs over 20 years by strategy selection.

Strategy Alternatives	Mean ($B)	Min ($B)	Median ($B)	Max ($B)	Std. Dev ($M)
Install own Green Gen	$1.69	$1.19	$1.68	$2.30	$169
Cheapest Retail Energy Cost	$1.68	$1.11	$1.66	$2.81	$229
No CAPEX w/ Green short-term	$1.68	$1.15	$1.66	$2.70	$220
No CAPEX w/ Green long-term	$1.44	$1.01	$1.44	$2.17	$174
Offset with RECs	$1.72	$1.19	$1.72	$2.78	$225

Figure 34: Cumulative distribution of total costs by strategy.



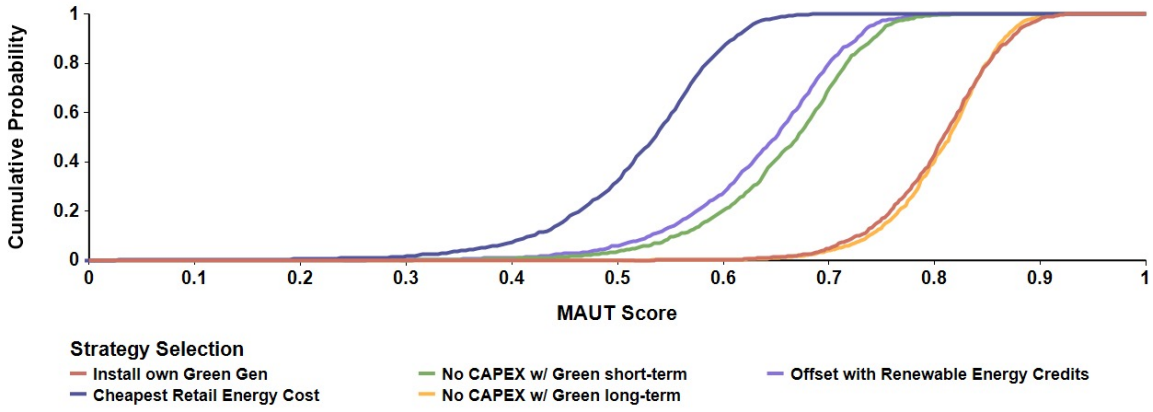
the cumulative distribution in Figure 35, the two strategies are overlapping each other. More interestingly, the max value of "Install Own Green Gen" actually is larger at 0.95, meaning if conditions are right, it is the best option. That overlap is seen clearly in Figure 35, with the red line surpassing the yellow near the top of the probability curves. Also, note that "Cheapest Retail Energy Cost" has lost ground on all other strategy alternatives as its expected utility value of 0.52 is the smallest, even though its expected value for cost was in the middle of the

pack. This reduction is not surprising and is attributable to this option having limited utility for both Green Value and Prestige Value.

Table 30: Statistics for multi-attribute utility scores by strategy selection.

Strategy Alternatives	Mean	Min	Median	Max	Std. Dev
Install own Green Gen	0.80	0.57	0.81	0.95	0.057
Cheapest Retail Energy Cost	0.52	0.01	0.53	0.68	0.080
No CAPEX w/ Green short-term	0.65	0.21	0.67	0.81	0.075
No CAPEX w/ Green long-term	0.80	0.55	0.81	0.92	0.052
Offset with RECs	0.63	0.16	0.63	0.80	0.080

Figure 35: Cumulative distribution of multi-attribute utility scores by strategy.



The final results of interest are from the sensitivity analysis output of select input variables for each strategic alternative. We apply the technique of varying input variables one-by-one across three values of min (50%), baseline (100%), and max (150%) while holding all other variables static to observe the sensitivity of the utility values to each inputs' changes (Eschenbach and McKeague 1989). This method enables us to get the plots shown in Figure 36. If the

graphs were sorted top to bottom based on the bars' width, we would have a classic Tornado Plot (Lumina Decision Systems 2018). We do not sort the graphs in that manner here to enable easier comparison of the sensitivity variables between the different strategic alternatives.

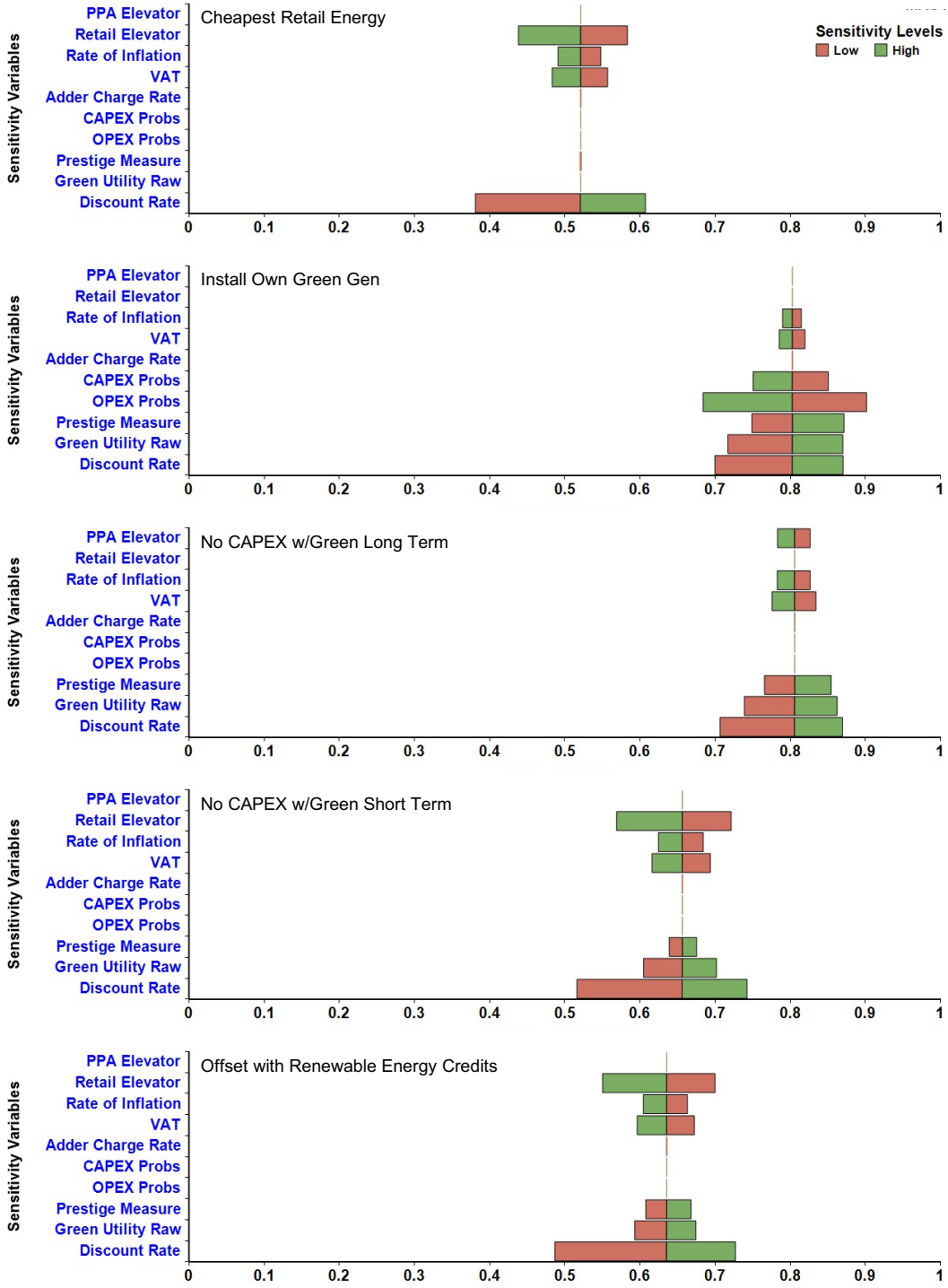
To read the graph in Figure 36, we will walk through a quick example explaining one of the bars. When looking at the "No CAPEX w/ Green short-term" strategy and the bar for "VAT," the reader will notice that the green bar shows a reduction of utility to $\sim$0.6 and an increase to $\sim$0.7 with the red bar. These changes can be interpreted to mean that all else being held constant; if the VAT increased by a factor of 50% above the values used in the model, the overall utility for this strategy would drop to $\sim$0.6. Conversely, if the VAT were reduced by 50%, the overall utility for this strategy would increase to $\sim$0.7. With that understanding, here are a few takeaways from the sensitivity plots:

1. The "No CAPEX w/ Green long-term" strategy does not show much sensitivity variation to the input variables. Its lowest values on the chart are nearly all higher than three of the other strategies' highest values (except "Install Own Green Gen"), which lends more strength to it being the strategy with the most utility and the least volatility.

2. The "Install Own Green Gen" strategy is highly sensitive to changes in the OPEX and CAPEX values. This sensitivity makes sense, given that OPEX is one of the most significant contributors to the expected cost. If OPEX increases by 50%, the utility value for this strategy drops by $\sim$20%. This strategy is the only one that can compete with the "No CAPEX w/ Green long-term" strategy based on the substantial overlap between their utility sensitivities, but has significantly more volatility, so the OPEX and CAPEX should be evaluated closely if pursuing this option.

3. The "No CAPEX w/ Green short-term" and "Offset with Renewable Energy Credits" strategies are two lower-scored options that look very similar in their sensitivities. They are sensitive to several variables, specifically: Retail Elevator, Rate of Inflation, and

Figure 36: Strategic alternatives sensitivity analysis of selected parameters.



VAT. If any of these values increase, the utility of these strategies drops. On the flip side, decreases in these values could push their utility as high as 0.7 and make them look competitive with the low end of the "Install Own Green Gen" strategy.

4. The final strategy of "Cheapest Retail Energy" shows similar sensitivity to Retail Elevator, Inflation Rate, and VAT. Of note is that varying the Prestige and Green Measures shows limited sensitivity on the "Cheapest Retail Energy" strategy's utility score due to the low utility values assigned to this strategy from the start.

5. Finally, the discount rate shows the largest sensitivity to change for all inputs across strategies, except for OPEX for "Install Own Green Gen." This sensitivity cutting across all the strategies may be a gating factor for any final decisions to choose a strategy. However, it should have the same impact regardless of which strategy is chosen, given its consistency in sensitivity across the different strategies.

4.5 Discussion and conclusion

The case study results suggest a best choice to pursue a sustainable energy strategy based on more than just the cost by incorporating multiple utility values for non-cost attributes. In fact, when looking at solely costs, there appears to be one winning strategy of "No CAPEX w/ Green long-term," but when looking at utility scores, the "Install Own Green Gen" strategy looks equally as attractive except for a wider sensitivity range due to OPEX and CAPEX uncertainty. The remaining alternatives are sufficiently far away on both the Energy Costs and MAUT Scores that in this particular case study looking out over a 20-year time horizon, we would not recommend pursuing them as viable alternatives. We construct this model for this case study's particulars, meaning the model output results could change significantly for another customer with different electricity market conditions, sustainability goals, and marketing strategies.

As a first effort, the model has some limitations. Many of the energy cost input variables are based on estimates or simplified probability distributions to provide uncertainty bands around the forecast. Future applications in specific electricity markets would ideally utilize a history of energy prices, forecasted future prices, rate structures, known rates, and a band of expected values for other variables tied to the location and customer needs. These values are all key contributors to the determination of energy cost and need to be closely aligned with actual customer situations to choose the best alternative. Additionally, for the sake of simplicity, the current model does not include parameters for foreign exchange rates, but this could be a helpful extension of the model when looking across countries and regions.

In developing the Green and Prestige utility value functions, we simplify the scoring to a scale from 0 to 5 and assign illustrative values based on our experience in this industry. In practice, decision-maker utility or value functions ought to be elicited and codified through the techniques available to decision analysts (Ayyub 2001). Additionally, the shaping variables for the MAUT calculation in our case study are all values we estimate based on the energy advisor's experience.

The application of Analytica to this decision-support model proves to be very useful. Analytica has an initial learning curve – specifying models takes time beyond creating the graphical representation. Nevertheless, it is well-documented and well-supported. This model's complexity could be replicated in a spreadsheet environment, but Analytica's interface allows the combination of developers, experts, users, and decision-makers to work more smoothly, with less potential for user error. Similarly, the heavy lifting of the model's array mathematics and simulation could be coded in a programming language, but again, that results in a black box that the decision-maker will need to trust without being able to walk through the model visually. Some decision-makers may not trust models that they cannot see and understand. Analytica's object-oriented approach, indexing of arrays, and visual/tabular outputs make it simple to walk through the model and explain how the model works to non-technical decision-makers and in-

terested stakeholders. Note, Analytica can also serve as a front-end that can pass data to and from spreadsheets or other common programming environments as needed.

We see several potential directions forward from our research in this study. The development of general modules in Analytica for both Strategy Tables and Multi-attribute Utility Theory provides a new application for decision modeling using these techniques. We applied these modules in a particular case study for a strategic energy purchasing decision, but the generalized module can be used in any complex decision-making scenario being modeled within Analytica. When looking more closely at our case study, we only evaluate the decision by looking across five decision points. With significant developments in energy storage over the last few years, there is an opportunity to include energy storage as part of the analysis, which introduces considerable complexity but also presents the opportunity to substantially improve the costs and efficiency of energy purchases. For practitioners, the ability to assess the purchase as we have in this model, combined with the real-time energy storage optimization products on the market (Baxter 2021), could enable significant cost reductions and more focus on intermittent renewable energy generation capability for the energy purchase decision.

The tool developed here supports sustainable energy strategy decisions entailing multiple attributes and business objectives. Much literature exists showing the application of multi-criteria decision-making in the sustainable energy space; however, there is limited applied research aimed at the challenge global companies face in effectively managing costs while evaluating the utility of meeting sustainability requirements and capturing added brand prestige. Incorporating multi-attribute utility theory helps to capture both the risk attitude of the decision-maker as well as the importance of the three key attributes of Energy Cost, Green Value, and Prestige Value, each of which can be calculated based on other business information. The use of a Strategy Table reduces the decision scope to a reasonable selection of strategic options, thereby reducing the complexity of the problem presented to the decision-maker without loss of generality. By combining these two decision analysis methods into a

functioning model with useful summary visualizations and output, we have created a model and toolkit that can be used to support company executives in evaluating their decisions. This approach has potential for broader application in the field of sustainable energy procurement strategic decision-making.

Beyond its topical domain, this case study demonstrates the potential for MCDA modelers to leverage Analytica software's rich scientific and financial modeling environment. In particular, Analytica's modular capability makes it straightforward to connect strategic MCDA methods with modeling activities associated with building decision-support tools (e.g., visualizations, simulations, array calculations, probability distributions). We introduced two general modules that can be applied to many different decision-making challenges. We then applied these modules in a functioning model using Analytica for a real-world use case that incorporated reducing the numerous decision combinations into a few distinct strategies that will be easier to explain to a time-strapped executive decision-maker. Finally, Analytica's visual interfaces and layered model views help explain the case study model's inputs and outputs in a manner that is easy to understand. Through the generic modules and the decision-analytic methods they facilitate, the bridge to Analytica, and the sustainable energy procurement itself, this effort aims to enhance MCDA practice.

CHAPTER 5

CONCLUSION

In this dissertation, we study the challenge of climate-change-related issues through an applied business analytics lens. We apply management, decision, information, and data science theories and techniques to build and evaluate novel data-driven methodologies to improve our understanding of climate-change-related challenges. Each chapter's analysis applies one or more business analytics approaches (i.e., descriptive, prescriptive, and predictive). We highlight key characteristics of our data sources, demonstrating how to incorporate various datasets into the analyses, whether directly or after transforming the data. To help the reader understand the results of each analysis, we apply varied methods of explaining outputs, from simple tables to detailed visualizations such as correlation heatmaps and sensitivity analysis using tornado plots. Our analyses focused on a broad swath to demonstrate the agility of using business analytics on varied problems. We start by first describing the climate change adaptation literature in a comprehensive manner using natural language processing. We then evaluate distinct areas of climate adaptation (i.e., flood protection) and mitigation (i.e., sustainable energy purchasing) to apply business analytics approaches to develop general decision support models that we then apply to to our specific case case studies.

In Chapter 2, we apply the natural language processing technique of Latent Dirichlet Allocation (LDA) to gain insights into the topics discussed in the climate change adaptation literature. Climate change adaptation research has grown substantially over the last two decades, seeing nearly 100x growth in published papers associated with this topic since the turn of the century. This study identifies 16 distinct topics from the literature through our supervised use

of LDA. Our modeling approach includes varying key input parameters (e.g., the number of potential topics) to construct hundreds of prospective models. We then use quantitative and qualitative techniques to narrow down to the one model that best fits and identifies a valid and representative breakdown of the discourse within this literature. When evaluating these topics over the period, it is clear that there is a substantial focus on policy and frameworks within the vast majority of the corpus. We observe a limited emphasis on financial elements as evidenced by the absence of majority weighting of finance-related issues. Therefore, we adopt a finance-oriented approach for the next two studies. We develop comprehensive models that leverage business analytics methodologies to address climate-change-related challenges from a financial perspective. In Chapter 3, we delve into the investment requirements and costs associated with mitigating the impact of climate-induced sea-level rise on an urban coastal region. In seeing the substantial and uncertain costs uncovered in Chapter 3, we see the need to move upstream and evaluate methods to mitigate climate change effects at the source. Therefore, in Chapter 4, we introduce a decision support model to assist businesses in reducing their contribution to greenhouse gas emissions, which are major drivers of climate change.

In Chapter 3, we evaluate a coastal city's flooding challenges caused by rising sea levels due to climate change. This problem is universal for all coastal cities and is projected to cost trillions of dollars of damage around the globe. We apply a prescriptive analytics method to create a cost-benefit analysis that simultaneously accounts for flood damages while building a flood protection network. We do this by superimposing a network layer on a coastal area, enhancing each network node with real-world parameters associated with monetary value and elevation, and subjecting that network to rising sea levels. In defining the problem, we also prove the decision version of this problem to be NP-complete. Due to the intractability of the core model, we use both a scenario-based and a simulation-based approach to assess the potential cost benefits across a range of input parameters. In our simulations, we apply stochastic values to the various prospective sea levels over time and optimize for the least cost option to

jointly build defenses to protect the city and evaluate the associated flood damages that will occur based on those defenses. In the scenarios, we assess the costs over a more comprehensive range of outcomes to help decision-makers assess their willingness to take on risks and handle associated costs. Our novel network flow approach to the problem of jointly building flood defenses and mitigating flood damages has not been used before. Our East Boston case study simulations show a significant reduction of up to 92.2% from the "do-nothing" strategy. The multi-stage character of our model and its recourse feature allows prospective community leaders to revise their adaptation measures as the sea level rise situation unfolds. Planners can use only open-source data to apply this methodology in any coastal area with the same open-source data availability. The work completed in chapter 3 demonstrates the ability to combine many varied datasets and apply all three business analytics approaches to provide novel decision support capability to support coastal decision-makers in the face of rising seas.

Chapter 3 showed significant uncertainty about the expected levels of climate-change-induced sea level rise. Therefore, in Chapter 4, we focus on upstream actions that could reduce the worst-case climate change scenarios and lower the overall cost required for future climate change adaptation. In this chapter, we develop a decision support tool to aid global companies in making long-term strategic sustainable energy purchases. The value of such a tool from the climate change perspective is that through increased sustainable energy usage, there should be a requisite reduction in greenhouse gas emissions and improvement in the future prognostication of worst-case impacts of climate change. We find many studies on multicriteria decision analysis in the sustainable energy space, but the vast majority only focus on technological or economic aspects of energy usage. Seeing this gap in the literature, we approach the problem from a more generalized business perspective for any global company looking to make a long-term strategic energy purchase. Our approach focuses not only on costs and sustainability requirements but also on how a company could enhance its brand and improve its competitiveness by aligning its sustainable energy purchasing strategies. If global companies incorporate

more than just cost into their decision-making, they potentially gain more utility value in purchasing sustainable energy in the long term. Compared to the work done in Chapter 3, we have significantly fewer data points in demonstrating this case study. However, we highlight how to apply subject matter expertise to provide reasonable estimates and probabilities to conduct the simulation. Finally, in building the model for our case study, we create new generalizable modules to incorporate Strategy Tables and Multi-attribute Utility Theory into the decision support software provided by Lumina Analytica.

As the introduction states, climate change is an existential threat facing humanity, civilization, and the natural world. Our primary purpose in conducting this research was to apply management, decision, information, and data science theories and techniques to propose, build, and evaluate novel data-driven methodologies to improve understanding of climate-change-related challenges. We start by assessing climate change adaptation literature to understand the main covered topics. After identifying less focus on finance-related issues, we pursue business analytics approaches to assess both climate change adaptation and mitigation challenges. We accomplish what we set out to do in this work by demonstrating enhanced data-driven decision-support methods to help inform businesses and society as they seek to address the deep uncertainty and incomplete knowledge of climate change issues.

APPENDIX A

SETS, PARAMETERS AND VARIABLES USED IN THE FLOOD RISK MITIGATION (FRM) MODEL IN CHAPTER 3

- t_{max}: Number of periods within the planning horizon.
- $\mathcal{T}$: The set containing all time periods within the planning horizon, i.e., $\mathcal{T} = \{1, \ldots, t_{max}\}$.
- s: The sea level during a period solely due to the climate change effects.
- $\hat{s}$: The sea level during a period due to both climate change effects and hurricane storm surge factors.
- $\mathcal{S} = (s, \hat{s})$: The sea level state during a period.
- Ξ_t.: The set containing all possible sea level states during a period t.
- $p_t^{\mathcal{S}\mathcal{S}'}$: The probability that the sea level state during period t is $\mathcal{S}$ and during period $t+1$ is $\mathcal{S}'$ for a given $t \in \{0, \ldots, t_{max} - 1\}$, $\mathcal{S} \in \Xi_t$ and $\mathcal{S}' \in \Xi_{t+1}$.
- $\hat{s}_{max}$: The highest sea level across all sea level states, i.e., $\hat{s}_{max} = \max\{\hat{s} : (s, \hat{s}) \in \Xi_{t_{max}}\}$.
- Φ: The set containing the land grids forming the "area of interest", which are the flooded land grids in the region of interest under sea level $\hat{s}_{max}$.
- Ψ: The set containing the land grids forming the "area of relevance", which are the flooded land grids in the surrounding region with a water path to some flooded land grid in the region of interest without going through the sea under sea level $\hat{s}_{max}$.
- h_i: The initial elevation for a land grid $i \in \Phi \cup \Psi$.
- O: The set containing the hexagonal grids formed on the sea at time zero.

- $R_t^{\mathcal{S}}$: The subset of land grids in Φ at risk of permanent inundation flooding during period t and under sea level state $\mathcal{S} \in \Xi_t$.

- $Q_t^{\mathcal{S}}$: The subset of land grids in Ψ at risk of permanent inundation flooding during period t and under sea level state $\mathcal{S} \in \Xi_t$.

- $\hat{R}_t^{\mathcal{S}}$: The subset of land grids in Φ only at risk of temporary flooding during period t and under sea level state $\mathcal{S} \in \Xi_t$.

- $\hat{Q}_t^{\mathcal{S}}$: The subset of land grids in Ψ only at risk of temporary flooding during period t and under sea level state $\mathcal{S} \in \Xi_t$.

- o: The vertex representing all sea-based grids within a given at-risk network.

- $N_t^{\mathcal{S}}(i)$: The set containing vertices that are adjacent to a vertex $i \in R_t^{\mathcal{S}} \cup \hat{R}_t^{\mathcal{S}} \cup Q_t^{\mathcal{S}} \cup \hat{Q}_t^{\mathcal{S}}$ within the at-risk network during a period t and under a sea level state $\mathcal{S} \in \Xi_t$. This set is also referred to as the neighbors of i within the aforementioned at-risk network.

- c: The cost to elevate a grid in Φ by one meter at the start of a given period.

- g_i: The cost of losing a grid $i \in \Phi$ due to inundation if the grid is in $R_t^{\mathcal{S}}$ during a period t and under a sea level state $\mathcal{S} \in \Xi_t$, and is permanently flooded.

- f_i: The cost of hurricane storm surge flood damage to a grid i in $R_t^{\mathcal{S}} \cup \hat{R}_t^{\mathcal{S}}$ during a period t and under a sea level state $\mathcal{S} \in \Xi_t$ when the grid is temporarily flooded to a depth of 1m.

- $\bar{f}_i$: The estimated value for f_i based on tax parcel data and is basis for sensitivity analysis of f_i.

- λ: The discount rate per period to incorporate realistic costs over time.

- d: The standard annual discount rate.

- b_t: The fixed construction and maintenance budget for a period $t \in \mathcal{T}$ that does not carry over into other periods.

