Thesis Report

Exploring equity weighting at the household level for flood risk assessment: a case study at Charleston, United States

MSc Thesis in Environmental Engineering

Dwiva Anbiya Taruna

Delft University of Technology | Deltares | LPDP

TU Delft

Thesis Report

Exploring equity weighting at the household level for flood risk assessment: a case study at Charleston, United States

by

Dwiva Anbiya Taruna

to obtain the degree of Master of Science
at the Delft University of Technology,
to be defended publicly on Tuesday August 27, 2024 at 10:45 AM.

Student number:	5849578	
Thesis committee:	Dr. Ir. Maurits W. Ertsen,	TU Delft, supervisor
	Dr. Ir. Olivier Hoes,	TU Delft, supervisor
	Dr. Ir. Cong Mai Van,	TU Delft, supervisor
	Dr. Ir. Juan Pablo Aguilar-Lopez,	TU Delft, supervisor
	Dr. Kathryn Roscoe,	Deltares, External supervisor
	Dr. Ir. Panagiotis Athanasiou,	Deltares, External supervisor

TU Delft lpdp lembaga pengelola dana pendidikan Deltares

Acknowledgement

I want to express my deepest gratitude to all those who contributed to completing this thesis on exploring equity weighting at the household level for flood risk assessment.

Firstly, I thank Maurits W. Ertsen, my chair, for guiding me throughout this journey from the internship to the thesis project and forming the committee. I sincerely thank my first supervisor, Olivier Hoes, for his guidance, feedback, and brainstorming sessions that helped shape the thesis project. I am also profoundly grateful to my second and third supervisors, Cong Mai Van and Juan Pablo Aguilar-Lopez, for their detailed feedback and support.

I am thankful to Kathryn Roscoe for allowing me to explore the equity method from the internship through the thesis project and for her direction and feedback. I also would like to thank Panagiotis Athanasiou for his time, guidance, and detailed feedback, which inspired me to keep working on my thesis.

I am so grateful to have these thesis committees in my thesis journey, who always gave constructive feedback and encouragement while I was working on this thesis project.

I extend my heartfelt thanks to the LPDP (Indonesia Endowment Fund) for providing the scholarship that made my studies possible. Also, Deltares, specifically the Climate Adaptation and Disaster Risk Management Department, will allow me to use this topic as my thesis project.

Special thanks to my parents, Agus and Wiwi, my brother Pascal and his wife Shabrina and my sister Mutiara, and my parents-in-law, Rochiman and Enny. Your endless support and prayers from afar have been my strength. To my wife, Merry, thank you for your unwavering support and encouragement throughout my studies. Lastly, I thank my friends in Delft, housemates, and colleagues in Deltares for creating an enjoyable and supportive environment.

Dwiva Anbiya Taruna
Delft, August 2024

Preface

When I first arrived at TU Delft, my primary motivation was to undertake a thesis in the field of flood risk assessment. This thesis fulfills that initial goal. Initially, I had expected to work more within the field of hydrology, hydraulics, and modeling areas that have long intrigued me. However, I challenged myself to shift towards the vulnerability aspect of flood risk assessment when this subject offered, which allowed me to explore different perspectives of flood risk.

My perspective on flood risk expanded when I was introduced to the concept of equity weighting during my internship at Deltares. This new focus on incorporating equity into flood risk assessment was unexpected but enriching. It allowed me to explore how equity weighting can address the needs of vulnerable populations in flood risk management. I was particularly intrigued by how equity weighting could help identify and prioritize those most at risk due to socioeconomic factors, thereby offering a more just and inclusive approach to flood risk management. This focus on vulnerability provided a deeper understanding of how flood risks are not evenly distributed and how different communities are affected in disparate ways.

Completing this work would not have been possible without the invaluable guidance and contributions of my supervisors. I am also deeply grateful for the unwavering support and prayers from my family and friends. Their encouragement has been instrumental in helping me navigate this journey.

I am grateful for the decision to explore this specific field and am satisfied with the results of my research. This experience has reinforced my commitment to continuing in this area. In my future roles, I aim to develop further and apply these insights to enhance flood risk management practices.

Dwiva Anbiya Taruna
Delft, August 2024

Summary

Flood risk management is essential in water resource planning to reduce potential damages and losses in flood-prone areas. Traditional risk assessments often focus on monetary values and ignore equity, leading to unequal distribution of protective measures. The equity weighting method has been implemented in past studies, but at certain administrative levels, it may not capture the income heterogeneity within those areas. This research investigates how accounting for income heterogeneity within administrative areas affects equity-weighted risk and impact estimates and explores practical ways to implement this approach, given privacy-related data limitations.

The study focuses on Charleston County, South Carolina, using disaggregated methods at the household level to integrate detailed socioeconomic data into equity-weighted flood risk assessments. The methodology involves assuming an equal distribution of income within predefined brackets and optimizing upper and lower bounds for open income brackets. Additionally, a linear relationship is assumed between the structure value and household income, allowing for the spatial allocation of income within census block groups.

The results demonstrate that the disaggregated method identifies more vulnerable households and shows significant differences in equity-weighted damage (EWD) and equity-weighted expected annual damage (EWEAD) metrics compared to the aggregated method. These findings suggest that the disaggregated method provides useful evidence for considering the importance of taking into account income heterogeneity when assessing flood risks, particularly in areas with significant income inequality.

Sensitivity and uncertainty analyses indicate that the income distribution parameters show both low sensitivity and low uncertainty when using a fitted log-normal distribution, suggesting that this distribution is a good fit for the data and provides stable results. In contrast, in the spatial allocation of income with structural values, the analysis results show high sensitivity and significant uncertainty, which means small changes in structure values lead to substantial variations in equity weight estimates.

The study concludes with recommendations for stakeholders to adopt household-level analysis and incorporate equity weights in cost-benefit analyses. Future research should explore additional socioeconomic factors, develop dynamic risk assessment models, and apply the methodology to diverse geographic and economic contexts.

Contents

List of Figures

List of Tables

Nomenclature

Abbreviations

Abbreviation	Definition
ACS	American Community Survey
CBA	Cost-Benefit Analysis
DEM	Digital Elevation Model
EAD	Expected Annual Damage
EW	Equity-Weighted
EWD	Equity-Weighted Damage
EWEAD	Equity-Weighted Expected Annual Damage
FEMA	Federal Emergency Management Agency
FIAT	Flood Impact Assessment Tool
GIS	Geographic Information Systems
HAZUS	Hazards U.S. Multi-Hazard
NSI	National Structure Inventory
UNSAFE	Uncertain Structure and Fragility Ensemble

1 Introduction

1.1. Background

Flood risk management is fundamental to water resource planning, requiring comprehensive strategies to effectively mitigate potential damage and losses. However, the current risk assessment landscape often lacks equity consideration, leading to disparities in developing prevention measures (Sanders et al., 2023). Typically, decision-making processes rely on cost-benefit analysis, resulting in an uneven distribution of protective measures across flood-prone areas (Ciullo et al., 2020). This analysis focuses on asset losses without accounting for the varying impacts of identical losses on households with different socio-economic statuses. For instance, a $100 loss has a more significant effect on a low-income household than a high-income household.

Equity weighting methods provide a mechanism to adjust disaster impacts, redistributing losses based on income levels (Kind et al., 2017). This approach aims to reflect more accurately the welfare consequences of such losses. Research by Soden et al. (2023) highlights important elements, such as the differential impacts of disasters on vulnerable groups, that are often overlooked when equity weighting is excluded from disaster risk methods. Equity weighting ensures a fair distribution of benefits and burdens among affected communities (Soden et al., 2023).

This perspective is supported by the investigation of equity-weighted metrics, such as the Equity Weighted Expected Annual Damage (EWEAD) proposed by Kind et al. (2017), which considers income differences to create equity weighting factors. This metric attempts to provide a more equitable assessment of flood damage by recognizing the varying capacities of households to recover from such events. Current implementations using equity weighting in flood risk assessment are often applied at specific administrative aggregation levels, such as village levels (Kind et al., 2020) and census sectors (Frontuto et al., 2020; Markhvida et al., 2020; Walsh and Hallegatte, 2020).

However, this aggregation can lead to a failure to capture the true conditions of social vulnerability, as the income determining the equity weighting is averaged out, masking income heterogeneity. Therefore, a household-level analysis is proposed. The motivation for using a more disaggregated method at the household level arises from the need to capture social vulnerabilities that are not visible with aggregate methods. By focusing on individual households, this analysis can reflect more fully the socio-economic disparities, including income heterogeneity, that affect the capacity to recover from flood damage. While increasing resolution is recommended, it also adds complexity to the risk assessment (Ciullo et al., 2020) and introduces uncertainty due to using high-resolution data.

Despite these challenges, the household-level approach is expected to promote a more detailed and equitable understanding of flood risks. The approach will lead to a more effective and equitable distribution of resources and protective measures, enhancing the resilience of vulnerable communities. This study will demonstrate the applicability of the equity weighting method at the household level under various scenarios and explore the sensitivities and uncertainties due to the assumptions and approaches used during implementation.

1.2. Problem definition

Income distribution is a key component in the implementation of equity weighting. Research by Markhvida et al. (2020) indicates that the use of income aggregation levels can distort the characteristics of households in highly heterogeneous areas, which can result in the detection of very low-income households being overlooked.

For example, if there are two areas with the income distributions shown in Figure 1.1, both areas would have the same mean household income, which will be used in the aggregated method. However, this approach would mask differences in the actual distribution of income within each area. In the top distribution, a normal income distribution is shown, where most households have incomes close to the mean. This shows that the average income can reflect the economic status of the area.

In contrast, the bottom distribution in Figure 1.1 shows a skewed log-normal distribution with a long tail towards higher income levels. The distribution has a large proportion of households have below-average incomes, while a small proportion have much higher incomes. However, the aggregated method results in a single average value, which can hide income disparities within this area.

When applying the equity weighting method, since the average income in a skewed distribution does not represent the economic situation of most households, the needs of low-income households can be overlooked and lead to increased inequality. This example highlights the need to incorporate income heterogeneity when implementing equity weighting. It is important to not only look at average income but also understand how income is distributed across different segments of the population.

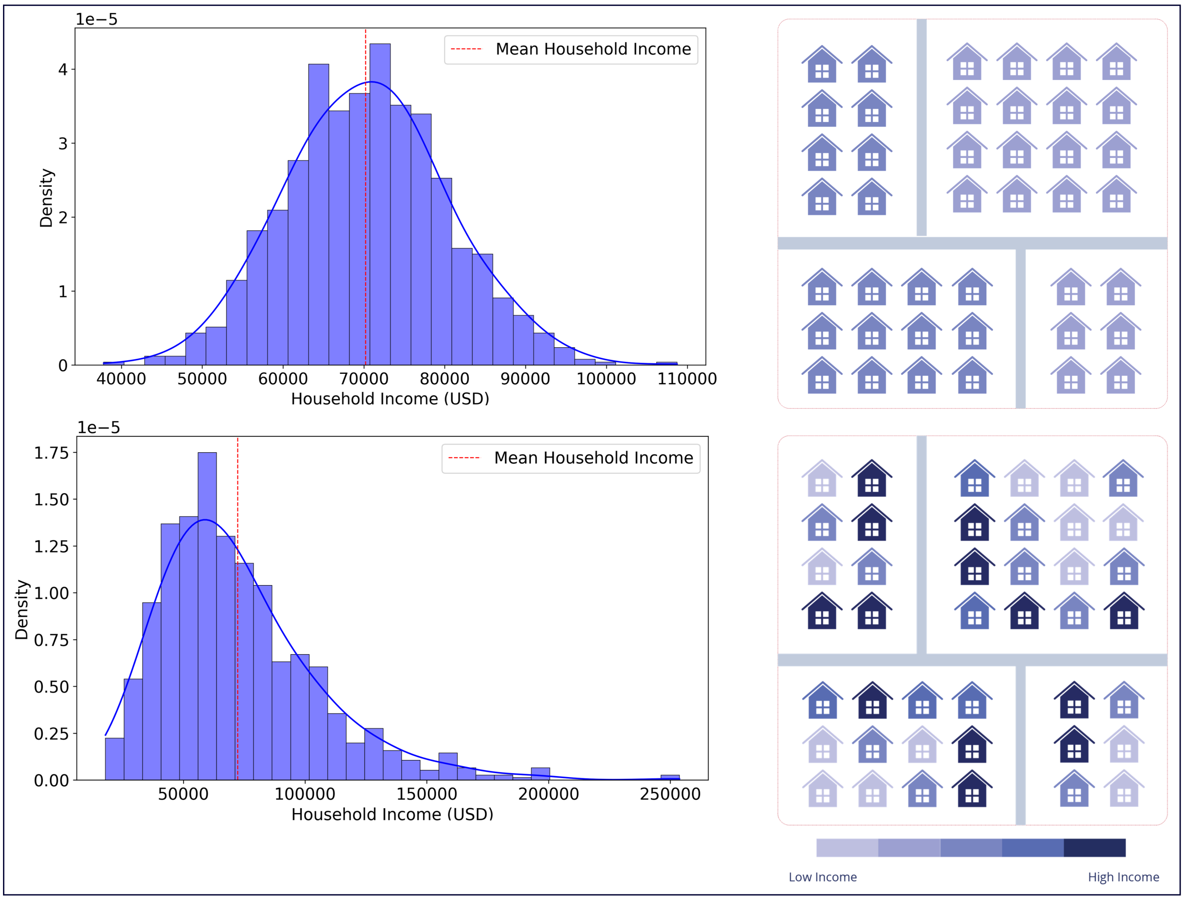


Figure 1.1: Example of household income distribution

Income heterogeneity is the variation in income levels among households in a given area. This variation is important as it shows the diversity of income within an area, including high and low income households.

Incorporating income heterogeneity at a more granular level, such as the household level, also presents challenges. The lack of publicly available income datasets at this level requires assumptions and methods to derive individual household income and allocate household income spatially.

1.3. Research objectives and questions

The objective of this study is to investigate how equity weighting is applied at the household level and how flood risk assessment is affected by taking equity into account at this finer level compared to an aggregated method.

Based on the research objective, the research questions can be formulated as follows:

1. How can the heterogeneity of income be accounted for in practice given limitations on income data due to privacy rules?
   (a) What assumptions can be made?
   (b) What are the uncertainties surrounding this?
2. What is the added value of considering the heterogeneity of income within a census block group for equity-weighted risk and impact estimations?

1.4. Reader guidance

The purpose of this section is to provide an overview of the thesis structure and guide readers to specific sections of interest. The following diagram illustrates the main chapters and their key components, making it easier to navigate through the document.

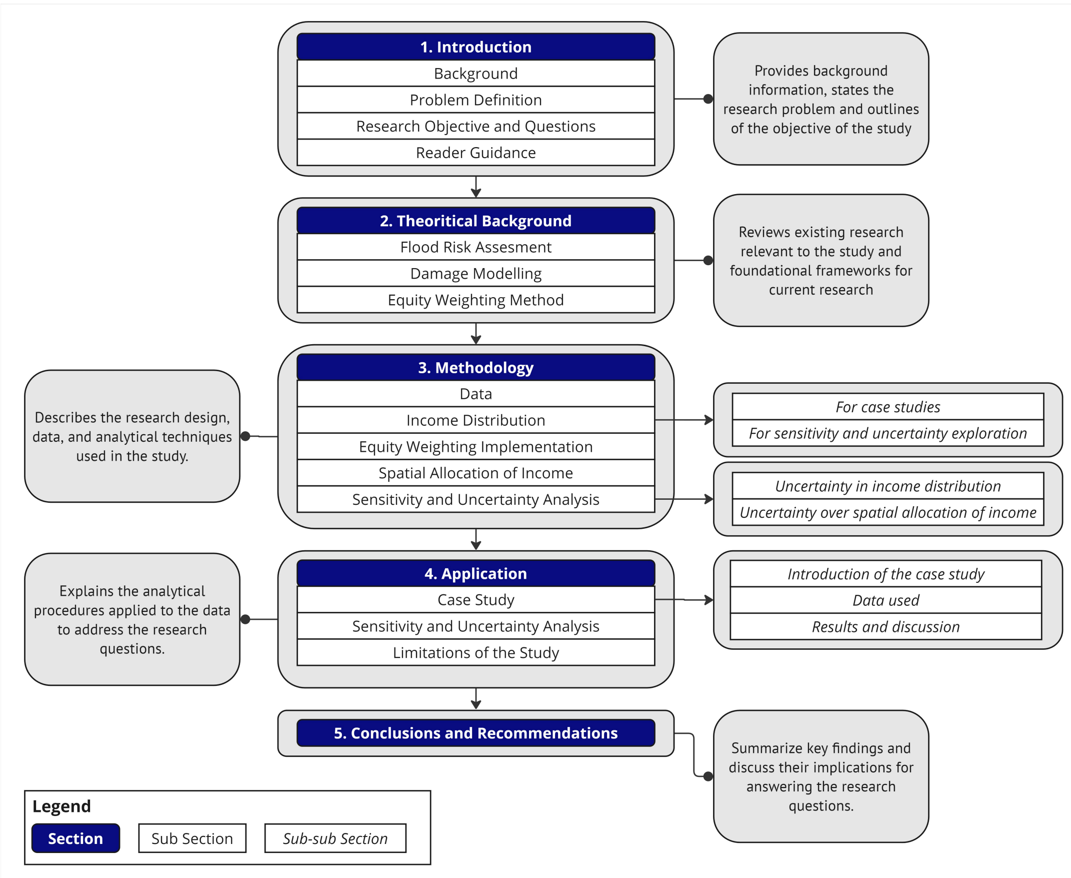


Figure 1.2: Overview of thesis report structure

2

Theoritical Background

This chapter explores the theoretical background underlying the study of flood risk assessment and its associated components. It explains the flood risk assessment process, which informs the estimation of damage and expected annual damages (EAD). Furthermore, the damage modeling technique is investigated, and the use of Delft-FIAT together with software back-end processing is investigated. Finally, the equity weighting method is introduced, with a specific focus on its utility function and relevance within our research context.

2.1. Flood risk assessment

Flood risk assessment is a multifaceted process for water resources planning, which aims to understand the potential impacts of flooding on various elements within a community. The process involves a systematic evaluation of flood hazard, exposure and vulnerability. Hazard assessment focuses on identifying and analyzing flood events' frequency, magnitude, and spatial extent using techniques such as hydrological modeling, hydraulic modeling, and historical flood data analysis, where the appropriate level of detail is crucial to balancing complexity and practicality (Apel et al., 2009).

Exposure assessment involves evaluating the elements at risk, such as populations, buildings, infrastructure, and economic activities. This process often utilizes Geographic Information Systems (GIS) and remote sensing technologies to map and quantify these assets. The validation of flood risk models remains critical to ensure their reliability and accuracy in predicting flood impacts, supporting effective decision-making (Molinari et al., 2019).

Vulnerability assessment determines the susceptibility of exposed elements to flood damage, considering factors such as building materials, construction quality, and the socio-economic status of households. Long-term trends and socio-economic changes significantly influence global exposure to river and coastal flooding, emphasizing the need to incorporate these factors into flood vulnerability assessments (Jongman et al., 2012).

Traditional methodologies in flood risk assessment include deterministic approaches, which use specific flood scenarios to evaluate potential damages, and probabilistic approaches, which estimate the likelihood and impact of various flood events through statistical analyses (Hall and Solomatine, 2008; Merz et al., 2010). A common tool in decision-making is cost-benefit analysis (CBA), which compares the costs of flood mitigation measures against expected benefits in reduced damages.

2.2. Damage modelling

This study uses the Delft Flood Impact Assessment Tool (Delft-FIAT) developed by Deltares (Slager et al., 2016) for damage modelling. Delft-FIAT assesses the potential impacts of flooding on various sectors, including residential, commercial, industrial, and infrastructure. The tool integrates hydrodynamic flood models with economic data to estimate potential damages.

Delft-FIAT is a flood impact assessment tool that does not include exposure or hazard data by default,

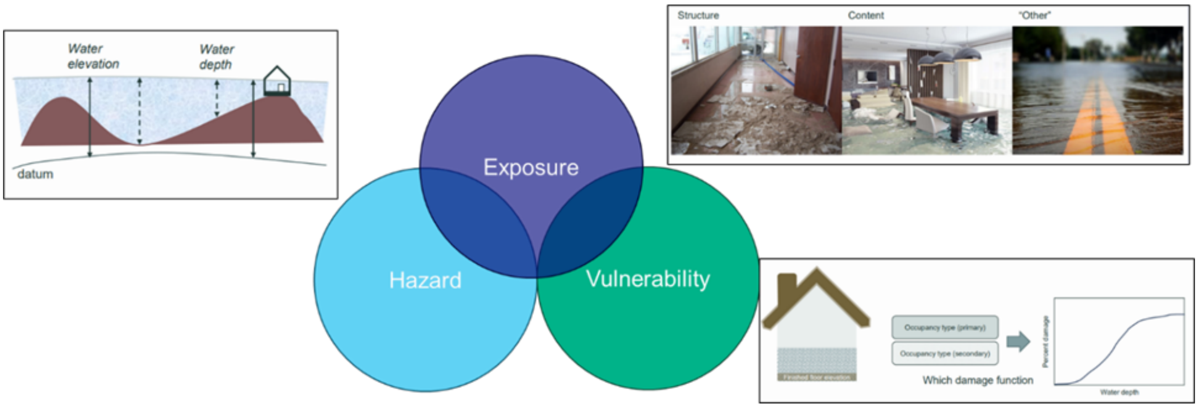


Figure 2.1: Overview of Delft-FIAT Model. Source:
`https://www.deltares.nl/software-en-data/producten/delft-fiat-tool-voor-overstromingseffecten`

which must be supplied by the user. Risk is computed as the product of hazard, exposure, and vulnerability, and Delft-FIAT can be used to calculate monetary values of flood risk (Burzel et al., 2017). The versatile tool can evaluate no mitigation vs. mitigation scenarios by incorporating asset-based regulations into the exposure (Peña et al., 2023). In Delft-FIAT, damage assessment is performed using the unit-loss method, which calculates the direct economic impacts on buildings, utilities, and roads based on flood maps provided by the user. (Burzel et al., 2017).

Damage modelling with Delft-FIAT involves several key components. Hydrodynamic modelling simulates flood scenarios to predict water flow and inundation extents, which are crucial for understanding the spatial distribution and depth of flooding. Exposure analysis identifies the assets and population at risk by mapping buildings, infrastructure, and other critical assets within flood-prone areas using geographic information systems (GIS). Vulnerability assessment evaluates the susceptibility of different assets to flood damage by assigning vulnerability curves or damage functions to various asset types, indicating the expected damage at different flood depths. The vulnerability of human and natural assets is quantified through simplifications, which can lead to uncertainties in the damage modelling (Parodi et al., 2020).

Combining hydrodynamic modelling, exposure analysis, and vulnerability assessment allows for estimating potential damages. Delft-FIAT uses these inputs to calculate the financial impacts of flooding, including repair and replacement costs. The tool also supports the Expected Annual Damage (EAD) calculation, representing the average annual financial loss due to flooding. EAD is computed by integrating the probabilities of different flood events with their corresponding damage estimates, which is essential for cost-benefit analyses of flood mitigation measures.

2.3. Equity weighting method

Equity weighting methods in flood risk management assessments aim to incorporate considerations of utility functions, social welfare, and equity weights to address income disparities and social vulnerabilities.

The utility function follows the law of diminishing marginal utility of consumption, is non-linear and has a concave shape. As consumption decreases, marginal utility or well-being increases Figure 2.2 (Kind et al., 2017). This study assumes that an individual's annual income is fully consumed within the same year, so annual consumption is equal to annual income. The diminishing marginal utility of consumption is important for evaluating social welfare in flood risk assessments in two ways. First, it helps to understand why individuals tend to be risk-averse: an additional consumption unit has more value during low than high consumption periods. Second, it supports the use of equity weights in cost-benefit analyses, where an additional unit of consumption is given greater value for poorer individuals compared to wealthier ones Kind et al. (2017). The utility function is described as follows:

$$U(C) = \frac{C^{1-\gamma}}{1-\gamma}, \quad \gamma \geq 0, \gamma \neq 1 \tag{2.1}$$

Where U is utility, C is consumption, and γ is the elasticity of marginal utility of consumption. The utility is linear if $\gamma = 0$. For $\gamma \geq 0$, the marginal utility of consumption diminishes, resulting in a concave utility function (see Figure 2.2).

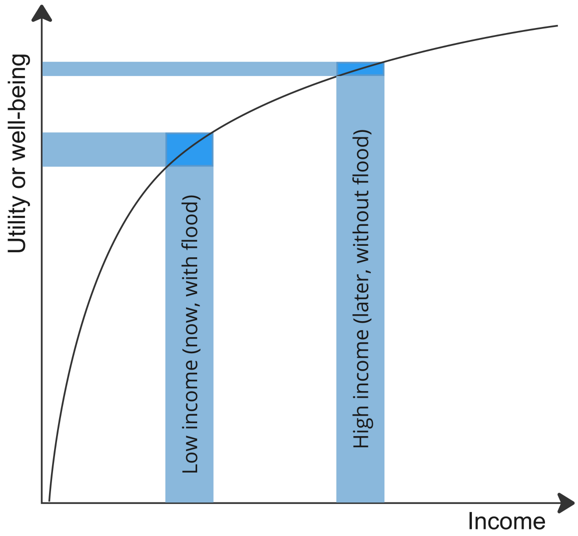


Figure 2.2: Utility function curve. Adapted from Kind et al. (2017).

The elasticity of marginal utility determines the curvature of the utility function, with suggested values ranging from 0.5 to 2.0. The recommended value is the middle value of 1.2 (Kind et al., 2017). In addition, Kind et al. (2017) assumes the same marginal utility elasticity value for everyone, meaning that an increase in income by one dollar has the same value for everyone.

Social welfare functions aggregate individual utilities into a measure of societal well-being. A utilitarian social welfare function is often assumed, summing the utilities of all individuals:

$$W = W(U_1, U_2, \ldots, U_N) \tag{2.2}$$

This approach implies that a marginal increase in income yields the same increase in social welfare regardless of the recipient's income level. This approach is also often modified by introducing equity weights to reflect preferences for equity and fairness. These weights adjust the contributions to social welfare based on the marginal utility of income for different individuals, giving more weight to increases in income for poorer individuals compared to richer ones.

Equity weights ω_Y are applied in cost-benefit analyses to account for income disparities among different groups. These weights are derived from the marginal utility of income, as shown:

$$\omega_Y = \left(\frac{Y_i}{Y_{avg}}\right)^{-\gamma} \tag{2.3}$$

where Y_i is the individual's income, Y_avg is the average income, and $gamma$ is the elasticity of marginal utility. Equity weights ensure that benefits and costs are adjusted to reflect the higher relative income value for lower-income groups. For example, in flood risk assessments, equity weighting means that the damage suffered by lower-income households is more significant than that suffered by higher-income households.

For equity weight is more than one, an additional dollar is valued more for low-income groups than high-income groups, reflecting higher marginal utility for poorer individuals. For instance, an equity weight of 2.28 for a low-income group values a dollar loss for this group as $2.28 in the analysis. When equity weight is equal to one, it indicates that the marginal utility of income is the same for everyone, assuming that a dollar is equally valuable across all income levels.

In evaluating the benefits of equity weighting, Kind et al. (2017) proposed four frameworks that can be used; this study only focuses on two main frameworks, which are the Expected Value Framework or the conventional method and the Equity Weighted Expected Value Framework.

The Expected Value (EV) framework is the most commonly used approach in cost-benefit analyses (CBA) for flood risk management. This framework calculates the benefits based on the expected annual damages (EAD) reduction. The primary objective of the EV framework is to maximize aggregated wealth, assuming a utilitarian social welfare function. The utilitarian approach aims to maximize the total sum of individual utilities across society, where utility is usually tied directly to income or wealth. The framework values each dollar of benefit or cost equally, regardless of who gains or loses that dollar. As a result, the EV framework focuses on overall efficiency. It does not consider social vulnerability or equity concerns, as it treats all individuals' utility similarly, without considering differences in socioeconomic status or ability to recover from flood losses. The EV framework is most suitable under the following conditions:

- **Flood damages are relatively small compared to income**
  When flood damages represent only a small fraction of individual or household income, the impact on overall well-being is minimal. In such cases, the focus on maximizing total wealth without adjusting for social vulnerability is less problematic, as the damages do not significantly alter the utility of those affected.
- **Compensation for flood damages is sufficient**
  If the flood affected receive appropriate compensation to cover their losses, there would be less need to account for risk aversion or equity concerns. Compensation helps restore individuals to their pre-flood financial status, in line with the EV framework's emphasis on aggregate wealth without the need to adjust for different levels of vulnerability.
- **The income distribution is fair or income is redistributed through other means**
  The EV framework assumes that all individuals derive the same utility from an additional dollar, which is more valid in a context where income distribution is relatively equal. Suppose income is already fairly distributed, or there are effective mechanisms for income redistribution. In that case, the framework's lack of focus on equity is less of a concern, as the broader economic system helps balance disparities.

The Equity Weighted Expected Value (EWEV) framework addresses equity concerns by applying equity weights to the expected damages. This weight adjusts the valuation of damages to reflect disparities in income and social vulnerability. Benefits in this framework are valued in terms of the equity-weighted EAD (EWEAD). The EWEV framework is appropriate under the following conditions:

- **Compensation for flood damages is insufficient**
  Unlike the EV framework, which assumes adequate compensation mechanisms, the EWEV framework is specifically designed for situations lacking compensation. In such cases, lower-income groups are more severely impacted due to their limited ability to recover. The EWEV framework addresses this by assigning greater value to the benefits received by these groups, ensuring that flood risk management measures better reflect the actual economic burden on those most affected.
- **Social vulnerability is low**
  The EV framework does not account for social vulnerability, treating all individuals as if they derive the same utility from an additional dollar. In cases where social vulnerability is low, the EWEV framework shifts the focus toward correcting income inequality. Applying equity weights ensures that the benefits of flood risk management are distributed more equitably, particularly favoring those with lower incomes.
- **The income distribution is unfair and not adjusted through other means**
  The EV framework assumes a fair or otherwise adjusted income distribution, which may not always hold. The EWEV framework, on the other hand, is designed to address these imbalances

directly. In contexts where income inequality is significant and not mitigated by other policies, the EWEV framework incorporates equity weights to ensure that flood risk management measures contribute to reducing income inequality, aligning the outcomes with broader social welfare goals.

While the EV framework is commonly used due to its focus on maximizing economic efficiency, it does not consider the different impacts of flood damage across different income groups. The EWEV framework addresses this gap by integrating equity considerations, ensuring that the benefits of flood risk management are distributed to reflect the diverse financial and social vulnerabilities within a community. Shifting from a purely economic perspective to one that also considers social equity allows for more fair decision-making, especially when income inequality and inadequate compensation mechanisms are significant issues.

3 Methodology

This chapter outlines the study methodology. The first section covers the required and available data, including discussing data gaps and the assumptions made. The second section details the income distribution scenarios involving using the dataset directly and fitting a distribution to the data. The third section describes implementing the equity weighting with aggregated and disaggregated methods. The fourth section explains the spatial allocation of income to households. The final section discusses sensitivity and uncertainty analysis related to income distribution and spatial allocation of income within the context of equity weighting.

Figure 3.1 shows the general workflow of this study. The main result of this study is the equity-weighted loss using the disaggregated method, which will be compared with the results obtained using the aggregated method.

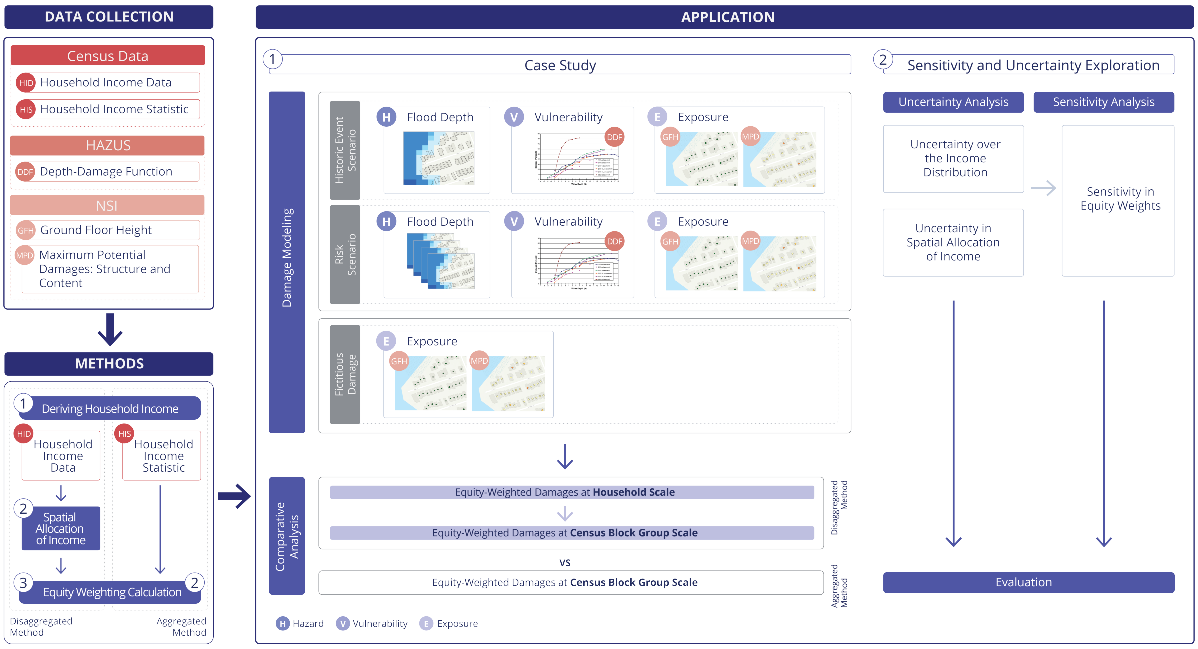


Figure 3.1: General workflow of the study

3.1. Data

In this study, the main processes are deriving income distribution, implementing equity weighting, spatially allocating income, and damage modelling. The necessary data for starting such an analysis will be described.

The dataset to be used for deriving income distribution is collected from the Census Dataset, specifically the product of the American Community Survey (ACS). This dataset includes household income distribution and statistics such as mean, median, and aggregate or total household income. This dataset will be used to derive the income distribution for the disaggregated method.

Household income includes the total income of all individuals aged 15 years and above living in the household, regardless of their relationship to each other, assessed over the past 12 months (Bureau, 2023). Median household income is the midpoint of the income distribution, where half of the households earn more and the other half earn less (Bureau, 2023). Aggregate household income refers to the cumulative income of all households in a defined area (Bureau, 2023). The average household income is obtained from the aggregate household income by dividing it by the total households in a given area.

