Assessing Climatic Hazards in Coastal Socio-Ecological Systems using Complex System Approaches

Zahra Nourali

Dissertation submitted to the faculty of the Virginia Polytechnic Institute and State University in partial fulfillment of the requirements for the degree of

Doctor of Philosophy
In
Biological Systems Engineering

Julie E. Shortridge, Chair
Zachary Easton
John Little
Anamaria Bukvic

19 April 2024
Blacksburg, VA

Keywords: Socio-Ecological Systems, Climate Change Impacts, Sea-Level Rise, Coastal Flooding, Human Relocation, Extreme Weather, Agricultural Losses, Crop Damage, Data-Driven Modeling

Assessing Climatic Hazards in Coastal Socio-Ecological Systems using Complex System Approaches

Zahra Nourali

ABSTRACT

Coastal socio-ecological systems face unprecedented challenges due to climate change, with impacts encompassing long-term, chronic changes and short-term extreme events. These events will impact society in many ways and prompt human responses that are extremely challenging to predict. This dissertation employs complex systems methods of agent-based modeling and machine learning to simulate the interactions between climatic stressors such as increased flooding and extreme weather and socio-economic aspects of coastal human systems. Escalating sea-level rise and intensified flooding has the potential to prompt relocation from flood-prone coastal areas. This can reduce flood exposure but also disconnect people from their homes and communities, sever longstanding social ties, and lower the tax base leading to difficulties in providing government services. Chapter 2 demonstrates a stochastic agent-based model to simulate human relocation influenced by flooding events, particularly focusing on the responses of rural and urban communities in coastal Virginia and Maryland. The findings indicate that a stochastic, bottom-up social system simulator is able to replicate top-down population projections and provide a baseline for assessing the impact of increasingly intense flooding. Chapter 3 leverages this model to assess how incorporating heterogeneity in relocation decisions across socio-

economic groups impacts flood-induced relocation patterns. The results demonstrate how this heterogeneity leads to a decrease in low-income households, yet a rise in the proportion of elderly individuals in flood-prone regions by the end of the simulation period. Flood-prone areas also exhibit distinct income clusters at the end of simulation time horizon compared to simulations with a homogenous relocation likelihood. Lastly, Chapter 4 explores relationships between extreme weather and agricultural losses in the Delmarva Peninsula. Existing research on climatic impacts to agriculture largely focuses on changes to major crop yields, providing limited insights into impacts on diverse regional agricultural systems where human management and adaptation play a large role. By comparing various multistep modeling configurations and machine learning techniques, this work demonstrates that machine learning methods can accurately simulate and predict agricultural losses across the complex agricultural landscape that exists on the Delmarva peninsula. The multistep configurations developed in this work are able to address data imbalance and improve models' capacity to classify and estimate damage occurrence that depend on multiple geographical, seasonal, and climatic factors. Collectively, this work demonstrates the potential for advanced modeling techniques to accurately replicate and simulate the impacts of climate on complex socio-ecological systems, providing insights that can ultimately support coastal adaptation.

Assessing Climatic Hazards in Coastal Socio-Ecological Systems using Complex System Approaches

Zahra Nourali

GENERAL AUDIENCE ABSTRACT

Coastal areas are facing increasing challenges from climate change, including rising sea levels and extreme weather conditions. This dissertation explores socio-economic consequences of these adverse environmental changes for coastal communities. Disruptive repetitive flooding due to exacerbated rise in sea levels is one of these consequences that may eventually leave some highly exposed coastal communities no alternative but migrating from their residences. Focusing on coastal Virginia and Maryland, Chapter 2 develops a data-informed model that can simulate individual relocation decisions and assess how they impact population changes and migration patterns. Chapter 3 employs this model to investigate how future changes in sea levels affect diverse socio-economic groups, their relocation decisions, and the resulting collective migration flows in flood-prone areas. We found that considering demographic differences leaves highly flood-prone areas with less low-income households, higher elderly individuals, and more economic clusters compared to simulations where these differences are not accounted for. Chapter 4 uses machine learning models to simulate the economic impact of extreme weather events as another manifestation of climate change

on the agriculture in the Delmarva Peninsula. Through data-based modeling techniques, we identify the climatic conditions most responsible for agricultural losses and recognize modeling choices that enhance our predictive ability. Collectively, this dissertation demonstrates how sophisticated modeling techniques can be used to better understand the complex ways in which climate change will impact human society, with the ultimate goal of supporting adaptation strategies that can better address these impacts.

Dedication

To my family

For your endless love and support.

Acknowledgements

I extend my deepest gratitude to Dr. Julie Shortridge, my advisor, for her support and mentorship throughout my PhD journey. Her continuous guidance, encouragement, and invaluable insights have been instrumental in shaping my research and personal growth.

I would like to thank Dr. Zachary Easton, Dr. John Little, and Dr. Anamaria Bukvic for their contributions as members of my committee. Their wisdom and feedback have been invaluable in helping me develop as an engineer and researcher.

I would like to thank to Dr. Jennifer Irish and Dr. Shao for their assistance and contributions to my research endeavors.

I would like to thank the faculty, staff members, and student organizations within the Department of Biological Systems Engineering at Virginia Tech for creating a memorable experience during my time in Blacksburg.

I would like to thank the Disaster Resilience and Risk Management (DRRM) Interdisciplinary Graduate Education Program at Virginia Tech for their support and resources.

Table of Contents

List of Figures

List of Appendix Figures

List of Tables

1. Introduction

Climate change is expected to impact many aspects of human society through a variety of changes in long-term climatic conditions and the frequency and intensity of extreme events. Coastal and low-lying areas are one of the ecosystems highly vulnerable to many impacts of climate change which are anticipated to include: sea-level rise of up to 0.6 m or more by 2100; an increase in sea surface temperatures by up to 3°C; intensified tropical and extratropical cyclones; larger extreme waves and storm surges; altered precipitation; and ocean acidification (Nicholls et al., 2007). These impacts, combined with considerable growth in the coastal residing human population, are increasing socio-economic vulnerability to extreme events. These consequences are significantly different at regional and local scales, but are expected to be predominantly negative and have intensified over the past decades.

The future changes in climatic events can be classified in different ways, which suggests different perspectives that focus on climate-induced coastal problems. According to Stephenson (2008), climate events can be distinguished using the terms of acute extremes and chronic events: the former term leads to a rapid onset and follows a short but severe course. Some high-impact, acute events are cyclones and convective storms with extreme wind speeds and precipitation that can lead to devastating impacts such as flood damage or total yield loss in agriculture. Chronic events are those that last for a long period of time or are marked by frequent recurrence with generally low intensity (Stephenson, 2008). Higher average temperature, less rainfall, and increased nuisance flooding are some examples of chronic events that can cause socio-ecological

impacts such as critical water shortages, agricultural failure, low life quality of the residing population, and migration. Climate change will impact both chronic conditions and extreme events, as well as their potential impacts on socio-ecological systems, such as economic disruptions in agricultural regions, displacement of communities, and damage to urban infrastructure. Building resilience capacity in coastal communities requires distinguishing between these impacts, understanding their interconnectedness and their cascading effects, especially considering the differential vulnerability of populations.

Characterizing and simulating the socioeconomic impacts of climate events can also differ based on geographical settings (such as metropolitan, urban or more rural) event when exposed to similar risks. Despite a high potential vulnerability due to reliance on climate-sensitive resources and limited capacities for adaptation, rural areas have been the topic of much less research and climate change policy making in comparison with urban settings (Kraybill & Lobao, 2001). Rural areas have already experienced serious climate change-related effects, including crop and livestock loss from severe drought, flooding, and deforestation (Neupane et al., 2016), damage to roads from extreme storms, and a reduction in families' livelihood options (Chumky et al., 2022). These all can lead them to leaving resource-dependent rural areas and creating new migration patterns (Bates & Rudel, 2004) which is a tremendous force of social, cultural, demographic, and economic changes. The potential for climate change to affect rural communities deserves careful consideration and policy attention, due to their climate-sensitive livelihood, less-developed infrastructure (Singh et al., 2019; Žurovec et al., 2017), and potentially lower climate literacy (Mycoo, 2015). These impacts also require consideration of the

complexity and interdependence between hazardous physical events and social and economic dimensions of coastal systems.

More recently, the concept has been extended to the context of socio-environmental systems (SES) where humans interact with the natural (ecosystems, geological formations, and climatic patterns) and engineered components (infrastructure, built environments, and technological interventions) within their environments (Elsawah et al., 2017). Cross scale feedbacks, adaptive dynamics, and interactions between subsystems and their components in SES are best characterized using a complex systems paradigm (Filatova et al., 2013). This paradigm tries to understand systems' highly interconnected components and circular causalities, giving rise to emergent properties and non-equilibrium phenomena that cannot be derived from a reductive analysis of the system in terms of its isolated components. Analytical and Statistical Modeling, Cellular Automata, Artificial Learning, and Agent-based Modeling (ABM) have emerged as different ways to better capture the complexity of human-natural dynamics. The complex nature of environmental challenges and adaptation planning requirements presents several obstacles for integrated research. Firstly, achieving an integrated assessment of socio-ecological systems necessitates input from diverse sources and forms of knowledge (Jakeman et al., 2006), involving the utilization and synthesis of various methods such as conceptual, computational, and primary research approaches. Secondly, attaining the promised holistic perspective of systems and processes relies significantly on system representation form and levels of aggregation (Kelly et al., 2013), with coupling models offering a potential solution to this issue. Thirdly, integrated modeling of socio-ecological systems often contends with spatially distributed biophysical and socio-

economic dynamics, where spatial heterogeneity significantly influences system behavior (Hamilton et al., 2015). These issues create numerous challenges in simulation of socio-ecological systems that require theoretical and methodological advancements. This dissertation endeavors to address some of the research gaps by developing approaches for modeling the socio-economic consequences of climate change impacts on coastal communities.

1.1. Goals and Objectives

The overall goal of this dissertation is to assess how climate change can affect complex coastal socio-ecological systems using complex systems methods in order to simulate and predict human-environment interactions. This will be accomplished through the following three primary objectives:

1. Assess the relocation response of rural and urban municipalities to sea-level rise and repetitive flooding in coastal Virginia using agent-based modeling.
2. Explore how socio-demographic heterogeneity in relocation decisions impacts municipalities under sea-level rise using a stochastic agent-based approach.
3. Develop data-analytic approaches to investigate the agricultural losses caused by extreme weather events in Delmarva peninsula.

1.2. Dissertation Organization

This dissertation is composed of four chapters. Chapter 1 is an introduction to the overall dissertation. Chapter 2 discusses the research performed to accomplish the first objective

and was published in *Water*. Chapter 3 discusses the research performed to accomplish second objective and is under preparation for submission in *Climatic Change* journal. Chapter 4 discusses the research performed to accomplish third objective and is in preparation to be submitted for publication in *Climatic Change*. Chapter 5 discusses overarching conclusions and findings from the three projects encompassed in the dissertation.

2. Simulation of Flood-Induced Human Migration at the Municipal Scale: A Stochastic Agent-Based Model of Relocation Response to Coastal Flooding

Zahra Nourali, Julie E. Shortridge, Anamaria Bukvic, Yang Shao and Jennifer L. Irish

Water 2024, *16*(2), 263; https://doi.org/10.3390/w16020263
Published: 11 January 2024

Abstract

Human migration triggered by flooding will create sociodemographic, economic, and cultural challenges in coastal communities, and adaptation to these challenges will primarily occur at the municipal level. However, existing migration models at larger spatial scales do not necessarily capture relevant social responses to flooding at the local and municipal levels. Furthermore, projecting migration dynamics into the future becomes difficult due to uncertainties in human–environment interactions, particularly when historic observations are used for model calibration. This study proposes a stochastic agent-based model (ABM) designed for the long-term projection of municipal-scale migration due to repeated flood events. A baseline model is demonstrated initially, capable of using stochastic bottom-up decision rules to replicate dynamics of county-level population distribution. This approach is then combined with physical flood-exposure data to simulate how population projections diverge under different flooding assumptions. The methodology is applied to a study area comprising 16 counties in coastal Virginia and Maryland, U.S., and include rural areas which are often overlooked in adaptation research. The results show that incorporating flood impacts results in divergent population growth patterns in both urban and rural locations, demonstrating potential municipal-level migration response to coastal flooding.

Keywords: agent-based modeling (ABM); human migration; climate change; coastal flooding; socio-ecological systems

2.1. Introduction

Changing precipitation patterns, droughts, land degradation, flood events, and sea-level rise already affect many coastal socio-ecological systems (SES). These and other potential climate change impacts on human systems can drive complex, uncertain changes to the population size, resource use, and economic activity of future societies (Cairns et al., 2013), and in some cases, exacerbate human migration patterns (Döös, 1994; Hugo, 1996; MacKellar et al., 1998; Magadza, 2000; Meze-Hausken, 2000; Myers, 2002). The fact that climate-induced migration can itself cause subsequent environmental consequences such as deforestation and degradation of natural resources implies the bidirectional relationship between environment and migration (de Sherbinin et al., 2008). Accordingly, predicting human migration and understanding how relocation dynamics will be influenced by different environmental changes is crucial for both adaptation planning and environmental conservation, as an important input in the decision-making process of many governments and organizations.

