UNIVERSITY OF OKLAHOMA

GRADUATE COLLEGE

INTEGRATING ML-BASED REVERSE ENGINEERING OF CGE MODELS WITH OPTIMIZATION FOR ENHANCING RESILIENCE

A THESIS

SUBMITTED TO THE GRADUATE FACULTY

in partial fulfillment of the requirements for the

Degree of

MASTER OF SCIENCE

BY

NUSHRA ZANNAT

Norman, Oklahoma

2024

INTEGRATING ML-BASED REVERSE ENGINEERING OF CGE MODELS WITH OPTIMIZATION FOR ENHANCING RESILIENCE

A THESIS APPROVED FOR THE
GALLOGLY COLLEGE OF ENGINEERING

BY THE COMMITTEE CONSISTING OF

Dr. Charles D. Nicholson, Chair

Dr. Andres Gonzalez Huertas

Dr. Talayeh Razzaghi

Acknowledgements

I am profoundly grateful to Dr. Charles Nicholson for his unwavering support and invaluable guidance throughout this journey. Dr. Nicholson's kindness, patience, passion for his field, and the dedication he provides for his students have inspired me to push beyond my own expectations. I am incredibly lucky to have had him as a mentor. And it has been a privilege to learn from him.

I am also sincerely thankful to Dr. Andres Gonzalez for his steadfast support, helpfulness with any issues, and patience. I owe him much of my growth during this time. I would like to express my gratitude to Dr. Talayeh Razzaghi for her time, and support as a member of my committee.

To my friends, Himadri Sen Gupta and Mansura Roly – who have always been there to help me, no matter the situation, since my time here in the US.

Lastly, I owe everything to my family, my mother, and my sister, Rafia Bushra—I owe my deepest gratitude. Thank you for your constant love and support. None of this would be possible without you.

Sincerely,

Nushra Zannat

Signature

Contents

List of Figures

List of Tables

Abstract

Natural hazards pose significant risks to economic stability and community well-being, particularly in regions vulnerable to seismic activity. This research develops a novel framework integrating machine learning (ML) and optimization techniques to enhance community resilience against natural hazards, with a specific focus on earthquake impacts. A Computable General Equilibrium (CGE) model is employed to simulate economic outcomes, such as domestic supply, income, employment, and migration, under various earthquake scenarios. To overcome the computational challenges of CGE models, an Elastic Net regression-based ML model is utilized as a surrogate, achieving a predictive accuracy of over 98% (R^2). These predictive insights are incorporated into an integer linear programming (ILP) optimization model to identify optimal retrofit strategies for minimizing economic loss while addressing resource constraints. The framework is applied to Salt Lake City (SLC), Utah, a region highly vulnerable to seismic activity. Using 5,000 earthquake simulations, the approach evaluates economic impacts across 11 sectors and seven geographic regions. The optimization model ensures equitable resource allocation and effectively prioritizes infrastructure retrofits based on sector-specific vulnerabilities and budget limitations.

This work makes significant contributions by: (i) introducing a scalable framework that integrates CGE models, ML, and optimization for disaster resilience, (ii) demonstrating the effectiveness of Elastic Net regression in balancing model accuracy and interpretability, and (iii) providing actionable insights for policymakers to enhance disaster preparedness and recovery strategies. The findings emphasize the value of advanced computational tools in informing data-driven decision-making for community resilience. Future research will focus on extending the methodology to multi-hazard scenarios and dynamic resilience metrics to broaden its applicability and impact.

1 Introduction

Natural hazards have posed a threat throughout human history, causing widespread devastation and leaving long-term socio-economic impacts. These events, ranging from earthquakes and hurricanes, to floods and wildfires, have the potential to devastate entire communities, disrupt economies, and cause significant loss of life. Aside from immediate physical damage, these disasters lead to prolonged economic damage, abrupt population dislocation, and disruption in the natural flow of life. Recovering from such impacts takes years, if not decades, and can test the capabilities and strain the resources of even the most developed nations. Globally, there have been over 1.6 million fatalities caused by natural hazards since 1990, with an estimated USD 260–310 billion in annual economic losses [2]. According to the National Centers for Environmental Information [3], over the past three years (2021-2023), the United States faced 41 severe storms, 9 tropical cyclones, and 16 other natural hazards including floods, droughts, wildfires, and winter storms. These events led to approximately 1,690 deaths and caused an average annual damage of 146.9 billion dollars. Earthquakes, in particular, represent one of the most destructive types of natural hazards due to their sudden onset, widespread impact, and potential to cause catastrophic loss of life and economic damage.

Even with all the technological and societal advances, natural hazards remain one of the most formidable obstacles that civilization faces. The frequency and severity of these events are exacerbated by factors such as climate change, deforestation, and urbanization. Earthquake hazards are further amplified by urbanization in high-risk zones and inadequate mitigation strategies for aging infrastructure, as evidenced by past seismic events like the 1995 Kobe earthquake and the 1994 Northridge earthquake [4], [5]. The interconnected nature of infrastructure systems—such as energy, transportation, and water supply—creates cascading risks, where the failure of one system can lead to the collapse of others. For example, disruptions in transportation systems following the 1994 Northridge earthquake significantly hindered emergency response efforts and prolonged economic recovery [5]. This highlights the critical need for proactive and integrated disaster mitigation strategies. While these events cannot be completely prevented, there are a variety of steps that communities can take to better prepare and protect themselves and enhance resilience [6]. Community resilience, as discussed in Koliou, Lindt, McAllister, *et al.* [7], encompasses the ability of communities to prepare for, resist, absorb, and recover from the impacts of natural hazards. It includes not only the physical infrastructure but also the social and economic systems that support the community. In the context of earthquakes, resilience encompasses designing buildings and infrastructure to withstand seismic forces, improving early warning systems, and ensuring equitable resource allocation during recovery efforts. [8] By implementing strategic planning, retrofitting infrastructure, and robust disaster preparedness frameworks, communities become more resilient and grow stronger over the long term [9], [10]. This paper focuses on Salt Lake City (SLC), which is one of the most seismically hazardous areas due to its location within the Intermountain Seismic Belt and its proximity to the active Wasatch fault. Earthquakes are one of the most unpredictable hazards, capable of causing widespread devastation in mere seconds with little to no warning. Past seismic events and recent studies suggest that Salt Lake City is overdue for a major earthquake, which could result in catastrophic damage and severe economic repercussions. Defining the scope of this problem and developing effective mitigation strategies are crucial for enhancing the city's resilience to such events.

This research proposes an integrated framework combining machine learning (ML) and optimization techniques to enhance earthquake resilience in SLC. The framework uses a Computable General Equilibrium (CGE) model to assess economic impacts, employing penalized regression models, such as Elastic Net, to derive coefficients quantifying relationships between economic indicators and seismic impacts. These coefficients are then integrated into a multi-objective optimization model to prioritize retrofit strategies, minimizing economic losses and ensuring effective resource allocation. Unlike traditional methods, this framework bridges economic modeling and practical decision-making, offering a data-driven approach to disaster preparedness.

The following sections of this paper provide a detailed account of the methodology, findings, and implications of this research. Chapter 2 reviews the existing literature on disaster resilience and state-of-the-art mitigation strategies. Chapter 3 introduces the case study of SLC and outlines the proposed CGE-ML framework, including the specific machine learning models and optimization techniques employed. Chapter 4 presents a comprehensive quantitative analysis of the results, including scenario-based evaluations, sensitivity analyses, and their implications for policy and practice. Practical recommendations for resource allocation and disaster preparedness are also discussed in this chapter. Finally, Chapter 5 summarizes the contributions of this study and identifies potential avenues for future research, including extending the framework to multi-hazard scenarios and incorporating dynamic resilience metrics.

2 Background and Related Work

The concept of community resilience has accumulated substantial attention in recent years as a vital component of disaster preparedness and response. This chapter delves into the field of resilience planning, showcasing its complex nature, exploring various definitions, and reviewing key research literature.

2.1 Exploring the Multidimensional and Intricate Nature of Community Resilience

"Resilience" is a term constantly used and discussed and is perhaps one of the most dominating factors in Disaster Risk Reduction (DRR) and Disaster Risk Management (DRM). Despite its wide usage and being the focal point of discussion surrounding natural hazards, DRR and DRM, it is a complex, misunderstood, and ever-changing term. Its meaning evolved through time, with different variations, definitions, and unique perspectives offered by people across various fields. For years, the concept of resilience in general and in terms of community and hazards has been thought of as a process for withstanding disturbances and bouncing back to stability — its prior state. For instance, in general, a resilient material, when put under pressure, will bend and bounce back, rather than breaking [11]–[13]. In the context of natural hazards and community resilience, the early works of Dovers and Handmer [14], Mileti and Noji [15], and Paton and Johnston [16] collectively showcased resilience as a community's ability to cope, strive and return to a stable state after a hazard event. These studies have provided a foundational layout for several DRM frameworks. However, the notion of returning to stability — aligning with the "bounce-back" idea of resilience has often been critiqued as narrow and counterproductive in terms of reducing vulnerability [7], [17]–[19].Timmerman [20] is one of the earliest to define resilience in terms of disasters (man-made or natural) and focuses on the ability of (resilient) systems not only to recover but also adapt to new circumstances if needed. Timmerman [20] grounds much of his work from the famous paper by Holling [21], who is acknowledged to be a pioneer of resilience theory and one of the first to distinguish the inextricably linked concepts of "resilience" and "stability" as two separate ideas gracefully. Handmer and Dovers [22] build on their earlier work from 1992 by refining and proposing a 3-tier typology of resilience: Type I (Resilience as Stability; resistant to change; the traditional bounce-back approach), Type II (Resilience through Adaptation; change at the margins; Adaptive Resilience), and Type III (Resilience through Transformation; openness and adaptation; Transformative Resilience). Most frameworks in recent studies for community resilience, DRR, DRM focus on adaptive resilience and fall in the Type II category, aiming to gradually improve existing systems to better withstand and recover from disasters. While Type III is increasingly gaining recognition in current studies, its complexity and the need for large-scale transformation and changes in established norms make it difficult to achieve, and it remains relatively uncommon in practice.

Defining community resilience precisely, given its broad and varied understandings, is challenging as it is inherently multidimensional. Each viewpoint— economic, social, environmental, ecological, etc. — adds a unique dimension to resilience. Thus, resilience, especially for hazard events, spans across disciplines such as psychology, public health, environmental science, engineering, infrastructural, economics, and social sciences [11], [23]–[28]. While social scientists may consider social responsibility as major factor for enhancing resilience [29], [30], civil engineers might look into improving infrastructure systems to build stronger communities [31], [32]; both are important aspects to consider. Therefore, no single persepective is suffiecient to evaluate and enahance resilience; rather a multidisciplinary and integrated approach is required that acknowledges the intricacies between these dimensions.

2.2 Relevant Studies for Enhancing Community Resilience: Multifaceted Approaches and Computational Tools

This section reviews key studies in community resilience and disaster management, exploring a range of methodologies such as theoretical frameworks, quantitative methods, or computational approaches employed by several studies conducted throughout the years. The studies highlight several dimensions of resilience, such as economic, social, and infrastructural, offering insights into resilience assessment, planning, and decision-making to address the challenges of natural hazards.

For instance, Norris, Stevens, Pfefferbaum, *et al.* [11] discusses four dimensions - Economic Development, Social Capital, Information and Communication, and Community Competence to equip communities with proper resources and reduce risks. It lays the groundwork for several future studies in this area with several conceptual insights. Their work is theoretical and limited in quantitative resilience measurement. In contrast, Bruneau, Chang, Eguchi, *et al.* [33] presents both a conceptual and quantitative framework for assessing seismic resilience across four dimensions — technical, organizational, social, and economic. They focused on metrics like robustness, redundancy, resourcefulness, and rapidity, creating a structured way of analyzing resilience within communities. Another study by Wang, Tang, Qi, *et al.* [34] focus on the built environment, particularly in rural areas, to show how infrastructural resilience interacts with social vulnerabilities.

Cutler, Shields, Tavani, *et al.* [35] developed a dynamic CGE model to estimate the economic impacts of natural disasters. They work with engineering and economic data to simulate the effects of disasters on a virtual community — Centerville, providing a comprehensive tool for policymakers. Kajitani and Tatano [4] applied a CGE model to examine the short-term economic effects of the 2011 Great East Japan Earthquake and Tsunami. Guha [36] applied a CGE model to simulate the economic impacts of electricity outages for a hypothetical earthquake in Memphis, Tennessee. CGE models are great tools for quantitative insights into the effects of disasters across sectors, but they are also computationally intensive and complex.

Several studies leverage ML approaches and predictive models for analyzing disaster-prone areas, providing data-driven insights into resilience planning. Bastaminia, Rezaei, and Saraei [37] examined social and economic resilience across earthquake-affected communities Bam and Rudbar in Iran. They used household surveys and data analysis methods like multiple linear regression and neural networks. Their study is comprehensive and well-established for the specific regions. However, it may not apply to other areas or disasters. The reliance on self-reported survey data also raises concerns about subjective biases. Nonetheless, their research provides insight into socioeconomic factors like income and social capital, showing that Rudbar had higher resilience than Bam. Nachappa, Ghorbanzadeh, Gholamnia, *et al.* [38] conducted multi-hazard exposure mapping for Salzburg, Austria, focusing on floods and landslides. Using machine learning models like Support Vector Machine (SVM) and Random Forest (RF), they analyzed 13 influencing factors. Their study provides a detailed approach for this specific alpine region, with RF achieving high accuracy in exposure predictions. However, similar to the previous study, it may be limited to other areas due to model complexity and data requirements.

Mathematical models and decision frameworks can help find optimal solutions to complex problems and guide complex choices to help enhance community resilience. For example - Zhang and Nicholson [39] developed a multi-objective optimization model for retrofitting strategies to minimize direct economic losses and population dislocation following an earthquake. Their model allows detailed resource allocation decisions within a hypothetical community (Centerville) comprised of over 15,000 buildings. Building on this, Gupta, Adluri, Sanderson, *et al.* [40] introduced a similar multi-objective optimization framework to enhance resilience against multiple hazards, like earthquakes and tsunamis, focusing on Seaside, Oregon. Both of their work offers detailed strategies for resource allocation. Nonetheless, an integrated approach to resilience has yet to be provided by the studies that have been discussed thus far. Wen [41] builds on these advancements by proposing a multi-objective optimization model focusing on infrastructural resilience for Joplin. Wen makes a significant and fundamental addition to this topic in her final chapter, where she briefly presents a hybrid model technique for combining ML predictions with the multi-optimization framework. Although Wen's main focus, the multi-objective framework for Joplin and the ML methods, is well-documented, the results for the MOO model that utilizes a hybrid ML-optimization model are absent, leaving its practical application untested.

2.3 Research Gap and Contribution

Despite substantial progress in the field of resilience planning, there are notable gaps in research. Existing studies often tackle individual dimensions of resilience or lack comprehensive integration of tools that may help enhance resilience.

This research aims to bridge these gaps by developing an integrated framework that combines CGE model outputs with interpretable ML models, specifically Elastic Net Regression, and incorporates these into an optimization model for disaster response. Wen [41] provides a foundation for this work, and it is the closest to the work shown in this research. However, several things need to be addressed. Her approach integrates ML coefficients in the optimization model without considering the direct economic loss resulting from disasters for buildings. As a result, mitigation strategies may not accurately account for actual economic consequences of retrofitting decisions. This study explicitly incorporates direct economic losses into the decision-making framework, ensuring that optimization strategies are grounded in predictive relationships and their real economic implications. In addition, this research also addresses scalability and granularity by applying the framework to Salt Lake City, a larger and more complex urban environment than Joplin that is discussed in [41]. Furthermore, our work deals with both community-wide and sector-specific models, offering precision and enhancing their applicability for targeted resilience strategies. While Wen's multi-objective framework provides the flexibility to balance competing priorities, this research focuses on single-objective optimization to simplify implementation and produce concrete results.

To conclude, by combining ML techniques, optimization frameworks, and CGE models, this research bridges critical gaps in scalability, integration, and practical implementation while offering a thorough and actionable tool for enhancing community resilience. The high dimensionality of Salt Lake City, encompassing 70 sectors, enables an analysis that primarily captures economic interdependencies through the CGE model. When integrated into an optimization framework for retrofitting buildings, this analysis includes infrastructural considerations, bridging the economic and infrastructural domains. This work builds on and extends Wen's contributions by addressing the limitations, ensuring that resilience strategies are both informed and effective.

