Efficient Computational Evaluation of Vulnerability, Resilience, and Sustainability of Social-Environmental Systems in the City of Nashville

Bowen He

Dissertation Proposal in partial fulfillment of the requirements for the degree of

Doctor of Philosophy
In
Environmental Engineering

Department of Civil and Environmental Engineering
Vanderbilt University
Nashville, TN

Sep. 30, 2021

Doctoral Committee:
Dr. Jonathan Gilligan
Dr. Janey Camp
Dr. Mark Abkowitz
Dr. Hiba Baroud
Dr. J.B. Ruhl
Dr. Arleen Hill

Contents

Background

1. Problem Statement

Natural disasters such as floods, cyclones, and storms have long been part of the heritage (COAG, 2011). More and more efforts have been made to mitigate natural disasters. However, over the past few decades, the frequency and severity of climate-related hazardous events have progressively increased (Leaning and Guha-Sapir, 2013). For example, flooding has been the most ubiquitous and costly natural hazard in the U.S. with an approximately $17 billion loss annually between 2010 and 2018 based on the recent testimony from Federal Emergency Management Agency (FEMA) representative. The main causes of flood are climatic changes, changes in land use, and other anthropogenic interventions that include urban growth, deforestation, etc. Therefore, building and enhancing community resilience to these disasters and improving natural hazard management strategies are urgently needed to reduce losses in the future and minimize the reduction in our life's well-being.

For years, voluntary property acquisitions ("home buyouts") are playing an increasingly important policy role after flood disasters (Zavar 2015). Through buyout programs, owners of severely damaged properties are offered pre-flood fair market values for their properties and given a chance to move from difficult circumstances (Fraser et al., 2003). When acquired parcels are permanently converted to green space or wetlands, they enhance hazard mitigation by eliminating future property losses, reduce threats to residents and provide environmental benefits through expanded ecological habitat, flood storage and conveyance, and recreational opportunities. However, due to limited funds and criteria, this mitigation effort can often lead to a checkerboard effect in neighborhoods that have experienced flooding. When enough residents are bought out to leave the community, the social fabric of the community can break down, the tax base for supporting community schools and maintaining public infrastructure can also collapse. In rural communities, the side effect of the loss of community members can be more significant than in urban areas. In general, the adverse impact of the property acquisitions effort is seldomly investigated.

Previous studies only evaluated the effects of the home buyout program from an individual or household perspective. For instance, Baker et al. (2018) did a survey to gather information of the home buyout participated individual's experience with the acquisition process that was implemented in their community after Hurricane Sandy in 2012. McGhee et al., (2020) quantified change-in-exposure to coastal flood hazards and social vulnerability among households that participated in a buyout program in New York State following Hurricane Sandy. They found that 99% of the household studied relocated to an area of higher social vulnerability and over 20% relocated to an area exposed to coastal flood hazards, suggesting that significant uncertainty remains regarding the extent to which buyout programs reduce household vulnerability (McGhee et al., 2020).

Studies have also found that past voluntary buyout programs generated resident feelings of coercion, degradation of trust, and loss of attachment to place (Fraser et al., 2003). However, to my knowledge, no research has been conducted on evaluating the home buyout program's negative effects on community social vulnerability status. Besides, the transparency in the current environmental policy-making process is inferior as the public has no idea how the community's social vulnerability is measured and why the proposed policy like home buyout would bring benefits to their life's will-being. Two overarching research questions that motivate this research

project are: for the current post-hazard recovery policy decision such as home buyout, is it a cost-effective way to reduce future flood damage as well as decreasing the community's social vulnerability? Can efforts be made to improve the transparency and flexibility in the policy-making process so that the public can have a better understanding of how the natural hazard mitigation decision like home buyout is made by policymakers?

2. Proposed Research

To better evaluate post-hazard policies such as home buyout program and their potential effects on community's social vulnerability status, several research gaps still linger: (1) lack of reliable spatial disaggregation approach to implementing social vulnerability analysis in any spatial resolution of interest, (2) lack of a comprehensive assessment framework that can evaluate the validity and the reliability of the community social vulnerability index coupled with dasymetric mapping technique, (3) lack of a strategy to increase the transparency and flexibility in the environmental policy-making process between policymakers and public.

Thus, I propose a dissertation research project that will contribute to the state-of-the-art by addressing these three research gaps mentioned above. The proposed research consists of 3 main chapters. Chapter 1 involves a critical review of spatial data disaggregation techniques that utilize deterministic, logic-based functions to reapportion data to spatial scales consistent with phenomena of interest. The purpose of using spatial data disaggregation techniques such as intelligent dasymetric mapping (IDM) is to address some confounding effects resulting from conducting analyses at improper spatial scales. Specifically, a cross-validation incorporated regularized hierarchical Bayesian multiple linear regression algorithm that uses National Land Cover Data (NLCD) as ancillary information will be developed and elaborated. Chapter 1 contributes to solving the Modifiable Areal Unit Problem (MAUP) in the context of social vulnerability index (SoVI) analysis.

In chapter 2, a review of uncertainty and global sensitivity analysis that utilizes the variance-based approach to identify the most influential uncertain input factors in the modeling process will be conducted. Applications of the uncertainty and global sensitivity analysis will be implemented to evaluate the robustness of the SoVI model coupled with the dasymetric mapping technique developed in Chapter 1 to understand uncertainty distribution caused by input factors within the model and how the dasymetric mapping technique affects the SoVI model's uncertainty and sensitivity distribution. Chapter 2 will contribute to providing useful insights during the process of building composite indicators. Combined with the proposed cross-validation incorporated regularized hierarchical Bayesian spatial linear regression algorithm developed in chapter 1, the use of uncertainty and global sensitivity analysis will increase transparency and confidence, enabling the policy inference process more defensible in any customized spatial scale regardless of indicators' original spatial scale.

Following in chapter 3, to better address the challenge of increasing the transparency and flexibility in the environmental policy inferencing process, I will build a Rmarkdown report and an interactive web-based visualization and sustainability decision support tool using Rshiny app by performing Nashville metro city's greenhouse gas (GHG) emissions analysis as a case study. The detailed Rmarkdown report consists of Nashville's current GHG emissions and the evaluation of the impacts of many possible GHG emissions reduction actions that explain the models and calculations within the Nashville metro sustainability analysis. The interactive web-based visualization and decision support tool uses the Rshiny app to explore the potential opportunities

of various combinations of GHG emissions reduction actions. This tool that utilized, including Rmarkdown and Shiny, allows for direct integration of data, analysis, and text, and is expected to ensure flexibility, transparency, and adaptability in the evaluation of different possible environmental policy combinations.

The goal of this dissertation project is to investigate the following research questions:

- **Chapter 1 - Spatial Disaggregation Modeling Approach using Dasymetric Multilevel Regression Analysis**

    a) How much information from land cover ancillary data should be extracted to interpolate population distribution to avoid overfitting?
    b) Does the arbitrary nature of administrative boundaries such as census tract and census block exhibit hierarchical structure in terms of population distribution characteristics?
    c) How many advancements the proposed spatial multilevel autoregressive dasymetric model will result in terms of prediction accuracy compared to the traditional ordinary least squares model (OLS)?

- **Chapter 2 - Global Sensitivity and Uncertainty Analysis on SoVI coupled with the Dasymetric Mapping Technique**

    a) How much uncertainty is associated with the SoVI model coupled with the dasymetric mapping technique developed in chapter 1?
    b) What is the relationship between SoVI model coupled with the dasymetric mapping technique and uncertainty?
    c) Which modeling decision contributes the most to variability in SoVI model coupled with the dasymetric mapping technique? Does the most influential modeling decision in the traditional census tract scale SoVI model differ from the SoVI model coupled with the dasymetric mapping technique?

- **Chapter 3 - Towards more transparency and flexibility in environmental policy and decision-making process: an interactive Decision Support Tool using RShiny app**

    a) Can the proposed interactive Rshiny decision support tool aid in identifying potential opportunities of various combinations of GHG emissions reduction policies?

Literature Review

1. Community Resiliency and Vulnerability Indices

Community resilience is defined as the ability of groups or communities to cope with external stresses and disturbances because of social, political, and environmental change (Adger, 2000). The increasing risk of natural hazards combined with a lack of community preparedness has revealed the necessity for effective quantitative measurement for assessing and improving a community's resilience (Johansen et al., 2017). Indices are an important approach to evaluate a community's resilience ability to withstand, and recover from all types of natural hazards, to compare across different communities, and to track any changes that occur in a community's resilience ability from a perspective of time (Johansen et al., 2017). Numerous studies have made efforts on producing a quantitative approach to measure community resilience to help them better manage themselves during the natural disaster (CARRI, 2013; Arup International Development, 2011; Arup, 2014; Sempier et al., 2010; EPA, 2014; FEMA, 2008; Plyer, 2013; OSSPAC, 2013). For example, the Community and Regional Resilience Institute (CARRI) Community Resilience System adopts a computer program to evaluate a community resilience for various disruptions based on quantifying the community's functional capacity (CARRI, 2013). City Resilience Index is invented as a framework of 12 indicators that define resilient city indicators segmented into people, place, organization, and knowledge (Arup, 2014). Coastal Resilience Index applies a community self-assessment strategy for a bad storm based on the historical records and summarizes results in resilience indices with low, medium, and high for each evaluated sector (Sempier et al., 2010). The New Orleans Index uses economic growth, inclusion, quality of life, and sustainability indicators to track the recovery of New Orleans neighborhoods since Hurricane Katrina in 2007 (Plyer et al., 2013). While these resilience measurement approaches provide valuable insights to aid the process of preparedness and recovery, limitations such as the specificity for a certain geographic area or a particular hazard, lack of an explicit quantitative outcome can prevent their generalized application across all communities in the country (Johansen et al., 2017). Thus, more work needs to be done to improve the breadth, utility as well as scientific merit in the community resilience index development process (Johansen et al., 2017).

Community vulnerability is defined in terms of the characteristics of a community that affect its capacity to anticipate, confront, and recover from the effects of a natural disaster (Flanagan et al., 2018). Examples of factors that might affect a community's social vulnerability status include its socioeconomic condition, gender composition, race and ethnicity composition, family structure, education, and medical services, etc. (Cutter et al., 2003). Many indices related to vulnerability have been created, with some of the more prominent ones featuring the Social Vulnerability Index (SoVI) to natural hazards (Cutter et al., 2003), the Social Vulnerability Index (SVI) for disaster management (Flanagan et al., 2011), the Environmental Vulnerability Index (EVI) (Kaly et al., 2014), the Coastal City Flood Vulnerability Index (CCFVI) (Balica et al.,2012), and the Human Development Index (HDI) (UNDP, 2016). Cutter et al. (2003) constructed an index of social vulnerability to environmental hazards using county-level socioeconomic and demographic data for the U.S. based on 1990 data. In their SoVI, the original 42 variables were reduced to 11 independent components that accounted for 76 percent of the variance using a principal component analysis (PCA) (Cutter et al., 2003). These components were placed in an additive model to compute a summary score for each county-the Social Vulnerability Index (Cutter et al., 2003). Kaly et al. (2014) constructed an index to describe the vulnerability of the

environment because they believe that human and environmental systems are dependent on one another so that risks to the environment of a state will eventually translate into risks to humans. The EVI was constructed based on a theoretical framework that identified three aspects of environment vulnerability: risks to the environment, the innate ability of the environment to cope with the risks and ecosystem integrity (Kaly et al., 2014). These three aspects correspond to three sub-indices, which are the Risk Exposure sub-Index, Intrinsic Resilience sub-Index, and Environmental Degradation sub-Index, respectively (Kaly et al., 2014). The EVI was calculated as a weighted average of scores allocated in the range of 0-7 derived from a total of 57 original indicators (Kaly et al., 2014). Balica et al. (2009, 2012) developed a Coastal City Flood Vulnerability Index (CCFVI) based on exposure, susceptibility, and resilience to coastal flooding. The CCFVI was constructed by an additive model that consists of three main component sub-indices: hydro-geological component sub-index, social and economic component sub-index, and politico-administrative component sub-index (Balica et al., 2009, 2012). Each component index is calculated based on selected indicators that are associated with that component (Balica et al., 2009, 2012). The final CCFVI gives a number from 0 to1 to indicate comparatively low or high coastal flood vulnerability to coastal cities within a future climate change context (Balica et al., 2009, 2012). Similar to resilience indices, while these community vulnerability indices provide valuable measurement tools that can reflect a community's social vulnerability status from multi-dimensional aspects such as social system, economic system, and environmental system, confusions still exist in the indicators selection process, challenges still exist in the reduction process regarding redundancy within and across indicators in terms of what to measure and how to measure it, and inconsistency still presents in the systems of interest where some indices focus largely on one aspect of community resilience as opposed to other aspects such as economic or environmental aspects (Gillespie-Marthaler et al., 2019). Thus, recognizing the need for consolidation of indicators within a consistent structure that facilitates identification and selection for use in evaluation processes, developing a consistent framework to reduce the redundancy across selected indicators are the two important directions of current community vulnerability research.

For most of these community resilience and vulnerability indices, the construction process begins with the identification of capital systems defined by a theory that are pertinent to the evaluation context. Examples of these capital systems could include categories related to social, environmental, and economic resources as foundational demarcation (Gillespie-Marthaler et al., 2019). Indicators that are related to community vulnerability or resilience are selected to compose the identified capital systems. Examples of indicators include voter percent in presidential election for the social theme (Sherrieb et al., 2010), percent of land under agriculture including plantation/forestry for the environmental theme (Kaly et al., 2014), and community's per capita income for the economic theme (Cutter et al. 2003). By means of surveys, remote sensing, and open-accessed data, these indicators are populated with each other to generate an index to represent a community's vulnerability or resilience status.

Although the construction process of these indices is relatively similar, there are myriad options at each step in the process without a consensus framework for identifying capital systems, selecting indicators, or acquiring data. As a result, this creates considerable confusion and redundancies between studies (Gillespie-Marthaler et al., 2019). For instance, some studies identify three capital systems (Pendall et al., 2010), others identify four (Balica et al., 2009; Norris et al., 2008; Ebisudani and Tokai, 2017) or even five (Cutter, 2016; Shaw et al., 2010; Tapia et al., 2017; Yoon et al, 2016). Even if the number of categories is the same, their compositions vary

significantly. Cutter (2016) uncovered those 10 indicators that appear in 40% of studies, indicating that great potential exists in reducing complexity and redundancy in this research field. Gillespie-Marthaler et al. (2019) developed a classification scheme and organizational structure to aid in the identification, selection, and application of indicators within a dynamic assessment framework. A nonduplicative set with a number of over 550 community sustainable resilience indicators and metrics are identified, where similar issues of specification and redundancy are still present (Gillespie-Marthaler et al., 2019).