In response to these challenges, there is a growing body of research on climate-influenced migration that can generally be categorized into three approaches: theoretical frameworks, empirical methods, and agent-based modeling (ABM). Early migration theories were based on economic reasoning (e.g., employment opportunities and income) as the main driver triggering the migration processes (Massey, 2015). It is also conceptualized using a pull–push model accounting for aspects that attract and repel people

to move (Lee, 1966). This foundation led to further studies that consider other drivers of migration such as environmental, social, political, and demographic forces (Black et al., 2011; Hunter et al., 2015; McLeman, 2014; Piguet, 2010). Empirical approaches typically use statistical methods to quantitatively identify relationships between environmental change and migration, emphasizing the role of climatic event type and migration characteristics (Bohra-Mishra et al., 2014; Feng et al., 2010; Gray & Mueller, 2012; Gray & Wise, 2016; Marchiori et al., 2012; Mastrorillo et al., 2016; Mueller et al., 2014; Thiede & Gray, 2017). However, empirical approaches require large volumes of data to identify causal relationships, which limits their applicability in contexts with sparse data. While useful to characterize factors influencing migration on a large-scale population level, such methods cannot capture complex granular factors that influence migration decisions at the individual or household level (Kniveton et al., 2011). Finally, empirical analyses of migration behavior can only assess historically observed relationships, which are not necessarily reflective of future migration patterns. For instance, Hauer (2017) assesses the potential influence of SLR migrants on the U.S. population distribution across all coastal and in-land areas by merging projected populations at risk of SLR with pre-existing county-level migration pathways. This is based on the assumption that SLR-induced migrations happen only among locations that have already experienced migrations historically. However, research that integrates future climate change projections suggests that migration patterns will not necessarily follow historically observed patterns (Robinson et al., 2020). Finally, empirical methods do not capture heterogeneous decision making and behaviors across large populations and only find general patterns and relationships out of collective data. While broad-scale data-driven assessments provide overarching trends in relocation

responses to coastal flooding, they may not fully capture the wide range of factors influencing individual decision-making, such as socioeconomic status, personal preferences, and risk perceptions.

Agent-based modeling (ABM) provides an alternative mechanism to represent heterogeneous decision-making processes and human responses to natural disasters where individual-level variation in behaviors leads to emergent collective outcomes, such as evacuation (Chen et al., 2023; Joo et al., 2013; Lee et al., 2021; Silva & Eleutério, 2023) and migration (Hassani-Mahmooei & Parris, 2012; Karanci et al., 2017). Numeric simulations of multi-agent systems can capture effects stemming from heterogeneity across different types of actors (De Cian et al., 2020). ABM allows social, economic, and environmental factors to be incorporated into a migration model that considers social-ecological feedbacks, evolutionary learning, and out-of-equilibrium dynamics to better capture complex system characteristics of migration patterns due to climate change (Thober et al., 2018). These factors can be combined with spatial data to develop a spatially explicit simulation of climatic risks and associated individual responses. The latter is important as SES often shows high levels of spatial variability. Through spatial ABMs, possible interactions between social systems and spatially explicit environmental systems (such as the dynamics of sea-level rise and inundated areas across spatial scale) can be better tracked. The added value of ABM is its ability to explore dynamic paths of SES, which usually involve abrupt changes and transitions emerging from cumulative effects of social interactions and adaptive behaviors, since other modeling approaches that assume perfect information and static or rational behavior may be misleading (Filatova et al., 2013).

ABM has been recently used to study migration in response to different environmental factors, including studies that assess different theoretical foundations (Silveira et al., 2006) and incorporate temporal dimensions (Hailegiorgis et al., 2018; Lin et al., 2016; Nguyen et al., 2018). One key challenge in ABM in any context is model validation (Crooks et al., 2008), with multiple studies calling for more systematic and rigorous model validation processes (Filatova et al., 2013; Mehdizadeh et al., 2022). Forward-looking ABMs that consider sea-level rise as one of the environmental drivers in their model (Hassani-Mahmooei & Parris, 2012; Karanci et al., 2017) usually validate simulations only by feeding historical data to the initialization step. However, developing and validating forward-looking models using historical data may not produce valid projections of migration into the future. The lack of calibration to compare model results in forward-looking ABMs can affect the credibility of their long-term migration projections in the context of climate change.

Existing empirical models of climate-induced migration primarily focus on livelihood-based migration at international or national scales (Cai & Oppenheimer, 2013; Kniveton et al., 2012). Spatially explicit ABM can complement this work by tracking displacements for more detailed regional assessment and planning. For instance, Hassani-Mahmooei and Parris (Hassani-Mahmooei & Parris, 2012) investigate internal migration across all divisions of Bangladesh using a temporal-spatial ABM. Models of migration decision making over smaller scales (municipal to regional) can be beneficial in addressing multiple issues. First, national-scale findings are not necessarily representative of the specific migration drivers and their impacts on regional and municipal scales. This is particularly important in adaptation planning, as many actions in response to SLR and

flooding (such as infrastructure hardening or zoning changes) are taken by local governments or regional agencies. Local mobility in response to climate stressors could have significant implications for municipalities, for instance, by exacerbating population loss in rural communities or intensifying urban housing shortages. While national-scale studies mostly consider migration from a livelihood perspective (for example, subsistence farmers who must migrate when crop yields decline), many other factors combine to influence human spatial behavior and vary among regions and social groups (McLeman & Smit, 2006). At smaller spatial scales, people's displacements occur across shorter distances which may not lead to different jobs or wages. Thus, the factors that drive people to move at local scales may be quite different than at large scales and include factors such as housing, family, and educational opportunities (Frost, 2020). Existing ABMs of local mobility responses to sea-level rise or flooding typically adopt an economic focus on housing market dynamics based on utility maximization (Ettema, 2011; Y. Han et al., 2020; Hemmati et al., 2021; Karanci et al., 2017). However, this does not account for other factors that influence relocation decision making, such as disaster impacts on infrastructure and community services (Bukvic & Owen, 2017). Finally, a comparison between the portion of rural versus urban governance structures that do have adaptation and land-use planning sections (31% versus 71%) shows that rural settings are usually overlooked in adaptation research (Kraybill & Lobao, 2001). At the same time, rural settings are often even more vulnerable to climate impacts due to factors such as demography, occupations, earnings, literacy, poverty incidence, and dependency on government funds (Lal et al., 2011). A small spatial scale can allow distinguishing between urban and rural settings and their unique

characteristics, compare their migration response, and investigate outcomes on their municipal structure.

To address the aforementioned gaps, this study proposes a stochastic ABM for flood-induced migration at the municipal to regional level across rural and urban coastal areas. This ABM is based on the push–pull theory of migration which posits that the decision to migrate stems from numerous factors driving people away from certain locations and attracting them to others (Dorigo & Tobler, 1983; Hunter, 2005; Pan, 2019). A set of bottom-up agent-level decision rules is calibrated to replicate county-level population projections from NASA's Socioeconomic Data and Applications Center (SEDAC). This step provides a baseline model of population movement in the study region and demonstrates that individual decision rules can be formulated to replicate macro-scale behavior in best-available population projects. This model is then used to assess the impact on census tract-level population distribution under different assumptions about the relative role of flooding in individual migration decisions under multiple 50-year storylines of nuisance and extreme flooding. This model is then utilized to determine how flooding alters the spatial and temporal patterns of population change relative to baseline projections that do not consider flooding. This general approach can provide a foundation for further investigations on flood-induced migration based on available public data that can be expanded to accommodate various decision rules and projections of coastal flood risk.

2.2. Materials and Methods

The proposed method revolves around the concept of simulating migration behavior in response to flooding through a bottom-up mindset using ABM. In an ABM, a system is depicted as a group of independent decision-making entities known as agents.

Each agent autonomously evaluates its circumstances and makes decisions based on a predefined set of rules. These agents may engage in a set of actions based on the interactions defined in the system they represent. An ABM captures the emergent behavior arising from the interplay of individual entities within their environment. Emerging phenomena such as flood-induced human migration as a whole exceed the mere sum of its parts due to the inherent dynamics of this system.

This research developed a stochastic ABM based on the push–pull migration theory for flood-induced migration across flood-prone rural and urban areas of coastal Virginia and Maryland. Initially, a baseline model was developed and calibrated to simulate individual agent migration decisions as a stochastic process, relying on the underlying pull score of the census tract in which they currently live, without considering flood impacts. This model was then used to simulate migration patterns under different storm surge flood storylines derived from the U.S. Army Corps of Engineers' North Atlantic Coast Comprehensive Study storm surge hazard assessment (Cialone et al., 2015) and under different sea-level rise scenarios. Migration simulations were conducted with different assumptions about the importance of storm surge flooding relative to other factors that influence migration decisions and evaluated the degree to which storm surge flooding causes future projections to deviate from our baseline model (Figure 2-1). Model simulations were run for a 50-year simulation period from 2021 to 2070, to assess how repeated and aggregated individual relocation decisions impact municipal-scale population patterns in flood-prone areas over the long term.

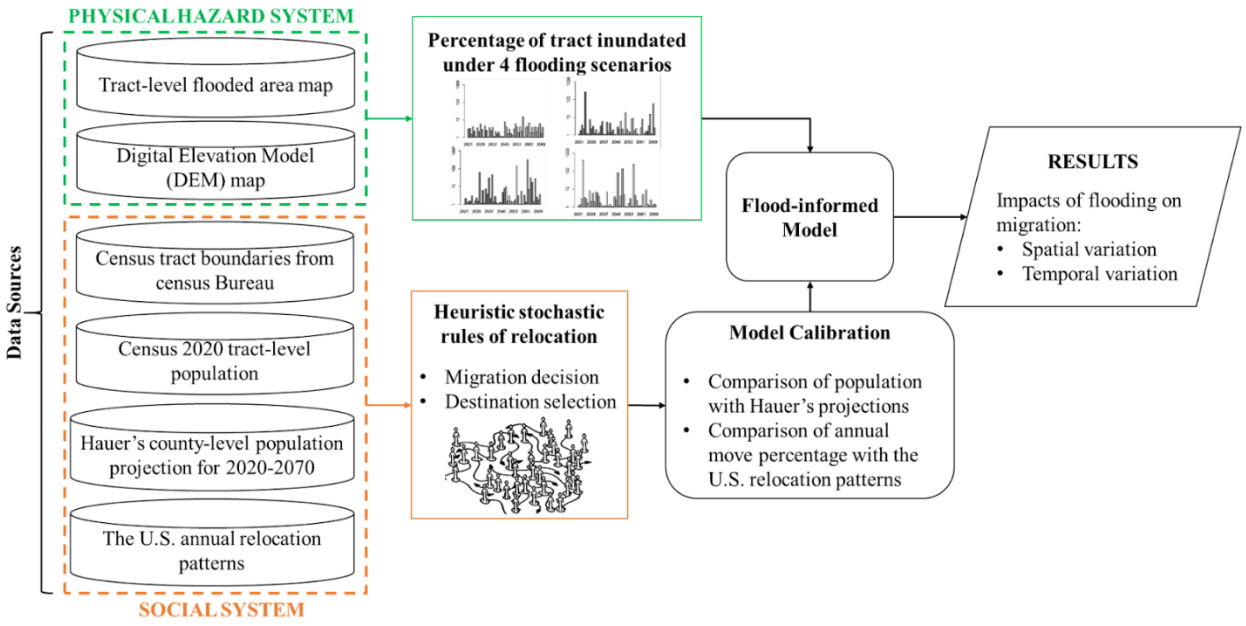


Figure 2-1. Schematic overview of proposed ABM

2.2.1. Study Area

The study area considered in this research comprises 16 coastal counties within the states of Virginia and Maryland, U.S. This study location includes both urban and rural coastal areas with both increasing and decreasing population over the past decade (Figure 2-2). Sea levels have risen significantly faster in the Mid-Atlantic U.S. region compared to the global average due to land subsidence from sediment compaction, glacial isostatic rebound, groundwater extraction, and weakening Gulf Stream currents, making this area the second largest population center at SLR-related risks in the U.S. (Atkinson et al., 2012; Engelhart et al., 2011; Parris et al., 2012; Tompkins & Deconcini, 2014). The rate of SLR in coastal Virginia is predicted to increase 13.1% to 71% by 2100 (Habete & Ferreira, 2017), which is exacerbated due to the low-lying topography in this area (Kleinosky et al., 2007). Higher sea level increases the extent of flooding even without the occurrence of rainfall. Even though sea levels have only risen by around 12.7 cm, the annual frequency of minor flooding events has increased by 33% in some areas of Virginia since 2000 (Sweet

et al., 2018). While less damaging, minor flooding can lead to notable disruptions affecting quality of life, such as school closures, infrastructure shutdowns, and difficulty accessing community services. Sea-level rise also increases the risks and potential impacts from larger flood events in the region, including tropical storms and hurricanes. The annual accelerating frequencies and cumulative effects of floods are becoming a threatening problem in several locations with strategic importance to national security, including Norfolk, Virginia (Sweet et al., 2018), which encompasses Naval and other defense department installations (Hall et al., 2016).