- $x_{it\mathcal{S}}$: The decision variable associated with the height of a grid $i \in \Phi$ during a period $t \in \mathcal{T}$ if the sea level state during period $t-1$ is $\mathcal{S} \in \Xi_{t-1}$.

- m: The minimum threshold of elevation increase in any grid in Φ during a period.

- M: A valid upper bound on the elevation increase in any grid in Φ during a period. We use $M = \max\{\hat{s}_{max} - \min\{h_i : i \in \Phi \cup \Psi\}, m\}$ in our FRM model.

- $w_{it\mathcal{S}\mathcal{S}'}$: A binary variable that captures if a grid $i \in R_t^{\mathcal{S}'} \cup Q_t^{\mathcal{S}'}$ is inundated during a period $t \in \mathcal{T}$ and under sea level states $\mathcal{S} \in \Xi_{t-1}$ and $\mathcal{S}' \in \Xi_t$ for which $p_{t-1}^{\mathcal{S}\mathcal{S}'} > 0$. $w_{it\mathcal{S}\mathcal{S}'} = 0$ if the grid is not inundated, and $w_{it\mathcal{S}\mathcal{S}'} = 1$ otherwise.

- $z_{it\mathcal{S}\mathcal{S}'}$: A continuous variable to capture the water depth used to calculate the temporary flood cost if a grid $i \in R_t^{\mathcal{S}'} \cup Q_t^{\mathcal{S}'} \cup \hat{R}_t^{\mathcal{S}'} \cup \hat{Q}_t^{\mathcal{S}'}$ is not inundated but faces hurricane storm surge related flooding during period t and under sea level states $\mathcal{S} \in \Xi_{t-1}$ and $\mathcal{S}' \in \Xi_t$ for which $p_{t-1}^{\mathcal{S}\mathcal{S}'} > 0$.

- $v_{it\mathcal{S}\mathcal{S}'}$: A binary variable indicating whether a grid $i \in \Phi$ is elevated at the start of a period $t \in \{2, \ldots, t_{max}\}$ under sea level states $\mathcal{S} \in \Xi_{t-2}$ and $\mathcal{S}' \in \Xi_{t-1}$ for which $p_{t-2}^{\mathcal{S}\mathcal{S}'} > 0$.

- $v_{i1(0,0)}$: A binary variable indicating whether a grid $i \in \Phi$ is elevated at the start of the first period.

- $y_{it\mathcal{S}\mathcal{S}'}$: A binary variable that indicates whether a grid $i \in R_t^{\mathcal{S}'} \cup Q_t^{\mathcal{S}'}$ is a neighbor of the sea grid or has an adjacent grid $i' \in N_t^{\mathcal{S}'}(i) \cap (R_t^{\mathcal{S}'} \cup Q_t^{\mathcal{S}'})$ that is inundated during period t and under sea level states $\mathcal{S} \in \Xi_{t-1}$ and $\mathcal{S}' \in \Xi_t$ for which $p_{t-1}^{\mathcal{S}\mathcal{S}'} > 0$. If grid i is not a neighbor of the sea grid and also does not have an inundated neighbor, then $y_{it\mathcal{S}\mathcal{S}'}$ is zero, and one otherwise.

- $\hat{y}_{it\mathcal{S}\mathcal{S}'}$: A binary variable that indicates whether a grid $i \in R_t^{\mathcal{S}'} \cup Q_t^{\mathcal{S}'} \cup \hat{R}_t^{\mathcal{S}'} \cup \hat{Q}_t^{\mathcal{S}'}$ is a neighbor of the sea grid or has a hydraulic path to the sea during period t and under sea level states $\mathcal{S} \in \Xi_{t-1}$ and $\mathcal{S}' \in \Xi_t$ for which $p_{t-1}^{\mathcal{S}\mathcal{S}'} > 0$. $\hat{y}_{it\mathcal{S}\mathcal{S}'}$ is equal to one if grid i is a neighbor of the sea or if such a path exists, and zero otherwise.

APPENDIX B

SEA LEVEL STATES PATHS, PROBABILITIES, AND SEA LEVEL VALUES FOR EACH PERIOD IN CHAPTER 3

Table 31: Sea level states paths, probabilities, and sea level values in cm for each period.

Path	RCP	SLR Curve	Surge Curve	Path Probability	2030 $(s, \hat{s})$	2040 $(s, \hat{s})$	2050 $(s, \hat{s})$	2060 $(s, \hat{s})$	2070 $(s, \hat{s})$
1	rcp85	1.0%	99%	0.001912500	(154, 210)	(156, 212)	(159, 215)	(161, 217)	(162, 218)
2	rcp85	5.0%	99%	0.005004375	(156, 212)	(159, 215)	(164, 220)	(169, 225)	(172, 228)
3	rcp85	16.7%	99%	0.014343750	(157, 213)	(162, 218)	(168, 224)	(175, 231)	(181, 237)
4	rcp85	50.0%	99%	0.021228750	(159, 215)	(167, 223)	(176, 232)	(186, 242)	(196, 252)
5	rcp85	83.3%	99%	0.014343750	(162, 218)	(172, 228)	(184, 240)	(197, 253)	(212, 268)
6	rcp85	95.0%	99%	0.005004375	(163, 219)	(175, 231)	(190, 246)	(207, 263)	(226, 282)
7	rcp85	99.0%	99%	0.001434375	(165, 221)	(181, 237)	(200, 256)	(222, 278)	(246, 302)
8	rcp85	99.5%	99%	0.000286875	(167, 223)	(184, 240)	(205, 261)	(230, 286)	(258, 314)
9	rcp85	99.9%	99%	0.000191250	(170, 226)	(194, 250)	(223, 279)	(256, 312)	(294, 350)
10	rcp60	1.0%	99%	0.001912500	(152, 208)	(154, 210)	(154, 210)	(156, 212)	(157, 213)
11	rcp60	5.0%	99%	0.005004375	(154, 210)	(156, 212)	(158, 214)	(163, 219)	(166, 222)
12	rcp60	16.7%	99%	0.014343750	(156, 212)	(160, 216)	(164, 220)	(169, 225)	(175, 231)
13	rcp60	50.0%	99%	0.021228750	(158, 214)	(163, 219)	(170, 226)	(179, 235)	(187, 243)
14	rcp60	83.3%	99%	0.014343750	(160, 216)	(168, 224)	(178, 234)	(189, 245)	(202, 258)
15	rcp60	95.0%	99%	0.005004375	(163, 219)	(172, 228)	(185, 241)	(199, 255)	(215, 271)
16	rcp60	99.0%	99%	0.001434375	(166, 222)	(178, 234)	(195, 251)	(212, 268)	(235, 291)
17	rcp60	99.5%	99%	0.000286875	(168, 224)	(181, 237)	(200, 256)	(220, 276)	(246, 302)
18	rcp60	99.9%	99%	0.000191250	(171, 227)	(190, 246)	(214, 270)	(245, 301)	(281, 337)
19	rcp45	1.0%	99%	0.001912500	(154, 210)	(155, 211)	(156, 212)	(157, 213)	(156, 212)
20	rcp45	5.0%	99%	0.005004375	(156, 212)	(159, 215)	(161, 217)	(165, 221)	(167, 223)
21	rcp45	16.7%	99%	0.014343750	(157, 213)	(161, 217)	(166, 222)	(171, 227)	(175, 231)
22	rcp45	50.0%	99%	0.021228750	(158, 214)	(165, 221)	(173, 229)	(181, 237)	(188, 244)
23	rcp45	83.3%	99%	0.014343750	(160, 216)	(170, 226)	(181, 237)	(192, 248)	(203, 259)
24	rcp45	95.0%	99%	0.005004375	(162, 218)	(174, 230)	(187, 243)	(202, 258)	(216, 272)

Continued on next page

Table 31: Sea level states paths, probabilities, and sea level values in cm for each period. (Continued)

Path	RCP	SLR Curve	Surge Curve	Path Probability	2030 $(s, \hat{s})$	2040 $(s, \hat{s})$	2050 $(s, \hat{s})$	2060 $(s, \hat{s})$	2070 $(s, \hat{s})$
25	rcp45	99.0%	99%	0.001434375	(164, 220)	(179, 235)	(197, 253)	(216, 272)	(236, 292)
26	rcp45	99.5%	99%	0.000286875	(166, 222)	(183, 239)	(203, 259)	(224, 280)	(249, 305)
27	rcp45	99.9%	99%	0.000191250	(169, 225)	(192, 248)	(219, 275)	(248, 304)	(281, 337)
28	rcp26	1.0%	99%	0.001912500	(152, 208)	(152, 208)	(151, 207)	(149, 205)	(150, 206)
29	rcp26	5.0%	99%	0.005004375	(154, 210)	(155, 211)	(156, 212)	(158, 214)	(160, 216)
30	rcp26	16.7%	99%	0.014343750	(155, 211)	(159, 215)	(162, 218)	(165, 221)	(168, 224)
31	rcp26	50.0%	99%	0.021228750	(158, 214)	(164, 220)	(170, 226)	(176, 232)	(182, 238)
32	rcp26	83.3%	99%	0.014343750	(161, 217)	(170, 226)	(180, 236)	(189, 245)	(197, 253)
33	rcp26	95.0%	99%	0.005004375	(163, 219)	(175, 231)	(188, 244)	(200, 256)	(210, 266)
34	rcp26	99.0%	99%	0.001434375	(166, 222)	(181, 237)	(198, 254)	(215, 271)	(231, 287)
35	rcp26	99.5%	99%	0.000286875	(167, 223)	(185, 241)	(204, 260)	(223, 279)	(243, 299)
36	rcp26	99.9%	99%	0.000191250	(170, 226)	(193, 249)	(221, 277)	(250, 306)	(278, 334)
37	rcp85	1.0%	50%	0.003337500	(154, 235)	(156, 237)	(159, 240)	(161, 242)	(162, 243)
38	rcp85	5.0%	50%	0.008733125	(156, 237)	(159, 240)	(164, 245)	(169, 250)	(172, 253)
39	rcp85	16.7%	50%	0.025031250	(157, 238)	(162, 243)	(168, 249)	(175, 256)	(181, 262)
40	rcp85	50.0%	50%	0.037046250	(159, 240)	(167, 248)	(176, 257)	(186, 267)	(196, 277)
41	rcp85	83.3%	50%	0.025031250	(162, 243)	(172, 253)	(184, 265)	(197, 278)	(212, 293)
42	rcp85	95.0%	50%	0.008733125	(163, 244)	(175, 256)	(190, 271)	(207, 288)	(226, 307)
43	rcp85	99.0%	50%	0.002503125	(165, 246)	(181, 262)	(200, 281)	(222, 303)	(246, 327)
44	rcp85	99.5%	50%	0.000500625	(167, 248)	(184, 265)	(205, 286)	(230, 311)	(258, 339)
45	rcp85	99.9%	50%	0.000333750	(170, 251)	(194, 275)	(223, 304)	(256, 337)	(294, 375)
46	rcp60	1.0%	50%	0.003337500	(152, 233)	(154, 235)	(154, 235)	(156, 237)	(157, 238)
47	rcp60	5.0%	50%	0.008733125	(154, 235)	(156, 237)	(158, 239)	(163, 244)	(166, 247)
48	rcp60	16.7%	50%	0.025031250	(156, 237)	(160, 241)	(164, 245)	(169, 250)	(175, 256)
49	rcp60	50.0%	50%	0.037046250	(158, 239)	(163, 244)	(170, 251)	(179, 260)	(187, 268)
50	rcp60	83.3%	50%	0.025031250	(160, 241)	(168, 249)	(178, 259)	(189, 270)	(202, 283)
51	rcp60	95.0%	50%	0.008733125	(163, 244)	(172, 253)	(185, 266)	(199, 280)	(215, 296)
52	rcp60	99.0%	50%	0.002503125	(166, 247)	(178, 259)	(195, 276)	(212, 293)	(235, 316)
53	rcp60	99.5%	50%	0.000500625	(168, 249)	(181, 262)	(200, 281)	(220, 301)	(246, 327)

Continued on next page

Table 31: Sea level states paths, probabilities, and sea level values in cm for each period. (Continued)

Path	RCP	SLR Curve	Surge Curve	Path Probability	2030 $(s, \hat{s})$	2040 $(s, \hat{s})$	2050 $(s, \hat{s})$	2060 $(s, \hat{s})$	2070 $(s, \hat{s})$
54	rcp60	99.9%	50%	0.000333750	(171, 252)	(190, 271)	(214, 295)	(245, 326)	(281, 362)
55	rcp45	1.0%	50%	0.003337500	(154, 235)	(155, 236)	(156, 237)	(157, 238)	(156, 237)
56	rcp45	5.0%	50%	0.008733125	(156, 237)	(159, 240)	(161, 242)	(165, 246)	(167, 248)
57	rcp45	16.7%	50%	0.025031250	(157, 238)	(161, 242)	(166, 247)	(171, 252)	(175, 256)
58	rcp45	50.0%	50%	0.037046250	(158, 239)	(165, 246)	(173, 254)	(181, 262)	(188, 269)
59	rcp45	83.3%	50%	0.025031250	(160, 241)	(170, 251)	(181, 262)	(192, 273)	(203, 284)
60	rcp45	95.0%	50%	0.008733125	(162, 243)	(174, 255)	(187, 268)	(202, 283)	(216, 297)
61	rcp45	99.0%	50%	0.002503125	(164, 245)	(179, 260)	(197, 278)	(216, 297)	(236, 317)
62	rcp45	99.5%	50%	0.000500625	(166, 247)	(183, 264)	(203, 284)	(224, 305)	(249, 330)
63	rcp45	99.9%	50%	0.000333750	(169, 250)	(192, 273)	(219, 300)	(248, 329)	(281, 362)
64	rcp26	1.0%	50%	0.003337500	(152, 233)	(152, 233)	(151, 232)	(149, 230)	(150, 231)
65	rcp26	5.0%	50%	0.008733125	(154, 235)	(155, 236)	(156, 237)	(158, 239)	(160, 241)
66	rcp26	16.7%	50%	0.025031250	(155, 236)	(159, 240)	(162, 243)	(165, 246)	(168, 249)
67	rcp26	50.0%	50%	0.037046250	(158, 239)	(164, 245)	(170, 251)	(176, 257)	(182, 263)
68	rcp26	83.3%	50%	0.025031250	(161, 242)	(170, 251)	(180, 261)	(189, 270)	(197, 278)
69	rcp26	95.0%	50%	0.008733125	(163, 244)	(175, 256)	(188, 269)	(200, 281)	(210, 291)
70	rcp26	99.0%	50%	0.002503125	(166, 247)	(181, 262)	(198, 279)	(215, 296)	(231, 312)
71	rcp26	99.5%	50%	0.000500625	(167, 248)	(185, 266)	(204, 285)	(223, 304)	(243, 324)
72	rcp26	99.9%	50%	0.000333750	(170, 251)	(193, 274)	(221, 302)	(250, 331)	(278, 359)
73	rcp85	1.0%	10%	0.001837500	(154, 261)	(156, 263)	(159, 266)	(161, 268)	(162, 269)
74	rcp85	5.0%	10%	0.004808125	(156, 263)	(159, 266)	(164, 271)	(169, 276)	(172, 279)
75	rcp85	16.7%	10%	0.013781250	(157, 264)	(162, 269)	(168, 275)	(175, 282)	(181, 288)
76	rcp85	50.0%	10%	0.020396250	(159, 266)	(167, 274)	(176, 283)	(186, 293)	(196, 303)
77	rcp85	83.3%	10%	0.013781250	(162, 269)	(172, 279)	(184, 291)	(197, 304)	(212, 319)
78	rcp85	95.0%	10%	0.004808125	(163, 270)	(175, 282)	(190, 297)	(207, 314)	(226, 333)
79	rcp85	99.0%	10%	0.001378125	(165, 272)	(181, 288)	(200, 307)	(222, 329)	(246, 353)
80	rcp85	99.5%	10%	0.000275625	(167, 274)	(184, 291)	(205, 312)	(230, 337)	(258, 365)
81	rcp85	99.9%	10%	0.000183750	(170, 277)	(194, 301)	(223, 330)	(256, 363)	(294, 401)
82	rcp60	1.0%	10%	0.001837500	(152, 259)	(154, 261)	(154, 261)	(156, 263)	(157, 264)

Continued on next page

Table 31: Sea level states paths, probabilities, and sea level values in cm for each period. (Continued)

Path	RCP	SLR Curve	Surge Curve	Path Probability	2030 $(s, \hat{s})$	2040 $(s, \hat{s})$	2050 $(s, \hat{s})$	2060 $(s, \hat{s})$	2070 $(s, \hat{s})$
83	rcp60	5.0%	10%	0.004808125	(154, 261)	(156, 263)	(158, 265)	(163, 270)	(166, 273)
84	rcp60	16.7%	10%	0.013781250	(156, 263)	(160, 267)	(164, 271)	(169, 276)	(175, 282)
85	rcp60	50.0%	10%	0.020396250	(158, 265)	(163, 270)	(170, 277)	(179, 286)	(187, 294)
86	rcp60	83.3%	10%	0.013781250	(160, 267)	(168, 275)	(178, 285)	(189, 296)	(202, 309)
87	rcp60	95.0%	10%	0.004808125	(163, 270)	(172, 279)	(185, 292)	(199, 306)	(215, 322)
88	rcp60	99.0%	10%	0.001378125	(166, 273)	(178, 285)	(195, 302)	(212, 319)	(235, 342)
89	rcp60	99.5%	10%	0.000275625	(168, 275)	(181, 288)	(200, 307)	(220, 327)	(246, 353)
90	rcp60	99.9%	10%	0.000183750	(171, 278)	(190, 297)	(214, 321)	(245, 352)	(281, 388)
91	rcp45	1.0%	10%	0.001837500	(154, 261)	(155, 262)	(156, 263)	(157, 264)	(156, 263)
92	rcp45	5.0%	10%	0.004808125	(156, 263)	(159, 266)	(161, 268)	(165, 272)	(167, 274)
93	rcp45	16.7%	10%	0.013781250	(157, 264)	(161, 268)	(166, 273)	(171, 278)	(175, 282)
94	rcp45	50.0%	10%	0.020396250	(158, 265)	(165, 272)	(173, 280)	(181, 288)	(188, 295)
95	rcp45	83.3%	10%	0.013781250	(160, 267)	(170, 277)	(181, 288)	(192, 299)	(203, 310)
96	rcp45	95.0%	10%	0.004808125	(162, 269)	(174, 281)	(187, 294)	(202, 309)	(216, 323)
97	rcp45	99.0%	10%	0.001378125	(164, 271)	(179, 286)	(197, 304)	(216, 323)	(236, 343)
98	rcp45	99.5%	10%	0.000275625	(166, 273)	(183, 290)	(203, 310)	(224, 331)	(249, 356)
99	rcp45	99.9%	10%	0.000183750	(169, 276)	(192, 299)	(219, 326)	(248, 355)	(281, 388)
100	rcp26	1.0%	10%	0.001837500	(152, 259)	(152, 259)	(151, 258)	(149, 256)	(150, 257)
101	rcp26	5.0%	10%	0.004808125	(154, 261)	(155, 262)	(156, 263)	(158, 265)	(160, 267)
102	rcp26	16.7%	10%	0.013781250	(155, 262)	(159, 266)	(162, 269)	(165, 272)	(168, 275)
103	rcp26	50.0%	10%	0.020396250	(158, 265)	(164, 271)	(170, 277)	(176, 283)	(182, 289)
104	rcp26	83.3%	10%	0.013781250	(161, 268)	(170, 277)	(180, 287)	(189, 296)	(197, 304)
105	rcp26	95.0%	10%	0.004808125	(163, 270)	(175, 282)	(188, 295)	(200, 307)	(210, 317)
106	rcp26	99.0%	10%	0.001378125	(166, 273)	(181, 288)	(198, 305)	(215, 322)	(231, 338)
107	rcp26	99.5%	10%	0.000275625	(167, 274)	(185, 292)	(204, 311)	(223, 330)	(243, 350)
108	rcp26	99.9%	10%	0.000183750	(170, 277)	(193, 300)	(221, 328)	(250, 357)	(278, 385)
109	rcp85	1.0%	1%	0.000412500	(154, 295)	(156, 297)	(159, 300)	(161, 302)	(162, 303)
110	rcp85	5.0%	1%	0.001079375	(156, 297)	(159, 300)	(164, 305)	(169, 310)	(172, 313)
111	rcp85	16.7%	1%	0.003093750	(157, 298)	(162, 303)	(168, 309)	(175, 316)	(181, 322)

Continued on next page

Table 31: Sea level states paths, probabilities, and sea level values in cm for each period. (Continued)

Path	RCP	SLR Curve	Surge Curve	Path Probability	2030 $(s, \hat{s})$	2040 $(s, \hat{s})$	2050 $(s, \hat{s})$	2060 $(s, \hat{s})$	2070 $(s, \hat{s})$
112	rcp85	50.0%	1%	0.004578750	(159, 300)	(167, 308)	(176, 317)	(186, 327)	(196, 337)
113	rcp85	83.3%	1%	0.003093750	(162, 303)	(172, 313)	(184, 325)	(197, 338)	(212, 353)
114	rcp85	95.0%	1%	0.001079375	(163, 304)	(175, 316)	(190, 331)	(207, 348)	(226, 367)
115	rcp85	99.0%	1%	0.000309375	(165, 306)	(181, 322)	(200, 341)	(222, 363)	(246, 387)
116	rcp85	99.5%	1%	0.000061875	(167, 308)	(184, 325)	(205, 346)	(230, 371)	(258, 399)
117	rcp85	99.9%	1%	0.000041250	(170, 311)	(194, 335)	(223, 364)	(256, 397)	(294, 435)
118	rcp60	1.0%	1%	0.000412500	(152, 293)	(154, 295)	(154, 295)	(156, 297)	(157, 298)
119	rcp60	5.0%	1%	0.001079375	(154, 295)	(156, 297)	(158, 299)	(163, 304)	(166, 307)
120	rcp60	16.7%	1%	0.003093750	(156, 297)	(160, 301)	(164, 305)	(169, 310)	(175, 316)
121	rcp60	50.0%	1%	0.004578750	(158, 299)	(163, 304)	(170, 311)	(179, 320)	(187, 328)
122	rcp60	83.3%	1%	0.003093750	(160, 301)	(168, 309)	(178, 319)	(189, 330)	(202, 343)
123	rcp60	95.0%	1%	0.001079375	(163, 304)	(172, 313)	(185, 326)	(199, 340)	(215, 356)
124	rcp60	99.0%	1%	0.000309375	(166, 307)	(178, 319)	(195, 336)	(212, 353)	(235, 376)
125	rcp60	99.5%	1%	0.000061875	(168, 309)	(181, 322)	(200, 341)	(220, 361)	(246, 387)
126	rcp60	99.9%	1%	0.000041250	(171, 312)	(190, 331)	(214, 355)	(245, 386)	(281, 422)
127	rcp45	1.0%	1%	0.000412500	(154, 295)	(155, 296)	(156, 297)	(157, 298)	(156, 297)
128	rcp45	5.0%	1%	0.001079375	(156, 297)	(159, 300)	(161, 302)	(165, 306)	(167, 308)
129	rcp45	16.7%	1%	0.003093750	(157, 298)	(161, 302)	(166, 307)	(171, 312)	(175, 316)
130	rcp45	50.0%	1%	0.004578750	(158, 299)	(165, 306)	(173, 314)	(181, 322)	(188, 329)
131	rcp45	83.3%	1%	0.003093750	(160, 301)	(170, 311)	(181, 322)	(192, 333)	(203, 344)
132	rcp45	95.0%	1%	0.001079375	(162, 303)	(174, 315)	(187, 328)	(202, 343)	(216, 357)
133	rcp45	99.0%	1%	0.000309375	(164, 305)	(179, 320)	(197, 338)	(216, 357)	(236, 377)
134	rcp45	99.5%	1%	0.000061875	(166, 307)	(183, 324)	(203, 344)	(224, 365)	(249, 390)
135	rcp45	99.9%	1%	0.000041250	(169, 310)	(192, 333)	(219, 360)	(248, 389)	(281, 422)
136	rcp26	1.0%	1%	0.000412500	(152, 293)	(152, 293)	(151, 292)	(149, 290)	(150, 291)
137	rcp26	5.0%	1%	0.001079375	(154, 295)	(155, 296)	(156, 297)	(158, 299)	(160, 301)
138	rcp26	16.7%	1%	0.003093750	(155, 296)	(159, 300)	(162, 303)	(165, 306)	(168, 309)
139	rcp26	50.0%	1%	0.004578750	(158, 299)	(164, 305)	(170, 311)	(176, 317)	(182, 323)
140	rcp26	83.3%	1%	0.003093750	(161, 302)	(170, 311)	(180, 321)	(189, 330)	(197, 338)

Continued on next page

Table 31: Sea level states paths, probabilities, and sea level values in cm for each period. (Continued)

Path	RCP	SLR Curve	Surge Curve	Path Probability	2030 $(s, \hat{s})$	2040 $(s, \hat{s})$	2050 $(s, \hat{s})$	2060 $(s, \hat{s})$	2070 $(s, \hat{s})$
141	rcp26	95.0%	1%	0.001079375	(163, 304)	(175, 316)	(188, 329)	(200, 341)	(210, 351)
142	rcp26	99.0%	1%	0.000309375	(166, 307)	(181, 322)	(198, 339)	(215, 356)	(231, 372)
143	rcp26	99.5%	1%	0.000061875	(167, 308)	(185, 326)	(204, 345)	(223, 364)	(243, 384)
144	rcp26	99.9%	1%	0.000041250	(170, 311)	(193, 334)	(221, 362)	(250, 391)	(278, 419)

APPENDIX C

FULL RESULTS FOR SIMULATION AND SCENARIO RUNS CONDUCTED IN CHAPTER 3

In this appendix, we present the full results for the sensitivity analysis in our simulation-based approach discussed in Section 3.5.1 and the scenario-based approach discussed in Section 3.5.2. The 81 plots in each of the following six figures represent results across all 81 parameter combinations, with an overall cost shown with stacked bar charts of the three cost types. Figure 37 shows the full results for the simulation runs discussed in Section 3.5.1. The remaining Figures in this appendix show the full results for the scenarios discussed in Section 3.5.2. We omit the zero budget cost curves because the large magnitude of these curves overwhelms the curves for the other budgets, and prevents an effective visualization of the budget effect in each cost benefit curve. At the top of each chart is the combination of parameters for that chart, with the values shown in Table 32. To effectively show each combination of 81 charts, we have split each multi-paneled figure across two pages on the following pages.