3 Research Methodology

The research combines large numeric models, machine learning, and optimization techniques to enhance resilience against natural hazards, such as earthquakes. First, a Computable General Equilibrium model is used to estimate economic impacts across the sectors based on simulated damage scenarios. Then, machine learning models, particularly Elastic Net, is trained to replicate CGE outputs, creating a simplified surrogate model. Finally, this surrogate model is integrated into a single-objective optimization framework, where decision variables like building retrofits are optimized to minimize impacts on certain economic factors (CGE outputs) under given constraints.

3.1 Approach

The surrogate model developed from the CGE framework aims to work within an optimization model for building retrofits, similar to the approach in Zhang and Nicholson [39]. There are several critical characteristics necessary for a successful integration of ML with optimization. These are summarized below and follow the guidelines provided in Wen [41].

3.1.1 Model-based Structure

A variety of machine learning methods can estimate the economic outputs of a CGE model. However, some approaches may be unsuitable for integrating into a mathematical program. Supervised learning techniques that generate functional forms are likely more appropriate. Lazy learners (like k-nearest neighbors and locally weighted regression) lack functional forms, relying on stored training data until an explicit request for information is received. This requires significant memory and makes predictions more difficult than with eager learners, such as ordinary least squares (OLS) regression, decision trees, or neural networks.

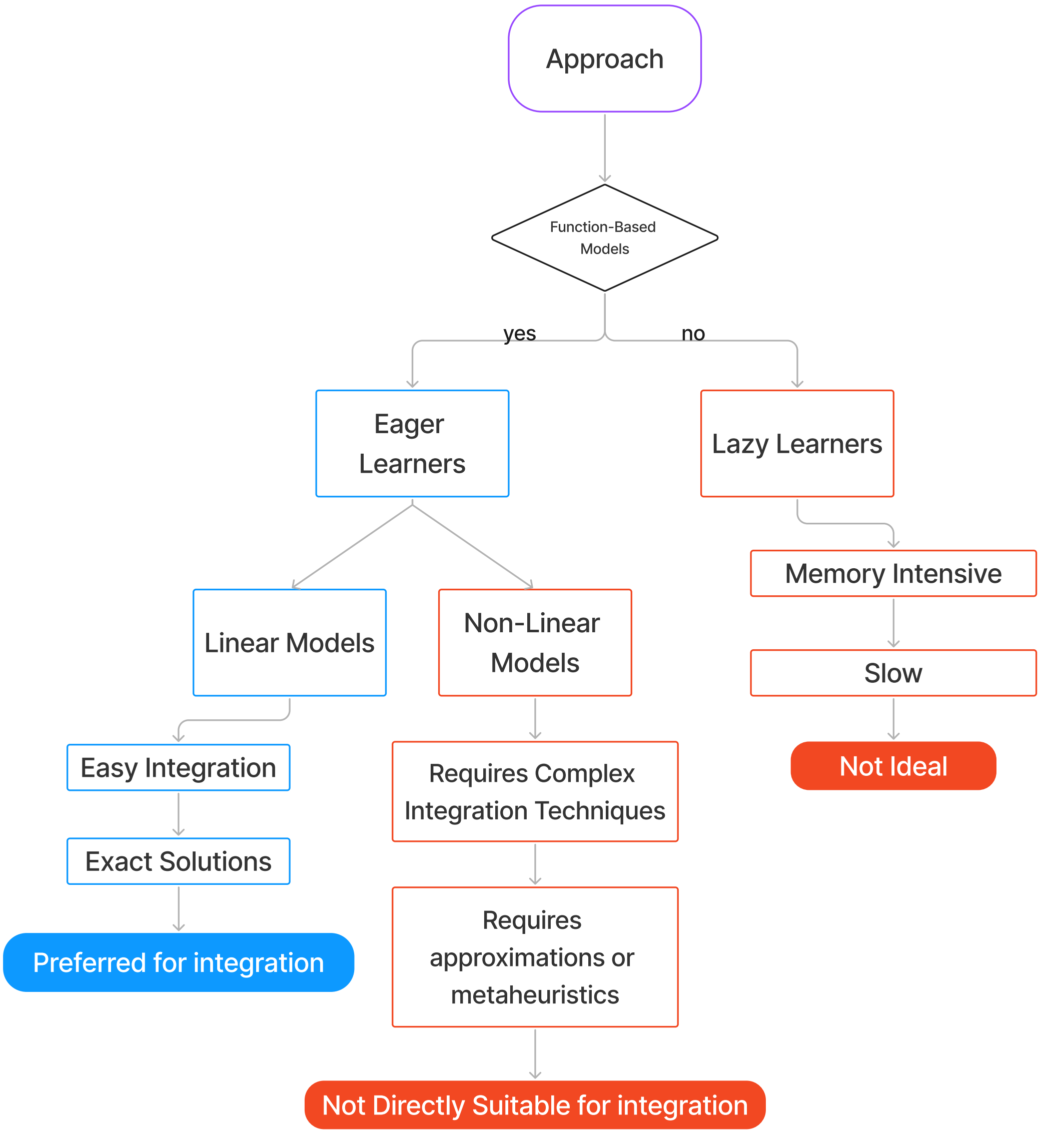


Figure 1: Choosing the right ML approach

3.1.2 Consistency of Inputs and Decision Variables

To effectively link the CGE model with the ML surrogate and optimization models, inputs and decision variables must align. For example, if shocks to capital stock are considered at sector level in the CGE model, then the surrogate model inputs cannot be more granular than this. Decision variables in the optimization model should ultimately address these shock values. Retrofitting buildings could serve as a decision variable if buildings contribute to an economic sector and retrofits reduce the impact on capital stock. Input features must be scaled consistently across models; for instance, if features are range-scaled for the ML model, then the optimization model's decision variables must follow this scale as well.

3.1.3 Functional Structure of the Model

An effective ML surrogate model needs to be both accurate and structured in a way that fits the optimization model's requirements. A linear model, such as one created using OLS regression, can support linear or integer programming. However, using nonlinear transformations (like products, ratios, or box-cox transformations) may prevent the model from being compatible with linear programs. Nonlinear models like neural networks or support vector machines with complex kernels would require either nonlinear programming or advanced metaheuristics to approximate solutions. Other nonlinear models like decision trees or gradient boosting methods may require rule-based logic and would likely need metaheuristic optimization techniques.

3.1.4 Predictive accuracy

The ML model must provide accurate approximations of the CGE model for scenarios related to decision-making. While simple surrogate models are often desirable, they are ineffective if they lack predictive accuracy needed for decision support. If advanced ML models also fall short in accuracy within the decision-making domain, then a surrogate model may not be feasible.

3.1.5 Interpretability

Since the surrogate model will integrate directly into an optimization framework, it's crucial for the model to be both predictive and meaningful. For instance, an OLS model with highly correlated inputs may predict well, but due to high variance, its parameter estimates may lack interpretability. This could lead to counterproductive decisions in the optimization process, where increasing damage to a sector might be incentivized. Methods like LASSO, which may exclude one of two highly correlated features, could have high predictive accuracy but may limit the optimization model's effectiveness by excluding sectors that might have benefited from resources. Ensuring the parameter estimates are directionally accurate and appropriately scaled is essential to avoid irrational resource allocations.

3.2 Fundamental Terms and Techniques

Three fundamental computational models are rooted in this research as described above. Below is a general description of all key foundational elements used to build the framework.

3.2.1 CGE: Computable General Equilibrium

Computable General Equilibrium models are powerful numerical tools that simulates how an economy responds to changes, such as policy shifts, technological advancements, or exogenous shocks. That is, it can help predict how an entire economy - like that of a city or specific region - might react to changes, such as new policies, big economic shifts, financial crises, natural hazards etc. to simulate economic outcomes [27]. Economic outcomes in CGE models can refer to changes in production or supply, household income, employment and labor etc. These outcomes show how the events(shocks) might impact the economy. While most CGE models cover core economic outcomes like the ones mentioned earlier, the terminology for these outcomes may vary. Throughout this research, the CGE model deals with four core economic outcomes - changes in domestic supply is denoted as DDS, changes in income levels are denoted DY, changes in number of households is Migration or MIGT, and changes in employment is DFFD. DDS(Domestic Supply) measures how much the overall supply in the economy changes after a shock, DY(Income) reflects how household income levels shift due to economic changes, Migration tracks changes in the number of households moving in or out, and DFFD (Employment), shows how job numbers vary across different sectors and regions.

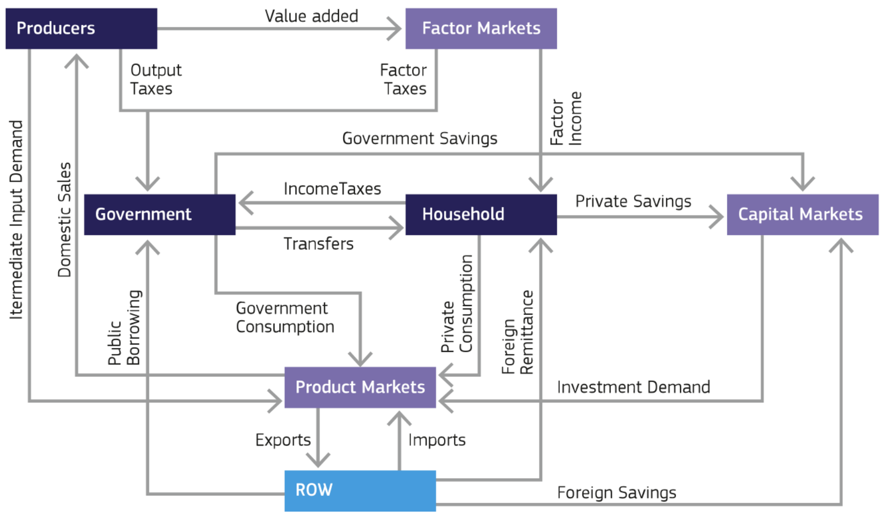


Figure 2: Overview of typical elements of a CGE model.(Helgeson, et al [1]) Note: “ROW” refers to “Rest of the world”.

The image above shows the core components and interactions in a CGE model, highlighting flows between producers, households, government, factor and product markets, capital, and the rest of the world (ROW), capturing the flow of goods, services, and financial transactions in the economy.

Before CGE models, economic modelling of natural hazards typically relied on Input-Output (I-O) models, such as the Leontief models. Traditional I-O models represents economic activities in fixed, linear relationships. This is intuitively easier to grasp, however, I-O models are only good at capturing demand-side impacts but struggles with supply-side impacts.

Unlike I-O models, CGE models incorporate both utility-maximizing households and profit-maximizing firms to capture both demand and supply dynamics effectively [35]. Both households and firms can make decisions to adjust to changes in prices, income, and other economic factors. Households tries to maximize utilities, for example, if the prices of a certain product increases, a household might reduce spending on that product and allocate more towards other necessities. Firms, on the other hand, try to maximize profit. This flexibility helps to capture the real-world economies, particularly when examining diverse economic shocks, such as natural disasters, but it also makes them highly complex and non-linear in nature. The Social Accounting Matrix (SAM) is crucial in organizing the data needed to build these complex models. It provides a detailed record of economic transactions between households, firms, the government, and external sectors, and serve as the core of CGE analysis. [1].

3.2.2 Machine Learning

Machine Learning is a subset of Artificial Intelligence that allows to systems to learn patterns in data using various algorithms [42]. It encompasses multiple approaches: supervised learning, where models are trained on labeled data to make predictions; unsupervised learning, which identifies hidden patterns in unlabeled data; and reinforcement learning, where agents learn optimal behaviors through trial and error in dynamic environments. In this study, supervised learning is applied, training models on labeled data to predict outcomes from the CGE model. Given the need to eventually integrate these predictions into optimization frameworks, we prioritize models that provide interpretability and maintain simplicity, leaning toward linear models over more complex, non-linear approaches.

Ordinary Least Squares (OLS) Regression: Ordinary Least Squares (OLS) is a foundational method in statistics and machine learning .It is one of the simplest and most widely used regression methods, that helps to understand the relationship between variables. It works by fitting a line through the data aiming to minimize the squared difference between actual values y and the predicted values $\hat{y}$. OLS can be described by the following equation [43]:

$$\text{minimize} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \tag{1}$$

While valuable for its simplicity, OLS has limitations when applied to high-dimensional data, like the complex outputs from CGE models, where high feature counts or multicollinearity among features can lead to overfitting, unstable predictions, and difficulty with interpretations.

Penalized Regression: Penalized regression models builds on OLS by adding a penalty to the model. This penalty term controls model complexity and helps to avoid overfitting when dealing with high-dimensional data, taking care of the limitations of OLS discussed above.

Penalized regression is particularly suited for this study's needs, where model stability, simplicity, and interpretability are critical for integrating predictive outputs into optimization frameworks. Among these methods, Elastic Net, is well-suited for our purpose as it allows effective feature selection while maintaining a linear structure and helps in seamless integration with the framework.

The penalized regression function is expressed as Hastie, Tibshirani, and Friedman [44]:

$$\text{minimize} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 + \lambda \left(\alpha \sum_{j=1}^{p} |\beta_j| + (1 - \alpha) \frac{1}{2} \sum_{j=1}^{p} \beta_j^2 \right) \tag{2}$$

The first term represents the OLS objective, minimizing the difference between actual and predicted values. The second term, the penalty, is controlled by the regularization parameter $\lambda \geq 0$, which adjusts the strength of regularization. When $\lambda = 0$, the equation reduces to standard OLS. As λ increases the impact of the penalty increases. The parameter α determines the type of penalty: setting $\alpha = 1$ results in LASSO (Least Absolute Shrinkage and Selection Operator), applying an L1 penalty that encourages sparsity by allowing some coefficients to be exactly zero. Setting $\alpha = 0$ leads to Ridge regression, which applies an L2 penalty that shrinks coefficients but does not zero them out, helping to manage multicollinearity in high-dimensional data. When $0 < \alpha < 1$, the model is Elastic Net, a blend of LASSO and ridge regression, balancing feature selection and coefficient shrinkage.

This approach helps build interpretable models that work well with integer linear optimization frameworks. Elastic Net offers control over feature selection and multicollinearity and can be used directly within the resilience framework, ensuring compatibility with decision-making systems without adding non-linear complexities.

3.2.3 Optimization Models

Optimization models are an essential tool in decision-making frameworks. They help to identify the best solution for a specific goal or objective while satisfying various constraints. These models are adaptable and can take multiple forms depending on the nature of the problem, objective functions, and constraints involved. Broadly, they can be categorized into linear programming (LP), integer programming (IP), integer linear programming (ILP), mixed-integer linear programming(MILP) and nonlinear programming (NLP), among others. Let's consider the general form of a single-objective optimization model:

$$\text{minimize} \quad f(x) \tag{3}$$

subject to:

$$g_i(x) \leq b_i, \quad \text{for } i = 1, 2, \ldots, m \tag{4}$$

$$x \geq 0 \tag{5}$$

Here, $f(x)$ represents the objective function to be minimized (or maximized), and $g_i(x) \leq b_i$ denotes the set of constraints that define the feasible region for the decision variables x. The decision variables symbolize the choices or actions available, such as the number of buildings to retrofit in disaster-prone areas. Constraints are conditions the solution must satisfy, such as cost or budget limits, resource availability or specific defined criteria, ensuring that the solution is feasible and practical within the context. Parameters within the model are fixed values that influence the optimization process, such as costs associated with retrofitting, the estimated economic losses or the coefficients derived from ML models.

This general framework is used to solve multiple types of problems, where the constraints are often expressed in a standard matrix form for simplicity and computational efficiency:

$$Ax \leq b \tag{6}$$

A is a matrix representing the coefficients of the constraints, x is the vector of decision variables, and b is the vector of bounds or limits on the feasible region. This is the foundational representation for various types of optimization models, each of which modifies or extends this structure to solve specific requirements and complexities.

Linear programming is the most straightforward application of this framework, where both the objective function and constraints in eq(3) and eq(4) are linear. It is computationally efficient, with solution methods like the Simplex algorithm or Interior Point methods providing reliable results even for large-scale problems. However, LP assumes that decision variables are continuous, which makes it unsuitable for problems involving discrete decisions, such as determining the number of buildings to retrofit.

To handle discrete decisions, integer linear programming introduces constraints that requires decision variables to be integers, i.e, $x \in \mathbb{Z}$. ILP is particularly useful when decisions are inherently discrete, as it is in this study. Compared to LP, ILP problems are generally more challenging as the solution space becomes discrete, often non-convex, where the feasible region is composed of isolated points. This makes ILP computationally more demanding.

Mixed-integer linear programming builds upon ILP by allowing some decision variables to remain continuous while others are restricted to integers. This flexibility makes MILP valuable for problems that combine discrete and continuous decisions, such as optimizing retrofit strategies alongside continuous resource allocation. The constraints remain linear, but the hybrid feasible region increases computational complexity. However, since our study only deals with discrete decision variables, MILP is irrelevant in this case.