Other datasets from the Census are per capita income and total population, which will be used for the aggregated method when applying the equity weighting method.

Per capita income represents the average income earned by each individual in a specific area, including all people living in group quarters. It is determined by dividing the total income of the area by its total population, offering a measure of the economic well-being of the residents (Bureau, 2023). The total population includes every individual living in a particular geographical area, thus facilitating a comprehensive assessment of the demographic and economic characteristics of the area (Bureau, 2023).

As shown in Figure 3.2, the most detailed scale dataset available from the Census is the census block groups, which are part of the census tracts. From this, it can be determined that aggregation methods are at the scale of census block groups, while disaggregation methods are within census block groups at the household level.

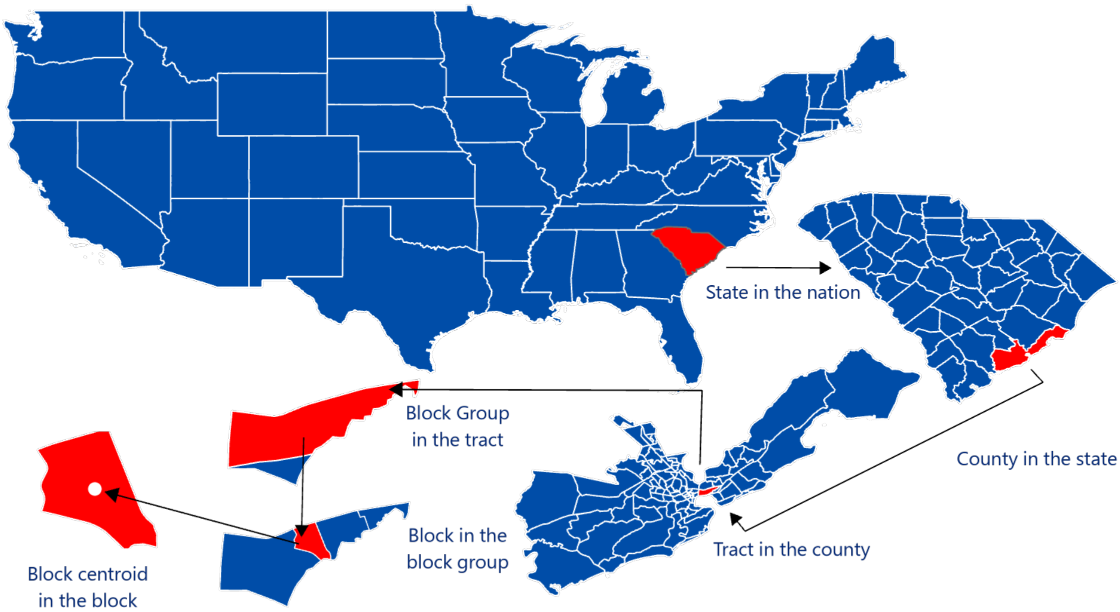


Figure 3.2: Spatial scale of census dataset

In Figure 3.3, the census data only provides the distribution of household income in 16 predefined income brackets. Then, the first and last income brackets are open-ended. Therefore, to derive individual household income, assumptions need to be made. These assumptions will be further explained in the next section on income distribution.

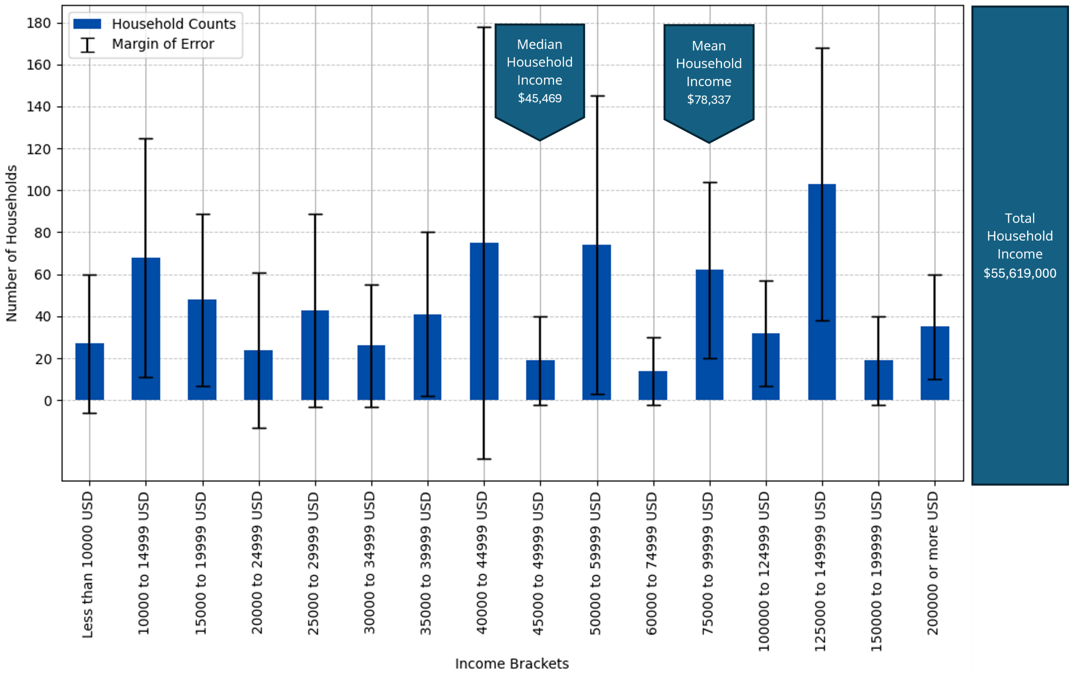


Figure 3.3: Household Income at census block group

Notes: In this context, income refers to the pre-tax cash earnings of the householder and all individuals aged 15 and older residing in the household, regardless of their relationship to the householder (Bureau, 2023). However, it excludes periodic income, such as capital gains, and transfers of goods. Notably, the top income brackets, labeled as '$200,000 or more' represents a broader range of incomes compared to the groups that cover most of the distribution. Additionally, the 'Less than $10,000' brackets includes households with zero or negative income.

The National Structure Inventory (NSI) dataset contains points representing several types of structures, such as residential, commercial, industrial, and public buildings. In this study, only residential structure types will be used. The structure and content values from this dataset will be used as inputs for damage modeling, and only the structure values will be used to allocate income distribution spatially. The occupancy type in the points will be used to link the value of total housing units from the census dataset. Figure 3.4 provides an example of a building point from this dataset.

No.	Variable	Value
1	Structure Value	$ 577,112
2	Content Value	$ 288,556
3	Occupancy Type	RES3A
4	Foundation Height of the structure	2 feet

Figure 3.4: Example of building points from the NSI dataset

The spatial scale of this dataset is based on structure inventories, which need to be downscaled when applying the equity weighting with the disaggregated method. The downscaling will be explained in the section on equity weighting implementation.

For damage modeling, the required input parameters are specified as follows:

1. Hazard: Multiple maximum water depths of the respective return period flood or maximum water depth of a single flood event, depends on which scenarios will be utilized
2. Exposure: Maximum potential damage to the structure and content, and ground floor height of the building
3. Vulnerability: Vulnerability data includes physical vulnerability and socioeconomic vulnerability that influence susceptibility to flood damage

3.2. Income distribution

Two methods will be utilized for determining individual household income. First is to use directly from the census dataset and the second method is trying to fit the distribution by using the income statistic (see Appendix A for more details on the explanation of the method)

3.2.1. For case studies

First, the assumption of using uniform distribution within predefined income brackets will be assessed. The assessment is based on a comparative analysis with household income statistics from the census dataset, such as the mean, median, and total household income in the census block groups (see Appendix A.1.3).

Optimization was performed by changing the lower and upper limits of the open income brackets that were not determined by the census dataset, with the optimization objective being to minimize the sum of squared errors between the household income statistics and the generated household income statistics. Afterward, the optimized lower and upper limits are used to perform individual household income calculations.

The main objective of this comparison and optimization process is to mimic better the census dataset in distributing household income into individual household income. By comparing various income metrics, the goal is to achieve a more accurate distribution of household income across different census blocks. The optimization ensures that the income distribution for each household closely matches the census data, thereby improving the reliability of the income estimates.

3.2.2. For sensitivity and uncertainty exploration

Another method used to generate individual household income is to use Monte Carlo simulation using a log-normal distribution. This distribution is suitable as it inherently produces non-negative values, fitting the nature of income data. It is often used in economics due to its ability to model the skewness of real-world income distributions. Empirical studies confirm that the log-normal distribution approximates income distribution well across various social classes (Bartosˇova, 2006; Battistin et al., 2009).

In the context of the United States, studies have fitted U.S. family income data to log-normal distributions among other models, demonstrating that the log-normal distribution can capture the variability in income data across different years (Azzalini et al., 2002). This study further supports the suitability of the log-normal distribution for modeling household income.

To determine the log-normal distribution parameters, μ and σ must be determined, this study uses the log transform method. The log transform involves taking the natural logarithm of the dataset and transforming the data into a normal distribution. The parameters μ and σ are then calculated from the mean and standard deviation of the log-transformed data. For more details, see Appendix A.2.1.

3.3. Equity weighting implementation

The implementation of equity weighting for case study will be conducted with two methods aggregated and disaggregated. The aggregated method is defined as the analysis conducted at the census block group scale and the disaggregated method is conducted at the household scale.

Figure 3.6 shows the detailed workflow of how to implement equity weighting in the aggregated and disaggregated methods.

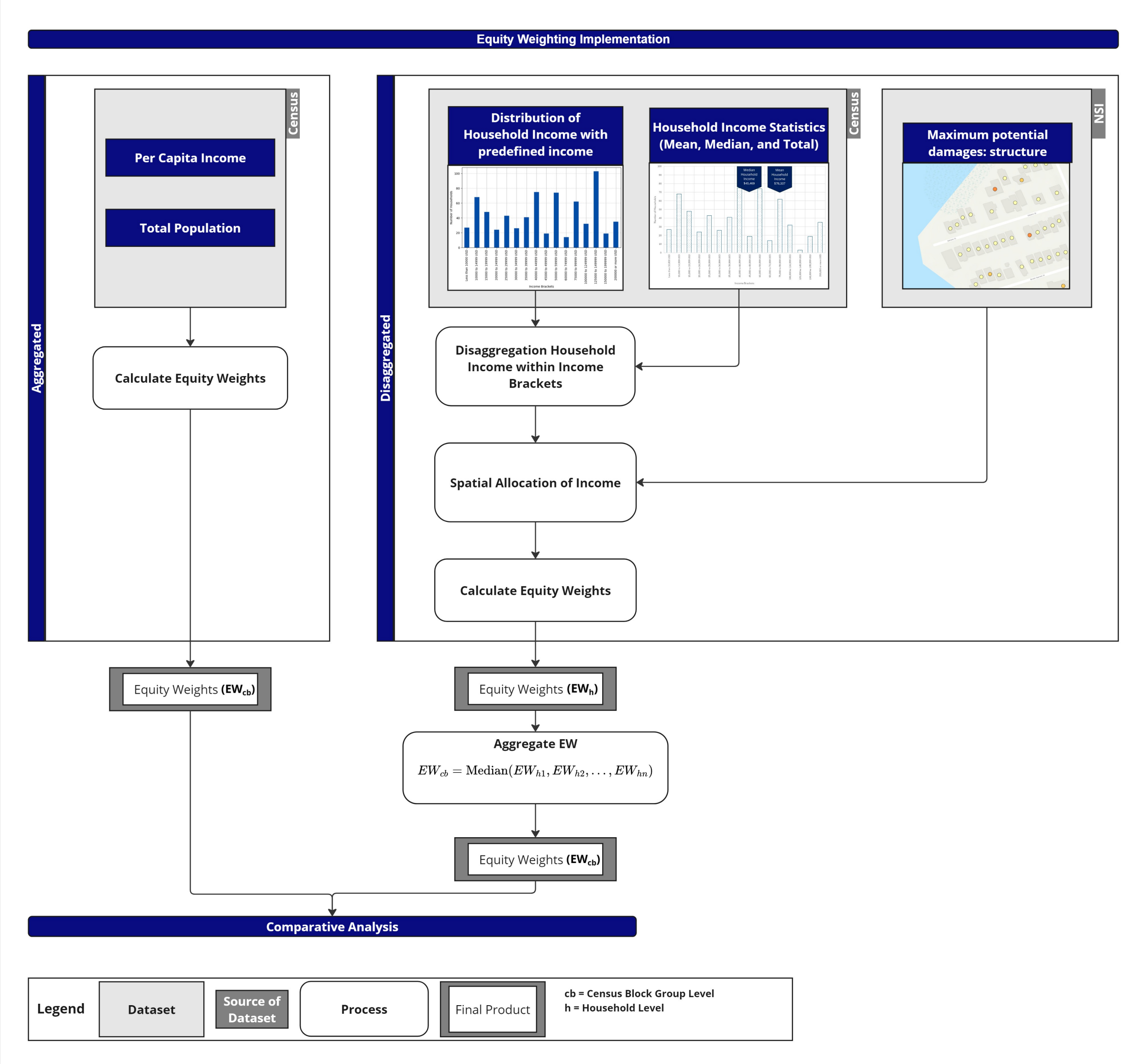


Figure 3.5: Workflow of equity weighting implementation

To apply the equity weighting method in flood risk assessment practice, a case study with multiple scenarios will be conducted. Figure 3.6 describe the detail workflow of the implementation of case study.

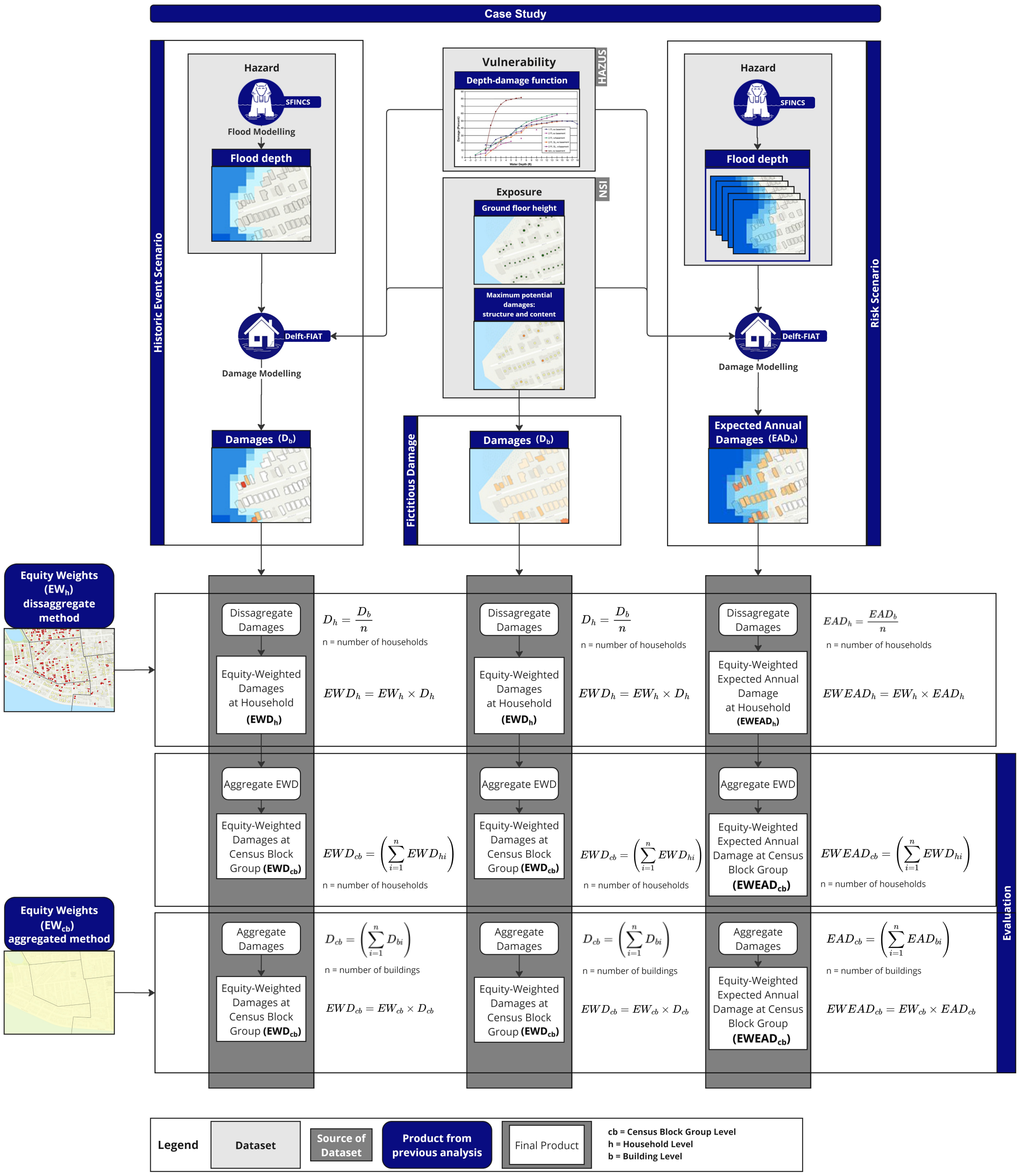


Figure 3.6: Workflow of case study

The explanation of the equations used in the workflow in Figure 3.6 will be explained as follows:

To calculate the equity weighting method for aggregated method, the equation 2.3 will be implemented as follows:

$$EW_{cb} = \left(\frac{I_p}{I_{pwavg}}\right)^{-\gamma} \tag{3.1}$$

where:

- EW_{cb} is the equity weighted factor for aggregated income or census block group income,

- I_p is the per capita income in the census block group,
- I_{pwavg} is the weighted average per capita income for Charleston County, with the weight being the total population of each census block group.
- γ is the elasticity of marginal utility of income

To be able calculate the equity weighting method at the household level, the equation 2.3 of equity weight will be adapted as follows:

$$EW_h = \left(\frac{I_h}{I_{havg}}\right)^{-\gamma} \tag{3.2}$$

where:

- EW_h is the equity weighted factor for household income,
- I_h is the household income,
- I_{havg} is the average household income for the Charleston County,
- γ is the elasticity of marginal utility of income

This EW will be compared with the EW calculated at the census block group scale. To evaluate both, aggregation from the household level is required, here is how to aggregate the EW from the household scale.

$$EW_{cb} = \text{Median}(EW_{h1}, EW_{h2}, \ldots, EW_{hn}) \tag{3.3}$$

where:

- EW_{cb} is the equity weight for the census block group.
- $EW_{h1}, EW_{h2}, \ldots, EW_{hn}$ are the equity weights for individual households within the census block group. Here, EW_{h1} represents the equity weight for the first household, EW_{h2} for the second household, and so on, up to the n-th household within one census block group.

To apply the Equity-Weighted Expected Annual Damage (EWEAD) and Equity-Weighted Damage (EWD) metrics, the spatial scale of the damage must be adjusted. The scale of damages should be the same as the equity weighting factor. Since the damages obtained from damage modeling are at the building scale, disaggregation is required. In this study, the damages are evenly distributed among all households living in the same building. The adapted calculation for the metrics is as follows:

For risk scenario:

$$EWEAD_h = EW_h \times \frac{EAD_b}{n} \tag{3.4}$$

For flood event-based damage:

$$EWD_h = EW_h \times \frac{D_b}{n} \tag{3.5}$$

where:

- $EWEAD_h$ is the equity weighted expected annual damage at household level,
- EWD_h is the equity weighted damage at household level,
- EW_h is the equity weighted factor for household income,
- D_b is the damage at building level,
- n is number of households

After calculating the household scale, the results will be aggregated to the census block group scale. This aggregation is needed to conduct a comparative analysis between the methods. The aggregation process is described as follows:

For risk scenario:

$$EWEAD_{cb} = \left(\sum_{i=1}^{n} EWEAD_{hi} \right) \tag{3.6}$$

For flood event-based damage:

$$EWD_{cb} = \left(\sum_{i=1}^{n} EWD_{hi} \right) \tag{3.7}$$

where:

- $EWEAD_{cb}$ is the equity weighted expected annual damage at household level,
- EWD_{cb} is the equity weighted damage at household level,
- $EWEAD_{hi}$ is the equity weighted expected annual damage for household i,
- EWD_{hi} is the equity weighted damage for household i,
- n is number of households

3.4. Spatial allocation of income

The input parameter used to spatially distribute household income is the maximum potential structure damage or structure replacement cost. This parameter is assumed to have a linear relationship with income. This allows sorting from the lowest to the highest structure replacement cost and household income. Therefore, the household with the lowest income will be assigned the lowest structure replacement cost.

Figure 3.7 shows the workflow of how the spatial allocation of income is implemented.

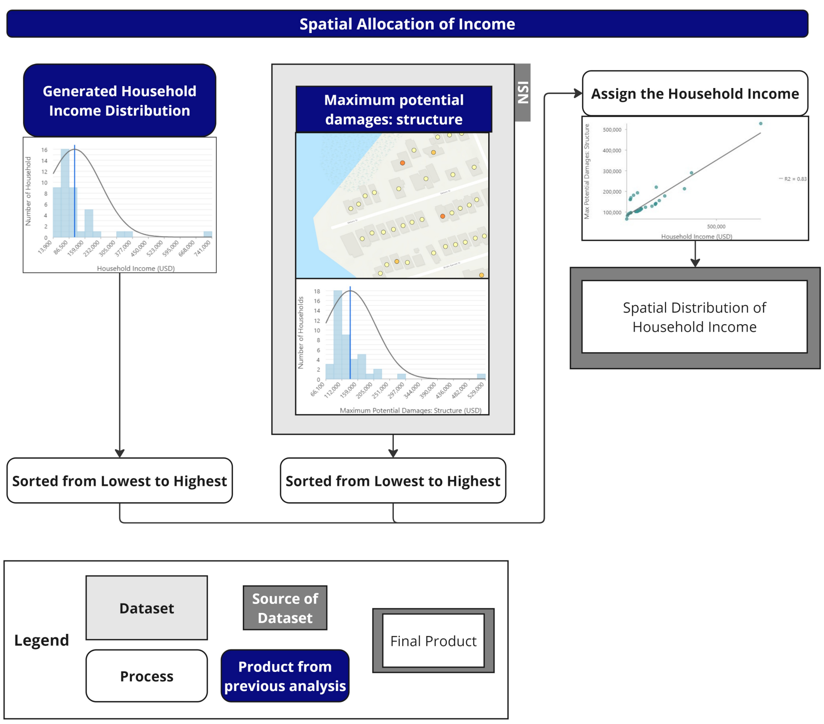


Figure 3.7: Workflow for spatial allocation of income

From the NSI dataset, the occupancy type for each building (point) was determined (U.S. Army Corps of Engineers, 2022). This occupancy type determines the number of housing units. A housing unit is defined as a house, apartment, mobile home, group of rooms, or a single room that is occupied (or, if vacant, intended to be occupied) as an independent living space (Bureau, 2023) and household refers to all people who occupy a housing unit (Bureau, 2023). The dataset only provides the type and not the number of housing units in each type. Table 3.1 describes the assumed number of housing units.

Table 3.1: Occupancy type of building point and assumption

Occupancy Type	Description	Assumption Unit
RES1	Single Family Residential Structures	1
RES2	Mobile Home	1
RES3A	Multi-Family Housing Duplex	2
RES3B	Multi-Family Housing 3-4 Units	3.5
RES3C	Multi-Family Housing 5-9 Units	7
RES3D	Multi-Family Housing 10-19 Units	14.5
RES3E	Multi-Family Housing 20-49 Units	34.5
RES3F	Multi-Family Housing 50+ Units	50

Note: The assumption unit represents the median value of the number of units within each occupancy type category, except the last type of occupancy

From the Census dataset, the total number of households and housing units within the census block group was estimated (Bureau, 2022c,g). This dataset was used to calculate the ratio of housing units to households, which will be used for linking with the NSI dataset.

A process of optimization, normalization, and random correction was performed to link the NSI dataset with the census dataset and determine the specific locations to assign individual household incomes. The method of Bick et al. (2021) implements normalization and random correction on undistributed housing units. However, the research also included an additional dataset from CoreLogic, which provided estimates of housing units in some building parcels in their study area. Since such data is unavailable in this study, optimization is performed first before normalization and random correction.

The optimization aims to adjust the initial values applied to each occupancy type in each census block group according to the housing unit statistics available from the census. In this process, the assumed values shown in Table 3.1 were used as initial values. Several occupancy type values, such as RES3B, RES3C, RES3D, and RES3F, were optimized. The objective function was to minimize the sum of squared errors between the total housing units from the census and the total housing units generated. This optimization resulted in a refined set of housing units per occupancy type for each census block group (see Appendix B for more details).

After the optimization, normalization and random correction were performed using the same method from Bick et al. (2021) with some adjustments. The input directly uses the number of households per building instead of housing units per building and adds a prioritization rule when performing random correction (see Appendix B).

3.5. Sensitivity and uncertainty analysis

To analyze the sensitivity and uncertainty of the determination of individual household income using Monte Carlo simulations with log normal distribution as described in section 3.2.2. Sensitivity analysis will be conducted on the EW variable with variations in household income distribution. The uncertainty analysis will focus on the uncertainty of the income distribution and spatial allocation of income. Figure 3.8 shows the sensitvity and uncertainty exploration workflow.

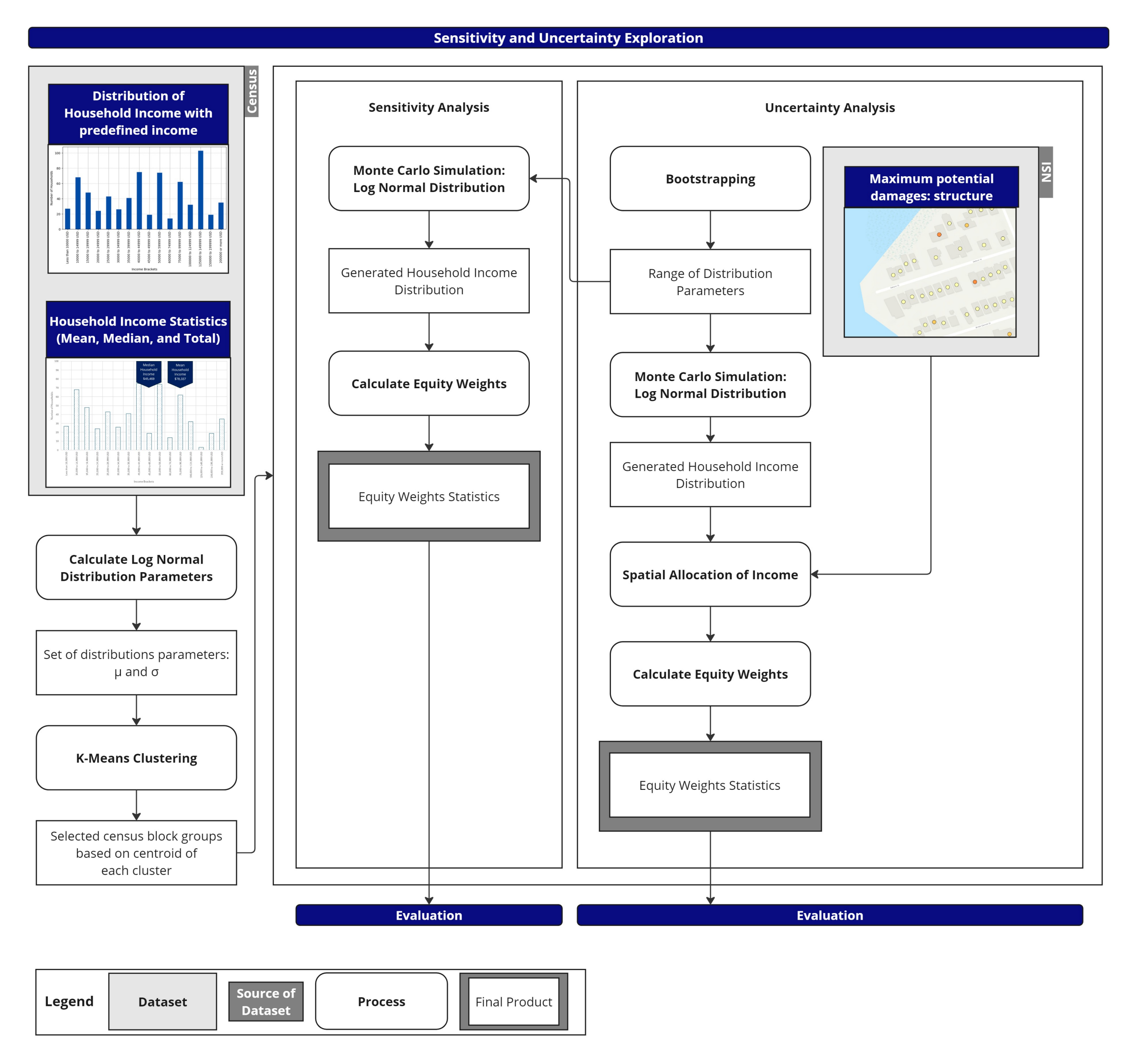


Figure 3.8: Workflow for sensitivity and uncertainty exploration

First, the log-normal distribution parameters, μ, and σ are calculated for all census block groups, resulting in a set of distribution parameters. Second, the data is clustered using k-means clustering to facilitate sensitivity and uncertainty analysis. The elbow and silhouette methods determine the optimal number of clusters (k). This process identifies the clusters and their centroids.

The census block groups with parameter distributions closest to these centroids are then selected for further uncertainty and sensitivity analysis.

3.5.1. Uncertainty over the income distribution

The bootstrapping method will be applied to the parameters μ and σ of the income distribution. This involves repeatedly sampling with replacement from the original data to create multiple simulated samples, allowing for the estimation of the distribution of these parameters and assessing their variability.

James et al. (2021) highlighted that bootstrapping is a versatile and effective statistical method for measuring the uncertainty of estimators or statistical learning models. The method allows estimation of the standard errors of the coefficients and can be applied to a wide range of statistical learning methods.

For this analysis, the number of bootstrapping samples is set to 1000. This is a common choice to minimize experimental randomness and achieve reliable results, as supported by previous studies (Davidson and MacKinnon, 2000; Dwivedi et al., 2017).

3.5.2. Uncertainty in spatial allocation of income

The uncertainty analysis of the spatial allocation of income will focus on the variable of maximum potential damage to structures, which serves as the basis for the spatial allocation of individual household income. Part of the UNSAFE (Uncertain Structure and Fragility Ensemble) framework will be utilized to assess the extent of uncertainty in these values, as described by Pollack et al. (2024).

The UNSAFE framework uses the maximum potential damage to structures, or structure values, from the NSI, considered depreciated replacement costs. This framework introduces uncertainty by assuming a normal distribution around the structure values, with the additional assumption that these values are mostly unbiased and precise (Pollack et al., 2024). A lower bound was included to ensure structure values do not fall below $1.

The value of a structure v_i is represented with uncertainty as follows:

$$v_i \sim \begin{cases} N(\text{NSI\_val}_i, 0.2 \times \text{NSI\_val}_i) & \text{if } v_{i,j} \geq 1 \\ 1 & \text{otherwise} \end{cases} \tag{3.8}$$

Where NSI_val$_i$ denotes the structure value provided in the NSI record as-is, and $v_{i,j}$ represents the value of structure i in the j-th realization of an ensemble of plausible realizations.

The ensemble approach introduces normally distributed noise around the NSI value, with a standard deviation of 20% of the NSI value (Pollack et al., 2024). This standard deviation figure was originally derived from the assumption that NSI follows the model of the automated valuation model for property market transactions (Krause et al., 2020).

Results from (Krause et al., 2020) show that about 40% of the estimated structural value can be captured by a linear model within 10% of its value, and about 83% can be captured within 30% of its value. Based on this information, a standard deviation of 20% was chosen as a balanced estimate to introduce noise, ensuring that the variability reasonably reflects the uncertainty captured by the model (Pollack et al., 2024).

The number of ensembles to be used is 10,000. This method will allow the analysis to show how structure valuation uncertainties impact the spatial allocation of income.

4
Application

This chapter presents the application of the equity weighting method. The first section presents the general case study information, the results, and a discussion of this case study analysis. The second section describes the sensitivity and uncertainty analysis across the main process in this application. The last section explains the limitations of this study.

4.1. Case study

4.1.1. Introduction to the case study

The case study area selected for this research is Charleston County, State of South Carolina, United States, as shown in Figure 4.1. The county has a diverse socioeconomic landscape and varying levels of exposure to environmental risks. It offers a unique opportunity to explore the implications of income heterogeneity on risk assessment and management due to several factors. One key reason is the availability of detailed census data, which allows for a comprehensive analysis of how different income groups are affected by and respond to various risks. Additionally, Charleston County's mix of high-income and low-income households, with a Gini index value of 0.5011 (Bureau, 2022j) higher than the overall United States average in 2022 (Bureau, 2022j) highlights the pronounced income inequality within the region. This disparity is critical in understanding the challenges and disparities in risk exposure and vulnerability among different socioeconomic groups.

Another consideration for selecting this case study location was based on the framework described in section 2.3, which states that equity weighting will place a value on benefits when several criteria are achieved in the study area. This framework further supports the choice of Charleston County as an ideal location to investigate the intersection of income heterogeneity and risk management strategies.

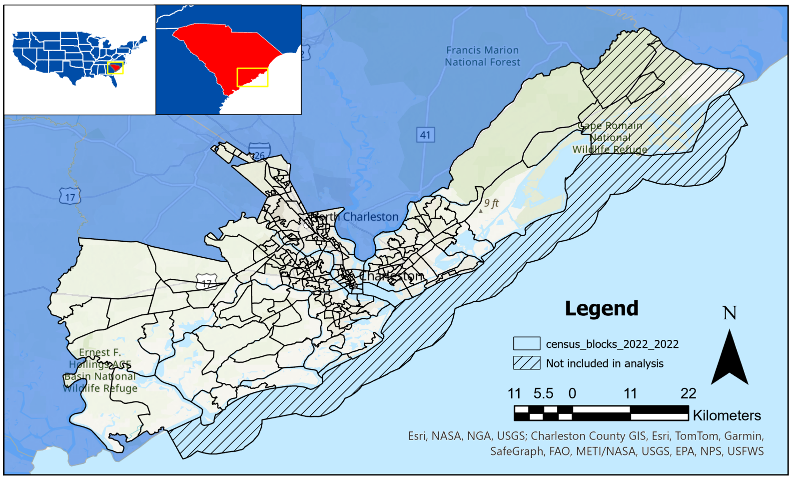


Figure 4.1: Location of Case Study Area

The case study area is generally divided into 261 census block groups (U.S. Census Bureau, 2022), with a total population of 410,000 people in 2022 and a median household income of $78,795. Total households are 179,597 with 201,884 housing units (Bureau, 2022i).