Another great challenge in the index construction process lies in the correlation analyses to address the redundancy and categorization issue of selected indicators. For a typical index, there are usually more than 20 relevant indicators that are selected to be incorporated in the first place. Principal Component Analysis (PCA) is one of the most popularized approaches to aid in indicator selection and to avoid redundancy of variables in some composite indices' construction processes (Cutter et al., 2003; Cutter et al., 2008; Sherrieb et al., 2010). One beneficial point of PCA is that the component loadings associated with each primary component calculated by varimax rotation can provide informative categorical measures for those primary components that are basically linear combinations of original indicators (Cutter et al., 2003). Thus, decision-makers can identify the cardinal relationship between each primary component and the final index. However, based on the results of Cronbach's alpha tests (Western et al., 2005) that are designed to see how well their indicators fit into their prior defined capital themes by Cutter et al. (2014) and Tapia et al. (2017), a corresponding PCA can provide a completely different grouping while each classification was sufficient in Cutter et al. (2014) study, suggesting an alternative classification may be more appropriate. Another important approach is Confirmatory Factor Analysis (CFA). In the Communities Advancing Resilience Toolkit (CART) that was developed by Pfeferbaum et al. (2013, 2015), Likert Scale surveys and CFA were applied to validate that whether their capital systems were evident in stakeholder responses. Shim and Kim (2015) applied a CFA methodology to operationalize the biophysical, built environment, and socioeconomic resilience dimensions for local jurisdictions in large urban metropolitan areas in South Korea, and found that mapping the factor scores of the dimensions revealed great spatial variations. Cui and Han (2019) conducted a study to cross-culturally adapt and validate the original version of the 10-Item Conjoint Community Resiliency Assessment Measurement (CCRAM-10) in China and the CFA was performed to test the evaluation instrument's practicability and applicability. In Bec et al. (2019) study that developed an 18-item index for measuring resilience in the context of applying resilience as an emerging change management approach to respond and manage economic structural change for sustainable regional development, they used the CFA methodology to assess the reliability of the measurements and confirmed the extent to which the measurements were consistent with the Principal Component Factor (PCF) analysis.

2. Global Sensitivity and Uncertainty Analysis on Community Vulnerability and Resilience Indices

The usefulness of any model output is dependent on accuracy and reliability of its output. Nonetheless, since all models are eventually a form of abstraction of reality, not only the precise input data are rare in the case, but also the modeling process is subject to imprecision, leading to imperfect model output. As a result, the final model product is always associated with certain level of uncertainties and imprecisions which need to be assessed, interpreted, and visualized.

Sensitivity and uncertainty analysis are great tools to investigate the imprecisions of the model outputs for user's to be more confident when implementing activities associated with model's results. The difference between the two approaches lie in that uncertainty analysis only evaluates and represents the uncertainty in model outputs that derives from uncertainty in inputs, while sensitivity analysis focuses on evaluating the contributions of the uncertain inputs to the total uncertainty in analysis outcomes.

Uncertainty analysis (UA) is an important process to assess the entire set of possible outcomes associated with their probabilities of occurrence. The uncertainty of model output values results from modest changes in model input values as well as different model configuration. The goal of uncertainty analysis in models of complex systems is to produce output metrics with a greater degree of precision, transparency, and credibility, with an underlying aim of improving user's confidences in implementing activities associated with model's output. Uncertainty performance has been widely studied in model predictions of sea level rise (Haasnoot et al., 2020), hurricane paths (Cox et al., 2013), and communities' social vulnerability (Tate, 2013). Two general forms of uncertainty have been well understood: aleatoric and epistemic uncertainty. Aleatoric uncertainty occurs because of the heterogeneity or the intrinsic model randomness. Epistemic uncertainty arises from things that cannot be known but could potentially be measured from the limited accuracy and precision of our measurement (Jakeman, Eldred, and Xiu, 2010). For example, in terms of social index research, aleatoric uncertainty affects the input data used for social indexes, but epistemic uncertainty accompanies every stage of index construction process that can include conceptual framework development, selection of indicators, collection of data, consideration of measurement error, data transformation, weighting, and aggregation (Tate, 2013). As modeling decisions during the development of the index proceed, epistemic uncertainty associated with each step propagates and potentially interacts with that from previous steps (Tate, 2013). Uncertainty and sensitivity analysis are usually combined to quantitatively validate the social index where uncertainty analysis help evaluate the robustness of index values, and sensitivity analysis decomposes the uncertainty to determine which index construction decisions have the greatest influence on the output rank variability.

Different from the uncertainty analysis, sensitivity analysis (SA) attempts to determine the change in model output values that sources from modest changes in model input value. While the context where the sensitivity analysis is conducted could be complex, the term SA often refers to a 'what-if' analysis where the model parameterization or the process configuration are varied one at a time (Pianosi et al., 2016). SA also tells us about how the uncertainties (aleatoric and epistemic) in the independent variables affect the accuracy of our model's predictions of the dependent variables. Sensitivity analysis has been widely studied in human-environmental models such as weather and climate forecasts and simulations (Stephenson and Doblas-Reyes, 2000; Collins et al., 2012), sea level rise (Anthoff et al., 2006), projection of hurricane losses (Iman et al., 2005), evaluation of river water quality (Van Griensven et al., 2002), multizone air flow evaluation (Firrantello et al., 2007) and communities' social vulnerability (Schmidtlein et al, 2008). Besides, some emerging trends in the sensitivity analysis have been found that the application of SA can be used to analyze the impact of non-numerical uncertain factors like model spatial resolution or structure (Baroni and Tarantola, 2014). In this proposal, I mainly focus on reviewing the uncertainty analysis (Tate, 2013) and sensitivity analysis (Schmidtlein et al, 2008) associated with the SoVI developed by Cutter et al. (2003).

Schmidtlein et al. (2008) examined the sensitivity of quantitative features underlying the SoVI approach (Cutter et al., 2003) to changes in its construction, the scale at which it is applied,

the set of variables used, and to various geographic contexts. Specifically, multiple aggregation levels in the State of South Carolina and different subsets of the original variables were used to determine the impact of scalar and variable changes on SoVI construction process. For example, to determine the impact of changing the level of aggregation on the social vulnerability analysis, they constructed SoVI and applied the principal component analysis (PCA) on three different spatial scales: the county level that was original SoVI scale adopted in Cutter et al. (2003), census tract level scale, and a manually created intermediate level of aggregation (Schmidtlein et al., 2008). They uncovered that as the level of aggregation at which the principal component analysis (PCA) was conducted decreased, the variance explained per component also decreased, and the number of components selected using the Kaiser criterion thus increased, echoing the result found by Clark and Avery (1976). This is because with increasing level of aggregation, more and more spatial frequencies may be lost, and the same amount of important information can be modelled by fewer numbers of independent variables. Besides, they also tested the sensitivity of the algorithm to changes in construction and determined if that sensitivity was constant in various geographic contexts (Schmidtlein et al., 2008). In terms of SoVI construction algorithm sensitivity analysis, they first identified three categories: PCA component selection, PCA rotation, and the weighting scheme used to combine the components to create the final index (Schmidtlein et al., 2008). Within each of the three different categories for index construction, several options were considered. For instance, in terms of PCA component selection, they considered Kaiser criterion, percentage variance explained, Horn's parallel analysis as different methods (Schmidtlein et al., 2008). For PCA rotation methods, they accounted 4 rotation strategies which are unrotated solution, varimax rotation, quartimax rotation, and promax rotation (Schmidtlein et al., 2008). For weighting schemes, three approaches were considered: sum the component scores, first component only, and weighted sum using explainable variance from PCA to weigh each component (Schmidtlein et al., 2008). Factorial analysis with partial ("Type III") sums of squares approach was conducted to assess the impact of changes in the algorithm construction on the index values and found that the algorithm is robust to minor changes in variable composition and scale but is sensitive to changes in its quantitative construction such as weighting scheme (Schmidtlein et al., 2008). Their sensitivity analysis plays a critical role in understanding the impacts of changes in index construction as well as the scale on the final index representation and increasing users' confidence in metrics designed to represent the extremely complex phenomenon of social vulnerability (Schmidtlein et al., 2008).

Tate (2012, 2013) investigated the uncertainty associated with the methods of SoVI construction process including decisions related to indicator selection, scale of analysis, measurement error, data transformation, normalization, and weighting. Each of these stages was imbued with uncertainty due to choices made by the index developer (Tate, 2013) to answer the research questions of how much uncertainty is associated with the SoVI, how does the index perform when alternative configurations are considered, and what is the relationship between vulnerability and uncertainty (Tate, 2013). For example, in terms of data transformation method, options of raw counts (none), population (percentage), and density (area) were applied (Tate, 2013). For areas with low uncertainty, decision makers can proceed to implement vulnerability reduction strategies with a greater degree of confidence since the index representation is reliable (Tate, 2013). His study applied Monte Carlo-based uncertainty analysis to assess and visualize uncertainty for a hierarchical SoVI (Tate, 2013). During each run in the simulation, a single option within each stage of index development is selected and used to compute a social vulnerability index. The vulnerability ranks from that run are saved and the process is repeated 7,168 times (Tate, 2013).

In the uncertainty measurement and representation stage, three statistics of particular use to measure the uncertainty magnitude are the confidence levels (CIs), the median to evaluate bias, and the coefficient of variation (CV) to assess the index precision. They found that for areas with higher vulnerability, there tends to be greater index uncertainty, suggesting a potential alternative use of the index not as an approach to identify high-vulnerability areas but as a screening too to eliminate low-vulnerability areas from consideration (Tate, 2013). They also suggested that it is the weighting stage that is the key driver of uncertainty for the hierarchical model (Tate, 2013). Besides, they also concluded that the scale of analysis might not be an uncertain factor for an index explicitly designed for use at a particular administrative scale (Tate, 2013).

3. Spatial Disaggregation Approach: The Dasymetric Mapping Analysis

The spatial scale and resolution of interests have been verified by many studies to play a critical role in evaluating the spatial distribution of natural hazards' risks and the effectiveness of the mitigation efforts when examining the intersection between natural hazards, mitigation efforts, and community resilience in a multi-level spatial scales and extents (Chakraborty, 2011; Mennis, 2003). Different interpretations of intersection and overlap of spatial units with hazard places have been verified to affect the results of hazard risk spatial analysis, leading to both overestimation and underestimation of at-risk zones, an issue that has been widely recognized as the Modifiable Areal Unit Problem (MAUP) (Maantay et al., 2007; Mennis, 2002). Especially for the social vulnerability, community resilience, and environmental justice study that heavily rely on the census data from the American Census Survey (ACS), their spatial interpretation of the demographic data in the census tracts (CTs) and census blocks (CBs) are not aligned well with the spatial scale of the hazard of interest such as the floodplains and superfund site zones. In addition, the demographic variable such as population, single-person household in a census tract, or a census block spatial is usually too coarse to evaluate the effectiveness of the hazard mitigation efforts. For instance, the home buyout program is developed by the Federal Emergency Management Agency (FEMA) as a flood hazard mitigation tool to buy out those flood-affected properties in the floodplain and evacuate the flood-affected population. Those buyout activities usually occur in a property or building scale which is much finer than the census tract or block spatial scale. Thus, the spatial scale misalignment problem that occurs in the interface between the real natural hazard and the census boundary zone remains a huge challenge in the current community resilience and natural disaster management research field.

Dasymetric mapping approach is one of the cartographic techniques that recently experienced a renewed interest because of the rapid progress in computation algorithm, Geographic Information System (GIS), and remote sensing technologies (Mennis, 2009; Petrov, 2012). Recently, new development of new approaches regarding using innovative ancillary data (e.g. tax parcel data (TP), building footprint data (BDF), and night light data (NTL)) as well as calculation process has received much attentions since they provide a way to reliably disaggregate socio-demographic information to a finer scale (e.g. tax parcel or building scale) which is more suitable for environmental justice analyses and hazard mitigation efforts evaluation (Chakraborty, 2011; Bozheva et al., 2005; Eicher and Brewer, 2001; Maantay et al., 2007; Mennis, 2003; Holt and Lu, 2011; Mennis and Hultgren, 2006; Wu et al., 2005). For example, Mennis and Hultgren (2006) invented an intelligent dasymetric mapping (IDM) technique that combines an analyst's subjective knowledge with an empirical sampling approach to parameterize the relationship between the ancillary data (e.g., land cover land use data from National Land Cover Database

(NLCD)) and underlying statistical surface for purposes of reapportionment. They also identified the superiority of IDM compared to the traditional areal weighting and 'binary' dasymetric mapping approaches in terms of estimation accuracy (Mennis and Hultgren, 2006). The IDM was widely adopted in the environmental justice and assessment field. For instance, Giordano and Cheever (2010) used IDM to identify communities at risk from hazardous waste generation in San Antonio, Texas, and uncovered that people living near the generators are more often Black, non-Hispanic, or Hispanic, and more likely to live below the poverty level and not own their home. In conclusion, it is the IDM that helped them find that minorities, the poor, and non-homeowners have a higher potential for exposure than the general population to large-quantity hazardous waste generation in Bexar County, Texas (Giordano and Cheever, 2010).

Another important dasymetric mapping technique is the Cadastral-based Expert Dasymetric System (CEDS) (Maantay et al., 2007). Rather than using ancillary information in a uniform spatial resolution such as the land cover data from NLCD that is a 30m resolution, Maantay et al. (2007) developed the CEDS that uses non-uniformly distributed tax parcel information (e.g. property value, property type, and land use information) to delineate the heterogeneity spatial distribution of demographic data such as population in more urban areas such as the New York City (NYC). They conducted a case study of asthma hospitalizations in the Bronx, NYC to show the impact that a more accurate estimation of population distribution that the CEDS can provide on the current environmental justice and health disparities evaluation (Maantay et al., 2007). Nelson et al. (2015) developed a hybrid method for creating a social vulnerability index (SoVI) at a tax parcel level by utilizing supplementary information about tax parcels to link CEDS technique and the established social vulnerability indexing method. It is the CEDS that resolve the problem regarding the significant underestimation of vulnerable populations as environmental hazards typically occur on a finer spatial scale than census units such as census tracts or block groups (Nelson et al., 2015).

In terms of using the nighttime light (NTL) as an ancillary data, Zhou et al. (2014) adopted the Defense Meteorological Satellite Program/Operational Linescan System (DMSP/OLS) nighttime stable light data (NTL) as an ancillary data and used a cluster-based method to estimate optimal thresholds of the urban area. Then, they used the optimal thresholds of the urban extent to map the dynamics of urban areas in an economic and timely manner for both U.S. and China. Anther novel dasymetric mapping approach includes using built-area and height data as ancillary data (Alahmadi et al., 2014), using high-resolution address point datasets to conduct the dasymetric mapping (Zandbergen, 2011), developing multi-layer multi-class dasymetric mapping to estimate population distribution (Su et al., 2010), using means of raster pixel maps to rapidly facilitate dasymetric-based population interpolation (Langford, 2007), applying the hybrid model with different ancillary data combination (e.g. land cover data combined with tax parcel data, land cover data combined with NTL) (Briggs et al., 2007; Jia and Gaughan, 2016), and using machine learning model like random forests combined with remotely-sensed and ancillary data to project a finer spatial scale of demographic information distribution (Stevens et al., 2015). Here, we focus on reviewing the IDM and CEDS techniques in this proposal.

The IDM is an 'intelligent' dasymetric mapping technique that can combine an analyst's domain knowledge with a data-driven methodology to specify the functional relationship of the ancillary land cover classes with the underlying statistical surface being mapped (Mennis and Hultgren, 2006). The data-driven part of IDM employs a flexible empirical sampling approach to acquire information on the census data densities of individual ancillary classes, and it uses the ratio of different class densities to redistribute the census data of interest like population to the target

spatial zone areas controlled by the ancillary data (Mennis and Hultgren, 2006). In their intelligent dasymetric sampling strategy, they developed three sampling methods: the 'containment' method, the 'centroid' method, and the 'percent cover' method (Mennis and Hultgren, 2006). The 'containment' method samples those source zones that are wholly contained within an individual ancillary class (Mennis and Hultgren, 2006). The 'centroid' method selects those source zones that have their centroids contained within an individual ancillary class (Mennis and Hultgren, 2006). The 'percent cover' method allows the user to set a threshold percentage value and then selects those source zones whose area of occupation of a single ancillary class is equal to or exceeds that threshold (Mennis and Hultgren, 2006). Particularly, for those ancillary classes that go unsampled during the sampling process, which can usually occur using the containment or the percent cover sampling method, the unsampled classes' densities are estimated using the 'refined' areal weighting (Mennis and Hultgren, 2006). The IDM Toolbox was developed for ArcGIS Pro and can be found at: https://github.com/USEPA/Dasymetric-Toolbox-ArcGISPro.