Nonlinear programming extends this framework by allowing the objective function or constraints in eq(3) and eq(4) to become nonlinear, with an added constraint:

$$h_j(x) = 0 \quad \text{for } i = 1, 2, \ldots, p \tag{7}$$

Nonlinear systems often involve exact relations, which can be represented using equality constraints expressed in eq(7).

Nonlinear systems often involve exact relations, represented using equality constraints expressed in eq(7). NLP can deal with a comprehensive range of real-world problems due to its flexibility in handling nonlinear relationships. However, solving NLP problems is often computationally intensive and may not guarantee a globally optimal solution. It can also be limiting when quick and reliable solutions are required, such as in resilience planning under time constraints.

3.2.4 The Framework

Here is an overview of the framework that combines the components discussed so far to construct a reliable resilience framework.

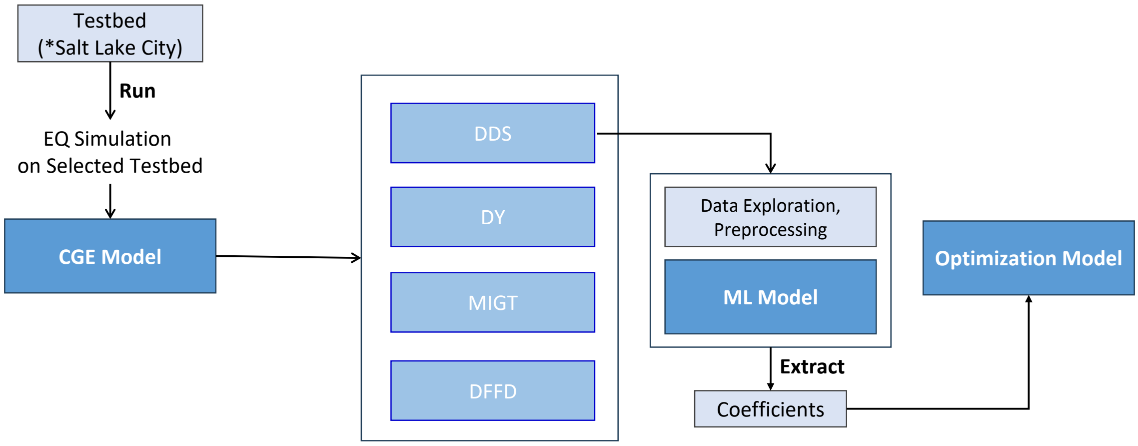


Figure 3: Overview of the CGE-ML Optimization Framework

Once a testbed is selected, multiple hazard simulations—in this case, earthquake scenarios—are generated for the chosen location. These simulations are fed into the CGE model to analyze the economic impact on various sectors of Salt Lake City across the region. The CGE model calculates shifts in key economic outcomes like employment, household income, migration and domestic supply in response to the earthquake shocks. These outcomes together provide an in-depth look into how a hazard might affect a city's economy and population.

Following the CGE simulations, the output data undergoes exploration and preprocessing before feeding into a machine-learning model. The ML model is trained on these CGE outcomes, allowing it to learn patterns and relationships within the data. By extracting coefficients from this model, we quantify the effects of different sectors on each outcome. These coefficients are then passed into an optimization model for finding optimal building retrofit strategies.

This allows us to use predictive insights from ML within a structured optimization framework, ensuring that each decision—such as building retrofits or resource allocation—aims to minimize negative impacts in a practical, data-driven manner. The framework thus combines predictive accuracy and decision-making clarity to support resilience planning tailored to the specific attributes of the testbed.

3.3 Application

3.3.1 Case Study: Salt Lake City - A general Overview

Salt Lake City is Utah's capital and most populous city. It is very close to the Wasatch Fault Zone (WFZ), one of the longest and most seismically active fault lines in the Intermountain West, extending about 240 miles from Fayette, central Utah, to Malad City, Idaho [45].According to Earthquake Engineering Research Institute [46], over 75 percent of Utah's economy is primarily based in Salt Lake, Utah, Davis, and Weber counties along the Wasatch fault, and it is also home to a high population density, with nearly 80 percent of the state's population living within 15 miles of the Wasatch Front.

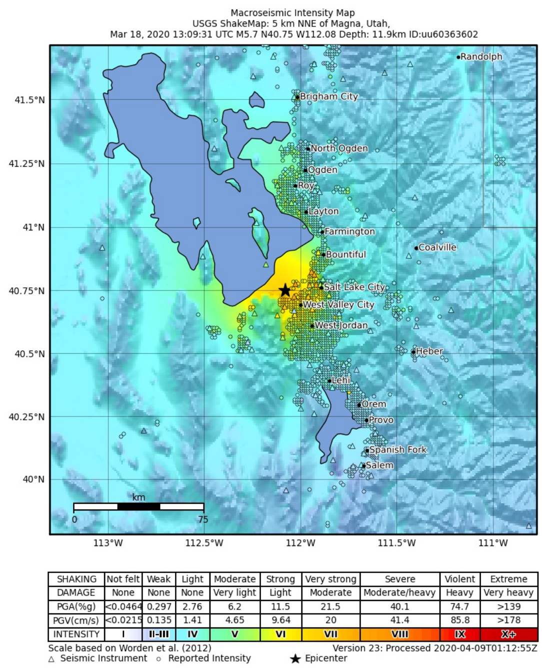
In the last 6000 years, there have been records of at least 22 earthquakes of magnitude over 7.0 or higher (commonly referred to as "Big Ones"), once every 300 years on average, along the five central segments of WFZ between Brigham City and Nephi [46]. On the Salt Lake City segment, the chances of "Big Ones" occurring have a repetition gap, averaging about 1,300–1,500 years, with the last one occurring around 1,400 years ago [46]. Therefore, it has been relatively quiet and very close to reaching the end of the gap. This quiet period, known as a seismic gap, raises concerns about a large earthquake shortly.

According to the Utah Seismic Safety Commission (USSC) [47], in the next 50 years, the Wasatch Front will face an earthquake of magnitude 6.75 or higher with a likelihood of 43% or an earthquake with magnitude 6.0 or higher with a probability of 57%, underlining the urgent need for preparedness in the region.

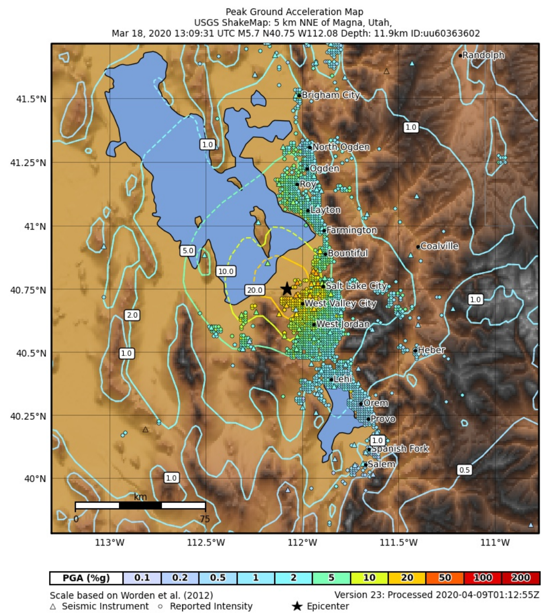
Figure 4: Wasatch Fault. (Source: Utah Geological Survey[48])

The Wasatch Fault Zone, as shown in Figure 4, stretches all the way from Fayette to Malad City—visible in red on the map. It is divided into distinct segments, each of which can produce independent seismic activity. These segments are near major areas like Salt Lake City, Ogden, and Provo, which increases the fault's potential risk to Utah's populated areas.

Only four years ago, on March 18, 2020, a magnitude 5.7 earthquake struck near Magna, Utah, around 5 km northeast of the town and close to Salt Lake City (SLC) [49]. This earthquake was the largest seismic event in the area in decades and was felt widely across the region. Additionally, it caused structural damage, particularly to older and unreinforced masonry buildings, and temporarily disrupted infrastructure such as power and transportation.



(a) Intensity of EQ of 5.7 Magnitude in Magna on March 18th, 2020



(b) PGA

Figure 5: Intensity and PGA. (Source: U.S. Geological Survey (USGS)[49])

The earthquake's effects are shown in two types of maps: the Intensity Map (a) and the Peak Ground Acceleration (PGA) Map (b).

The first figure(a) depicts the level of shaking felt by residents around the affected areas using color gradients from light blue (weak) to yellow(strong) and orange (quite strong) near the epicenter.

The second figure (b) gives a quantitative measure of ground movement as a percentage of gravity. Higher PGA values, shown in yellow and green near the epicenter, indicate significant ground forces that could impact buildings structurally. The PGA values decrease further away from the epicenter and show reduced impacts on distant areas.

3.3.2 Data Generation for CGE and ML Models

Salt Lake City is geographically divided into seven regions and comprises 11 unique economic sectors—nine commercial sectors and two housing services. The building inventory includes approximately 285,000 structures, with data from the IN-CORE data service (ID: 62fea288f5438e1f8c515ef8). Most buildings are either unreinforced masonry (URML) or wood frame (W1) constructions. Residential buildings constitute 90.6% of the total buildings, while commercial buildings comprise 9.3%.

Earthquake Scenario Development A series of earthquake simulations are developed to assess the potential economic impacts of earthquakes on SLC by civil engineers. A Ground Motion Prediction (GMP) model estimates the shaking intensity of an earthquake based on magnitude, depth, and location. An initial set of 38 ground GMP models is utilized to simulate earthquake events involving all buildings in SLC. These models encompass a range of earthquake magnitudes from 5.5 to 7.3 and consider four epicenters at varying distances from the city. Using fragility curves specific to each building type, the simulations estimate the remaining functional capital stock post-earthquake at the individual structure level. These results are then aggregated by sector and region to determine the proportion of remaining capital stock for each of the 70 sector-region combinations.

Based on the initial 38 GMP simulations, 5000 more damage scenarios are created using a statistical approach to generate a comprehensive dataset used for CGE modeling and the machine-learning training dataset. A critical assumption in preparing this data is that all buildings were uniform before the earthquakes, i.e., no additional mitigation strategies or retrofits beyond existing building codes have been implemented. This assumption may lead to similar damage patterns across sectors due to consistent vulnerability levels. However, to build an effective machine learning model, it is crucial that the model is trained on a vast range of potential outcomes with variability. Therefore, the shocks generated with this approach are uncorrelated across simulations to ensure we capture various potential outcomes while remaining consistent with the general scale and footprint of the original GMP model outputs.

For each sector-region combination $i \in I$, the minimum min_i and maximum max_i damage values from the GMP simulations were determined. These bounds were slightly adjusted by a factor γ (where $0 \leq \gamma \leq 1$) to enhance variability:

$$min_i' = (1 - \gamma)min_i \tag{8}$$

$$max_i' = (1 + \gamma)max_i \tag{9}$$

The adjusted bounds min_i' and max_i' are then used to generate random damage values for each of the 5,000 scenarios, ensuring a vast and diverse set of examples for the machine learning model.

A damage value of 1.0 for a sector-region combination suggests no loss, whereas a value of 0.5 represents a 50% reduction in capital stock. In this study, a value of $\gamma = 0.05$ is chosen. For each sector-region combination $i \in I$, capital shocks are generated by sampling from a uniform distribution $U(min_i', max_i')$. If any sampled values exceed 1.0, they are restricted to 1.0.

CGE Input for Economic Modeling The generated 5,000 earthquake damage scenarios are then used as inputs to the CGE model for economic impact analysis. The CGE model is run 5,000 times using GAMS, focusing on the overall impacts of domestic supply, employment, real household income, and migration. As shown in Table 1 below, the data represents a sample of 5000 damage scenarios for all 70 sector-regions in SLC.

Table 1: CGE Input Data - 5000 Damage Scenarios

Sector-Region/Scenarios	1	2	3	.	5000
ART_ACC_R1	0.828	0.941	0.514	.	0.669
COMMER_R1	0.953	0.921	0.575	.	0.960
CONS_R1	0.863	0.581	0.508	.	0.777
EDU_R1	0.433	0.859	0.977	.	0.473
HS1_R1	0.968	0.865	0.819	.	0.864
HS2_R7	0.771	0.781	0.923	.	0.873
HEALTH_R1	0.799	0.832	0.461	.	0.652
MANU_R1	0.883	0.773	0.430	.	0.984
RELIG_R1	0.808	0.504	0.666	.	0.530
UTIL_R1	0.900	0.610	0.711	.	0.804
.	.	.	.	.	.
.	.	.	.	.	.
UTIL_R7	0.758	0.783	0.682	.	0.922

Machine Learning Model Data The 5000 damage scenarios are structured as a matrix, with each row representing different damage scenarios and columns corresponding to the 70 sector-region combinations. The damage values are converted from percent capital stock remaining to percent capital stock lost and scaled by the base capital stock to transform them into dollar values (unit in $M) for better comprehension. Expressing losses in monetary terms enhances interpretability, practical analysis, and evaluation of direct comparisons of economic losses across sectors and regions.

These damage values are used as inputs for the machine learning (ML) model, remaining consistent with the inputs used in the CGE model. In the ML model, each sector-region combination serves as a predictor variable, while the corresponding total economic outcome from the CGE model—such as DDS, DY, DFFD, or MIGT—is the target variable. Together, the 5000 observations of sector-region shocks and their CGE analysis outputs form the training dataset for the ML model.

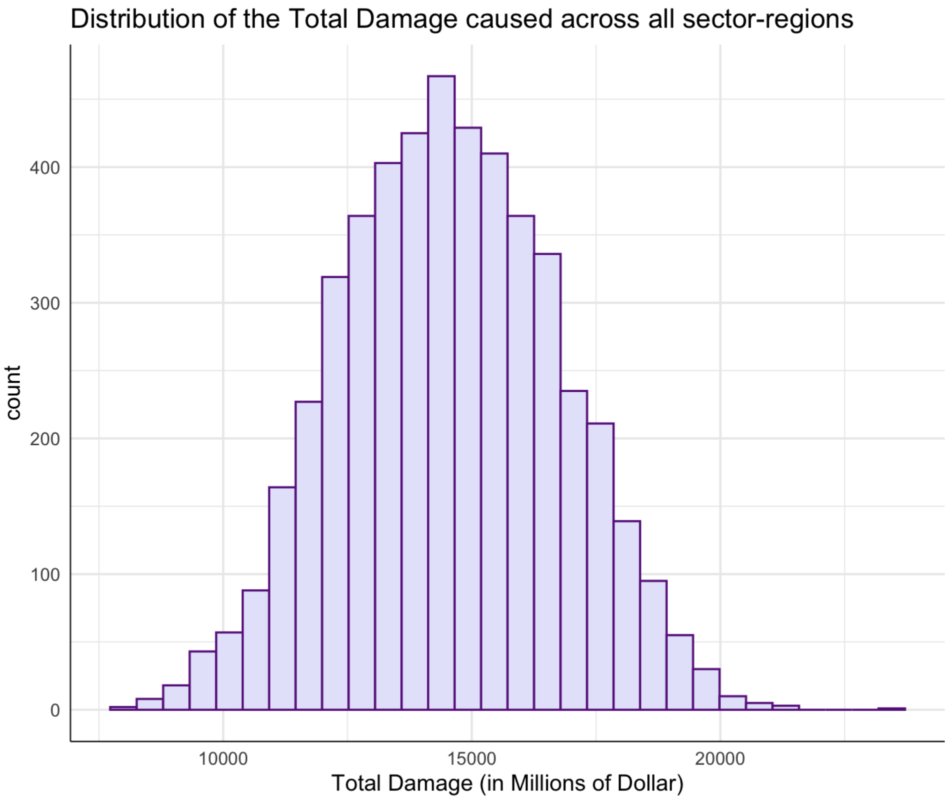
3.3.3 CGE Model Data Analysis

The CGE model provides a structured framework to evaluate the economic impacts of disasters across Salt Lake County (SLC). It divides the county into seven regions and captures the changes in fundamental economic factors discussed before - DDS, DY, migration, and DFFD.