Table 32: Parameter values used in sensitivity analysis for charts shown in Appendix C.

Parameter	High	Mid	Low
Discount rate (d) [%]	7	5	3
Minimum elevation increase (m) [meters]	5	3	1
Grid elevation cost (c) [$M/km per m]	25	15	5
Storm flood damage curve (f_i) [$M/m]	$1.25\bar{f}_i$	$\bar{f}_i$	$0.75\bar{f}_i$

Figure 37: East Boston simulation expected cost benefit curves by per-period budget with expected costs averaged across full 144 simulated sea level states paths for each of the 81 parameter combinations. (1/2)

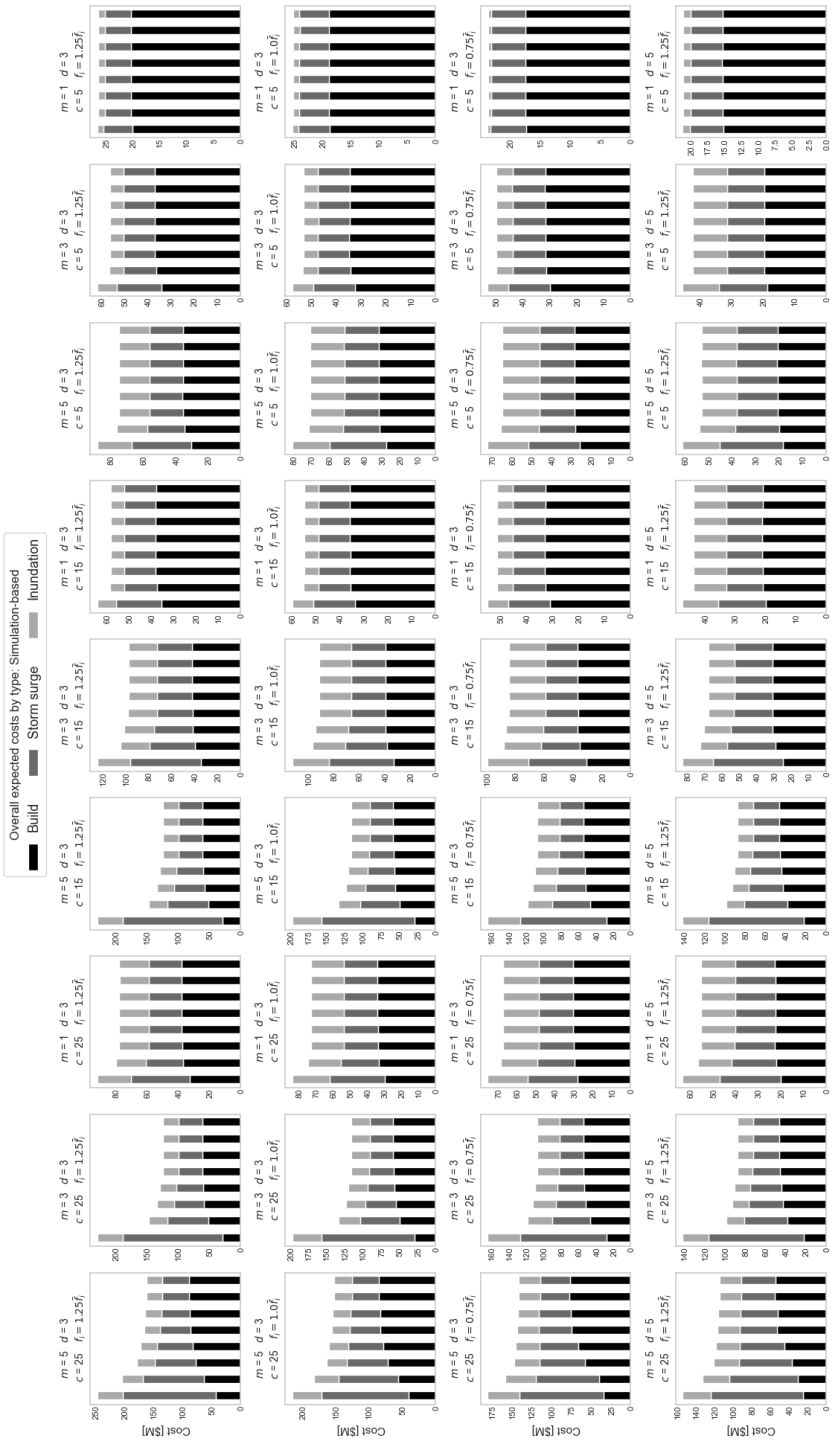


Figure 37: East Boston simulation expected cost benefit curves by per-period budget with expected costs averaged across full 144 simulated sea level states paths for each of the 81 parameter combinations. (2/2)

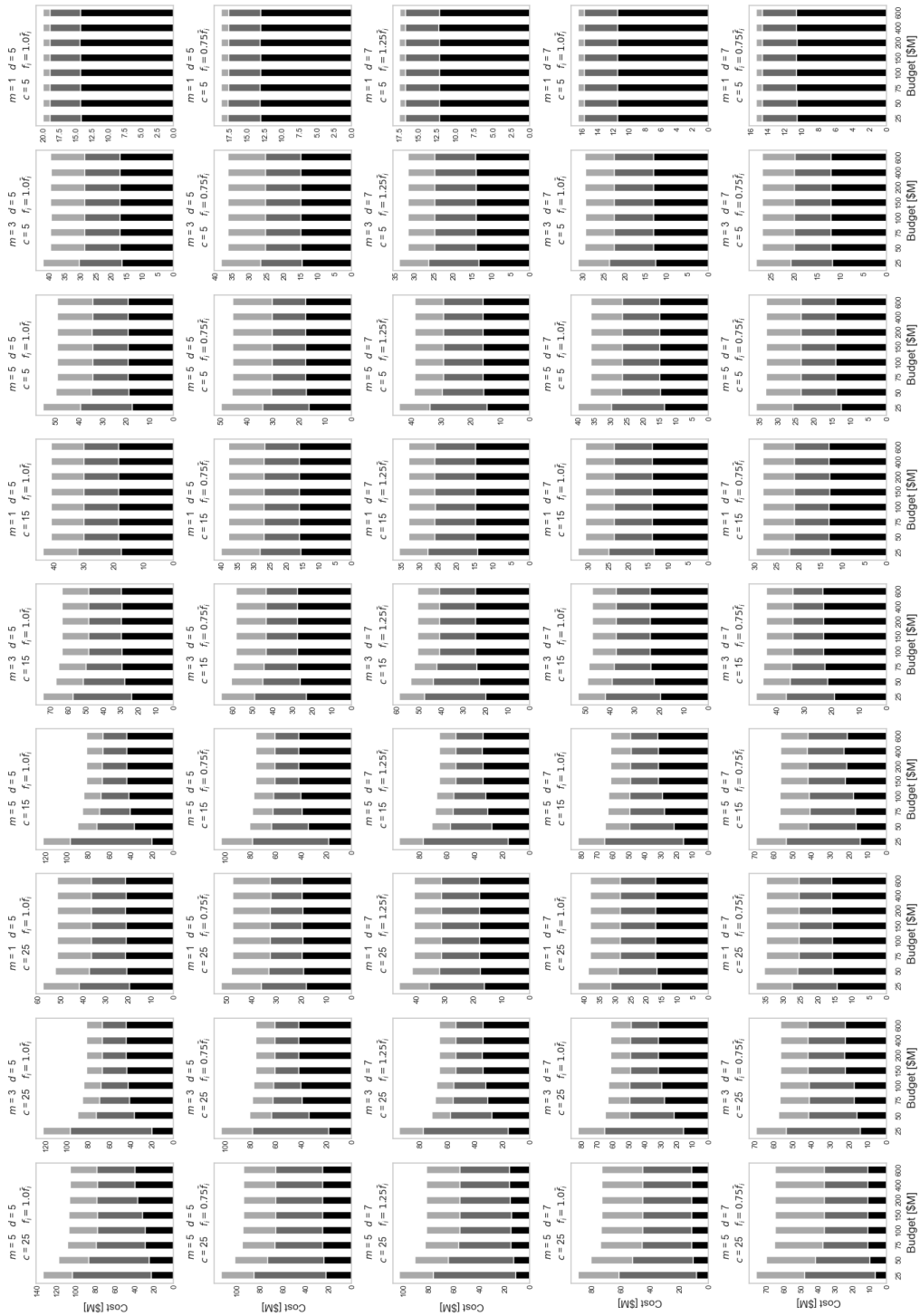


Figure 38: East Boston High scenario cost benefit curves by per-period budget for each of the 81 parameter combinations. (1/2)

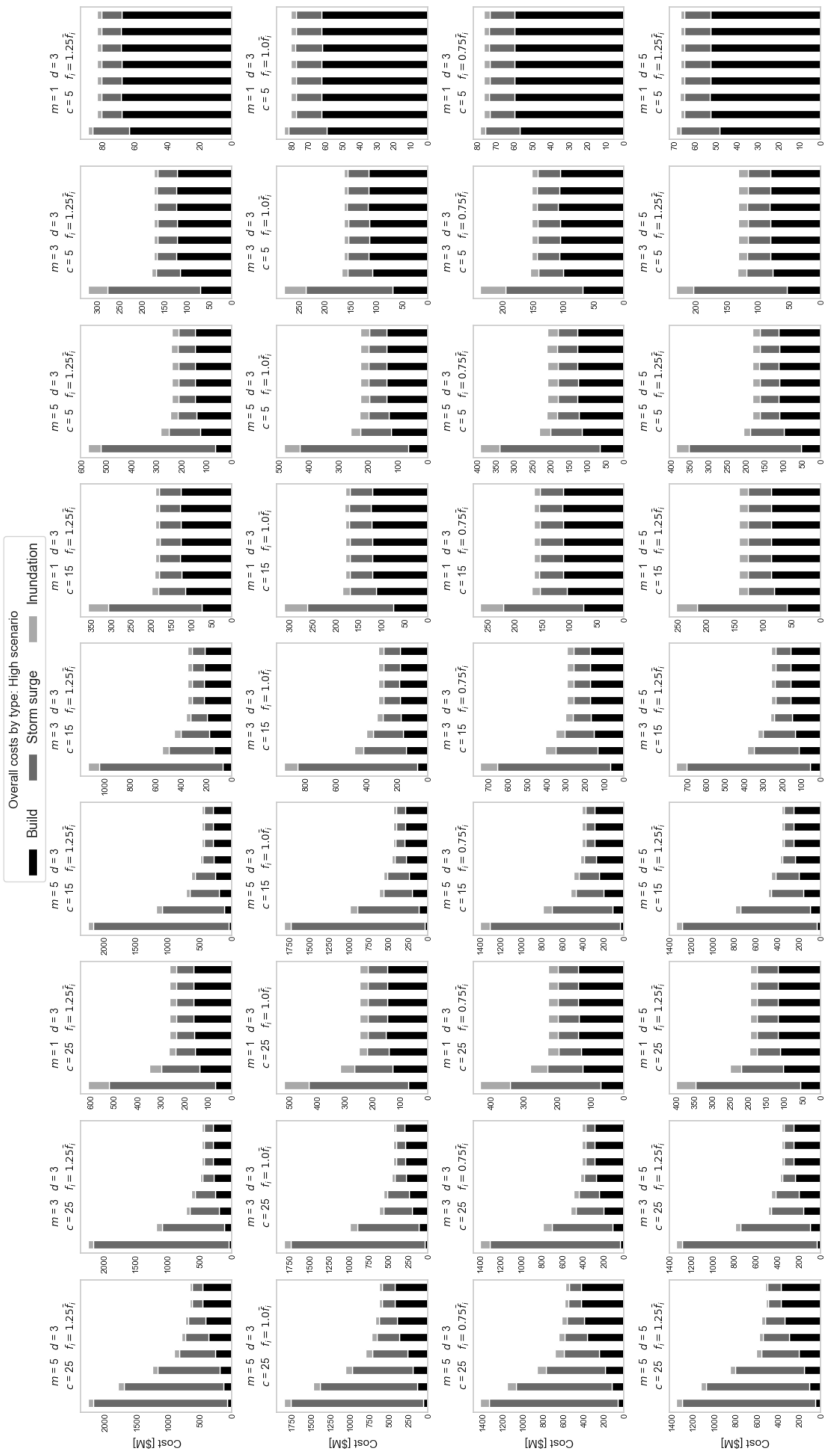


Figure 38: East Boston High scenario cost benefit curves by per-period budget for each of the 81 parameter combinations. (2/2)

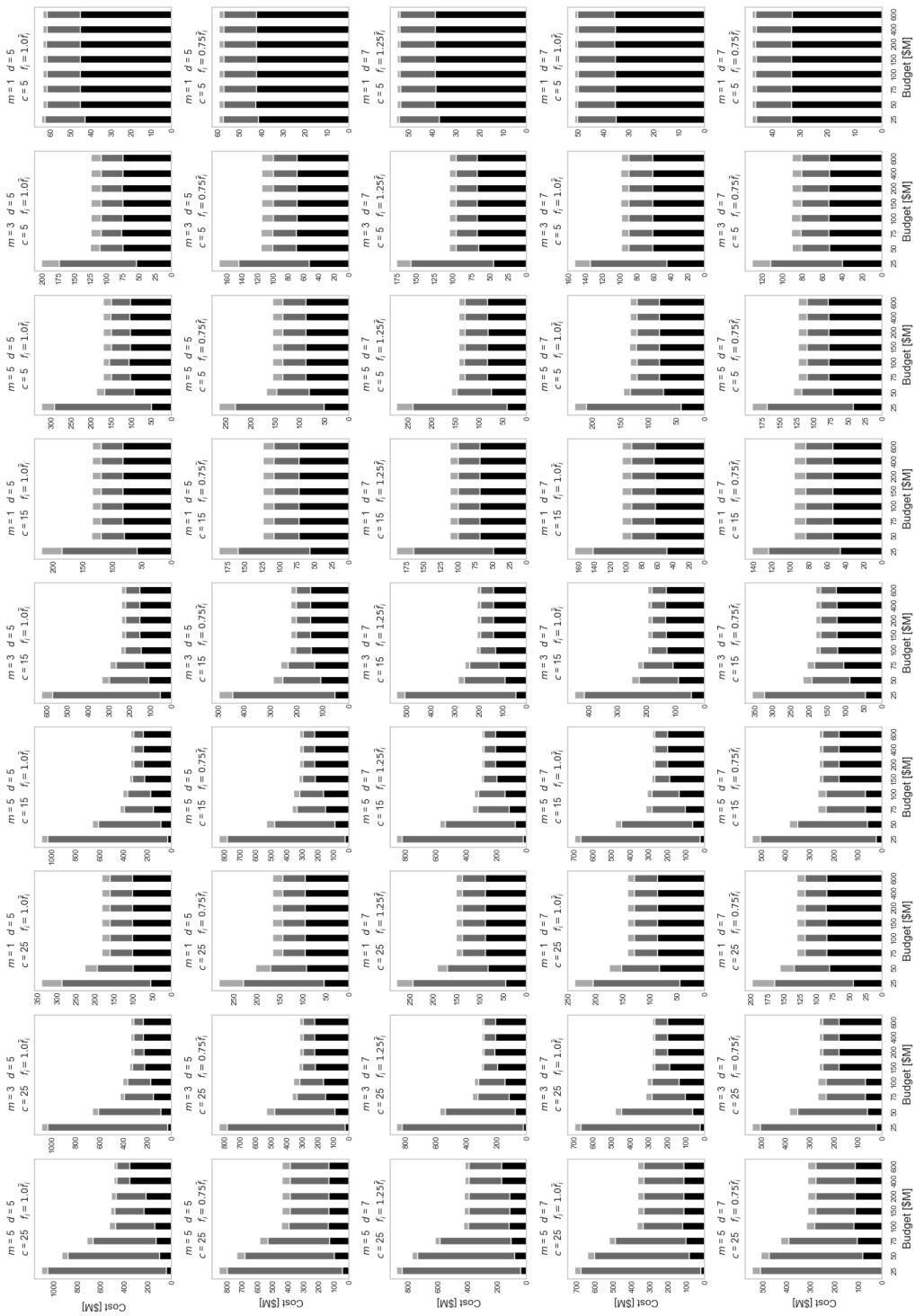


Figure 39: East Boston Expected-high scenario cost benefit curves by per-period budget for each of the 81 parameter combinations. (1/2)

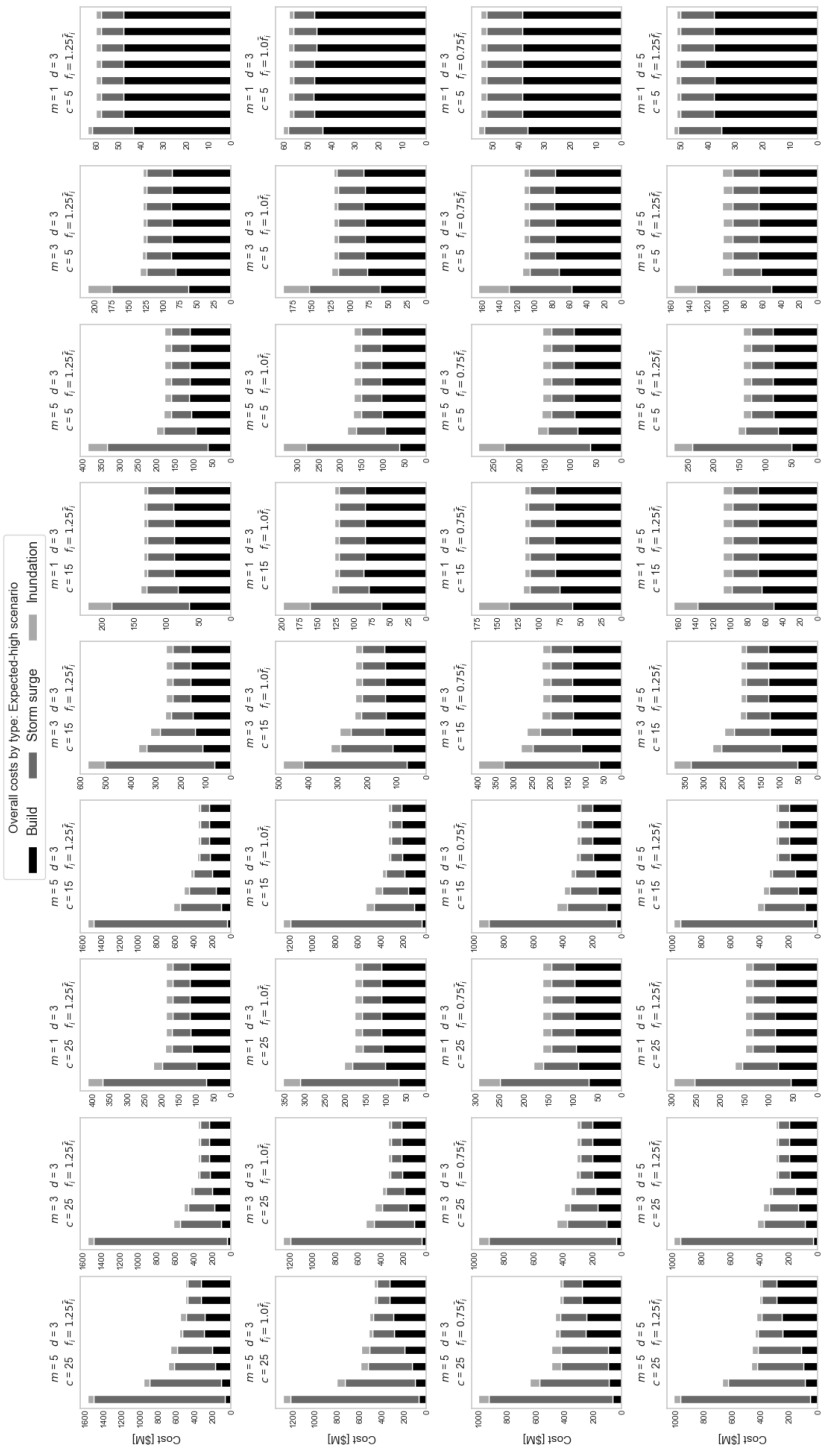


Figure 39: East Boston Expected-high scenario cost benefit curves by per-period budget for each of the 81 parameter combinations. (2/2)

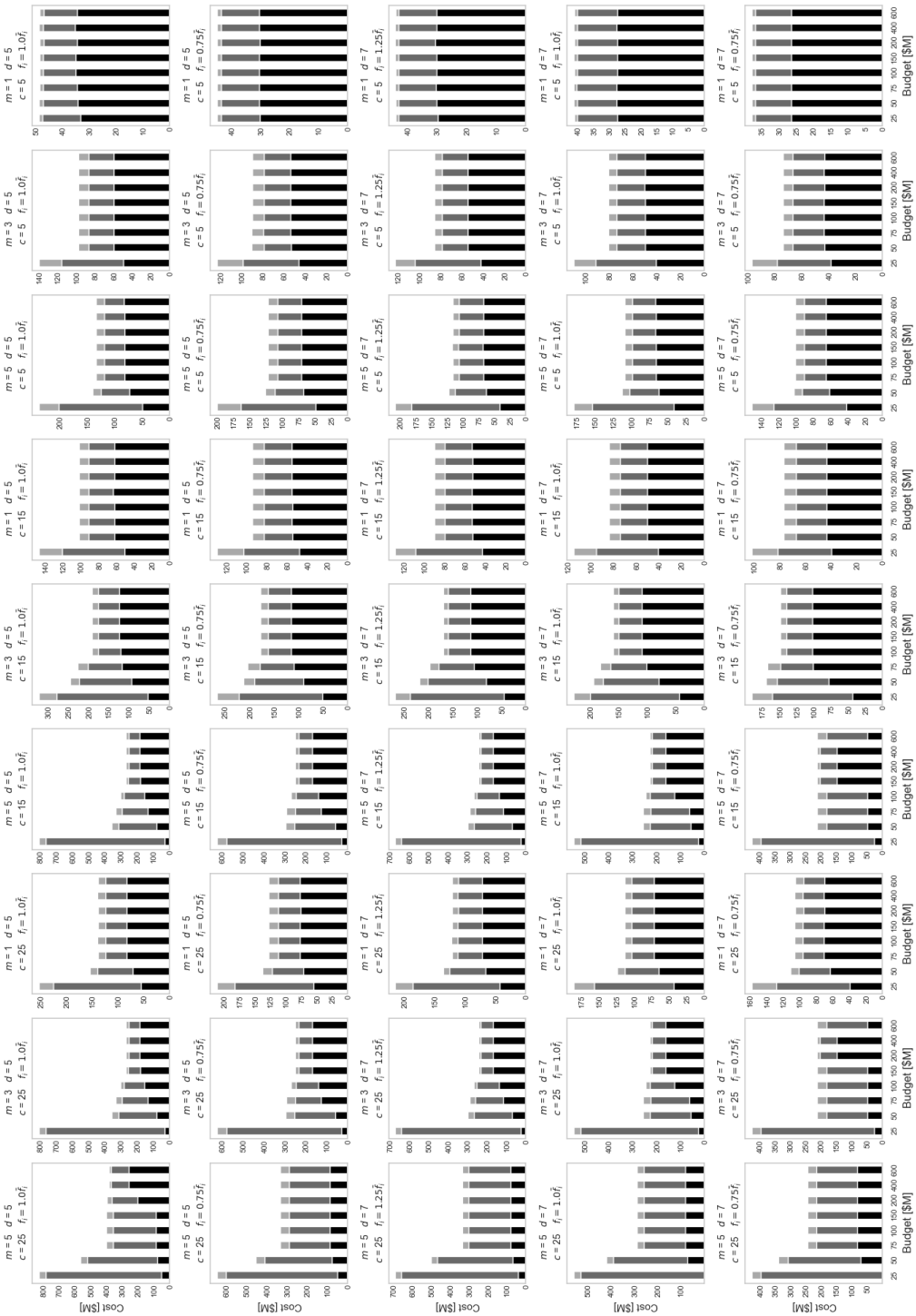


Figure 40: East Boston Expected-low scenario cost benefit curves by per-period budget for each of the 81 parameter combinations. (1/2)