Cadastral-based dasymetric mapping systems (CEDS) is an important technique for mapping the census data of interest to a finer spatial scale in urban areas (Maantay et al., 2007). The CEDS method uses specific cadastral data, land-use filters, modeling by expert system routines, and validation against various census enumeration units and other data (Maantay et al., 2007). For example, the residential units (RU) and residual area (RA) can be used as population proxy units to derive populations within each tax parcel (Maantay et al., 2007). Within the lot-level data, the RU variable did not require additional processing; however, there were many instances of missing data values for the RA variable in the original data (Maantay et al., 2007). Thus, they created a new variable, adjusted residential area (ARA) to replace the missing RA variable value to be used as proxy units associated with census data of interest (Maantay et al., 2007). The expert system was designed to determine which proxy unit-number of residential units (RU) or adjusted residential area (ARA) accurately predicts the population distribution on a tract-by-tract basis by reaggregating the tax lot level population value to the census block group level (Maantay et al., 2007). By comparing the reaggregated census block group value derived from predicted population value in the tax lot level for both the RU- and ARA-based CEDS techniques and the true census block group value from the ACS source, the expert system would then select the superior proxy unit that has a smaller difference between the two data as the disaggregation technique (Maantay et al., 2007). In essence, it is the performance of the tract-level disaggregation that defines the proxy units used for each block group disaggregation, ultimately resulting in a final dasymetrically derived value individually tailored for each census tract (Maantay et al., 2007). However, the disadvantage of this method is that its applicability is restricted by the tax parcel data availability of the study area of interest.

4. Social Vulnerability Index (SoVI)

In Cutter et al. (2003), the county-level socioeconomic and demographic data were used to construct an index of social vulnerability to environmental hazards, called the Social Vulnerability Index (SoVI) for the United States based on 1990 data. Using a factor analytic approach, the original 42 variables were reduced to 11 independent factors that accounted for about 76 percent of the total variance. These factors were placed in an equal weight additive model to compute a summary score to inform the social vulnerability status of each county within the U.S.

In an updated version of SoVI 2010-14, it marks a change in the formulation of the SoVI metric from earlier versions. New directions in the theory and practice of vulnerability science

emphasize the constraints of family structure, language barriers, vehicle availability, medical disabilities, and healthcare access in the preparation for and response to natural disasters, thus necessitating the inclusion of such factors in SoVI. Table 1 below shows the updated list of variables of the current version of SoVI 2010-14. All variables have pycnophylactic property (mass or volume preserve property) that value allocated to any spatial resolution pixel can be summed up to obtain the original polygon counts except for median age, per capita income, median housing value, and median gross rent. All studies in this proposal are based on this updated 2010-14 SoVI.

The SoVI has been applied in many natural hazard evaluation studies as an important tool to evaluate a region's vulnerability in the context of natural hazards. For instance, Guillard-Gonçalves et al. (2015) applied SoVI to Greater Lisbon (Portugal) to delineate natural risk zones. Recently, Aksha et al. (2019) applied SoVI to analyze social vulnerability to natural hazards in Nepal. While these studies provide general information of social vulnerability of populations at different locations, the restrictions of using standard census areal units' data can lead to significant underestimation of vulnerable populations as environmental hazards as well as home buyout policies typically occur on a finer spatial scale than standard census units such as district boundary, causing the MAUP problem. Thus, looking for a reliable dasymetric mapping approach that can reliably scale the SoVI in any spatial resolution of interest is critical to inform a valid measurement of vulnerable populations and help devise precise natural hazard mitigation plan.

Table 1. List of SoVI 2010-2014 Variables (n = 30)

Variable	Description
QASIAN	Percent Asian
QBLACK	Percent Black
QHISP	Percent Hispanic
QNATAM	Percent Native American
QAGEDEP	Percent Population under 5 years or 65 and over
QFAM	Percent Children Living in 2-parent families
MEDAGE	Median Age
QSSBEN	Percent Households Receiving Social Security Benefits
QPOVTY	Percent Poverty
QRICH	Percent Rich
PERCAP	Per Capital Income
QESL	Percent Speaking English as a Second Language with Limited English Proficiency
QFEMALE	Percent Female
QFHH	Percent Female Headed Households
QNRRES	Nursing Home Residents Per Capital
HOSPTPC	Hospitals Per Capital
QNOHLTH	Percent population without health insurance
QED12LES	Percent with Less than 12th Grade Education
QCVLUN	Percent Civilian Unemployment
PPUNIT	People per Unit
QRENTER	Percent Renters
MDHSEVAL	Median Housing Value
MDGRENT	Median Gross Rent

QMOHO	Percent Mobile Homes
QEXTRCT	Percent Employment in Extractive Industries
QSERV	Percent Employment in Service Industry
QFEMLBR	Percent Female Participation in Labor Force
QNOAUTO	Percent of Housing
QUNOCCHU	Percent Unoccupied Housing Units
QMORTBRDN	Percent of all households spending more than 40% of their income on housing expenses

Research Design

Chapter 1: Spatial Disaggregation Modeling Approach using Dasymetric Multilevel Spatial Autoregressive Analysis

Perform a critical review of literatures in spatial disaggregation techniques, related dasymetric mapping approach and its application in natural disaster management and environmental justice field. Develop a cross-validation incorporated land-cover based multilevel spatial dasymetric autoregressive approach that can improve the current IDM's sub-population reallocation accuracy. (Completion by June 2022)

In chapter 1, I will focus on addressing the research gap associated with the lack of a reliable spatial disaggregation approach to implementing SoVI analysis that can mitigate the MAUP problem. Previous studies showed tremendous amounts of examples that adopted the dasymetric mapping technique to disaggregate the index indicators to the spatial resolution of interest (Nelson et al., 2015, Bian and Wilmot, 2015). The most popular two dasymetric mapping techniques used are NLCD-based IDM (Mennis and Hultgren, 2006) and Cadastral-based CEDS (Maantay et al., 2007).

Figure 1(b), (c), (d) display the preliminary results of applying these two popular dasymetric mapping approaches to disaggregate the SoVI-related indicators from their original spatial scale to two customized spatial scales so that the decisionmaker can obtain the social vulnerability information of the resolution of their interest.

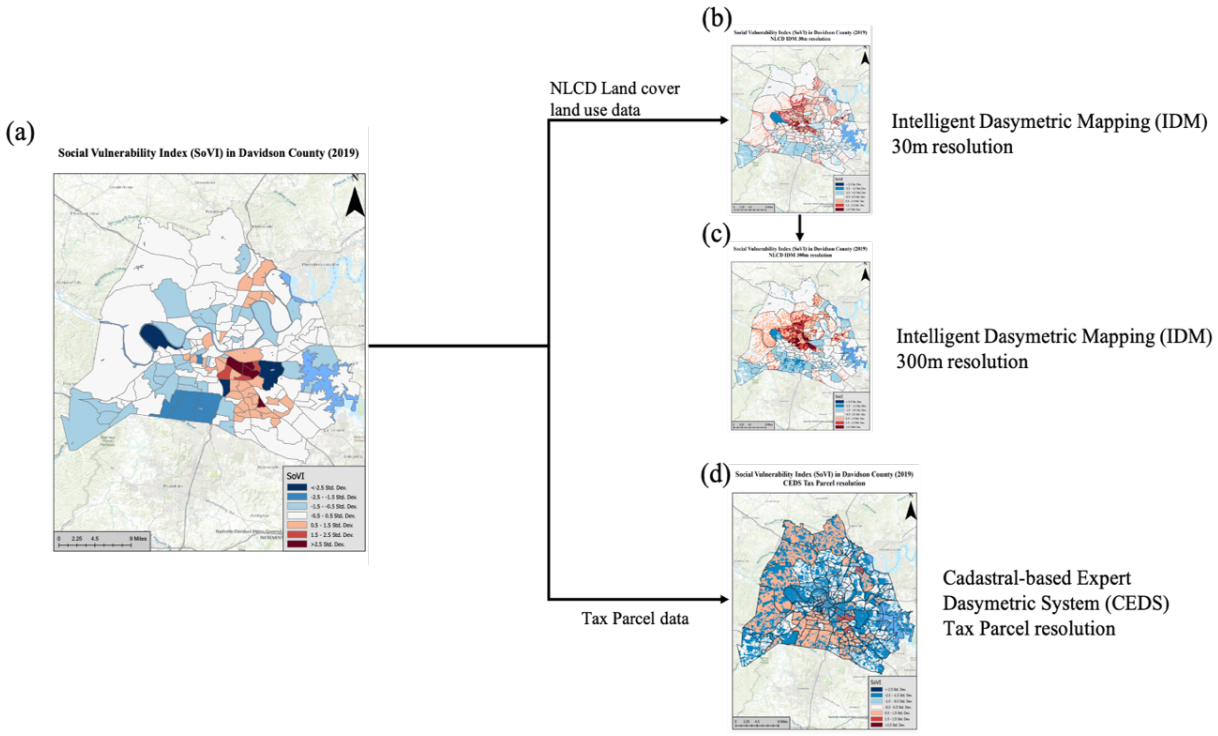

Figure 1 (a) SoVI in Davidson County using 2019 ACS data; (b) SoVI in 30m resolution derived from IDM using 2019 ACS data; (c) SoVI in 300m resolution derived from IDM using 2019 ACS data; (d) SoVI in tax parcel resolution derived from CEDS using 2019 ACS data.

However, several challenges are existing in these two popular dasymetric mapping methods. In terms of the CEDS, the biggest challenge lies in that the dasymetric mapping process is highly dependent on the fine resolution tax parcel data which is not available across the country. For IDM, although the ancillary land cover data is available across the country, the traditional dasymetric sampling method suffers from defective prediction accuracy. Here, a preliminary error analysis associated with the Intelligent Dasymetric sampling approach is conducted on two spatial scales: (1) census tract level using all 2016-year census tracts population data in the TN state; (2) census block group level using all 2016-year census block groups population data in the Davidson County. For census tract level evaluation, the TN's county population data is derived from the ACS, and then the IDM is applied to approximate the population distribution in a 30m resolution. Following, the IDM dasymetrically sampled population is reaggregated to the census tract level to compare with the TN's census tract population data derived from the ACS served as the benchmark. Similarly, for census block group level evaluation, Davidson County's census block group population is derived from the ACS, and then the IDM is applied to predict the population distribution in a 30m resolution. Following, the IDM dasymetrically sampled population is reaggregated to the census block group level to compare with Davidson County's census block group population data derived from the ACS served as the benchmark. The difference between the reaggregated dasymetrically sampled population value and the ACS-derived value is calculated for both spatial scales and the error analysis results are summarized in Figure 2. Figure 2(a) reveals that for census tract level evaluation, the raw population count error by IDM sampling can be as large as 20000. From an error percentage perspective, Figure 2(b) suggests that for some regions, the IDM is subjected to produce a population with 10 magnitudes larger than the ACS-derived

value. Similarly, for the census block group evaluation shown in Figure 2(c) and (d), IDM sampling also predicts a significantly larger population than the benchmark for some of the census block groups in Davidson County. In addition, the R^2 is calculated for the census block group level evaluation and the value is only 0.5, indicating that the traditional intelligent dasymetric sampling model can only explain 50% of the variances of the census block group population using the NLCD land cover information as ancillary data.

Besides, the IDM also suffers from a lack of understanding of its 'black-box' sampling parameter design. For IDM applications in the different study regions, the optimal sampling configuration and design are usually unknown, making the model sampling design inference extremely difficult. Thus, the model design and its prediction accuracy of the traditional IDM are to be revised and improved.

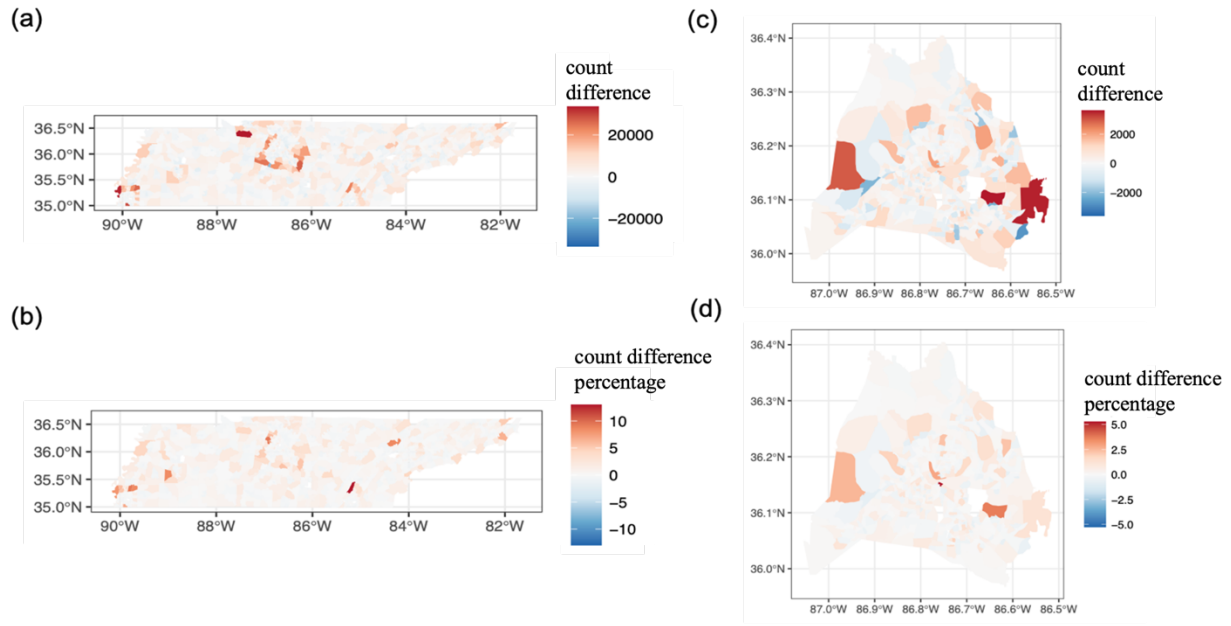

Figure 2. (a) Population count difference at the census tract level between reaggregated population derived from IDM and the census tract population from ACS using TN state data; (b) Population count difference percentage at the census tract level between reaggregated population derived from IDM and the census tract population from ACS using TN state data; (c) Population count difference at the census block group level between reaggregated population derived from IDM and the census block group population from ACS using Davidson County data; (d) Population count difference percentage at the census block group level between reaggregated population derived from IDM and the census block group population from ACS using the Davidson County data.