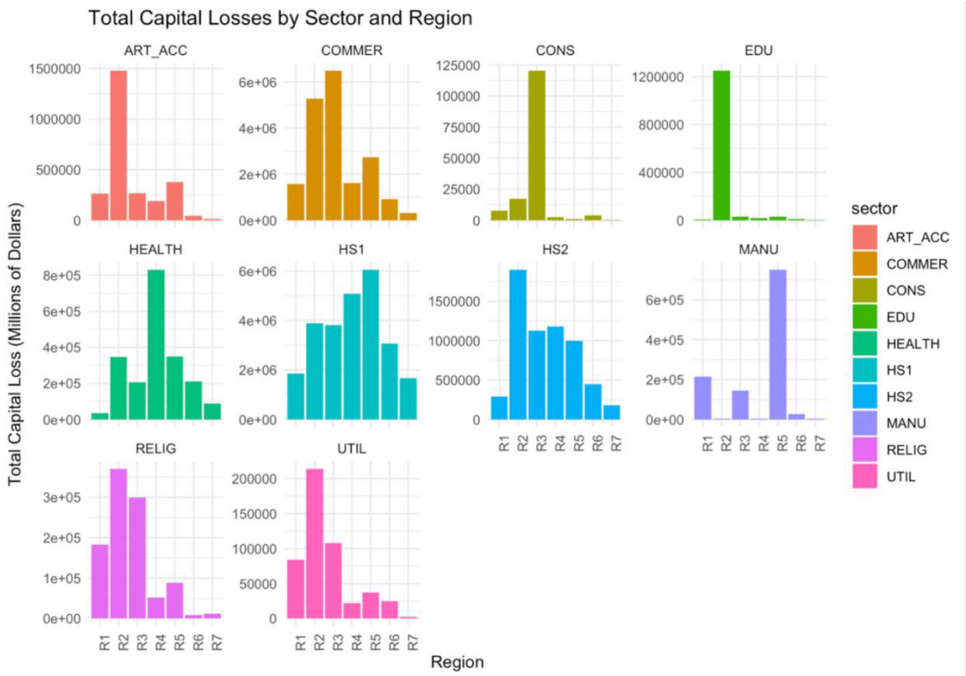
The model explores nine key sectors. These sectors include agriculture and mining, utilities, construction, manufacturing, commerce (encompassing wholesale and retail trade, transportation, and professional services), education, health care, arts, entertainment, accommodations, and religious organizations.

Additionally, the CGE model incorporates household groups divided by income levels and housing types. Households are categorized into four income groups, from lower-income (HH1) to higher-income (HH4), while housing is split between single-family homes (HS1) and multi-family units (HS2). This granularity is what allows the model to explore not only the direct economic losses, such as damage to physical capital, but also broader distributional impacts, like changes in employment and income.

Taking a look into the total capital stock loss for one of the output: DDS from the CGE model below:



(a) Distribution of Total Capital Stock Loss Overall



(b) Total Capital Stock Loss by sector and region

Figure 6: Distribution of Total Capital Stock Loss

Figure (a) on the left shows the histogram of total simulated disaster damages across sector-regions, with most values centered around 15,000 million dollars, indicating most scenarios produce moderate losses. A right-skewed tail extending to 20,000 million dollars shows some high-impact cases, suggesting the potential for extreme events with high economic damage. Figure (b) on the right breaks down total capital losses by sector across regions (R1-R7), revealing that regions R1, R2, and R5 consistently face higher losses across sectors.

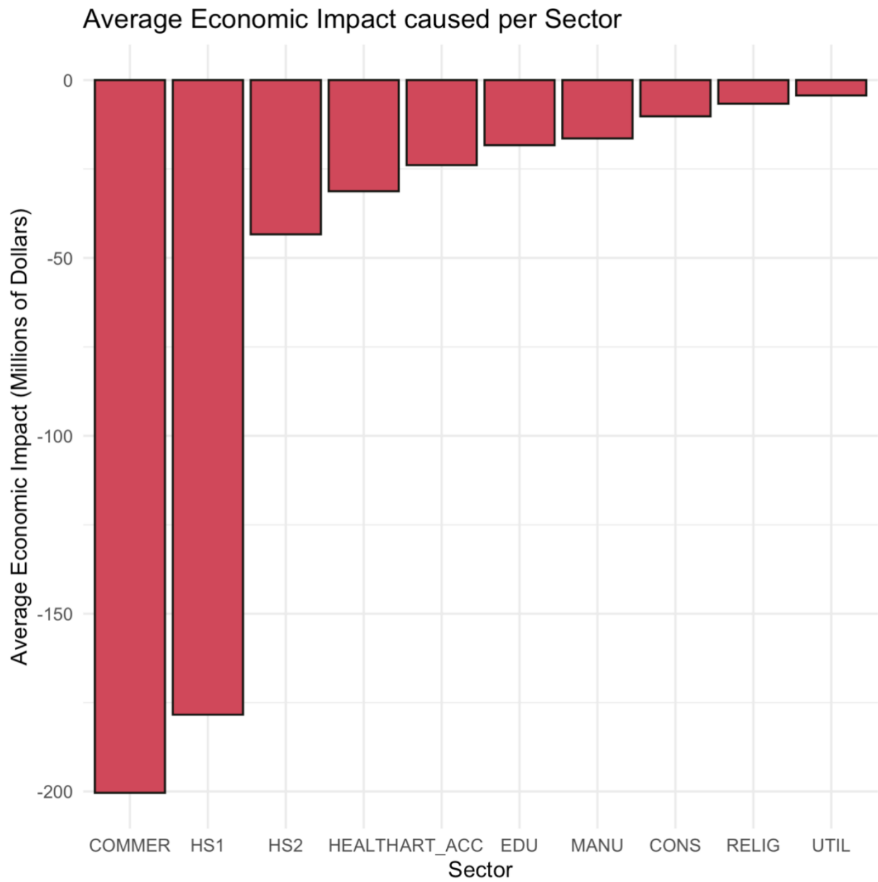


Figure 7: Impact per Sector(DDS)

Figure 4 shows the average economic impact per sector, measured in millions of dollars for DDS. It is evident that the Commercial and HS1 sectors face the highest average economic losses (around -200 million dollars).

3.3.4 ML Model Development and Training

This section details model selection, development, and training. As discussed in Section 3.1, we are leaning more towards straightforward, linear models rather than non-linear ones that require approximations or lazy learners that lack functional forms and rely on stored training data. Some fundamental concerns regarding finding the best model for this study are that it needs to be efficient, accurate, and meaningful. One of the primary concerns is that it needs to integrate seamlessly into an optimization framework, specifically, an ILP model.

We briefly discussed Ordinary Least Square regression in Section 3.2.2 due to its simplicity and widespread use in statistical modeling. However, in the context of our study, which involves high-dimensional data derived from CGE models, OLS is not the ideal choice as it cannot deal with feature selection and multicollinearity and can lead to unstable predictions.

Dimensionality reduction techniques, including Principal Component Regression (PCR) and Partial Least Squares (PLS) regression, were considered to address these issues. Both methods extract linear combinations of the original predictor variables to project the data into a lower-dimensional space. In PCR, Principal Component Analysis (PCA) is first performed on the predictor data to identify orthogonal components that capture the maximum variance. A subset of these principal components is then selected for the regression model, which reduces dimensionality and addresses multicollinearity issues. However, PCR is inherently unsupervised—it solely considers the predictor variables without incorporating the response variable. Thus, it may not capture the relation between the predictor and response variable.

On the other hand, PLS regression constructs its components based on both the input (predictors) and output (response) data through a supervised learning approach. While both methods are similar in practical performance, PLS specifically tailors these components to enhance the prediction of the response variable. Therefore, between PCR and PLS, PLS is a better candidate for the surrogate model due to its ability to capture the relationship between predictors and the response variable effectively.

Despite its strengths, effectiveness in handling high-dimensional data and multicollinearity and having high model performance, PLS shows difficulty in interpretability. Each component of PLS is a linear combination of all predictors, making it challenging to pin down specific effects to specific predictors.

Given the need for good predictive performance, as well as, good interpretability, elastic net regression was an alternate choice to consider. In Section 3.2.2, we defined elastic net as an extension of OLS that introduces L1 (LASSO) and L2 (Ridge) regularization penalties, allowing to control model complexity and improve reliability simultaneously. The L1 penalty performs the variable selection by shrinking some coefficients to zero, while the L2 penalty ensures that groups of correlated predictors are maintained together, effectively handling multicollinearity. This combination of L1 and L2 penalty aligns well with our objective of deriving a linear model that balances predictive power with clear coefficient interpretability and meaning.

A 5-fold cross-validation (CV) procedure with 10 repeats is utilized to evaluate PLS and elastic net. Cross-validation is a resampling method used to estimate the performance of a model on unseen data. In a 5-fold CV, the data is divided into five subsets (folds), where the model is trained on four folds and tested on the remaining fold, repeating this process for each fold. The results are then averaged to provide a robust estimate of model performance. Repeating this process 10 times ensures that the performance metrics are stable and reliable.

The performance of PLS and elastic net for predicting Delta Domestic Supply (DDS) is shown in Table 2 below.

Table 2: Performance Metrics for DDS Prediction Using PLS and Elastic Net

Model	Hyperparameters	RMSE	$\mathbf{R^2}$	MAE
PLS	$ncomp = 6$	68.34	0.9834	54.84
Elastic Net	$\alpha = 0.275$, $\lambda = 1.1684$	25.07	0.9978	20.33

The above table summarizes the performance metrics Root Mean Square Error (RMSE), R^2, and Mean Absolute Error (MAE) for both models using the optimal values of hyperparameters. PLS uses only one hyperparameter - "ncomp", the number of components, while elastic net uses two hyperparameters α for balancing the regularization parameters L1 and L2, and λ for controlling the strength of regularization. Elastic net outperforms PLS across all metrics, suggesting a more accurate prediction of DDS.

Aside from having good performance, elastic net retains a direct mapping between predictors and response variables, unlike PLS as mentioned earlier. It allows for drawing meaningful relationships between variables without sacrificing predictive accuracy, making it more suitable for applications requiring clear insights into variable importance. The model performance of the elastic net is excellent for all outcomes of the CGE model, which we'll examine further in the next chapter.

Elastic Net is the most promising candidate as it aligns well with our goals. It is a linear model with high accuracy and offers interpretability. It handles high-dimensional data like ours and can be easily integrated into a linear optimization framework without introducing nonlinear complexities or exploring heuristics and approximate solutions. Thus, it is a strong choice for modeling the complex outputs from CGE models.

3.3.5 Development of Community-Wide and Sector-Specific Models

With Elastic Net regression selected as the modeling approach, we developed two types of models: community-wide and sector-specific.

Community-Wide Models The community-wide models aim to predict the overall economic impact of earthquakes on Salt Lake City by considering the aggregated effects across all sectors. These models use capital stock losses from all sectors as predictors to estimate total economic outcomes, such as total change in domestic supply (totDDS), total income change (totDY), total migration (totMIGT), and total employment change (totDFFD). Community-wide models provide a holistic view of economic vulnerabilities by focusing on the collective impact.

Sector-Specific Models Sector-specific models, on the other hand, enable a more granular and detailed viewpoint. These models predict the economic outcome for a particular industry, using capital stock losses from all sectors as predictors. It helps to understand the relationship and dependencies between sectors, demonstrating how damages in one sector can influence the performance of others. For example, a sector-specific model for the manufacturing sector would predict the change in domestic supply for manufacturing based on capital losses not only in manufacturing but also in other sectors like construction or utilities.

3.3.6 Optimization Model Integration

The optimization framework builds on the predictive insights provided by the machine learning models. The framework allows for precise and efficient resource allocation for retrofitting building codes in SLC by incorporating ML-derived coefficients into the optimization problem. The objective function, decision variables, constraints, and other aspects of the framework formulation are covered in detail in this part and illustrate how adding ML coefficients improves the model's performance.

Model Input Data Parameters

The decision support framework is built for SLC, where the earthquake simulated magnitude 7.1 with an epicenter in East Bench. It is structured around the sets sector-regions (Z), structure types (S), and retrofitting levels (K). These sets define the model's granularity and ensure that all input data align with this level of detail. The set Z (sector-region) is consistent with the sector-region used in the CGE and ML models. This consistency is essential for the seamless integration of predictive coefficients, which is vital for maintaining the reliability and validity of the results. The set structure type S defines the category a building belongs to based on its design, material, or even purpose - residential, commercial, or industrial buildings. The dataset used for this model has approximately 279,861 SLC buildings after data pre-processing and cleaning. Among these, there are 33 distinct structure types for buildings in SLC. Most of the buildings belong to W1 (Woodframe 1), URML (Unreinforced Masonry Low-Rise), RM1L (Reinforced Masonry 1 Low-Rise), or RM2L (Reinforced Masonry 2 Low-Rise). W1 alone comprises approximately 40% of the buildings in SLC, URML consists of around 30%, RM1L 18%, and RM2L 11%. The Set K, retrofit levels, defines the current or enhanced building codes, categorized into four levels: "Pre-Code (PC)," "Low-Code (LC)," "Mid-Code (MC)," and "High-Code (HC)."

Table 3 details an overview of the building inventory at baseline with the number of buildings b_{ijk} in sector-region $i \in Z$, structure type $j \in S$, retrofit level $k \in K$. For example, from the table, HS1_R4 is the sector HS1(Housing Service 1) in Region 1, with structure type URML, in retrofit level at PC currently has 18,317 buildings. This number is the highest number of buildings belonging to a specific structure type in a particular sector-region in our dataset.Table 4 shows the direct economic loss for buildings in sector-region $i \in Z$, structure type $j \in S$, retrofit level $k \in K$. These losses are estimated by a building's vulnerability to hazards modeled using fragility curves. Fragility curves quantify the likelihood of damage under specific hazard scenarios. These likelihood values also depend on the building structure type and the current retrofit level. Table 5 specifies the strategy cost required for upgrading a building in retrofit level $k \in K$ to retrofit level $k' \in K'$.

Table 3: Building Inventory Before Retrofitting

Sector-Region ($i \in Z$)	**Structure Type** ($j \in S$)	**Retrofit Level** ($k \in K$)	**Baseline Building Count** (b_{ijk})
ART_ACC_R1	RM1L	MC	67
ART_ACC_R1	RM2L	HC	14
ART_ACC_R1	S2L	MC	7
ART_ACC_R2	C2L	MC	25
ART_ACC_R2	URML	PC	219
ART_ACC_R7	URML	PC	2
COMMER_R1	RM1L	MC	214
COMMER_R1	URML	PC	156
COMMER_R1	W1	MC	27
CONS_R1	C2L	MC	1
CONS_R1	S2L	LC	9
EDU_R2	RM1L	PC	188
EDU_R2	RM1L	LC	10
HS1_R3	W1	MC	15493
HS1_R4	RM1L	MC	3067
HS1_R4	URML	PC	18317
HS2_R3	W1	PC	1
HS2_R3	W1	LC	260
HEALTH_R1	RM1L	MC	9
HEALTH_R3	RM1L	MC	36
.	.	.	.
.	.	.	.
UTIL_R6	W1	MC	13
UTIL_R7	RM1L	MC	6

Table 4: Direct Economic Loss Values

Sector-Region ($i \in Z$)	**Structure Type** ($j \in S$)	**Retrofit Level** ($k \in K$)	**Direct Economic Loss** (L_{ijk})
ART_ACC_R1	RM1L	MC	329821.043
ART_ACC_R1	RM2L	HC	325678.398
ART_ACC_R1	S2L	MC	253270.975
ART_ACC_R2	C2L	MC	3131382.22
ART_ACC_R2	URML	PC	142340.128
ART_ACC_R7	URML	PC	125412.337
COMMER_R1	RM1L	MC	436591.558
COMMER_R1	URML	PC	101168.14
COMMER_R1	W1	MC	24858.808
CONS_R1	C2L	MC	38327.041
CONS_R1	S2L	LC	14756.3957
EDU_R2	RM1L	PC	424774.426
EDU_R2	RM1L	LC	403222.998
HS1_R3	W1	MC	18264.2631
HS1_R4	RM1L	MC	47586.2071
HS1_R4	URML	PC	32813.8396
HS2_R3	W1	PC	23504.1307
HS2_R3	W1	LC	21288.4976
HEALTH_R1	RM1L	MC	257950.127
HEALTH_R3	RM1L	MC	238379.048
.	.	.	.

Table 5: Strategy Cost Data Example

Sector-Region ($i \in Z$)	**Structure Type** ($j \in S$)	**Retrofit Level** ($k \in K$)	**Enhanced Retrofit Level** ($k' \in K'$)	**Strategy Cost** ($SC_{ijkk'}$)
ART_ACC_R1	URML	PC	LC	17287.9162
ART_ACC_R1	URML	PC	MC	57626.3872
ART_ACC_R1	URML	PC	HC	74914.3034
ART_ACC_R1	URML	LC	MC	40338.4711
ART_ACC_R1	URML	LC	HC	57626.3872
COMMER_R3	RM1L	PC	LC	22731.7505
COMMER_R3	RM1L	PC	MC	75772.5018
COMMER_R3	RM1L	PC	HC	98504.2524
MANU_R5	C3L	LC	HC	5171455.6
MANU_R5	C3L	LC	MC	3620018.92
COMMER_R2	S1H	LC	HC	3550780.8
EDU_R2	C2M	PC	HC	3327837.5
COMMER_R3	C2H	MC	HC	3150826.8
EDU_R2	S1L	PC	HC	2799463.29
EDU_R2	C2M	PC	MC	2559875
HEALTH_R4	C2L	LC	HC	2189054.77
EDU_R2	S1L	PC	MC	2153433.3
MANU_R1	C3L	LC	HC	2114815.6
EDU_R2	C3M	PC	MC	1950028.57
EDU_R2	C3H	PC	HC	1946880
COMMER_R2	C3M	PC	HC	1852490.21
.	.	.	.	.