Using this case study, the application of equity weighting and equity-weighted damages at two scales, household scale (disaggregated) versus census block group scale (aggregated), will be compared and evaluated. Specifically, this case study will examine the following:

1. Equity weights calculated using disaggregated method compared to those calculated using aggregated method.
2. Equity-weighted damages based on household income versus those based on aggregated method under various scenarios:
    - A fictitious damage scenario with the same relative damage, the relative damage was set to 10% for all households.
    - Damages observed during a simulated historic flood event.
    - Expected annual damages as determined through a comprehensive risk assessment.

By exploring these comparisons, this case study aims to provide insights into the added value and potential limitations of incorporating income heterogeneity at different spatial scales in flood risk assessment.

The spatial scale used in this case study analysis will be determined by two methods: the aggregated method, which uses the scale of census block groups, and the disaggregated method, which uses the scale of households within census block groups. For the comparative analysis, the results from the disaggregated method will be aggregated to the same spatial scale as the aggregated method, which is at the census block group level. To perform aggregation will follow the section 3.3.

As mentioned in section 3.2, the income distribution in this case study analysis will only use methods that directly use census datasets. The uncertainty of those datasets will be explored in section 4.2.

4.1.2. Data used

In this case study, several datasets were used to perform damage modeling with Delft-FIAT, incorporating input parameters of hazard, exposure, and vulnerability. The specific datasets used include:

Hazard Data:

- The study uses flood hazard inputs from validated simulations conducted by Deltares using the SFINCS (Super-Fast INundation of CoastS) model. SFINCS efficiently calculates compound floods in coastal systems caused by fluvial, pluvial, tidal, wind, and wave processes, resulting in flood depth maps (Leijnse et al., 2021).
- Due to the limited area of the flood model, some census block groups in Charleston County are excluded from the analysis.

Exposure Data:

- This dataset includes information on the distribution and characteristics of residential buildings within the study area. It incorporates data on ground floor height and occupancy type of the structures, obtained from the National Structure Inventory (U.S. Army Corps of Engineers, 2022).
- The digital elevation model (DEM) from U.S. Geological Survey (2019) is used to derive the ground elevation.
- The maximum potential damage to structures and contents is derived from the occupancy type data and projected with the depth damage function from (Federal Emergency Management Agency (FEMA), 2022), based on maximum flood water depth.

Vulnerability Data:

- Physical vulnerability data is collected using the HAZUS depth damage function from (Federal Emergency Management Agency (FEMA), 2022).

- Socioeconomic vulnerability is assessed using income distribution, statistics of income, and total population from the census data. Further details are provided in Table 4.1.

Table 4.1: Census dataset codes and descriptions used in the study

No.	Code	Data Name	Period
1	B19001	Household Income in the Past 12 Months (in 2022 Inflation-Adjusted Dollars)(Bureau, 2022c)	2022
2	B19013	Median Household Income in the Past 12 Months (in 2022 Inflation-Adjusted Dollars)(Bureau, 2022a)	2022
3	B19025	Aggregate Household Income in the Past 12 Months (in 2022 Inflation-Adjusted Dollars)(Bureau, 2022d)	2022
4	B25001	Housing Units(Bureau, 2022g)	2022
5	B19301	Per Capita Income in the Past 12 Months (in 2022 Inflation-Adjusted Dollars)(Bureau, 2022f)	2022
6	B01003	Total Population(Bureau, 2022b)	2022

Socioeconomic vulnerability will not be used in damage modeling with Delft-FIAT, but will be used when applying the equity weighting.

4.1.3. Results and discussion

This sub-section explains the results and discussion based on the case study described in the previous section. It starts with the generated household income, spatial allocation of income, calculation of equity weights, and scenarios used to demonstrate the application of the equity weighting method in flood risk assessment.

The discussion will be based on a comparative analysis between the aggregated and disaggregated methods.

Income distribution

First, the household income distribution is derived to determine individual household income. The income distribution is equally distributed across all the predefined income brackets, using optimized lower and upper limits for the open-ended income brackets (see Appendix A for comparing the approaches). This method is applied to each census block group across the study area.

In Figure 4.3, the box plot shows the distribution of household income per census block group. Almost all household income statistics (represented by blue dots and green diamonds) of the census match the optimized income results. The mean optimized income per census block group is shown in the red line with crosses indicating that all averages correspond to the census.

For median household income, 21 out of 259 census block groups could not be compared with household income statistics as median household income was unavailable in the census.

The performance metrics for the optimization across all income statistics (mean, median, and total household income) for all census block groups are as follows: $R^2 = 0.99$, with mean RMSE values of 535.3 USD for the mean household income, 6,498.45 USD for the median household income, and 219,173.66 USD for the total household income.

The highest household income from the optimization results is 2,935,402 USD, and the lowest is 1,107 USD. Compared with the available census data, the highest of the top 5 percent mean household income by census tract for the entire Charleston area is 1,840,320 USD ($\pm$ 839,565 margin of error). The lowest quintile mean household income by census tract for the entire Charleston area is 1,062 USD ($\pm$ 20,684 margin of error) (Bureau, 2022e). Therefore, the results of the optimization are still within this range.

Income per capita was used for the aggregated method, but for comparison purposes, income per capita was converted to household income using the following equation:

$$\text{Household Income} = \text{per capita income} \times \frac{\text{total population}}{\text{total households}} \tag{4.1}$$

In Figure 4.3, the purple squares represent income per capita. The figure indicates that for some census block groups, the converted household income does not match the mean household income from the same source dataset. The discrepancy arises due to differences in calculation and aggregation methods. Income per capita is calculated by dividing total income by population, giving an average income per individual. In contrast, household income aggregates the total income of all members in a household, reflecting combined earnings and household size (Bureau, 2023).

Spatial allocation of income

This spatial allocation of income is only used in the disaggregated method (household scale) that needs to determine where individual household income is spatially distributed within census block groups.

Before conducting the spatial allocation of income, matching the total housing units from the Census and NSI data is required to obtain consistency between the household income and spatial data. Several steps have been taken to determine how many households are in each building point. These steps are detailed in Appendix B.

Figure 4.2 shows the realized spatial allocation of income, the higher the income (larger black dots), the greater the potential for structural damage or structural replacement costs (darker green in the building footprint).

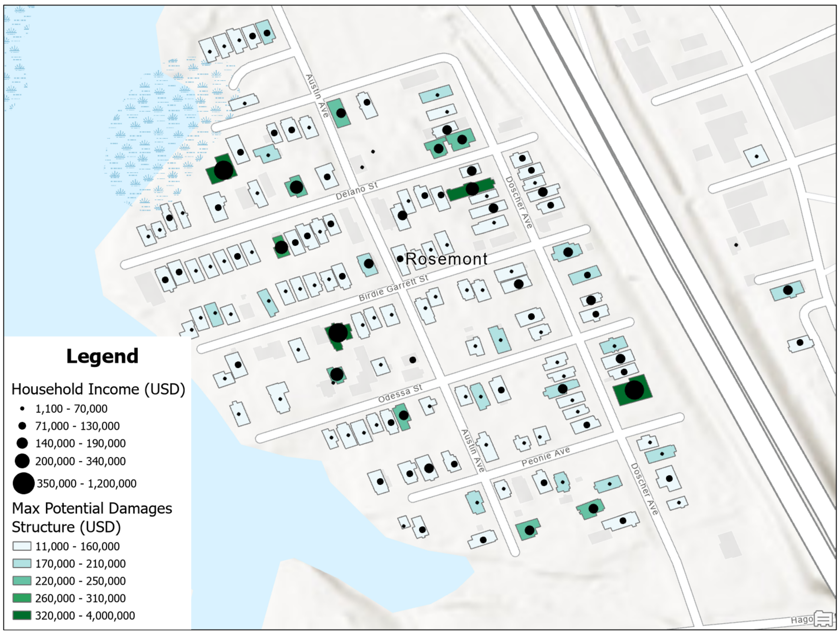


Figure 4.2: Example of implementing spatial allocation of income in Rosemont Neighborhood, Charleston County

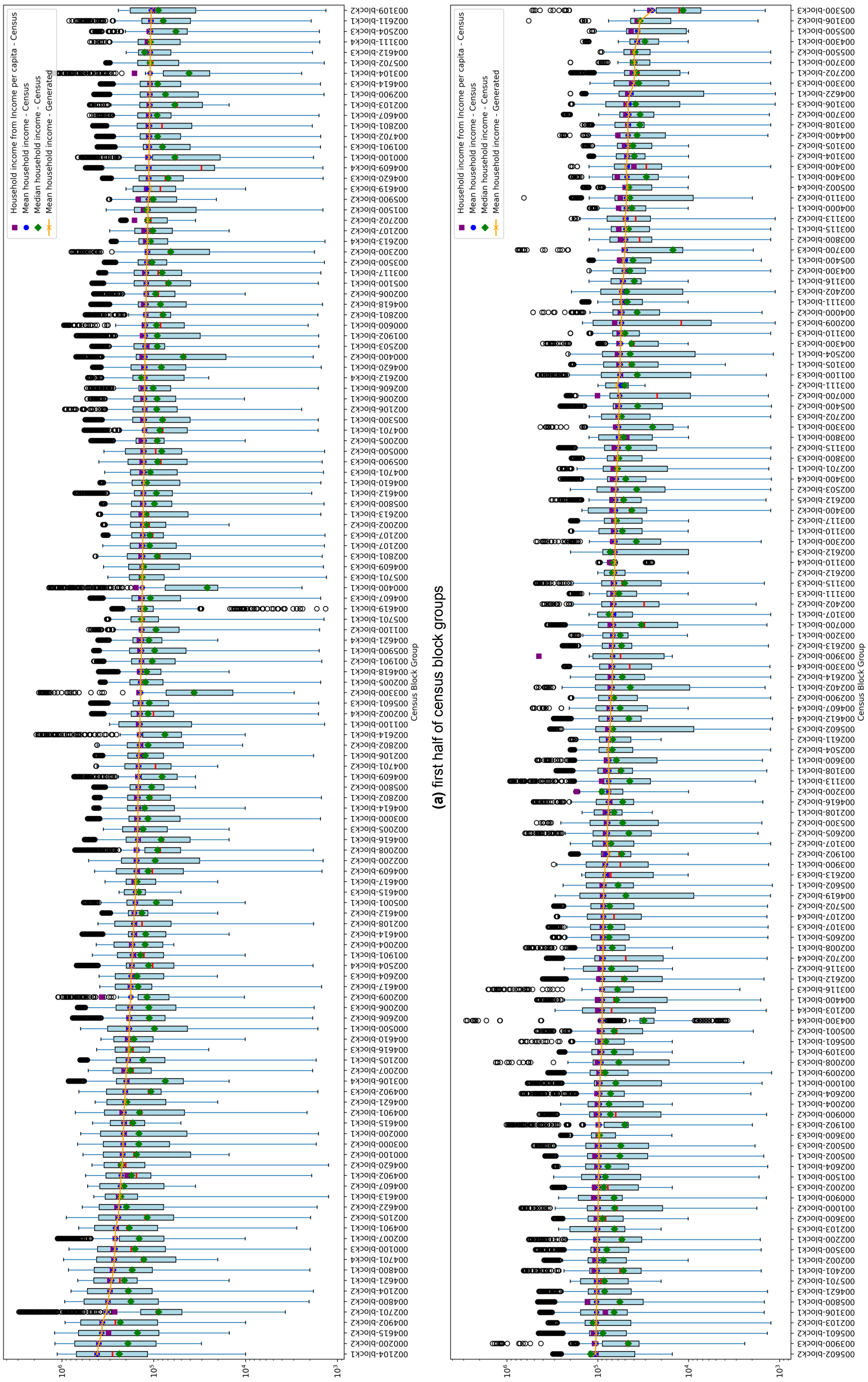
Equity weights

Equity weights are calculated using Equation 3.2 for the disaggregated, and using Equation 2.3 for the aggregated method.

For comparing the equity weights of the disaggregated method and the aggregated method, the disaggregated method needs to be aggregated first using the equation 3.3 which takes the median EW value for each census block group. The median value was selected because it reduces the impact of outliers on the aggregation.

As mentioned in section 2.3, equity weighting has a threshold that is also related to the utility function of Figure 2.2. Therefore, in the comparative analysis of EW between these two methods, thresholds of more than one and less than one will be applied.

As shown in Figure 4.4, incorporating income heterogeneity within census block groups using household income reveals that the Equity Weight (EW) values vary between greater than 1 and less than 1 on this scale. The number of census block groups with an EW value greater than 1 increases to 212 out of 259, approximately 82%, representing an increase of about 25% compared with aggregated method.

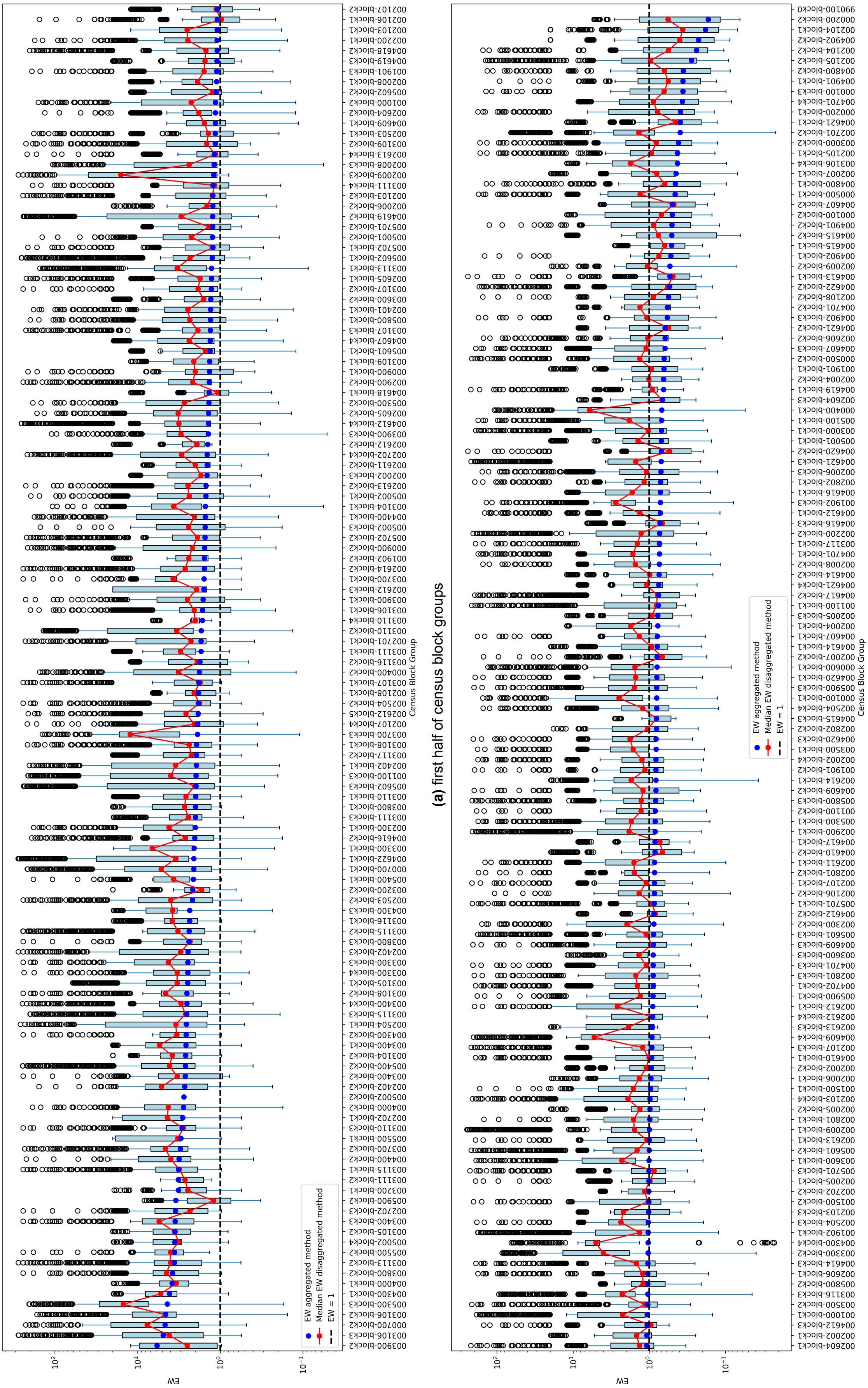


(a) first half of census block groups

(b) second half of census block groups

Figure 4.3: Generated household income distribution (box plot) and mean generated household income (orange line with crosses) across census block groups in Charleston County, compared with household income statistics (green diamonds for median, blue dots for mean)

Notes: purple squares shows the converted income per capita to household income



(a) first half of census block groups

(b) second half of census block groups

Figure 4.4: Equity Weights (EW) distribution: disaggregated method (box plot) vs. aggregated method (red dotted line)

Figure 4.5 illustrates the changes in census block groups where the EW metric shifted from less than 1 to greater than 1, or vice versa.

Figure 4.5: EW map with aggregated method (left) and Median EW map with disaggregated method (right), zoomed in on Charleston Peninsula region

To quantify the differences between the methods, the ratio of EW can be calculated as follows:

$$\text{Ratio of EW} = \frac{\text{EW}_2}{\text{EW}_1} \tag{4.2}$$

where EW_1 is the EW aggregated method, and EW_2 is the median EW disaggregated method.

Figure 4.6 shows the distribution of EW ratios across census block groups in the extent of the Charleston Peninsula area, with a color-coded scale indicating the magnitude of these ratios. Darker shades represent larger deviations from a ratio of 1, indicating significant differences between methods, while white indicates minimal differences.

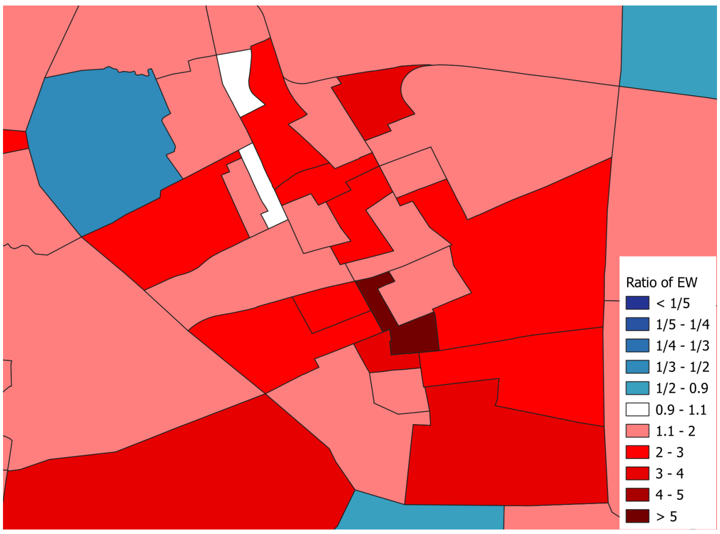


Figure 4.6: Map of the ratio of EW between methods, zoomed in on the Charleston Peninsula area

In Figure 4.7, The histogram displays the distribution of EW ratios across census block groups for Charleston County. The result shows that 150 of 259 census block groups have ratios between 1.1 and 2, indicating that the EW disaggregated method is close to or double the EW aggregated method.

This number of census block groups means the disaggregated method highlights more vulnerable households according to the larger EW value.

However, 36 out of 259 census block groups have ratios between 0.9 and 1.1, indicating a smaller EW disaggregated method than the aggregated one. This number of census block groups suggests that these census block groups have fewer vulnerable households.

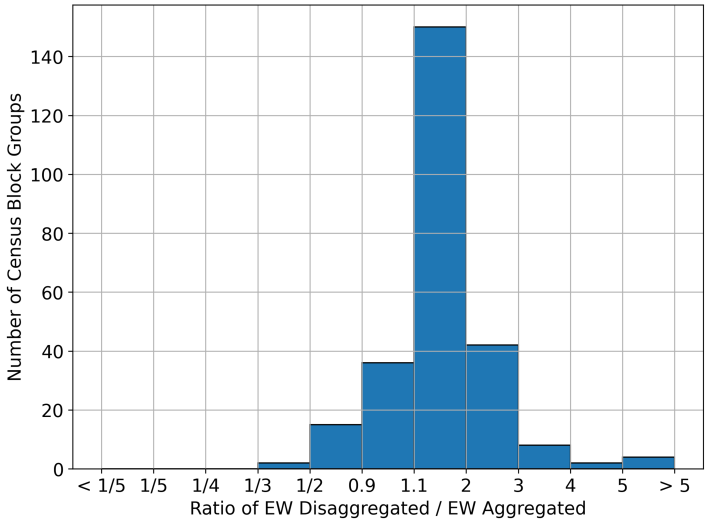


Figure 4.7: Histogram of the ratio of EW between methods for all census block groups in Charleston County

Therefore, to investigate how this might occur, the census block groups with the highest and lowest ratios and the census block groups with ratios close to one will be examined.

Figure 4.8 describes how the variability of household income influences the equity weight (EW) values, as indicated by the EW ratio between methods. If the household income is higher than the average household income for Charleston County, then the EW value will be greater than 1. In contrast, if the household income is lower than the average, the EW value will be less than 1. Therefore, the terms ”low income” (below the average household income) and ”high income” (above the average household income) are used to analyze the graph.

- **Left Graph**: This graph shows that in areas with a wide spread of household income skewed towards low income, the EW values from the disaggregated method tend to be higher than the aggregated method. This graph also shows the high impact of identifying that this census block group is more vulnerable than before (aggregated method) by including income heterogeneity in the analysis (disaggregated method).
- **Middle Graph**: In this graph, the distribution of household income has a longer spread but is still more concentrated towards low income. However, higher income in this census block group makes the median EW values from the disaggregated method lower than the aggregated method. These decreased EW values occur because higher-income households receive lower equity weights, reducing the overall EW values from a disaggregated method.

  By incorporating income heterogeneity, this graph shows that this census block group is less vulnerable than the aggregated method. This graph also highlights the complexity of vulnerability assessment: while the aggregated method indicates a uniform level of need census block group, the disaggregated method suggests that the presence of high-income households can balance or offset the vulnerability caused by low-income households.
- **Right Graph**: This graph demonstrates that when the spread of household income is limited, the ratio of EW values between the aggregated method and the disaggregated method is minimal or close to one. This example indicates that low-income heterogeneity leads to fewer differences.

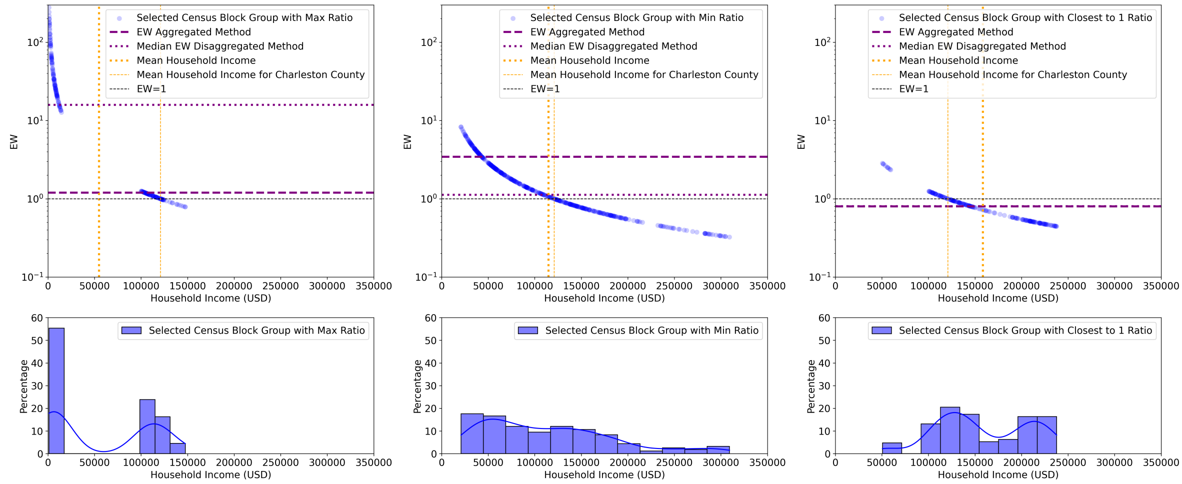


Figure 4.8: From left to right, the census block groups represent: the highest ratios (002009-block3), the lowest ratios (005900-block2), and the ratios closest to one (004615-block3)

Scenarios

The next step is to evaluate equity-weighted damages under different scenarios: a fictitious damage scenario, a historic flood event scenario, and a risk scenario. For the results of the Charleston overall coverage map, see Appendix C.

Starting with the fictitious damage scenario, the relative damage assumption used was 10% of the maximum potential damage to structures and contents for all households in Charleston County.

Figure 4.9 shows the non-equity weighted damages for the fictitious damage in different spatial scales.

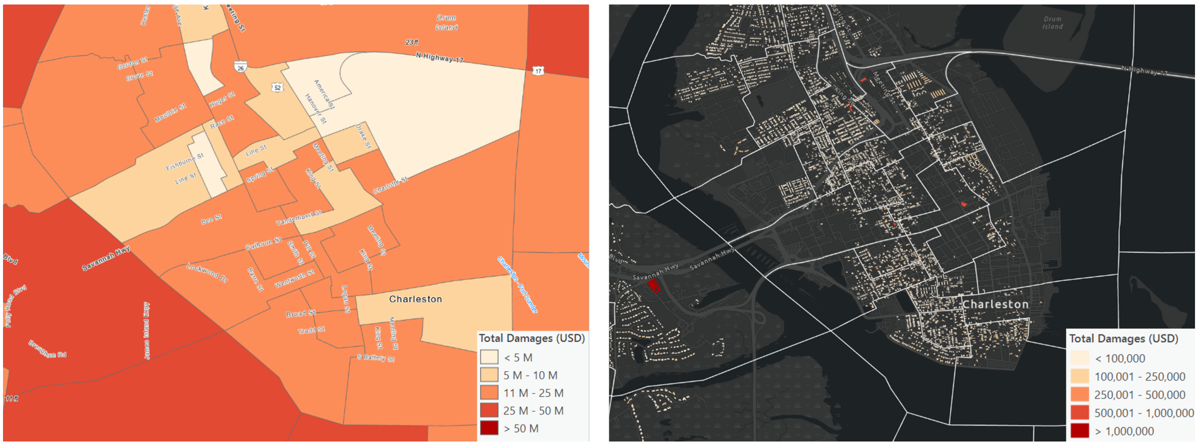


Figure 4.9: Left map damages at the census block group scale, and damages at the building footprint scale for fictitious damage scenario, zoomed in on the Charleston Peninsula area

From Figure 4.10, the left map, with the census block scale method, shows varying levels of EWD, with many areas experiencing damage mostly below 25 million and some areas reaching 50 million. With the disaggregated method, the right map shows an increase in EWD, with most areas now exceeding 25 million USD and many areas exceeding 50 million USD. This shift is consistent with the expectation that using the disaggregated method shows larger damages because EW as a multiplier of damages is generally larger.

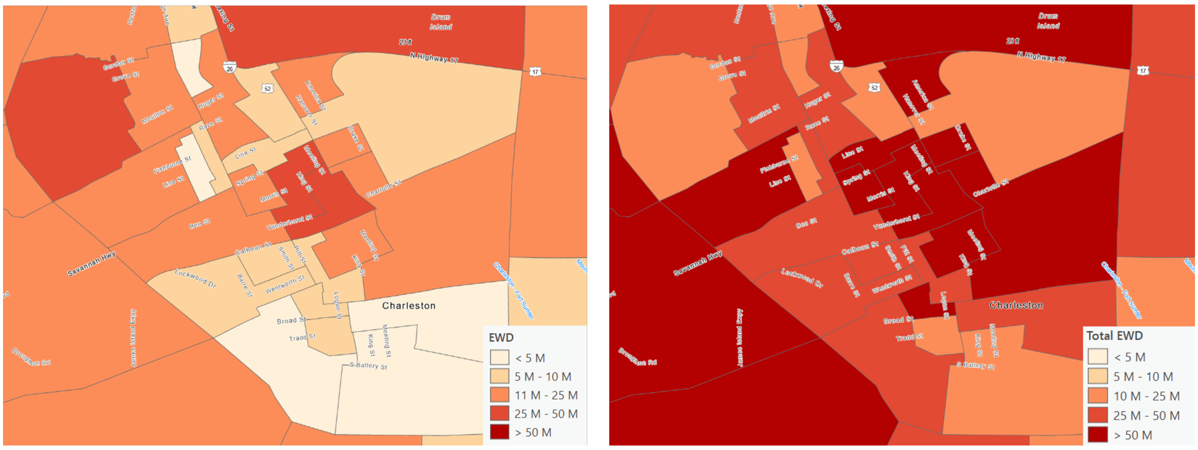


Figure 4.10: EWD map with aggregated method (left) and disaggregated method (right) for a fictitious damage scenario case, zoomed in on the Charleston Peninsula area

The two maps from Figure 4.11 illustrate the changes in EWD in the Charleston Peninsula region, showing absolute differences and ratio values. The EWD ratio follows the same equation 4.2 and the absolute differences are calculated by subtracting the values from the disaggregated method from the aggregated method.

The left map, showing total differences in USD, reveals widespread increases in EWD, especially in the central and southern regions, with some areas experiencing increases of over 25 million USD. Reductions in EWD are observed in one census block group, decreasing by 10 million - 25 million USD.

The right map, which describes the ratio of EWD, shows an overall increase of EWD indicated by ratios greater than 1, with most areas having ratios between 2 and 4, meaning the EWD from the disaggregated method is 2 to 4 times higher. Some areas have ratios between 4 and 5, indicating even higher increases. One census block group showed a ratio between half and one-third, meaning that the census block group has a lower EWD than the aggregate method. The lower EWD value in this census block group is due to the lower EW value of the disaggregated method.

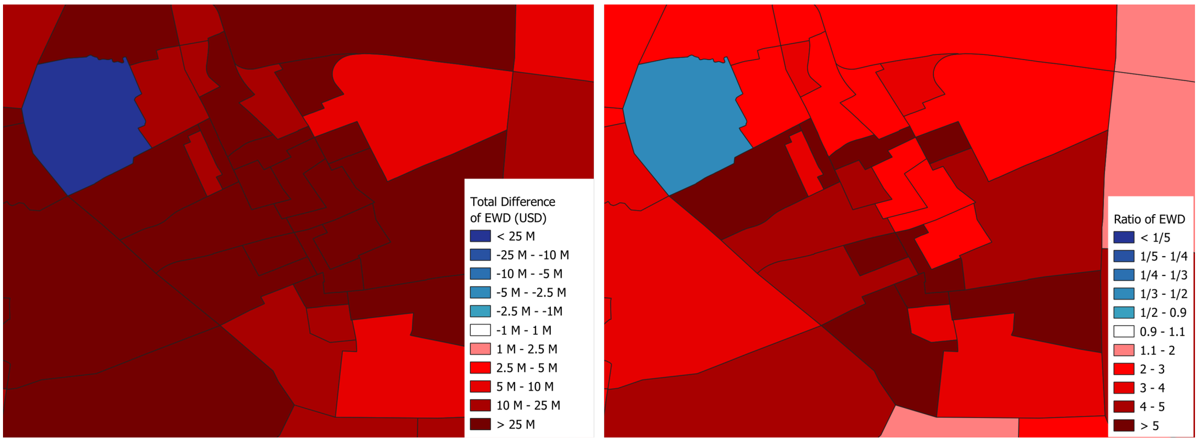


Figure 4.11: Map of EWD total differences between methods (left) and ratio EWD between methods (right) for fictitious damage scenario, zoomed in on the Charleston Peninsula area

The histogram in Figure 4.12 shows the distribution of the ratio of EWD between the disaggregated method and the aggregated method across census block groups. A total of 94 out of 259 census block groups have ratios between 1.1 and 2, indicating that the EWD from the disaggregated method is close to or up to double that of the aggregated method. This distribution suggests that, in most cases, the disaggregated method identifies a higher EWD than the aggregated method.

There are 36 out of 259 census block groups, or 14%, in Charleston County that have ratios less than 1.1, meaning the EWD from the disaggregated method is lower than the EWD from the aggregated method.

Overall, the pattern observed in this scenario is linear with the ratio of EW values shown in Figure 4.6 and Figure 4.7. This linear pattern means that if the ratio of EW in a census block group is greater than one, then the EWD value will also be higher than the aggregated method and vice versa. Moreover, this linear pattern is expected because the scenario equally applies relative damage to all households, with 10%

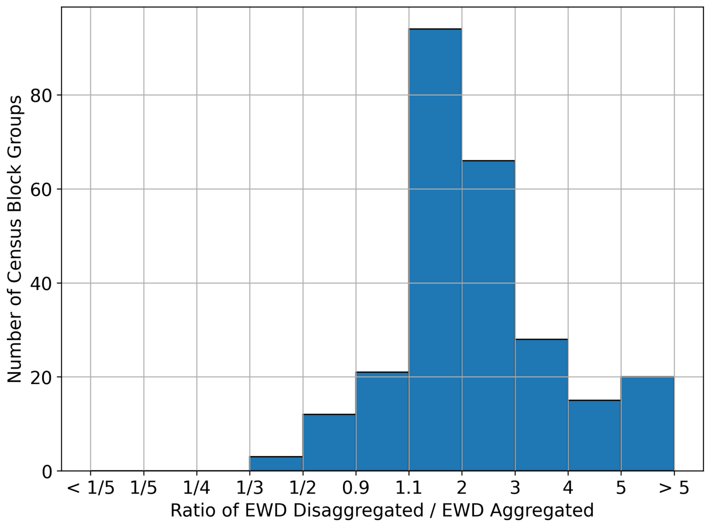


Figure 4.12: Histogram of EWD ratios between methods for all census block groups in Charleston County

For the next 2 cases, damages were modeled using Delft-FIAT, with the output shown per building point based on building points from the NSI. The damage results for the disaggregated method were disaggregated to households with an equal distribution for each household, and for the aggregated method, the income was aggregated by adding to get the total damage in each census block group.

The following is a comparative analysis of damage from historical flood events, focusing on the flood that impacted the east coast on December 17, 2023. This event occurred during the fourth highest tide on record, resulting in major coastal flooding along the southeast coast of South Carolina on December 17 (US Department of Commerce, NOAA, 2023).

In Figure 4.13, the results of the damage modeling are shown, with the flood hazard input displayed on the first map on the left. The results are then aggregated into a census block group scale(second map in the center) and a building footprint scale map (right map). The white color on the aggregated maps indicates no damage to residential buildings in the area. The building footprint scale demonstrates how the results are spatially distributed.