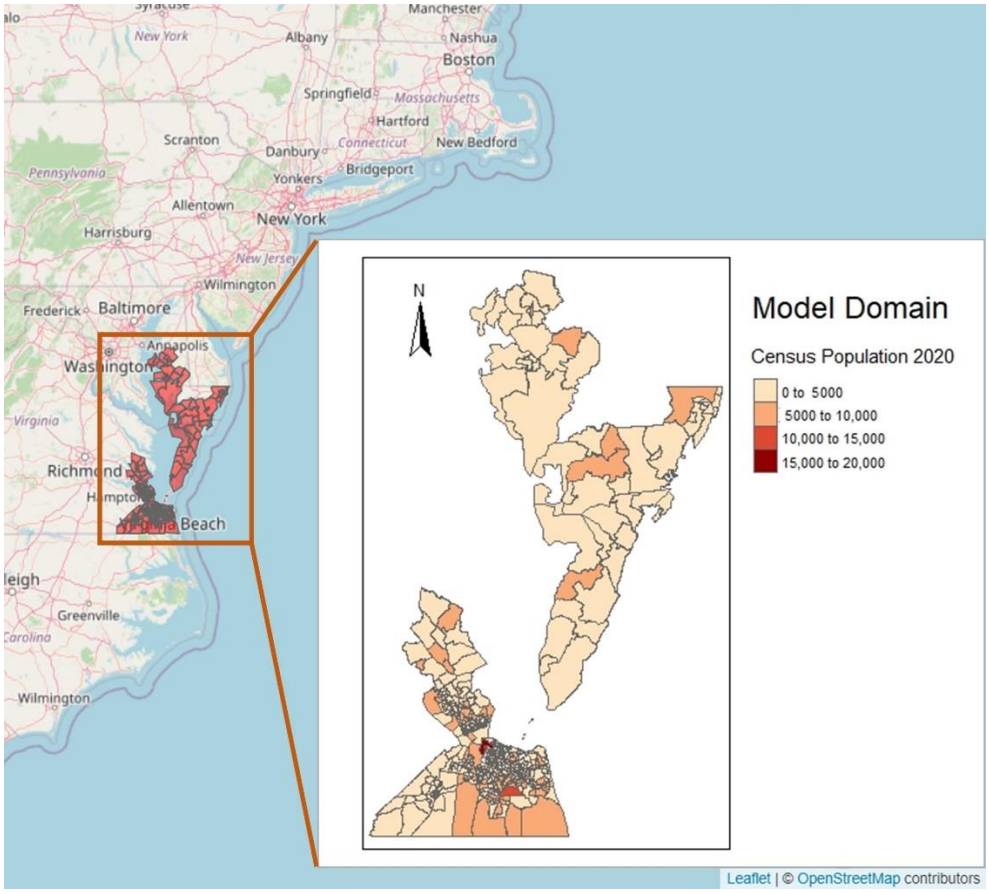


Figure 2-2. The model domain delineated by U.S. census tracts and colored by the real population in 2020 (U.S. Census Bureau, 2021).

2.2.2. Baseline Model

The goal of the baseline model was to develop a set of stochastic agent decision rules that could effectively replicate broad-scale population projections in the study area. For this purpose, best-available dataset of county-level population projections based on SSPs (Hauer & CIESIN, 2020; Hauer, 2019) was utilized. These aggregate data-driven projections were obtained from transforming county-level historic U.S. census data to demographic analyses measures using Autoregressive Integrated Moving Average (ARIMA) models, and feeding them to Leslie matrix population projection models, subject to SSPs' control to avoid runaway population growth. A key advantage of the statistical approach used to generate these projections is the explicit ability to account for statistical error terms and quantify uncertainty in future projections. Consequently, incorporating these details into agent-based modeling simulations would be a valuable area of additional research.

In an ABM, agents can individually assess their situation and take actions based on a set of decision rules (Bonabeau, 2002), which can aggregate into emergent phenomena where group-level behavior differs from the individuals alone. The factors that influence population growth and decline in urban and rural areas are highly complex, including proximity to employment opportunities and wages (Karcagi Kováts & Katona Kovács, 2012), housing prices (Plantinga et al., 2013), and public services such as education, unemployment insurance benefits, and public safety (Charney, 1993). Explicitly simulating all of these factors is beyond the scope of this work. As an alternative, our baseline model combined a "pull score" that describes the relative desirability of each census tract based on population growth projections.

U.S. Census tracts were used as the spatial unit within the model, allowing for the integration of physical flood inundation data in later steps. The forward-looking simulation is conducted yearly so that each time step represents a year from 2021 to 2070. Each agent in the model represents a group of 200 individuals, which serves as a representative sample of the real population in the study area. The abstraction level of 200 individuals per agent was made to balance computational efficiency and statistical representation, aligning with prior research that successfully aggregated agents to simulate the climate-induced migration behavior of extensive populations (Hassani-Mahmooei & Parris, 2012). As the stochastic nature of the model required running multiple simulations to determine system behavior under each scenario, aggregating individuals into groups allows capturing the inherent heterogeneity within the population while reducing computational complexity. This value results in an initial population of 2602 agents in 2021, representing a study area population of 520,400 individuals. In each time step, agents are randomly assigned a satisfaction score based on the push–pull score of the census tract in which they currently live; these push–pull scores are derived from county-level population projections (Hauer & CIESIN, 2020; Hauer, 2019). The current model uses the satisfaction score as an aggregate measure summarizing the numerous factors that influence an individual's desire and capability to relocate, such as economic reasons and job opportunities, information concerning alternative localities, demographic characteristics such as age, education level, and parenthood, and psychic cost of migration (Greenwood, 1975). Because of the challenges inherent in measuring and simulating these factors explicitly at the individual scale, a stochastic approach is adopted that samples a random satisfaction value across a

known distribution, with the acknowledgment that individuals will vary in their satisfaction in ways that are impossible to predict.

NetLogo (Tisue & Wilensky, 2004) is used as the modeling toolkit which is a programmable environment for simulating natural and social phenomena. The overview of the modeling approach is shown in Figure 2-3. The model is initiated through a setup phase, where the spatial features of the model are integrated into the NetLogo environment, push–pull factors for each census tract are calculated, and agents are generated across the study area. For each year of the simulation, a two-phase process is used to simulate agent moves within the study area (blue box in Figure 2-3), and agent moves in and out of the study area (red box in Figure 2-3). The details of each step of the model process are discussed in more detail in the following sections.

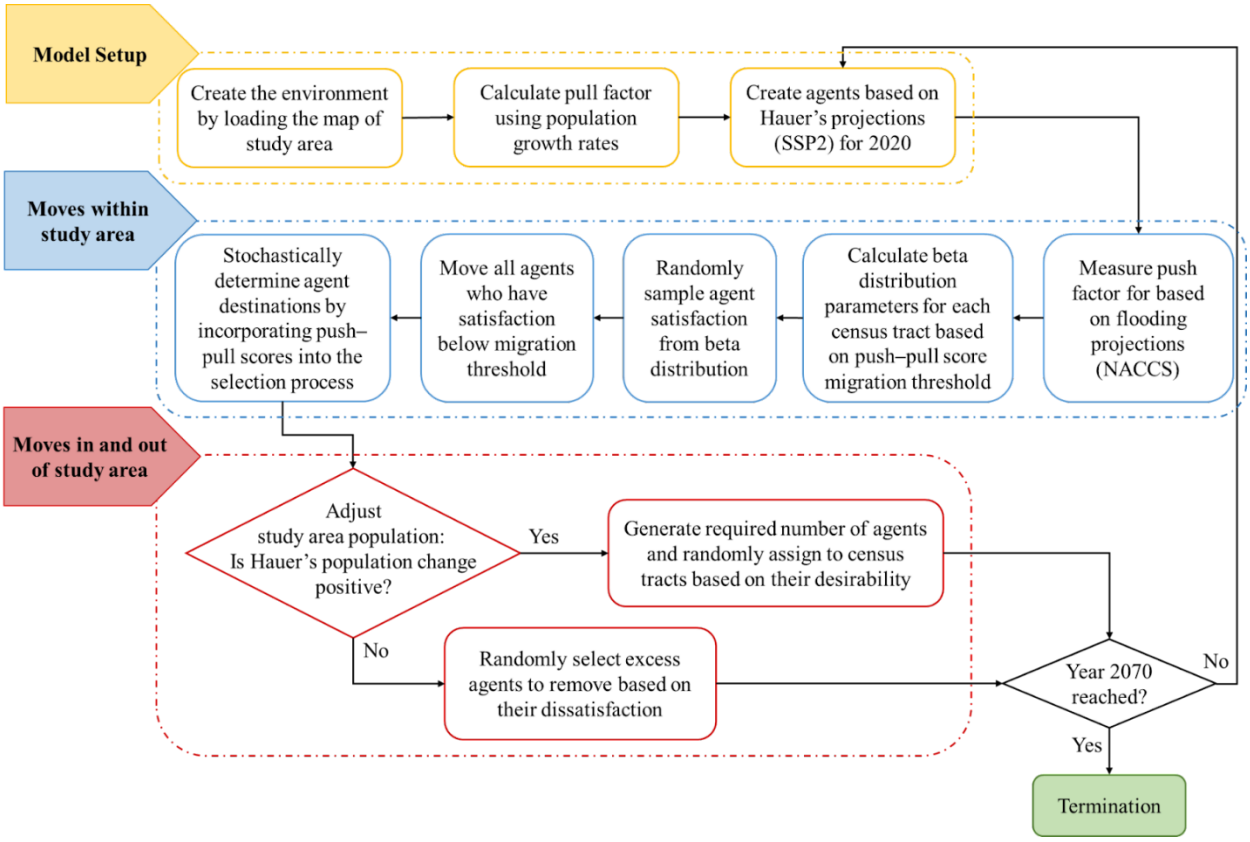


Figure 2-3. The flowchart overview of the agent-based mode.

2.2.2.1. Data Sources

Census tract-level data and geographic boundaries within the study area (U.S. Census Bureau, 2021) were used to define the spatial setting of the model. For the baseline model, push–pull scores for each census tract were defined based on population projections to 2070. A common critique of model-based projections of future conditions, including but not limited to ABMs, is that calibrating models using historical data inherently assumes that future events and consequences will follow historical trends. This is unlikely to be the case in situations with the multi-dimensional uncertainties and complexities inherent to socio-environmental systems. To address this issue, county-level population projections (Hauer & CIESIN, 2020; Hauer, 2019) for the study region were used to develop and calibrate the baseline model. These projections were created by developing autoregressive

models of projected rates of population change based on observed data from 1960–2016 (Hauer & CIESIN, 2020; Hauer, 2019). These projections do not explicitly consider potential environmental influences that could impact population changes in a given area, such as sea-level rise. Thus, these are used as baseline projections for how the population would likely evolve if flooding impacts are not considered, and then this baseline is used as a comparison for different flood scenarios.

Projections from the Shared Socioeconomic Pathway SSP2 were used to calculate an annual linear growth rate. These projections, representing population trends at the county level, were then disaggregated into census tracts within each county based on their respective 2020 population (Figure 2-4). While linear population trends will not be applicable in all cases, a comprehensive analysis using least squares regression of the 2020–2070 projections demonstrated that in all but one county in the study area, at least 74% of the projection could be explained by a linear trend. To convert the county-level growth rates to the individual census, the county growth rate was used as the mean of a normal distribution from which census tract growth rates were sampled, with higher-populated tracts receiving higher percentile values. This process facilitated an effective distribution of the county-level population projections to the census tract level, accounting for variations in population size and growth patterns within and across counties.

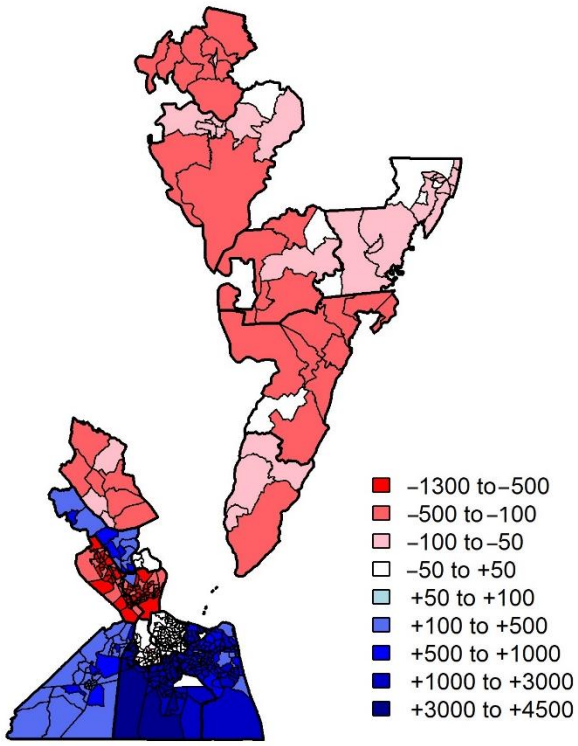


Figure 2-4. Annual population growth rates for the census tracts of the study area.