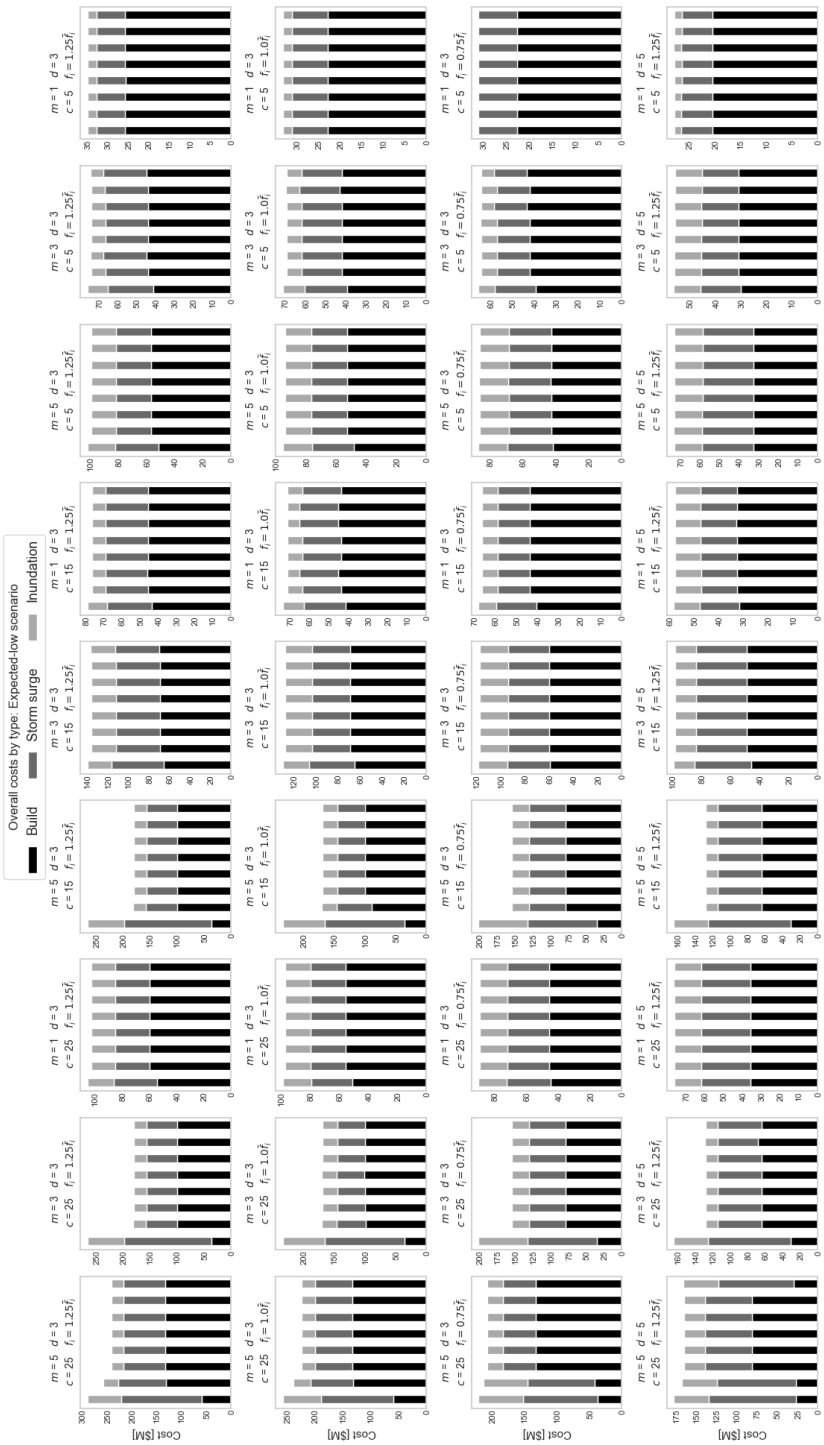


Figure 40: East Boston Expected-low scenario cost benefit curves by per-period budget for each of the 81 parameter combinations. (2/2)

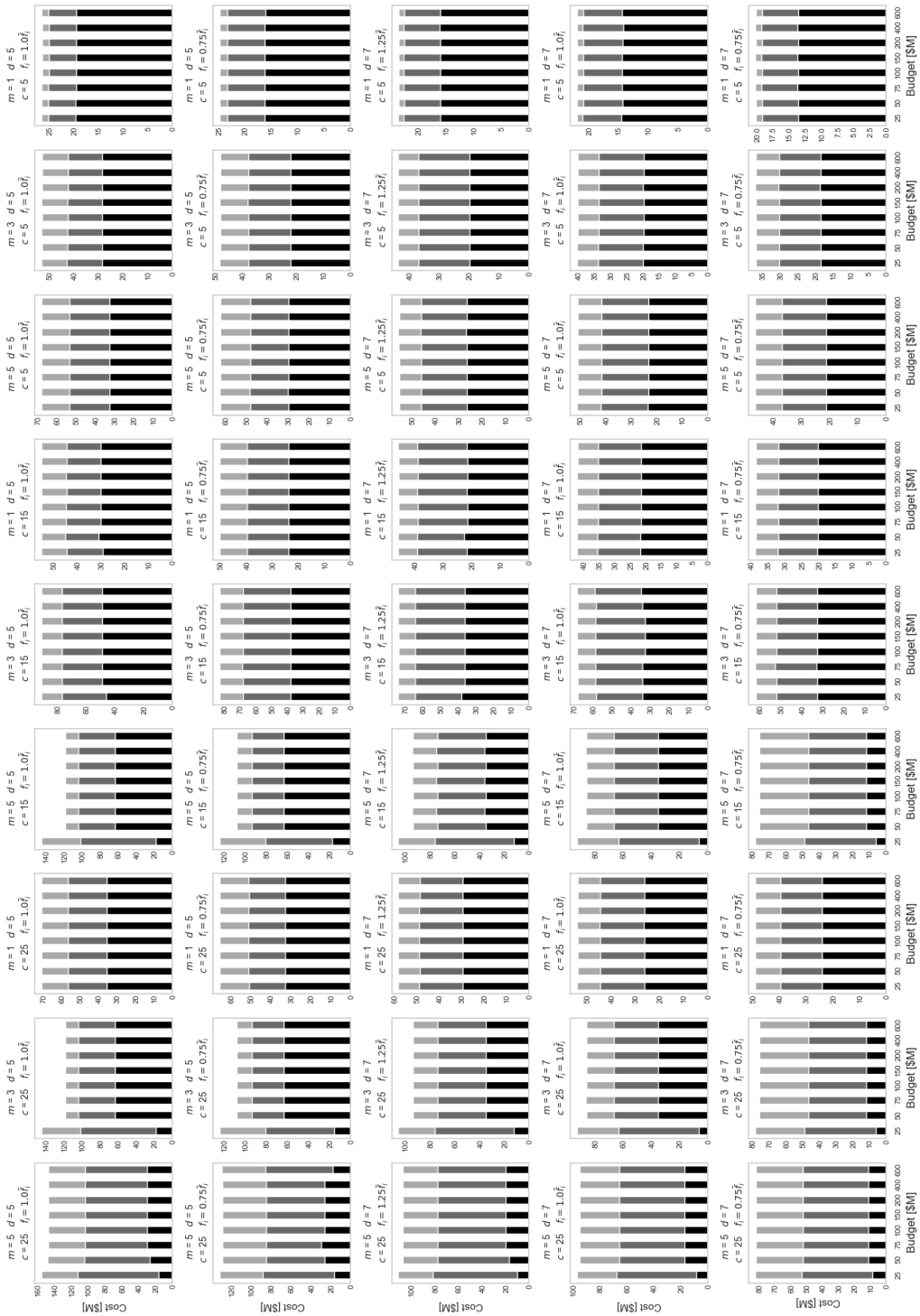


Figure 41: East Boston Optimistic scenario cost benefit curves by per-period budget for each of the 81 parameter combinations. (1/2)

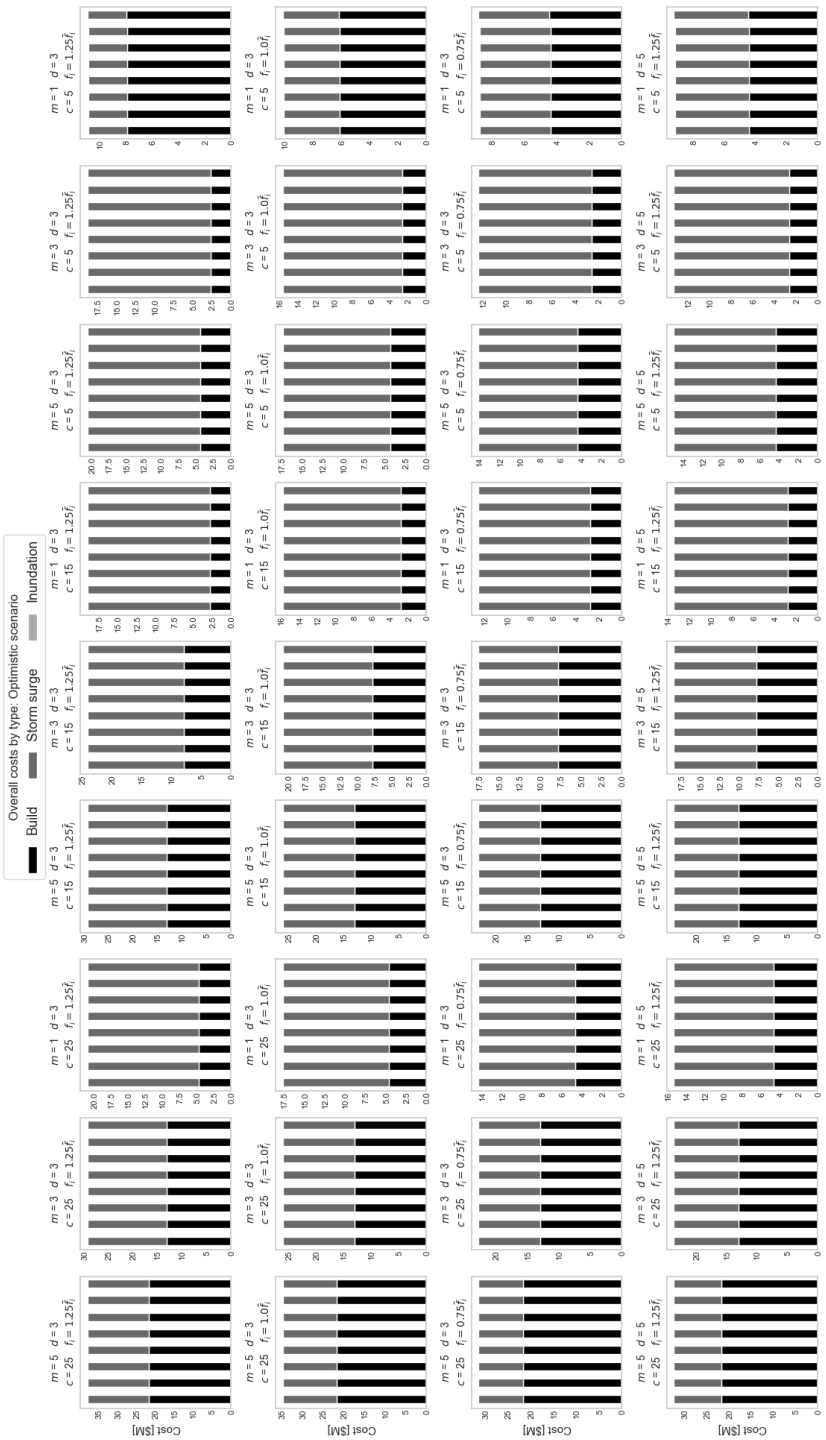


Figure 41: East Boston Optimistic scenario cost benefit curves by per-period budget for each of the 81 parameter combinations. (2/2)

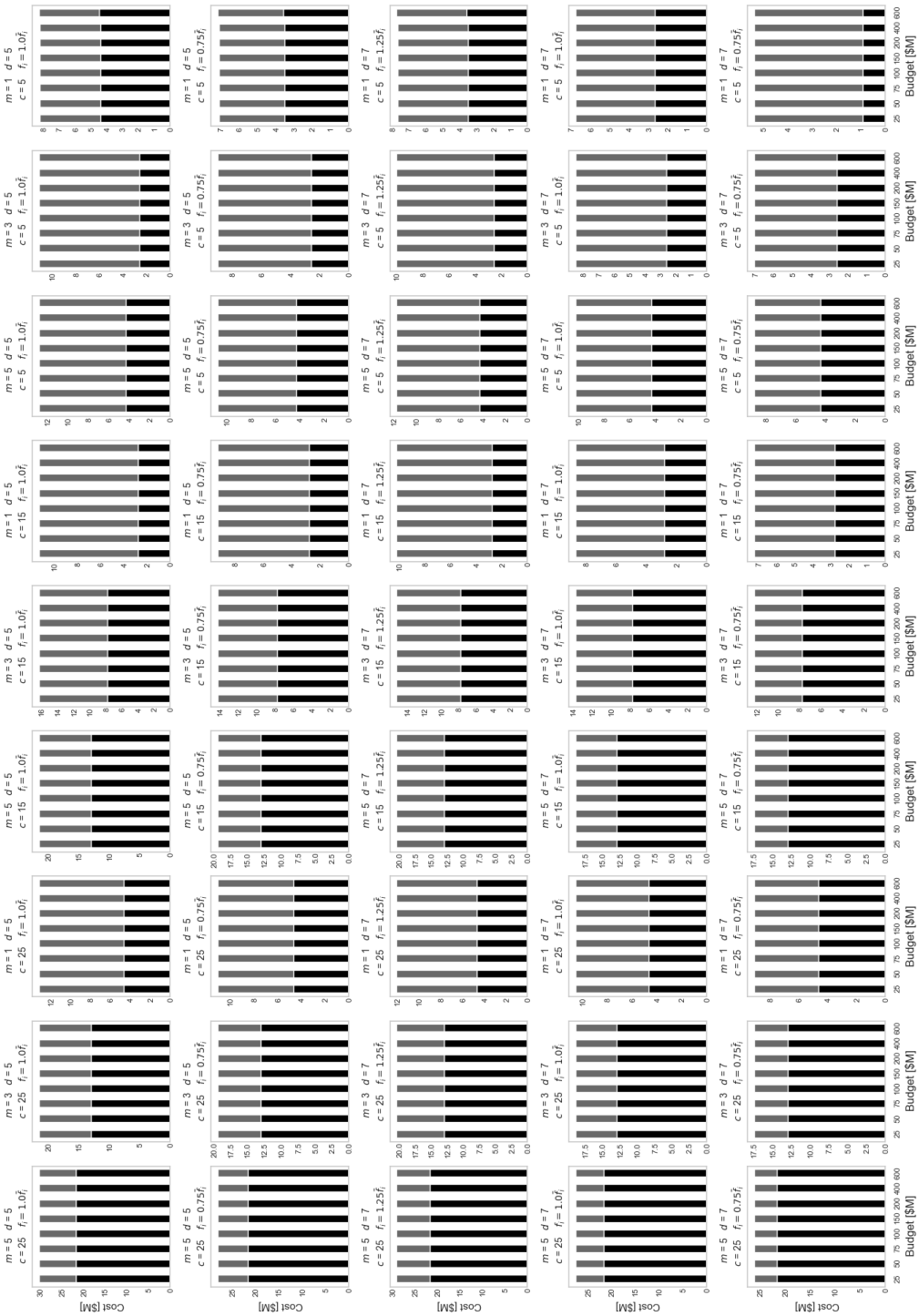


Figure 42: East Boston No-sea-level-rise scenario cost benefit curves by per-period budget for each of the 81 parameter combinations. (1/2)

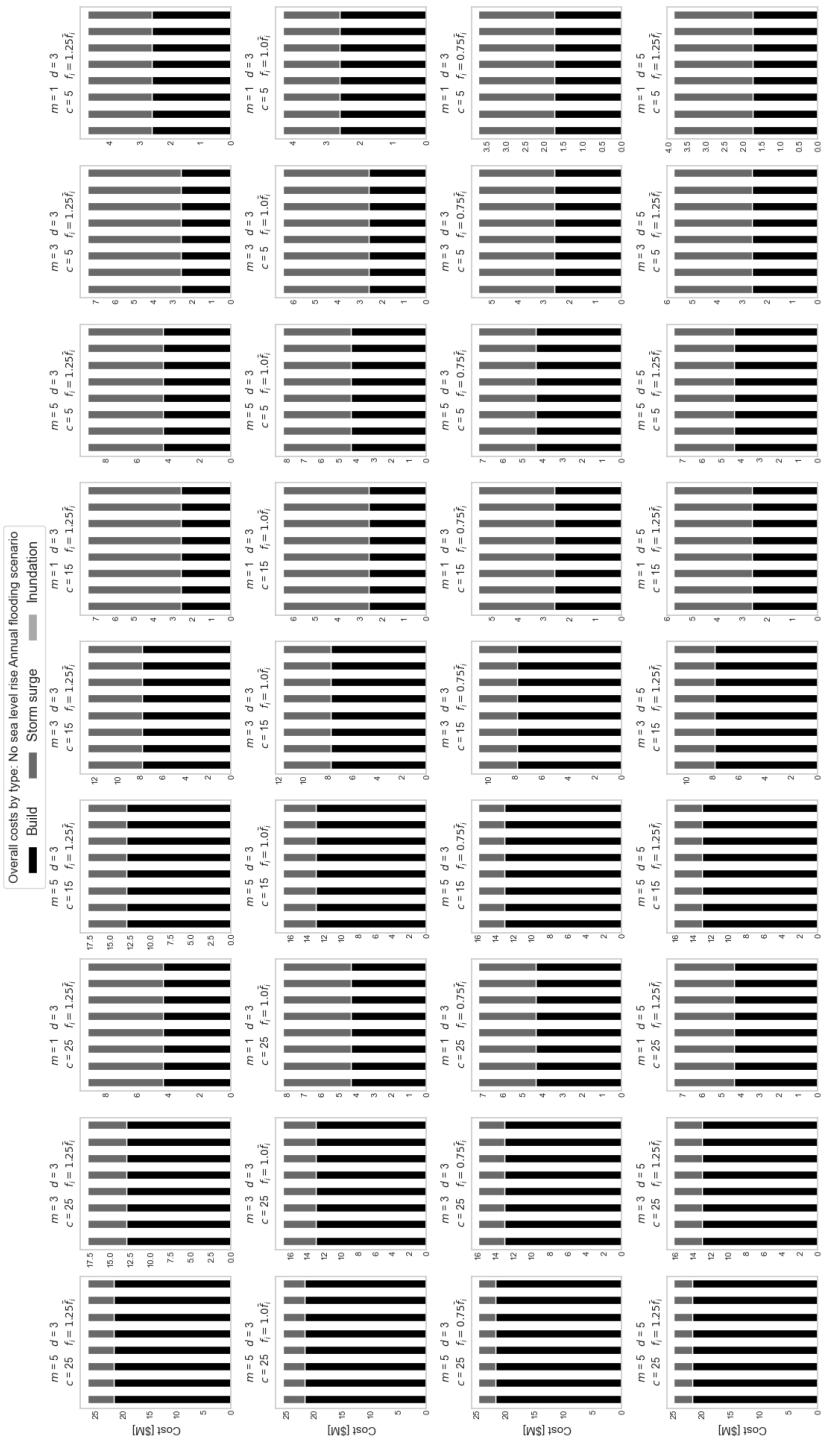
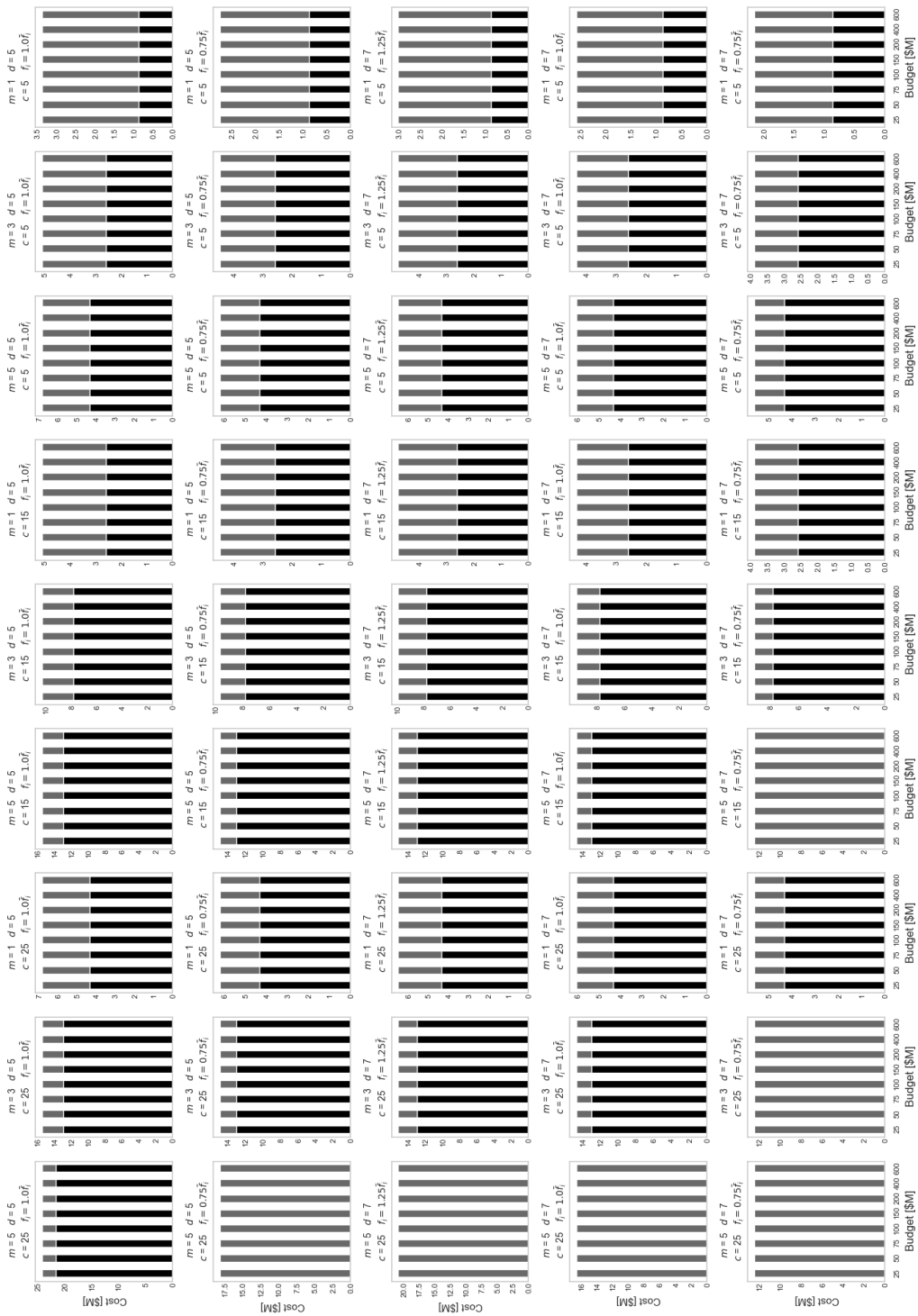


Figure 42: East Boston No-sea-level-rise scenario cost benefit curves by per-period budget for each of the 81 parameter combinations. (2/2)



APPENDIX D

RANDOM NETWORK CREATION AND EXPERIMENT DISCUSSION FOR RANDOM NETWORKS USED IN CHAPTER 3

In this section, we present the pseudocode in Algorithm 1 used to create the random networks for the analysis discussed in Section 3.5.3. We start with a square grid made up of 402 hex grids, randomly select sea nodes and Region of Interest (ROI) nodes, randomly assign land elevations h_i to ROI and the Surrounding Region (SR) nodes, and then randomly assign g_i and f_i to the ROI nodes. The result is a network used to run the same experiments as in the Boston case study.