Formulation of the Optimization Problem

Extending the work by Gupta, Adluri, Sanderson, *et al.* [40] and Gupta, Nofal, González, *et al.* [26], my optimization framework builds on the predictive insights provided by the machine learning models. The framework allows for precise and efficient resource allocation for retrofitting solutions by incorporating ML-derived coefficients into the optimization problem. The objective function, decision variables, constraints, and other aspects of the framework formulation are covered in detail in this part and illustrate how adding ML coefficients improves the model's performance.

Decision Variable The model deals with two decision variables. The variable x_{ijk} represents the number of buildings in sector-region i of type j, and retrofitting level k post-intervention. The variable $y_{ijkk'}$ represents the number of buildings in sector-region i of type j, transitioned from retrofit level k to k'.

Objective of the Model The objective of the optimization model is to minimize the total change in economic outcomes across all sector-regions, structure types, and building retrofit strategies.

$$\min \sum_{i \in I} \sum_{j \in J} \sum_{k \in K} \beta_i \, l_{ijk} \, x_{ijk} \tag{10}$$

The parameter β_i represents the coefficients of economic outcomes (e.g., Delta Domestic Supply) for sector-region $i \in Z$, l_{ijk} is the direct economic loss for buildings of type $j \in S$ in sector-region $i \in Z$ at retrofitting level $k \in K$, and x_{ijk} , as defined earlier, is the decision variable indicating the number of buildings after retrofitting.

Constraints The model is subject to several constraints in terms of feasibility and practicality. Two new sets, denoted U and T, are introduced to streamline transitions and improve model performance. U is a set of valid retrofitting transitions. Each element of U is a tuple (i, j, k, k') representing a feasible transition from k to k' for buildings of type j in sector-region i. Set U only allows for practical and possible upgrades, significantly improving model efficiency by limiting the decision space to viable and realistic possibilities, lowering computing complexity, and focusing on actionable tactics. Similarly, set T is defined as the set of valid building states after retrofitting, where each element (i, j, k) represents a specific combination of sector-region, structure type, and retrofit level. The explicit definition of sets U and T improves computational efficiency by limiting the scope to valid scenarios and ensures that the model complies with structural and retrofit feasibility in SLC.

The total cost of retrofitting must not exceed the available budget B.

$$\sum_{(i,j,k,k')\in U} SC_{ijkk'}\, y_{ijkk'} \leq B \tag{11}$$

Here, $SC_{ijkk'}$ represents the cost of upgrading a building from level k to k, and $y_{ijkk'}$ is the number of buildings undergoing this transition.

The number of buildings in each retrofitting state after intervention must account for the initial building count and retrofitting transitions:

$$x_{ijk} = b_{ijk} + \sum_{k':(i,j,k',k)\in U} y_{ijk'k} - \sum_{k':(i,j,k,k')\in U} y_{ijkk'} \tag{12}$$

This constraint ensures that the final building count x_{ijk} accurately reflects retrofitting actions applied to the initial number of buildings b_ijk.

The total number of buildings in each sector-region and structure type must remain constant to ensure no buildings are lost or created during the retrofitting process:

$$\sum_{k\in K} x_{ijk} = \sum_{k\in K} b_{ijk}, \quad \forall i \in I, \forall j \in J \tag{13}$$

Lastly, all decision variables must be non-negative integer, as negative and fraction number of buildings or retrofit actions are not physically or realistically meaningful:

$$x_{ijk} \in \mathbb{Z}_{\geq 0}, \quad \forall\, (i,j,k) \in T \tag{14}$$

$$y_{ijkk'} \in \mathbb{Z}_{\geq 0}, \quad \forall\, (i,j,k,k') \in U \tag{15}$$

Incorporation of ML-derived Coefficients into the Optimization Model

Integrating coefficients β_i from the ML models forms the backbone of the optimization framework, linking predictive analytics with mathematical decision-making. These coefficients quantify the sensitivity of economic outcomes in sector-region i to retrofitting interventions. For instance, a higher β_i implies that sector-region i is more economically vulnerable to disasters than lower β_i, making it a priority for retrofitting efforts. By incorporating β_i into the objective function, the optimization model chooses or prioritizes retrofit actions that significantly reduce the negative impacts of an economic outcome for a specific sector region. Another way to think about this is the optimization model, which uses the ML coefficients as a guide to choose retrofit actions that have the most significant positive impact on reducing the negative impacts of an economic outcome for a specific sector region. Not only does the integration guide the decision-making process, but also allows the optimization model to adapt to changing circumstances. For instance, the economic conditions and hazard scenarios can change. In that case, the model can be retrained with updated data. By recalibrating the ML models with updated data, the coefficients β_i can reflect changes in economic dynamics and hazard scenarios. As a result, the optimization framework remains robust and relevant over time, offering dynamic and data-driven solutions for disaster mitigation planning. Additionally, β_i helps interpret optimization results with clarity and transparency as decision-makers can trace the rationale for recommended retrofit strategies to the underlying sensitivities of the economic outcome captured by the coefficients.

4 Results and Discussion

The key findings of this research are presented in two parts: the performance of the ML models in predicting critical outcomes such as Domestic Supply, Household Income, Migration, and Employment, and following it, the optimization results that utilize these predictions to minimize total disruptions. Together, these results emphasize the value of a hybrid approach that combines data-driven predictions with decision-oriented frameworks.

4.1 Elastic Net Regression Model Performance

The Elastic Net regression is our chosen model due to its simplicity, transparency, and ability to handle complex datasets with high precision. The results for both the community-wide and sector-specific models are discussed below.

4.1.1 Community-Wide Models

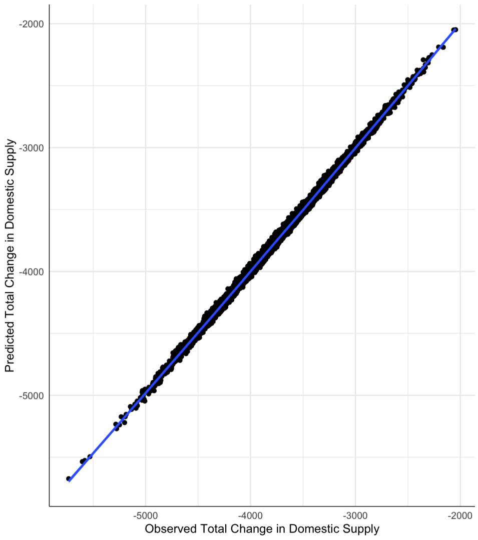
Table 6: Optimal Hyperparameters from 5-Fold CV with 10 Repeats

	DDS	**DY**	**MIGT**	**DFFD**
Alpha (α)	0.275	0.25	0.125	0.125
Lambda (λ)	1.169	0.72	1.473684	3

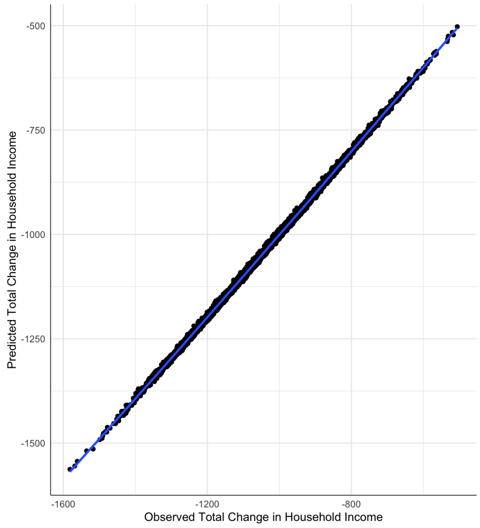
Table 7: Model Performance Metrics

	DDS	**DY**	**MIGT**	**DFFD**
RMSE	25.08106	5.382021	14.84268	72.0755
$\mathbf{R^2}$	0.9977903	0.9990747	0.9987012	0.9982809

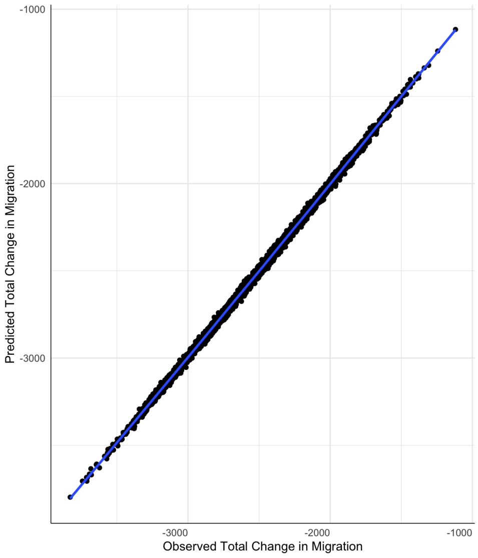
Table 6 shows the optimal values for α and λ for all outcomes. Table 7 summarizes the model performance for the chosen parameters. The RMSE values across all four outcomes (DDS, DY, MIGT, and DFFD) are remarkably low, indicating a solid understanding between predicted and observed values. For example, the RMSE for DDS is 25.08, while for DY, it is only 5.38, signifying minimal prediction deviation. The models' robustness and capacity to account for the great majority of variance in the outcome variables are confirmed by the R^2 values, which surpass 0.997.



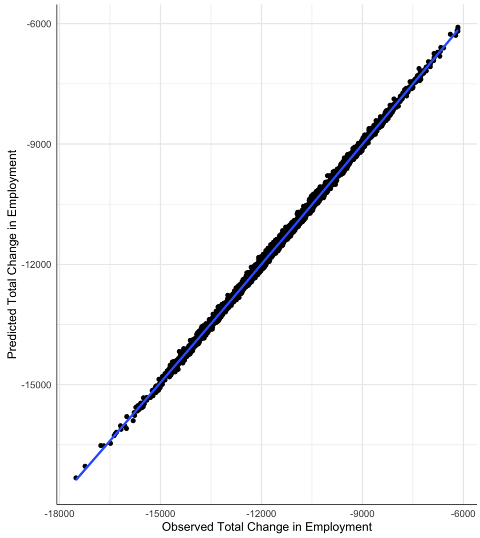
(a) Observed vs Predicted Total Change in Domestic Supply



(b) Observed vs Predicted Total Change in Household Income



(c) Observed vs Predicted Total Change in Migration



(d) Observed vs Predicted Total Change in Employment

Figure 8: Observed vs Predicted Changes

The above plots shows the actual vs predicted values for all outcomes.

Table 8: Community-Wide Coefficients for All Outcomes

Sector-Region	**Domestic Supply**	**Household Income**	**Migration**	**Employment**
(Intercept)	94.70476157	19.2803932	51.75368591	403.709859
ART_ACC_R1	-0.287883918	-0.0035188	-0.109027111	-0.940446
ART_ACC_R2	-0.28755274	-0.0396723	-0.071868419	-0.5519669
ART_ACC_R3	-0.265603553	-0.0324987	-0.074033468	-0.5120931
ART_ACC_R4	-0.270426297	-0.024522	-0.063925121	-0.4015873
ART_ACC_R5	-0.251441336	-0.0335814	-0.058659648	-0.3684516
ART_ACC_R6	-0.161053876	0	-0.085282681	-0.2773067
ART_ACC_R7	0	0	-0.327185221	-0.6672047
COMMER_R1	-0.279906684	-0.0058881	-0.089211018	-0.7698283
COMMER_R2	-0.278294827	-0.0452129	-0.069297226	-0.5206741
COMMER_R3	-0.248057808	-0.0451126	-0.080644908	-0.5466128
COMMER_R4	-0.252279283	-0.042042	-0.082202288	-0.4304422
COMMER_R5	-0.24405085	-0.0441968	-0.067234012	-0.4007471
COMMER_R6	-0.239123537	-0.0409094	-0.092094519	-0.4510846
COMMER_R7	-0.257225309	-0.0391481	-0.190873754	-0.7056949
CONS_R1	-0.840216991	0	-0.039865933	-2.6314792
CONS_R2	-0.163923516	0	0	-0.1648393
CONS_R3	-0.273974116	-0.054852	-0.063514666	-0.6108509
CONS_R4	0	0	0	0
CONS_R5	0	0	0	0
CONS_R6	0	0	0	0
CONS_R7	0	0	1.28915582	0
EDU_R1	0	0	0	0
EDU_R2	-0.234141415	-0.0406623	-0.058540094	-0.433874
EDU_R3	-0.035411428	0	0	-0.2129912
EDU_R4	-0.211278138	0	0	-0.4645992
EDU_R5	0	0	-0.03529706	-0.0885247
EDU_R6	-0.055452969	0	-0.226007432	-0.5430332
EDU_R7	0	0	0.49148889	0
HS1_R1	-0.426973108	-0.0719403	-0.29443449	-2.0322535
HS1_R2	-0.411642214	-0.1442109	-0.330322929	-1.5141024
HS1_R3	-0.365646029	-0.1179453	-0.342916017	-1.3928714
HS1_R4	-0.353496706	-0.1233088	-0.272167529	-1.1696781
HS1_R5	-0.295690617	-0.124733	-0.239264339	-0.9191531

HS1_R6	-0.378819835	-0.148884	-0.334591567	-1.3442914
HS1_R7	-0.549150075	-0.1727581	-0.886428782	-2.5520562
HS2_R1	-0.416412237	-0.066983	-0.286485019	-1.9904336
HS2_R2	-0.41325882	-0.1441493	-0.330467448	-1.5195709
HS2_R3	-0.360576429	-0.1159178	-0.340660603	-1.3814077
HS2_R4	-0.348094506	-0.120919	-0.268277995	-1.1539183
HS2_R5	-0.29041025	-0.1218768	-0.233796756	-0.9006456
HS2_R6	-0.380819188	-0.1460504	-0.332518006	-1.3353626
HS2_R7	-0.530257771	-0.1617883	-0.887709474	-2.5523443
HEALTH_R1	-0.353947495	0	-0.175861117	-1.9856044
HEALTH_R2	-0.275110453	-0.0501745	-0.073079256	-0.5839443
HEALTH_R3	-0.225904	-0.0411885	-0.071984771	-0.644539
HEALTH_R4	-0.22782175	-0.0397499	-0.068420528	-0.3624593
HEALTH_R5	-0.238088481	-0.0471171	-0.072468001	-0.464326
HEALTH_R6	-0.212711777	-0.0409048	-0.087702247	-0.4795301
HEALTH_R7	-0.218107182	-0.0314602	-0.161476407	-0.5966541
MANU_R1	-0.294327783	0	-0.060182042	-0.7171101
MANU_R2	-1.527949546	0	-0.809856394	-5.7698093
MANU_R3	-0.233125311	-0.0308295	-0.048916819	-0.4437598
MANU_R4	0	0	0	0
MANU_R5	-0.244884125	-0.0379209	-0.041948141	-0.2909434
MANU_R6	-0.124930092	0	-0.094932372	-0.386771
MANU_R7	-0.772223762	0	-0.270674625	-3.4519058
RELIG_R1	-0.17623337	0	-0.099131547	-0.7714677
RELIG_R2	-0.18408582	-0.0445703	-0.076764113	-0.5608929
RELIG_R3	-0.157656743	-0.0408451	-0.078118649	-0.4877325
RELIG_R4	-0.162229151	-0.0268683	-0.068574262	-0.5720801
RELIG_R5	-0.153482019	-0.0359653	-0.083110117	-0.5035363
RELIG_R6	0.132397127	0	0.096150033	0.84465583
RELIG_R7	-0.081993055	0	-0.72126245	-2.9327316
UTIL_R1	-0.33875764	0	-0.044556783	-0.4587073
UTIL_R2	-0.327710188	-0.0110954	-0.033922919	-0.334638
UTIL_R3	-0.309468516	-0.001675	-0.048584728	-0.3277286
UTIL_R4	-0.169771849	0	-0.007324942	0
UTIL_R5	-0.213460474	0	0	-0.0380509
UTIL_R6	-0.329639309	0	-0.0637707	-0.5712193
UTIL_R7	0	0	-0.862365981	-0.4915708