For the traditional IDM, the analyst has different options of sampling strategy to derive the relationship between the data density and any land cover class. For instance, the 'percent cover' sampling strategy allows the user to set a threshold percentage value and then selects those source zones whose area of occupation by a single land cover class is equal to or exceeds that threshold. However, the strategy suffers from the compromise between inadequate sampling and sampling bias. On one hand, when the threshold is set high, the sampled source zones that meet the criterion can sufficiently represent the relationship between the data density and the land cover type. However, the number of sampled source zones can be very low since few sources zones' areas of

occupation can be equal to or exceed that high threshold. This phenomenon is called inadequate sampling. On the other hand, when the threshold is low, there can be enough sampled source zones that meet the criterion. However, the quality of those sampled source zones that are used to derive the relationship between the population density and the land cover type can be low since many source zones can meet the criterion with a low threshold. This phenomenon is called sampling bias.

Another prediction error caused by the traditional IDM sampling process is global smoothing bias, or spatial non-stationarity. Important distinctions exist between land cover and land use information and lack of land use information in the traditional IDM sampling process is responsible for the error. For instance, for some census tracts filled with many public government buildings, commercial and industrial buildings, the land cover type is all identified as developed land areas, the same as the land where residential buildings locate. Nonetheless, the association coefficient between population density and the developed land areas in those census tracts is significantly lower than other census tracts filled with residential buildings because people don't reside in the government buildings or industrial buildings. The traditional IDM sampling process is not able to detect the variance between the association connection between population and government buildings as well as residential buildings, treating these buildings as the same based on global developed land areas because of the lack of land use information as input in the sampling process. As a result, the IDM sampling process easily overestimates the population for some census tracts shown in Figure 2(c).

Finally, the traditional IDM sampling process ignores the spatial autocorrelation information between census tracts, sampling them as spatially independent and identical data points. As the first law of geography by Waldo Tobler indicates that "everything is related to everything else, but near things are more related than distant things." (Tobler, 1970) Incorporating spatial effect in the dasymetric modeling will enable us to better understand the relationship between the ancillary data and population distribution.

Based on this, three levels of efforts will be made to improve the current dasymetric modeling in both the prediction and model inference perspectives. The first effort will be adopting cross-validation and regularized hierarchical linear regression algorithm to identify the optimal dasymetric linear regression model structure. The second effort will be incorporating a multilevel/hierarchical model structure with exchangeable (iid) random effect to account for variances between the association relationship between ancillary data and population distribution across census tracts. The third effort will be including spatial autocorrelation effects into the multilevel mixed-effect linear regression model to better understand how spatial effect affects the association relationship between ancillary data and population distribution across census tracts. The detailed elaboration of these three efforts is described below.

Effort I: *Cross-validated multiple linear regression with regularization scheme*

To quantitatively explore the relationship between each different land cover type and population density, I intend to implement a cross-validated multiple linear regression with regularization scheme including Ridge, Least Absolute Shrinkage and Selection Operator (Lasso), and Elastic Net linear regression analysis to overcome the issues of inadequate sampling, sampling bias, and sampling inconsistency in the conventional IDM dasymetric approach. Based on our knowledge, the developed land cover type with low, medium, and high intensity are confidently believed to relate to population density while open water and perennial ice/snow are believed not to relate to population density. However, we know little about the relationship between other land

cover types such as barren lands or pastures and the population density. Thus, I propose to use the regularized multiple linear regression which is a good method of fitting regression models by adding a tuning parameter λ and a regularization term (L1, L2, or a weighted combination of L1 and L2 norm) in the original least-squares cost function for the purpose of penalizing the size of the regression coefficients and thus to reach a balance between variance and bias.

For a dataset with relatively small data points, the abnormal data points can significantly alter the result and lead to inconsistencies. To address this issue, I propose to use the k-fold Cross-Validation (KFCV) method because the method tests the prediction accuracy on all k subsamples while systematically removing single fold subsamples from the full dataset. Combined with the regularized multiple linear regression analysis, this cross-validated regularized multiple linear regression analysis will guide to selection of the set of significant independent variables with the minimum KFCV mean error (CVME). After the significant independent variables set is identified, the final linear multiple regression model is to be fitted in a Bayesian context so that the whole range of inferential values and the posterior distribution associated with each model coefficient/parameter can be uncovered and assessed.

Here, an initial 10-fold cross-validated regularized multiple linear regression analysis was conducted on the 2010-year Davidson County's census block population data combined with 2008-year NLCD land cover information as ancillary data. Three different regularization schemes were tested at this stage: (1) LASSO whose regularization term (L1 regularization norm) penalizes absolute value of the regression coefficients ($\lambda \sum_{j=1}^{m} |\beta_j|$); (2) Ridge whose regularization term (L2 regularization norm) penalizes the square of the magnitude of the regression coefficients ($\lambda \sum_{j=1}^{m} \beta_j^2$); (3) Elastic Net regression whose penalization algorithm uses a weighted combination of LASSO and Ridge regularization term ($\lambda(\frac{1-\alpha}{2}\sum_{j=1}^{m} \beta_j^2 + \alpha \sum_{j=1}^{m} |\beta_j|)$, where α controls the weighting scheme between L1 and L2 norm).

For LASSO regression analysis, with the increasing of the turning parameter λ, the 10-fold CVME first decreases to reach the minimum value of 12977 at the optimal λ value of 11.22, and then increases (Figure 3(a)). Figure 3(b) displays that for LASSO regression, at the optimal λ value of 11.22 shown in Figure 3(a), only three land covers including developed open space, developed low intensity, and developed high intensity land cover predictors are found significant. In terms of Ridge regression analysis, the optimal turning parameter λ is identified at the value of 22.03 to produce the minimum 10-fold CVME at the value of 14083 (Figure 3(c)). Figure 3(d) reveals that for the Ridge regression model, at the optimal λ value of 22.03 suggested in Figure 3(c), four land covers including barren, developed low intensity, developed high intensity, and developed medium intensity are found significant. Following in the Elastic Net regression experiments, the optimal weighting scheme α is found at 0.4 that can produce the minimum 10-fold CVME at the value of 12173 along with the optimal λ value at the value of 6.95 (Figure 3(e)). According to the minimum CVME value as the standard, the Elastic Net regularization scheme has the smallest value of minimum CVME among the three regularization schemes, indicating that it is the optimal regularization algorithm to reach the balance between variance and bias.

The detailed information regarding selected significant land cover predictors and their associated optimal regression coefficients along with the calculated minimum CVME for three regularization schemes are summarized in the following Table 2.

Table 2. Summary results of significant regression coefficients and minimum CVME for each 10-fold cross-validated regularized multiple linear regression model

Lasso Regression		**Ridge Regression**		**Elastic Net Regression**	
$CVME_{min}$ = 12977.13		**$CVME_{min}$ = 14083.81**		**$CVME_{min}$ = 12173.1**	
Land Cover	**Coefficient**	**Land Cover**	**Coefficient**	**Land Cover**	**Coefficient**
Intercept	17.883	Intercept	17.206	Intercept	14.087
Developed open space	0.027	Barren	0.437	Developed open space	0.145
Developed low intensity	2.856	Developed low intensity	2.216	Developed low intensity	2.635
Developed high intensity	-0.265	Developed high intensity	-1.739	Developed high intensity	-1.762
		Developed medium intensity	1.556	Developed medium intensity	1.254

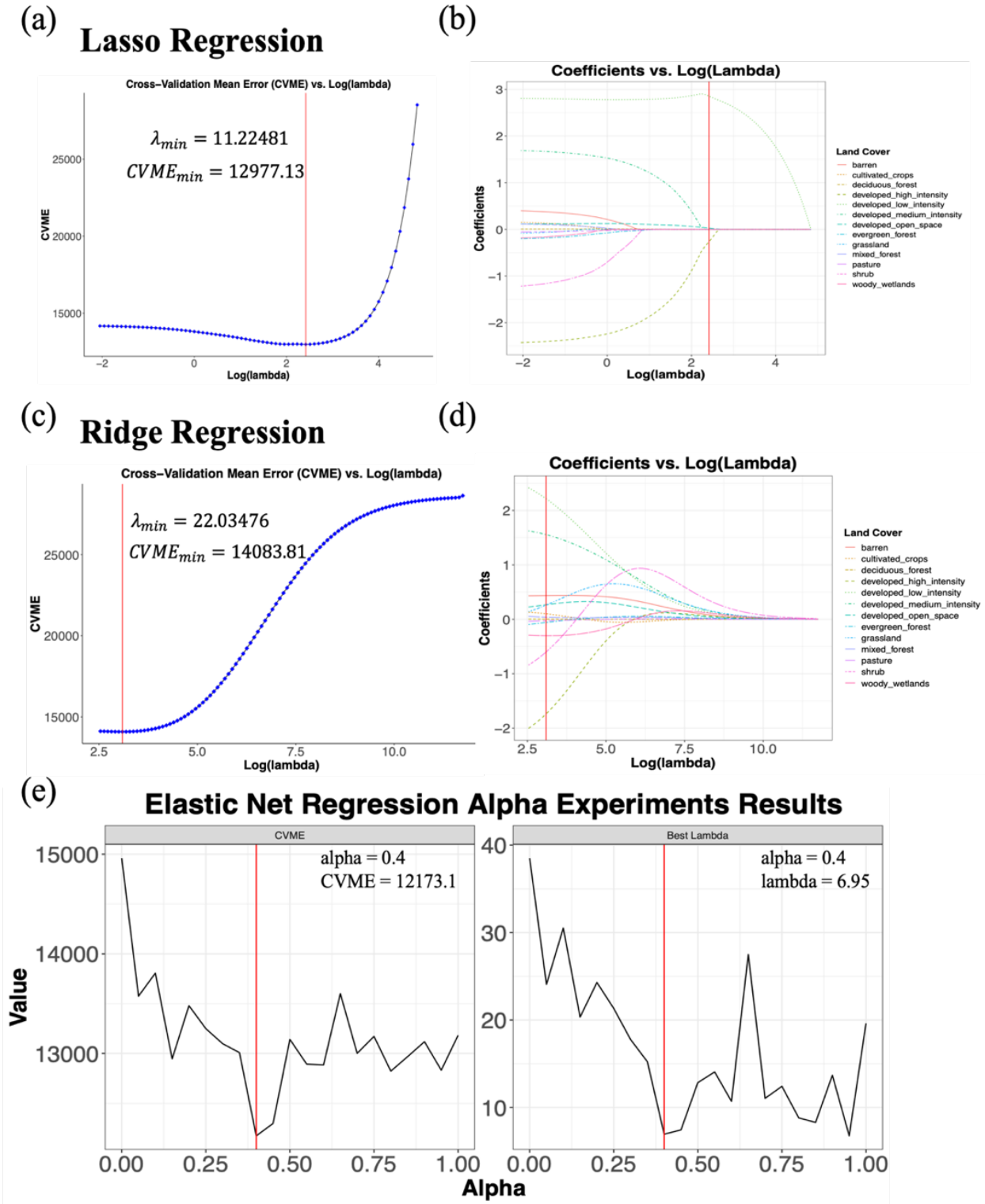

Figure 3. 10-fold cross-validated regularized multiple linear regression analysis results summary: (a) LASSO regression: the relationship between λ and CVME; (b) LASSO regression: the relationship between λ and regression coefficients; (c) Ridge regression: the relationship between λ and CVME; (d) Ridge regression: the relationship between λ and regression coefficients; (e)

Elastic Net regression experiments to find the optimal α and associated optimal λ in the weighted regularization term.

Effort II: *Multilevel/Hierarchical linear regression model structure with exchangeable (iid) random effect*

To mitigate the statistical error caused by global smoothing bias of the association coefficient between the land cover type and population density induced by the original dasymetric sampling, I intend to incorporate a multilevel hierarchical linear model structure with an exchangeable (iid) random effect so that the association coefficient between the land cover type and population density can vary across the census tracts.

$$Y_{ij} \sim N(\beta_{0[j]} + \boldsymbol{\beta_{1[j]}} \boldsymbol{X_{ij}},\ \sigma^2_{y_{ij}}), \text{ for } i = 1,\ldots,n \text{ (census blocks)} \tag{1}$$

$$\begin{pmatrix} \beta_{0j} \\ \beta_{1j} \end{pmatrix} \sim \mathrm{N}\left(\begin{pmatrix} \mu_{\beta_0} \\ \mu_{\beta_1} \end{pmatrix},\ \begin{pmatrix} \sigma^2_{\beta_0} & \rho\sigma_{\beta_0}\sigma_{\beta_1} \\ \rho\sigma_{\beta_0}\sigma_{\beta_1} & \sigma^2_{\beta_1} \end{pmatrix} \right),\ \text{for } j = 1,\ldots,J \text{ (census tracts)} \tag{2}$$

where $\beta_{0[j]}$ is an intercept term for census tract j that is modeled as gaussian distribution, $\boldsymbol{\beta_{1[j]}}$ is a vector of coefficients that describe the effects of each different land cover types on population for census tract j that is modeled as gaussian distribution, $\boldsymbol{X_{ij}}$ is a vector of covariates of land cover information that are identified in the previous cross-validated regularized hierarchical linear regression algorithm.

Effort III: *Area-specific spatial autocorrelation model coupled with exchangeable (iid) random effect*

I intend to include a spatial effect into the model that accounts for the spatial autocorrelation information at the census tract level. Spatial autocorrelation effects ($\boldsymbol{S_{[j]}}$) at the census tracts level used an intrinsic conditional autoregressive (iCAR) model coupled with an exchangeable (iid) random effect, also known as a Besag-York-Mollié (BYM) model. The addition of the spatial random effects can be interpreted as the addition of a random intercept term such that the latent field of the population mean is controlled by census tracts' spatial autocorrelation effects in addition to an independent and identically distribution (iid).

$$Y_{ij} \sim N(\beta_{0[j]} + \boldsymbol{\beta_{1[j]}} \boldsymbol{X_{ij}} + \boldsymbol{S_{[j]}},\ \sigma^2_{y_{ij}}), \text{ for } i = 1,\ldots, n \text{ (census blocks)} \tag{3}$$

where $\boldsymbol{S_{[j]}}$ represents both the spatially exchangeable random differences across the census tracts and spatial autocorrelation effects between neighboring census tracts:

$$s_j | s_{-j} \sim \text{Normal}\ (\mu_j + \frac{1}{Ne_j} \sum_{i=1}^{n} a_{ij}(s_i - \mu_i),\ \sigma_j^2) \tag{4}$$

$$\sigma_j^2 = \frac{\sigma_s^2}{Ne_i} \tag{5}$$

$Ne_i = \#\ Ne_i$, the number of neighbors of the i^{th} area

a_{ij} is 1 if areas *i* and *j* are neighbors and 0 otherwise (adjacency matrix, Figure 4)

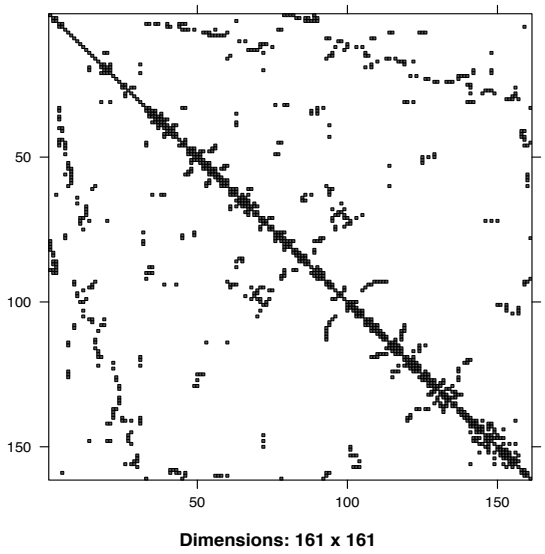

Figure 4. Adjacency matrix ***A*** for the Davidson County: rows and columns identify areas and squares identify neighbors.