Below, we present the cost benefit curves for the experiments on the 50 random networks. Figure 43 shows the combined average of all 50 networks with a cost benefit curve looking at the costs for each budgeted run for one of the 81 parameter combinations. We omit the zero budget cost curves to provide better visualization of the budget effect in each cost benefit curve. To better understand where the model starts being constrained by the budget, we added lower budgets for some model runs. These added per-period budget runs include \$5M and \$12.5M per period budgets for all instances of runs with m at 1m, and \$12.5M where m is at 3m, and c is at \$5M/km per m or \$15M/km per m. Similar to Figure 43, we show the cost benefit curves averaged across all 50 networks for each of the four scenarios discussed in Section 3.5.2 across the 81 parameter combinations and for each budget. At the top of each chart is the combination of parameters for that chart, with the values shown in Table 33. One note of interest in the random network scenario charts, the inundation costs are a fraction of the overall costs, so in many of the bar charts below the reader will notice that the costs are mostly made up of build and storm surge costs, with very small inundation contribution in the stacked bar charts.

Algorithm 1 Random Network Generation

Initialize *baseNetwork* of 402 hex-grid nodes (*i.e.*, (23×17)+11 extra nodes for the last row)
repeat 50 times
    randomNetwork $\leftarrow$ *baseNetwork*
    Randomly select up to 3 nodes at the boundary of the *baseNetwork* as sea nodes
    while Number of sea nodes < 100 **do**
        Randomly select a non-sea node from the *baseNetwork* that is adjacent to some sea node
        Add the selected node to the set of sea nodes
    end while
        ▷ A *Sea subcomponent* is a connected component composed of sea nodes.
    for each *Sea subcomponent* **do**
        if $|$*Sea subcomponent*$| < 4$ **then**
            while $|$*Sea subcomponent*$| < 4$ **do**
                Randomly select a non-sea node from *baseNetwork* adjacent to *Sea subcomponent*
                Add the selected node to the set of sea nodes and the *Sea subcomponent*
            end while
        end if
    end for
    Randomly select one non-sea node bordering the set of sea nodes
    Add the selected node to the Region of Interest (*ROI*) set
    while $|$*ROI*$| < 100$ **do**
        Randomly select a non-sea node from the *baseNetwork* that is adjacent to *ROI*
        Add the selected node to *ROI*
    end while
      ▷ A *SR subcomponent* is a connected component composed of Surrounding Region nodes
    for each *SR subcomponent* **do**
        if $|$*SR subcomponent*$| < 4$ **then**

Algorithm 1 Random Network Generation - Continued

```
            if all SR subcomponent nodes are adjacent to some ROI node then
                Re-label the SR subcomponent nodes as ROI nodes
            else
                Re-label the SR subcomponent nodes as sea nodes
            end if
        end if
    end for
    for all non-sea nodes do
        Randomly sample one h_i elevation from East Boston population of h_i
        Assign sampled h_i to the chosen node
    end for
    for all ROI nodes do
        randomly sample one f_i and g_i pair from East Boston population of f_i and g_i
        Assign sampled f_i and g_i pair to the chosen node
    end for
    save randomNetwork
until
```

Table 33: Parameter values used in sensitivity analysis for charts shown in Appendix D.

Parameter	High	Mid	Low
Discount rate (d) [%]	7	5	3
Minimum elevation increase (m) [meters]	5	3	1
Grid elevation cost (c) [$M/km per m]	25	15	5
Storm flood damage curve (f_i) [$M/m]	$1.25\bar{f}_i$	$\bar{f}_i$	$0.75\bar{f}_i$

Figure 43: Random network simulation expected cost benefit curves by per-period budget with expected costs averaged across all 50 networks for each of the 81 parameter combinations. (1/2)

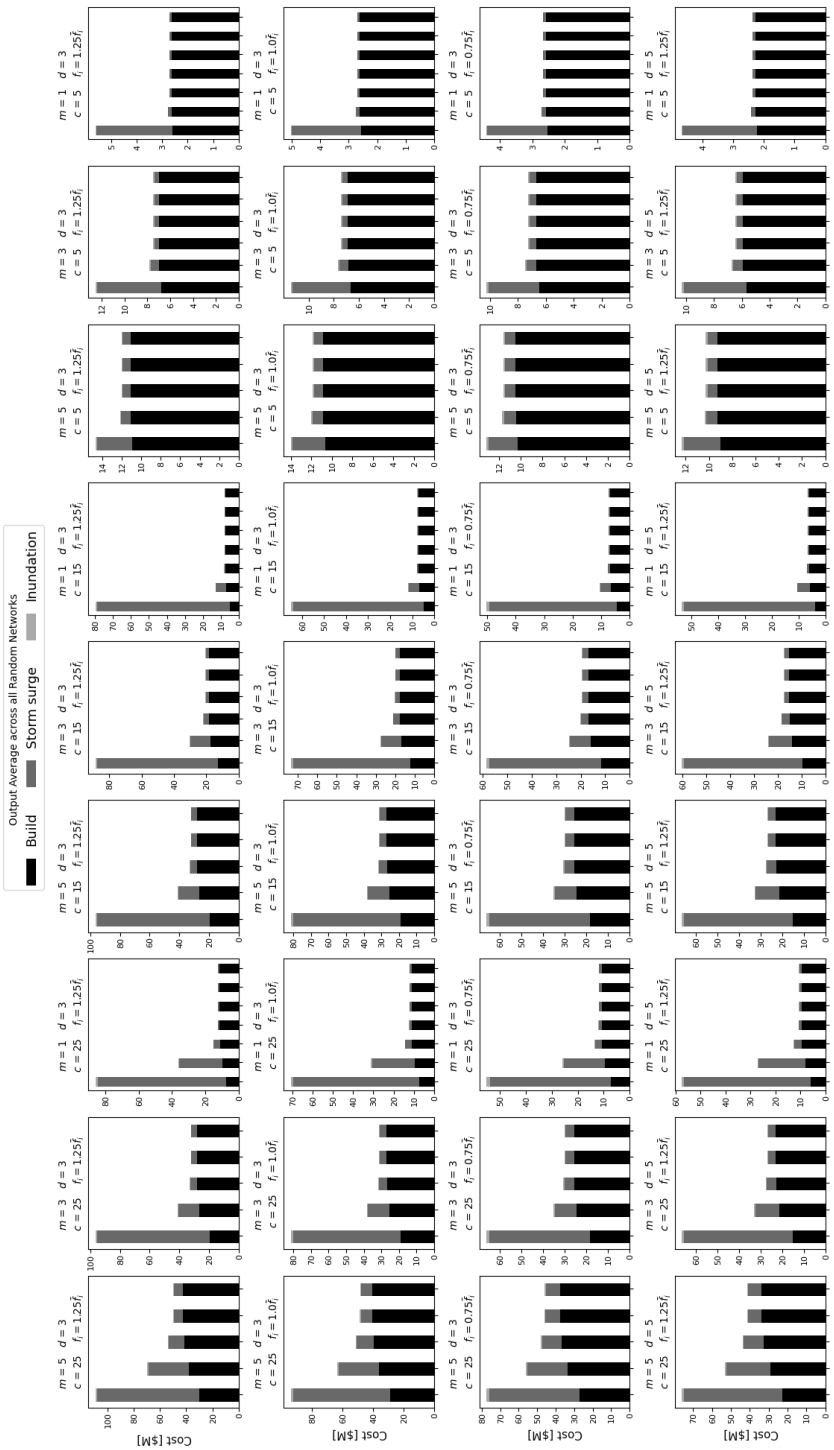


Figure 43: Random network simulation expected cost benefit curves by per-period budget with expected costs averaged across all 50 networks for each of the 81 parameter combinations. (2/2)

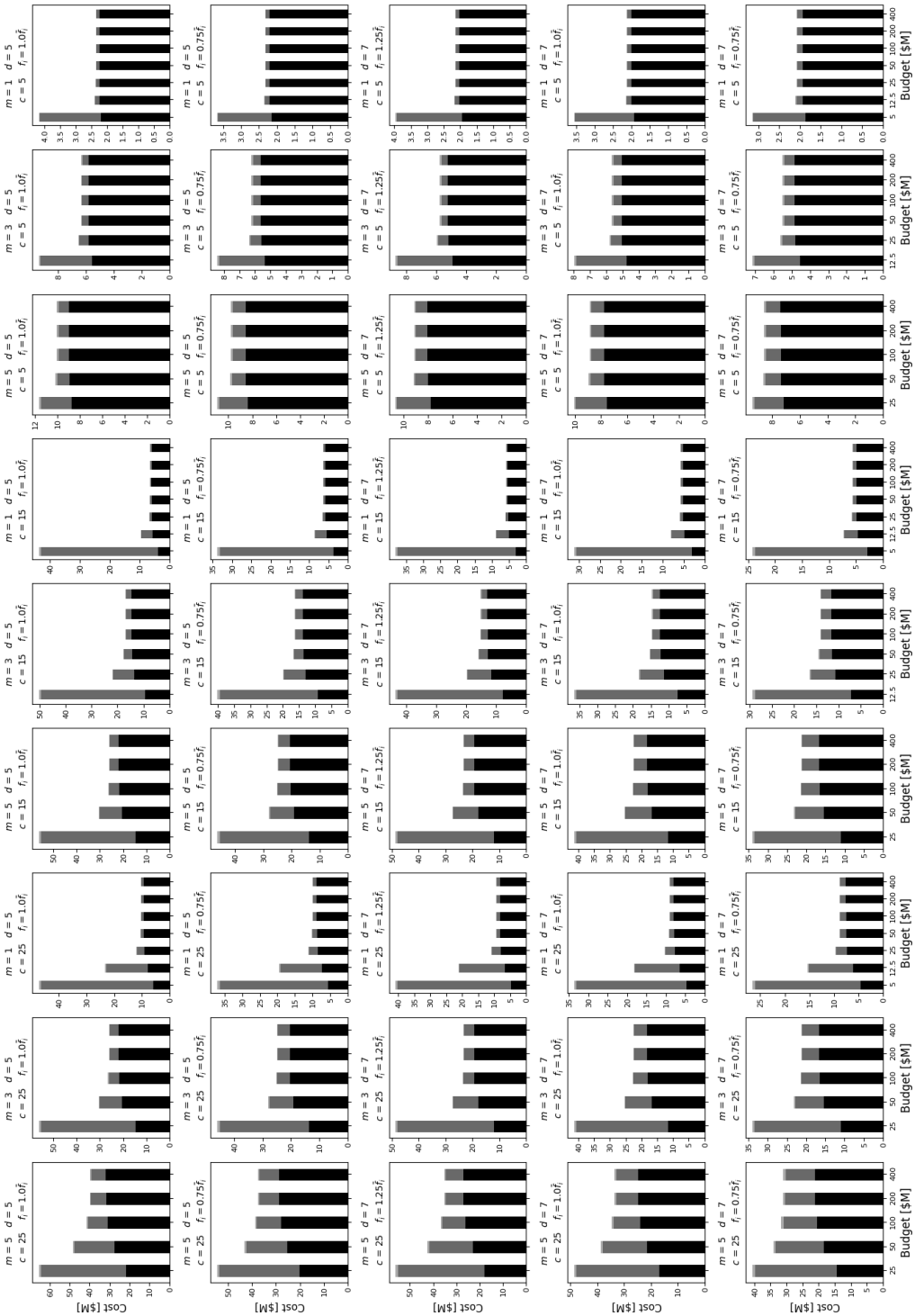


Figure 44: Random network High scenario cost benefit curves by per-period budget with scenario costs averaged across all 50 networks for each of the 81 parameter combinations. (1/2)

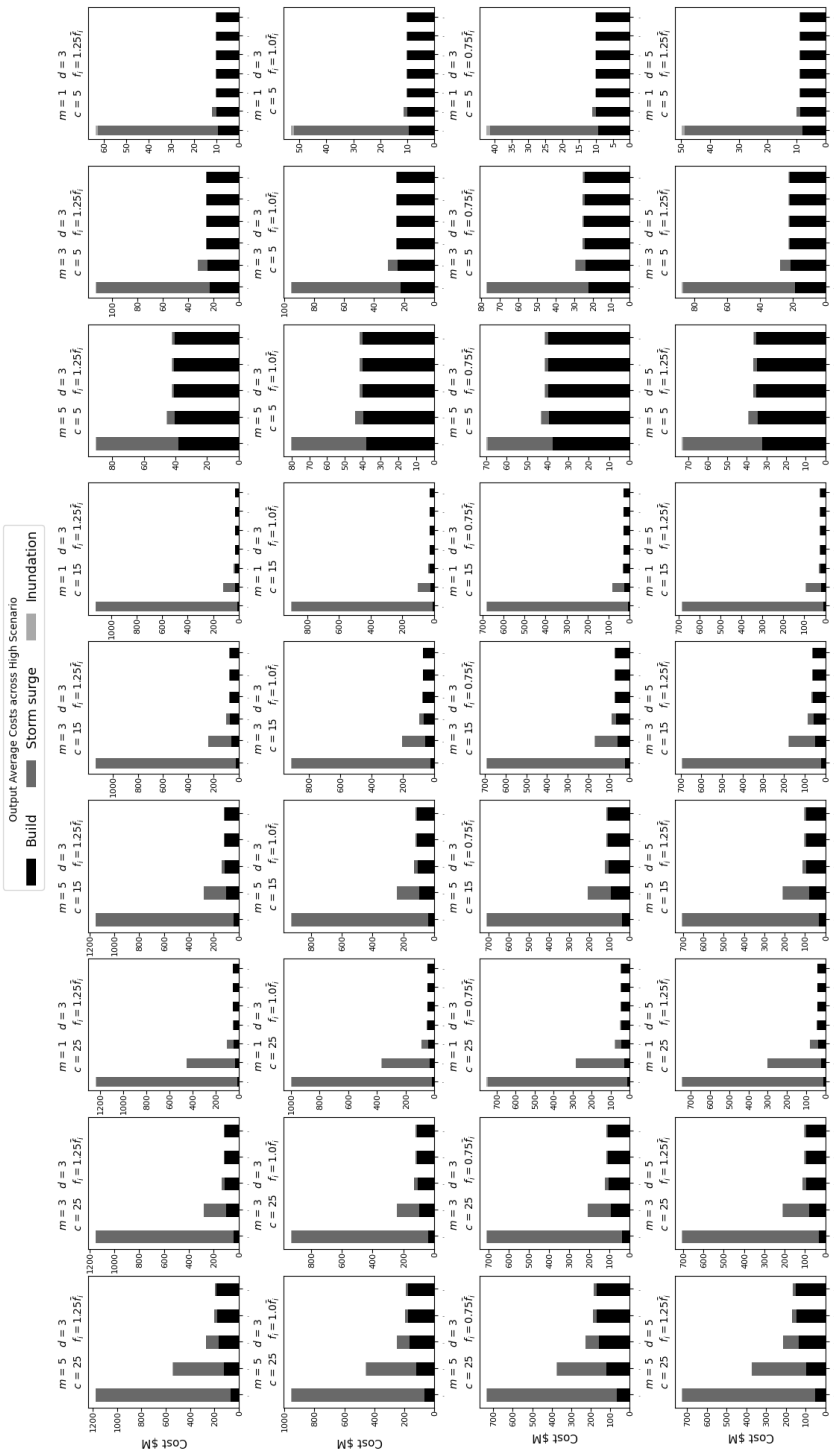


Figure 44: Random network High scenario cost benefit curves by per-period budget with scenario costs averaged across all 50 networks for each of the 81 parameter combinations. (2/2)

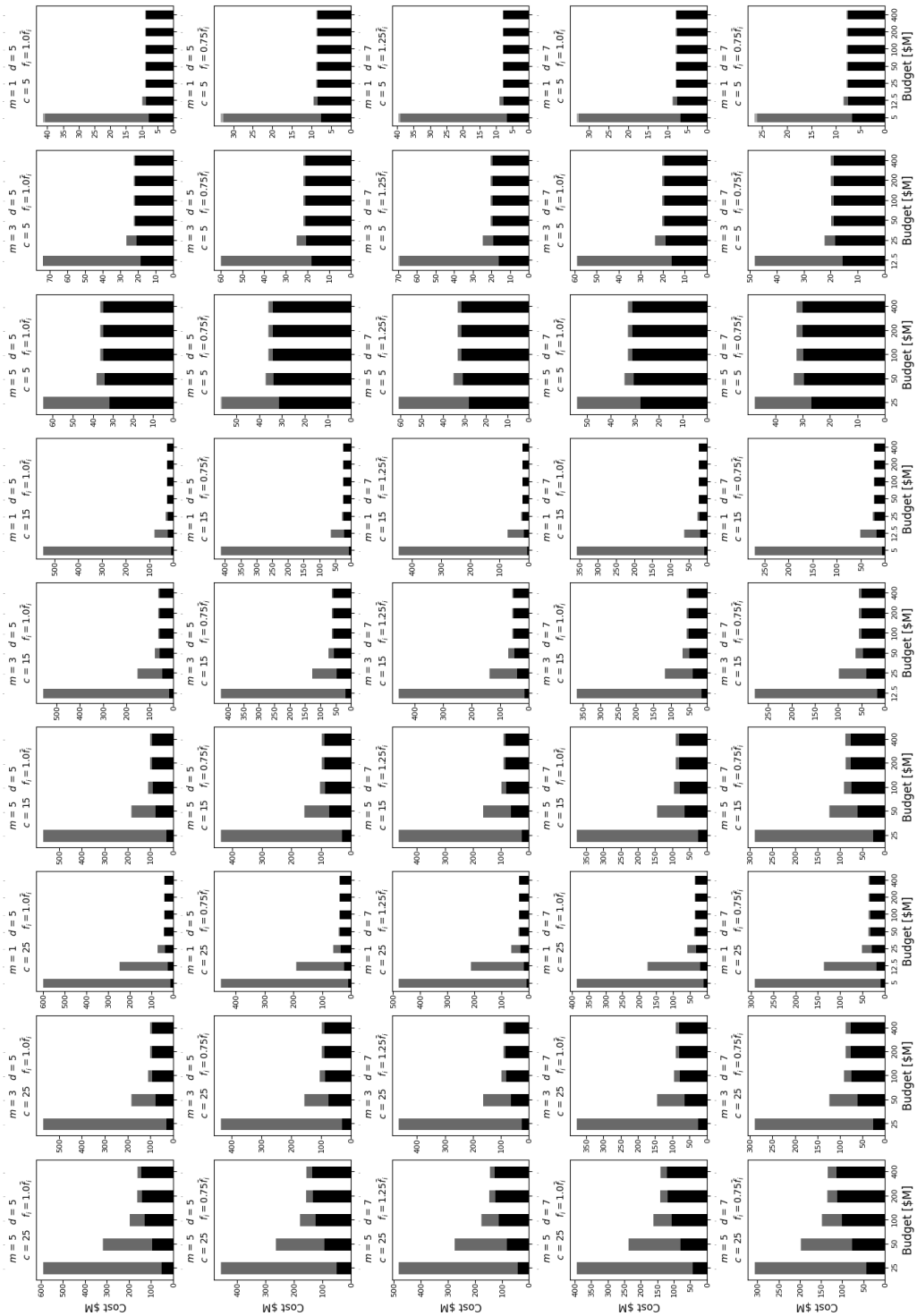


Figure 45: Random network Expected-high scenario cost benefit curves by per-period budget with scenario costs averaged across all 50 networks for each of the 81 parameter combinations. (1/2)

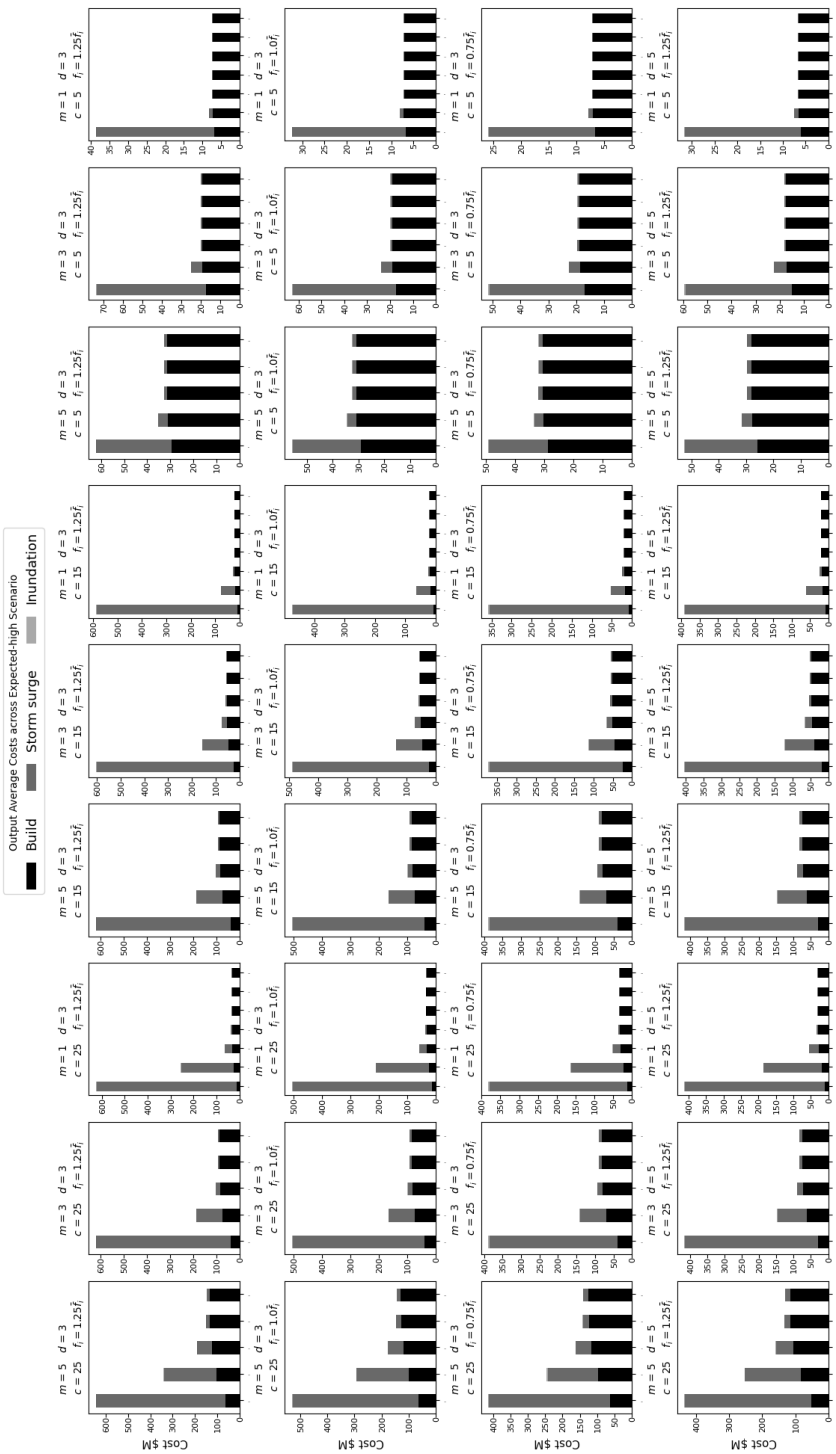


Figure 45: Random network Expected-high scenario cost benefit curves by per-period budget with scenario costs averaged across all 50 networks for each of the 81 parameter combinations. (2/2)