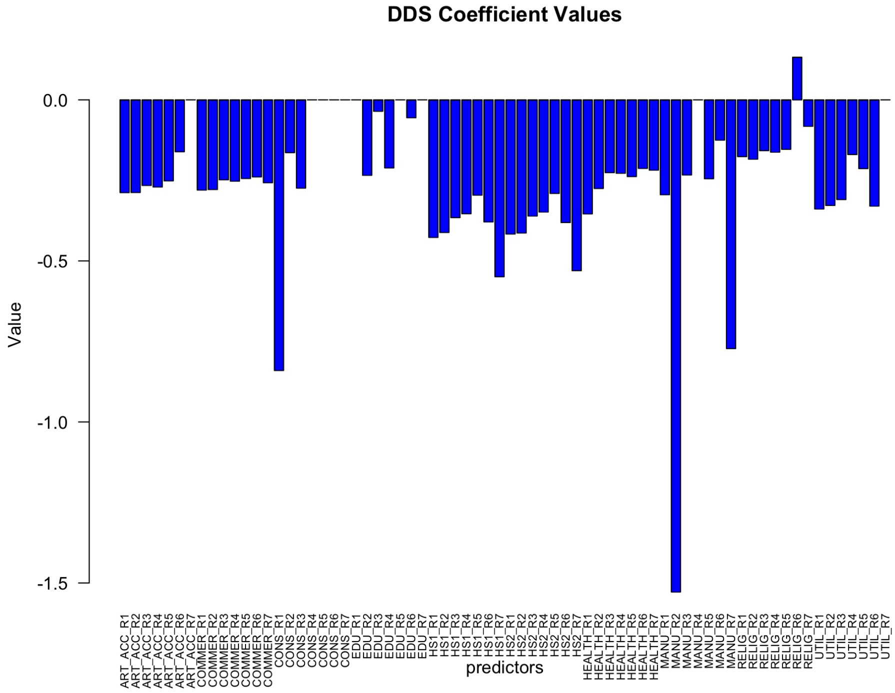


Figure 9: Bar Chart for DDS

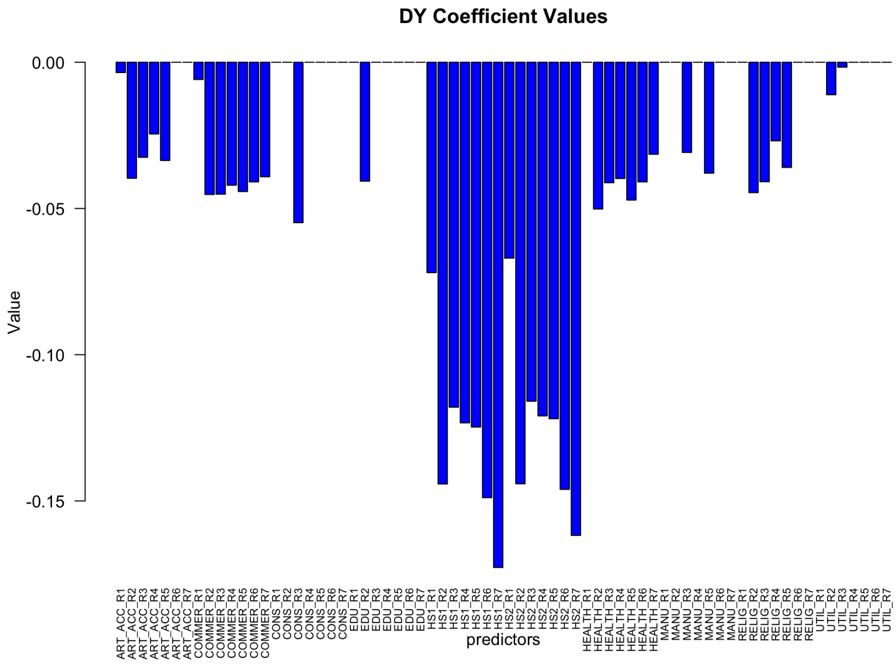


Figure 10: Bar Chart for DY

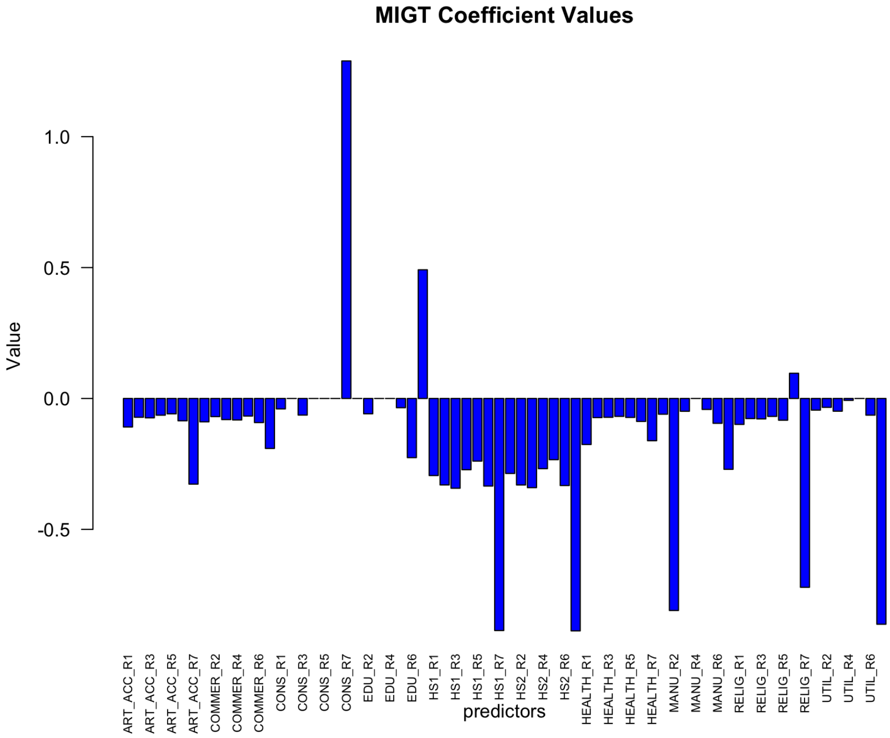


Figure 11: Bar Chart for MIGT

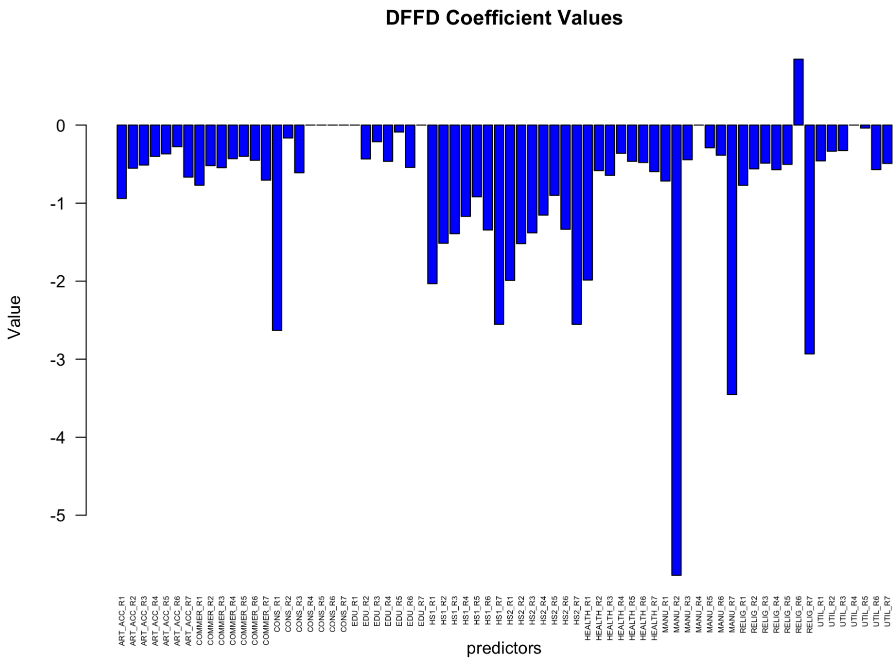


Figure 12: Bar Chart for DFFD

The coefficients in the community-wide model provide insight into how different sector-regions (predictors) influence the outcomes—Domestic Supply, Household Income, Migration, and Employment. These coefficients represent the estimated impact of changes in sector-regions on each outcome. Figure 9 - 12 visualizes all outcomes. From the table 8, sectors such as CONS_R1 have a significant negative coefficient (-0.84). This means that for every 1 million dollar loss in capital stock for CONS_R1, the total change in domestic supply decreases by 0.84. Sector-region Manu_R2 has the highest negative impact on DDS. In contrast, some coefficients, like RELIG_R6 (0.13), are positive but small, suggesting negligible contributions to DDS. DY demonstrates relatively smaller coefficients compared to DDS. However, the most impactful sectors in DY are mostly residential sectors HS1 and HS2. Migration coefficients, such as HS1_R7 (-0.886) and HS2_R7 (-0.887), highlight that housing sectors in Region 7 contribute significantly to migration. These large negative values suggest that housing availability or affordability disruptions in this region lead to population movements, possibly due to job loss or uninhabitable conditions. Employment outcomes display some of the most significant negative coefficients out of all outcomes.

Overall, sectors like construction and manufacturing consistently show significant impact across all outcomes. DDS and DFFD show the most significant coefficients. Areas with high coefficients for migration and employment require targeted interventions to stabilize housing and job markets. Similarly, sectors like manufacturing and construction should receive infrastructure support to minimize disruptions to DDS, and housing services in DY for stability.

4.1.2 Sector-Specific Models Models

Outcome	Sector Model	R-squared	RMSE	MAE
Domestic Supply	UTIL_R7	0.9709	0.0484	0.0413
	UTIL_R4	0.9897	0.1214	0.0957
Household Income	HH1_R1	0.9960	0.0885	0.0738
	HH1_R6	0.9834	0.0622	0.0523
Migration	HH2_R6	0.9995	0.3697	0.2912
	HH1_R7	0.9962	0.0849	0.0717
Employment	ART_ACC_R1_L4	0.9559	0.0718	0.0574
	ART_ACC_R1_L2	0.9725	1.4172	1.1942
	UTIL_R7_L2	0.3073	0.0109	0.0088
	UTIL_R7_L3	0.9539	0.0559	0.0448

Table 9: Sector-specific Model Performance

As shown in Table 9, the sector-specific models perform well overall, providing granular insights into the impacts of shocks at the sector-region level.

High-performing models, such as those for DDS, DY and MIGT, demonstrate high R^2 values with low RMSE and MAE values, indicating their reliability and accuracy. Nevertheless, there are instances of poor performance, as UTIL_R7_L2 for Employment, which has a low RMSE and MAE but an R-squared of 0.3073. Although the errors are low, the model has trouble capturing variability, indicating that the model needs further fine-tuning or modifications.

Overall, these sector-specific models add a layer of granularity to the analysis, provides a more detailed and specialized insights into how specfic sectors and regions respond to shocks.

4.2 Single Objective Model Performance

4.2.1 SLC Baseline Inventory

A quick overview of the current status of buildings in different regions for Salt Lake City.

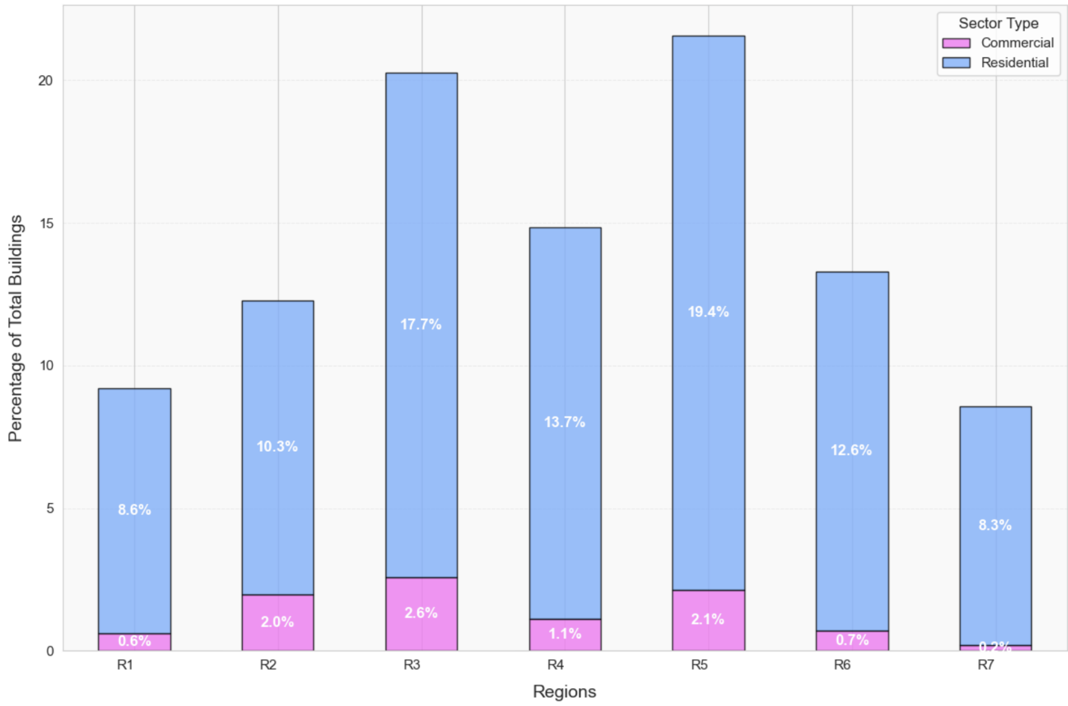


Figure 13: Baseline number of buildings by Region (in percentages)

Figure 13 shows the percentage distribution of Residential and Commercial buildings across Salt Lake City (SLC) regions. Residential buildings dominate all regions, with R5 (19.4%) and R3 (17.7%) having the largest shares. Commercial buildings make up a small portion, peaking at 2% in R2, R3, and R5, while regions like R1, R6, and R7 have less than 1%, reflecting limited commercial infrastructure. The distribution suggests Regions like R5 and R3, with majority of residential and commercial buildings, will likely need prioritized retrofitting, while areas like R7, with fewer buildings, may require less immediate intervention.

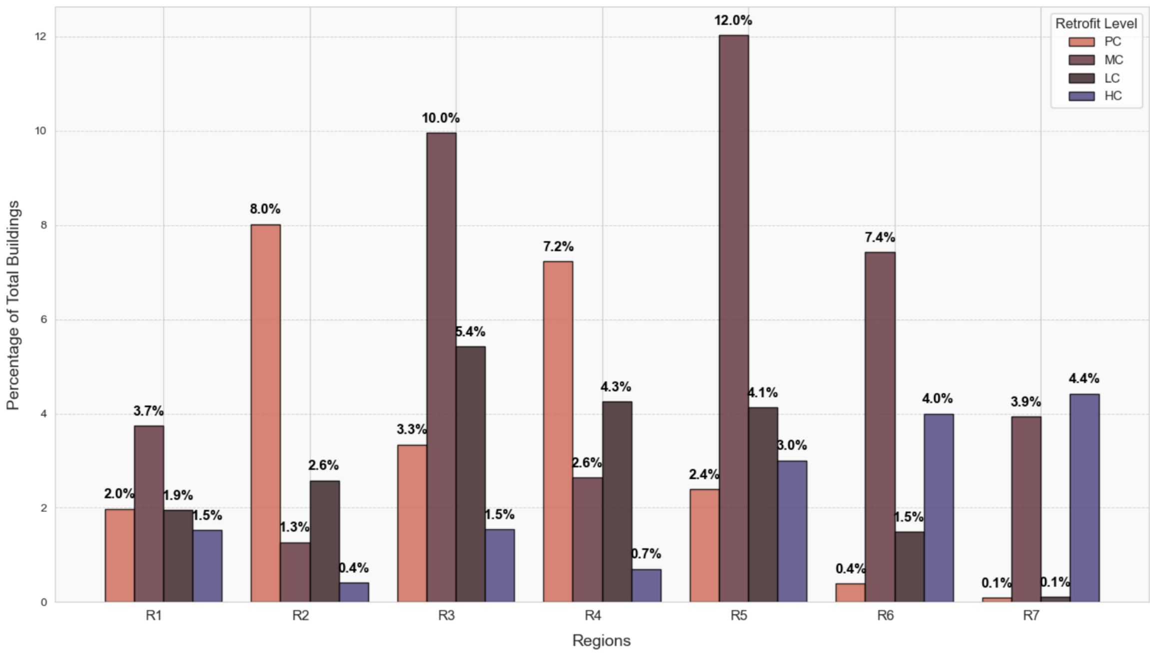


Figure 14: Baseline number of buildings by Region and Retrofit Level(in percentages)

Figure 14 shows the percentage distribution of buildings across retrofit levels (PC, LC, MC, HC). LC and MC dominate most regions, with R5 leading in MC (12%) and LC (4.1%). R2 has the highest share of PC buildings (8%), while R6 and R7 invest more in HC retrofits, at 4.4% and 4%, respectively. Regions such as R5 show foundational retrofits, while R6 and R7 focus more on higher retrofit levels despite having smaller building stocks. These differences highlight diverse retrofit priorities already in place, providing insights for targeted strategies.