2.2.2.2. Individuals' Behavior and Decision Making

In this model, the approach taken for the agents to make migration decisions is associated with two groups of factors, including (1) push factors that encourage an agent to leave their current location (e.g., lower inherent desirability or larger area percentage flooded), and (2) pull factors that determine where agents relocate. Agents will evaluate their current location based on pull–push factors to decide if they prefer to migrate. Then, they evaluate other locations again based on pull–push factors to decide where to move. To account for the numerous factors not explicitly represented in the model, a stochastic approach is employed where agents' satisfaction is randomly sampled based on a beta distribution with parameters that vary depending on the push–pull score of the agents' location. A summary of the model parameters and state variables used to represent agent behavior are presented in Table 2-1 and described in more detail in the following sections.

Table 2-1. State variables and global model parameters.

	Name/Symbol	Numeric Domain	Definition and Interpretation
State Variables	Growth Rate (*G*)	[−1300, 4500]	Applies to census tracts; annual rate of population change (persons per year) derived from Hauer projections (Hauer & CIESIN, 2020; Hauer, 2019) for the study area.
	Push–Pull Score (*P*)	(0, 1)	Applies to census tracts in baseline model; derived from state variable *G* and parameter k_{pull}.
	Satisfaction (*S*)	(0, 1)	Applies to agents; random variable sampled from a beta distribution. The beta distribution parameters are calculated from *P* of their current location and global parameters δ and υ
	Flood Extent (*F*)	[0, 1]	Applies to census tracts in the flood model (not included in the baseline model). Describes the percentage of a census tract inundated in each year *y*.
	Flood-influenced Push–Pull (*P**)	(0, 1)	Applies to census tracts in flood model; derived from state variables *P* and *F*.
Global Model Parameters	k_{pull}	(0, 1)	Used to convert census tract growth rates into pull scores. Lower values result in more equal pull scores across census tracts; higher values result in more unequal pull scores. Calibrated value of 0.0006.
	δ	(0, 1)	Used to convert census tract *P* into beta distribution for sampling agent *S*. Higher values result in more discrepancy in *S* scores across high and low *P* census tracts. Calibrated value of 0.2.
	υ	(4, 20)	Used to convert census tract *P* into beta distribution for sampling agent *S*. Higher values result in more variance in *S* scores within a single census tract. Calibrated value of 15.
	MT	(0.1, 0.5)	Move threshold. Agents decide to move if $S < MT$. Higher values result in more agents moving during each model year. Calibrated value of 0.35.

Push–pull scores for each census tract were calculated based on linear rates of population growth or decline. Rates of census tract population change *G,* which ranged from -1300 to 4500 persons per year in the study area, were translated into a pull score P for each census tract *i* ranging between 0 and 1 using a logistic function governed by a parameter, k_{pull} (Equation (2-1)). The logistic function allows transforming the growth

rates into a [0, 1] bounded probability scale, while the use of a k_{pull} parameter allows the process to be replicated in other study areas using the same projection source.

$$P_i = \frac{1}{1 + exp\left(-k_{pull} \times G_i\right)} \qquad (2-1)$$

A k_{pull} value of zero results in a uniform distribution and leads to random movement between census tracts regardless of their projected growth rate (Hauer, 2019). Greater k_{pull} values result in greater differences in the pull scores of high and low growth rate tracts. Thus, increasing k_{pull} results in greater migration flows to high-growth tracts.

Within a given census tract, agents will vary in terms of their desire and ability to move based on numerous factors that are difficult to measure, let alone simulate. This inclination to move, referred to as agent satisfaction *S*, was thus represented using a random variable sampled from a beta distribution, the parameters of which would change based on the pull score *P* of agent's location. The beta distribution is confined to the interval (0, 1) and governed by two shape parameters α and β. Agents' satisfaction scores were modeled using a beta distribution because its interval aligns with the range of pull scores and the distribution offers great flexibility in modeling various shapes of distributions, including symmetrical, skewed, U-shaped, inverted U-shaped, and straight lines. This flexibility was desirable as there was no predetermined assumption about the distribution of satisfaction levels within the study area and we sought to calibrate it with a high level of flexibility. The mean (μ) and variance (var) of beta distribution are formulated by Equation (2-2) and Equation (2-3), respectively:

$$\mu_i = E[S_i] = \frac{\alpha}{\alpha + \beta} \qquad (2-2)$$

$$var[S_i] = \frac{\alpha\beta}{(\alpha + \beta)^2(\alpha + \beta + 1)} \qquad (2-3)$$

To calculate a beta distribution for agents within a census tract based on the pull score *P*, two parameters δ and υ were defined. The parameter δ defines the relationship between tract pull score *P* and the mean μ of the beta distribution as in Equation (2-4).

$$\mu_i = 0.5 + \delta\,(P_i - 0.5); \qquad \forall\delta \in (0,1) \qquad (2-4)$$

A value of δ equal to 0 leads to a mean value for the distribution of agent satisfaction $\mu = 0.5$ regardless of *P*, meaning that tracts' pull score has no influence on expected agent satisfaction. As the value of δ increases, the mean value of the distribution for *S* will move from 0.5 toward the value of *P*, resulting in higher satisfaction scores for agents in high pull-score tracts. A value of δ equal to 1 leads to $\mu = P$ for tract i.

The parameter υ represents the sum of shape parameters α and, β (Equation (2-5)) and indirectly controls the variance of the beta distributions across each census tract. Higher values of υ lead to lower variance in the distribution of satisfaction *S*. Thus, a high value of both δ and υ would result in a high correspondence between census tract pull *P* and agent satisfaction *S*, due to low-variance distributions of *S* centered on different mean values. Figure 2-5 demonstrates the effect of changing δ and υ on the distributions used to sample agent satisfaction within low (*P* = 0.3) and high (*P* = 0.8) pull-score tracts. For the purposes of this model, υ was constrained to the interval (4, 20). Given a set of model

parameters δ and υ and a census tract pull score P, the beta distribution parameters α and β can be calculated as in Equations (2-6) and (2-7):

$$\upsilon = \alpha + \beta; \qquad \forall \upsilon \in (4,20) \tag{2-5}$$

$$\alpha = \upsilon[0.5 + \delta\,(P - 0.5)] \tag{2-6}$$

$$\beta = \upsilon - \alpha \tag{2-7}$$

Since the distribution of satisfaction is unknown and dependent on the study area and perceptions and values of the residing people, this distribution is allowed to have different shapes based on the inherent desirability of a location through parameters δ and υ, while considering the possibility of its being bell-shaped ($\mu = 0.5$) as well. Several factors can contribute to the possibility of considering people's satisfaction distribution as normally distributed, similar to various other phenomena. The first significant factor is the level of complexity and randomness. In complex systems or processes involving a large number of random variables, the cumulative effect tends to produce a normal distribution. This is because the likelihood of extreme values occurring in multiple variables simultaneously decreases, promoting a bell-shaped curve (Sornette, 2007). Second is the effect of statistical aggregation. Many measurements in science, economics, and social sciences involve the aggregation of numerous small, independent influences. When these influences are combined, the resulting distribution often conforms to the normal pattern.

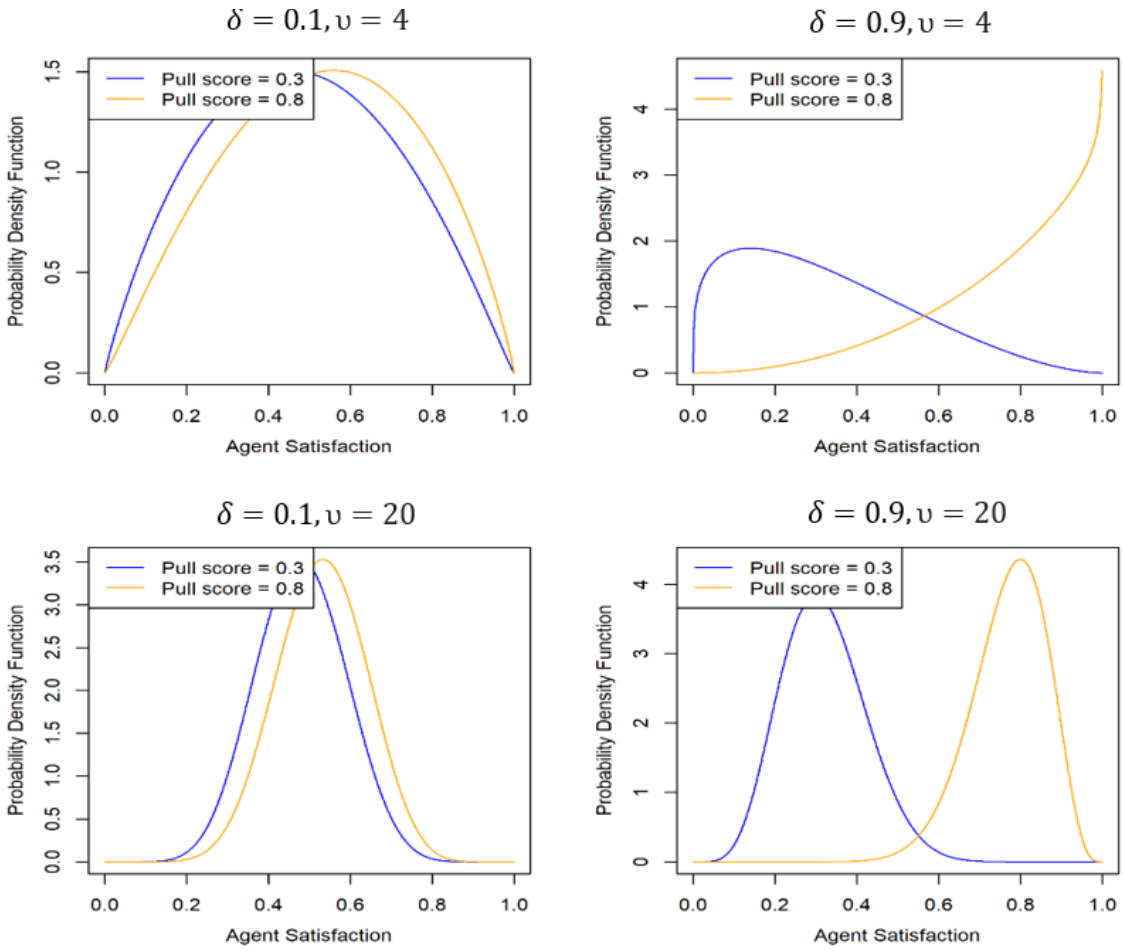


Figure 2-5. Examples of how adjusting global model parameters δ and υ influence the distributions used to sample agent satisfaction in high pull-score tracts and low pull-score tracts.

At each time step of simulating pull–push effects, agents' satisfaction scores are sampled from the associated distributions. A threshold was considered for migration decisions to be made, with which agents' satisfaction scores S were compared. If their satisfaction was less than this move threshold MT, they decided to migrate. Agents' decisions about their migration destination were based on the pull score P of the tracts, meaning that tracts with higher pull scores are more likely to be migrants' destinations. Agents did so, using NetLogo's built-in weighted random draw method in destination selection, where the likelihood of a census tract being selected as a destination is directly

proportional to that census tract's pull score. In summary, numeric ranges were established as follows: tract-level *G* scores in our study area varied from −1300 to 4500. *P*, *S*, *F*, and *P** are assigned values within the range of 0 to 1 for standardization purposes. The parameter δ modifies the mean of the satisfaction distribution based on P, resulting in a mean value between 0.5 and *P*. Additionally, the parameter υ adjusts the sum of shape parameters in the satisfaction distribution, with a specified range (4, 20). This range is selected to account for distribution variances, leading to a diverse array of distribution shapes. *MT* is a level to which satisfaction score is compared, which is set to values below 0.5 to mitigate excessive migration flows.

2.2.2.3. Baseline Model Calibration

The global parameters k_{pull}, δ, υ, and MT were calibrated by systematically testing all combinations of values for parameters within the bounds presented in Table 1. For each combination of parameter values, the relative root mean square error (Relative RMSE) between the modeled and projected population (Hauer & CIESIN, 2020; Hauer, 2019) in each simulation year was calculated as in Equation (2-8).

$$Relative\ RMSE = \sqrt{\frac{\sum_{t=1}^{T}\sum_{y=1}^{Y}\left(\frac{MPop_{t,y} - SPop_{t,y}}{SPop_{t,y}}\right)^2}{T \times Y}} \qquad (2-8)$$

where $MPop_{t,y}$ is county's simulated population obtained from the model in year y and tract t, $SPop_{t,y}$ is county's projected population projection for the corresponding year and tract, T is the total number of census tracts in the study area, and Y is the number of simulation years.