Preliminary testbed model with synthetic population data

To better illustrate the idea of model structures described above, I create a synthetic population distribution data for each census block of the Davidson County using three model structures described above as a preliminary study to see whether the proposed model structures can find the designed model parameters and hyperparameters:

Model A (Fixed-effect linear regression): y_i ~ Normal ($\beta_0 + \beta_1 x, \sigma^2$)
Model A is a fixed-effect linear regression model with land cover predictors identified in the cross-validated regularized multiple linear regression algorithm. For the sake of simplicity, I only incorporate the developed open space as the only land cover predictor in this model. The synthetic population distribution data is created based on y_i ~ Normal ($\beta_0 + \beta_1 x, \sigma^2$).

Model B (Mixed-effect linear regression with exchangeable (iid) random effect): y_{ij} ~ Normal ($\beta_0 + \upsilon_{0j} + \beta_1 x, \sigma^2$)
Model B is a mixed-effect linear regression model with an exchangeable (iid) random effect. For the sake of simplicity, the random effect only exists in the intercept term and is governed by a gaussian distribution. The synthetic population distribution data is created based on y_{ij} ~ Normal ($\beta_0 + \upsilon_{0j} + \beta_1 x, \sigma^2$).

Model C (Mixed-effect linear regression with spatial effect coupled with an exchangeable (iid) random effect): y_{ij} ~ Normal ($\beta_0 + \upsilon_{0j} + s_j + \beta_1 x, \sigma^2$)
Model C is a mixed-effect linear regression with spatial effect coupled with exchangeable (iid) random effect. For the sake of simplicity, the random effects only exist in the intercept term. The

spatial exchangeable random effect v_{0j} is governed by a random distribution and the spatial random effect s_j is governed by the BYM model. It is difficult to generate synthetic population data since the intrinsic conditional autoregressive (iCAR) prior is conditional distribution, where the structure spatial random effects depend on its neighboring values. So, Model C uses the same synthetic population distribution data as Model B. Thus, identification of the hyperparameters of Model C is not expected.

Parameters and hyperparameters for both fixed and random effects of those three models described above are summarized below:

Fixed intercept term: β_0 = **43**
Fixed slope term: β_1 = **1.2**
Exchangeable random effect: v_{0j} ~ Normal (0, $\sigma_{v_0}^2$ = 1/ **0.25**)
$s_j|s_{-j}$ ~ Normal ($\mu_j + \frac{1}{Ne_j}\sum_{i=1}^{n} a_{ij}(s_i - \mu_i),\ \sigma_j^2$)
Variance of Gaussian observation: σ^2 = 1/**3**

Table 3. Posterior Bayes estimates for preliminary models using synthetic population distribution

	Model A (Fixed effect linear regression)	**Model B (mixed-effect linear regression with exchangeable random effect)**	**Model C (mixed-effect linear regression with spatial effect)**
Estimate (95% Bayes Credibility Intervals)			
β_{00}	43.005 (42.991.951, 43.019)	42.98 (42.66, 43.30)	42.98 (42.95, 43.01)
β_1	1.200 (1.168, 1.255)	1.200 (1.17, 1.26)	1.200 (1.168, 1.256)
Precision for Gaussian observation	3.03 (2.94, 3.14)	2.988 (2.812, 3.094)	2.996 (2.877, 3.095)
Precision for exchangeable random effect (iid)	None	0.247 (0.199, 0.305)	54.155 (39.597, 79.757)
Precision for spatial effect	None	None	0.113 (0.084, 0.165)
Marginal Log-Likelihood	-6602.51	-7150.35	-7075.64
WAIC	13154.05	13413.93	13414.67
DIC	13153.45	13415.83	13414.94
β_{00}: Intercept, β_1: developed open space fixed effect			

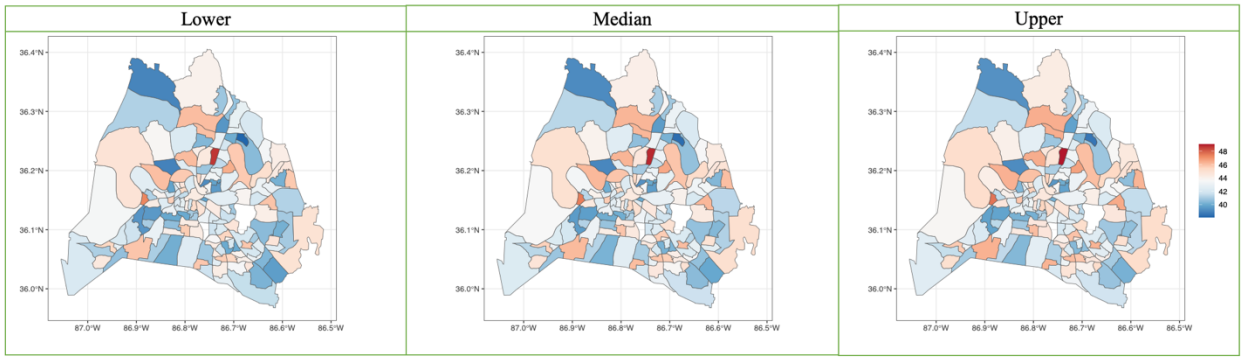

Figure 5. Posterior median of the random intercept by census tracts with upper and lower 95% credibility intervals for Model B: Mixed-effect linear regression with exchangeable (iid) random effect.

Table 3 shows that both Model A and Model B successfully identified the designed model parameters and hyperparameters. For Model C, the fixed effects are successfully identified, however, the random effects are identified differently from the designed parameters since the synthetic population in Model C is the same as Model B. Figure 5 displays that a multilevel model structure in the model can successfully identify the variances between census tracts, mitigating the issue of global smoothing caused by arbitrarily sampling.

Following the model derivative process, the model evaluation value for any given grid *i* in census tract *j* can be calculated based on the regression model:

$$y_{ij(c)} = f_j(x_{ij}) \tag{6}$$

where $f_j()$ is the mixed-effect multilevel linear regression model for the census tract *j*, x_{ij} is the land cover information for any given spatial resolution grid *i* inside the census tract *j*.

The reallocation value for that grid *i* in census tract *j* is calculated based on a reallocation model:

$$y_{ij(p)} = \frac{y_{ij(c)}}{\sum_{i=1}^{n} y_{ij(c)}} Y \tag{7}$$

where *n* is the total number of grids in the census tract *j*, *Y* is the total value from the source zone (census tract), and $y_{ij(p)}$ is the value that is predicted/reallocated to the grid *i* in the census tract *j*.

These works redesign the traditional IDM sampling process, suggesting that consideration of the hierarchical structure and spatial autocorrelation effect in dasymetric modeling can systematically improve the understanding between the ancillary data and the spatial data of interest. In the meantime, I hypothesize that this proposed multilevel Bayesian spatial modeling approach has the potential to improve the spatial reallocation accuracy and may lead to greater confidence in the identification of the spatial distribution of community vulnerability, resilience, and adaptive capacity indicators at any spatial resolution of interests.

Chapter 2: Global Sensitivity and Uncertainty Analysis on the SoVI Model coupled with Dasymetric Mapping Technique

Perform a global sensitivity and uncertainty analysis on Social Vulnerability Index (SoVI) in two spatial scales using Davidson County, Nashville as a case study: (1) census tracts scale that is derived from ACS data source, defined as a census tract scale context; (2) customized spatial scale that is derived from cross-validation incorporated regularized hierarchical Bayesian spatial linear regression algorithm developed in chapter 1, defined as a dasymetric mapping context. (Completion by August 2022)

The evaluation of uncertainty in models of complicated system has long been an area of research interest. Models are now understood to be with two general forms of uncertainty. The first is aleatoric, which occurs due to heterogeneity or the inherent randomness of input parameters and processes (Kiureghian and Ditlevsen, 2009). The second is epistemic uncertainty that stems from an incomplete and imprecise understanding of parameters that are modeled with fixed but poorly known values (Helton et al., 2010). Chapter 3 focuses on uncertainty and global sensitivity analysis, intending to characterize the magnitude and spatial distribution of epistemic uncertainty in the modeling of SoVI, as well as determining the model decision with the greatest influence on the model's final output.

Traditionally, every stage of the SoVI construction step is arbitrarily selected, ignoring any epistemic uncertainties that accompany every step of the index construction process. However, as modeling decisions during the index development proceed, epistemic uncertainty associated with each step propagates and potentially interacts with that of previous steps, causing large uncertainties to the final model output. To fully understand the effect of multiple and potentially interacting modeling stages, especially when the model is coupled with a dasymetric mapping technique, uncertainty and global sensitivity analysis are required. Uncertainty analysis helps assess the robustness of model output, and global sensitivity analysis decomposes the uncertainty to determine which index construction decision has the greatest influence on the model output variability.

Thus, I will perform a global sensitivity and uncertainty analysis on the SoVI model coupled with the newest dasymetric mapping technique developed in chapter 1 to answer the following research questions: (a) How much uncertainty is associated with the SoVI model coupled with the dasymetric mapping technique? (b) What is the relationship between the dasymetric mapping coupled SoVI model and uncertainty? (c) Which modeling decision contributes the most to variability in the dasymetric mapping coupled SoVI model? To address these questions, model options for an inductive SoVI model for the Davidson County, Nashville, are subjected to an uncertainty analysis using Quasi-Monte Carlo simulation and variance-based global sensitivity analysis. The uncertain model decisions that are evaluated are summarized in the following Table 4.

Table 4. Uncertainty analysis model factors.

Construction stage	Options	Probability density function
Indicator selection	Pycnophylactic indicators All indicators	Discrete (1, 2)
Indicator transformation	Raw Data Averaged by area	Discrete (3, 4)

Indicator normalization	Raw Data Z-score normalization Min-Max normalization	Discrete (5, 6, 7)
PCA component selection	Kaiser selection Percentage variance explained Horn's Parallel analysis	Discrete (8, 9, 10)
PCA rotation methods	Unrotated Varimax rotation Quartimax rotation Promax rotation (m = 2, 3, 4)	Discrete (11, 12, 13, 14, 15, 16)
Weight scheme	Equal weight sum First component only Weight sum using explainable variance	Discrete (17, 18, 19)

Indicator Selection

1. Pycnophylactic indicators: pycnophylactic is equivalent to volume or mass preserving property. Indicators count that is reallocated to finer spatial scales can be summed up to obtain the original value in the original coarse spatial scale. Here, I only select a subset of indicators with pycnophylactic properties which excludes median age, per capita income, median housing units, median gross rent, from the overall SoVI indicators.
2. All indicators: select the overall SoVI indicators.

Indicator Transformation

1. Raw data: no transformation method is applied.
2. Averaged by area: every indicator value is divided by the corresponding spatial unit area.

Indicator Normalization

1. Raw data: no normalization method is applied.
2. Z-score normalization: z score is calculated based on $z = \frac{x-\mu}{\sigma}$, where x is the individual indicator value, μ and σ are the mean and standard deviation of that indicator, respectively.
3. Min-Max normalization: $x' = \frac{x - \min(x)}{\max(x) - \min(x)}$

PCA Component Selection

1. Kaiser criterion (Kaiser, 1960): select only those components whose eigenvalues are greater than one.
2. Percentage variance explained: retain as many components as needed to account for an 80% amount of variation in the original data.
3. Horn's parallel analysis (Horn, 1958): Instead of using a fixed threshold, one retains those factors whose eigenvalues are larger than the expected eigenvalue for that component. Horn's parallel analysis uses 100 randomly generated data sets on which a PCA is performed and then averages over the resulting eigenvalues.

PCA Rotation Methods

1. Unrotated solution: no rotation is applied.

2. Varimax rotation (Kaiser, 1958): load each variable highly on just one component.
3. Quartimax rotation (Neuhaus, 1954; Carroll, 1953; Ferguson, 1954; Saunders, 1953): increase large loadings and decrease small ones, so that each variable will load only a few factors.
4. Promax rotation (Hendrickson & White, 1964): this method represents an oblique rotation. Thus, the rotated components are no longer orthogonal. The method also requires the specification of a power parameter, which is typically taken between 2 and 4. I chose the values 2, 3, and 4 for the algorithm.

Weight Scheme

1. Equal weight sum: a simple approach to sum the selected principals.
2. First component only: only select the first principal component to compose the SoVI.
3. Weighted sum using explainable variance: each component is weighted based on the proportion of total variation that the principal explains and then sums together to get the SoVI.

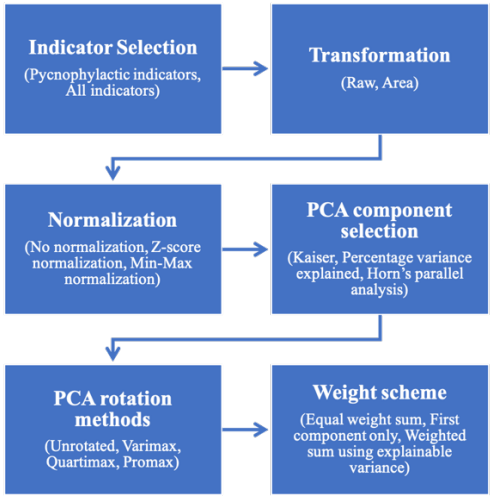

Figure 6. Diagram of uncertain construction factors associated with social vulnerability index (SoVI) composition process.

For non-linear mathematical models, such as the index composite model when normalization methods, PCA computation methods, and weighting schemes are all sampled, variance-based techniques for sensitivity analysis are the most appropriate (Saltelli, Tarantola and Campolongo, 2000). The importance of a given input factor X_i can be measured via the so-called sensitivity index, which is defined as the fractional contribution to the model output variance due to the uncertainty in X_i (Saltelli et al., 2004). For k independent input factors, the sensitivity indices can be computed by using the following decomposition formula for the total output variance $V(Y)$ of the output Y (Saltelli et al., 2004):

$$V(Y) = \sum_i V_i + \sum_i \sum_{j>i} V_{ij} + \cdots + V_{12\ldots k} \tag{8}$$

where

$$V_i = V_{X_i}\{E_{X_{-i}}(Y|X_i)\} \tag{9}$$

$$V_{ij} = V_{X_i X_j}\left\{E_{X_{-ij}}(Y|X_i, X_j)\right\} - V_{X_i}\left\{E_{X_{-i}}(Y|X_i)\right\} - V_{X_j}\{E_{X_{-j}}(Y|X_j)\} \quad (10)$$

Based on the law of total variance:

$$V_y = V_{X_i}\left\{E_{X_{-i}}(Y|X_i)\right\} + E_{X_i}\left\{V_{X_{-i}}(Y|X_i)\right\} \quad (11)$$

A first measure of the fraction of the unconditional output variance $V(Y)$ that is accounted for by the uncertainty in X_i is the first-order sensitivity index for the factor X_i defined as (Saltelli et al., 2004):

$$S_i = V_i/V \quad (12)$$

Terms above the first term in equation (8) are known as interactions effects among input factors. A model without interactions among its input factors is said to be additive. In this case, $\sum_{i=1}^{k} S_i = 1$ and the first-order conditional variances of equation (8) are all that need to be known to decompose the model output variance. Nonetheless, for a non-additive model, higher-order sensitivity indices, which are responsible for interaction effects among sets of input factors, must be computed. For a model with k independent input factors, the total number of indices (including $S_i s$) that should be estimated is as high as $2^k - 1$. For this reason, a more compact sensitivity measure which is the total effect order of sensitivity index is usually adopted (Saisana et al., 2005). The total effect order of sensitivity index concentrates all the interactions involving a given factor X_i in one single term S_{Ti}, defined as (Homma and Saltelli, 1996, Saltelli et al., 2004):

$$S_{Ti} = 1 - \frac{V_{X_{-i}}\{E_{X_i}(Y|X_{-i})\}}{V(Y)} \quad (13)$$

The conditional variance in the equation expresses the total contribution to the variance of Y due to non-X_i, i.e., the $k-1$ remaining factors, so that $1 - \frac{V_{X_{-i}}\{E_{X_i}(Y|X_{-i})\}}{V(Y)}$ includes all terms, i.e. a first-order term and interactions with all the other factors that involve factor X_i. In general, $\sum_{i=1}^{k} S_{Ti} \geq 1$, with equality if there are no interactions. For any given factor X_i, a significant difference between S_i and S_{Ti} indicates a critical role of interactions for that factor in Y. Highlighting interactions between input factors helps us to improve our understanding of the model structure (Saisana et al., 2005).