4.2.2 Objective Values, Budget and Regional Investments Analysis for DDS

This section examines how different budget levels affect the DDS objective value, highlighting budget efficiency, diminishing returns, and the economic value of retrofitting for resilience. It also shows budget distribution between residential and commercial sectors across regions for DDS.

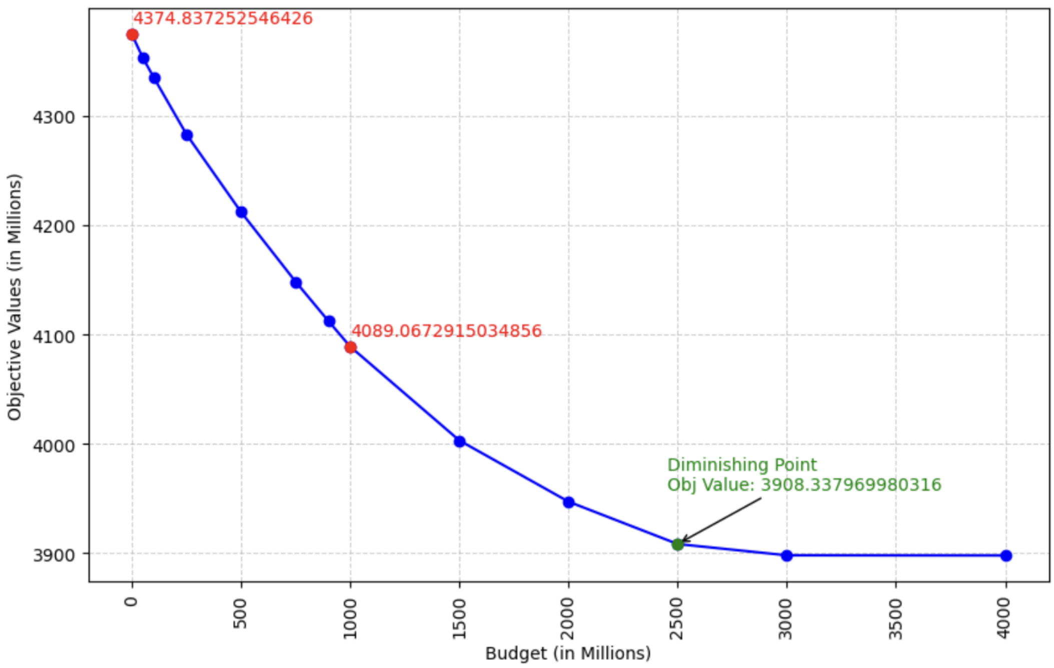


Figure 15: DDS objective values against budget levels

From the above plot, it is clear that as budgets increase from \$0 to \$4 billion, the DDS Objective Value (in millions) decreases, showcasing retrofitting's impact on reducing economic disruptions. At a baseline budget (\$0) with no retrofits, the DDS Objective Value is 4374 million, reflecting maximum vulnerability. Significant gains are seen with early investments, particularly those up to \$1.5 billion; for instance, a \$1 billion budget lowers the value to 4089 million. But at 2.5 billion, the objective value starts to decline, and the curve flattens out around \$4 billion, or about 3900 million, indicating diminishing returns. This pattern shows the need to use resources wisely, prioritizing early budgets for high-impact interventions.

Retrofitting buildings helps reduce disruptions caused by disasters, preventing both displacement and fatalities due to building collapses and severity caused by disaster impacts. Although the initial costs may appear high, the savings on disaster response and economic recovery far exceed these expenses. It is a long-term investment. Communities that undergo retrofitting become safer and more resilient, which minimizes devastation during disasters. In Salt Lake City, it is essential to balance residential and commercial priorities to ensure the protection of both individuals and the economy.

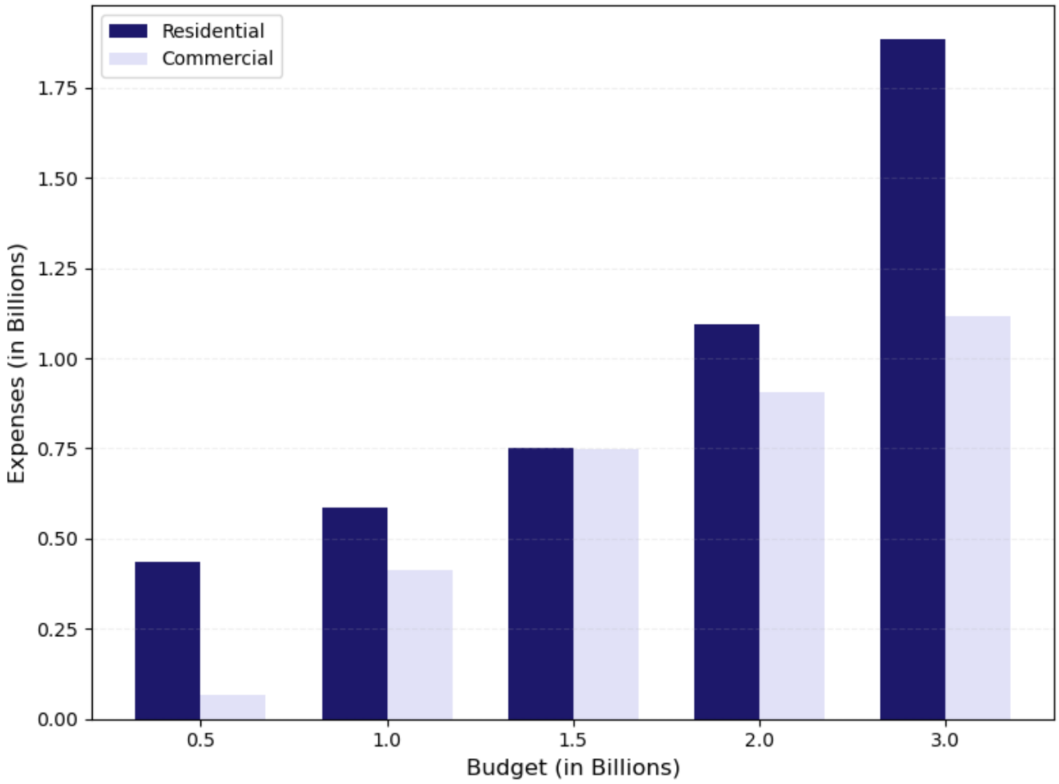


Figure 16: Regional Investments in Residential vs Commercial Sectors

Figure 16 shows how investments shift between residential and commercial sectors across budgets. Lower budgets (\$0.5–\$1 billion) focus on residential areas to stabilize and reduce vulnerabilities. At \$1.5B, the investments made for commercial and residential sectors are the same. By \$3 billion, commercial spending surpasses residential, reflecting a shift toward protecting critical economic infrastructure as residential needs are addressed. Early investments focus on residential retrofits to reduce displacement and social vulnerability, while later budgets target commercial infrastructure to sustain economic continuity.

These plots collectively reveal vital insights, such as initial investments under \$2 billion yielding the highest reduction in DDS economic losses and prioritizing foundational vulnerabilities in residential and commercial sectors. Beyond \$3 billion, diminishing returns call for strategically targeting remaining resilience gaps. The shift from residential to commercial spending shows the priorities, focusing first on stabilizing homes and then protecting businesses.

4.2.3 Regional Analysis of Retrofitting Priorities and Vulnerabilities Across All Outcomes

Examples of building retrofits across regions for various budget values and outcomes are shown in the following tables.

Table 10: Retrofit Analysis Across Budgets by Region and K level for DDS

Budget	Region	Retrofit Level	Number of Buildings	Percentage (%)
0	R1	HC	4275.00	1.53
	R2	PC	22440.00	8.02
	R3	MC	27866.00	9.96
	R5	LC	11562.00	4.13
	.	.	.	.
	.	.	.	.
1500 M	R1	HC	17971.67	6.42
	R2	PC	22207.00	7.94
	R3	MC	137.00	0.05
	R5	LC	11497.00	4.11
	.	.	.	.
	.	.	.	.
2000 M	R1	HC	19359.00	6.92
	R2	PC	22007.00	7.86
	R3	MC	136.00	0.05
	R5	LC	11463.00	4.10
	.	.	.	.
	.	.	.	.
3000 M	R1	HC	25013.00	8.94
	R2	PC	0.00	0.00
	R3	MC	1.00	0.00
	R5	LC	11764.00	4.20
	.	.	.	.

Table 11: Retrofit Analysis Across Budgets by Region and K level for DY

Budget	Region	Retrofit Level	Number of Buildings	Percentage (%)
0	R1	HC	4275.00	1.53
	R2	PC	22440.00	8.02
	R3	MC	27866.00	9.96
	R5	LC	11562.00	4.13
	.	.	.	.
	.	.	.	.
1500 M	R1	HC	13901.00	4.97
	R2	PC	22369.00	7.99
	R3	MC	548.80	0.20
	R5	LC	5620.00	2.01
	.	.	.	.
	.	.	.	.
2000 M	R1	HC	13934.00	4.98
	R2	PC	2385.00	0.85
	R3	MC	248.00	0.09
	R5	LC	5575.00	1.99
	.	.	.	.
	.	.	.	.
3000 M	R1	HC	19275.00	6.89
	R2	PC	12.00	0.00
	R3	MC	38.00	0.01
	R5	LC	11760.00	4.20
	.	.	.	.
	.	.	.	.

Table 12: Retrofit Analysis Across Budgets by Region and K level for MIGT

Budget	Region	Retrofit Level	Number of Buildings	Percentage (%)
0	R1	HC	4275.00	1.53
	R2	PC	22440.00	8.02
	R3	MC	27866.00	9.96
	R5	LC	11562.00	4.13
	.	.	.	.
	.	.	.	.
1500 M	R1	HC	19144.00	6.84
	R2	PC	22361.00	7.99
	R3	MC	550.68	0.20
	R5	LC	11549.00	4.13
	.	.	.	.
	.	.	.	.
2000 M	R1	HC	19209.00	6.86
	R2	PC	3730.85	1.33
	R3	MC	300.00	0.11
	R5	LC	5620.00	2.01
	.	.	.	.
	.	.	.	.
3000 M	R1	HC	19439.00	6.95
	R2	PC	9.00	0.00
	R3	MC	126.00	0.05
	R5	LC	11760.00	4.20
	.	.	.	.
	.	.	.	.

Table 13: Retrofit Analysis Across Budgets by Region and K level for DFFD

Budget	Region	Retrofit Level	Number of Buildings	Percentage (%)
0	R1	HC	4275.00	1.53
	R2	PC	22440.00	8.02
	R3	MC	27866.00	9.96
	R5	LC	11562.00	4.13
	.	.	.	.
	.	.	.	.
1500 M	R1	HC	19306.00	6.90
	R2	PC	22366.00	7.99
	R3	MC	343.00	0.12
	R5	LC	11549.00	4.13
	.	.	.	.
	.	.	.	.
2000 M	R1	HC	19367.00	6.92
	R2	PC	22088.00	7.89
	R3	MC	169.00	0.06
	R5	LC	5717.59	2.04
	.	.	.	.
	.	.	.	.
3000 M	R1	HC	25336.00	9.05
	R2	PC	0.00	0.00
	R3	MC	14.00	0.01
	R5	LC	11764.00	4.20
	.	.	.	.
	.	.	.	.

Only a few excerpts of the optimization results are shown in the above tables 10- 13. The results of each objective provide a detailed understanding of how different regions within Salt Lake City contribute to disaster resilience. Region R5 stands out as a significant hub for both residential and commercial activities and is consistently ranked as a top priority across all objectives. For DDS, R5 mostly has HC retrofits, with MIGT and DY focusing on MC upgrades. DFFD is more evenly distributed, with upgrades ranging from LC to HC. Region R3 is also a critical and vulnerable region, with approximately 18% of the buildings belonging to residential sectors. Much of the buildings in R3 at the base level are in PC(3.3%) or LC(5.5%). These vulnerabilities are most pronounced under Migration (MIGT) and Income (DY) outcomes. The framework prioritizes MC retrofits in R3 for cost-effective household protection and mitigating large-scale migration. Region R2 occupies a unique role, blending residential and commercial structures like R5 but in smaller numbers. With many residential buildings at PC and LC levels, it faces moderate risks, particularly in MIGT and DY. Retrofitting focuses on MC and LC upgrades.

Other regions, such as R6 and R7, have a solid baseline, with most buildings already being at MC or HC before retrofitting. This may indicate that these sectors are important and already implemented retrofits for disaster preparedness. R6 is mainly impacted by DDS, while R7 has more attention from DFFD, with its contribution being limited to other objectives. R4, although not as prioritized as R5 or R3, demonstrates vulnerabilities under MIGT and DY.

Based on these results, Region R5 should have a substantial share of retrofitting resources, focusing on High-Code upgrades for its commercial infrastructure and Mid-Code upgrades for its residential areas. This dual approach will ensure that R5 remains an economic and social anchor during disasters. R3 needs targeted Mid-Code retrofits in residential areas to prevent displacement and stabilize incomes, supported by selective retrofits in R2. Investments in R6 should focus on domestic supply, while R4 requires measures to enhance residential stability. A data-driven, region-specific approach ensures effective retrofitting and promotes resilience and equity in disaster management for Salt Lake City.

5 Challenges, Limitations and Future Work

This section discusses some challenges the existing framework may face and the options to overcome them. It questions and challenges all three models (CGE, ML, Optimization) used in the framework and tries to address each of the issues to ensure the reliability of our framework.

First, ensuring that the CGE data is consistent with real-life scenarios is essential, as it is the base of our analysis. The reliability of our framework depends on the quality of this data. One way to validate the CGE data is to compare the model outcomes with historical data or past observations to find patterns and trends and assess their alignment with the real world. Additionally, other conventional methods like input-output (IO) models, although far less complex and comprehensive than the CGE model, can be an excellent tool for comparing, cross-checking, and validating the CGE results—mainly if there are inconsistencies that require attention. Furthermore, reviewing and discussing the results with domain experts and policymakers are crucial to help evaluate the practicality and applicability of the CGE results.

Second, the chosen model for this research, elastic net regression, is excellent for linear relationships and maintaining interpretability. Still, it can fall short when dealing with the non-linear dependencies inherent in CGE models. To address this limitation, residual analysis, and diagnostic plots can be crucial first steps that help uncover where the linear model struggles. For example, patterns in the residual might reveal non-linear relationships that the model cannot capture. Once the problem is discovered, we can think of alternative techniques and models to consider and deal with the issue. For example, piecewise modeling can divide the data into smaller, more manageable subsets based on meaningful thresholds, such as different levels of damage thresholds (e.g., high or low) to handle non-linearity. A separate linear model is trained for each subset, enabling the model to approximate complex relationships by treating smaller subsets as approximately linear. Aside from that, ensemble methods or hybrid modeling can be considered. Non-linear models such as the random forest or gradient boosting can be used alongside elastic net to address complex relationships and maintain simplicity and interpretability.

Third, the current optimization framework does not allow buildings to be demolished. This is not necessarily an issue but rather an exploration of choices and options we can have in the future. The machine learning coefficients obtained can be positive, negative or zero as discussed in the results section. Negative coefficients identify the sector-regions that are most vulnerable, and the buildings in these sector-regions must be prioritized. The priority depends on the magnitude of the coefficient. A coefficient of zero implies that retrofitting a particular building has no impact on the objective function (e.g., DDS). Retrofitting or not retrofitting the building neither increases nor decreases the economic outcome being optimized. Positive coefficients in the ML model signify buildings less vulnerable than negative coefficients or where retrofitting would not meaningfully reduce DDS. In fact, not retrofitting these buildings minimizes the objective function, as retrofitting them would either have a neutral or counterproductive impact on DDS. The current optimization model is set up so that it does not allow buildings to be destroyed. This is done in two of the constraints defined in Equation (12) and (13) in Section 3.3.6, where both equations ensure that no building is lost or created during the process of retrofitting and after retrofitting. All buildings are accounted for. If these two constraints were not in place, the positive coefficients would allow buildings to be destroyed, as that would minimize the objective function. In the current formulation, these buildings are not prioritized for retrofitting, leaving them unprotected but not physically removed.