The ABM model was also evaluated to be consistent with nationwide statistics on local and regional scale mobility developed by the Harvard Joint Center for Housing Studies (Frost, 2020). Based on this data, 13% of the U.S. population moves each year. Of these moves, 65% are within the same county, and 17% are towards different counties, but in the same state. For the baseline model, the combination of parameter values with the lowest RMSE was selected that also had a movement percentage within 2% of the HJCHS estimate of 10.7%.

2.2.3. Flood-Informed Model

2.2.3.1. Physical Flood Inundation Modeling

To represent potential flood events over the period of simulation, statistical coastal flood hazard data were utilized from the U.S. Army Corps of Engineers' 2015 North Atlantic Coast Comprehensive Study (Cialone et al., 2015; Nadal-Caraballo et al., 2015). The NACCS statistics represent the combined flood hazards from Nor'easters, tropical storms, and hurricanes and include the influence of astronomical tides. The NACCS statistical values are reported for present-day sea level at return periods (one over annual exceedance probability) ranging from 1 to 10,000 years. The NACCS probabilistic surge hazard methodology is consistent with the methodology now adopted by FEMA for establishing Flood Insurance Rate Maps. The NACCS present-day flood statistics were used to estimate a percentage of each census tract inundated for the floods with return periods of 1, 2, 5, 10, 20, 50, 100, 200, and 500 years. These percentage inundation values were linearly interpolated to yield estimates of inundation percentages for storms with return periods between the nine return periods listed above. This study assumes flooding scenarios under current sea levels as an initial step in demonstrating how flood impacts can

be incorporated into a baseline ABM calibrated to long-term projections that do not explicitly consider sea-level rise or other environmental factors. This general approach could accommodate multiple methods for representing future flood scenarios under different assumptions about sea-level rise informed by different data sources. Under different SLR projections, it is expected that the proposed overall approach would not change but that the increasing intensity of floods may intensify migration flows and patterns observed here.

While the NACCS hazard statistics are presented in terms of the return period, the ABM required sequences of flood events through time to simulate the impact that repeated flood events of differing intensities could have on migration behavior. To address this requirement, four 50-year flood storylines were generated representing different long-term scenarios of flood occurrence. To create these storylines, uniform distribution was used to generate 300 sequences of fifty (the number of years in our simulation) random numbers between 0 and 1 and inverted them to obtain sequences of flood return periods. For example, a random number of 0.5 would equate to a two-year flood occurring, whereas a random number of 0.1 would equate to a 10-year flood occurring. Then, a search was conducted across this ensemble of statistically feasible 50-year flood sequences to identify sequences that demonstrated behavior consistent with four conceptual storylines that represent feasible long-term sequences of flood events in the region:

- Storyline 1: Frequent small floods (with a return period of 2–10 years).

- Storyline 2: Frequent small floods and one severe storm (with a return period of 100 years or more) occurring early in the horizon (within the first 15 years).

- Storyline 3: Frequent small floods, one large storm (with a return period of 10–100 years), and two severe storms (with a return period of 100 years or more) occurring late in the horizon (within the last 15 years).

- Storyline 4: Frequent low-intensity flooding, one large storm (with a return period of 10–100 years), and three severe storms (with a return period of 100 years or more).

Figure 2-6 illustrates these storylines. Values represented in y-axes are log-transformed return periods to better demonstrate the minor flooding events. Because the time step of the ABM is yearly, these 50-year flood storylines represent a single instance of flooding per year, and thus cannot account for the effect of multiple flood events in a year. While this may underestimate the impact of minor flooding with a return interval of less than one year, as well as the occurrence of multiple large floods in a single year, these details could be incorporated into more sophisticated flood storylines in further work.

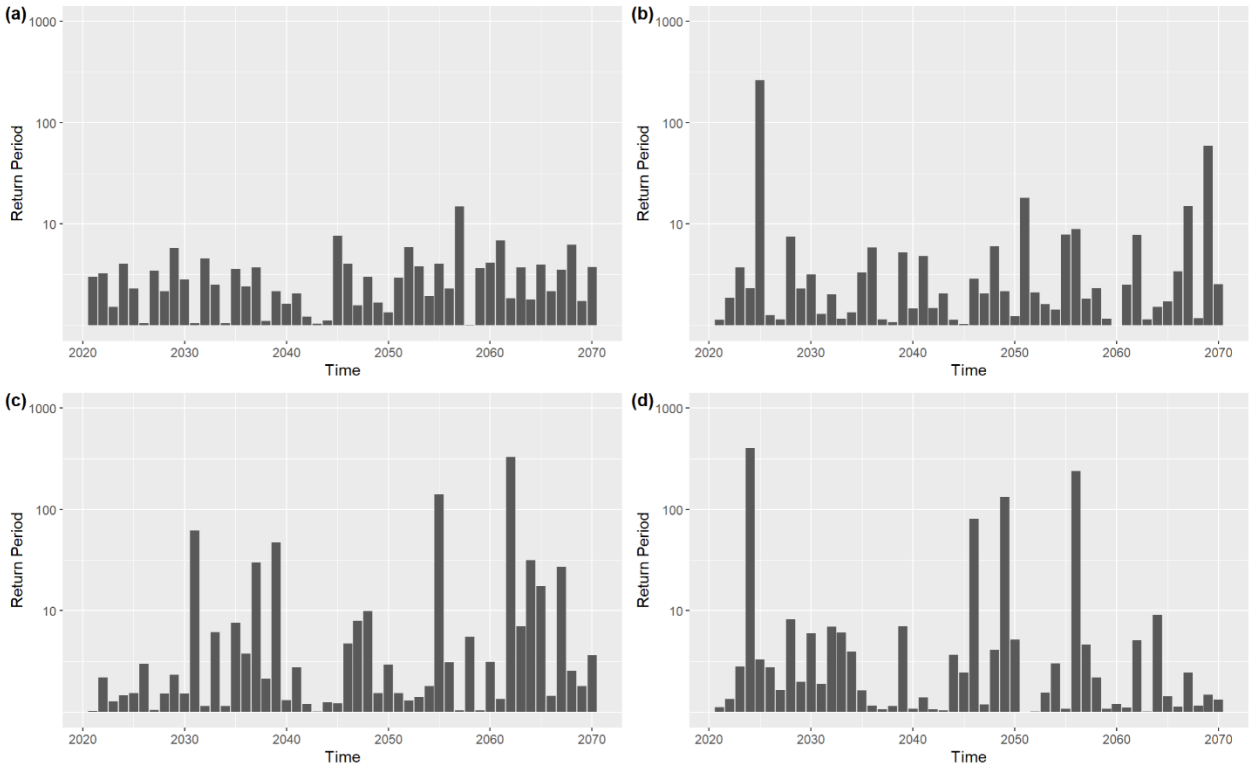


Figure 2-6. Conceptual storylines: (a) minor flooding, (b) early severe flooding event, (c) late severe flooding event, (d) two severe flooding events.

2.2.3.2. Adjustment of Decision Rules

As a simple initial decision rule structure, it was assumed that agents would make decisions about moving from or staying in a census tract using a weighted average of the pull–push score (P) from the baseline model and the severity of flooding they experience. Flooding severity was represented by the percentage of the census tract that is inundated in a given year (F). While the pull–push score is calculated the same as the way it was in the baseline model (Equation (2-1)), flood-influenced push–pull score for census tract i is updated as Equation (2-9).

$$P^*{}_i = (1 - W_i) \times P_i + W_i \times (1 - F_i) \qquad (2-9)$$

The weight W represents the relative importance of flooding in comparison with the census tract's baseline desirability. Different values of W were tested to see how the migration response would differ from baseline model projections as flood impacts become an increasing influence on the decision-making process.

${P^*}_i$ is then used to obtain the mean of satisfaction beta distribution and its shape parameters. From this distribution, S_i is calculated for its residing agents to compare with the threshold MT as in Equations (2-2)–(2-7) but calculated as a function of ${P^*}_i$.

$$\mu_i = 0.5 + \delta({P^*}_i - 0.5); \qquad \forall \delta \in (0,1) \qquad (2-10)$$

It should be mentioned that agents' decisions about migration destinations are made only based on the pull–push score of destination options. This assumes that agents do not have a detailed awareness of flood impacts or vulnerability in their destination. However, this advancement could be incorporated into our general framework in future research.

Comprehensive description of the proposed agent-based model according to ODD + D (Overview, Design, Details + Decisions) (Müller et al., 2013) can be found in the Appendix.

2.2.3.3. Model Simulations and Analysis

Because the model leverages a stochastic approach to modeling agent behavior, 50 simulations were conducted for each weighting scheme, using the mean and standard deviation of which to characterize average population projections under different scenarios and variability stemming from stochastic behavior. The performance of the flood-informed ABM was compared with the baseline model by calculating the deviation of each stochastic

iteration under the weight of W from the average of stochastic iterations in the baseline scenario. Equation (2-11) shows baseline population projections against which the flood-informed models were compared:

$$\overline{Pop}_{t,y,0} = \frac{\sum_{n=1}^{N}\left(Pop_{t,y,n,0}\right)}{N} \qquad (2-11)$$

where N is the number of stochastic iterations, equal to 50. The root mean square of relative deviations from the baseline was then used to characterize spatial and temporal variations:

$$RD_{w,n} = \sqrt{\frac{\sum_{y=1}^{Y}\sum_{t=1}^{T}\left(\frac{Pop_{t,y,n,W} - \overline{Pop}_{t,y,0}}{\overline{Pop}_{t,y,0}}\right)^2}{T \times Y}} \qquad (2-12)$$

where $RD_{W,n}$ is the relative deviation from baseline of stochastic iteration n under flood weight of W, $Pop_{t,y,n,W}$ is the population of stochastic iteration n in tract t and year y under flood weight of W, Y is the number of years simulated (50), and T is the total number of census tracts (430).

To measure temporal variability across the study area, the relative deviations were aggregated over all census tracts in each simulated year under each storyline.

$$ARD_{w,y} = \sqrt{\frac{\sum_{t=1}^{T}\left(\frac{Pop_{t,y,W} - \overline{Pop}_{t,y,0}}{\overline{Pop}_{t,y,0}}\right)^2}{T}} \qquad (2-13)$$

where $ARD_{W,y}$ is the annual relative deviation for year y under flood weight of W from baseline averaged across all census tracts. Additionally, a measure was defined for relative deviation in each census tract in year y. This measure was used to show spatial variation of deviation in the study area for a specific year (2070) and flood weight.

$$TRD_{w,t,y} = \frac{Pop_{t,y,W} - \overline{Pop}_{t,y,0}}{\overline{Pop}_{t,y,0}} \qquad (2-14)$$

$TRD_{W,t,y}$ is the tract-level relative population deviation for tract t from year 2021 through year y under flood weight of W.

2.3. Results

2.3.1. Baseline Model Calibration

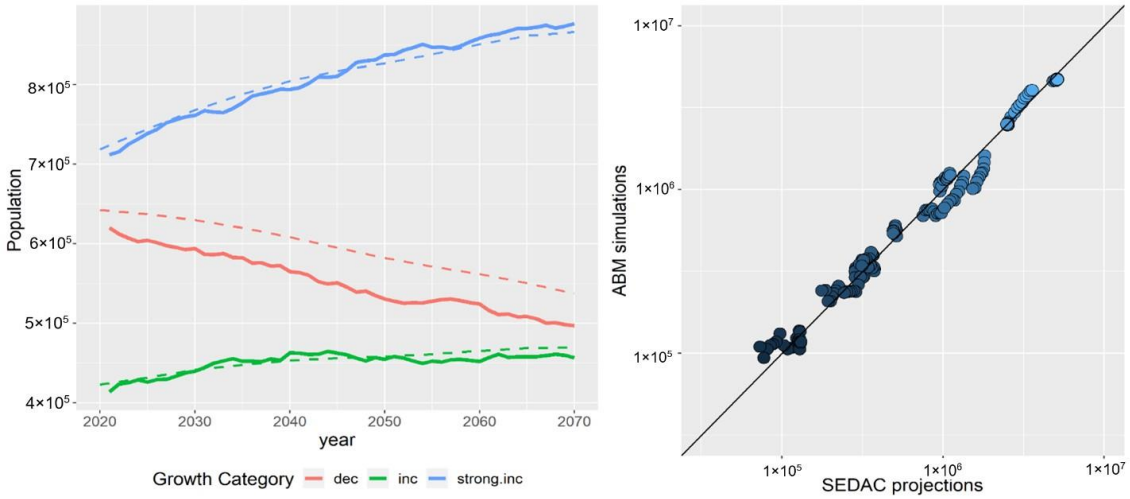
The baseline model was calibrated by running 256 simulations in a full factorial arrangement of the four global model parameters (k_{pull}, δ, υ, and MT) to identify parameter values that best replicated population projections (Hauer & CIESIN, 2020; Hauer, 2019) and broad-scale movement patterns. The ranges of values tested for each parameter and the final results of calibration are mentioned in Table 2-2.