In this project, the method of Saltelli et al. (2010) will be used to calculate the first order of sensitivity index S_i:

$$S_i = \frac{\frac{1}{N}\sum_{v=1}^{N} f(\boldsymbol{B})_v[f(\boldsymbol{A}_B^{(i)})_v - f(\boldsymbol{A})_v]}{V(y)} \quad (14)$$

the method of Jansen (1999) will be used to calculate the total order of sensitivity index S_{Ti}, following the best practices identified in the sensitivity analysis (Saltelli et al. 2010; Puy et al., 2020):

$$S_{Ti} = \frac{\frac{1}{2N}\sum_{v=1}^{N}[f(\boldsymbol{A})_v - f\left(\boldsymbol{A}_B^{(i)}\right)_v]^2}{V(y)} \quad (15)$$

where any sampling point in either A or B sampling matrix can be indicated as x_{vi}, where v and i respectively index the row and the column. Estimators for both S_i and S_{Ti} are reviewed in Chan et al. (2000).

Sobol's (1967) method of Quasi-Random sampling will be used as the sampling algorithm for selecting the input factors since it fills the input space quicker and more evenly, leaving smaller unexplored volumes. The pair (S_i, S_{Ti}) can give a good description of the model sensitivities at a reasonable cost with a $n(k+2)$ model evaluations, where n represents the sample size that is required to approximate the multidimensional integration implicit in an equation to a plain sum (Saisana et al., 2005). N typically varies in the 100-1000 range. All the sensitivity computation processes will be conducted in the R package sensobol (Puy et al., 2021). The detailed experiment design is summarized in the following Table 5.

Table 5. Experiment design of uncertainty and sensitivity analysis.

	Uncertainty Analysis		**Sensitivity Analysis**
N	2^9	**Estimator**	First order: "saltelli", Saltelli et al. (2010)
K	6		Total order: "jansen", Jansen (1999)
Model evaluation	$N(k+2)$ $(6+2) \times 2^9 = 4096$	**Matrices**	c ("A", "B", "AB")
Input Factor PDF	Uniform Discrete	**Sample Algorithm**	Quasi-Random sampling

Both the uncertainty analysis and the global sensitivity analysis will be conducted in two spatial scales using the Davidson County, Nashville as a case study: (1) census tract level; (2) customized spatial scale derived from dasymetrically calculated values. For census tract level, all the SoVI variables are directly sourced from American Census Survey (ACS) or open street map resource and both UA and SA will be directly conducted at the census tract-level SoVI. For the customized spatial scale, all the SoVI variables in the census tract scale are firstly obtained from the ACS and the open street map source, then the cross-validated regularized hierarchical Bayesian multiple linear regression algorithm developed in chapter 1 will be applied to create a customized resolution map for every SoVI variable. Then, the SoVI at this customized spatial scale will be synthesized. Following this, both UA and SA will be conducted at this customized scale. The UA and SA results will be compared with the previous results conducted at the census tract level to learn how the newly developed dasymetric mapping technique affects the SoVI model's uncertainty and sensitivity distribution of different uncertain model inputs.

The innovation of this work lies in that it refines the traditional uncertain and global sensitivity analysis on the SoVI model, taking the dasymetric mapping technique into consideration. I hypothesize that the dasymetric mapping approach developed in chapter 1 will not affect the uncertain and sensitivity distribution of the SoVI model, suggesting the same robustness and validity of the SoVI model in both the original censu tract version and dasymetric mapping coupled version.

Chapter 3: Towards more transparency and flexibility in environmental policy making process: an interactive policy decision support tool using RShiny app

Develop an interactive decision support tool using Rshiny app by performing Nashville metro city's greenhouse gas (GHG) emissions analysis. Two deliverables will be produced: (1) a detailed Rmarkdown report that consists of Nashville's current GHG emissions and the evaluation of the impacts of many possible GHG emissions reduction actions; (2) an interactive web-based visualization and decision support tool using Rshiny app to explore the potential opportunities of various combinations of GHG emissions reduction actions. (Completion by November 2022)

The increasing demand for transparency is becoming an important feature of global sustainability governance (Gupta and Mason, 2014). Such transparency is increasingly 'multi-directional', with information disclosure demanded from but also provided to states, private actors, and citizens alike (Michener and Bersch, 2013). Motivated by this, Chapter 3 focuses on developing an approach that aims to improve the transparency and flexibility in the environmental policy-making process in the context of global climate change and sustainability that can be applied to any other metro city across the United States. In this project, Nashville metro city's GHG emissions data will be used as a case study and two products will be delivered: (1) A detailed Rmarkdown report and (2) an interactive, web-based visualization and environmental decision support system (EDSS) using Rshiny web application will be available to the public as well as the policymakers.

Recently, the development of new web frameworks offers an opportunity to create web applications directly from common scientific languages such as R and Python. This further increases the accessibility of scientific research and modeling tools because researchers can create web applications based on programming languages that they are already familiar with. Taking the R language program as an example, originally designed for statistical analysis, R is now widely applied in various fields of research and is particularly popular among environmental scientists. Among those R packages, Shiny (Chang et al., 2017) is the one providing a framework for building interactive web applications. Using the Shiny package, researchers can easily turn their R-based models and functions into EDSS as web-based apps targeting stakeholders as well as users. Publications of Shiny web have appeared in a wide range of scientific fields. For instance, Guillaume and Fu (2013) developed a Shiny web application to assist in knowledge elicitation about water requirements of floodplain and wetland vegetation. Whateley et al. (2015) developed a Shiny app for performing climate risk evaluations of small-scale water utilities in the northeastern U.S. Ye et al. (2018) proposed to use Shiny for developing cloud-based water resource analysis tools with a demonstration of a case study in China. The role of Shiny in all those applications is to take complex scientific concepts and models, process them in the backend and present them through an intuitive and user-friendly interactive front-end interface. Specifically, the server-side framework, which offers reactive expressions that automatically regenerate output data and figures at meantime when changes are made to the input, allows users to intuitively learn the relationship between the output and input. Thus, the Shiny web application framework offers a promising new method for developing decision support applications in various research communities.

To my knowledge, there are no Shiny web-based tools designed for exploring opportunities of environmental policy combinations to help restrict metro cities' future GHG emissions in the context of global climate change. This study addresses this gap by developing a Shiny web-based

decision support tool to rapidly evaluate various environmental policy combinations' impact on future GHG emissions using Nashville metro's GHG emissions data as a case study. To build an interactive EDSS to analyze Nashville metro's GHG emissions data, several important inputs are used throughout the Nashville sustainability and policy/wedge analysis. These include conversion factors, emissions factors, global warming potentials (GWP). Two alternative sources are used for this: (1) Nashville's 2009 Greenhouse Gas Emissions inventory (GHG inventory for short) (Gresham Smith and Partners, 2009); (2) The Environmental Protection Agency's Greenhouse Gas Emissions Inventory (Environmental Protection Agency, 2014, 2020). The EPA figures are representative of the nation in 2014, whereas the Nashville GHG inventory provides emissions factors specific to Nashville in 2006.

In terms of the emissions factors for electricity generation, three different scenarios are defined: (1) The emissions factors used in the Nashville 2009 GHG inventory (Gresham Smith and Partners, 2009); (2) The emissions factors used in the TVA's 2015 Integrated Resource Plan (IRP) (Tennessee Valley Authority, 2015); (3) The emissions factors used in the TVA's 2019 IRP (Tennessee Valley Authority, 2019). After the emissions data are read in, I will use the sum of emissions factors weighted by GWP (global warming potential) to determine the equivalent carbon dioxide/GHG emissions:

$$CO_2e = \sum_{i \in gases} GWP_i E_i \tag{16}$$

where GWP_i is the global warming potential for gas *i* and E_i is the emissions factor for gas *i*.

The main drivers of activity in this sustainability model are population and employment increase. These come from several sources of historical data and projected future growth. For population analysis, data is obtained from US Census for the historical population together with forecasts of the future population from two different growth forecasts for Davidson County through 2040. These forecasts were conducted for the NashvilleNext Planning process by Woods & Poole and by the University of Tennessee's Boyd Center for Business and Economic Research (CBER) and are described in a report on demographic trends submitted to NashvilleNext. The population forecast model interpolates between the given years using a constant exponential growth rate between these years and extrapolates beyond 2040 to estimate the 2050 population. For instance, to interpolate annual population numbers from 2010-2015, 2015-2025, and 2025-2040, I will calculate the implied annual growth rate between each pair and apply that rate for the growth in each year. I will also use the 2025-2040 growth rate to extrapolate from 2035-2050. I start with the formula for growth

$$P_n = P_0 \, (1 + r)^n \tag{17}$$

where P_0 is the initial population, P_n is the population *n* years later, and r is the growth rate. From this, I can derive the growth rate given P_0 and P_n:

$$r = \left(\frac{P_n}{P_0}\right)^{1/n} \tag{18}$$

In the data, the rate for year *n* means the increase from year *n* – 1 to year *n*. For employment data, similar to population, there is a combination of historical data from 2005-2017, provided by Shannon Hall at Metro Nashville Human Resources and projections, which are calculated in two ways but yield identical results. The first approach performs a linear regression to determine the

coefficient relating the number of Metro employees to the county population and then extrapolates the number of employees based on the increase in population after 2017. This coefficient is approximately 0.0017. This approach is defined as the regression model approach. The other approach that increases the number of Metro employees at the same rate as the county population is defined as the level of service (LOS) model approach.

Once the main drivers of activity and emissions factors are determined, the energy demand of each sector in the Nashville metro area can be computed. In this model, 7 sources of energy demand are included that are: streetlights, metro buildings, metro vehicles, wastewater treatment plants, solid waste plants, employee commuting, and public transit system. Each of these energy demand sectors has its own model to predict the future GHG emissions associated with some proposed policy alternatives to restricting its GHG emissions shown in Figure 7. Taking the simplest streetlight sector as an example, streetlight energy consumption projection is based on the Davidson County population whose increase rate of electricity consumption is assumed as the same as the population growth rate. The proposed environmental policy associated with streetlights is converting them to LEDs. By calculating different kinds of energy demand from those 7 sectors, the modified demand for electricity can be obtained.

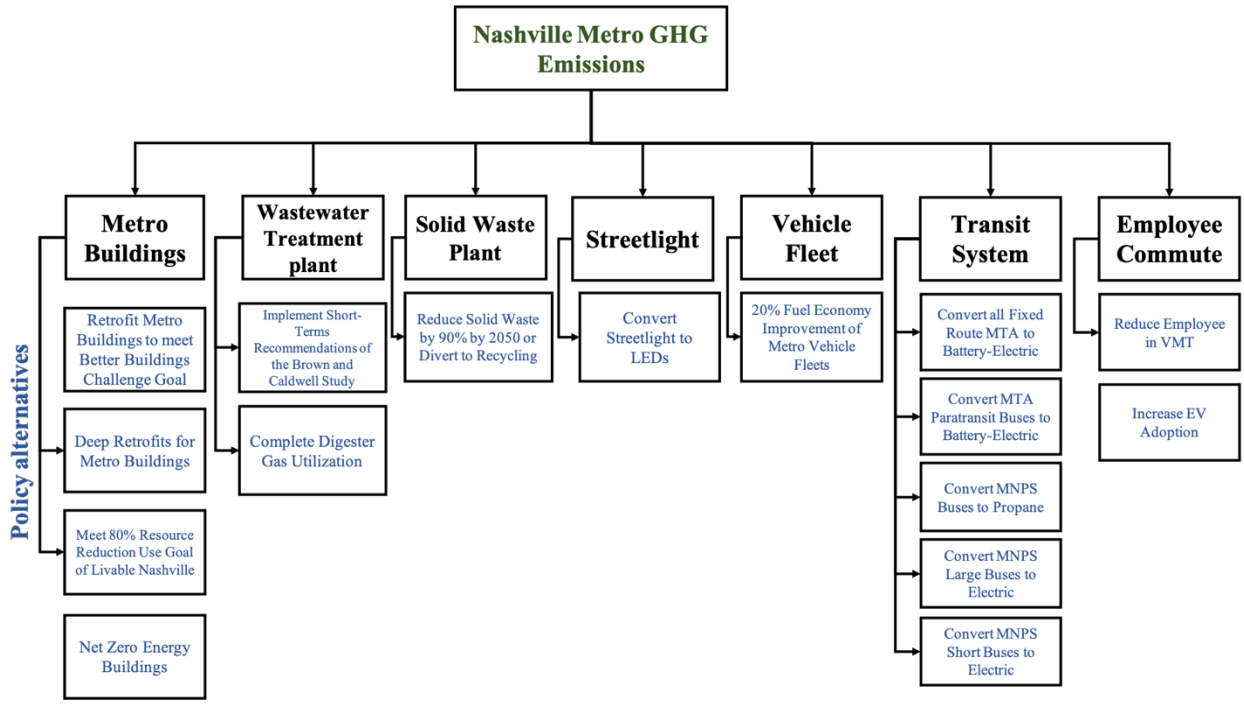

Figure 7. The seven sources of energy demand sectors and their associated proposed policy alternatives to restrict future GHG emissions in Nashville Metro GHG Emissions model.

After the emissions factors for each type of electricity generation and the total modified future demand for electricity from those 7 energy demand sectors are obtained, forecasts for the mix of generation of types providing electricity to Nashville can be computed. Then, the total greenhouse gas emissions in the baseline scenario as well as the wedge scenario can be calculated. The detailed calculation procedure logic flow of the interactive Rshiny app is shown in Figure 8. All the calculations and analyses will be published in a Rmarkdown document.