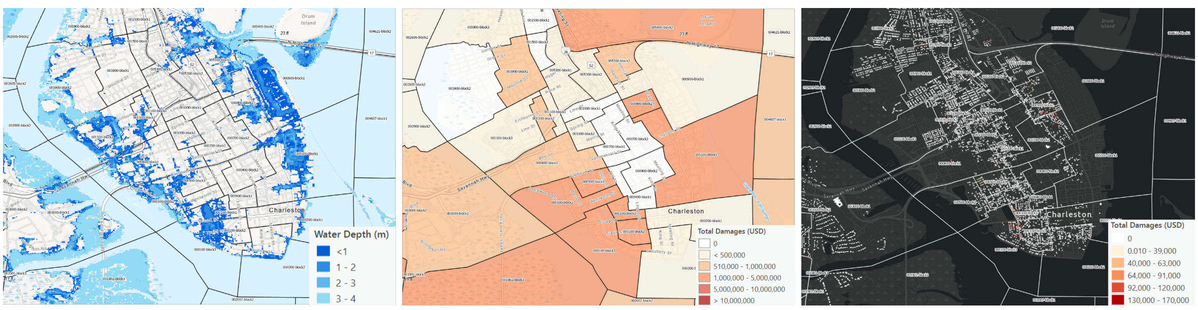


Figure 4.13: From left to right: Flood map, damages at the aggregated method, and damages at the building footprint scale for historic flood event, zoomed in on the Charleston Peninsula area

The EWD for disaggregated method will be used the equations 3.5 and 3.7, with the input damage from damage modelling result.

From Figure 4.14, the left map, with the aggregated method, almost all census block groups that have a range of 5 million USD - 25 million USD change if using the disaggregated method as shown in the right map. The white-colored census block groups remain the same in both maps because there is no damage because of no flood inundation in these census block groups.

In addition, some census block groups experience decreased EWD between the two methods due to lower EW values from the disaggregated method compared to the aggregated method.

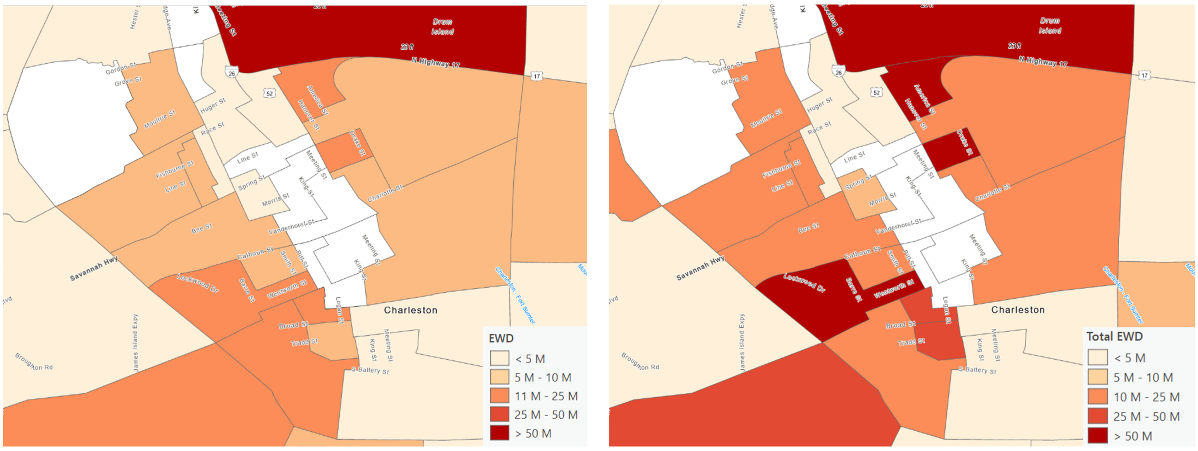


Figure 4.14: EWD map with aggregated method (left) and disaggregated method (right) for a historic flood event case, zoomed in on the Charleston Peninsula area

These two maps from Figure 4.15 illustrate the changes in EWD in the Charleston Peninsula region, showing absolute differences and ratios of EWD between methods. The left map shows the total difference in EWD, where most census block groups have increases in EWD, with changes ranging from 250,000 USD to over 2.5 million USD. These increases correspond to significant shifts in equity-weighted damages, highlighting the areas where the disaggregated method identifies substantially more damage than the aggregated method.

The left map also reveals areas with decreased EWD values, where reductions range from 50,000 USD to over 2.5 million USD. These decreases are primarily observed in census block groups that are less vulnerable to flooding, indicating that the disaggregated method may identify lower damages in areas with high-income households.

The right map displays the ratio of EWD between the disaggregated and aggregated methods. A diverse pattern is observed: many census block groups have ratios greater than 1, indicating higher EWD values with the disaggregated method. In some areas, extreme ratios are shown, with EWD values more than five times greater than those from the aggregated method. This extreme ratios can happen if the flooded areas mainly affect low-income households with high EW values. In other census block groups, ratios less than one are observed, particularly in less vulnerable areas, where the disaggregated method produces lower EWD values.

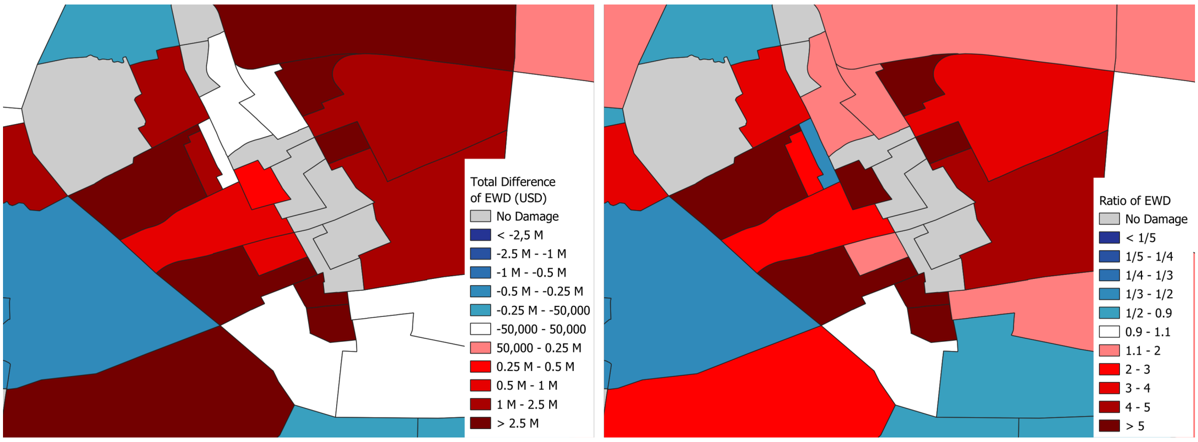


Figure 4.15: Map of EWD total differences between methods (left) and Ratio of EWD between methods (right) for a historic flood event case, zoomed in on the Charleston Peninsula area

The pattern of ratios greater than one and less than one differs from Figure4.6. This pattern difference happens because of the non-linearity of flood-inundated areas, making the damages depend on the location of the flood-inundated area. This non-linearity will be discussed further after the last case study, where similar damage patterns driven by flood-inundated areas will be explored.

In the full extent of all census block groups in Charleston County, the histogram in Figure 4.16 shows the distribution of the ratio of EWD value between the methods. 115 of the 259 census block groups have not suffered flood damage. Therefore, these census block groups were not included in the calculation of the EWD ratio.

Most census block groups have ratios of 1/2 to 0.9, including 36 out of the 144 census block groups analyzed. However, when considering ratios greater than 1.1, 77 out of 144 census block groups, or more than half, are given higher EWD values using the disaggregated method in this scenario.

When considering the extreme values of the EWD ratio, 13 out of the 144 census block groups have a ratio greater than 5, indicating that the EWD in these areas is more than five times higher when using the disaggregated method. This extreme ratio suggests that the flood extent has affected households with high vulnerability (high EW values), resulting in a much greater impact on the EWD calculated by the disaggregated method. In these cases, the disaggregated method highlights the concentrated impact on more vulnerable households, leading to significantly higher damage estimates than the aggregated method.

The reasons why some census block groups have extremely high or low ratios of EWD will be discussed further after the last case study, where similar damage patterns driven by flood-inundated areas will be explored.

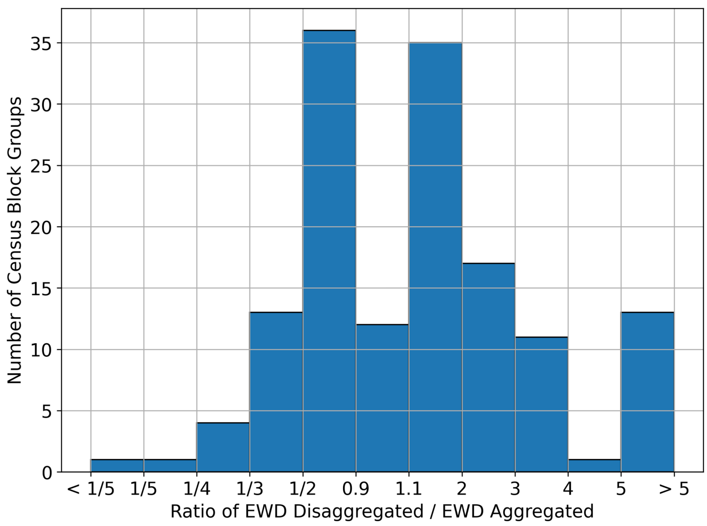


Figure 4.16: Histogram of EWD ratios between methods for all census block groups in Charleston County

Continue to the last case study, which involves a risk scenario. For the risk scenario, several flood hazard maps with different return periods were used, including return periods of 1, 2, 5, 10, 25, 50, and 100 years. The damages were derived into the metric EAD, which is one of the options available in the Delft-FIAT method for damage modeling (Burzel et al., 2017).

Figure 4.17 shows the damage modeling results. The first map on the left displays the flood hazard input, while the center and right maps show aggregated results at the census block group scale and building footprint scale, respectively. White areas on the aggregated maps indicate no residential building damage.

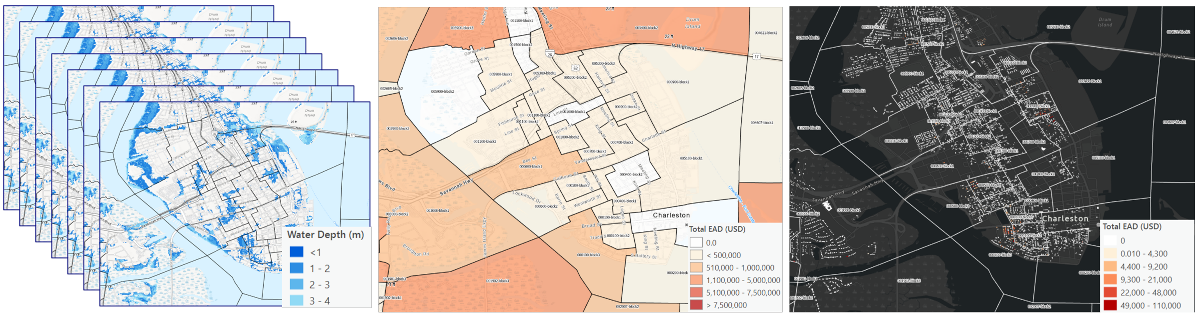


Figure 4.17: From left to right: several flood maps with different return periods, EAD at the census block group scale, and EAD at the building footprint scale for historic risk scenario, zoomed in on the Charleston Peninsula area

From Figure 4.18, the general pattern is similar to the previous case shown in Figure 4.14, where the disaggregated method increases damages for some census block groups, and the white-colored census block groups remain unchanged due to no damages in those areas.

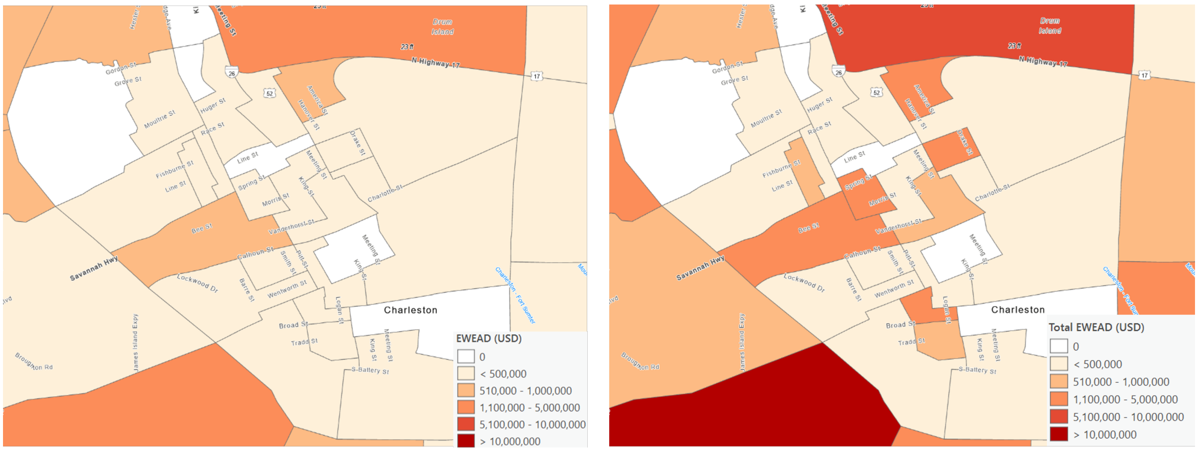


Figure 4.18: EWEAD map with aggregated method (left) and disaggregated method (right) for risk scenario, zoomed in on the Charleston Peninsula area

The absolute differences and ratio of EWEAD between the two methods, as shown in Figure 4.19, there are some census block groups with decreased damages, but the majority still show an increase.

In addition, there is a different pattern where census block groups experience a decrease in damage. In this case, the number of census block groups experiencing decreased is less compared to the previous case, as shown in Figure 4.15. This different pattern occurs because the risk scenario has wider flood-inundated areas than the historic flood event (see Appendix C).

The EWEAD pattern increases and decreases compared to the EW pattern in Figure 4.6 in this scenario is also different. This different pattern means there are census block groups with a ratio of EW greater than one but a ratio of EWEAD less than one, or vice versa.

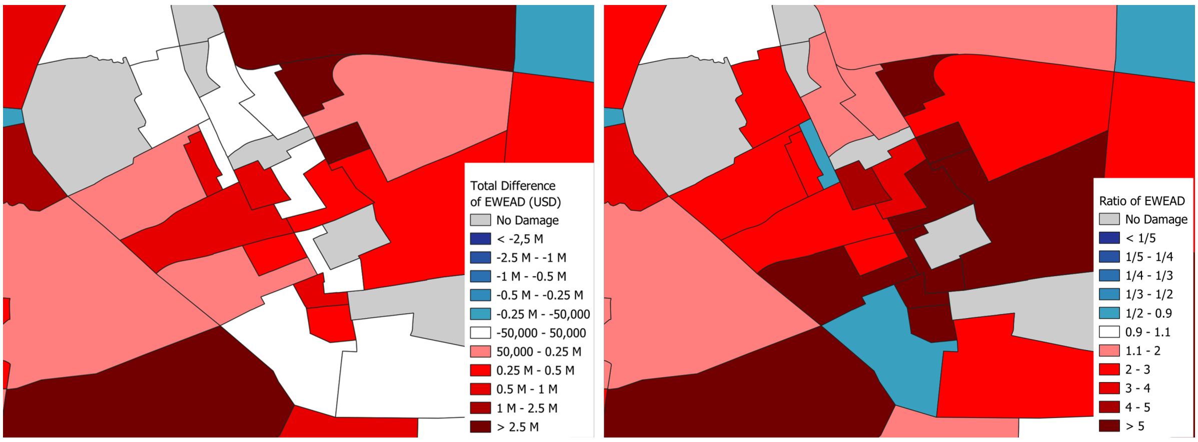


Figure 4.19: Map of EWEAD total differences between methods (left) and Ratio of EWEAD between methods (right) for a risk scenario, zoomed in on the Charleston Peninsula area

The histogram in Figure C.12 shows the distribution of the ratio of EWEAD between the disaggregated and aggregated methods across all census block groups in Charleston County. 32 of the 259 census block groups have not suffered flood damage, which is also not included in the histogram.

For the ratio of EWEAD greater than one, 151 out of 227 census block groups, or 66% of the census block groups in Charleston County, have higher EWEAD values using the disaggregated method compared to the aggregated method. The highest frequency is observed in the range of 1.1 to 2, which includes 65 census block groups.

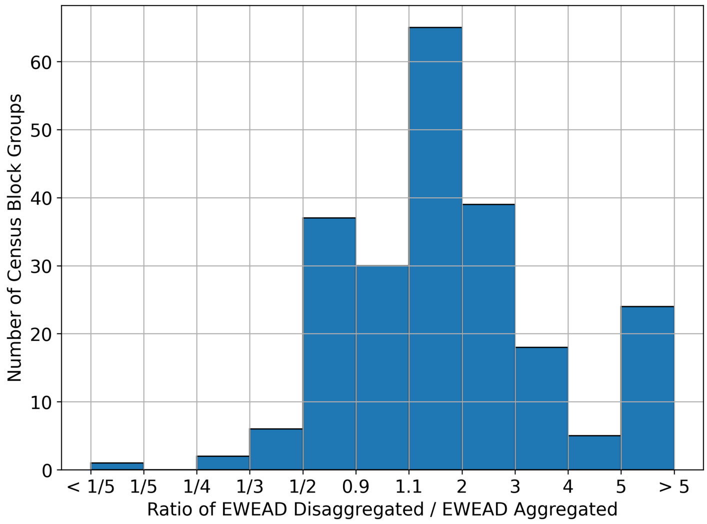


Figure 4.20: Histogram of EWEAD ratios between methods for all census block groups in Charleston County

One census block group will be analyzed to investigate the non-linearity of flood extent that influences the ratios of EWD and EWEAD. Figure 4.21 illustrates an example of a census block group where the ratio of EWD is 0.3 (less than one), the ratio of EWEAD is 1.5 (greater than one), and the ratio of EW is 1.5 (greater than one).

The ratio equal to one is used as the threshold for identifying different patterns. It would be expected that if the ratio of EW is greater than one, all scenario ratios would also be greater than one. However, in this case, the influence of flood extent causes the ratio patterns to differ.

The left map shows that the flood extent occurred in residential buildings during the historic flood event, most likely affecting less vulnerable households (EW less than one). On the other hand, the right map shows that the flood risk scenario involves a wider flood extent, which also impacts more vulnerable households (EW greater than one). These two maps explain why the patterns in both scenarios differ.

This example further illustrates how the flood extent can lead to extreme variations in the EWD ratio. When a flood affects areas with differing levels of household vulnerability, the disaggregated method, which applies individual equity weights, can reveal significant differences in damage assessments. In particular, extreme EWD ratios occur when the flood impacts households with higher vulnerability, which are given greater weight in the disaggregated method. This analysis highlights the important need to properly validate the spatial allocation of income, as this directly impacts the determination of equity weights and equity-weighted damages.

Moreover, this example highlights the methodological difference where non-equity-weighted damages in the aggregated method are assessed similarly to those in the disaggregated method. When a single EW is used across the entire area in the aggregated method, the final equity-weighted results can differ significantly from those in the disaggregated method, where each household is assigned an individual EW. This difference in applying equity weights can lead the aggregated method to either underestimate or overestimate the impacts on specific households, particularly in areas with varying levels of vulnerability. As a result, while the actual damages remain consistent across both methods, the disaggregated method reflects equity-adjusted impacts more, providing a clearer picture of how households are affected based on their vulnerabilities.

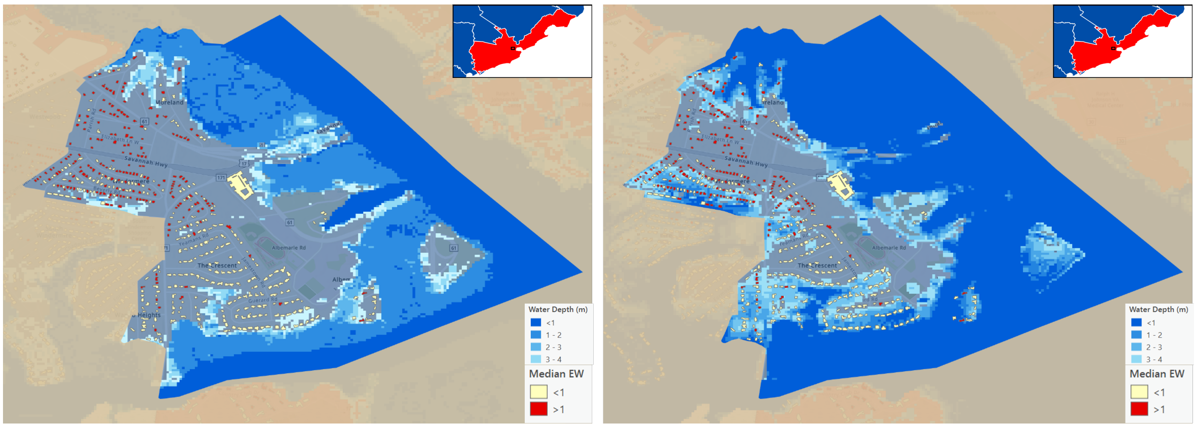


Figure 4.21: Map of a census block group (003000-block1) showing different patterns in EWD and EWEAD ratios between the historic flood event scenario (left) and the risk scenario (right)

Figure 4.22 presents the general results of the case study in Charleston County through different scenarios and methods.

The analysis of three scenarios, fictitious damage scenario, historical flood event, and risk scenario, demonstrates that the disaggregated method tends to yield higher damage estimates in both EWD and EWEAD than the aggregated method. For example, the disaggregated method captures 15% more vulnerable households and 25% more vulnerable census block groups in Charleston County overall. This increased percentage leads to an increase in total damages of two times higher for the fictitious damage scenario, three times higher for the historic flood event scenario, and the risk scenario.

Although the results might seem expected given the detailed approach of the disaggregated method, the main point is the magnitude of the observed difference. The significant increases in equity-weighted damages could lead stakeholders, such as government agencies and insurance companies, to consider adopting this disaggregated method.

For example, while the rise in damages might result in higher absolute financial losses for wealthier households due to the greater value of their assets, the relative impact is often more severe for low-income households. It occurs because low-income households have fewer resources to absorb such losses, which the equity weighting approach recognizes as critical in assessing overall impact.

These results also highlight the importance of using a disaggregated method that accounts for income heterogeneity in equity-weighted flood risk assessment, as this approach provides a different perspective on equity weighting.

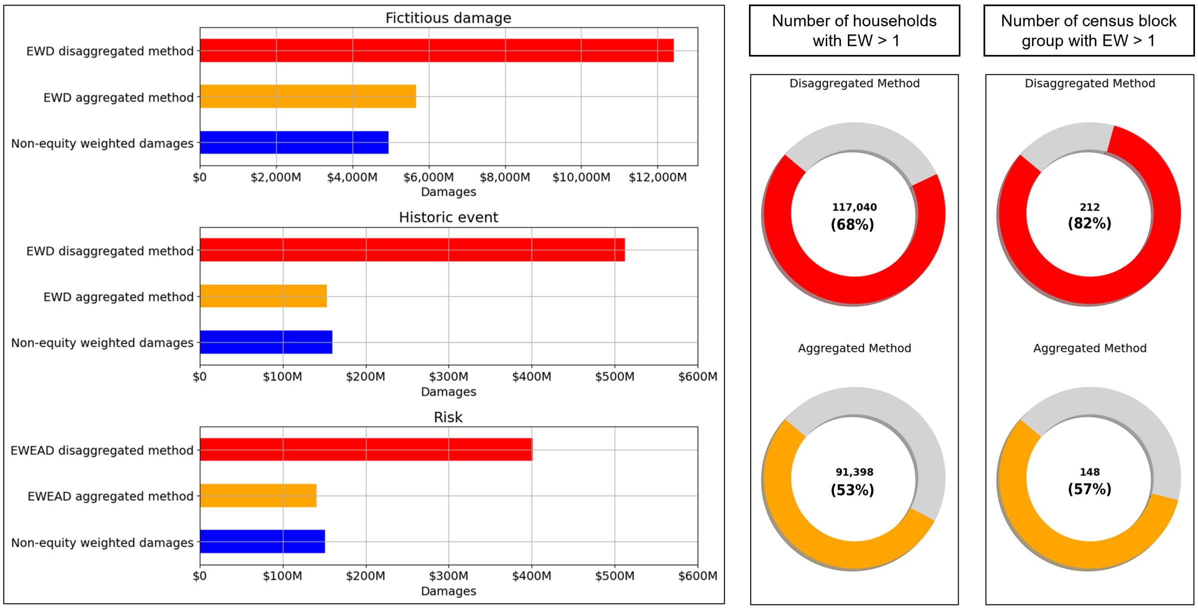


Figure 4.22: General results of case study in Charleston County

4.2. Sensitivity and uncertainty analysis

Sensitivity and uncertainty analyses will be conducted to evaluate the impact of incorporating income heterogeneity in the calculation of equity weights. This approach is necessary due to the uncertainties in the census dataset and other data sources that influence the equity weights.

A parameterized household income distribution will be used to account for the variability and potential inaccuracies within the data.

Uncertainty over the income distribution

Figure 3.3 shows the uncertainty in the household income dataset, which is the main factor for determining the EW values. The evaluation of uncertainty in income distribution aims to examine the variability in income distribution, which can differ across different census block groups, depending on the average income level and the spread of income distribution. Therefore, several representative census block groups with different average incomes and income distributions were examined.

The representative census block groups will be selected based on clustering analysis. K-means clustering was used for this purpose, and the optimal number of clusters (k) was determined using the elbow method and silhouette analysis.

The inputs for clustering and determining the optimal number of clusters are the parameter sets μ and σ per census block group. The parameter sets are obtained from the optimized individual household income, which is also used in the 4 section and directly from census income statistics. The parameters are then applied using the log transform method (see Appendix A.2.1).

Based on elbow and silhouette methods, the optimal k of 5 clusters was chosen (see Appendix A.3). These 5 clusters will then be used for conducting k-means clustering.

Figure 4.23 shows that the clustering result divides census block groups into five clusters based on μ and σ.

1. **Cluster 0 (red)** has low μ values and low σ values, indicating low average incomes with low-income variability.
2. **Cluster 1 (green)** has low μ values and high σ values, indicating low average incomes with moderate income variability.
3. **Cluster 2 (blue)** has moderate values for μ and σ, representing census block groups with balanced average incomes and variability.

4. **Cluster 3 (purple)** has high μ values and low σ values, indicating high average incomes with low-income variability.
5. **Cluster 4 (orange)** has high μ and high σ values, indicating high average incomes and high-income variability.

The results of this clustering can be assigned to several household income categories, such as low-income households for clusters 0 and 1, middle-income households for cluster 2, and high-income households for clusters 3 and 4.

The black dots are the centroids of each cluster, labeled with the nearest census block group identifiers, such as 000700-block1, 004300-block3, 000990-block1, 002613-block4, and 004622-block2. These census block groups will represent each cluster.

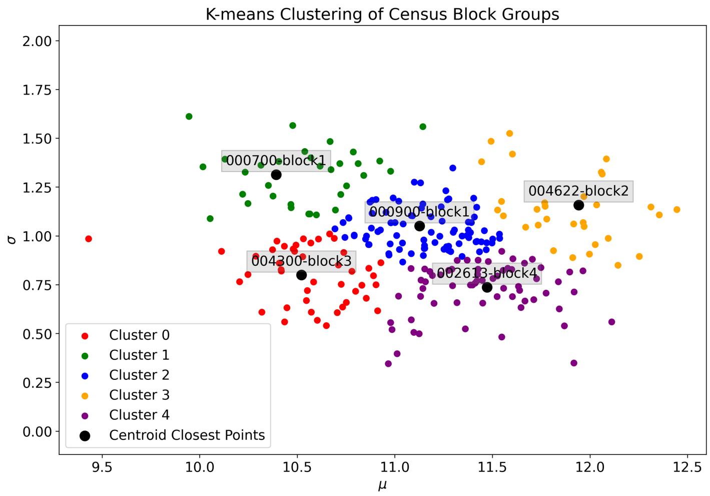


Figure 4.23: K-means clustering of census block groups

Figure 4.24 shows the spatial distribution of the clustering results across Charleston County. Each colour represents a different cluster-ID, corresponding to the household income distribution of the census block groups. The representative census block groups for each cluster are highlighted in black. Areas not included in the analysis are shown with hatched lines.

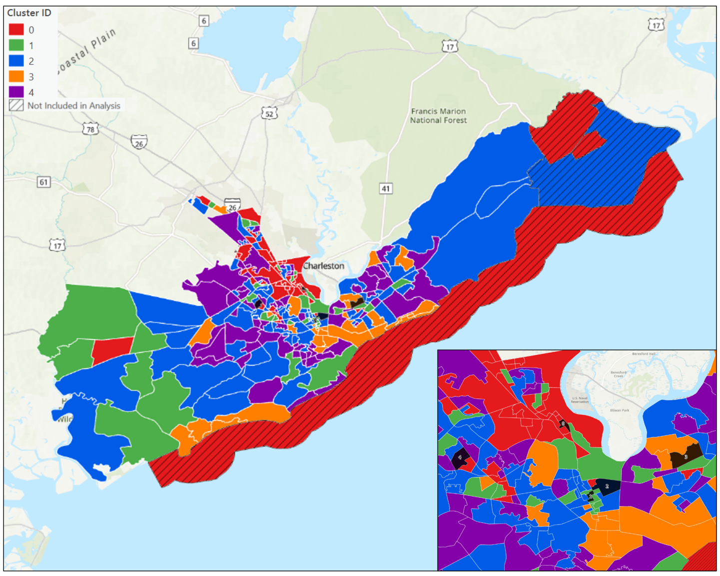


Figure 4.24: Spatial distribution of census block groups based on clustering analysis, with selected census block groups representing each cluster shown in black color

Figure 4.25 compares EW distribution per census block group using the actual distribution used in case study analysis with the log-normal distribution. The EW values from log-normal distribution were derived by median household income from 10,000 Monte Carlo simulations (see Appendix A for further details of log-normal household income distribution).

Bootstrapping will then be applied to determine the range of distribution parameters μ and σ for the selected census block groups (see Appendix A.3). This method assesses the uncertainty in the fitted distribution parameters, which will be utilized for equity weights sensitivity analysis.

The range of distribution parameters from bootstrapping will be used for the sensitivity analysis using the confidence interval 95% for the μ and σ.

The Table 4.2 shows the distribution parameters in both log-transformed and actual values. Census block groups 004300-block3 and 002613-block4 shows a narrow confidence interval (95%) for the μ and σ parameters, indicating low uncertainty. In contrast, census block groups 000700-block1, 000900-block1, and 004622-block2 show moderate uncertainty, with a slightly wider range for the μ and σ parameters, particularly in σ.

Table 4.2: Range of household income distribution parameters from bootstrapping results

No.	Census Block Group	Distribution Parameters		Distribution Parameters (Actual value in USD)	
		μ	σ	Mean	Standard Deviation
1	004300-block3	10.47 - 10.65	0.74 - 0.86	46,342 - 61,072	39,570 - 63,909
2	000700-block1	10.43 - 10.63	1.24 - 1.38	73,043 - 107,173	139,613 - 256,217
3	000900-block1	11.02 - 11.28	0.95 - 1.13	95,918 - 150,009	116,127 - 241,207
4	004622-block2	11.82 - 11.96	1.09 - 1.19	246,237 - 317,442	371,878 - 560,805
5	002613-block4	11.4 - 11.51	0.68 - 0.76	112,555 - 133,093	86,300 - 117,676

The mean of each distribution parameter from bootstrapping results was used to calculate the EW, which was then compared with the EW calculated using the actual distribution.

Figure 4.25 compares the EW values derived from the actual distribution vs. the log-normal distribution. This comparison shows relatively similar results to those using the actual distribution. This similarity is supported by the performance metrics, with an R^2 value of 0.911, a Kolmogorov-Smirnov (KS) test statistic of 0.2, and a p-value of 1.0.

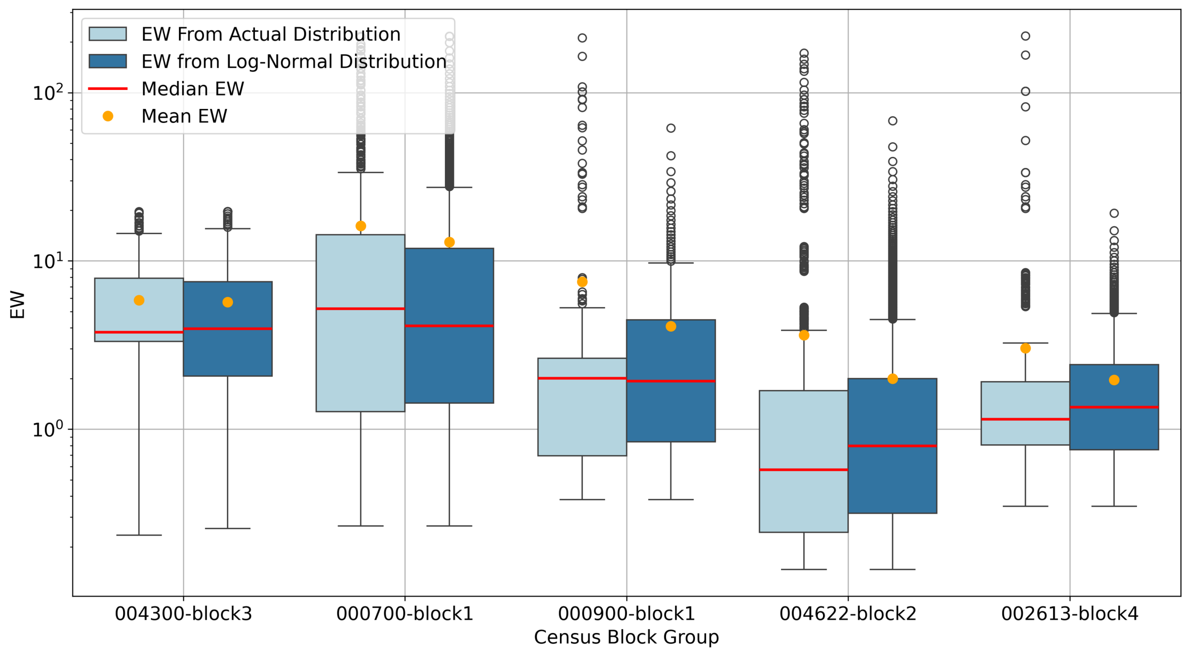


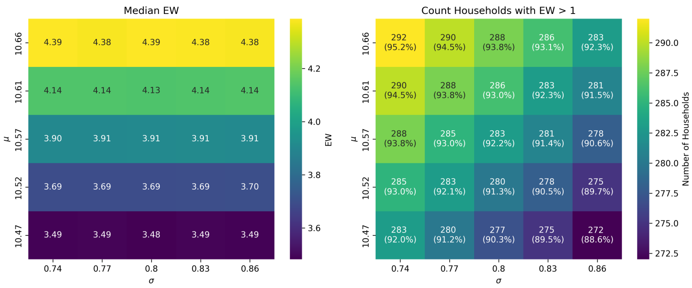
Figure 4.25: Equity weights distribution by representative census block groups: using actual vs. log-normal distribution

Figure 4.26 presents the heat maps of the sensitivity analysis results. The sensitivity analysis was conducted in 5 intervals for each parameter, μ and σ. Each cell value in the heat maps represents the outcome of Monte Carlo simulations with a total of 10,000 simulations, which were then used to derive the median EW and count households with EW greater than one.