Table 2-2. Calibration parameters, ranges, and calibrated values.

Parameter	Minimum Value	Maximum Value	Increment	Calibrated Value
$\boldsymbol{k_{pull}}$	0.0002	0.0008	0.0001	0.0006
$\boldsymbol{\delta}$	0.2	0.8	0.2	0.2
$\boldsymbol{\varepsilon}$	10	16	2	12
$\boldsymbol{MT}$	0.25	0.40	0.05	0.35

The comparison between the projected and simulated populations resulted in a minimum relative RMSE of 0.17 in 2070 for the whole study area and an average of 12.6% of agents migrating each year. To ensure that the calibrated ABM was accurately replicating population trajectories, projected and modeled population trajectories were compared for counties with high growth (G > 700 persons per year), moderate growth (0 < G < 700 persons per year) and declining (G < 0 persons per year) populations. This comparison

along with a scatter plot of projected versus modeled population across all census tracts and years is presented in Figure 2-7. While the model reliably replicates population trajectories in the counties experiencing rapid growth (blue lines) and moderate growth (green lines), its performance is poorer in counties experiencing population decline. This is possibly due to the format of the logistic function and its lower bound being set to zero for the push–pull score of negative-growth counties; refining this process would be a valuable area for further research. Despite this limitation, these results indicate that the bottom-up stochastic decision rules of the ABM can reasonably replicate top-down projections from (Hauer & CIESIN, 2020; Hauer, 2019).



(a) (b)

Figure 2-7. (a) Study area population simulated by the ABM (solid lines) in comparison with Hauer's projections (dashed lines) through time. (b) Scatter plot of ABM population simulation results for the study area versus top-bottom population projections

2.3.2. Flood-Informed System Behavior

In order to see how flooding impacts the system behavior given the uncertainty about the relative role of flood impacts in decision making, multiple values of flood weight were simulated and compared with baseline model projections ($W = 0$). Figure 2-8 demonstrates the difference between model projections under each flood storyline and the

baseline scenario for different values of flood weight. The starting point of all the lines is the same, which is our baseline scenario where the weight of flooding is equal to zero. As the weight applied to the census tract-flooded extent in calculating the push–pull score (Equation (2-10)) increases, different types of relative divergence from the baseline model emerge from individual-level decision making. Although it is unlikely that high weights of flooding represent realistic preferences, the purpose here is to explore how the system responds under different assumptions about flood importance.

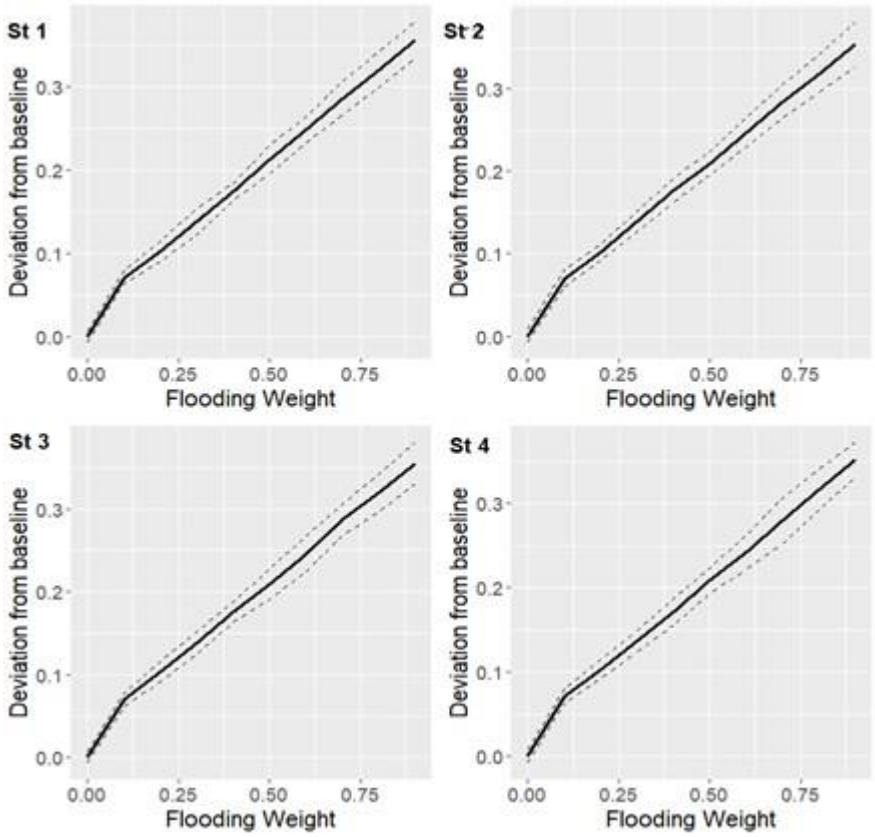


Figure 2-8. Relative deviation from baseline scenario under different storylines and values of flood weight. Solid lines represent the average deviation across 50 stochastic iterations with the range *(μ-2σ, μ+2σ)* illustrated by dashed lines

The response under different storylines follows a trend similar to deviation under nuisance flooding of storyline 1, and system behavior shows sensitivity to severe flooding events with subtle bumps. Two severe flooding events in storyline 4 do not show a significant synergic effect. This could potentially be due to the fact that highly vulnerable areas to flooding have already experienced high flows of out-migration when another large flooding event occurs late in the time horizon. It is possible that considering experience memory for agents results in more different behavior under different flood sequences, which needs to be further investigated; however, with the simple assumptions in the decision rules of this model, the system's migration response to different flood weighting schemes is not highly sensitive to the severity of flooding events.

The relative deviation measures formulated in Equations (2-13) and (2-14) only represent the degree to which flood-informed model simulations diverge from the baseline model but do not reveal the direction of population deviations or the specific years or locations where these deviations occur. To spatially characterize flood-induced changes in population projections, Figure 2-9 presents a tract-level map of relative deviations from baseline to identify areas more prone to in- and out-migration flows. The maps in Figure 2-9 demonstrate this spatial variation in 2070 under four flood storylines and two different flood weight values. In general, spatial patterns of population deviation were consistent across flooding storylines. The locations that lose population when flood impacts are incorporated, indicated in shades of red, mostly include urban areas. However, the locations that gain population include both urban and rural settings. Comparing migration maps with flood exposure in the study area also reveals that highly inundated areas do not necessarily lose population. Thus, the migration flow cannot be fully explained by urban/rural

classifications or flooding alone. In a small number of census tracts, the direction of change relative to baseline projections changes from positive to negative depending on the flood storyline. This demonstrates the complexity of climate migration in the face of multiple intervening factors affecting general migration patterns. The increase in the flooding factor from 0.3 to 0.6 triggers stronger migration flows and more change in population distribution highlighted as relative deviation values from baseline in Figure 2-9, but the general spatial patterns are consistent.

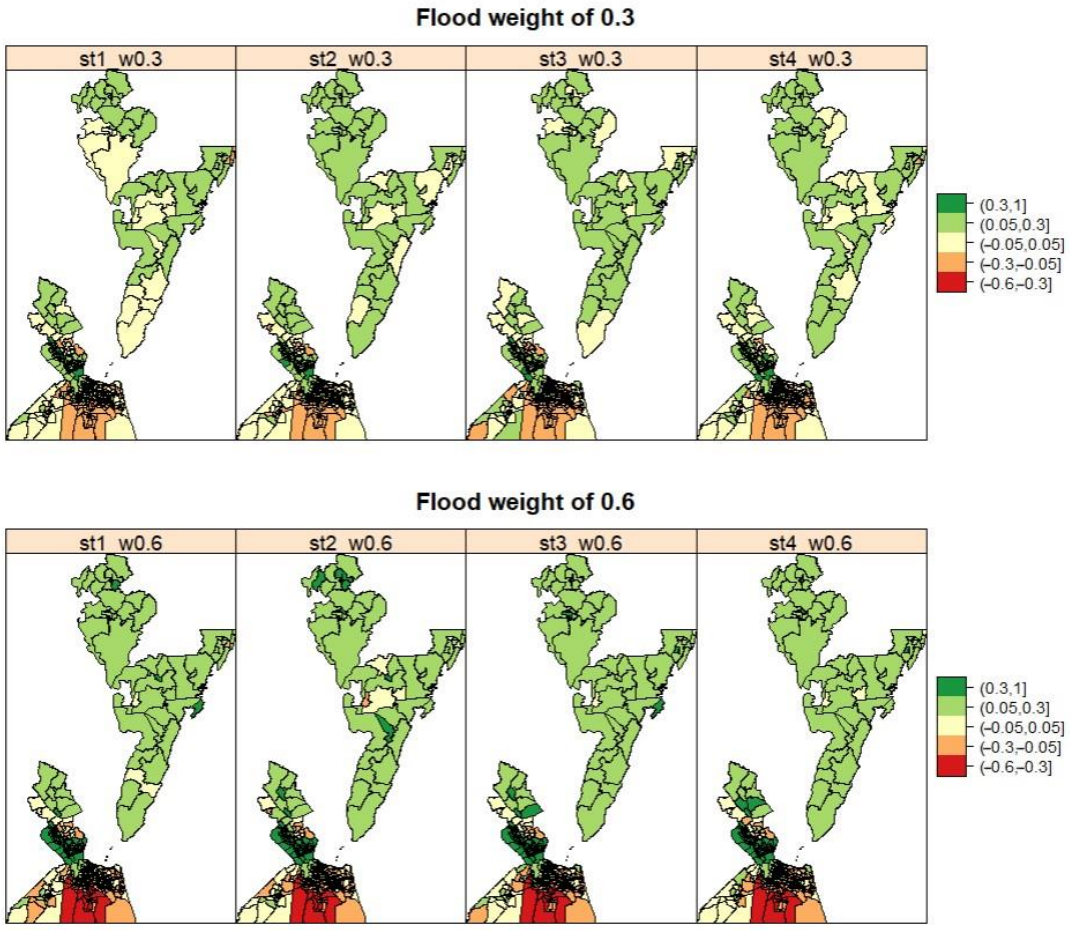


Figure 2-9. Spatial variation of relative population deviation from baseline scenario in 2070 under different storylines with flooding weight values of 0.3 and 0.6.

The time series of relative deviation across the study area are illustrated in Figure 2-10. Major flood events (>200 years) occur in 2025 in storyline 2, 2062 in storyline 3, and both 2024 and 2056 in storyline 4. Relative deviation time series shows a nonlinear increase through time, which is enhanced under a higher weight of flooding in agent decision making. When the flood weight W equals 0.3, storyline 1 (nuisance flooding with no extreme events) results in a significantly lower deviation from baseline than the storylines that do entail large flood events. However, at higher values of W, the deviation through time in the nuisance flooding storyline is more aligned with the others. In addition, the gradient of change in deviation from baseline in both graphs shows a decreasing trend through time for all storylines. This is notable because large flood events included in storylines 2, 3, and 4 do not result in a significant spike in deviation, and the gradient of change is almost similar over different storylines. This may stem from the fact that all storylines include frequent low-intensity flooding that occurs in most years and acts as a chronic, long-term influence on location push–pull scores that outweighs the impact of infrequent large floods. The decreasing gradient of change could indicate that people have already moved from areas highly exposed to flooding to safer locations toward the end of the time horizon. The same time steps ahead do not lead to high out-migration flows anymore as in the past.