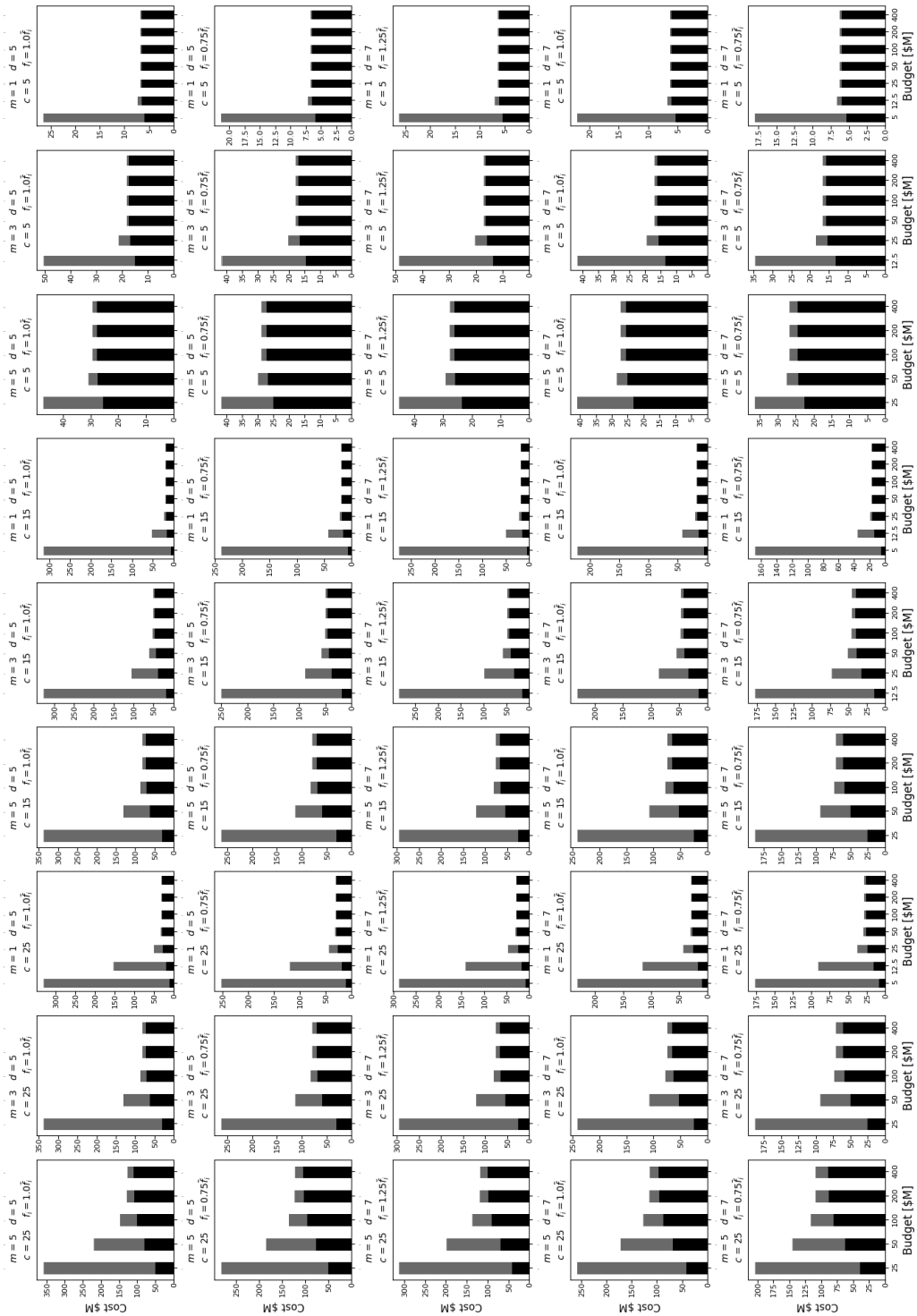


Figure 46: Random network Expected-low scenario cost benefit curves by per-period budget with scenario costs averaged across all 50 networks for each of the 81 parameter combinations. (1/2)

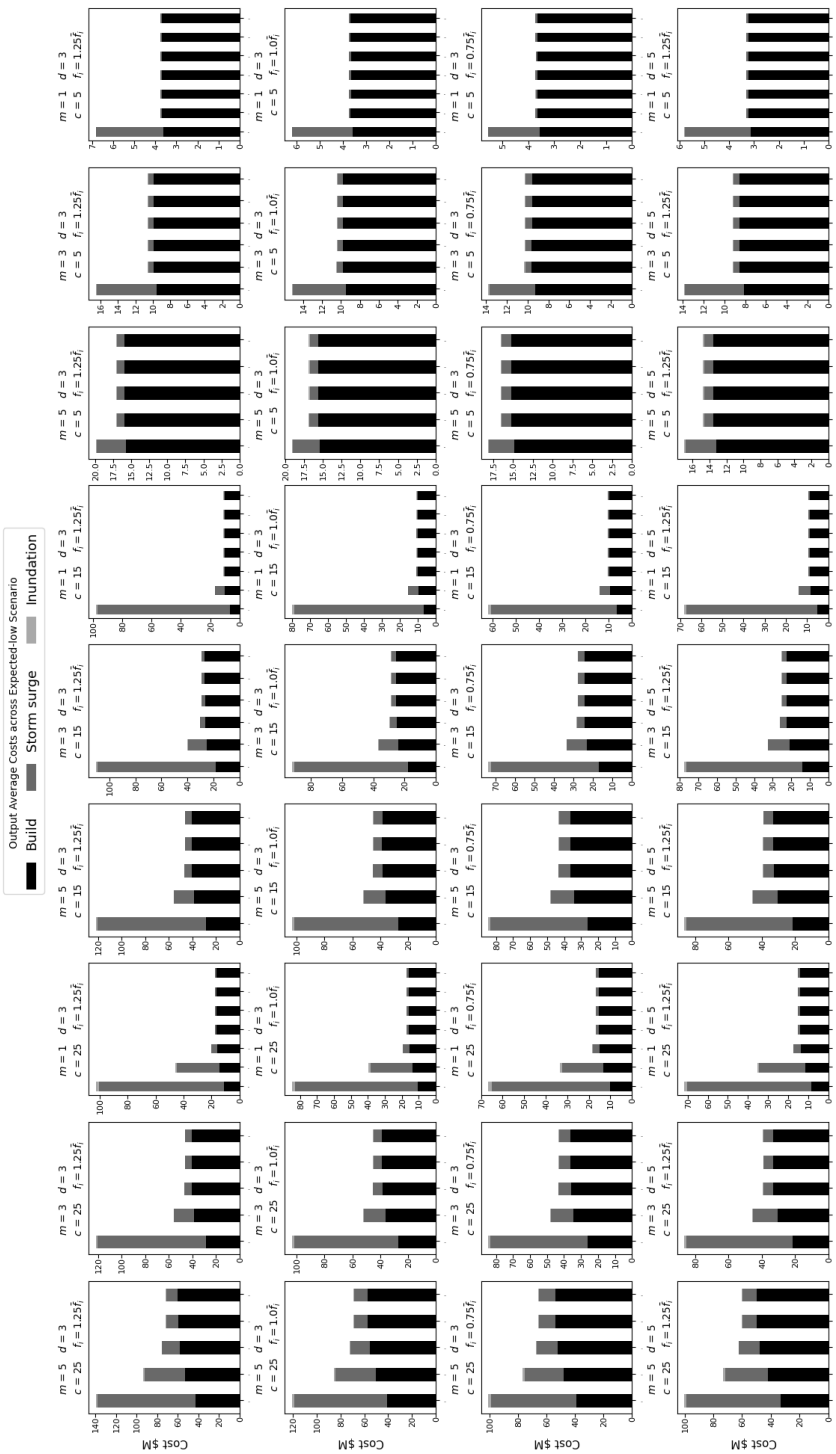


Figure 46: Random network Expected-low scenario cost benefit curves by per-period budget with scenario costs averaged across all 50 networks for each of the 81 parameter combinations. (2/2)

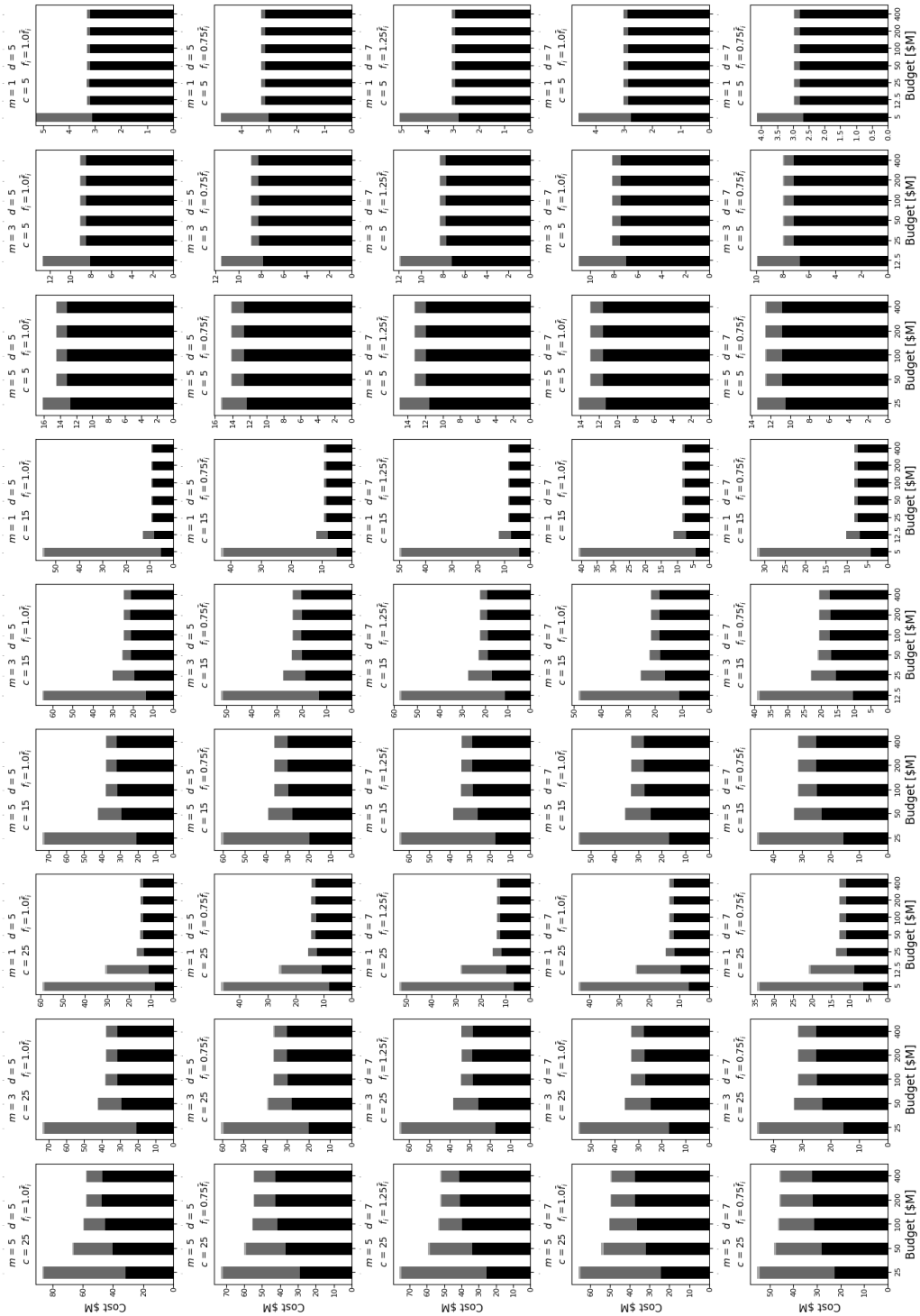


Figure 47: Random network Optimistic scenario cost benefit curves by per-period budget with scenario costs averaged across all 50 networks for each of the 81 parameter combinations. (1/2)

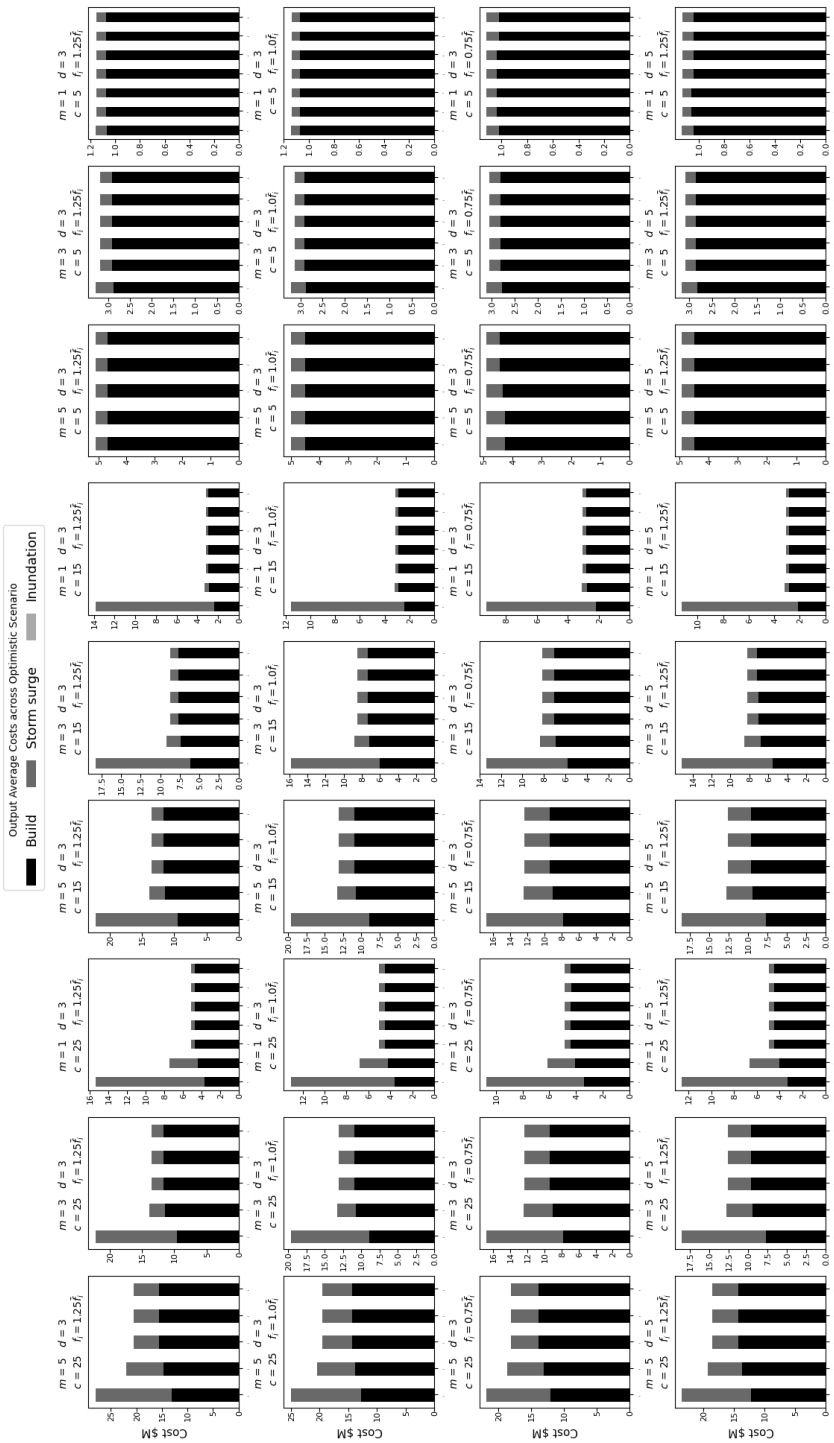
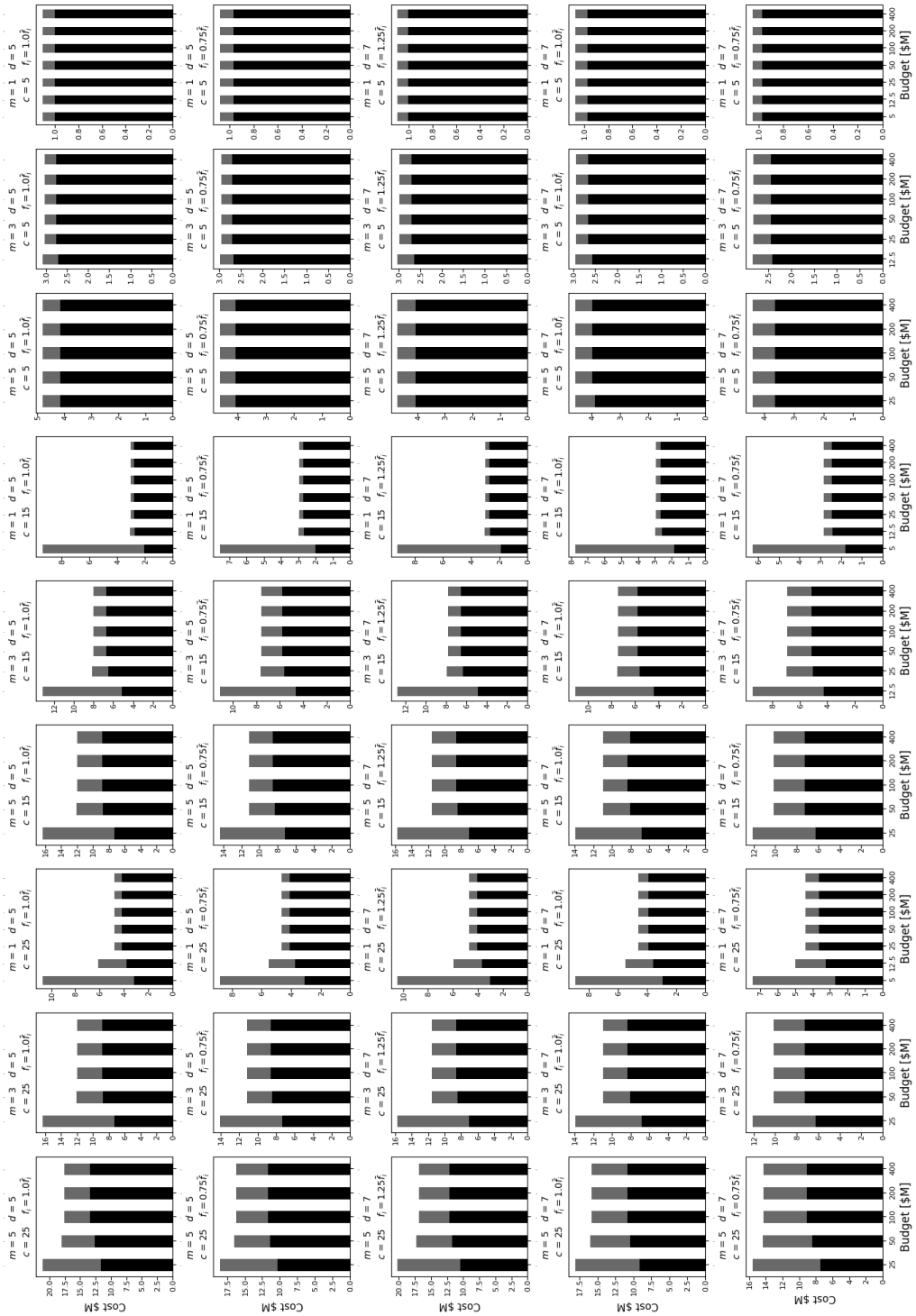


Figure 47: Random network Optimistic scenario cost benefit curves by per-period budget with scenario costs averaged across all 50 networks for each of the 81 parameter combinations. (2/2)



REFERENCE LIST

Abadie, Luis M. 2018. Sea Level Damage Risk with Probabilistic Weighting of IPCC Scenarios: An Application to Major Coastal Cities. *Journal of Cleaner Production* 175 (February): 582–598.

Abel, David. 2021. *Costs of floods will soar, study warns: Advocates call for plans to manage a retreat from low-lying coastal areas.* https://www.proquest.com/newspapers/costs-floods-will-soar-study-warns/docview/2491776772/se-2. Accessed: 2023-02-03.

Aerts, Jeroen C.J.H., W.J. Wouter Botzen, Kerry Emanuel, Ning Lin, Hans De Moel, and Erwann O. Michel-Kerjan. 2014. Evaluating Flood Resilience Strategies for Coastal Megacities. *Science* 344, no. 6183 (May): 473–475.

Akter, Shahriar, and Samuel Fosso Wamba. 2019. Big Data and Disaster Management: A Systematic Review and Agenda for Future Research. *Annals of Operations Research* 283, no. 1 (August): 939–959.

Aloisi, Jim. 2017. *Massport at 60: Shaping the Future Since 1956.* Boston: Massport.

Amazon Web Services, Inc. 2021. *Amazon EC2 Instances.* https://aws.amazon.com. Accessed: 2021-11-14.

———. 2023. *Amazon EC2 C5 Instances.* https://aws.amazon.com/ec2/instance-types/c5/. Accessed: 2023-02-15. Amazon Web Services.

Anguelov, Dragomir, Carole Dulong, Daniel Filip, Christian Frueh, Stéphane Lafon, Richard Lyon, Abhijit Ogale, Luc Vincent, and Josh Weaver. 2010. Google Street View: Capturing the World at Street Level. *Computer* 43 (6): 32–38.

Armal, Saman, Jeremy R. Porter, Brett Lingle, Ziyan Chu, Michael L. Marston, and Oliver E.J. Wing. 2020. Assessing Property Level Economic Impacts of Climate in the US, New Insights and Evidence from a Comprehensive Flood Risk Assessment Tool. *Climate* 8, no. 10 (October): 116.

Asmussen, Claus Boye, and Charles Møller. 2019. Smart Literature Review: A Practical Topic Modelling Approach to Exploratory Literature Review. *Journal of Big Data* 6, no. 93 (October): 1–18.

Ayyub, Bilal M. 2001. *Elicitation of Expert Opinions for Uncertainty and Risks.* Boca Raton, FL, USA: CRC Press.

Bagchi, Aniruddha, and Jomon Aliyas Paul. 2014. Optimal Allocation of Resources in Airport Security: Profiling vs. Screening. *Operations Research* 62, no. 2 (March): 219–233.

Baranes, Hannah E., Jonathan D. Woodruff, Stefan A. Talke, Robert E. Kopp, Richard D. Ray, and Robert M. DeConto. 2020. Tidally driven interannual variation in extreme sea level frequencies in the Gulf of Maine. *Journal of Geophysical Research: Oceans* 125 (10): e2020JC016291.

Baxter, Richard. 2021. *Energy Storage Financing: Project and Portfolio Valuation.* Technical report. Albuquerque, NM, USA: Sandia National Lab. (SNL-NM).

Berman, Oded, Arieh Gavious, and Mozart B.C. Menezes. 2012. Optimal Response Against Bioterror Attack on Airport Terminal. *European Journal of Operational Research* 219, no. 2 (June): 415–424.

Berrang-Ford, Lea, Tristan Pearce, and James D. Ford. 2015. Systematic Review Approaches for Climate Change Adaptation Research. *Regional Environmental Change* 15, no. 5 (February): 755–769.

Besiou, Maria, and Luk N. Van Wassenhove. 2020. Humanitarian Operations: A World of Opportunity for Relevant and Impactful Research. *Manufacturing & Service Operations Management* 22, no. 1 (November): 135–145.

Biesbroek, Robbert, Lea Berrang-Ford, James D. Ford, Andrew Tanabe, Stephanie E. Austin, and Alexandra Lesnikowski. 2018. Data, Concepts and Methods for Large-n Comparative Climate Change Adaptation Policy Research: A Systematic Literature Review. *Wiley Interdisciplinary Reviews: Climate Change* 9, no. 6 (July): 548–562.

Birch, Colin P.D., Sander P. Oom, and Jonathan A. Beecham. 2007. Rectangular and Hexagonal Grids Used for Observation, Experiment and Simulation in Ecology. *Ecological Modelling* 206, nos. 3-4 (August): 347–359.

Bird, Steven, Ewan Klein, and Edward Loper. 2009. *Natural Language Processing with Python: Analyzing Text with the Natural Language Toolkit.* USA: O'Reilly Media, Inc.

Blei, David, and John Lafferty. 2006. *Correlated Topic Models.* In *Advances in Neural Information Processing Systems,* edited by Y. Weiss, B. Schölkopf, and J. Platt, 18:147–154. Cambridge, MA, USA: MIT Press.

Blei, David M. 2012. Probabilistic Topic Models. *Communications of the ACM* 55, no. 4 (April): 77–84.

Blei, David M., Andrew Y. Ng, and Michael I. Jordan. 2003. Latent Dirichlet Allocation. *Journal of Machine Learning Research* 3 (January): 993–1022.

Bohr, Jeremiah. 2020. Reporting on Climate Change: A Computational Analysis of US Newspapers and Sources of Bias, 1997–2017. *Global Environmental Change* 61 (March): 102038.

Boussalis, Constantine, and Travis G. Coan. 2016. Text-Mining the Signals of Climate Change Doubt. *Global Environmental Change* 36 (January): 89–100.

Brede, Markus, and Bert J.M. de Vries. 2013. The Energy Transition in a Climate-Constrained World: Regional vs. Global Optimization. *Environmental Modelling & Software* 44 (June): 44–61.

Brereton, Pearl, Barbara A. Kitchenham, David Budgen, Mark Turner, and Mohamed Khalil. 2007. Lessons from Applying the Systematic Literature Review Process within the Software Engineering Domain. *Journal of Systems and Software* 80, no. 4 (April): 571–583.

Campbell, Ann Melissa, and Philip C. Jones. 2011. Prepositioning supplies in preparation for disasters. *European Journal of Operational Research* 209, no. 2 (March): 156–165.