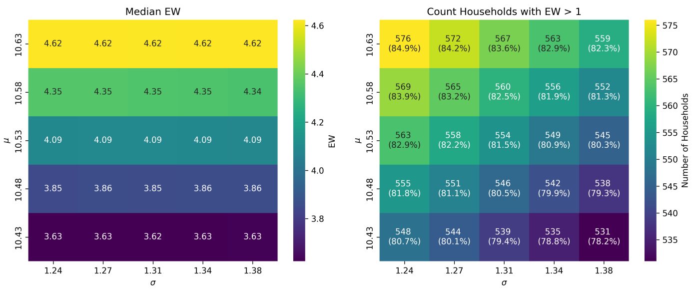
The results of the heat maps show that none of the distribution parameters (μ and σ) have a high sensitivity to both metrics. The highest difference in median EW from all census block groups is 1 with μ variate and almost zero differences if the σ variate. For the percentage of households counted with more than one EW, the highest difference is 12$\%$ with both parameters variate.

All heatmaps show the same pattern: a lower value of σ results in a higher count of households with EW greater than one, except for one census block group (004622-block2) in Figure 4.26d. This census block group shows the opposite pattern, where a lower value of σ results in a lower metric value. This opposite pattern occurs because this census block group has the highest range of actual values for σ and μ, which are also greater than the average household income for Charleston County (120,838 USD), as shown in Table 4.2.

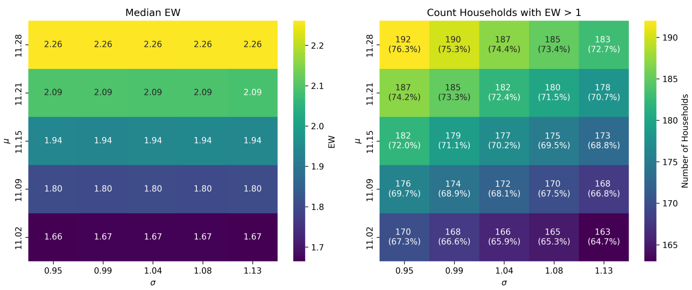
However, the μ parameter generally shows slightly higher sensitivity than the σ parameter. This greater difference means that changes in the mean (μ) have a slightly greater impact on the median EW and the count of households with EW greater than one than changes in the standard deviation (σ). Despite this, the overall sensitivity remains low, suggesting that the variability in these income distribution parameters does not significantly affect the EW.



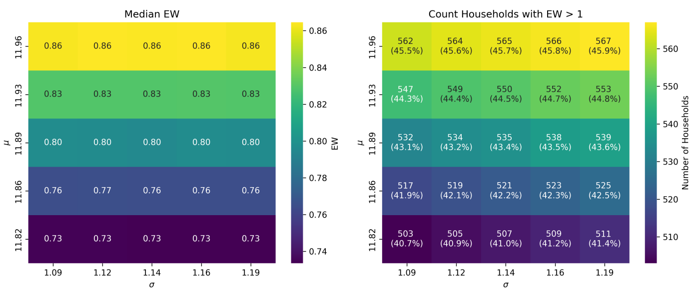
(a)



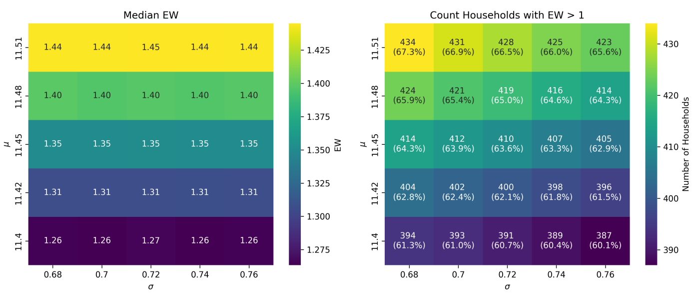
(b)



(c)



(d)



(e)

Figure 4.26: Sensitivity results for EW with varying parameters μ and σ for the following Census Block Groups: (a) '004300-block3', (b) '000700-block1', (c) '000900-block1', (d) '004622-block2', and (e) '002613-block4'. The left figures display heat maps of median EW values, while the right figures show heat maps of the number of households with EW greater than 1

The standard deviation and coefficient of variation were calculated to quantify the change of median EW across different parameters μ and σ. Figure 4.27 shows the results of these calculations. The standard deviation of the median EW for all representative census block groups shows a low value (less than 0.5), and the coefficient of variation shows a low value with an average value of 8% for all representative census block groups. These two metrics also highlight that the parameters are low sensitivity.

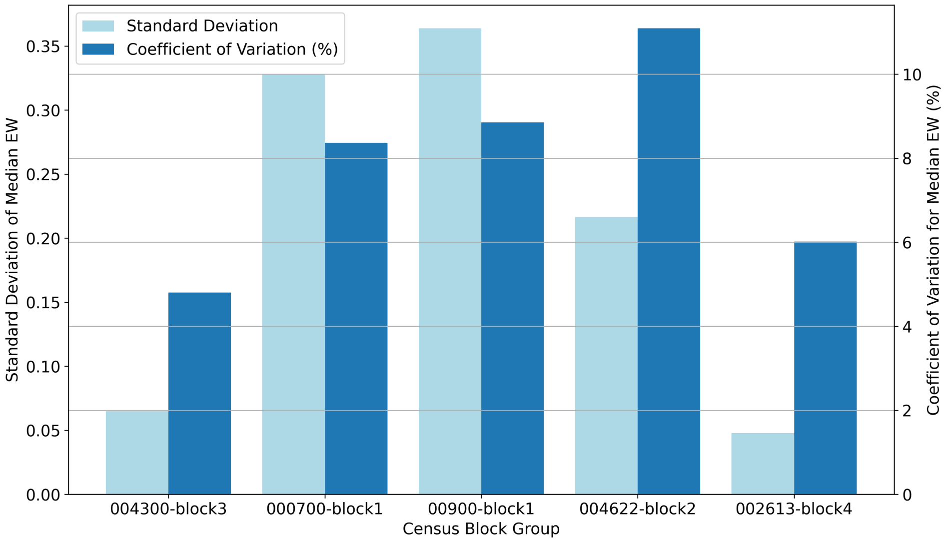


Figure 4.27: Comparison of standard deviation and coefficient of variation for median EW across census block groups

Therefore, it can be concluded that using a log-normal distribution to derive individual household income is appropriate for calculating equity weights, as the sensitivity analysis shows that the metrics are not sensitive to variations in μ and σ. Despite the uncertainties inherent in census datasets, such as potential inaccuracies and data limitations, the log-normal distribution offers a practical estimation method. When detailed distribution data is unavailable, relying on general statistics such as the mean and standard deviation to estimate the income distribution can be effective. However, this uncertainty should be recognized in the interpretation of the results.

Uncertainty spatial allocation of income

The NSI structure value is the depreciated replacement cost calculated on a dollar per square foot basis. This dollar per square foot is multiplied by the estimated square meter area for each structure(National Structure Inventory, 2022). However, in some cases the structure value has uncertainty as shown in Figure 4.28.

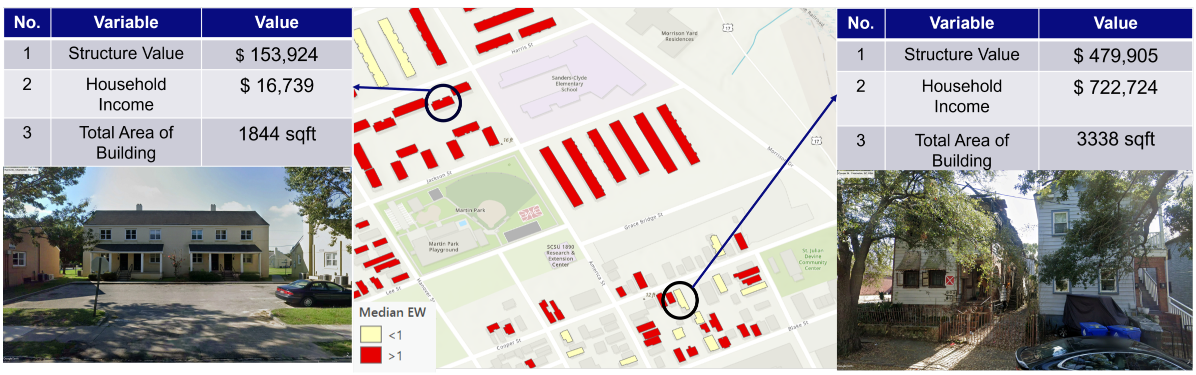


Figure 4.28: Example of uncertainty in spatial allocation of income

Using a total of 10,000 ensembles in the Monte Carlo simulation, the uncertainty of the structure value determined by the NSI will be assessed.

Figure 4.29 demonstrates the implementation of equation 3.8, showing the distribution of single building structure values. The red dashed line indicates the deterministic or NSI-derived values for comparison.

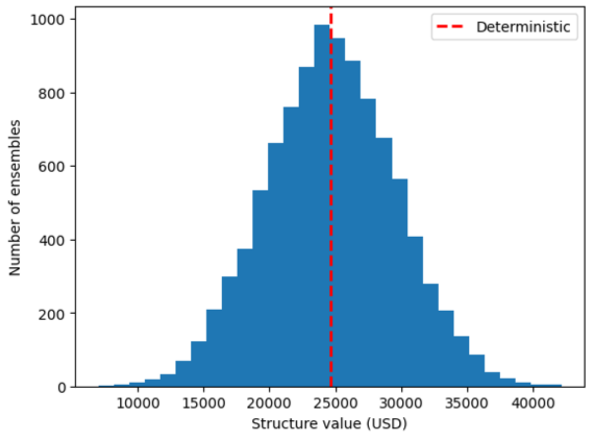


Figure 4.29: Example of structure value distribution of single building

This noise was applied to all building points for the selected census block groups and reassigned household incomes in order from lowest to highest for both variables. After that several metrics were calculated as follows:

The box plot in Figure 4.30 shows the standard deviation of EW for each census block. Census block group 00430-block3 has a minimal variation with a low standard deviation, indicating less impact of structure values to EW value in this census block group. 00700-block1 shows a wider spread and a higher median, indicating more variability. 00090-block1, 00462-block2, and 002613-block4 have varying degrees of spread, with 00462-block2 having many outliers, indicating wide variability in EW within that block.

Reflecting the box plot with Figure 4.23, the 00090-block1, 00462-block2, and 00700-block1 census block groups represent clusters with high standard deviations of household income, resulting in a high standard deviation of EW. Meanwhile, the 00430-block3 and 002613-block4 census block groups represent clusters with lower standard deviations of household income, resulting in a low standard deviation of EW. The highest EW standard deviation is observed in block 00700-block1, which represents the cluster with the lowest average income (μ) and the highest income standard deviation (σ).

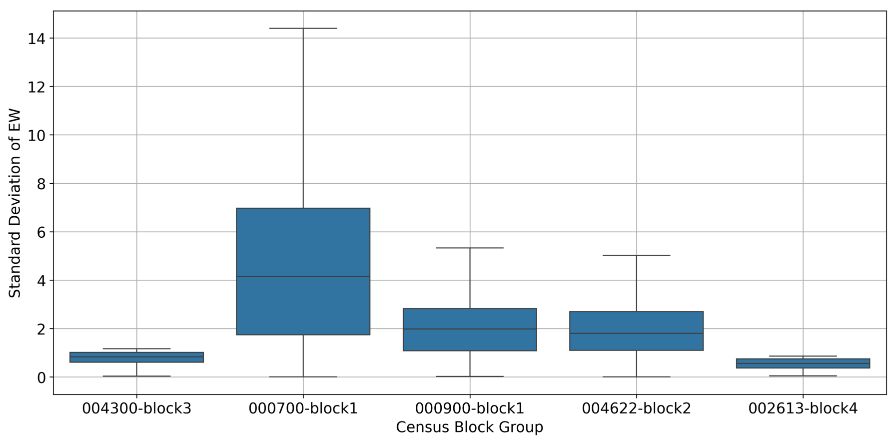


Figure 4.30: Standard deviation of EW across all ensembles for each selected census block group

The second box plot in Figure 4.31, which illustrates the coefficient of variation for EW, highlights different levels of consistency across census block groups. In 3 out of 5 census block groups, the coefficient of variation is higher than one, indicating that the structural value has high uncertainty relative to EW.

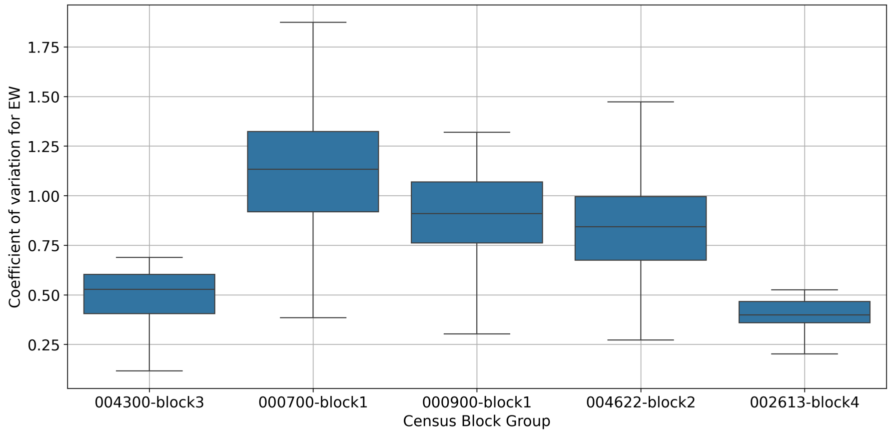


Figure 4.31: Coefficient of variation EW across all ensembles for each selected census block group

The third box plot in Figure 4.32 shows the ratio of EW values greater than one across different census block groups, which indicates that the census block group is more vulnerable. All census block groups show a median ratio above 0.7, indicating that most EW values exceed 1. The wide interquartile range and the presence of a lower whisker extending to 0 indicate that while many values exceed 1, there are still some cases where EW falls below this threshold, highlighting the varying vulnerability across census block groups.

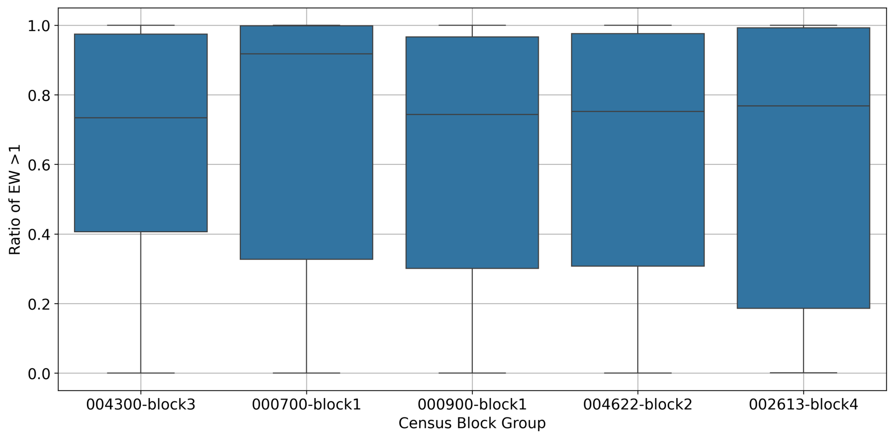


Figure 4.32: Ratio of EW greater than 1 across all ensembles for each selected census block group

In conclusion, the three box plots show that structure value uncertainty significantly affects the spatial allocation of income, indirectly impacting EW with varying degrees of impact across census block groups.

The sensitivity and uncertainty analysis regarding the calculation of equity weighting can be summarized as follows:

1. The comparison between the EW calculated from the log-normal distribution of income and the EW calculated from the actual income distribution shows a high degree of similarity. This result suggests that the log-normal distribution is a good fit for modeling the income data.
2. The analysis using a parameterized log-normal distribution for income, which derives the μ and σ parameters, showed moderate to low uncertainty with no significant sensitivity in equity weights

calculation. This result suggests that the variability in income distribution, as modelled by the log-normal distribution, does not significantly impact the equity weights.

3. The spatial allocation of income using structure values derived from the NSI shows high uncertainty and high sensitivity in equity weight calculations. This finding highlights the need for alternative methods for spatially distributing income, as the current method may introduce significant variability in the equity weight outcomes.

In flood risk assessment with the equity weighting method at the household level, the spatial allocation of income tends to be more important. Accurate spatial allocation ensures that the geographic distribution of equity weights better represents vulnerable households, which is vital for planning and implementing flood risk mitigation strategies. However, this does not diminish the importance of accurately deriving household income, which is essential for understanding vulnerability and ensuring equitable outcomes. Both aspects are important, but spatial allocation has a more direct impact in the context of flood risk.

4.3. Limitations of the Study

This study has several limitations, which can be categorized into three categories: general limitations, data limitations, and method limitations.

4.3.1. General limitations

This study has a general limitation in that it is limited to analysis in Charleston County, so the results may not be generalized to other areas with different social and economic conditions. However, if the characteristics of other areas, such as high-income inequality, are similar to this location, similar results may also be obtained. The equity weighting method at the household scale used in this study will likely be effective in areas with high-income inequality.

4.3.2. Data limitations

The distribution of household income from the census dataset only provides predefined income brackets and open-ended brackets with the number of households. This limitation was addressed by assuming that income is equally distributed within each predefined income bracket and by performing optimization to determine the upper and lower bounds of the open categories.

The income census dataset used in this analysis has a high margin of error, which may affect the accuracy of the results. The census survey technical documentation indicates that some aspects of the income data may need to be more accurate due to reliance on the respondent's memory (Bureau, 2023). Despite potential issues with data completeness or data collection accuracy, the log-normal distribution is introduced because it has a good fit with the income data available in this case.

4.3.3. Method limitations

Generating household income distributions directly from census data involves applying equally distributed income within predefined income brackets and optimizing upper and lower bounds for open-ended income brackets. However, this method may not accurately represent the actual income of individual households, although the resulting income statistics in aggregate are close. This discrepancy is due to the possibility of significant variation within each income bracket.

The spatial allocation of income based on structure replacement costs assumes a correlation between property value and household income, which may not always be accurate. This assumption ignores other influencing factors, such as ownership status and differentiating between rented and owned households. There are 63% owner households and 37% renter households in Charleston County, which must be considered to differentiate income allocation. While owner households tend to have higher property values and are often associated with higher incomes, this is not always consistently true. For example, retirees with high-value property may have lower annual incomes. In contrast, renter households may live in properties with lower values but can have significant incomes, especially in urban areas with high rental costs.

5

Conclusion and Recommendations

5.1. Conclusion

In this study, the equity weighting method was applied and tested through a case study analysis and an exploration of sensitivity and uncertainty. This research addressed the gaps in current flood risk assessment methodologies by incorporating household-level income heterogeneity into equity-weighted risk and impact assessments. The main findings and implications are summarized below:

1. How can the heterogeneity of income be accounted for in practice given limitations on income data due to privacy rules?

Income heterogeneity can be accounted for using statistical methods such as Monte Carlo simulations, bootstrapping, and optimization techniques. These methods allow for the estimation of household income distribution despite the limitations imposed by privacy rules.

a. What assumptions can be made?

Two main assumptions can be used to account for income heterogeneity. First, it is assumed that the household income is equally distributed within the predefined income brackets, and the upper and lower bounds of the open income brackets are obtained from the optimization of income statistics in the census block group. This method effectively reflects the income distribution from the census, evidenced by an R^2 value of 0.99 and a low average RMSE across all available census income statistics. .

Second, it is assumed that there is a linear relationship between the structure value (or the maximum potential damages) and the income at each residential building, which allows for the spatial allocation of household income within the census block group.

b. What are the uncertainties surrounding this?

The uncertainties related to accounting for income heterogeneity include uncertainties in the distribution of household income and uncertainties in the spatial allocation of income.

For the uncertainty over the household income distribution, a log-normal distribution was used on representative census block groups. Monte Carlo simulations explored the uncertainty and sensitivity using the range of distribution parameters obtained from bootstrapping. It was found that the change in median EW has a low range (average change less than 0.5), with an average coefficient of variation (CV) of median EW being 0.08. These values indicate a low level of uncertainty and low sensitivity. This also suggests that the log-normal distribution fits the data well, even if only general statistics are used.

For uncertainty in the spatial allocation of income, uncertainty in the structure value on which income was spatially distributed was explored for representative census block groups. The coefficient of variation for equity weights shows that 3 out of 5 representative census block groups exceed 1, indicating high relative uncertainty and high sensitivity. This high uncertainty can be compounded by the non-linearity of flood extent, where small changes in flood dynamics can lead to disproportionate impacts in different areas. Such non-linearity can lead to the assumption of linearity in the spatial allocation of income to underestimate or overestimate the actual risk and impact.

2. What is the added value of considering the heterogeneity of income within a census block group for equity-weighted risk and impact estimations? Considering income heterogeneity using a disaggregated method provides a different perspective on equity-weighted risk and impact estimation.

In terms of policy implications, the added value is that the disaggregated method can give more information for vulnerable households that have previously been overlooked, even if equity weighting has already been implemented with the aggregated method. This value is supported by the result of the disaggregated method, which can identify 15% more vulnerable households, equivalent to 25% more vulnerable census block groups than the aggregated method.

In terms of practical application, disaggregated methods show added value in evaluating the equity weighting of flood risk management strategies. Metrics such as Equity-Weighted Damage (EWD) and Equity-Weighted Expected Annual Damage (EWEAD) show significant variations in how risks and impacts are distributed among households within the same census block group. EWD shows a two-times higher value in total equity-weighted damage in the fictitious damage scenario and a three-times higher value in a historic flood event scenario compared to the aggregated method. Moreover, EWEAD also showed that the total equity-weighted damage estimate was three times higher than the aggregated method. These insights can be useful for practitioners in determining resource allocation and intervention priorities, especially in areas with high-income inequality, where aggregated methods may not be sufficient to capture different levels of social vulnerability related to income.

In terms of theoretical contributions, this application of the disaggregated method in equity-weighted risk and impact estimations shows the importance of incorporating income heterogeneity when it comes to areas with high-income inequality. With observed significant differences in EWD and EWEAD values, this application captures more vulnerable households that are in line with the purposes of the equity weighting method, prioritizing investments that benefit socially vulnerable and low-income groups.

5.2. Recommendations

Based on the findings of this study, several recommendations are proposed to improve flood risk assessment and policy implementation in the future:

Policymakers and practitioners should consider adopting a disaggregated method of flood risk assessment with an equity weighting method to capture the heterogeneity of income. This approach provides a better understanding of the impact of flood risk on different socio-economic groups, allowing for more targeted and equitable mitigation strategies. To make this approach feasible, it is also important to improve data collection practices, particularly in gathering data on household income and structure value.

Equity weights should be incorporated into the Cost-Benefit Analysis (CBA) of flood risk management. This approach could yield different results compared to CBA analyses previously conducted using aggregated methods, offering an different perspective on flood risk reduction strategies. This could provide an important consideration, particularly in prioritizing vulnerable households within these strategies.

For technical recommendations on methods, consider using 'Tenure by Units in Structure' (Bureau, 2022h) data. This data can be used with methods similar to (Bick et al., 2021) with adjustment to distinguish between renter and owner households. This data can also link NSI and census data to distribute total housing units per NSI building point by occupancy type.

5.3. Future Research Directions

Several future research directions are suggested to build on the findings of this study. First, further research should explore additional socioeconomic factors, such as education, employment, and health status, can be integrated into equity-weighted flood risk assessments. Incorporating these factors, which are included in the Social Vulnerability Index (SoVI), can provide a more comprehensive assessment of social vulnerability. The SoVI integrates various socioeconomic and demographic indicators to quantify and map the susceptibility of populations to adverse impacts from natural hazards, including floods.

For instance, the SoVI has been used to explain variations in flood-related deaths and property damage in the USA, demonstrating that social vulnerability significantly influences these outcomes (Tellman et al., 2020). Additionally, integrating SoVI into flood risk management planning can highlight spatial variations in socioeconomic vulnerabilities, helping to direct resources and mitigation efforts more effectively (Cutter et al., 2013).

Furthermore, developing dynamic models that account for temporal changes in socioeconomic conditions and flood risk exposure could improve the relevance and applicability of risk assessments. For example, such models could be particularly valuable in countries with limited resources, where socioeconomic conditions are rapidly changing, and where flood risk management often faces significant challenges. Applying the methodology to diverse geographic and economic contexts, including developing countries, can help validate its generalizability and practicality. Comparative studies across different regions could offer valuable insights into the effectiveness of equity-weighted risk and impact assessments in various settings, ensuring that the methodologies developed are both applicable and beneficial, especially in areas with high vulnerability.

References

H. Apel, G. T. Aronica, H. Kreibich, and A. H. Thieken. Flood risk analyses—how detailed do we need to be? *Natural Hazards*, 49(1):79–98, April 2009. ISSN 1573-0840. doi: 10.1007/s11069-008-9277-8. URL https://doi.org/10.1007/s11069-008-9277-8.

Adelchi Azzalini, Tomas Dal Cappello, and Samuel Kotz. Log-skew-normal and log-skew-t distributions as models for family income data. *Journal of Income Distribution®*, 11(3-4), Dec. 2002. doi: 10.25071/1874-6322.1249. URL https://jid.journals.yorku.ca/index.php/jid/article/view/1249.

Jitka Bartosˇova. Logarithmic-Normal Model of Income Distribution in the Czech Republic. *Austrian Journal of Statistics*, 35(2), 2006.

Erich Battistin, Richard Blundell, and Arthur Lewbel. Why is consumption more log normal than income? gibrat's law revisited. *Journal of Political Economy*, 117(6):1140–1154, 2009. doi: 10.1086/648995. URL https://doi.org/10.1086/648995.

I. Avery Bick, Adrian F. Santiago Tate, Katherine A. Serafin, Alex Miltenberger, Ifeoma Anyansi, Max Evans, Leonard Ortolano, Derek Ouyang, and Jenny Suckale. Rising Seas, Rising Inequity? Communities at Risk in the San Francisco Bay Area and Implications for Adaptation Policy. *Earth's Future*, 9(7):e2020EF001963, 2021. ISSN 2328-4277. doi: 10.1029/2020EF001963. URL https://onlinelibrary.wiley.com/doi/abs/10.1029/2020EF001963. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1029/2020EF001963.

U.S. Census Bureau. Gini index of income inequality. U.S. Census Bureau, 2022a. URL https://data.census.gov/table/ACSDT1Y2022.B19083?q=B19083:GINIINDEXOFINCOMEINEQUALITY&g=040XX00US45$0500000_160XX00US4513330&moe=false. Accessed on 28 May 2024.

U.S. Census Bureau. Total population. U.S. Census Bureau, 2022b. URL https://data.census.gov/table/ACSDT5Y2022.B01003?q=population&g=050XX00US45019$1500000. Accessed on 27 May 2024.

U.S. Census Bureau. Household income in the past 12 months (in 2022 inflation-adjusted dollars). U.S. Census Bureau, 2022c. URL https://data.census.gov/table/ACSDT5Y2022.B19001?q=householdincome&g=050XX00US45019$1500000. Accessed on 27 May 2024.

U.S. Census Bureau. Aggregate household income in the past 12 months (in 2022 inflation-adjusted dollars). U.S. Census Bureau, 2022d. URL https://data.census.gov/table/ACSDT5Y2022.B19025?q=householdincome&g=050XX00US45019$1500000. Accessed on 27 May 2024.

U.S. Census Bureau. Mean household income of quintiles. U.S. Census Bureau, 2022e. URL https://data.census.gov/table/ACSDT5Y2022.B19081?q=B19081:MEANHOUSEHOLDINCOMEOFQUINTILES&g=050XX00US45019$1400000. Accessed on 27 May 2024.

U.S. Census Bureau. Per capita income in the past 12 months (in 2022 inflation-adjusted dollars). U.S. Census Bureau, 2022f. URL https://data.census.gov/table/ACSDT5Y2022.B19301?q=B19301&g=050XX00US45019$1500000. Accessed on 27 May 2024.

U.S. Census Bureau. Housing units. U.S. Census Bureau, 2022g. URL https://data.census.gov/table/ACSDT5Y2022.B25001?q=housingunits&g=050XX00US45019$1500000. Accessed on 27 May 2024.

U.S. Census Bureau. Tenure by units in structure. U.S. Census Bureau, 2022h. URL https://data.census.gov/table/ACSDT5Y2022.B25032?t=Housing&g=050XX00US45019$1500000&y=2022. Accessed on 18 August 2024.

U.S. Census Bureau. American Community Survey, 2022i. URL https://data.census.gov/table/ACSDT1Y2022.B19083?q=B19083:%20GINI%20INDEX%20OF%20INCOME%20INEQUALITY&g=040XX00US45$0500000_160XX00US4513330&moe=false.

U.S. Census Bureau. Gini Index of Income Inequality, 2022j. URL https://data.census.gov/table/ACSDT1Y2022.B19083?q=B19083:GINIINDEXOFINCOMEINEQUALITY&g=040XX00US45$0500000_160XX00US4513330&moe=false.

U.S. Census Bureau. American community survey and puerto rico community survey 2022 subject definitions, 2023. URL https://data.census.gov/table/ACSDT5Y2022.B19301?g=050XX00US45019$1500000&tp=false. Accessed: 2024-07-06.

Andreas Burzel, Laurens Bouwer, Karin De Bruijn, Joan Sala Calero, and Dennis Wagenaar. *Delft-FIAT: An open-source flood impact analysis tool*. Deltares, January 2017. doi: 10.5281/zenodo.1400183.

City of Charleston, Illinois. Charleston, Illinois Code of Ordinances, Section 10-7-4: Accessory Buildings and Structures, 2023. URL https://codelibrary.amlegal.com/codes/charlestonil/latest/charleston_il/0-0-0-6751. Accessed: 2024-07-08.

Alessio Ciullo, Jan H. Kwakkel, Karin M. De Bruijn, Neelke Doorn, and Frans Klijn. Efficient or Fair? Operationalizing Ethical Principles in Flood Risk Management: A Case Study on the Dutch-German Rhine. *Risk Analysis*, 40(9):1844–1862, 2020. ISSN 1539-6924. doi: 10.1111/risa.13527. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/risa.13527. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/risa.13527.

S.l. Cutter, C.t. Emrich, D.p. Morath, and C.m. Dunning. Integrating social vulnerability into federal flood risk management planning. *Journal of Flood Risk Management*, 6(4):332–344, 2013. ISSN 1753-318X. doi: 10.1111/jfr3.12018. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/jfr3.12018. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/jfr3.12018.

Russell Davidson and James G. MacKinnon. Bootstrap tests: how many bootstraps? *Econometric Reviews*, 19(1):55–68, January 2000. ISSN 0747-4938, 1532-4168. doi: 10.1080/07474930008800459. URL http://www.tandfonline.com/doi/abs/10.1080/07474930008800459.

Alok Kumar Dwivedi, Indika Mallawaarachchi, and Luis A. Alvarado. Analysis of small sample size studies using nonparametric bootstrap test with pooled resampling method. *Statistics in Medicine*, 36(14):2187–2205, June 2017. ISSN 0277-6715, 1097-0258. doi: 10.1002/sim.7263. URL https://onlinelibrary.wiley.com/doi/10.1002/sim.7263.

Federal Emergency Management Agency (FEMA). *Hazus 5.1 Flood Model Technical Manual*, July 2022.

Vito Frontuto, Silvana Dalmazzone, Francesco Salcuni, and Alessandro Pezzoli. Risk Aversion, Inequality and Economic Evaluation of Flood Damages: A Case Study in Ecuador. *Sustainability*, 12(23):10068, January 2020. ISSN 2071-1050. doi: 10.3390/su122310068. URL https://www.mdpi.com/2071-1050/12/23/10068. Number: 23 Publisher: Multidisciplinary Digital Publishing Institute.

Jim Hall and Dimitri Solomatine. A framework for uncertainty analysis in flood risk management decisions. *International Journal of River Basin Management*, 6(2):85–98, June 2008. ISSN 1571-5124. doi: 10.1080/15715124.2008.9635339. URL https://doi.org/10.1080/15715124.2008.9635339. Publisher: Taylor & Francis _eprint: https://doi.org/10.1080/15715124.2008.9635339.

Gareth James, Daniela Witten, Trevor Hastie, and Robert Tibshirani. Resampling Methods. In Gareth James, Daniela Witten, Trevor Hastie, and Robert Tibshirani, editors, *An Introduction to Statistical Learning: with Applications in R*, pages 197–223. Springer US, New York, NY, 2021. ISBN 978-1-07-161418-1. doi: 10.1007/978-1-0716-1418-1_5. URL https://doi.org/10.1007/978-1-0716-1418-1_5.

Brenden Jongman, Philip J. Ward, and Jeroen C. J. H. Aerts. Global exposure to river and coastal flooding: Long term trends and changes. *Global Environmental Change*, 22(4):823–835, October 2012. ISSN 0959-3780. doi: 10.1016/j.gloenvcha.2012.07.004. URL https://www.sciencedirect.com/science/article/pii/S0959378012000830.

Jarl Kind, W.j. Wouter Botzen, and Jeroen C.J.H. Aerts. Accounting for risk aversion, income distribution and social welfare in cost-benefit analysis for flood risk management. *WIREs Climate Change*, 8(2): e446, 2017. ISSN 1757-7799. doi: 10.1002/wcc.446. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/wcc.446. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/wcc.446.