In some cases, the removal of buildings may be beneficial, depending on the location, available resources, emergency, or other adversaries. For example, structurally unsafe buildings might need to be demolished to reduce risk or allow for better resource allocation. Whether to allow such removal depends on the specific case, as well as valuable input and decisions of professionals, stakeholders, and policymakers. To handle such situations, modifications to the model are necessary. One approach is to relax the balance constraint in Equation (13) to be:

$$\sum_{k \in K} x_{ijk} \leq \sum_{k \in K} b_{ijk}, \quad \forall i \in I, \forall j \in J \tag{16}$$

Here, the optimization model implicitly excludes buildings by not requiring that the number of buildings post-retrofit matches the original count of buildings. It is simple and easy to implement. However, it lacks structure and does not account for buildings lost or removed. As a result, it might need more clarity in terms of decision-making and reporting.

A better alternative is introducing a new decision variable that explicitly represents and considers any building excluded or removed. For that, the balance constraint is reformulated as:

$$\sum_{k \in K} x_{ijk} + \sum_{k \in K} e_{ijk} \leq \sum_{k \in K} b_{ijk}, \quad \forall i \in I, \forall j \in J \tag{17}$$

Where e_ijk is the new decision variable that allows buildings in $\imath inI$, $\jmath inJ$, and inK to be excluded or removed and allows for tracking the buildings removed in the process. This method provides better clarity by ensuring that the total count of buildings still matches the initial number while clearly accounting for buildings removed. This approach is slightly more complex than simply relaxing a constraint, but it offers precision and clear results, making it a more robust option.

Furthermore, this research focuses on single objective optimization. However, the trade-offs between competing outcomes might not be adequately captured by the single objective model. A more thorough and balanced solution can be obtained by extending the framework into a multi-objective optimization (MOO) strategy, which takes into account several outcomes at once. For example, DDS and DY are separate objectives that can be regarded as a single MOO (making it a bi-objective model) instead of having two separate single-objective models. This can help policymakers address multiple competing priorities and ensure that no single objective is pursued at the expense of others. Several techniques can be employed to handle two or more competing objectives. Two standard methods are using weights or using the epsilon-constraint method. Weights allow both objectives to be combined into a single optimization function, with each objective scaled by relative importance. This method provides flexibility, enabling the exploration of trade-offs between objectives by adjusting the weights. For example, to prioritize DDS or DY, using a higher weight with the preferred objective would prioritize it. However, choosing appropriate weights can be subjective and require expert opinion and understanding. The objectives should be relative to one another to select the weights effectively. In the second method, the epsilon-constraint method, one objective is optimized while treating the other(s) as constraints with specified limits. For instance, DDS could be minimized while ensuring DY does not fall below an unacceptable objective value with a defined threshold. Both approaches improve the model's ability to balance trade-offs: the epsilon-constraint approach guarantees rigorous commitment to predetermined limits, which makes it perfect when specific outcomes must satisfy particular policy or practical requirements. On the other hand, weights provide a flexible framework for investigating different priorities.

Overall, resolving these issues enhances the framework's ability to provide useful, trustworthy insights for disaster management.

6 Summary and Conclusion

This research demonstrates the potential of merging predictive analytics with optimization methods to address the complex issues of disaster resilience. Through a focus on optimizing a single objective, we have created a system that turns data into specific and practical plans for reducing the economic effects of natural disasters. A significant contribution of this research lies in its ability to identify sectoral vulnerabilities and overall impacts, providing policymakers with the tools to enforce targeted, high-impact interventions.

Despite its strengths, this work acknowledges the complexity of resilience planning and the need for continuous improvement. While the current focus on single-objective optimization offers clarity and accuracy, it also leaves room for multi-objective optimization models for future work. It can explore trade-offs between different objectives, allowing decision-makers to balance competing priorities in real-world scenarios. Additionally, the framework can be extended to new testbeds and hazard scenarios. Like sector-specific models, it can work on impact-based models that allow analysis of outcomes for different damage levels for a more focused study.

As natural disasters grow in frequency and intensity, the methodologies developed here serve as a critical foundation for more thoughtful, innovative, and resilient planning on both local and global scales.

Bibliography

[1] J. Helgeson, J. Fung, C. O'Fallon, D. Webb, and H. Cutler, "Identifying and quantifying the resilience dividend using computable general equilibrium models: A methodological overview," *Identifying and Quantifying the Resilience Dividend using Computable General Equilibrium Models: A Methodological Overview*, 2017. [Online]. Available: `https://tsapps.nist.gov/publication/get_pdf.cfm?pub_id=924427`.

[2] United Nations International Stratety for Disaster Reduction, *Global Assessment Report on Disaster Risk Reduction 2015: Making Development Sustainable: the Future of Disaster Risk Management*. UN, 2015.

[3] A. B. Smith, *U.s. billion-dollar weather and climate disasters, 1980 - present (ncei accession 0209268)*, NOAA National Centers for Environmental Information. Dataset. Subset used:https://www .ncei.noaa.gov/access/billions/summary-stats/US/2021-2023. Accessed: 2024-06-13, 2020. [Online]. Available: `https://doi.org/10.25921/stkw-7w73`.

[4] Y. Kajitani and H. Tatano, "Applicability of a spatial computable general equilibrium model to assess the short-term economic impact of natural disasters," *Economic Systems Research*, vol. 30, no. 3, pp. 289–312, 2017. DOI: `10.1080/09535314.2017.1369010`.

[5] H. Cutler and S. Davies, "The economic consequences of productivity changes: A computable general equilibrium (cge) analysis," *Regional Studies*, vol. 44, no. 10, pp. 1415–1426, 2008. DOI: `10.1080/00343400701654210`.

[6] H. S. Gupta, A. D. Gonzalez, R. Jnad, and S. Kameshwar, "Fairness-Driven Multi-Objective Optimization for Evacuation Planning in Natural Disasters," in *International Conference on Transportation and Development 2024*, Atlanta, GA, 2024, pp. 170–180.

[7] M. Koliou, J. W. van de Lindt, T. P. McAllister, B. R. Ellingwood, M. Dillard, and H. Cutler, "State of the research in community resilience: Progress and challenges," *Sustainable and Resilient Infrastructure*, vol. 5, no. 3, pp. 131–151, 2018. DOI: `10.1080/23789689.2017.1418547`. [Online]. Available: `https://www.tandfonline.com/doi/full/10.1080/23789689.2017.1418547`.

[8] B. R. Ellingwood and K. Kinali, "Quantifying and communicating uncertainty in seismic risk assessment," *Structural safety*, vol. 31, no. 2, pp. 179–187, 2009.

[9] H. S. Gupta, "Optimal selection of short-and long-term mitigation strategies for flooding hazard," Master's Thesis, University of Oklahoma, Norman, OK, 2022.

[10] H. S. Gupta, A. Gonzalez, and C. Nicholson, "Building flood-resilient communities through strategic retrofitting of buildings under uncertainty," in *2024 IISE Annual Conference and Expo*, IISE, 2024.

[11] F. H. Norris, S. P. Stevens, B. Pfefferbaum, K. F. Wyche, and R. L. Pfefferbaum, "Community resilience as a metaphor, theory, set of capacities, and strategy for disaster readiness," *American Journal of Community Psychology*, vol. 41, pp. 127–150, 2008. DOI: `10.1007/s10464-007-9156-6`.

[12] P. Bodin and B. Wiman, "Resilience and other stability concepts in ecology: Notes on their origin, validity, and usefulness," *ESS Bulletin*, vol. 2, pp. 33–43, Jan. 2004.

[13] J. E. Gordon, *Structures: Or Why Things Don't Fall Down*. Da Capo Press, 1978, ISBN: 0306801515.

[14] Dovers and Handmer, "Uncertainty, sustainability and change," *Global Environmental Change*, vol. 2, no. 4, pp. 262–276, 1992. DOI: `10.1016/0959-3780(92)90044-8`.

[15] D. Mileti and E. Noji, *Disasters by Design: A Reassessment of Natural Hazards in the United States*. Jul. 1999, ISBN: 978-0-309-26173-9. DOI: `10.17226/5782`.

[16] D. Paton and D. Johnston, "Disasters and communities: Vulnerability, resilience and preparedness," *Disaster Prevention and Management: An International Journal*, vol. 10, no. 4, pp. 270–277, 2001.

[17] J. Barnett, "Adapting to climate change in pacific island countries: The problem of uncertainty," *World Development*, vol. 29, no. 6, pp. 977–993, 2001.

[18] E. Jordan and A. Javernick-Will, "Indicators of community recovery: Content analysis and delphi approach," *Natural Hazards Review*, vol. 14, no. 1, pp. 21–28, 2013.

[19] N. Doorn, "Resilience indicators: Opportunities for including distributive justice concerns in disaster management," *Journal of Risk Research*, vol. 20, pp. 1–21, Oct. 2015. DOI: `10.1080/13669877.2015.1100662`.

[20] P. Timmerman, *Vulnerability, Resilience and the Collapse of Society: A Review of Models and Possible Climatic Applications*. Toronto, Canada: Institute for Environmental Studies, University of Toronto, 1981.

[21] C. S. Holling, "Resilience and stability of ecological systems," *Annual Review of Ecology and Systematics*, vol. 4, pp. 1–23, 1973. DOI: `10.1146/annurev.es.04.110173.000245`. [Online]. Available: `http://www.jstor.org/stable/2096802`.

[22] J. W. Handmer and S. R. Dovers, "A typology of resilience: Rethinking institutions for sustainable development," *Industrial & Environmental Crisis Quarterly*, vol. 9, no. 4, pp. 482–511, 1996.

[23] G. A. Bonanno, "Loss, trauma, and human resilience: Have we underestimated the human capacity to thrive after extremely aversive events?" *American Psychologist*, vol. 59, no. 1, pp. 20–28, 2004.

[24] R. J. Klein, R. J. Nicholls, and F. Thomalla, "Resilience to natural hazards: How useful is this concept?" *Global Environmental Change Part B: Environmental Hazards*, vol. 5, no. 1-2, pp. 35–45, 2003.

[25] Y. Y. Haimes, "On the definition of resilience in systems," *Risk Analysis: An International Journal*, vol. 29, no. 4, pp. 498–501, 2009.

[26] H. S. Gupta, O. M. Nofal, A. D. González, C. D. Nicholson, and J. W. van de Lindt, "Optimal selection of short- and long-term mitigation strategies for buildings within communities under flooding hazard," *Sustainability*, vol. 14, no. 16, p. 9812, 2022. DOI: 10.3390/su14169812. [Online]. Available: https://doi.org/10.3390/su14169812.

[27] W. Wang, J. W. van de Lindt, B. Hartman, H. Cutler, J. L. Kruse, T. P. McAllister, and S. Hamideh, "Determination of individual building performance targets to achieve community-level social and economic resilience metrics," *Journal of Structural Engineering*, 2022. DOI: 10.1061/(ASCE)ST.1943-541X.0003338. [Online]. Available: https://tsapps.nist.gov/publication/get_pdf.cfm?pub_id=932848.

[28] S. B. Manyena, "The concept of resilience revisited," *Disasters*, vol. 30, no. 4, pp. 433–450, 2006.

[29] W. E. Highfield, W. G. Peacock, and S. Van Zandt, "Mitigation planning: Why hazard exposure, structural vulnerability, and social vulnerability matter," *Journal of Planning Education and Research*, vol. 34, no. 3, pp. 287–300, 2014.

[30] A. Mullins and R. Soetanto, "Enhancing community resilience to flooding through social responsibility," *International Journal of Safety and Security Engineering*, vol. 1, no. 2, pp. 115–125, 2011. DOI: 10.2495/SAFE-V1-N2-115-125.

[31] R. Guidotti, H. Chmielewski, V. Unnikrishnan, P. Gardoni, T. McAllister, and J. W. van de Lindt, "Modeling the resilience of critical infrastructure: The role of network dependencies," *Sustainable and Resilient Infrastructure*, vol. 1, no. 3-4, pp. 153–168, 2016. DOI: 10.1080/23789689.2016.1254999.

[32] D. Elms, I. McCahon, and R. Dewhirst, "Improving infrastructure resilience," *Civil Engineering and Environmental Systems*, vol. 36, no. 1, pp. 83–99, 2019. DOI: 10.1080/10286608.2019.1615479.

[33] M. Bruneau, S. Chang, R. Eguchi, G. Lee, T. O'Rourke, A. Reinhorn, M. Shinozuka, K. Tierney, W. Wallace, and D. Winterfeldt, "A framework to quantitatively assess

and enhance the seismic resilience of communities," *Earthquake Spectra - EARTHQ SPECTRA*, vol. 19, Nov. 2003. DOI: 10.1193/1.1623497.

[34] S. Wang, W. Tang, D. Qi, J. Li, E. Wang, Z. Lin, and C. Duffield, "Understanding the role of built environment resilience to natural disasters: Lessons learned from the wenchuan earthquake," *Journal of Performance of Constructed Facilities*, vol. 31, p. 04 017 058, Apr. 2017. DOI: 10.1061/(ASCE)CF.1943-5509.0001062.

[35] H. Cutler, M. Shields, D. Tavani, and S. Zahran, "Integrating engineering outputs from natural disaster models into a dynamic spatial computable general equilibrium model of centerville," *Sustainable and Resilient Infrastructure*, vol. 1, no. 3–4, pp. 169–187, 2016. DOI: 10.1080/23789689.2016.1254996.

[36] G. S. Guha, "Simulation of the economic impact of region-wide electricity outages from a natural hazard using a cge model," *Southwestern Economic Review*, vol. 32, pp. 101–124, 2011.

[37] A. Bastaminia, M. R. Rezaei, and M. H. Saraei, "Evaluating the components of social and economic resilience: After two large earthquake disasters rudbar 1990 and bam 2003," *Jàmbá: Journal of Disaster Risk Studies*, vol. 9, no. 1, pp. 1–12, 2017.

[38] T. G. Nachappa, O. Ghorbanzadeh, K. Gholamnia, and T. Blaschke, "Multi-hazard exposure mapping using machine learning for the state of salzburg, austria," *Remote Sensing*, vol. 12, no. 17, p. 2757, 2020.

[39] W. Zhang and C. Nicholson, "A multi-objective optimization model for retrofit strategies to mitigate direct economic loss and population dislocation," *Sustainable and Resilient Infrastructure*, vol. 1, no. 3-4, pp. 123–136, 2016.

[40] H. S. Gupta, T. Adluri, D. Sanderson, A. D. González, C. D. Nicholson, and D. Cox, "Multi-objective optimization of mitigation strategies for buildings subject to multiple hazards," *International Journal of Disaster Risk Reduction*, vol. 100, p. 104 125, 2024.

[41] Y. Wen, “Development of multi-objective optimization model of community resilience on mitigation planning,” Ph.D. dissertation, University of Oklahoma, 2021. [Online]. Available: https://shareok.org/items/2032a447-8e39-4ec0-9c95-8e56ff4b9933.

[42] IBM. “What is machine learning?” (2023), [Online]. Available: https://www.ibm.com/topics/machine-learning.

[43] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Germany: Springer New York, 2013.

[44] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2nd. Stanford, CA: Stanford University, 2009. DOI: 10.1007/978-0-387-84858-7. [Online]. Available: https://web.stanford.edu/~hastie/Papers/ESLII.pdf.

[45] L. Koenig, “Recent earthquake activity unlocks mysteries of the wasatch fault,” *Temblor*, 2021. DOI: 10.32858/temblor.184. [Online]. Available: http://doi.org/10.32858/temblor.184.

[46] U. C. Earthquake Engineering Research Institute, “Scenario for a magnitude 7.0 earthquake on the wasatch fault—salt lake city segment: Hazards and loss estimates,” Utah Seismic Safety Commission, Salt Lake City, UT, 2015, Prepared for the Utah Seismic Safety Commission and funded by the Federal Emergency Management Agency (FEMA).

[47] Utah Seismic Safety Commission (USSC), *Earthquake probabilities for the wasatch front region in utah, idaho, and wyoming*, 2016. [Online]. Available: https://ussc.utah.gov/pages/help.php?section=Utah+Earthquake+Probabilities.

[48] Utah Geological Survey, *The wasatch fault from above: Re-mapping the wasatch fault zone using airborne high-resolution topographic data*, 2017. [Online]. Available: https://geology.utah.gov/map-pub/survey-notes/wasatch-fault-above/.

[49] U.S. Geological Survey (USGS), *Magnitude 5.7 - 4 km n of magna, utah - executive summary*, 2020. [Online]. Available: https://earthquake.usgs.gov/earthquakes/eventpage/uu60363602/executive.