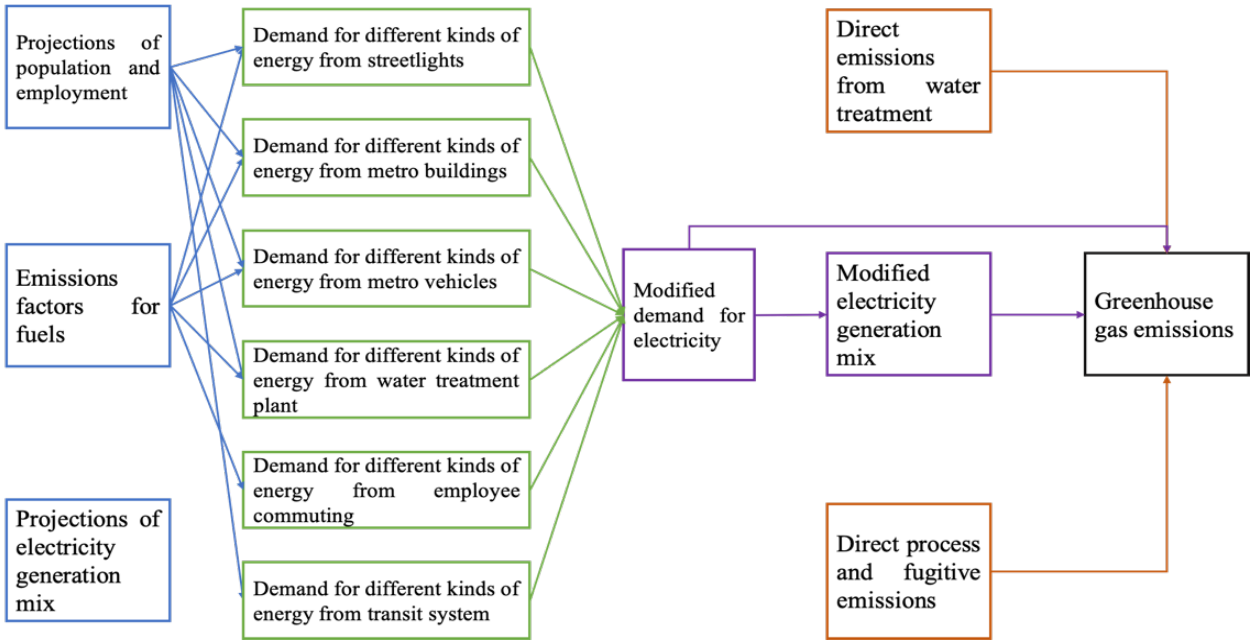

Figure 8. The interactive decision support tool Rshiny app calculation procedure logic flow diagram.

The Rmarkdown report will consist of two parts: the analysis of Nashville's current GHG emissions and several potential GHG reduction actions. The first part of this report will highlight Nashville's current GHG emissions, including the largest sources of emissions and how these emissions are computed. The second part of this report will analyze the impacts of a variety of possible GHG reduction actions, providing how these potential policies and actions are calculated to reduce the current GHG emissions and insights into which actions should be prioritized. This will allow both policymakers and the public to understand Nashville's current GHG emissions as well as different potential alternatives for reaching the city's reduction goal. The benefit of the Rmarkdown report format is that it allows for the direct integration of data, analysis, and text in the same document. Thus, all analyses conducted will be entirely transparent and reproducible. Due to this format, the report is also highly understandable and flexible. Data is integrated into the report, such that any future changes to the data or analysis will propagate into the report. The text serves to help both policymakers and the public understand how the analysis code works such that readers will understand how data is integrated, manipulated, and calculated.

The interactive Rshiny web application will be published online and open to the public. The application will allow users to understand the impacts of different policies on GHG emissions directly, as well as explore potential opportunities within interactions and combinations between policies. The application will be built using the "Shiny" tool which, like Rmarkdown, allows for transparency and flexibility by directly integrating the available data with the application. Users will be able to interact with the app to explore the impact of different policies and their combinations on the city's future GHG emissions. The application will be designed in such a manner that users may select policies to implement, rates of implementation, and scales of implementation and directly visualize how GHG emissions change under those different policies scenarios. Users may explore interactions between policies by selecting to implement multiple policies at the same time and viewing the impacts on emissions. This tool will be valuable for educating the public on the implications of different policy measures and how the combination of

those policies can reduce emissions. Also, combined with Rmarkdown document, this tool will be also valuable for decision-makers to make data-driven scientific-based climate resilience policies in a transparent and straightforward way. The Rshiny app user interface design and an example of the streetlights energy section user interface are displayed in Figure 9.

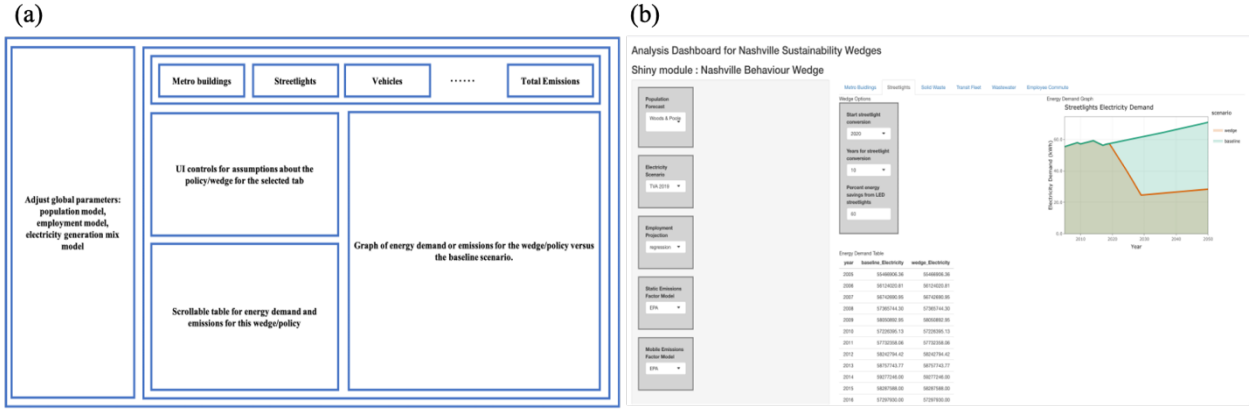

Figure 9. (a) Rshiny app user interface design; (b) Streetlights section user interface as an example.

Finally, a user study will be conducted to explore how this sustainability decision support system can help different groups of people better communicate future sustainability opportunities. Two groups of subjects will be studied in this study including decision-makers and the public. Surveys will be designed to test how these two groups of subjects respond to this Rshiny application. For decision-makers, the experiments will be designed to investigate how this Rshiny app helps them quickly grasp the information and make the decision. For the public, the focus will be assessing how this Rshiny app helps them understand the sustainability decisions made by policymakers.

Three key contributions of this project are summarized below:

1. **Understand current GHG sources and future opportunities.**

This work, both in the form of a detailed report and the interactive online application will increase policymakers' and the public's understanding of Nashville's current GHG emissions and future reduction opportunities. The detailed analysis will include recommendations of actions that should be prioritized to achieve Nashville's 80% emissions reduction target by 2050.

2. **Explore the impacts of various GHG reduction actions.**

In addition to recommendations from this work, the online Rshiny application will allow anyone to interactively explore pathways towards reducing emissions. The application will provide the public with a tool to visualize and investigate the implications of proposed actions, including interactions between multiple actions, on their own.

3. **Increase data transparency and flexibility.**

As all the deliverables will be built using open-source software and best practices in research reproducibility, the analysis and data will be completely transparent. The tools utilized, including Rmarkdown and Shiny, allow for direct integration of data, analysis, and text, which also ensures flexibility, transparency, and adaptability. This methodology and strategy of this work

have the potential to be transplanted to any metro city like Nashville metro area across the United States.

Schedule and Deliverables

A full list of dissertation tasks and corresponding schedule of activities is given in the Table 6 below.

Table 6: Proposed dissertation research tasks and schedules.

Tasks	2021					2022											
	Aug	Sep	Oct	Nov	Dec	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Synthesize various versions of Social Vulnerability Index (SoVI)																	
Qualifying Exam																	
Test the cross-validated regularized Bayesian hierarchical linear regression algorithm																	
Potential Publication: *"Cross-Validated Regularized Hierarchical Bayesian Linear Regression Application on Dasymetric Mapping Purpose (Chapter 1)"* (Target Journal: Applied Geography)																	
Uncertainty and Global Sensitivity Analysis coupled with dasymetric analysis on Social Vulnerability Index																	
Potential Publication: *"Uncertainty and Global Sensitivity Analysis coupled with Dasymetric Analysis of Social Vulnerability Index (Chapter 2)"* (Target Journal: Risk Analysis)																	
Develop an interactive Decision Support Tool using RShiny app																	
Potential Publication: "*Towards more transparency and flexibility in environmental policy making process: an interactive Decision Support Tool using RShiny app (Chapter 3)*" (Target Journal: Environmental Modeling and Software)																	
Submit draft of dissertation for review																	
Finalize and defend dissertation																	

References

Adger, W. N. (2000). Social and ecological resilience: Are they related? *Progress in human geography*, *24*(3), 347–364.

Ahlbrandt, R. (2013). *Neighborhoods, people, and community*. Springer Science & Business Media.

Aksha, S. K., Juran, L., Resler, L. M., & Zhang, Y. (2019). An analysis of social vulnerability to natural hazards in Nepal using a modified social vulnerability index. *International Journal of Disaster Risk Science*, *10*(1), 103-116.

Alahmadi, M., Atkinson, P. M., & Martin, D. (2014). A comparison of small-area population estimation techniques using built-area and height data, Riyadh, Saudi Arabia. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *9*(5), 1959-1969.

Anthoff, D., Nicholls, R. J., Tol, R. S., & Vafeidis, A. T. (2006). Global and regional exposure to large rises in sea-level: a sensitivity analysis.

Arup International Development. (2011). Characteristics of a safe and resilient community. International Federation of Red Cross and Red Crescent Societies, Geneva.

Arup (2014). City resilience framework, The Rockefeller Foundation, London.

Baker, C. K., Binder, S. B., Greer, A., Weir, P., & Gates, K. (2018). Integrating community concerns and recommendations into home buyout and relocation policy. *Risk, Hazards & Crisis in Public Policy*, *9*(4), 455-479.

Balica, S. F., Wright, N. G., & Van der Meulen, F. (2012). A flood vulnerability index for coastal cities and its use in assessing climate change impacts. *Natural hazards*, *64*(1), 73-105.

Balica, S., & Wright, N. G. (2009). A network of knowledge on applying an indicator-based methodology for minimizing flood vulnerability. *Hydrological Processes: An International Journal*, *23*(20), 2983-2986.

Baroni, G., & Tarantola, S. (2014). A General Probabilistic Framework for uncertainty and global sensitivity analysis of deterministic models: A hydrological case study. *Environmental Modelling & Software*, *51*, 26-34.

Bec, A., Char-lee, J. M., & Moyle, B. D. (2019). Community resilience to change: Development of an index. *Social Indicators Research*, *142*(3), 1103-1128.

Bian, R., Wilmot, C. G. (2015). Spatiotemporal population distribution method for emergency evacuation: case study of New Orleans, Louisiana. *Transportation Res. Rec*, *2532*(1), 99-106.

Bozheva, A. M., Petrov, A. N. and Sugumaran, R. (2005). The effect of spatial resolution of remotely sensed data in dasymetric mapping of residential areas. *GIScience and Remote Sensing*, *42*(2), 113–130.

Briggs, D. J., Gulliver, J., Fecht, D., & Vienneau, D. M. (2007). Dasymetric modelling of small-area population distribution using land cover and light emissions data. *Remote sensing of Environment*, *108*(4), 451-466.

Cantle, T. (2001). Community Cohesion: A Report of the independent Review Team (London: Home Office).

CARRI. (2013). Building Resilience in America's Communities: Observations and Implications of the CRS Pilots. Community and Regional Resilience Institute. http://www.resilientus.org/wp-content/uploads/2013/05/CRS-Final-Report.pdf.

Carroll, J. B. (1953). Approximating simple structure in factor analysis. *Psychometrika*, *18*(1), 23–38.

Chakraborty, J., Maantay, J. A., & Brender, J. D. (2011). Disproportionate proximity to environmental health hazards: methods, models, and measurement. *American journal of public health*, *101*(S1), S27-S36.

Chang, W., Cheng, J., Allaire, J., Xie, Y., McPherson, J. (2017). Shiny: web application framework for r. *R package version, 1*(5), 2017.

Chan, K., Tarantola, S., Saltelli, A. and Sobol', I. M. (2000). Variance based methods. In Sensitivity Analysis (eds A. Saltelli, K. Chan and M. Scott), pp. 167–197. New York: Wiley.

Clark, W. A. V., & Avery, K. L. (1976). The effects of data aggregation in statistical analysis. *Geographical Analysis*, *8*(4), 428–438.

COAG, National Strategy for Disaster Resilience: building the resilience of our nation to disasters. Attorney-Generals Department, Commonwealth of Australia. ISBN: 978-1-921725-42-5, 2011.

Collins, M., Chandler, R. E., Cox, P. M., Huthnance, J. M., Rougier, J., & Stephenson, D. B. (2012). Quantifying future climate change. *Nature Climate Change*, *2*(6), 403-409.

Cox, J., House, D., & Lindell, M. (2013). Visualizing uncertainty in predicted hurricane tracks. *International Journal for Uncertainty Quantification*, *3*(2).

Cruz, I., Stahel, A., & Max-Neef, M. (2009). Towards a systemic development approach: Building on the Human-Scale Development paradigm. *Ecological economics*, *68*(7), 2021-2030.

Cui, K., & Han, Z. (2019). Cross-Cultural Adaptation and Validation of the 10-Item Conjoint Community Resiliency Assessment Measurement in a Community-Based Sample in Southwest China. *International Journal of Disaster Risk Science*, *10*(4), 439-448.

Cutter, S. L. (2016). The landscape of disaster resilience indicators in the USA. *Natural hazards*, *80*(2), 741-758.

Cutter, S. L., Barnes, L., Berry, M., Burton, C., Evans, E., Tate, E., and Webb, J. (2008). A place-based model for understanding community resilience to natural disasters. *Global Environmental Change, 18*(4), 598–606.

Cutter, S. L., Boruff, B. J., & Shirley, W. L. (2003). Social vulnerability to environmental hazards. *Social science quarterly*, *84*(2), 242-261.

Cutter, S.L., Ash, K.D., and Emrich, C. T. (2014). The geographies of community disaster resilience. *Global Environmental Change*, *29*, 65–77.

Ebisudani, M., and Tokai, A. (2017). The Application of Composite Indicators to Disaster Resilience: A Case Study in Osaka Prefecture, Japan. *Journal of Sustainable Development*, *10*(1), 81.

Eicher, C. L. and Brewer, C. A. (2001). Dasymetric Mapping and Areal Interpolation: Implementation and Evaluation. *Cartography and Geographic Information Science*, *28*(2), 125–138.

EPA (Environmental Protection Agency). (2014). Flood resilience checklist. Vermont.

Environmental Protection Agency. (2014). *Emission Factors for Greenhouse Gas Inventories.* U.S. Environmental Protection Agency. https://www.epa.gov/sites/production/files/2015-07/documents/emission-factors_2014.pdf.

Environmental Protection Agency. (2020). *Inventory of U.S. Greenhouse Gas Emissions and Sinks: 1990–2018.* U.S. Environmental Protection Agency. https://www.epa.gov/sites/production/files/2020-04/documents/us-ghg-inventory-2020-main-text.pdf.

FEMA (Federal Emergency Management Agency). (2008). HAZUS-MH: Preparedness and response planning. Washington, DC.

Ferguson, G. A. (1954). The concept of parsimony in factor analysis. *Psychometrika*, *19*(4), 281–291.

Firrantello, J., Bahnfleth, W. P., Musser, A., Freihaut, J. D., & Jeong, J. W. (2007). Use of factorial sensitivity analysis in multizone airflow model tuning. *ASHRAE Transactions*, *113*, 642.

Flanagan, B. E., Gregory, E. W., Hallisey, E. J., Heitgerd, J. L., & Lewis, B. (2011). A social vulnerability index for disaster management. *Journal of homeland security and emergency management*, *8*(1).