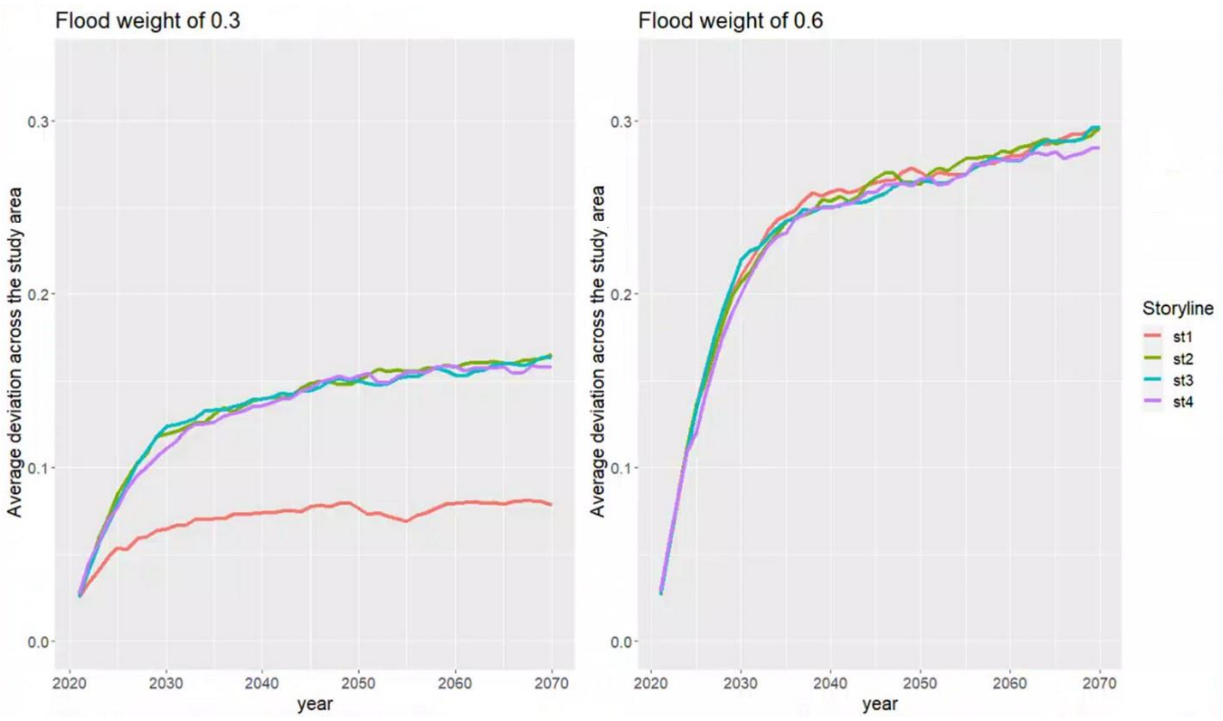


Figure 2-10. Temporal variation of relative population deviation from baseline scenario under different storylines with flooding weight values of 0.3 and 0.6.

2.4. Discussion

This study presents a novel stochastic ABM of long-term changes in local-scale migration patterns influenced by repeated coastal flooding. The ABM framework can enrich systems' understanding and modeling by incorporating bounded rationality, heterogeneity, interactions, evolutionary learning, and out-of-equilibrium dynamics of human behavior in a spatially explicit model (Filatova et al., 2013). Moreover, bottom-up approaches such as ABM can support a more detailed exploration of micro-level displacement patterns that usually emerge from differing individual decisions, but are masked through behavior generalization approaches commonly extracted from large amounts of data in top-bottom data-driven models.

This research demonstrates that bottom-up individual decision rules can replicate externally derived population projections, providing a baseline for incorporating local-scale factors such as nuisance and extreme flooding. This provides an alternative to models calibrated to historic population changes, and allows an understanding of how flooding could impact future population shifts relative to projections that do not include this hazard. This approach thus complements top-down data-driven models that have been widely applied to investigate migration response to climatic events (e.g. Davis et al., 2018; M. E. Hauer, 2017; M. E. Hauer et al., 2016; Robinson et al., 2020). Top-down approaches primarily rely on historic patterns for calibration. However, inherent uncertainties, the significant number of interactive subsystems and feedback loops (Aral, 2020) constrain the reliability of model projections. This is particularly true in the case of models that incorporate human behavior, as research has shown that structural decisions in terms of how human behavior is modeled can result in significant uncertainty that greatly exceeds uncertainty stemming from model parameterization and calibration (Yoon et al., 2023). Accordingly, variable human decision making suggests that good historical performance does not necessarily guarantee future accuracy (Barnsley, 2007), especially when migration projection is triggered by climate change over a long-term time horizon.

In the face of acknowledged uncertainties, it is helpful to be explicit about what a model (agent-based or otherwise) can and cannot tell us. Aligning with the existing literature, it is asserted that models of highly complex socio-environmental systems are unlikely to be useful as consolidative models that can be used in a predictive sense, as the significant uncertainty and unpredictability in human behavior and its drivers will undermine the reliability of any prediction derived from them (Moallemi et al., 2020).

As an alternative, researchers have proposed adopting an exploratory modeling approach that interprets models as computational experiments that reveal system-level outcomes if different assumptions embedded in the model are correct (Kwakkel & Haasnoot, 2019). Adopting this perspective, validating the model against best available population projections based on historical data provides a mechanism for assessing whether bottom-up decision rules can replicate these projections which were derived based on an autoregressive approach. Once this has been confirmed, this serves as a mechanism for experimentally evaluating how the introduction of a new factor (flooding) results in outcomes that diverge from the projections.

This approach also acts as a practical shortcut, circumventing potential uncertainties and intricacies linked to modeling individual population components such as fertility and mortality. This approach is especially beneficial given the frequent challenges associated with obtaining the required data. Another notable aspect of the population projection dataset (Hauer & CIESIN, 2020; Hauer, 2019) specifically used in this study is its capacity to offer projections under five socio-economic pathways (SSPs), each contributing to significant variations in geographic growth patterns. Beyond enhancing our comprehension of demographic shifts in small areas of the United States, utilizing this dataset amplifies the method's adaptability, making it applicable to diverse regions with varying data availability. Additionally, it facilitates the consideration of different SSPs in the planning process, further enriching the method's overall flexibility.

The increasing encounters with recurrent flooding and recent major coastal disasters are gradually changing attitudes toward relocation among residents and other stakeholders (Bukvic, 2015). A survey of coastal urban residents found that 87% would

consider relocating now or in the future due to flooding (Bukvic & Barnett, 2023). Understanding the impact this flood-driven mobility will have on municipalities and their adaptation capacities requires tools that can accurately replicate population-level behavior at fine spatial scales. Moreover, adaptive policy frameworks need to reflect behavioral processes and incentives at the individual, household, community, and organizational levels (Moench, 2010). Bottom-up approaches that build upon the proposed framework can thus serve as a valuable tool for evaluating and comparing diverse coastal adaptation strategies, of course within the constraints of the modeled representation of coastal community drivers and processes (Mills et al., 2021). For instance, by combining the proposed framework with simulations of coastal adaptation, (e.g., Y. Han & Peng, 2019), ABM can facilitate more informed policy-making processes in response to coastal flooding.

The spatial representation and diversity in the environmental context in ABMs are often needed to capture spatial heterogeneity of inputs and outputs. However, existing research mostly chooses the spatial representation based on the available data, without systematically justifying their spatial resolution or extent (Filatova et al., 2013). In this study, a detailed spatial scale (census tract) was intentionally selected and other data sources were rescaled to this level to discern potential implications of local-scale relocation. Detailed spatial representation allows for the incorporation of small-scale variations in flooding and tracking relocations over short distances, which can be stimulated by different migration purposes represented by stochastic agent behavior. Although census block level as the unit of analysis could provide a more spatially detailed representation of relocation behavior in comparison with census tract, their use over the

longer term presents methodological issues: first, census blocks' boundaries frequently change after the census is conducted each decade based on changes in block population. While this can also occur with census tracts, it happens much less frequently. For instance, while in the study area comprising 16 counties, the initial number of census tracts was 430 according to census 2010; this count increased to 481 in census 2020 due to boundary adjustments. Second, the proposed model was constructed as a framework adaptable for the inclusion of social and economic attributes, most of which are accessible at the census tract level. The census tract level was chosen due to data availability constraints, as several of these variables are not readily accessible at the census block level. Given these challenges, the loss in granularity from using the census tract level does not detract from the objectives of this study, which aims to conceptualize a bottom-up model that integrates population movement at the municipal level in response to flood events. Thus, instead of a high-resolution track of each household's movement, insights into potential changes in population distribution were gained. This information can be beneficial for the future use of policymakers in flood-prone municipalities, while keeping the required computational effort to apply this framework at a feasible level. Note that the use of census tract spatial scale in previous research on population and policy with ABM (Burger et al., 2017; Khansari et al., 2017) increases the meaningfulness of this choice of spatial unit for the purpose of informing policy decisions at the municipality scale. Nevertheless, the incorporation of finer-scale spatial processes and data is a potentially valuable area of future research.

This stochastic ABM framework is presented as a first step in modeling individual migration decisions at local scales. As such, a simple representation of human behavior

was adopted that could accommodate a more sophisticated representation of migration decision making in many ways. The current framework assumes that migration patterns in the study area follow the estimations provided by HJCHS as well as our suggested heuristic decision rules. Nonetheless, it is worth acknowledging that ABM frequently involves simplifications and assumptions about intricate social and ecological processes, potentially leading to an oversimplification of real-world dynamics and constraining the model's precision (Lawyer et al., 2023). The structural model choices made regarding the representation of human behavior significantly impact the credibility and validity of modeling insights. Therefore, evaluating the implications of these choices should be regarded as a crucial aspect of future modeling efforts (Yoon et al., 2023). Primary data collection such as surveys can further complement the ABM model by incorporating additional variables and yielding more robust results (Bilsborrow & Henry, 2012). Individual-level details of ABM in the context of climate-related migration decisions could be further advanced by incorporating behavioral psychology and decision theories in the computational modeling of climate migration phenomena. This approach also provides a ground to explore relocation complexities and categories of communities' response and comparison of the results with previously suggested theoretical classifications such as forced migrants, trapped population, and adaptive migrants (R. McLeman, 2018a).

The general framework presented here could also provide a foundation for the testing of different theoretical assumptions in a spatially explicit data-informed manner. For instance, how will migration response vary under sudden and unexpected hazards versus repeated nuisance events, and how does learning from experience and exposure to flood hazards at the individual level influence migration decisions? The proposed approach

could also accommodate a more detailed and comprehensive representation of physical hazards. For example, our approach could incorporate an explicit representation of sea-level rise to simulate migration response under different SLR projections, where greater flood intensities could exacerbate the migration flows. Another way to consider a more precise representation of the physical system is to capture heterogeneity in land use within a given census tract. Assuming that high levels of inundation result in a high impact of flooding on residents could result in overestimating the flood impact in areas with low population density but high percent inundated area. An alternative to solve this issue can be updating the model using downscaled population projections (e.g., Gao, 2017; Jones et al., 2020) in rural areas to overcome this overestimation. Finally, the proposed ABM approach rests on the assumption that the projections used for the baseline model are a reasonable representation of feasible future conditions. As such, the model can be tested against multiple population projections to account for uncertainty. There is a notable capacity in social system representation as well, which has not yet been of much consideration in population–environment assessments (McLeman, 2018a).

Ultimately, the modeling framework presented here could help advance the incorporation of coupled social, organizational, biophysical, and hazard systems in stochastic models that include location-specific interactions and feedback loops. Our current study focused on establishing a simple bottom-up framework with a forward-looking calibration step that can serve as a flexible foundation for future research that incorporates additional complexity, such as incorporating interactions among individuals. This enhances the model's ability to capture the complex interplay between individuals and their environment. In the context of flood-induced migration, the role of social

networks, community norms, information sharing, and collective decision making in shaping migration decisions can lead to cascading effects throughout the community. The role of considering interactions extends beyond the social system and also encompasses the interactions between the social and physical hazard systems. For example, one of the main reasons for land subsidence in coastal Virginia and Maryland is groundwater withdrawal. Localities can affect the extent of groundwater extraction, and accordingly, sea-level rise and migration, showing a bidirectional relationship in this area. An additional level of complexity emerges from the intricate interactions and feedback loops between the social system and infrastructure and governance factors. An example is between migration, tax revenues and the protection services provided by municipalities. The modeling approach presented here could serve as a means for incorporating more biophysical, political, and organizational components to yield more comprehensive results.

2.5. Conclusions

This study presents a stochastic ABM of flood-induced migration flows to simulate the role of individual decisions in relocation outcomes at municipal scales important for coastal adaptation. This approach aims to replicate the behavior of data-driven population projections with stochastic decision-making processes to simulate the impacts of repetitive flooding on a well-validated baseline projection. The proposed method integrates physical flood exposure data with a bottom-up simulated social system and provides detailed population projections in response to repeated flooding for 16 counties in coastal Virginia and Maryland, including both urban and rural settings. The results confirm the ability of ABM to match data-driven approaches even using simple heuristic decision rules. While simple heuristic rules were adopted in this study, the general framework can serve as a

foundation for the incorporation of more complex representations of behavior, policy, and physical risk to investigate complex drivers of migration in the context of climate risks. This stochastic agent-based modeling approach can be applied at municipal scales to integrate a wide range of information, observations, and theories into projections that can inform municipal adaptation decision making in the future due to recurrent and extreme flooding. This study suggests multiple directions for future research. These include the implication of modeling choices for the emergent migration patterns, as well as migration response under different theoretical assumptions in different model components. For example, an assessment of the physical hazard system could evaluate the impact of different types of flood scenarios, such as infrequent extremes versus nuisance events and different SLR projections. An evaluation of the social system could include incorporating behavioral psychology, decision theories, individuals' interactions, and social networks. The impact of human–environment interdependencies would benefit from considering alternative adaptation measures, incorporating primary data, exploring different community-level response categories, and the effect of learning from past experiences being exposed to flooding. Finally, coupling the proposed framework with organizational and biophysical systems could advance methods for systems integration within ABMs.