Jarl Kind, W. J. Wouter Botzen, and Jeroen C. J. H. Aerts. Social vulnerability in cost-benefit analysis for flood risk management. *Environment and Development Economics*, 25(2):115–134, April 2020. ISSN 1355-770X, 1469-4395. doi: 10.1017/S1355770X19000275. URL https://www.cambridge.org/core/journals/environment-and-development-economics/article/social-vulnerability-in-costbenefit-analysis-for-flood-risk-management/0F039808DF4EF50E4A64163A51C5C986.

A. Krause, A. Martin, and M. Fix. Uncertainty in automated valuation models: Error-based versus model-based approaches. *Journal of Property Research*, 37(4):308–339, October 2020. ISSN 0959-9916. doi: 10.1080/09599916.2020.1807587. URL https://doi.org/10.1080/09599916.2020.1807587. Publisher: Routledge _eprint: https://doi.org/10.1080/09599916.2020.1807587.

Tim Leijnse, Maarten van Ormondt, Kees Nederhoff, and Ap van Dongeren. Modeling compound flooding in coastal systems using a computationally efficient reduced-physics solver: Including fluvial, pluvial, tidal, wind- and wave-driven processes. *Coastal Engineering*, 163:103796, 2021. ISSN 0378-3839. doi: https://doi.org/10.1016/j.coastaleng.2020.103796. URL https://www.sciencedirect.com/science/article/pii/S0378383920304828.

Maryia Markhvida, Brian Walsh, Stephane Hallegatte, and Jack Baker. Quantification of disaster impacts through household well-being losses. *Nature Sustainability*, 3(7):538–547, July 2020. ISSN 2398-9629. doi: 10.1038/s41893-020-0508-7. URL https://www.nature.com/articles/s41893-020-0508-7. Publisher: Nature Publishing Group.

B. Merz, H. Kreibich, R. Schwarze, and A. Thieken. Review article "Assessment of economic flood damage". *Natural Hazards and Earth System Sciences*, 10(8):1697–1724, August 2010. ISSN 1684-9981. doi: 10.5194/nhess-10-1697-2010. URL https://nhess.copernicus.org/articles/10/1697/2010/.

Daniela Molinari, Karin M. De Bruijn, Jesica T. Castillo-Rodríguez, Giuseppe T. Aronica, and Laurens M. Bouwer. Validation of flood risk models: Current practice and possible improvements. *International Journal of Disaster Risk Reduction*, 33:441–448, February 2019. ISSN 2212-4209. doi: 10.1016/j.ijdrr.2018.10.022. URL https://www.sciencedirect.com/science/article/pii/S221242091830596X.

National Structure Inventory. NSI Technical Documentation, 2022. URL https://www.hec.usace.army.mil/confluence/nsi/technicalreferences/latest/technical-documentation.

M.U. Parodi, A. Giardino, A. Van Dongeren, S.G. Pearson, J.D. Bricker, and A.J.H.M. Reniers. Uncertainties in coastal flood risk assessments in small island developing states. *Natural Hazards and Earth System Sciences*, 20(9):2397–2414, 2020. ISSN 1561-8633. doi: 10.5194/nhess-20-2397-2020.

Francisco Peña, Jayantha Obeysekera, Robert Jane, Fernando Nardi, Carolina Maran, Andre Cadogan, Frederique de Groen, and Assefa Melesse. Investigating compound flooding in a low elevation coastal karst environment using multivariate statistical and 2D hydrodynamic modeling. *Weather and Climate Extremes*, 39:100534, March 2023. ISSN 2212-0947. doi: 10.1016/j.wace.2022.100534. URL https://www.sciencedirect.com/science/article/pii/S221209472200113X.

Adam Pollack, James Doss-Gollin, Vivek Srikrishnan, and Klaus Keller. UNSAFE: An UNcertain Structure And Fragility Ensemble framework for property-level flood risk estimation, May 2024. URL https://osf.io/jb9ta.

Brett F. Sanders, Jochen E. Schubert, Daniel T. Kahl, Katharine J. Mach, David Brady, Amir AghaKouchak, Fonna Forman, Richard A. Matthew, Nicola Ulibarri, and Steven J. Davis. Large and inequitable flood risks in Los Angeles, California. *Nature Sustainability*, 6(1):47–57, January 2023. ISSN 2398-9629. doi: 10.1038/s41893-022-00977-7. URL https://www.nature.com/articles/s41893-022-00977-7. Number: 1 Publisher: Nature Publishing Group.

K. Slager, A. Burzel, E. Bos, K. De Bruikn, D. J. Wagenaar, H. C. Winsemius, L. M. Bouwer, and M. Van der Doef. *User Manual Delft-FIAT version 1*. 2016.

Robert Soden, David Lallemant, Manveer Kalirai, Celine Liu, Dennis Wagenaar, and Sophia Jit. The importance of accounting for equity in disaster risk models. *Communications Earth & Environment*, 4(1):386, October 2023. ISSN 2662-4435. doi: 10.1038/s43247-023-01039-2. URL https://www.nature.com/articles/s43247-023-01039-2.

Beth Tellman, Cody Schank, Bessie Schwarz, Peter D. Howe, and Alex de Sherbinin. Using Disaster Outcomes to Validate Components of Social Vulnerability to Floods: Flood Deaths and Property Damage across the USA. *Sustainability*, 12(15):6006, January 2020. ISSN 2071-1050. doi: 10.3390/su12156006. URL https://www.mdpi.com/2071-1050/12/15/6006. Number: 15 Publisher: Multidisciplinary Digital Publishing Institute.

U.S. Army Corps of Engineers. National structure inventory (nsi), 2022. URL https://www.hec.usace.army.mil/confluence/nsi/. Accessed: 2024-05-28.

U.S. Census Bureau. Tiger/line shapefile, 2022, block groups, 2022. URL https://www.census.gov/cgi-bin/geo/shapefiles/index.php?year=2022&layergroup=Block+Groups. Accessed: 2024-06-15.

US Department of Commerce, NOAA. 2023 - December 17 Nor'easter Impacts: Flash Flooding, Wind, Coastal Flooding, and High Surf, 2023. URL https://www.weather.gov/chs/Dec2023. Publisher: NOAA's National Weather Service.

U.S. Geological Survey. 3d elevation program 1-meter resolution digital elevation model. Accessed: 2019-10-23, 2019. URL https://www.usgs.gov/the-national-map-data-delivery. Published 20200606.

Brian Walsh and Stéphane Hallegatte. Measuring Natural Risks in the Philippines: Socioeconomic Resilience and Wellbeing Losses. *Economics of Disasters and Climate Change*, 4(2):249–293, July 2020. ISSN 2511-1299. doi: 10.1007/s41885-019-00047-x. URL https://doi.org/10.1007/s41885-019-00047-x.

A

Income Distribution Method

A.1. Using household income distribution

A.1.1. Mid-point income bracket

For any given income bracket $[L, U]$, where L is the lower bound and U is the upper bound, the mid-point (M) can be calculated using the following formula:

$$M = \frac{L + U}{2}$$

1. For the income bracket $10,000 to $14,999:

$$M = \frac{10{,}000 + 14{,}999}{2} = \frac{24{,}999}{2} = 12{,}499.5 \approx 12{,}500$$

2. For the income bracket $45,000 to $49,999:

$$M = \frac{45{,}000 + 49{,}999}{2} = \frac{94{,}999}{2} = 47{,}499.5 \approx 47{,}500$$

For the highest income bracket, where the upper bound is not specified, an arbitrary large value is assumed. $250,000 is used as the upper bound for the bracket ”$200,000 or more”.

A.1.2. Uniform Distribution

Given:

- n_i: Number of households in the i-th income bracket
- L_i: Lower bound of the i-th income bracket
- U_i: Upper bound of the i-th income bracket
- x: Income

The uniform distribution method assumes that households are uniformly distributed across each income bracket. For a given income bracket $[L_i, U_i]$, the probability density function (PDF) $f_i(x)$ is constant and can be defined as:

$$f_i(x) = \frac{n_i}{U_i - L_i}, \quad \text{for } L_i \leq x \leq U_i$$

This implies that the number of households within any sub-interval $[a, b] \subseteq [L_i, U_i]$ is given by:

$$\text{Number of households between } a \text{ and } b = n_i \cdot \frac{b-a}{U_i - L_i}, \quad \text{for } L_i \leq a < b \leq U_i$$

Consider the income bracket \$10,000 to \$14,999 with $n_i = 35$ households. The lower bound $L_i = 10{,}000$ and the upper bound $U_i = 14{,}999$.

The PDF $f_i(x)$ is:

$$f_i(x) = \frac{35}{14{,}999 - 10{,}000} = \frac{35}{4{,}999}$$

If we want to find the number of households between \$12,000 and \$13,000 within this bracket:

$$\text{Number of households between } \$12{,}000 \text{ and } \$13{,}000 = 35 \cdot \frac{13{,}000 - 12{,}000}{14{,}999 - 10{,}000} = 35 \cdot \frac{1{,}000}{4{,}999} \approx 7$$

An example results using the uniform distribution as follows:

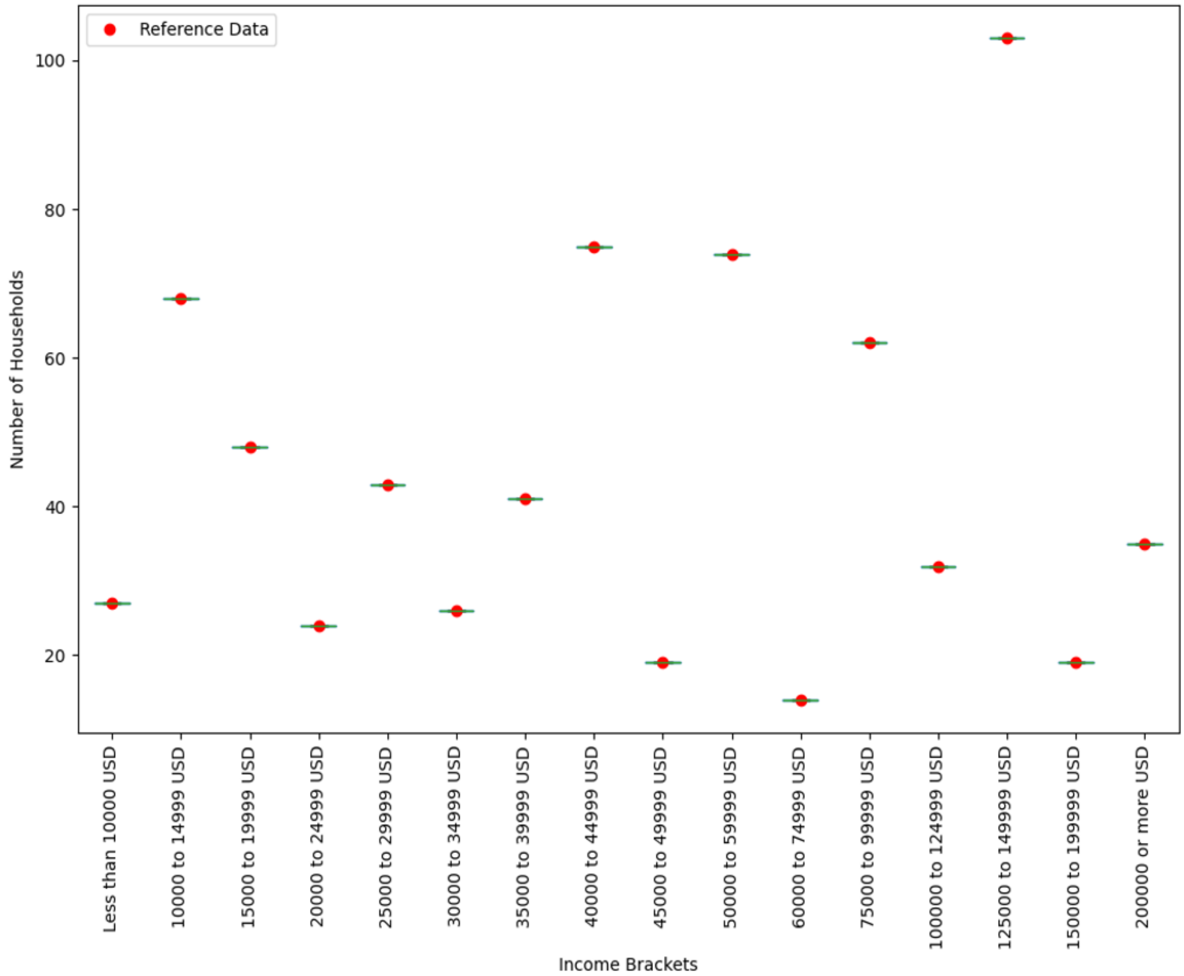


Figure A.1: Uniform Distribution of Household Income at selected Census Block Group

A.1.3. Comparative Analysis

In this section, the two methods above will be compared and it is decided which method will be used for further analysis in the application for case study.

From Figure A.2, it is shown that the uniform distribution has more accurate results compared to census block-scale income statistics.

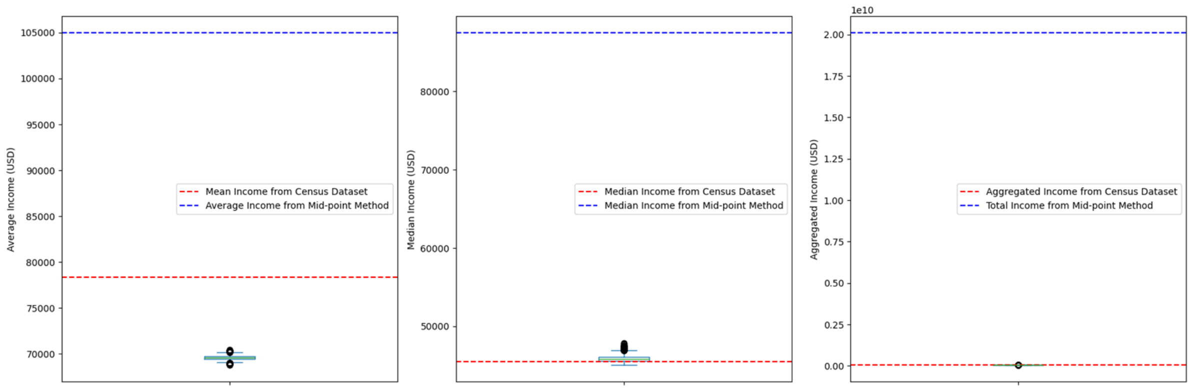


Figure A.2: Comparison between the uniform distribution household income and mid-point method with census household income statistic at selected Census Block Group

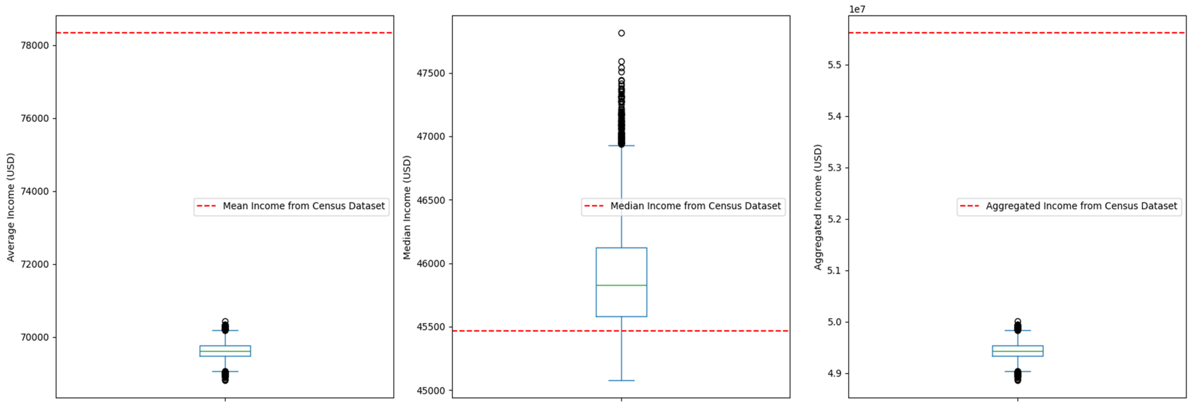


Figure A.3: Comparison between the uniform distribution household income with census household income statistic at selected Census Block Group

Therefore, further analysis for the case study application will use the uniform distribution.

A.2. Using household income statistics

A.2.1. Log normal distribution

In the log-normal distribution method, the distribution parameters μ and σ need to be defined. This study employs three methods to determine these distribution parameters. The assumption made in this determination is that the mid-point of each income bracket represents the individual household income.

First, the **log transform method** uses the individual household income data. Each income value is log-transformed, and the mean and standard deviation of these log-transformed incomes are calculated. The mean of the log-transformed incomes will be μ and the standard deviation will be σ, serving as the distribution parameters.

$$\mu_{\text{log transform}} = \frac{1}{N}\sum_{i=1}^{N}\log(x_i)$$

$$\sigma_{\text{log transform}} = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}\left(\log(x_i) - \mu_{\text{log transform}}\right)^2}$$

where x_i represents the individual income values and N is the total number of households.

Second, the **theoretical log-normal method** uses the individual household income data to calculate the mean (M) and standard deviation (S). These values are then applied to the following formulas to obtain the μ and σ distribution parameters:

$$\mu_{\text{theoretical}} = \ln\left(\frac{M^2}{\sqrt{M^2 + S^2}}\right)$$
$$\sigma_{\text{theoretical}} = \sqrt{\ln\left(1 + \frac{S^2}{M^2}\right)}$$

For example, if the mean income M is \$50,000 and the standard deviation S is \$20,000, the parameters are calculated as follows:

$$\mu_{\text{theoretical}} = \ln\left(\frac{50{,}000^2}{\sqrt{50{,}000^2 + 20{,}000^2}}\right) = \ln(45{,}662.66) \approx 10.726$$
$$\sigma_{\text{theoretical}} = \sqrt{\ln\left(1 + \frac{20{,}000^2}{50{,}000^2}\right)} = \sqrt{\ln(1.16)} \approx 0.378$$

Finally, the **empirical log-normal method** employs an empirical approach using tools from the `scipy` package to determine the scale, shape, and location parameters. These parameters are then transformed into μ and σ for the distribution parameters. The location parameter is set to 0, assuming there is no negative income, and the scale parameter is directly taken from the household income statistics.

Using `scipy.stats.lognorm` to fit the income data:

$$\text{shape, loc, scale} = \text{scipy.stats.lognorm.fit(data, floc=0)}$$

where shape $= \sigma$, loc is fixed to 0, and scale $= e^{\mu}$.

For example, if the fitted parameters are shape $= 0.4$, loc $= 0$, and scale $= 55{,}000$, then:

$$\sigma_{\text{empirical}} = 0.4$$
$$\mu_{\text{empirical}} = \ln(55{,}000) \approx 10.915$$

A.2.2. Comparative analysis

From the three methods of determining the log normal distribution parameters, in this section it will be determined which one method will be used for further analysis in the uncertainty and sensitivity analysis.

The following are the results for household income in the same census block group area:

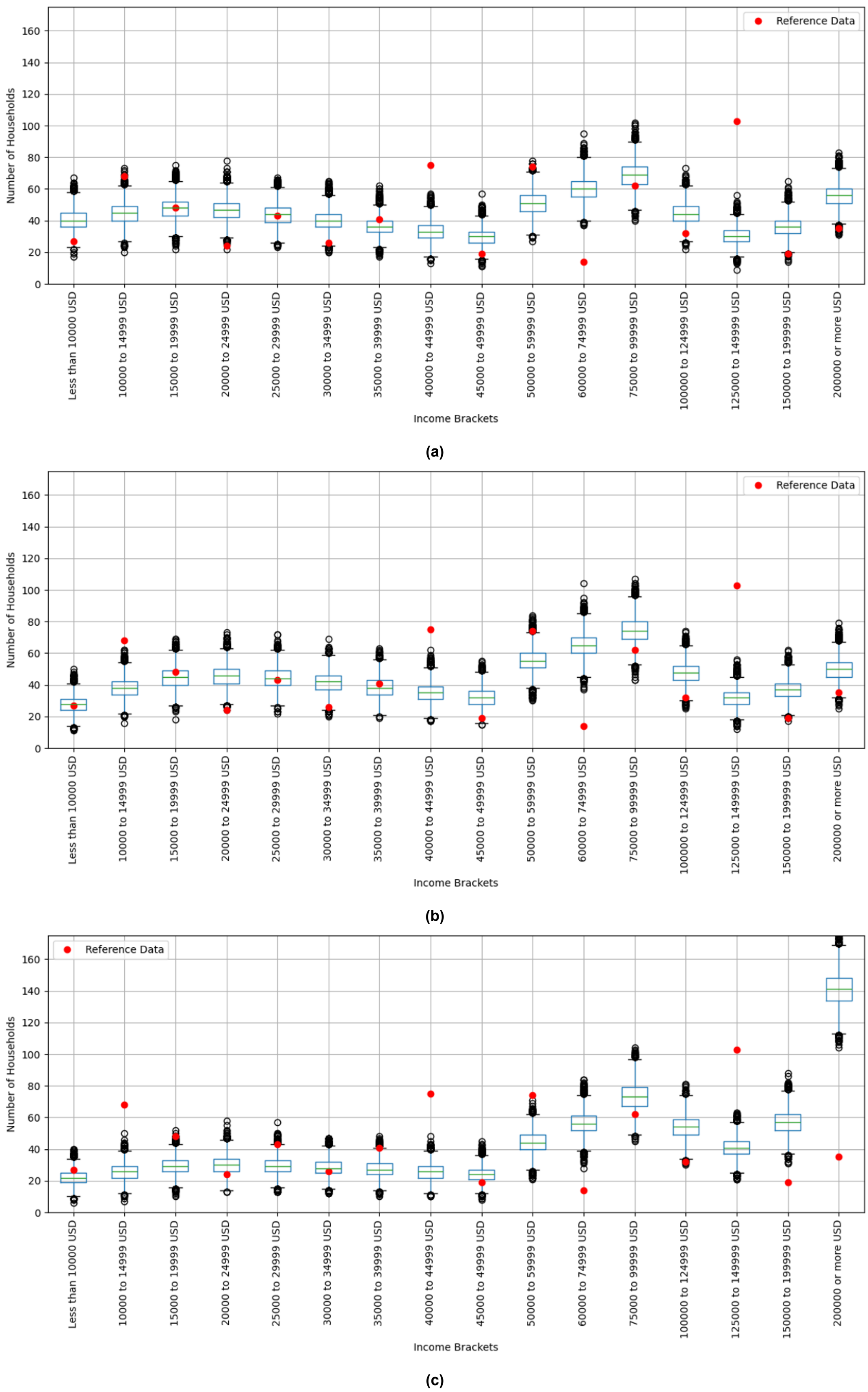


Figure A.4: Example of implementation distribution parameters log transform method (a), theoretical log normal method (b), and empirical log normal method (c)

The following is a comparison between the resulting household income statistics and the census income statistics:

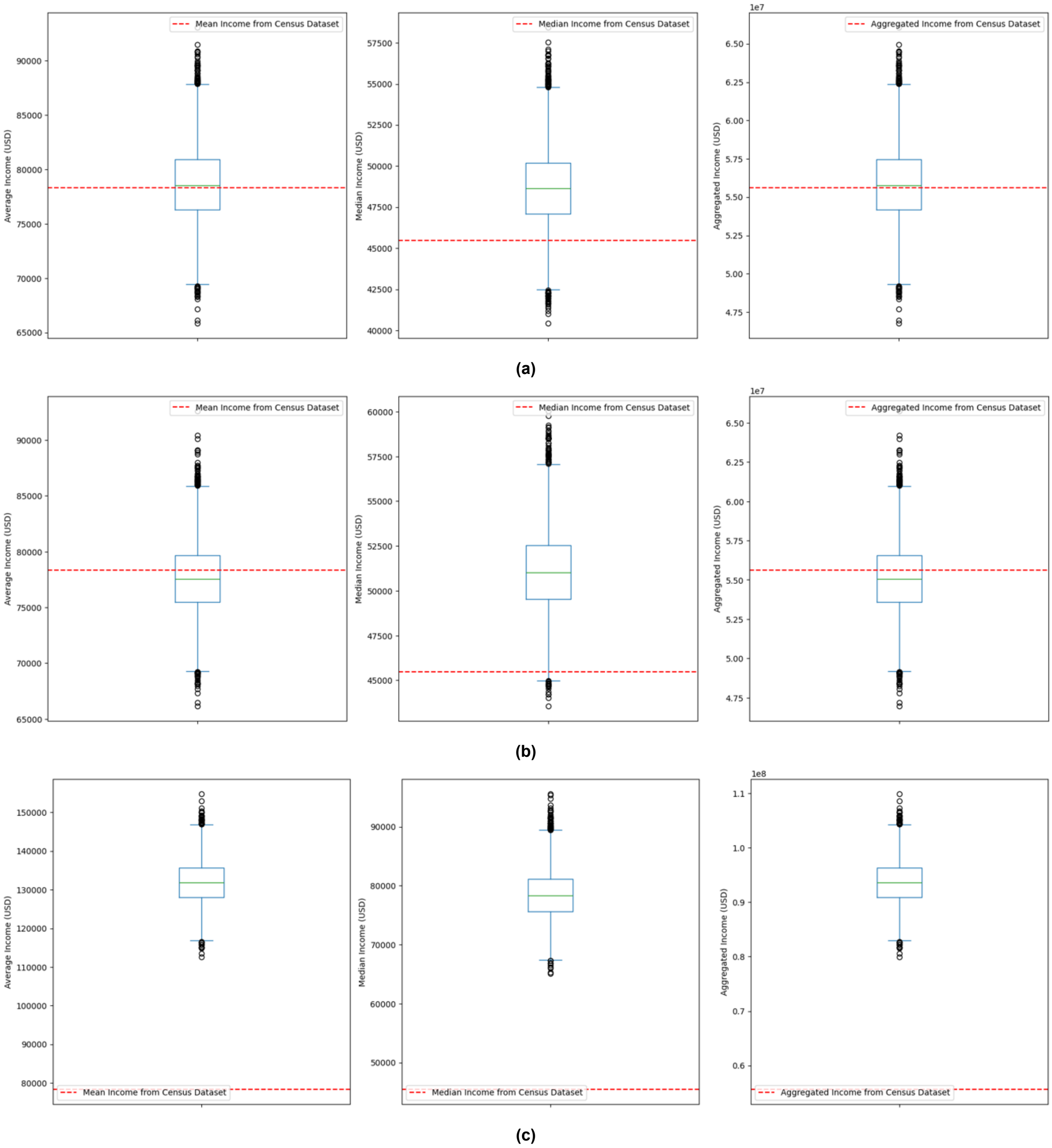


Figure A.5: Example of comparison distribution parameters log transform method (a), theoretical log normal method (b), and empirical log normal method (c) with census household income statistic

Figures A.4 and A.5 show household income with distribution parameters obtained from various methods. The figure shows that the household income with the log transform method closely matches the actual distribution.

FiguresA.6 also show that the log transform method can better capture the first peak of the lowest income compared to the other methods.

Therefore, based on this comparative analysis, the study decided to use the log transform method for further analysis.

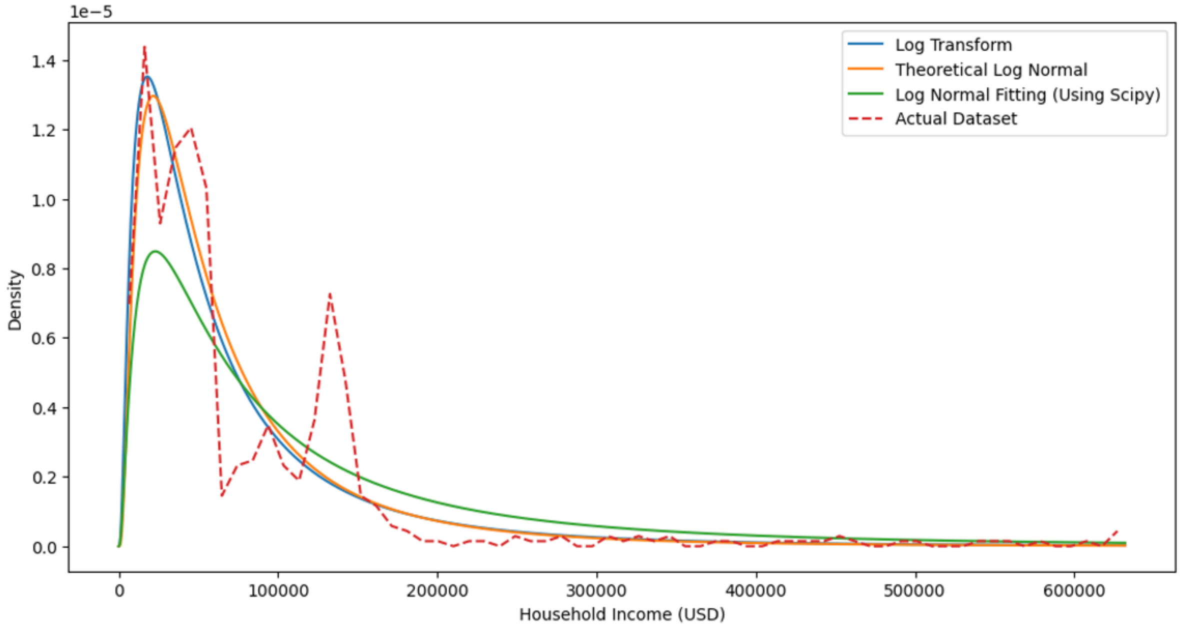


Figure A.6: PDF Comparison from all methods and the actual dataset

A.3. Selection of representative census block group

The selection of representative census blocks using k-means clustering, to determine the optimal K or the number of cluster, elbow and silhouette methods was conducted, the results are as follows:

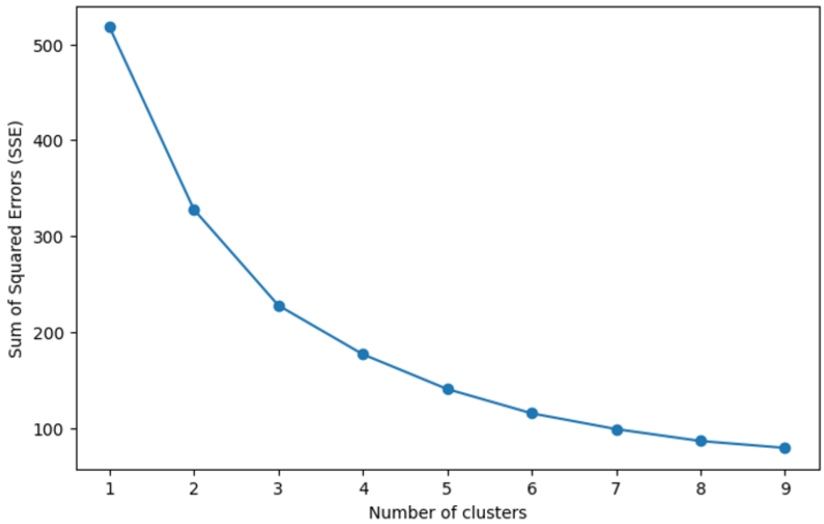


Figure A.7: Elbow method for optimal K

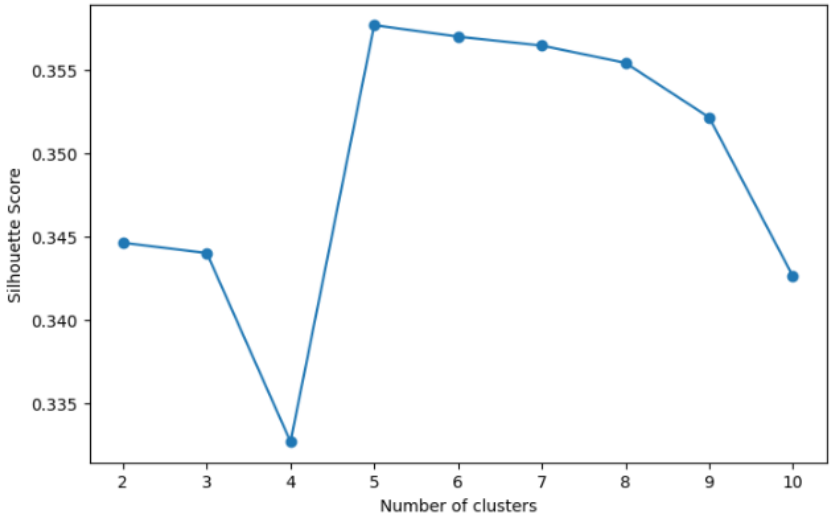


Figure A.8: Silhouette method for optimal K

As shown in Figure A.7, the optimal number of clusters (K) is around 4 or 5, as indicated by the less steep drop or smaller drop in the Sum of Squared Errors (SSE). Then for the results of the Silhouette method from Figure A.8 shows that the highest value of silhouette is with 5 clusters. For this study, the number of clusters was determined as 5, and the clustering results are as follows:

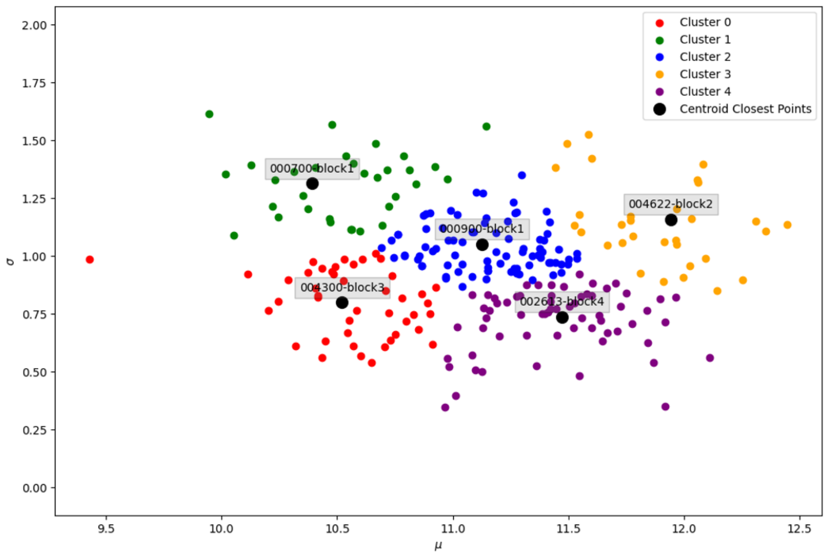
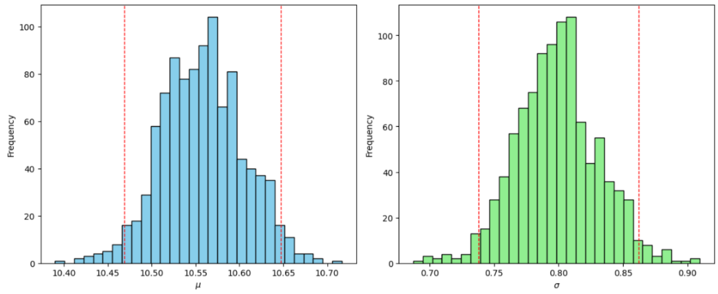


Figure A.9: K-means clustering of census block groups

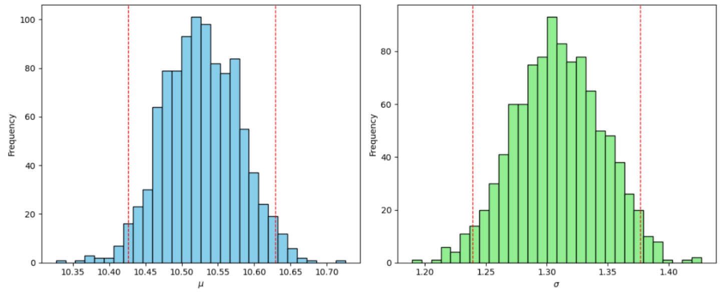
After performing clustering, the census block groups closest to each cluster centroid were selected. Bootstrapping will then be applied to determine the range of the distribution parameters μ and σ for these selected groups. This method is used to assess the uncertainty in the fitted distribution parameters, which will be utilized for equity weight sensitivity analysis.