Flanagan, B. E., Hallisey, E. J., Adams, E., & Lavery, A. (2018). Measuring community vulnerability to natural and anthropogenic hazards: the Centers for Disease Control and Prevention's Social Vulnerability Index. *Journal of environmental health*, *80*(10), 34.

Fraser, J., Elmore, R., Godschalk, D., Rohe, W. (2003). Implementing floodplain land acquisition programs in urban localities. *The Center for Urban and Regional Studies, University of North Carolina at Chapel Hill, Chapel Hill FEMA.*

Gallent, N., Mace, A., Tewdwr-Jones, M. (2003). Dispelling a myth? Second homes in Rural Wales. *Area, 35*(3), 271–284.

Gillespie-Marthaler, L., Nelson, K., Baroud, H., & Abkowitz, M. (2019). Selecting indicators for assessing community sustainable resilience. *Risk Analysis*, *39*(11), 2479-2498.

Giordano, A., & Cheever, L. (2010). Using dasymetric mapping to identify communities at risk from hazardous waste generation in San Antonio, Texas. *Urban Geography*, *31*(5), 623-647.

Gresham Smith and Partners. (2009). *Baseline Inventory of Greenhouse Gas Emissions for the Metropolitan Government of Nashville and Davidson County*. Metropolitan Government of Nashville-Davidson County. https://www.nashville.gov/Portals/0/SiteContent/Sustainability/2009GreenhouseGasInventory.pdf.

Guillard-Gonçalves, C., Cutter, S. L., Emrich, C. T., & Zêzere, J. L. (2015). Application of Social Vulnerability Index (SoVI) and delineation of natural risk zones in Greater Lisbon, Portugal. *Journal of Risk Research*, *18*(5), 651-674.

Guillaume, J.H.A., Fu, B. (2013). An Interactive Modelling Tool to Support Knowledge Elicitation Using Extreme Case Models.

Gupta, A., & Mason, M. (Eds.). (2014). *Transparency in global environmental governance: Critical perspectives*. MIT Press.

Haasnoot, M., Kwadijk, J., Van Alphen, J., Le Bars, D., Van Den Hurk, B., Diermanse, F., ... & Mens, M. (2020). Adaptation to uncertain sea-level rise; how uncertainty in Antarctic mass-loss impacts the coastal adaptation strategy of the Netherlands. *Environmental Research Letters*, *15*(3), 034007.

Helton, J. C., Johnson, J. D., Oberkampf, W. L., & Sallaberry, C. J. (2010). Representation of analysis results involving aleatory and epistemic uncertainty. *International Journal of General Systems*, *39*(6), 605-646.

Hendrickson, A. E., & White, P.O. (1964). Promax: A quick method for rotation to orthogonal oblique structure. *British Journal of Statistical Psychology*, *17*(1), 65–70.

Holt, J. B., & Lu, H. (2011). Dasymetric mapping for population and sociodemographic data redistribution. *Urban remote sensing: Monitoring, synthesis, and modeling in the urban environment*, *250*, 195-210.

Homma, T. and Saltelli, A. (1996). Importance measures in global sensitivity analysis of model output. *Reliability Engineering and System Safety*, *52*(1), 1–17.

Horn, J. L. (1958). A rationale and test for the number of factors in factor analysis. *Psychometrika*, *30*(2), 179–185.

Iman, R. L., Johnson, M. E., & Watson Jr, C. C. (2005). Sensitivity analysis for computer model projections of hurricane losses. *Risk Analysis: An International Journal*, *25*(5), 1277-1297.

Leaning, J., & Guha-Sapir, D. (2013). Natural disasters, armed conflict, and public health. *New England journal of medicine*, *369*(19), 1836-1842.

Jakeman, J., Eldred, M., & Xiu, D. (2010). Numerical approach for quantification of epistemic uncertainty. *Journal of Computational Physics*, *229*(12), 4648-4663.

Jansen, M. J. (1999). Analysis of variance designs for model output. *Computer Physics Communications*, *117*(1-2), 35-43.

Jia, P., & Gaughan, A. E. (2016). Dasymetric modeling: a hybrid approach using land cover and tax parcel data for mapping population in Alachua County, Florida. *Applied Geography*, *66*, 100-108.

Johansen, C., Horney, J., & Tien, I. (2017). Metrics for evaluating and improving community resilience. *Journal of Infrastructure Systems*, *23*(2), 04016032.

Kaiser, H. F. (1958). The varimax criterion for analytic rotation in factor analysis. *Psychometrika*, *23*(3), 187–200.

Kaiser, H. F. (1960). The application of electronic computers to factor analysis. *Educational and Psychological Measurement*, *20*(1), 141–151.

Kaly, U. L., Pratt, C. R., and Mitchell, J. (2014). The Environmental Vulnerability Index (EVI).

Kiureghian, A. D., and Ditlevsen, O. (2009). Aleatory or epistemic? Does it matter? *Structure Safety, 31*(2), 105-12.

Langford, M. (2007). Rapid facilitation of dasymetric-based population interpolation by means of raster pixel maps. *Computers, Environment and Urban Systems*, *31*(1), 19-32.

Latham, A., & Layton, J. (2019). Social infrastructure and the public life of cities: Studying urban sociality and public spaces. *Geography Compass*, *13*(7), e12444.

Local Government Association, Office of the Deputy Prime Minister, Commission for Racial Equality, The Inter-Faith Network (2002) Guidance on Community Cohesion London: LGA

Maantay, J. A., Maroko, A. R., & Herrmann, C. (2007). Mapping population distribution in the urban environment: The cadastral-based expert dasymetric system (CEDS). *Cartography and Geographic Information Science*, *34*(2), 77-102.

McGhee, D. J., Binder, S. B., & Albright, E. A. (2020). First, do no harm: evaluating the vulnerability reduction of post-disaster home buyout programs. *Natural Hazards Review*, *21*(1), 05019002.

Mennis, J and Hultgren, T. (2006). Intelligent dasymetric mapping and its application to areal interpolation. *Cartography and Geographic Information Science*, *33*(3), 179–194.

Mennis, J. (2009). Dasymetric mapping for small area population estimation. *Geography Compass, 3*, 727–745.

Mennis, Jeremy. (2003). Generating surface models of population using dasymetric mapping. *The Professional Geographer, 55*(1), 31-42.

Mennis, Jeremy. (2002). Using geographic information systems to create and analyze statistical surfaces of population and risk for environmental justice analysis. *Social science quarterly*, *83*(1), 281-297.

Michener, G., & Bersch, K. (2013). Identifying transparency. *Information Polity*, *18*(3), 233-242.

Nelson, K. S., Abkowitz, M. D., & Camp, J. V. (2015). A method for creating high resolution maps of social vulnerability in the context of environmental hazards. *Applied Geography*, *63*, 89-100.

Neuhaus, W. (1954). The quartimax method: An analytical approach to orthogonal simple structure. *British Journal of Statistical Psychology*, *7*(2), 81–91.

Norris, F.H., Stevens, S.P., Pfefferbaum, B., Wyche, K.F., and Pfefferbaum, R.L. (2008). Community Resilience as a Metaphor, Theory, Set of Capacities, and Strategy for Disaster Readiness. *American Journal of Community Psychology, 41*(1), 127–150.

OSSPAC (Oregon Seismic Safety Policy Advisory Commission). (2013). The Oregon resilience plan. Salem, OR.

Pendall, R., Foster, K.A., and Cowell, M. (2010). Resilience and regions: building understanding of the metaphor. *Cambridge J Regions Econ Soc, 3*(1), 71–84.

Petrov, A. (2012). One hundred years of dasymetric mapping: back to the origin. *The Cartographic Journal*, *49*(3), 256-264.

Pfefferbaum, R. L., Pfefferbaum, B., Nitiéma, P., Houston, J. B., and Van Horn, R. L. (2015). Assessing Community Resilience: An Application of the Expanded CART Survey Instrument with Affiliated Volunteer Responders. *American Behavioral Scientist, 59*(2), 181–199.

Pfefferbaum, R. L., Pfefferbaum, B., Van Horn, R. L., Klomp, R. W., Norris, F. H., & Reissman, D. B. (2013). The communities advancing resilience toolkit (CART): An intervention to build community resilience to disasters. *Journal of public health management and practice*, *19*(3), 250-258.

Pianosi, F., Beven, K., Freer, J., Hall, J. W., Rougier, J., Stephenson, D. B., & Wagener, T. (2016). Sensitivity analysis of environmental models: A systematic review with practical workflow. *Environmental Modelling & Software*, *79*, 214-232.

Plyer, A., Ortiz, E., Horwitz, B., and Hobor, G. (2013). The New Orleans index at eight. Greater New Orleans Community Data Center, New Orleans.

Puy, A., Becker, W., Piano, S. L., & Saltelli, A. (2020). The battle of total-order sensitivity estimators. *arXiv preprint arXiv:2009.01147*.

Puy, A., Piano, S. L., Saltelli, A., & Levin, S. A. (2021). sensobol: a R package to compute variance-based sensitivity indices. *arXiv preprint arXiv:2101.10103*.

Saisana, M., Saltelli, A., & Tarantola, S. (2005). Uncertainty and sensitivity analysis techniques as tools for the quality assessment of composite indicators. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, *168*(2), 307-323.

Saltelli, A., Annoni, P., Azzini, I., Campolongo, F., Ratto, M., Tarantola, S. (2010). Variance based sensitivity analysis of model output. Design and estimator for the total sensitivity index. *Computer Physics Communications*, *181*(2), 259-270.

Saltelli, A., Tarantola, S. and Campolongo, F. (2000). Sensitivity analysis as an ingredient of modelling. *Statist. Sci., 15*, 377–395.

Saltelli, A., Tarantola, S., Campolongo, F. and Ratto, M. (2004). Sensitivity Analysis in Practice, a Guide to Assessing Scientific Models. New York: Wiley.

Saunders, D. R. (1953). An analytical method for rotation to orthogonal simple structure (Research Bulletin 53–10). *Princeton, NJ: Educational Testing Service.*

Schmidtlein, M. C., Deutsch, R. C., Piegorsch, W. W., & Cutter, S. L. (2008). A sensitivity analysis of the social vulnerability index. *Risk Analysis: An International Journal*, *28*(4), 1099-1114.

Sempier, T. T., Swann, D. L., Emmer, R., Sempier, S. H., & Schneider, M. (2010). Coastal community resilience index: A community self-assessment. *Mississippi-Alabama Sea Grant Consortium*.

Shaw, R., Takeuchi, Y., Joerin, J., Fernandez, G., Tjandradewi, B. I., Wataya, E., ... & Matsuoka, Y. (2010). *Climate and Disaster Resilience Initiatives: Capacity-building Program*. United Nations International Strategy for Disaster Reduction (UNISDR).

Sherrieb, K., Norris, F. H., & Galea, S. (2010). Measuring capacities for community resilience. *Social indicators research*, *99*(2), 227-247.

Shim, J. H., & Kim, C. I. (2015). Measuring resilience to natural hazards: towards sustainable hazard mitigation. *Sustainability*, *7*(10), 14153-14185.

Sobol', I. Y. M. (1967). On the distribution of points in a cube and the approximate evaluation of integrals. *Zhurnal Vychislitel'noi Matematiki i Matematicheskoi Fiziki*, *7*(4), 784-802.

Stephenson, D. B., & Dolas-Reyes, F. J. (2000). Statistical methods for interpreting Monte Carlo ensemble forecasts. *Tellus A: Dynamic Meteorology and Oceanography*, *52*(3), 300-322.

Stevens, F. R., Gaughan, A. E., Linard, C., & Tatem, A. J. (2015). Disaggregating census data for population mapping using random forests with remotely sensed and ancillary data. *PloS one*, *10*(2), e0107042.

Su, M. D., Lin, M. C., Hsieh, H. I., Tsai, B. W., & Lin, C. H. (2010). Multi-layer multi-class dasymetric mapping to estimate population distribution. *Science of the Total Environment*, *408*(20), 4807-4816.

Tapia, C., Abajo, B., Feliu, E., Mendizabal, M., Martinez, J.A., Fernández, J.G., Laburu, T., and Lejarazu, A. (2017). Profiling urban vulnerabilities to climate change: An indicator-based vulnerability assessment for European cities. *Ecological Indicators, 78*, 142–155.

Tate, E. (2012). Social vulnerability indices: a comparative assessment using uncertainty and sensitivity analysis. *Natural Hazards*, *63*(2), 325-347.

Tate, E. (2013). Uncertainty analysis for a social vulnerability index. *Annals of the association of American geographers*, *103*(3), 526-543.

Tennessee Valley Authority. (2015). *Integrated Resource Plan*. Tennessee Valley Authority. https://tva-azr-eastus-cdn-ep-tvawcm-prd.azureedge.net/cdn-tvawcma/docs/default-source/default-document-library/site-content/environment/environmental-stewardship/irp/documents/2015_irp.pdf?sfvrsn=4892374_0.

Tennessee Valley Authority. (2019). *2019 Integrated Resource Plan*. Tennessee Valley Authority. https://tva-azr-eastus-cdn-ep-tvawcm-prd.azureedge.net/cdn-tvawcma/docs/default-source/default-document-library/site-content/environment/environmental-stewardship/irp/2019-documents/tva-2019-integrated-resource-plan-volume-i-final-resource-plan.pdf?sfvrsn=44251e0a_4.

Tobler, W. R. (1970). A computer movie simulating urban growth in the Detroit region. *Economic geography*, *46*(sup1), 234-240.

UNDP (2016). Human Development Index (HDI) | Human Development Reports.

Van Griensven, A., Francos, A., & Bauwens, W. (2002). Sensitivity analysis and auto-calibration of an integral dynamic model for river water quality. *Water Science and Technology*, *45*(9), 325-332.

Western, J., Stimson, R., Baum, S., & Van Gellecum, Y. (2005). Measuring community strength and social capital. *Regional Studies*, *39*(8), 1095–1109.

Whateley, S., Walker, J.D., Brown, C. (2015). A web-based screening model for climate risk to water supply systems in the northeastern United States. *Environ. Model. Software, 73*, 64–75.

Woolley, F. (2016). Social cohesion and voluntary activity: making connections. In *The economic implications of social cohesion* (pp. 150-182). University of Toronto Press.

Wu, S. S., Qiu, X. and Wang, L. (2005). Population estimation methods in GIS and remote sensing: a review. *GIScience and Remote Sensing, 42*(1), 80–96.

Ye, F., Chen, Y., Huang, Q., Li, L., 2018. Developing cloudbased tools for water resources data analysis using R and Shiny. In: Barolli, L., Zhang, M., Wang, X.A. (Eds.), Advances in Internetworking, Data & Web Technologies, vol. 6. Springer International Publishing, Cham, pp. 289–297.

Yoon, D. K., Kang, J. E., & Brody, S. D. (2016). A measurement of community disaster resilience in Korea. *Journal of Environmental Planning and Management*, *59*(3), 436-460.

Yuliastuti, N. (2018). Utilization of social facilities to reinforce social interaction in formal housing.

Zandbergen, P. A. (2011). Dasymetric mapping using high resolution address point datasets. *Transactions in GIS*, *15*, 5-27.

Zavar, E. (2015). Residential perspectives: the value of Floodplain-buyout open space. *Geographical Review*, *105*(1), 78-95.

Zhou, Y., Smith, S. J., Elvidge, C. D., Zhao, K., Thomson, A., & Imhoff, M. (2014). A cluster-based method to map urban area from DMSP/OLS nightlights. *Remote Sensing of Environment*, *147*, 173-185.