Capehart, Barney L., Wayne C. Turner, and William J. Kennedy. 2006. *Guide to Energy Management System.* 5th ed. New York: The Fairmont Press, Inc.

Carter, Jeremy G. 2018. Urban Climate Change Adaptation: Exploring the Implications of Future Land Cover Scenarios. *Cities* 77 (July): 73–80.

Chakravarty, Amiya K. 2018. Humanitarian response to hurricane disasters: Coordinating flood-risk mitigation with fundraising and relief operations. *Naval Research Logistics* 65, no. 3 (September): 275–288.

Chandramowli, Shankar, Frank A. Felder, Nancy Mantell, Will Irving, and Joseph Seneca. 2016. LP-CEM: A Modeling Tool for Power Systems Planning Incorporating Climate Change Effects and Macroeconomic Trends for New Jersey, United States. *Energy Strategy Reviews* 11 (June): 1–18.

Chang, Jonathan, Sean Gerrish, Chong Wang, Jordan Boyd-Graber, and David Blei. 2009. *Reading Tea Leaves: How Humans Interpret Topic Models.* In *Advances in Neural Information Processing Systems,* 22:288–296. Red Hook, NY, USA: Curran Associates, Inc.

Chen, Lichun, and Elise Miller-Hooks. 2012. Optimal Team Deployment in Urban Search and Rescue. *Transportation Research Part B: Methodological* 46, no. 8 (September): 984–999.

Chen, Wendy Y. 2015. The Role of Urban Green Infrastructure in Offsetting Carbon Emissions in 35 Major Chinese Cities: A Nationwide Estimate. *Cities* 44 (April): 112–120.

Chesney, Marc, Pierre Lasserre, and Bruno Troja. 2017. Mitigating Global Warming: A Real Options Approach. *Annals of Operations Research* 255, no. 1 (September): 465–506.

Chu, Xu, Ihab F. Ilyas, Sanjay Krishnan, and Jiannan Wang. 2016. *Data Cleaning: Overview and Emerging Challenges.* In *Proceedings of the 2016 International Conference on Management of Data,* 2201–2206. SIGMOD '16. San Francisco, California, USA: Association for Computing Machinery.

City of Boston. 2016. *Boston Tax Parcel Data.* https://bostonopendata-boston.opendata.arcgis.com/datasets/f3d274161b4a47aa9acf48d0d04cd5d4. Accessed: 2020-03-22.

———. 2016. *Climate Ready Boston: Final Report.* https://www.boston.gov/departments/environment/preparing-climate-change. Accessed: 2018-08-18.

———. 2020. *Property Assessment FY2016 Data Key.* https://data.boston.gov/dataset/property-assessment/resource/dbdc1bd8-60af-4913-a788-5f91cb68541b. Accessed: 2020-04-04.

Clarivate Analytics. 2020. *Web of Science Core Collection.* https://clarivate.com/products/web-of-science/. Accessed: 2020-12-10.

Climate Change, Intergovernmental Panel on. 2019. *IPCC Special Report on the Ocean and Cryosphere in a Changing Climate.* Cambridge, UK and New York, NY, USA: Cambridge University Press.

Commonwealth of Massachusetts. 2017. *MassGIS Data: Lidar Terrain Data.* https://www.mass.gov/info-details/massgis-data-lidar-terrain-data. Accessed: 2020-03-16.

Contreras, Sergio, Noel Finnerty, Raymond Sterling, Daniel Coakley, and Marcus M. Keane. 2017. *A Systematic Decision Support Framework and Prioritization Method for Energy Projects in Industrial Organisations.* In *10th International Conference on Sustainable Energy and Environmental Protection Energy Management and Policies.* Bled, Slovenia: University of Maribor Press.

EM-DAT. 2020. *The International Disaster Database.* http://www.emdat.be/. Accessed: 2022-04-04.

Davis, Lauren B., Funda Samanlioglu, Xiuli Qu, and Sarah Root. 2013. Inventory planning and coordination in disaster relief efforts. *International Journal of Production Economics* 141 (2): 561–673.

Dayeen, Fazle Rabbi, Abhinav S. Sharma, and Sybil Derrible. 2020. A Text Mining Analysis of the Climate Change Literature in Industrial Ecology. *Journal of Industrial Ecology* 24, no. 2 (March): 276–284.

De Sousa, Luís, Mike Gibson, Albert Chen, Dragan Savic, and João Leitão. 2017. *Exploring the Advantages of Hexagonal Raster for Flood Modelling Using Cellular Automata.* In *14th IWA/IAHR International Conference on Urban Drainage (ICUD).* Prague, Czech Republic: IWA/IAHR.

Debortoli, Stefan, Oliver Müller, Iris Junglas, and Jan Vom Brocke. 2016. Text Mining for Information Systems Researchers: An Annotated Topic Modeling Tutorial. *Communications of the Association for Information Systems* 39, no. 1 (July): 7.

Deerwester, Scott, Susan T. Dumais, George W. Furnas, Thomas K. Landauer, and Richard Harshman. 1990. Indexing by Latent Semantic Analysis. *Journal of the American Society for Information Science* 41, no. 6 (September): 391–407.

Desjardins, Jeff. 2019. *How much data is generated each day.* https://www.weforum.org/agenda/2019/04/how-much-data-is-generated-each-day-cf4bddf29f/. Accessed: 2023-03-15. World Economic Forum.

Devillers, Rodolphe, Yvan Bédard, Robert Jeansoulin, and Bernard Moulin. 2007. Towards spatial data quality information analysis tools for experts assessing the fitness for use of spatial data. *International Journal of Geographical Information Science* 21, no. 3 (March): 261–282.

DiMaggio, Paul, Manish Nag, and David Blei. 2013. Exploiting Affinities between Topic Modeling and the Sociological Perspective on Culture: Application to Newspaper Coverage of US Government Arts Funding. *Poetics* 41, no. 6 (December): 570–606.

Douglas, Ellen, and Paul Kirshen. 2022. *Climate change impacts and projections for the Greater Boston area.* Technical report. University of Massachusetts Boston.

Douglas, Ellen M., Paul H. Kirshen, Michael Paolisso, Chris Watson, Jack Wiggin, Ashley Enrici, and Matthias Ruth. 2012. Coastal Flooding, Climate Change and Environmental Justice: Identifying Obstacles and Incentives for Adaptation in Two Metropolitan Boston Massachusetts Communities. *Mitigation and Adaptation Strategies for Global Change* 17, no. 5 (December): 537–562.

Douglas, Ellen M., Kenneth M. Reardon, and Matthias C. Täger. 2018. Participatory Action Research as a Means of Achieving Ecological Wisdom Within Climate Change Resiliency Planning. *Journal of Urban Management* 7, no. 3 (December): 152–160.

Dowlatabadi, Hadi. 1998. Sensitivity of Climate Change Mitigation Estimates to Assumptions about Technical Change. *Energy Economics* 20, nos. 5-6 (December): 473–493.

Eftekhar, Mahyar, Hongmin Li, Luk N. Van Wassenhove, and Scott Webster. 2016. The Role of Media Exposure on Coordination in the Humanitarian Setting. *Production and Operations Management* 26, no. 5 (November): 802–816.

Eftekhar, Mahyar, Jeannette Song, and Scott Webster. 2022. Prepositioning and Local Purchasing for Emergency Operations Under Budget, Demand, and Supply Uncertainty. *Manufacturing & Service Operations Management* 24 (1): 315–332.

Eijgenraam, Carel, Ruud Brekelmans, Dick den Hertog, and Kees Roos. 2017. Optimal Strategies for Flood Prevention. *Management Science* 63, no. 5 (April): 1644–1656.

Eisensee, Thomas, and David Strömberg. 2007. News Droughts, News Floods, and US Disaster Relief. *The Quarterly Journal of Economics* 122 (2): 693–728.

Elgesem, Dag, Lubos Steskal, and Nicholas Diakopoulos. 2014. Structure and Content of the Discourse on Climate Change in the Blogosphere: The Big Picture. *Environmental Communication* 9, no. 2 (December): 169–188.

Eschenbach, Ted G., and Lisa S. McKeague. 1989. Exposition on Using Graphs for Sensitivity Analysis. *The Engineering Economist* 34, no. 4 (March): 315–333.

Esty, Daniel C., and Andrew S. Winston. 2008. *Green to Gold: How Smart Companies Use Environmental Strategy to Innovate, Create Value, and Build Competitive Advantage.* USA: John Wiley & Sons.

Evans, James R. 2016. *Business Analytics : Methods, Models, and Decisions.* 2nd ed. Boston: Pearson.

Ewing, Ben, and Erin Baker. 2009. Development of a Green Building Decision Support Tool: A Collaborative Process. *Decision Analysis* 6, no. 3 (July): 172–185.

Faghmous, James H., and Vipin Kumar. 2014. A Big Data Guide to Understanding Climate Change: The Case for Theory-Guided Data Science. *Big Data* 2, no. 3 (September): 155–163.

Fang, Debin, Haixia Yang, Baojun Gao, and Xiaojun Li. 2018. Discovering Research Topics from Library Electronic References using Latent Dirichlet Allocation. *Library Hi Tech* 36, no. 3 (February): 400–410.

Fernandez, Ricardo, and John Watterson. 2012. *End-User GHG Emissions from Energy. Reallocation of Emissions from Energy Industries to End Users 2005-2010.* Technical report. Denmark: European Environment Agency.

Finnerty, Noel, Raymond Sterling, Sergio Contreras, Daniel Coakley, and Marcus M. Keane. 2018. Defining Corporate Energy Policy and Strategy to Achieve Carbon Emissions Reduction Targets via Energy Management in Non-Energy Intensive Multi-Site Manufacturing Organisations. *Energy* 151 (May): 913–929.

Fleiss, Joseph L. 1971. Measuring Nominal Scale Agreement among Many Raters. *Psychological Bulletin* 76 (5): 378–382.

Frostell, Björn M., Rajib Sinha, Getachew Assefa, and Lars E. Olsson. 2015. Modeling Both Direct and Indirect Environmental Load of Purchase Decisions: A Web-Based Tool Addressing Household Metabolism. *Environmental Modelling & Software* 71 (September): 138–147.

Galbreth, Michael R., and Bikram Ghosh. 2013. Competition and Sustainability: The Impact of Consumer Awareness. *Decision Sciences* 44, no. 1 (December): 127–159.

Galindo, Gina, and Rajan Batta. 2013a. Prepositioning of supplies in preparation for a hurricane under potential destruction of prepositioned supplies. *Socio-Economic Planning Sciences* 47, no. 1 (March): 20–37.

———. 2013b. Review of Recent Developments in OR/MS Research in Disaster Operations Management. *European Journal of Operational Research* 230, no. 2 (October): 201–211.

Garner, Andra J., Jeremy L. Weiss, Adam Parris, Robert E. Kopp, Radley M. Horton, Jonathan T. Overpeck, and Benjamin P. Horton. 2018. Evolution of 21st Century Sea Level Rise Projections. *Earth's Future* 6, no. 11 (October): 1603–1615.

Geest, Kees van der, and Romy van den Berg. 2021. Slow-Onset Events: A Review of the Evidence from the IPCC Special Reports on Land, Oceans and Cryosphere. *Current Opinion in Environmental Sustainability* 50 (June): 109–120.

Gesch, Dean B. 2018. Best Practices for Elevation-Based Assessments of Sea-Level Rise and Coastal Flooding Exposure. *Frontiers in Earth Science* 6 (December): 230.

Gomes, Carla, Xiaoli Fern, Daniel Fink, Doug Fisher, Alexander Flecker, Daniel Freund, Angela Fuller, et al. 2019. Computational Sustainability: Computing for a Better World and a Sustainable Future. *Communications of the ACM* 62, no. 9 (September): 56–65.

Google. 2013. *Google's Green PPAs: What, How, and Why.* http://www.google.com/green/pdfs/renewable-energy.pdf. Accessed: 2020-09-07.

Greene, Derek, Derek O'Callaghan, and Pádraig Cunningham. 2014. *How Many Topics? Stability Analysis for Topic Models.* In *Machine Learning and Knowledge Discovery in Databases,* 498–513. Berlin, Heidelberg: Springer.

Grimmer, Justin, and Brandon M. Stewart. 2013. Text as Data: The Promise and Pitfalls of Automatic Content Analysis Methods for Political Texts. *Political Analysis* 21, no. 3 (July): 267–297.

Guikema, Seth, and Mark Milke. 1999. Quantitative Decision Tools for Conservation Programme Planning: Practice, Theory and Potential. *Environmental Conservation* 26, no. 3 (September): 179–189.

Guise, Amy, Jason Engle, Lynn Bocamazo, Kelly Burks-Copes, and Julie Rosati. 2015. *North Atlantic Coast Comprehensive Study: Resilient Adaptation to Increasing Risk.* Technical report. National Planning Center for Coastal Storm Risk Management, U.S. Army Corps of Engineers.

Gupta, Sushil, Martin K. Starr, Reza Zanjirani Farahani, and Niki Matinrad. 2016. Disaster Management from a POM Perspective: Mapping a New Domain. *Production and Operations Management* 25, no. 10 (June): 1611–1637.

Haasnoot, Marjolijn, Jan H. Kwakkel, Warren E. Walker, and Judith Ter Maat. 2013. Dynamic Adaptive Policy Pathways: A Method for Crafting Robust Decisions for a Deeply Uncertain World. *Global Environmental Change* 23, no. 2 (April): 485–498.

Hall, David, Dan Jurafsky, and Christopher D. Manning. 2008. *Studying the History of Ideas Using Topic Models.* In *Proceedings of the 2008 Conference on Empirical Methods in Natural Language Processing,* 363–371. Honolulu, Hawaii: Association for Computational Linguistics.

Hallegatte, Stéphane, Colin Green, Robert J. Nicholls, and Jan Corfee-Morlot. 2013. Future Flood Losses in Major Coastal Cities. *Nature Climate Change* 3 (August): 802–806.

Hallegatte, Stéphane, Ankur Shah, Casey Brown, Robert Lempert, and Stuart Gill. 2012. Investment Decision Making Under Deep Uncertainty – Application to Climate Change. *World Bank Policy Research Working Paper,* no. 6193 (April).

Hassani, Hossein, Xu Huang, and Emmanuel Silva. 2019. Big Data and Climate Change. *Big Data and Cognitive Computing* 3, no. 1 (February): 12.

Haunschild, Robin, Lutz Bornmann, and Werner Marx. 2016. Climate Change Research in View of Bibliometrics. *PLOS One* (San Francisco, CA USA) 11, no. 7 (July): e0160393.

He, Bing, Ying Ding, Jie Tang, Vignesh Reguramalingam, and Johan Bollen. 2013. Mining Diversity Subgraph in Multidisciplinary Scientific Collaboration Networks: A Meso Perspective. *Journal of Informetrics* 7, no. 1 (January): 117–128.

Hecht, Jory S., and Paul H. Kirshen. 2019. Minimizing Urban Floodplain Management Regrets under Deeply Uncertain Climate Change. *Journal of Water Resources Planning and Management* 145, no. 2 (February): 04018096.

Hecht, Sean B. 2008. Climate Change and the Transformation of Risk: Insurance Matters. *UCLA Law Review* 55 (July): 1559–1620.

Helicke, Nurcan Atalan. 2019. Markets and Collective Action: A Case Study of Traditional Wheat Varieties in Turkey. *Journal of Economy Culture and Society,* no. 59 (July): 13–30.

Hoffman, Matthew, Francis Bach, and David Blei. 2010. *Online Learning for Latent Dirichlet Allocation.* In *Advances in Neural Information Processing Systems,* 23:856–864. Red Hook, NY, USA: Curran Associates, Inc.

Hsiang, Solomon, Robert Kopp, Amir Jina, James Rising, Michael Delgado, Shashank Mohan, Daniel J. Rasmussen, et al. 2017. Estimating Economic Damage from Climate Change in the United States. *Science* 356, no. 6345 (June): 1362–1369.

Hyperreality. 2016. *American-British-English-Translator.* https://github.com/hyperreality/American-British-English-Translator. Accessed: 2020-07-29.

Ibáñez-Forés, Valeria, Maria D. Bovea, and Victoria Pérez-Belis. 2014. A Holistic Review of Applied Methodologies for Assessing and Selecting the Optimal Technological Alternative from a Sustainability Perspective. *Journal of Cleaner Production* 70 (May): 259–281.

Ignatow, Gabe, and Rada F. Mihalcea. 2017. *An Introduction to Text Mining: Research Design, Data Collection, and Analysis.* 1st ed. Los Angeles, CA, USA: SAGE Publications, Inc.

Intergovernmental Panel on Climate Change. 2014. *Climate Change 2014: Mitigation of Climate Change: Working Group III Contribution to the IPCC Fifth Assessment Report.* New York, NY, USA: Cambridge University Press.

International Energy Agency. 2021. "Industry." International Energy Agency. Accessed July 2, 2023. https://www.iea.org/topics/industry.

Inwood, Hilary, and Alysse Kennedy. 2020. Conceptualising Art Education as Environmental Activism in Preservice Teacher Education. *International Journal of Art & Design Education* 39, no. 3 (June): 585–599.

Jawid, Asadullah, and Menusch Khadjavi. 2019. Adaptation to Climate Change in Afghanistan: Evidence on the Impact of External Interventions. *Economic Analysis and Policy* 64 (December): 64–82.

Jelodar, Hamed, Yongli Wang, Chi Yuan, Xia Feng, Xiahui Jiang, Yanchao Li, and Liang Zhao. 2019. Latent Dirichlet Allocation (LDA) and Topic Modeling: Models, Applications, A Survey. *Multimedia Tools and Applications* 78 (11): 15169–15211.

Jonkman, Sebastiaan N., Marten M. Hillen, Robert J. Nicholls, Wim Kanning, and Mathijs van Ledden. 2013. Costs of adapting coastal defences to sea-level rise—new estimates and their implications. *Journal of Coastal Research* 29 (5): 1212–1226.

Jonkman, Sebastiaan N., Matthijs Kok, Mathijs Van Ledden, and Johannes K. Vrijling. 2009. Risk-based design of flood defence systems: a preliminary analysis of the optimal protection level for the New Orleans metropolitan area. *Journal of Flood Risk Management* 2, no. 3 (September): 170–181.

Kandakoglu, Ahmet, Anissa Frini, and Sarah Ben Amor. 2019. Multicriteria Decision Making for Sustainable Development: A Systematic Review. *Journal of Multi-Criteria Decision Analysis* 26, nos. 5-6 (November): 202–251.

Karp, Richard M. 1972. "Reducibility among Combinatorial Problems." In *Complexity of Computer Computations,* edited by Raymond E. Miller, James W. Thatcher, and Jean D. Bohlinger, 85–103. Boston, MA: Springer.

Keegan, Natalie, Rawle King, Nicole Carter, and Megan Stubbs. 2011. *Locally operated levees: Issues and federal programs.* Technical report. Washington D.C.: Congressional Research Service.

Keeney, Ralph L., Howard Raiffa, and Richard F. Meyer. 1993. *Decisions with Multiple Objectives: Preferences and Value Trade-Offs.* Cambridge, UK: Cambridge University Press.

Keim, Daniel A. 2002. Information visualization and visual data mining. *IEEE Transactions on Visualization and Computer Graphics* 8 (1): 1–8.

Keisler, Jeffrey M. 2020a. Multiattribute utility module for Analytica. *Working Paper.*

———. 2020b. Strategy table module for Analytica. *Working Paper.*

———. 2003. Attribute-based Differentiation of Alternatives. *Journal of Multi-Criteria Decision Analysis* 11, no. 6 (December): 315–326.

Kirshen, Paul, Thomas Ballestero, Ellen Douglas, Christine D. Miller Hesed, Matthias Ruth, Michael Paolisso, Chris Watson, Phil Giffee, Kim Vermeer, and Kirk Bosma. 2018. Engaging Vulnerable Populations in Multi-Level Stakeholder Collaborative Urban Adaptation Planning for Extreme Events and Climate Risks – A Case Study of East Boston USA. *Journal of Extreme Events* 5 (September): 1850013.

Kirshen, Paul, Mark Borrelli, Jarrett Byrnes, Robert Chen, Lucy Lockwood, Chris Watson, Rebecca Herst, et al. 2018. *Feasibility of Harbor-Wide Barrier Systems: Preliminary Analysis for Boston Harbor.* Technical report.

Kirshen, Paul, Samuel Merrill, Peter Slovinsky, and Norman Richardson. 2012. Simplified Method for Scenario-Based Risk Assessment Adaptation Planning in the Coastal Zone. *Climatic Change* 113, no. 3 (August): 919–931.

Kitchenham, Barbara A., David Budgen, and Pearl O. Brereton. 2011. Using Mapping Studies as the Basis for Further Research–A Participant-Observer Case Study. *Information and Software Technology* 53, no. 6 (June): 638–651.

Kopp, Robert E., Robert M. DeConto, Daniel A. Bader, Carling C. Hay, Radley M. Horton, Scott Kulp, Michael Oppenheimer, David Pollard, and Benjamin H. Strauss. 2017. Evolving Understanding of Antarctic Ice-Sheet Physics and Ambiguity in Probabilistic Sea-Level Projections. *Earth's Future* 5, no. 12 (December): 1217–1233.

Kovacs, Gyöngyi, and Mohammad Moshtari. 2019. A Roadmap for Higher Research Quality in Humanitarian Operations: A Methodological Perspective. *European Journal of Operational Research* 276, no. 2 (July): 395–408.

Krippendorff, Klaus. 2004. *Content Analysis: An Introduction to Its Methodology.* 2nd ed. Los Angeles, CA, USA: SAGE Publications, Inc.

Kuhl, Laura, Paul H. Kirshen, Matthias Ruth, and Ellen M. Douglas. 2014. Evacuation as a Climate Adaptation Strategy for Environmental Justice Communities. *Climatic Change* 127, no. 3 (October): 493–504.

Kulkarni, Shruti. 2020. *Using Machine Learning to Analyze Climate Change Technology Transfer (CCTT).* In *Workshop on Tackling Climate Change with Machine Learning at International Conference on Learning Representations (ICLR).* Charlottesville, Virginia: Center for Open Science.

Kumar, Abhishek, Bikash Sah, Arvind R. Singh, Yan Deng, Xiangning He, Praveen Kumar, and Ramesh C. Bansal. 2017. A Review of Multi Criteria Decision Making (MCDM) towards Sustainable Renewable Energy Development. *Renewable and Sustainable Energy Reviews* 69 (March): 596–609.

Lee, Sangno, Jaeki Song, and Yongjin Kim. 2015. An Empirical Comparison of Four Text Mining Methods. *Journal of Computer Information Systems* 51, no. 1 (December): 1–10.

Lerche, Nils, Ines Wilkens, Meike Schmehl, Swantje Eigner-Thiel, and Jutta Geldermann. 2019. Using Methods of Multi-Criteria Decision Making to Provide Decision Support Concerning Local Bioenergy Projects. *Socio-Economic Planning Sciences* 68 (December): 100594.

Lesnikowski, Alexandra, Ella Belfer, Emma Rodman, Julie Smith, Robbert Biesbroek, John D. Wilkerson, James D. Ford, and Lea Berrang-Ford. 2019. Frontiers in Data Analytics for Adaptation Research: Topic Modeling. *Wiley Interdisciplinary Reviews: Climate Change* 10, no. 3 (May): 576–590.

Li, Kai, Jason Rollins, and Erjia Yan. 2017. Web of Science Use in Published Research and Review Papers 1997–2017: A Selective, Dynamic, Cross-Domain, Content-Based Analysis. *Scientometrics* 115, no. 1 (December): 1–20.

Linnenluecke, Martina K., Tom Smith, and Brent McKnight. 2016. Environmental Finance: A Research Agenda for Interdisciplinary Finance Research. *Economic Modelling* 59 (December): 124–130.

Lodree, Emmett J., and Selda Taskin. 2008. An insurance risk management framework for disaster relief and supply chain disruption inventory planning. *Journal of Operational Research Society* 59, no. 5 (May): 674–684.

Lumina Decision Systems. 2018. *Tornado Plots – Analytica Wiki.* https://wiki.analytica.com/Tornado_Plots. Accessed: 2020-09-07.

———. 2020. *Analytica Software (ver 5.0).* https://lumina.com. Accessed: 2022-07-11.

Madnick, Stuart E., Richard Y. Wang, Yang W. Lee, and Hongwei Zhu. 2009. Overview and framework for data and information quality research. *Journal of Data and Information Quality* 1, no. 1 (June): 1–22.