A.3.1. Range of uncertainty distribution parameters

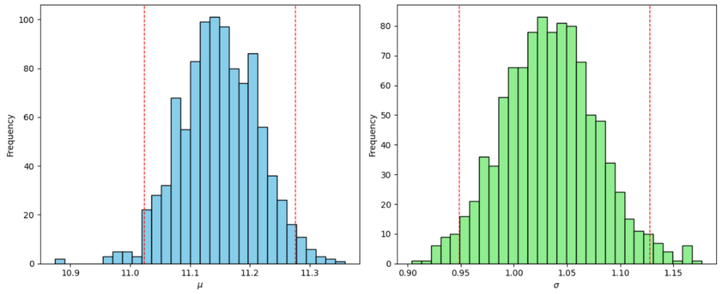
Figure A.10 presents the result of bootstrapping for representative census block groups that have been selected before.



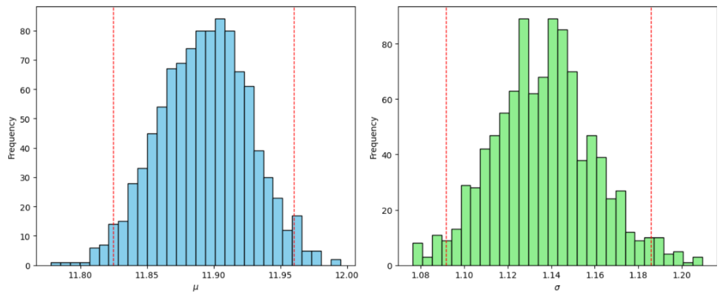
(a)



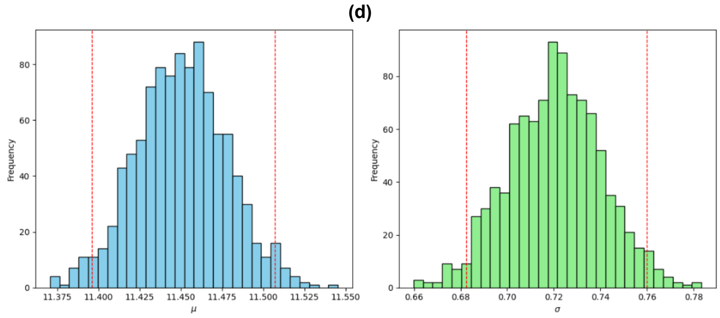
(b)



(c)



(d)



(e)

Figure A.10: Bootstrapping results for parameters μ and σ for the following Census Block Groups: (a) '004300-block3', (b) '000700-block1', (c) '000900-block1', (d) '004622-block2', and (e) '004622-block2'

A.3.2. Selected census block group

The following are the household income results generated for the selected census block groups that will be used for sensitivity and uncertainty exploration.

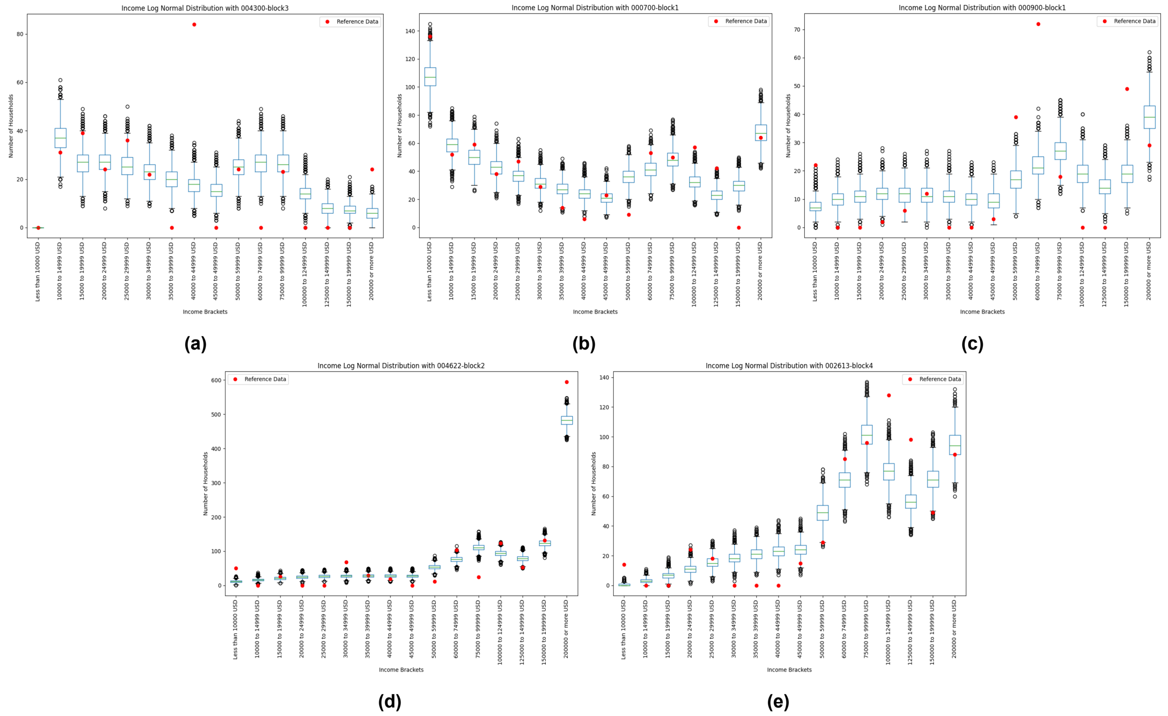


Figure A.11: Household Income with Log Normal Distribution for the following Census Block Groups: (a) '004300-block3', (b) '000700-block1', (c) '000900-block1', (d) '004622-block2', and (e) '002613-block4'

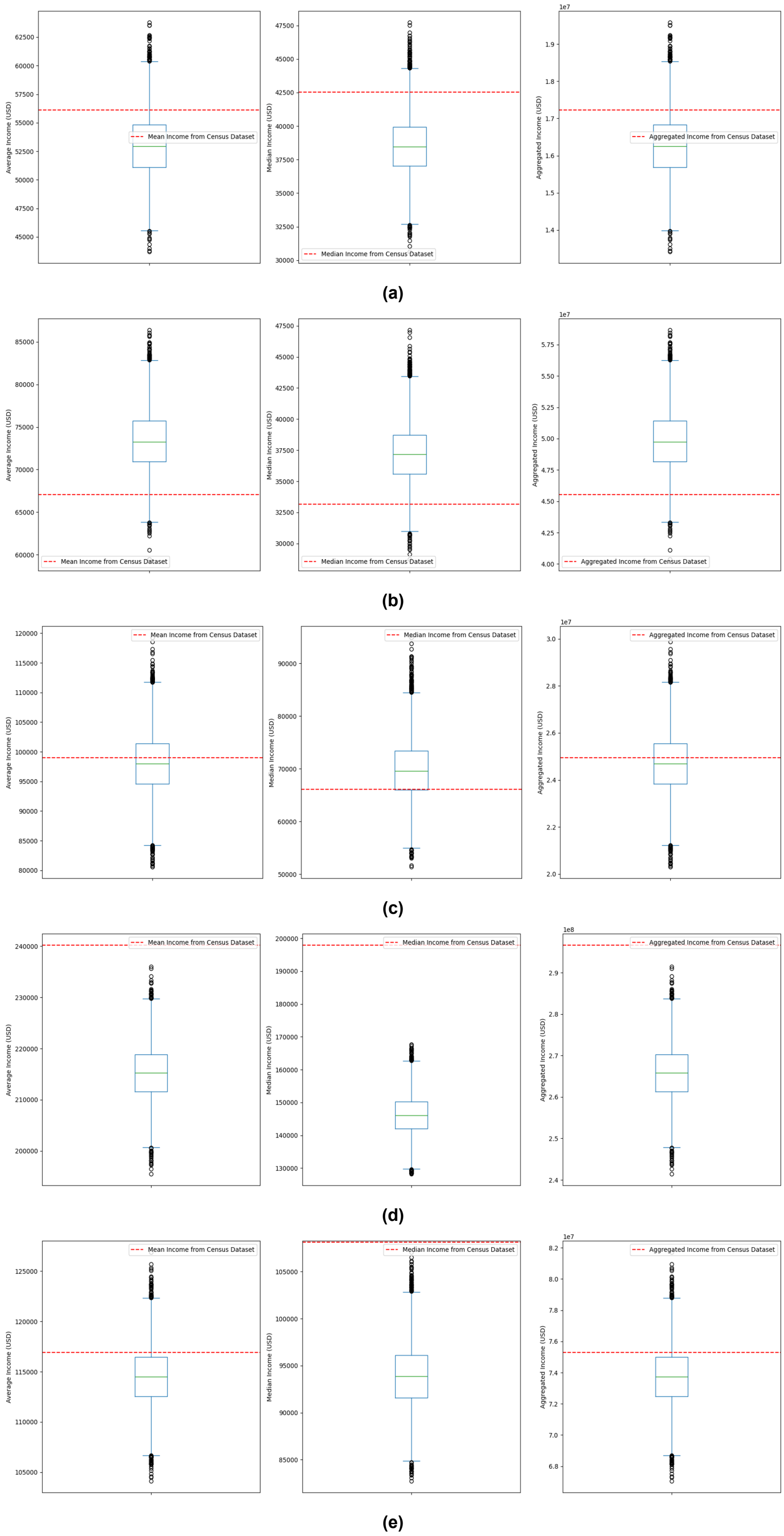


Figure A.12: Statistics on Household Income with Log Normal Distribution for the following Census Block Groups: (a) '004300-block3', (b) '000700-block1', (c) '000900-block1', (d) '004622-block2', and (e) '002613-block4'

B

Spatial Allocation of Income Methods

B.1. Matching total housing units and households

As mentioned in section 3.4 to link the datasets between spatial and statistical, several steps will be taken.

The first is to optimize the number of housing units per residential type as shown in the following figure:

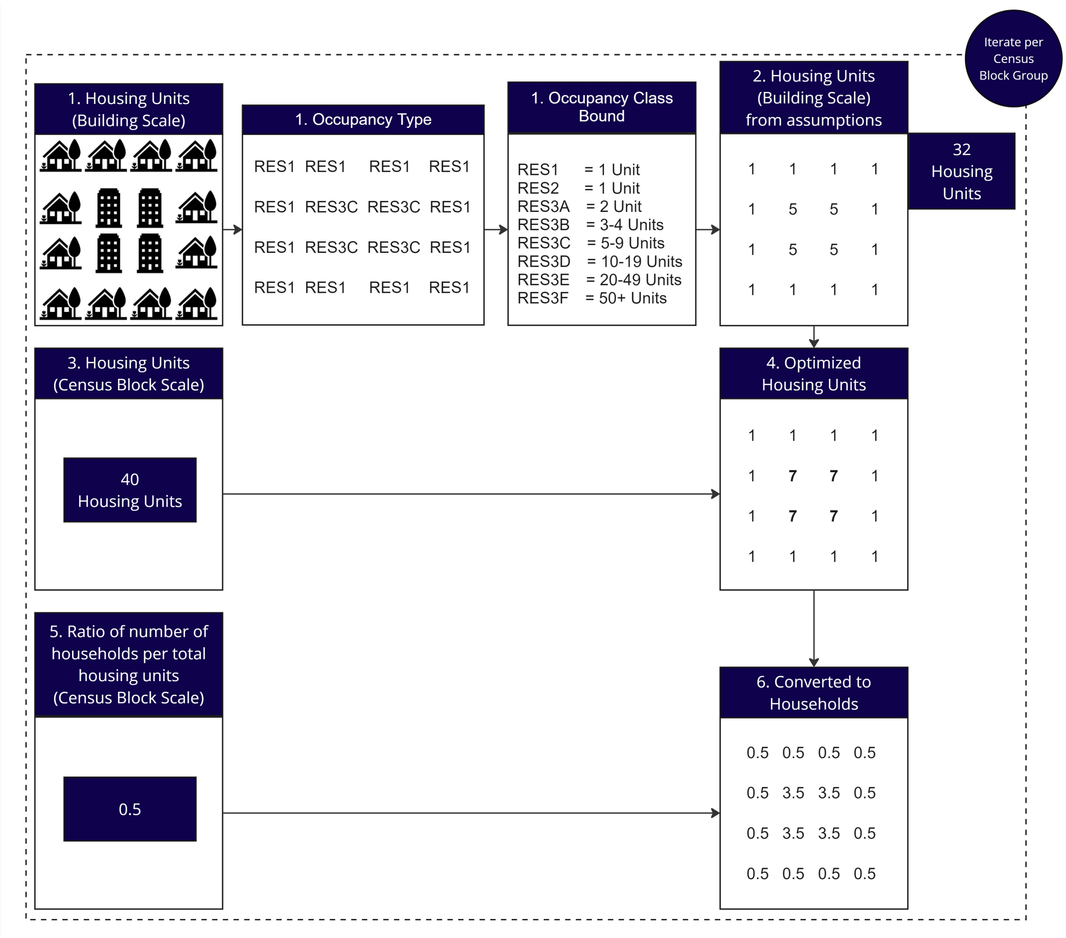


Figure B.1: Workflow for optimization number of housing units per occupancy type

From the NSI dataset, each building point has information on occupancy type (Figure B.1.1). NSI dataset also describes each occupancy type which derived into total or range of total housing units per occupancy type (Figure B.1.1). In census dataset it was specified the total housing units (Figure B.1.3) and the total households in census block group scale. Therefore, these 2 data can give the ratio number of households per total housing units (Figure B.1.5).

Optimization will be performed to minimize the sum of squared errors of total housing units from the census and total housing units from the generation. The variables that can change in this optimization are the total housing units of RES3B, RES3C, RES3D, RES3E, and RES3F. Since RES3F has an open range of housing units, 1325 housing units are assumed. This assumption is obtained from the maximum area of the RES3F building footprint divided by the minimum area for one housing unit in Charleston County, which is 640 sq-ft as specified in City of Charleston, Illinois (2023).

After determining the optimized housing units per occupancy type, this number will be converted to the number of households for each building point (Figure B.1.6) by multiplying it by the ratio (Figure B.1.5). This workflow B.1 is iterated across all census block groups in Charleston County resulting in the output of a set of number of households in each building point per census block group.

The second step is to normalize and randomize the number of households for each building point, as shown in the example workflow from Figure B.2 The household totals generated from the first step and the household totals at the census block scale from the census were used. The difference between the two household totals (Figures B.2.1 and B.2.2) will be normalized to the resulting household total. For example, a multi-family house has five households and the missing households are eight households, the fraction of this multi-family house (5/32) is multiplied by the number of missing households (8) resulting in 6.25 in Figure B.2.3. After normalization, rounding to the nearest integer is performed (Figure B.2.4) and followed by random correction with additional priority rules (Figure B.2.5) to add or subtract based on the capacity of the higher occupancy type first.

This step is repeated across all census block groups in Charleston County, resulting in a set of households in each building point per census block group.

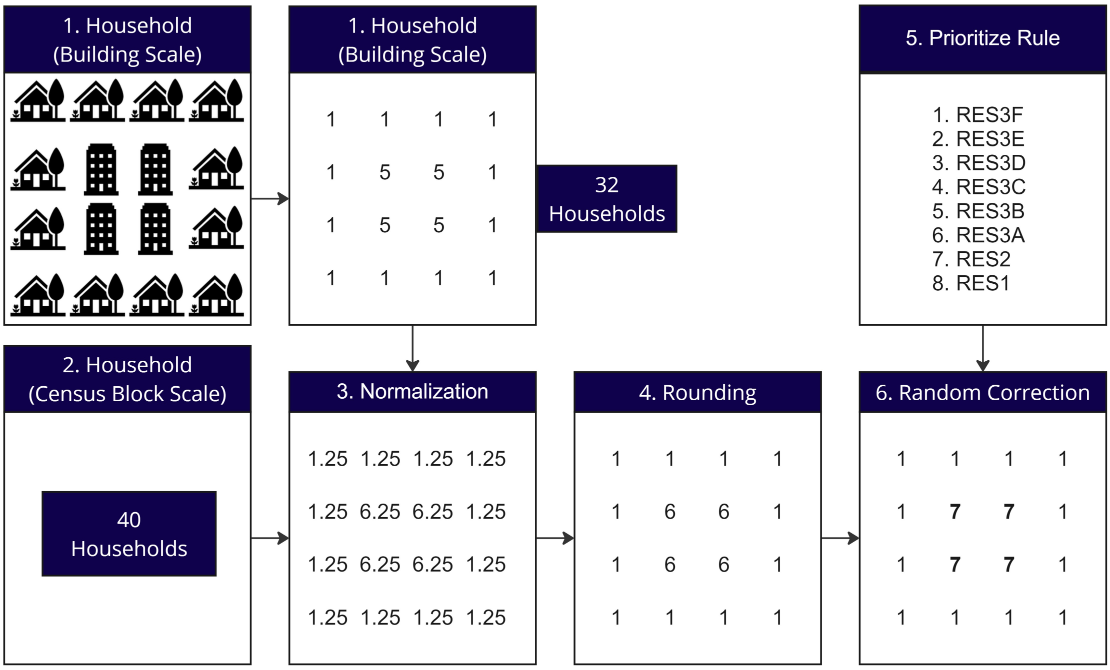
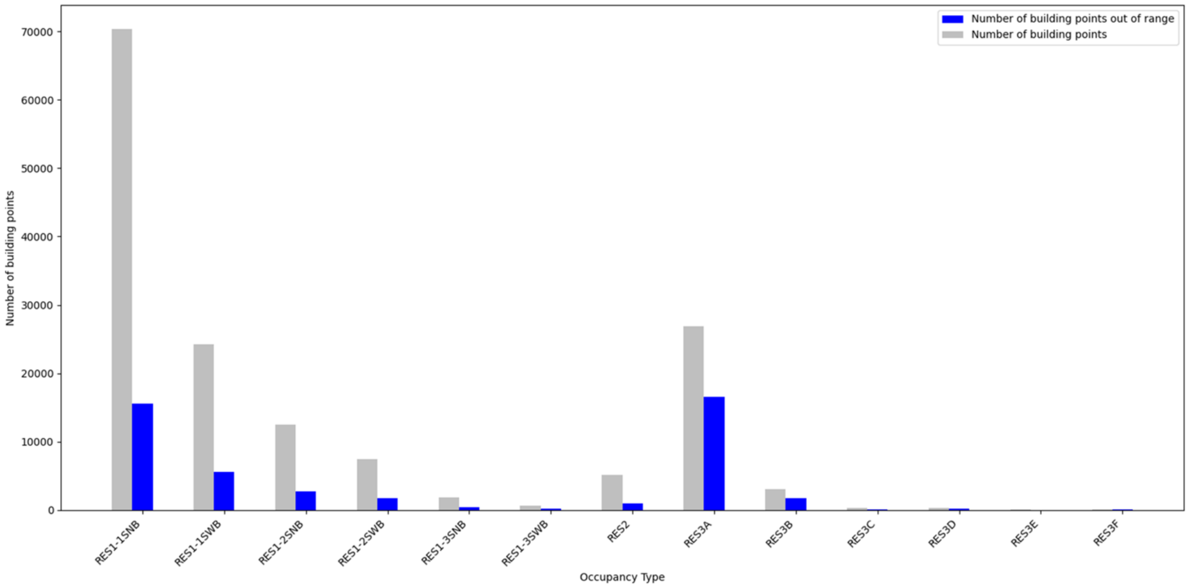
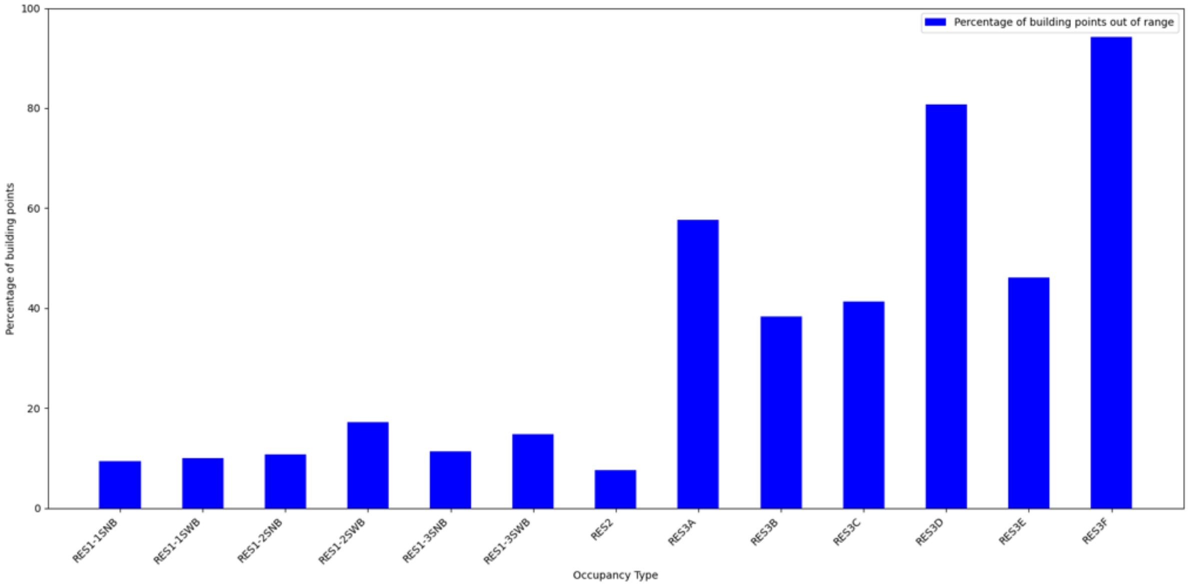


Figure B.2: Workflow for normalizing and random correction for number of households adapted from Bick et al. (2021).

In order to properly match the NSI building point spatial data with the census dataset, during the random correction step, some occupancy types were determined to be more or less than the range provided by the dataset itself.



(a) count of building points out of range



(b) percentage count of building points out of range

Figure B.3: Verification of total housing units in each building points for all census block groups

C
Map of Results

This appendix shows all maps with the full extent of Charleston County that has been discussed in section 4.

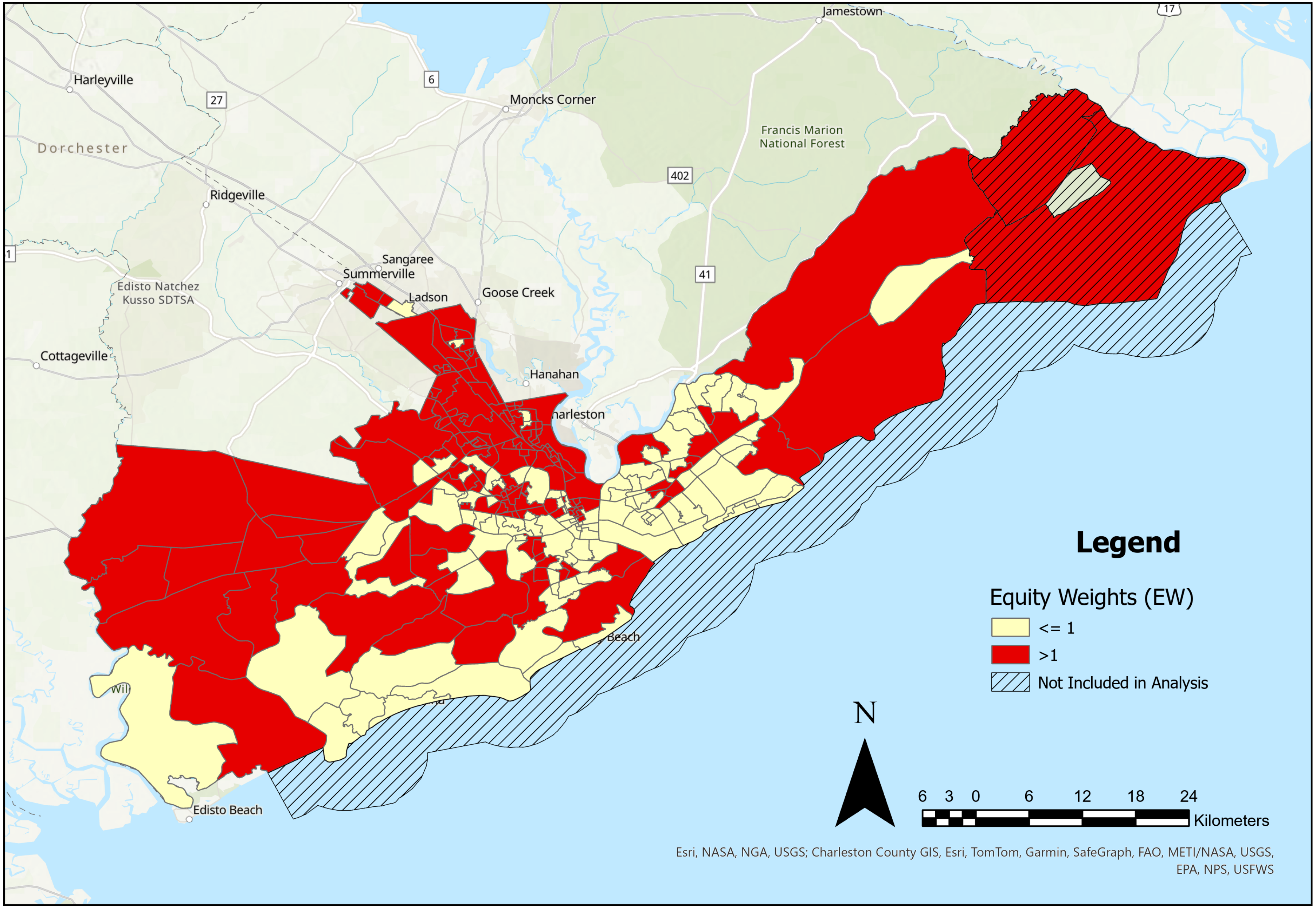


Figure C.1: EW map with aggregated method for entire Charleston County

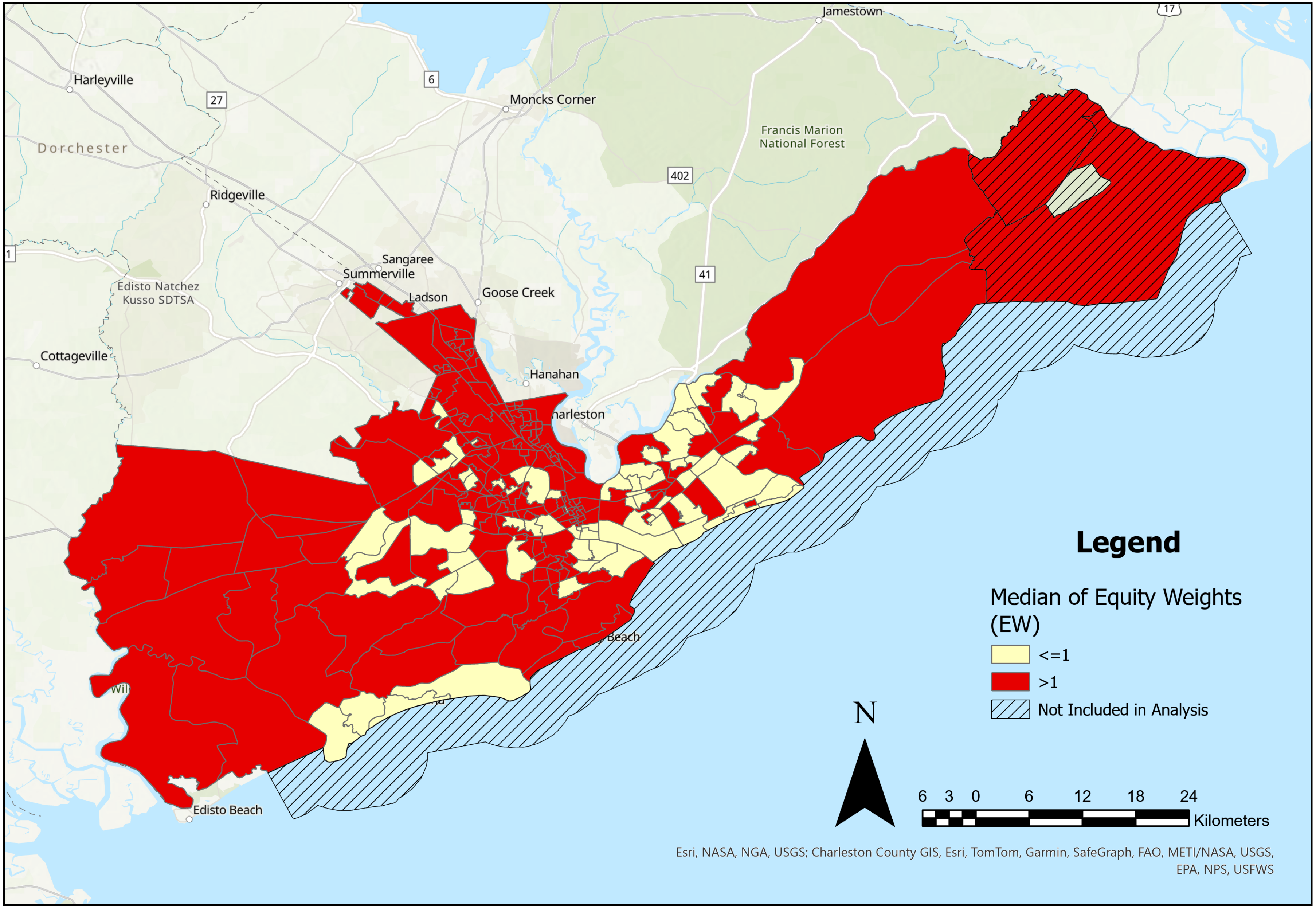


Figure C.2: EW map with disaggregated method for entire Charleston County

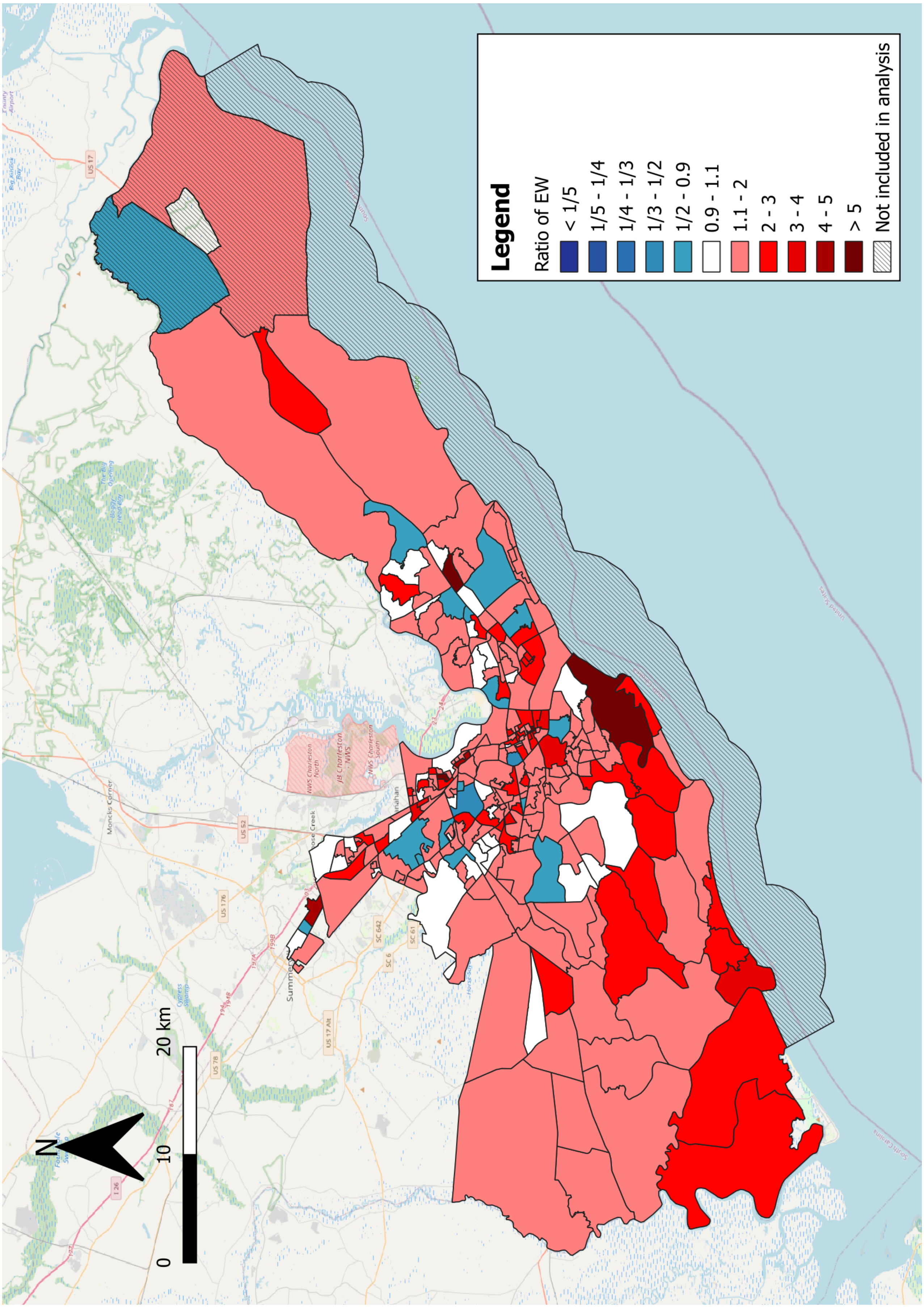


Figure C.3: Map of ratio EW between methods for entire Charleston County

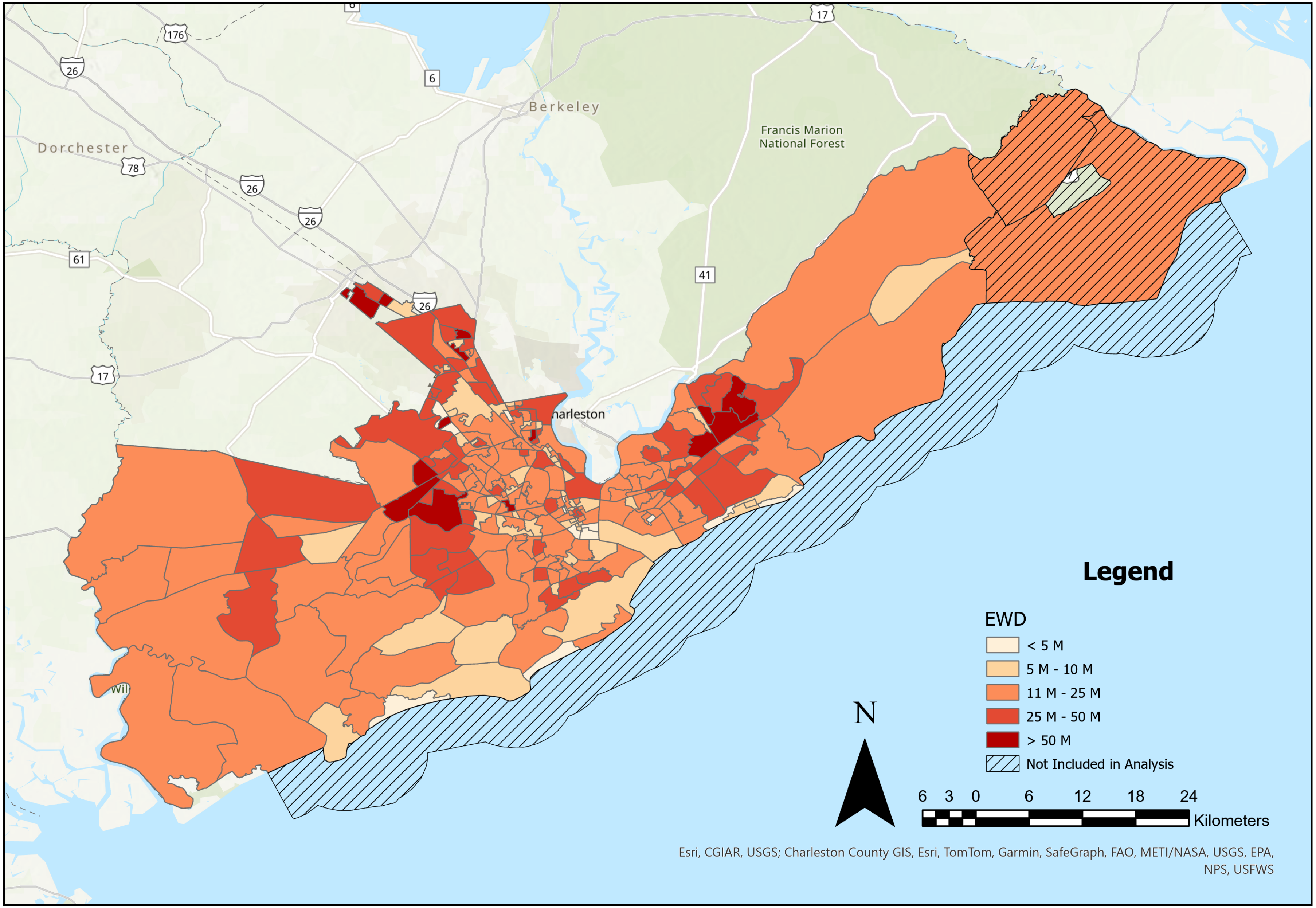


Figure C.4: EWD Result Map with aggregated method for fictitious damage scenario

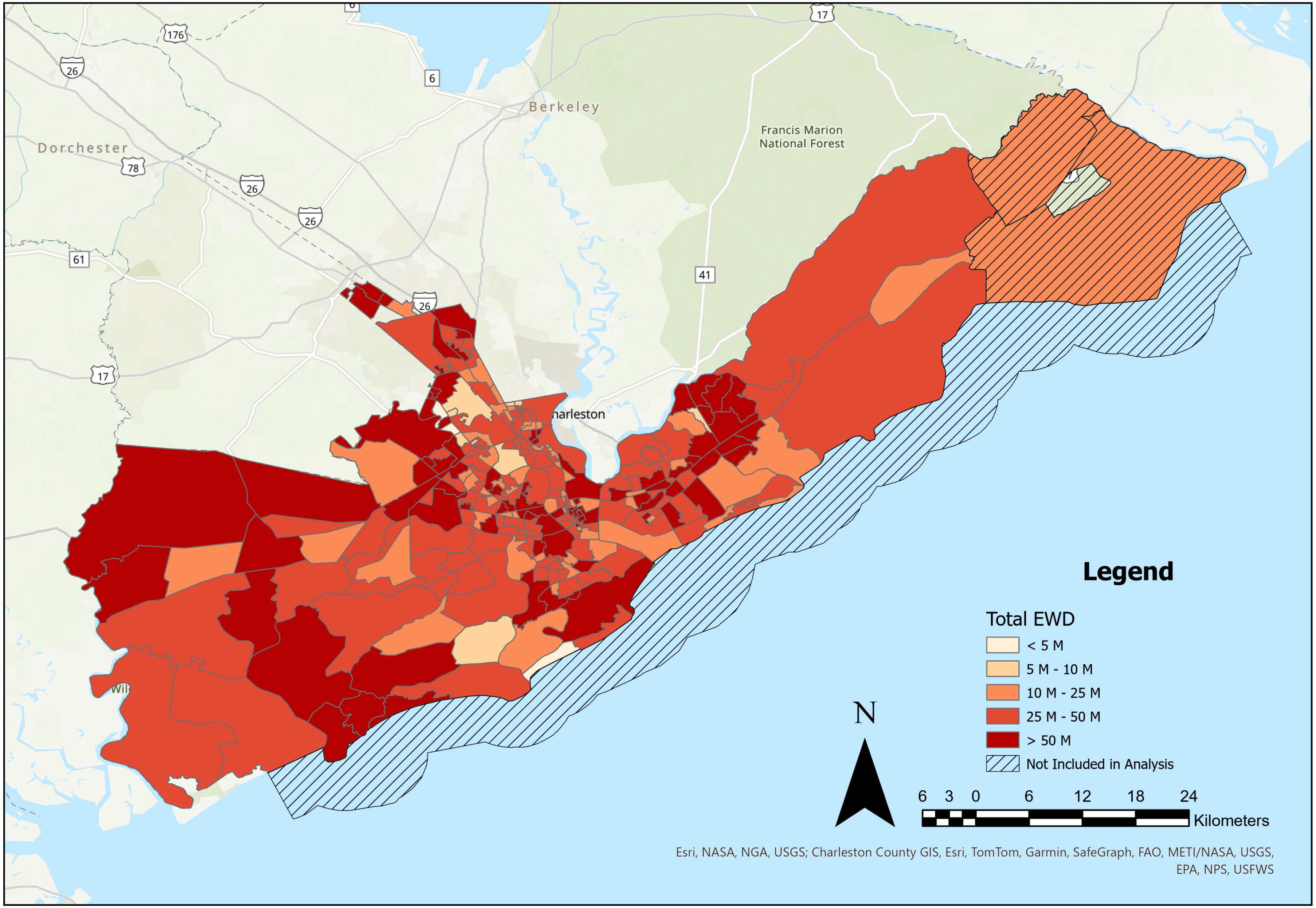


Figure C.5: EWD Result Map with disaggregated method for fictitious damage scenario

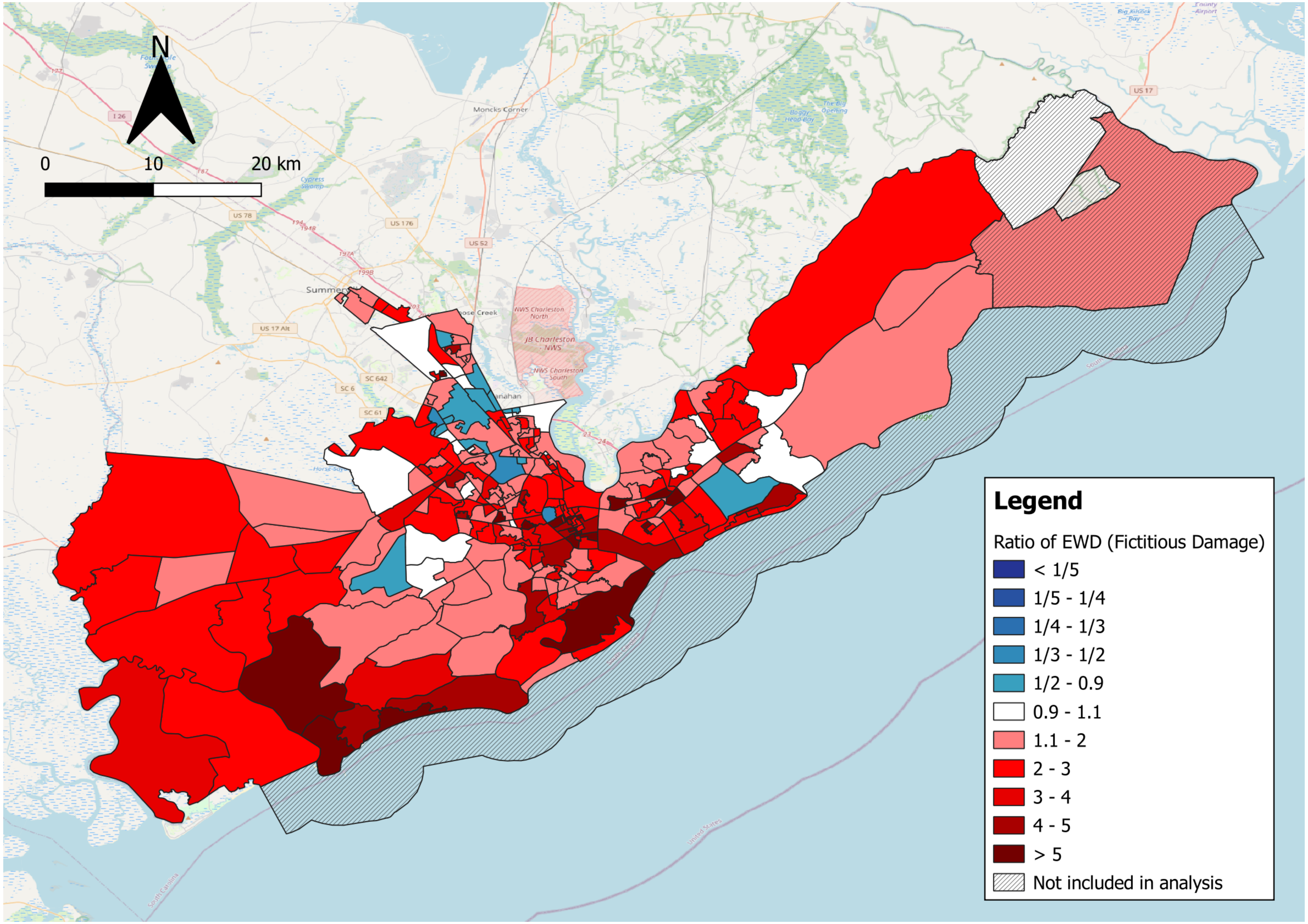


Figure C.6: Map of ratio EWD between methods for fictitious damage scenario

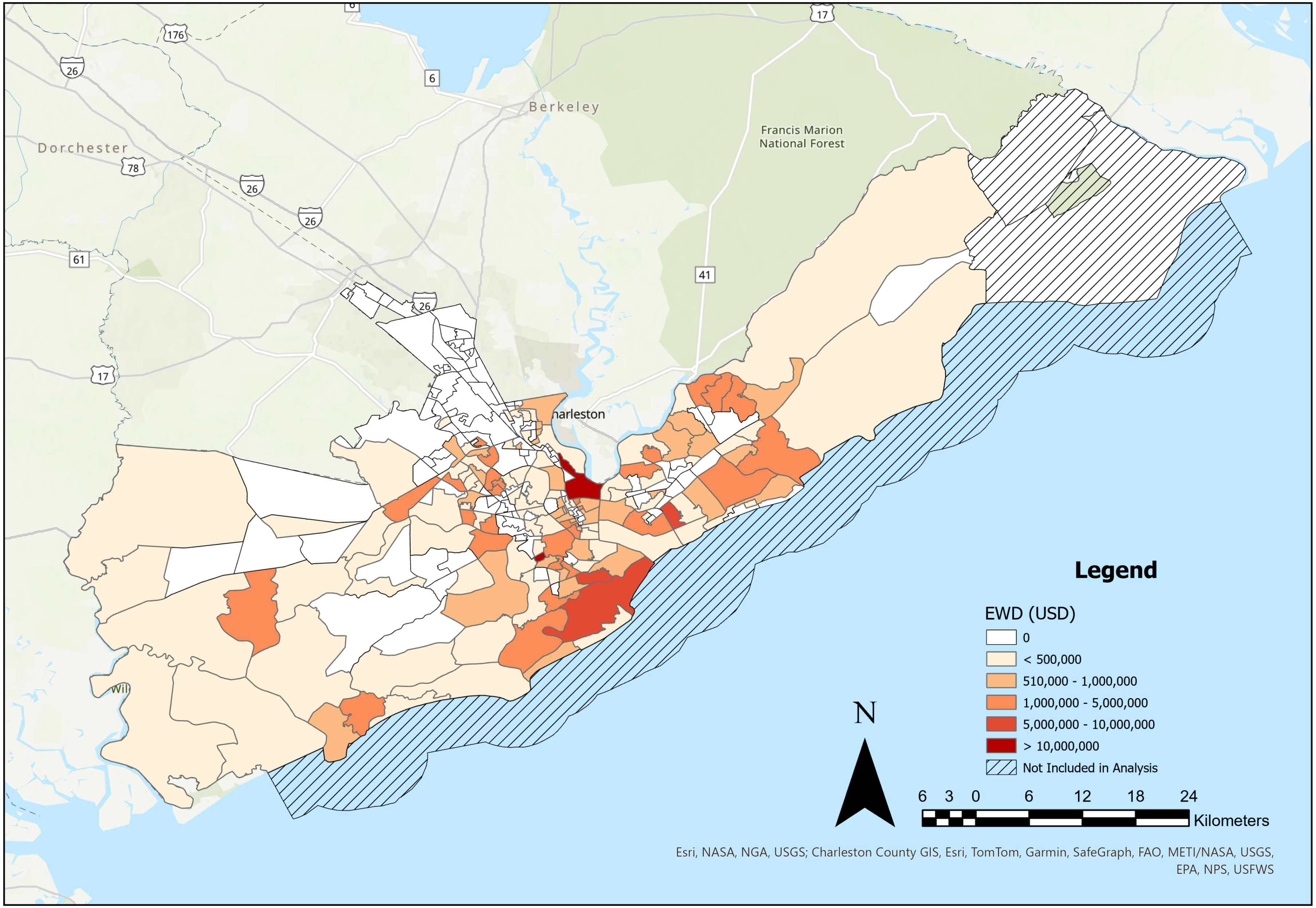


Figure C.7: EWD Result Map with aggregated method for historic event scenario

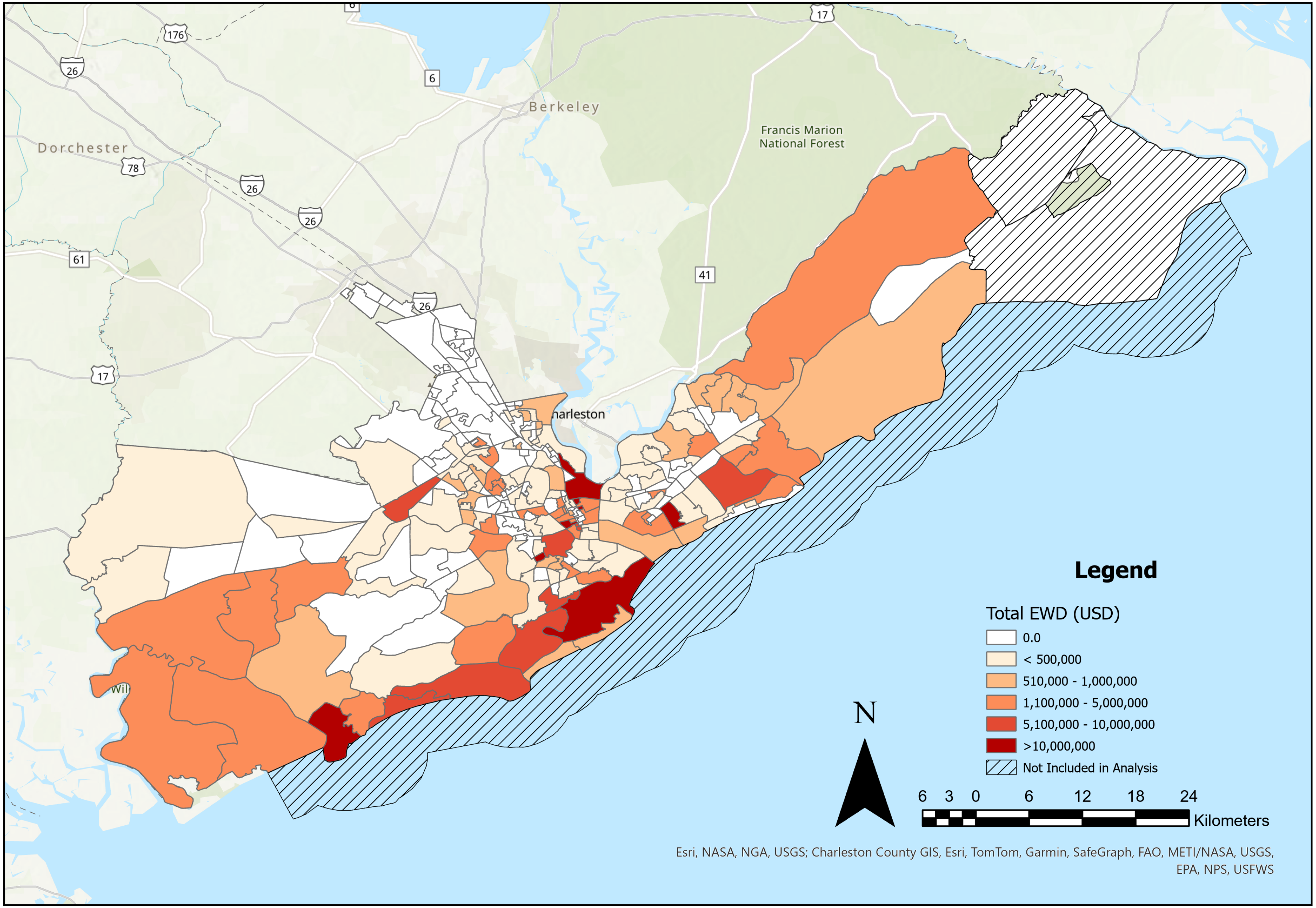


Figure C.8: EWD Result Map with disaggregated method for historic event scenario

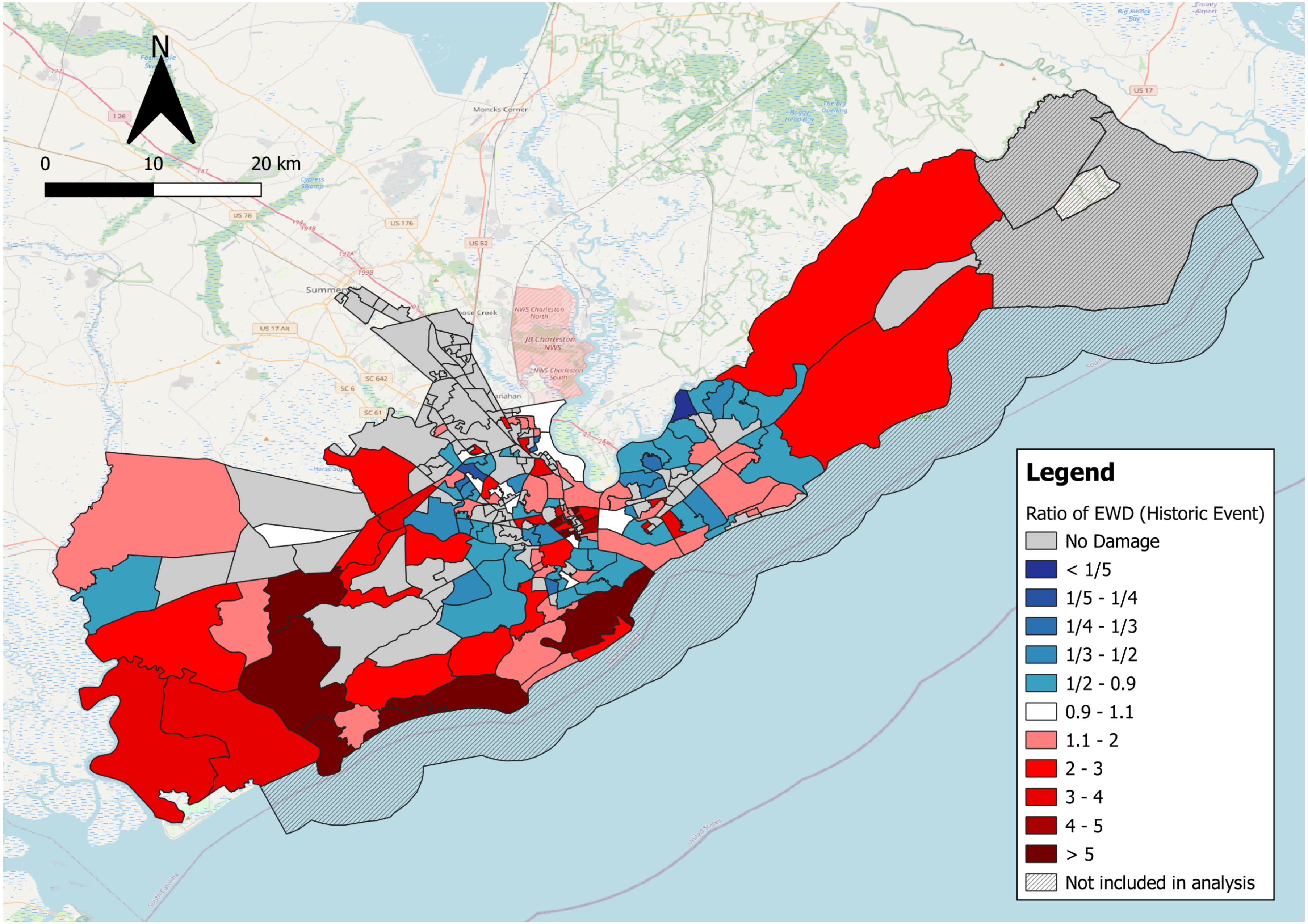


Figure C.9: Map of ratio EWD between methods for historic event scenario

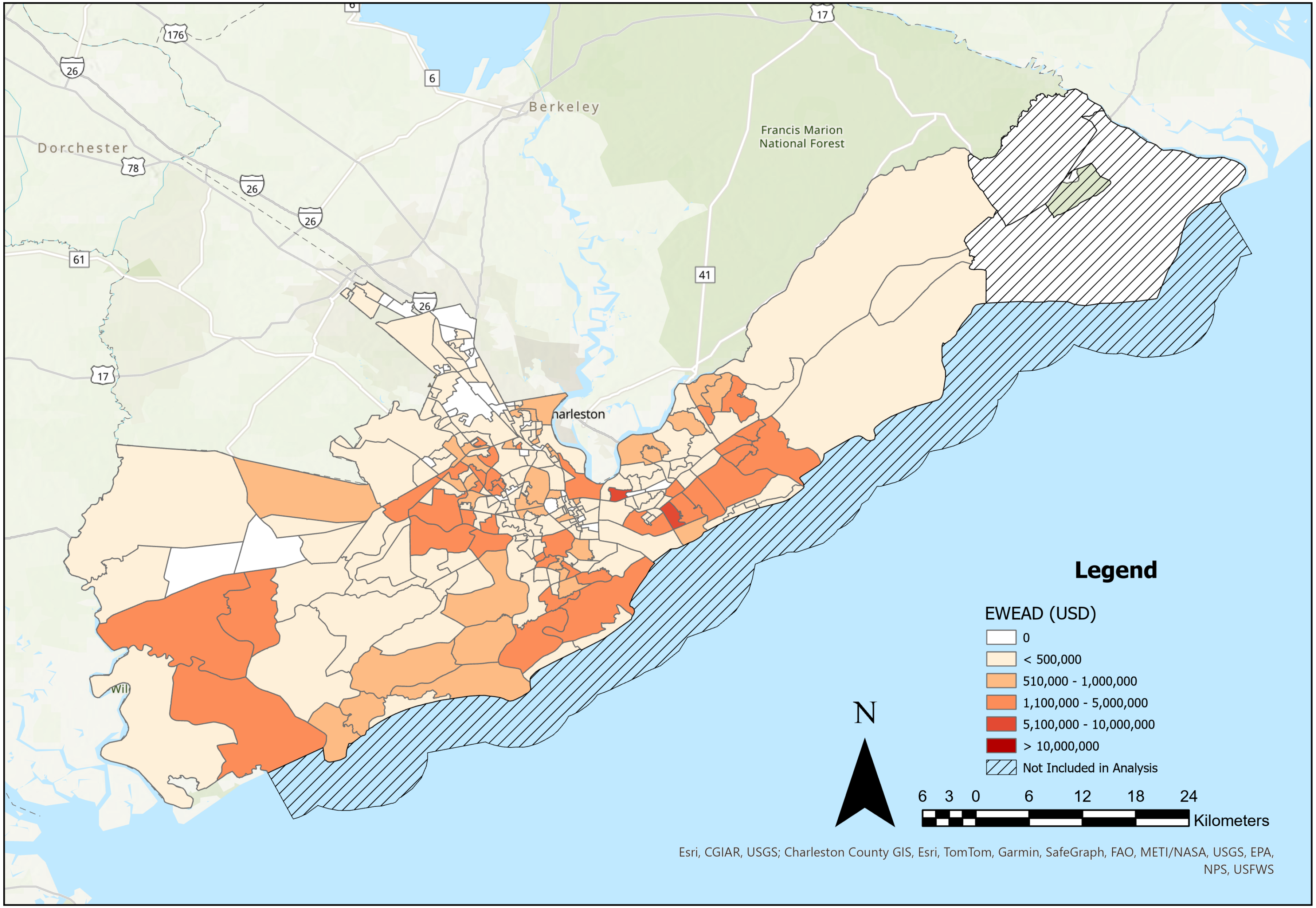


Figure C.10: EWEAD Result Map with aggregated method for risk scenario

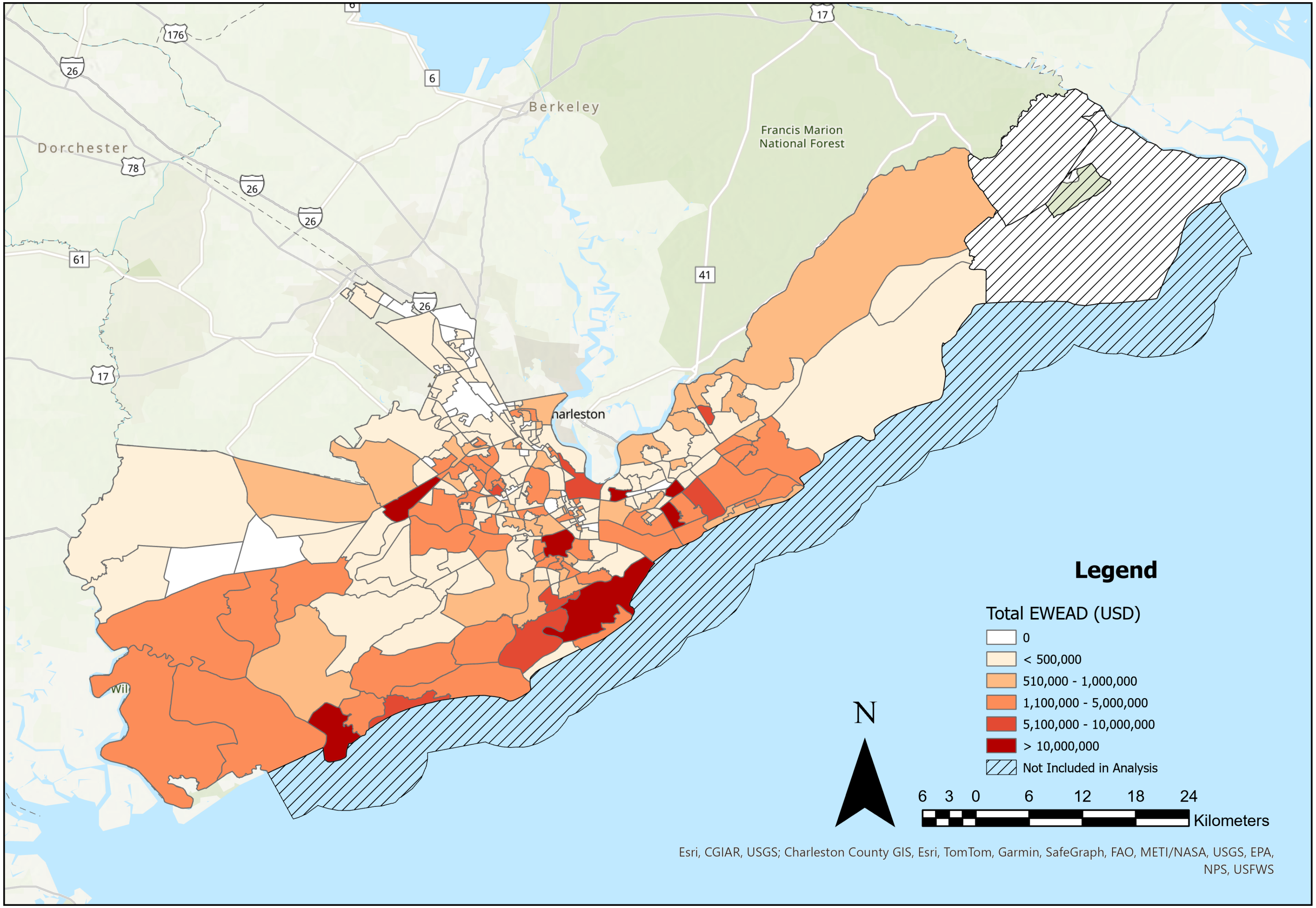


Figure C.11: EWEAD Result Map with disaggregated method for risk scenario

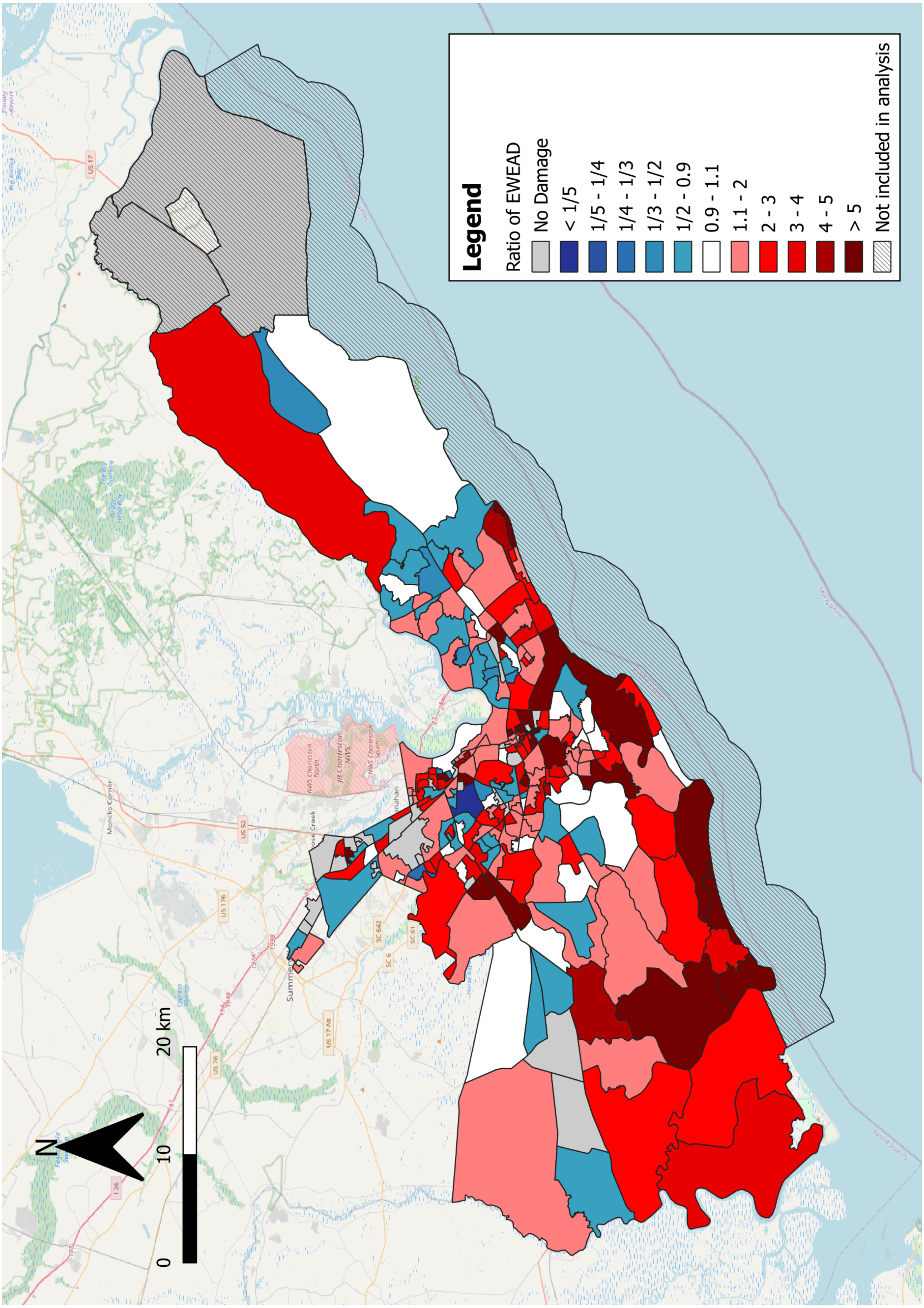


Figure C.12: Map of ratio EWEAD between methods for risk scenario