Maier, Daniel, Annie Waldherr, Peter Miltner, Gregor Wiedemann, Andreas Niekler, Alexa Keinert, Barbara Pfetsch, et al. 2018. Applying LDA Topic Modeling in Communication Research: Toward a Valid and Reliable Methodology. *Communication Methods and Measures* 12, nos. 2-3 (February): 93–118.

Maizlish, Neil, James Woodcock, Sean Co, Bart Ostro, Amir Fanai, and David Fairley. 2013. Health Cobenefits and Transportation-Related Reductions in Greenhouse Gas Emissions in the San Francisco Bay Area. *American Journal of Public Health* 103, no. 4 (March): 703–709.

Marjanac, Sophie, and Lindene Patton. 2018. Extreme Weather Event Attribution Science and Climate Change Litigation: An Essential Step in the Causal Chain? *Journal of Energy & Natural Resources Law* 36, no. 3 (February): 265–298.

Martínez-García, Miriam, Aida Valls, Antonio Moreno, and Arantza Aldea. 2018. A Semantic Multi-Criteria Approach to Evaluate Different Types of Energy Generation Technologies. *Environmental Modelling & Software* 110 (December): 129–138.

McKay, Michael D., Richard J. Beckman, and William J. Conover. 1979. A Comparison of Three Methods for Selecting Values of Input Variables in the Analysis of Output from a Computer Code. *Technometrics* 21, no. 2 (May): 239–245.

McKinley, Galen, Miriam Zuk, Morten Hojer, Montserrat Avalos, Isabel González, Rodolfo Iniestra, Israel Laguna Monroy, et al. 2005. Quantification of Local and Global Benefits from Air Pollution Control in Mexico City. *Environmental Science & Technology* 39, no. 7 (February): 1954–1961.

McNamee, Peter, and John Nunzio Celona. 2008. *Decision Analysis for the Professional.* 4th ed. Edited by Mimi Campbell, Bill Roehl, and Mary Story. USA: SmartOrg, Inc.

Mechler, Reinhard, Jeffrey Czajkowski, Howard Kunreuther, Erwann Michel-Kerjan, Wouter Botzen, Adriana Keating, Colin McQuistan, Nathan Cooper, and Ian O'Donnell. 2014. *Making Communities More Flood Resilient: The Role of Cost Benefit Analysis and Other Decision-support Tools in Disaster Risk Reduction.* Zurich, Switzerland: Zurich Flood Resilience Alliance.

Metz, Alexandre, Geoff Darch, and Mark Workman. 2016. Realising a Climate-Resilient UK Electricity and Gas System. *Proceedings of the Institution of Civil Engineers-Energy* 169, no. 1 (February): 30–43.

Mikolov, Tomas, Ilya Sutskever, Kai Chen, Greg S. Corrado, and Jeff Dean. 2013. *Distributed Representations of Words and Phrases and Their Compositionality.* In *Advances in Neural Information Processing Systems,* vol. 26. Red Hook, NY, USA: Curran Associates, Inc.

Mikulewicz, Michael. 2017. Politicizing Vulnerability and Adaptation: On the Need to Democratize Local Responses to Climate Impacts in Developing Countries. *Climate and Development* 10, no. 1 (March): 18–34.

Miller-Hooks, Elise, Xiaodong Zhang, and Reza Faturechi. 2012. Measuring and Maximizing Resilience of Freight Transportation Networks. *Computers & Operations Research* 39, no. 7 (July): 1633–1643.

Molenaar, Martin. 1998. *An Introduction to the Theory of Spatial Object Modelling for GIS.* 1st ed. Bristol, PA: Taylor & Francis.

Monteleoni, Claire, Gavin A. Schmidt, Francis J. Alexander, Alexandru Niculescu-Mizil, Karsten Steinhaeuser, Michael Tippett, Arindam Banerjee, et al. 2016. "Climate Informatics." In *Computational Intelligent Data Analysis for Sustainable Development,* 81–126. London, UK: Chapman / Hall, CRC Press.

Mooney, Christopher Z. 1997. *Monte Carlo simulation.* 116. California, USA: SAGE Publications, Inc.

Morgan, Millett Granger, Max Henrion, and Mitchell Small. 1990. *Uncertainty: A Guide to Dealing with Uncertainty in Quantitative Risk and Policy Analysis.* USA: Cambridge University Press.

Mydock III, Suni, Simon James Pervan, Alanoud F. Almubarak, Lester Johnson, and Michael Kortt. 2018. Influence of Made With Renewable Energy Appeal on Consumer Behaviour. *Marketing Intelligence & Planning* 36, no. 1 (January): 32–48.

National Oceanic and Atmospheric Administration. 2013. *NOAA Data Access Viewer.* https://coast.noaa.gov/dataviewer/. Accessed: 2021-08-04.

———. 2018. *NOAA Tides and Currents: Extreme Water Levels.* https://tidesandcurrents.noaa.gov/stationhome.html?id=8443970. Accessed: 2020-03-10.

Nelson, Laura K., Derek Burk, Marcel Knudsen, and Leslie McCall. 2018. The Future of Coding: A Comparison of Hand-Coding and Three Types of Computer-Assisted Text Analysis Methods. *Sociological Methods & Research* (Los Angeles, CA, USA) 50, no. 1 (May): 202–237.

Nerlich, Brigitte, Nelya Koteyko, and Brian Brown. 2010. Theory and Language of Climate Change Communication. *Wiley Interdisciplinary Reviews: Climate Change* 1, no. 1 (January): 97–110.

Neumann, James E., Kerry Emanuel, Sai Ravela, Lindsay Ludwig, Paul Kirshen, Kirk Bosma, and Jeremy Martinich. 2015. Joint Effects of Storm Surge and Sea-Level Rise on US Coasts: New Economic Estimates of Impacts, Adaptation, and Benefits of Mitigation Policy. *Climatic Change* 129, no. 1 (March): 337–349.

Nicholls, Robert J., Poh P. Wong, Virginia R. Burkett, Jorge O. Codignotto, John E. Hay, Roger F. McLean, Sachooda Ragoonaden, and Colin D. Woodroffe. 2007. "Coastal Systems and Low-Lying Areas." In *Climate Change 2007: Impacts, Adaptation and Vulnerability.* Cambridge, UK: Cambridge University Press.

Nordhaus, William D. 2017. Revisiting the Social Cost of Carbon. *Proceedings of the National Academy of Sciences* 114, no. 7 (January): 1518–1523.

Okoli, Chitu. 2015. A Guide to Conducting a Standalone Systematic Literature Review. *Communications of the Association for Information Systems* 37, no. 1 (November): 43.

Oloruntoba, Richard, and Richard Gray. 2006. Humanitarian Aid: An Agile Supply Chain? *Supply Chain Management: An International Journal* 11, no. 2 (March): 115–120.

Oppenheimer, Michael, Bruce Glavovic, Jochen Hinkel, Roderik van de Wal, Alexandre K. Magnan, Amro Abd-Elgawad, Rongshuo Cai, et al. 2019. "Sea Level Rise and Implications for Low-Lying Islands, Coasts and Communities." In *IPCC Special Report on the Ocean and Cryosphere in a Changing Climate,* 321–445. Cambridge, UK and New York, NY, USA: Cambridge University Press.

Pedersen Zari, Maibritt. 2014. Ecosystem Processes for Biomimetic Architectural and Urban Design. *Architectural Science Review* 58, no. 2 (November): 106–119.

Phillips-Wren, Gloria, Mary Daly, and Frada Burstein. 2021. Reconciling business intelligence, analytics and decision support systems: More data, deeper insight. *Decision Support Systems* 146 (July): 113560.

Piepenbrink, Anke, and Elkin Nurmammadov. 2015. Topics in the Literature of Transition Economies and Emerging Markets. *Scientometrics* 102, no. 3 (January): 2107–2130.

Pollard, James A., Tom Spencer, and Simon Jude. 2018. Big Data Approaches for Coastal Flood Risk Assessment and Emergency Response. *Wiley Interdisciplinary Reviews: Climate Change* 9, no. 5 (July): e543.

Porter, Michael E. 1998. *On Competition (A Harvard Business Review Book).* 1st ed. Boston, MA, USA: Harvard Business School Publishing.

Powell, Kendall. 2016. Does It Take Too Long to Publish Research? *Nature* 530, no. 7589 (February): 148–151.

Reed, Mark, Guillermo Podesta, Ioan Fazey, Nichola Geeson, Rudi Hessel, Hubacek Klaus, David Letson, et al. 2013. Combining Analytical Frameworks to Assess Livelihood Vulnerability to Climate Change and Analyse Adaptation Options. *Ecological Economics* 94 (October): 66–77.

Rehman, Muhammad Habib ur, Ibrar Yaqoob, Khaled Salah, Muhammad Imran, Prem Prakash Jayaraman, and Charith Perera. 2019. The role of big data analytics in industrial Internet of Things. *Future Generation Computer Systems* 99 (October): 247–259.

Rehurek, Radim, and Petr Sojka. 2011. *Gensim–Python Framework for Vector Space Modelling*. /https://radimrehurek.com/gensim/. Accessed: 2021-02-23. NLP Centre, Faculty of Informatics, Masaryk University, Brno, Czech Republic.

Reiblich, Jesse, Eric Hartge, Lisa M. Wedding, Sierra Killian, and Gregory M. Verutes. 2019. Bridging Climate Science, Law, and Policy to Advance Coastal Adaptation Planning. *Marine Policy* 104 (June): 125–134.

Resch, Bernd, Florian Usländer, and Clemens Havas. 2017. Combining Machine-Learning Topic Models and Spatiotemporal Analysis of Social Media Data for Disaster Footprint and Damage Assessment. *Cartography and Geographic Information Science* 45, no. 4 (August): 362–376.

Rezaeian, Mina, Hamid Montazeri, and Roel Loonen. 2017. Science Foresight Using Life-Cycle Analysis, Text Mining and Clustering: A Case Study on Natural Ventilation. *Technological Forecasting and Social Change* 118 (March): 270–280.

Röder, Michael, Andreas Both, and Alexander Hinneburg. 2015. Exploring the Space of Topic Coherence Measures. *WSDM 2015 – Proceedings of the 8th ACM International Conference on Web Search and Data Mining* (February): 399–408.

Rodríguez, J. Tinguaro, Begoña Vitoriano, Javier Montero, and Vojislav Kecman. 2011. A Disaster-Severity Assessment DSS Comparative Analysis. *OR Spectrum* 33, no. 3 (May): 451–479.

Rodríguez, Tinguaro, J, Begoña Vitoriano, and Javier Montero. 2012. A General Methodology for Data-Based Rule Building and Its Application to Natural Disaster Management. *Computers & Operations Research* 39, no. 4 (April): 863–873.

Rondinini, Carlo, and Piero Visconti. 2015. Scenarios of Large Mammal Loss in Europe for the 21st Century. *Conservation Biology* 29, no. 4 (May): 1028–1036.

Rosenzweig, Cynthia, William D. Solecki, Reginald Blake, Malcolm Bowman, Craig Faris, Vivien Gornitz, Radley Horton, et al. 2011. Developing Coastal Adaptation to Climate Change in the New York City Infrastructure-Shed: Process, Approach, Tools, and Strategies. *Climatic Change* 106, no. 1 (February): 93–127.

Roy, Abhra, and Jomon Aliyas Paul. 2013. Terrorism Deterrence in a Two Country Framework: Strategic Interactions Between R&D, Defense and Pre-Emption. *Annals of Operations Research* 211, no. 1 (July): 399–432.

Ruckert, Kelsey L., Vivek Srikrishnan, and Klaus Keller. 2019. Characterizing the Deep Uncertainties Surrounding Coastal Flood Hazard Projections: A Case Study for Norfolk, VA. *Scientific Reports* 9, no. 1 (August): 1–12.

Samaddar, Subhajyoti, Muneta Yokomatsu, Frederick Dayour, Martin Oteng-Ababio, Togbiga Dzivenu, Mujeeb Adams, and Hirohiko Ishikawa. 2015. Evaluating Effective Public Participation in Disaster Management and Climate Change Adaptation: Insights From Northern Ghana Through a User-Based Approach. *Risk, Hazards & Crisis in Public Policy* 6, no. 1 (August): 117–143.

Sarma, Gopal P. 2017. Scientific Literature Text Mining and the Case for Open Access. *arXiv preprint arXiv:1611.00097* (December).

Schlager, Edella, and Tanya Heikkila. 2011. Left High and Dry? Climate Change, Common-Pool Resource Theory, and the Adaptability of Western Water Compacts. *Public Administration Review* 71, no. 3 (May): 461–470.

Schütze, Hinrich, Christopher D. Manning, and Prabhakar Raghavan. 2008. *Introduction to Information Retrieval.* Vol. 39. Cambridge, UK: Cambridge University Press.

Scism, Leslie. 2021. *Ida Storm Damage Expected to Cost Insurers at Least $31 Billion.* https://www.wsj.com/articles/ida-storm-damage-expected-to-cost-insurers-atleast-31-billion-11632303002. Accessed: 2021-09-23.

Sebestyén, Viktor, Tımea Czvetkó, and János Abonyi. 2021. The applicability of big data in climate change research: The importance of system of systems thinking. *Frontiers in Environmental Science* 9 (March): 70.

Senbel, Maged, Timothy McDaniels, and Hadi Dowlatabadi. 2003. The Ecological Footprint: A Non-Monetary Metric of Human Consumption Applied to North America. *Global Environmental Change* 13, no. 2 (July): 83–100.

Sharma, Arun K. 2020. *Understanding Latent Dirichlet Allocation (LDA).* https://www.mygreatlearning.com/blog/understanding-latent-dirichlet-allocation/.

Siders, Anne R. 2019. Adaptive Capacity to Climate Change: A Synthesis of Concepts, Methods, and Findings in a Fragmented Field. *Wiley Interdisciplinary Reviews: Climate Change* 10, no. 3 (January): e573.

Sievert, Carson, and Kenneth Shirley. 2014. *A Method for Visualizing and Interpreting Topics.* In *Proceedings of Workshop on Interactive Language Learning, Visualization, and Interfaces, Association for Computational Linguistics,* 63–70. Baltimore, Maryland, USA: Association for Computational Linguistics.

Sills, George L., Noah D. Vroman, Ronald E. Wahl, and Neil T. Schwanz. 2008. Overview of New Orleans Levee Failures: Lessons Learned and Their Impact on National Levee Design and Assessment. *Journal of Geotechnical and Geoenvironmental Engineering* 134, no. 5 (May): 556–565.

Smit, Barry, Ian Burton, Richard J.T. Klein, and Roger Street. 1999. The Science of Adaptation: A Framework for Assessment. *Mitigation and Adaptation Strategies for Global Change* 4, no. 3 (September): 199–213.

Smith, Jordan W., Dorothy H. Anderson, and Roger L. Moore. 2012. Social Capital, Place Meanings, and Perceived Resilience to Climate Change. *Rural Sociology* 77, no. 3 (June): 380–407.

Sodhi, ManMohan S., and Christopher S. Tang. 2013. Buttressing Supply Chains Against Floods in Asia for Humanitarian Relief and Economic Recovery. *Production and Operations Management* 23, no. 6 (August): 938–950.

Solaun, Kepa, and Emilio Cerdá. 2017. The Impact of Climate Change on the Generation of Hydroelectric Power–A Case Study in Southern Spain. *Energies* 10, no. 9 (September): 1343.

Solera Jimenez, Maricruz. 2017. Green Walls: A Sustainable Approach to Climate Change, A Case Study of London. *Architectural Science Review* 61, nos. 1-2 (November): 48–57.

Stadler, Michael, Chris Marnay, Ines Lima Azevedo, Ryoichi Komiyama, and Judy Lai. 2009. *The Open Source Stochastic Building Simulation Tool SLBM and Its Capabilities to Capture Uncertainty of Policymaking in the US Building Sector.* Technical report. Berkeley, CA, USA: Lawrence Berkeley National Lab. (LBNL).

Steyvers, Mark, and Tom Griffiths. 2007. "Probabilistic Topic Models." In *Handbook of Latent Semantic Analysis,* 1st ed., 439–460. London, UK: Psychology Press.

Strantzali, Eleni, and Konstantinos Aravossis. 2016. Decision Making in Renewable Energy Investments: A Review. *Renewable and Sustainable Energy Reviews* 55 (March): 885–898.

Suddeth, Robyn J., Jeff Mount, and Jay R. Lund. 2010. Levee Decisions and Sustainability for the Sacramento-San Joaquin Delta. *San Francisco Estuary and Watershed Science* 8, no. 2 (August): 1–23.

Sutton-Grier, Ariana E., Kateryna Wowk, and Holly Bamford. 2015. Future of Our Coasts: The Potential for Natural and Hybrid Infrastructure to Enhance the Resilience of our Coastal Communities, Economies and Ecosystems. *Environmental Science & Policy* 51 (August): 137–148.

Sweet, William V., Robert E. Kopp, Christopher P. Weaver, Jayantha Obeysekera, Radley M. Horton, E. Robert Thieler, and Chris Zervas. 2017. *Global and Regional Sea Level Rise Scenarios for the United States.* Technical report 083. National Ocean Service (NOS) Center for Operational Oceanographic Products and Services (CO-OPS).

Syed, Shaheen, and Marco Spruit. 2017. *Full-Text or Abstract? Examining Topic Coherence Scores Using Latent Dirichlet Allocation.* In *2017 IEEE International Conference on Data Science and Advanced Analytics (DSAA),* 165–174. Tokyo, Japan: IEEE.

The Climate Group. 2019. *Budweiser Puts 100% Renewable Electricity Message at Heart of Super Bowl Commercial.* https://www.there100.org/our-work/news/budweiser-puts-100-renewable-electricity-message-heart-super-bowl-commercial-news. Accessed: 2022-07-11.

The Institute for Operations Research and the Management Sciences. 2020. *Energy, Natural Resources, and The Environment (ENRE).* https://connect.informs.org/energy-natural-resources-and-the-environment/home. Accessed: 2020-11-24.

Thompson, Philip R., Matthew J. Widlansky, Mark A. Merrifield, Janet M. Becker, and John J. Marra. 2019. A statistical model for frequency of coastal flooding in Honolulu, Hawaii, during the 21st century. *Journal of Geophysical Research: Oceans* 124, no. 4 (April): 2787–2802.

Toimil, Alexandra, Iñigo J. Losada, Robert J. Nicholls, Robert A. Dalrymple, and Marcel J.F. Stive. 2020. Addressing the Challenges of Climate Change Risks and Adaptation in Coastal Areas: A Review. *Coastal Engineering* 156 (March): 103611.

Truong, Chi, Stefan Trück, and Supriya Mathew. 2018. Managing Risks from Climate Impacted Hazards–The Value of Investment Flexibility under Uncertainty. *European Journal of Operational Research* 269, no. 1 (August): 132–145.

Tufte, Edward R. 2006. *Beautiful evidence.* Vol. 1. Cheshire, CT, USA: Graphics Press.

Tylock, Steven M., Thomas P. Seager, Jeff Snell, Erin R. Bennett, and Don Sweet. 2012. Energy Management Under Policy and Technology Uncertainty. *Energy Policy* 47 (August): 156–163.

U.S. General Services Administration. 2009. *Open Government.* https://www.data.gov/open-gov/. Accessed: 2021-08-04.

Uichanco, Joline. 2022. A Model for Prepositioning Emergency Relief Items Before a Typhoon with an Uncertain Trajectory. *Manufacturing & Service Operations Management* 24 (2): 766–790.

United Nations Treaty Collection. 2015. *Paris Agreement to the United Nations Framework Convention on Climate Change.* https://unfccc.int/process-and-meetings/the-paris-agreement/the-paris-agreement. Accessed: 2022-05-13.

Vanicek, Petr. 1991. Vertical Datum and NAVD88. *Surveying and Land Information Systems* 51 (2): 83–86.

Vasiliev, Yuli. 2020. *Natural Language Processing with Python and SpaCy: A Practical Introduction.* No Starch Press.

Vuuren, Detlef P. van, Jae Edmonds, Mikiko Kainuma, Keywan Riahi, Allison Thomson, Kathy Hibbard, George C. Hurtt, et al. 2011. The Representative Concentration Pathways: An Overview. *Climatic Change* 109, no. 1 (August): 5–31.

Wang, Jiang Jiang, You-Yin Jing, Chun-Fa Zhang, and Jun-Hong Zhao. 2009. Review on Multi-Criteria Decision Analysis Aid in Sustainable Energy decision-making. *Renewable and Sustainable Energy Reviews* 13 (9): 2263–2278.

Wang, Shuo, and Gordon H. Huang. 2016. Risk-based factorial probabilistic inference for optimization of flood control systems with correlated uncertainties. *European Journal of Operational Research* 249, no. 1 (February): 258–269.

Watson Marketing. 2017. *10 Key Marketing Trends for 2017 and Ideas for Exceeding Customer Expectations.* https://paulwriter.com/wp-content/uploads/2017/10/10-Key-Marketing-Trends-for-2017.pdf. Accessed: 2023-02-10. IBM.

Winston, Andrew, George Favaloro, and Tim Healy. 2017. Energy Strategy for the C-suite. *Harvard Business Review* 95 (1): 139–46.

Wirth, Rüdiger, and Jochen Hipp. 2000. *CRISP-DM: Towards a Standard Process Model for Data Mining.* In *Proceedings of the 4th International Conference on the Practical Applications of Knowledge Discovery and Data Mining,* 1:29–39. Waltham, MA, USA: Practical Application Company.

Wissman-Weber, Nichole, and David L. Levy. 2021. Organizing for Climate Adaptation: Competing Visions in Boston. *Economic Sociology – The European Electronic Newsletter* 22 (2): 24–29.

Wuebbles, Donald J., David W. Fahey, and Kathy A. Hibbard. 2018. *Climate Science Special Report: Fourth National Climate Assessment.* Technical report. Washington, DC, USA: U.S. Global Change Research Program.

Ying, Luwei, Jacob M. Montgomery, and Brandon M. Stewart. 2019. Inferring Concepts from Topics: Towards Procedures for Validating Topics as Measures. *Society for Political Methodology* 5.

Zandvoort, Mark, Nora Kooijmans, Paul Kirshen, and Adri van den Brink. 2019. Designing with Pathways: A Spatial Design Approach for Adaptive and Sustainable Landscapes. *Sustainability* 11 (3): 565.

Zheng, Bin, David C. McLean, and Xinghua Lu. 2006. Identifying Biological Concepts from a Protein-Related Corpus with a Probabilistic Topic Model. *BMC Bioinformatics* 7 (1): 1–10.

Zobel, Christopher W. 2011. Representing Perceived Tradeoffs in Defining Disaster Resilience. *Decision Support Systems* 50 (2): 394–403.

BIOGRAPHICAL SKETCH

DONALD J. JENKINS
Degree of DOCTOR OF PHILOSOPHY

Dissertation: APPLICATION OF BUSINESS ANALYTICS APPROACHES TO ADDRESS CLIMATE-CHANGE-RELATED CHALLENGES

Major: MANAGEMENT SCIENCE

Biographical:

Education:

- Received an Associates of Applied Sciences (A.A.S.) from Charter Oak Community College, Connecticut in April 1990
- Received dual B.S./B.A. in Electrical Engineering from the University of San Diego, California in May 1994
- Received M.S. Operations Research from the Naval Postgraduate School Monterey, California in June 2002
- Completed the requirements for the degree of Doctor of Philosophy with a major in Management Science at University of Massachusetts Boston in August 2023

Experience:

Don served in the US Marine Corps for eight years prior to serving as a US Naval Submarine Officer for another fourteen years. During this time, Don's interest in problem-solving led him to earn a dual BS/BA Electrical Engineering degree from the University of San Diego and an MS Operations Research degree from the Naval Postgraduate School Monterey. While in Monterey, Don completed his thesis on data mining for fraud detection with the Defense Finance and Accounting Service. Upon retiring

from active duty in 2007, Don joined the cleantech energy services company EnerNOC based in Boston. While there, he led global teams that implemented, delivered, and supported a full suite of Energy Advisory Solutions. When Enel-X acquired EnerNOC in 2017, Don moved on to become the Enel-X North America Chief Data Scientist while simultaneously beginning the pursuit of his Ph.D. at UMass Boston. Don left Enel-X in 2019 and started the data consultancy Data Driven Decision Advisors (D3 Advisors). In this new role, he helps commercial customers better utilize their business data for data-driven decision-making. Don also joined behavioral health startup Mirah in 2021. As Chief Operating Officer of Mirah, Don is helping mental health professionals develop evidence-based practices enabled by Mirah's *Software-as-a-Service* platform. Throughout his post-naval career, Don has led teams delivering services to clients in more than 100 countries. He acts as a tireless data champion, driving high-quality data collection and meaningful analysis to transform how his companies and their customers manage their businesses. Now that he has earned his Ph.D., Don will continue working in business leadership roles to apply his data-science-based knowledge and methods.