3. Exploring How Socio-Demographic Heterogeneity in Relocation Decisions Impacts Municipalities under Sea-Level Rise using a Stochastic Agent-based Approach

Zahra Nourali, Julie E. Shortridge, Anamaria Bukvic, Yang Shao and Jennifer L. Irish

Abstract

Rising sea levels prompt diverse responses from coastal communities, one of which is relocating from flood-prone areas which can significantly impact the distribution of exposed population. As individuals' relocation decisions are not homogenous and are likely to differ across demographic groups, emerging collective patterns may unexpectedly alter risk dynamics for various groups on a larger scale. To account for individuals' differences in relocation decisions, we examine the influence of households' demographic characteristics on aggregate migration patterns by the year 2070 in response to coastal flooding and sea level rise. Employing an agent-based approach, our study integrates census data on household age, income level, and urban/rural settings with survey data that captures the diverse responses of residents to sea level rise-induced risks. Through the analysis of three scenarios of projected sea level rise, we illustrate how these factors influence individuals' diverse decisions, resulting in emergent aggregate behaviors. The findings of this study demonstrate that accounting for diversity in relocation decision making in our study area results in fewer low-income households, but higher elderly proportion in highly flood prone areas at the end of simulation time horizon. Flood prone areas are clustered into groups of wealthy and less affluent compared to simulations where decision making diversity is unaccounted for.

Keywords: sea-level rise, human migration, agent-based modeling, socioeconomic factors, municipal adaptation

3.1. Introduction

Increasing sea levels pose a significant threat to coastal areas worldwide, with approximately 680 million individuals living in coastal regions at risk of flooding. This substantial figure, projected to exceed one billion by 2050 (IPCC, 2022), has sparked concerns and placed coastal adaptation at the forefront of climate resilience efforts. While there are various methods of adaptation to sea level rise (SLR) available and many communities are already planning to adapt in their current locations, the concept of human displacement has often been portrayed as an inevitable long-term outcome, suggesting that climate-related risks such as rising sea levels compel individuals to leave their coastal residences in many flood-prone areas (Adade et al., 2023). Relocation has the potential to avert fatalities or injuries in situations where it's challenging to manage risks through structural means, and it can also lead to cost savings when faced with recurring damages (R. W. Perry & Lindell, 1997). However, it also has the potential to disconnect people from their homes and communities, severing longstanding social ties and cultural bonds (Marino, 2018). It can result in a lower tax base and difficulties in providing government services for both the areas left behind and the new receiving communities. Furthermore, the process of relocation may disproportionately impact vulnerable populations who may not have the resources to effectively relocate, leaving them "stranded" in flood-prone locations (Bukvic & Barnett, 2023).

A holistic perspective on coastal adaptation to SLR risks should recognize that community vulnerability is influenced by not only public regulations, but also decisions and actions of individuals within a vulnerable community (Javeline & Kijewski-Correa, 2019). In this context, socio-demographic factors play a crucial role in shaping the decisions and outcomes of relocation plans (McMichael et al., 2020). Various characteristics of a population, such as age, income, and rural or urban residence, can significantly influence the feasibility, efficacy, and acceptance of relocation options. For instance, age demographics may determine the vulnerability of certain groups to the impacts of SLR, while income levels can affect access to resources necessary for relocation or other adaptive measures. Understanding the influence of demographics is key to understanding how SLR and flooding will impact coastal communities and developing adaptive strategies that are equitable, inclusive, and socially and economically viable.

The intricate dynamics between flood hazards and people occur across various spatial, temporal, and organizational scales, and their outcomes are influenced by factors such as imperfect information, bounded rationality, and the adaptive nature of human decision-making (Zhuo & Han, 2020). While the literature on environmentally-induced migration acknowledges the significance of demographic factors in modeling individual-level migration behavior and decision-making (Smith et al., 2010), the variation in migration behavior across distinct social groups remains an underexplored area of research (R. McLeman & Smit, 2006). It remains difficult to differentiate exposures and responses to environmental variables among population subgroups (Piguet, 2022). Analyzing relocation behavior within these complexities requires an approach that

facilitates the deep coupling of the human and flood subsystems across multiple dimensions and scales (An, 2012). Agent-Based Modeling (ABM) has gained prominence for simulating the decision-making and behavioral processes of humans in complex socio-environmental systems (Bousquet & Le Page, 2004). By accounting for factors such as heterogeneity, interactions, feedback loops, nonlinearity, and thresholds, ABMs have the capability to generate intricate behavioral patterns that cannot be accurately predicted solely by aggregating the behaviors of individual agents (Bonabeau, 2002; Dawson et al., 2011).

Demographic characteristics in flood-induced relocation have been incorporated into ABMs for adaptation assessment in different ways. Hassani-Mahmooei & Parris (2012) integrated attributes like family ties, land ownership status, wealth levels, education, and employment status to enhance the likelihood of migration in response to climatic shocks including floods within their model. Focused on rural settings, Trinh & Munro (2023) tailored a choice experiment within an ABM to capture the diversity of preferences among farmer communities facing climate-induced challenges such as flood frequency. Walsh et al. (2013) utilized a comprehensive set of variables, including village-level factors, household characteristics, and individual-level attributes, to simulate migration decisions in response to changing climate scenarios including floods. Koning & Filatova (2020) developed an ABM grounded in economic rationality to explore outmigration and gentrification in flood-exposed areas. Han & Peng, (2019) simulated adaptive behaviors based on households' risk perceptions, insurance coverage, and local flood mitigation strategies. Taken as a whole, these previous studies demonstrate how socio-economic characteristics can interact with physical hazards to influence migration

pattern. Zhang et al. (2024) simulated human response to flash flood warnings using opinion dynamics model to incorporate socio-demographic attributes such as age and gender and psychological processes, e.g. perceived safety and fear in human evacuation behavior under flash flooding. However, these studies are either event-based or largely based on an economic approach informed by housing market dynamics or individual incomes in specific cities to capture migration processes, which does not account for other factors that influence relocation decision making (Bukvic & Owen, 2017).

To develop a stochastic model of socio-economically heterogenous migration behavior in response to SLR based on the push-pull theory of migration, we consider a study area consisting of both urban and rural settings. We analyze relocation flows triggered by flooding from a broader perspective that transcends economic rationality and can stochastically encompasses other factors beyond economic incentives that may initiate, intensify, or influence migration patterns. We adopt a mixed approach that integrates three research methodologies (Piguet, 2022) of spatial, multilevel, and survey analyses. Using a spatially-explicit agent-based approach, we examine how differences in relocation decisions across income and age groups impact large-scale emergent relocation behavior and vulnerability profiles in flood-prone areas. Leveraging a recently developed stochastic ABM framework (Nourali et al., 2024), a social system is constructed based on push-pull theory of migration and calibrated to replicate future population projections to 2070. This model accounts for socio-economic heterogeneity in migration decisions based on relocation surveys conducted in the study area (Bukvic & Barnett, 2023) and simulates household-level migration in response to future flood hazard scenarios based on three SLR projection scenarios.

3.2. Methods

3.2.1. Study area

Our study area includes 16 coastal counties within Virginia and Maryland, United States, encompassing both urban and rural coastal areas with varying population trends over the past decade. Figure 3-1 demonstrates the spatial distribution of urban and rural settings across the 481 census tracts within the study area. As a region with one of the highest rates of SLR along the U.S. coasts, Virginia is experiencing a rise in flood risk, with 344,400 properties currently facing a significant risk of flooding (First Street Foundation, 2020). Over the next three decades, this risk is projected to increase by an additional 13.1%, resulting in a total of 389,700 properties with a substantial risk of flooding (First Street Foundation, 2020). The high local rate of SLR can be attributed to several factors, including land subsidence, ocean dynamics, human activities such as groundwater extraction, and Glacial Isostatic Adjustment (Boon et al., 2010; Eggleston & Pope, 2013). The impact of SLR can already be felt in many low-lying Virginia communities (Atkinson et al., 2012b; Mitchell et al., 2013). The frequency and duration of minor tidal flooding (nuisance flooding) has increased dramatically in recent decades along many U.S. coasts (Ezer & Atkinson, 2014; Sweet & Park, 2014), which disrupts daily life by, for example, blocking traffic and impeding access to work, hospitals, schools, and other essential services.

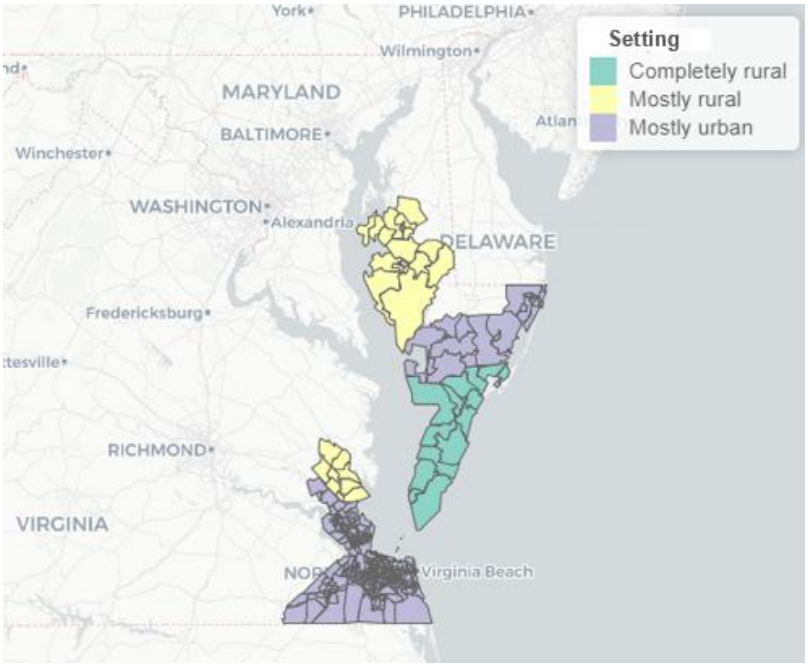


Figure 3-1. Census tract level map of study area. Census tracts are colored based on county-level data on setting type (Ratcliffe et al., 2016).

3.2.2. Modeling Approach

We initialized our model using 2020 census tract level geographic boundaries and data of population, average household size, household income, and age at census tract level for this region (U.S. Census Bureau, 2020) to generate the initial population in the model. Individuals are grouped into households using the average household size of their census tract. With the average household size of 2.5 across the whole study area, each agent represents approximately 80 households of the real population. This abstraction level was chosen to balance computational load and the accuracy of statistical representation, consistent with previous studies on the migration behavior of large populations affected by climate change (Hassani-Mahmooei & Parris, 2012). The model is initialized with a population of 22,440 agents in 2020, representing a population of 1,700,000 individuals. We simulate flood events, as well as households' migration and adaptation decisions with a yearly time step.

To simulate interactions between flooding and relocation behavior across demographic groups, we employed an ABM that incorporates stochastic flooding and uncertain SLR, along with human responses to assess patterns of relocation behavior through the years 2021-2070 (Nourali et al., 2024). This model is built upon the push-pull theory of migration, which asserts that migration decisions are shaped by a variety of factors that both push individuals away from certain locations and pull them towards others (Dorigo & Tobler, 1983; Hunter, 2005; Pan, 2019). This stochastic ABM included a baseline model where agents' migration decisions were stochastically made based on the push-pull scores of their current location, derived from population growth projections as a collective factor of relative desirability in different census tracts, without explicit consideration of flooding. An overview of the model structure is presented in Figure 3-2. The baseline model parameters were calibrated using county-level population projections (from Hauer & CIESIN, 2020; Hauer, 2019). To this baseline population projection model, flooding was incorporated as a factor affecting agents' decision on desirability of census tracts. Potential flood events were statistically generated using coastal flood hazard data from the U.S. Army Corps of Engineers' 2015 North Atlantic Coast Comprehensive Study (Cialone et al., 2015; Nadal-Caraballo et al., 2015). In the present study, several modifications are applied to the original modeling framework (Nourali et al., 2024) to account for agents' heterogeneities.

Agents are defined at the household level with age and income level attributes. Census tracts include rural/urban classification, demographic information, projected population growth rates and the percentage of area inundated by yearly flooding events. Agents migrate based on the relative significance of flooding impact compared to

desirability of locations resulting from various potential factors. We compare model outcomes under uniform and varied decision-making behavior to examine socioeconomic influences on relocation. In the uniform model, all agents have an equal likelihood of migration, whereas in the varied decision-making model, the probability of relocation varies based on agents' demographic characteristics. Heterogeneous migration choices relies on primary relocation preferences data (Bukvic & Barnett, 2023) to better account for the contemporary contextual attributes that significantly influence risk-related decision-making at the household level. We evaluate simulation results using temporal changes in displacement and spatial movement patterns.

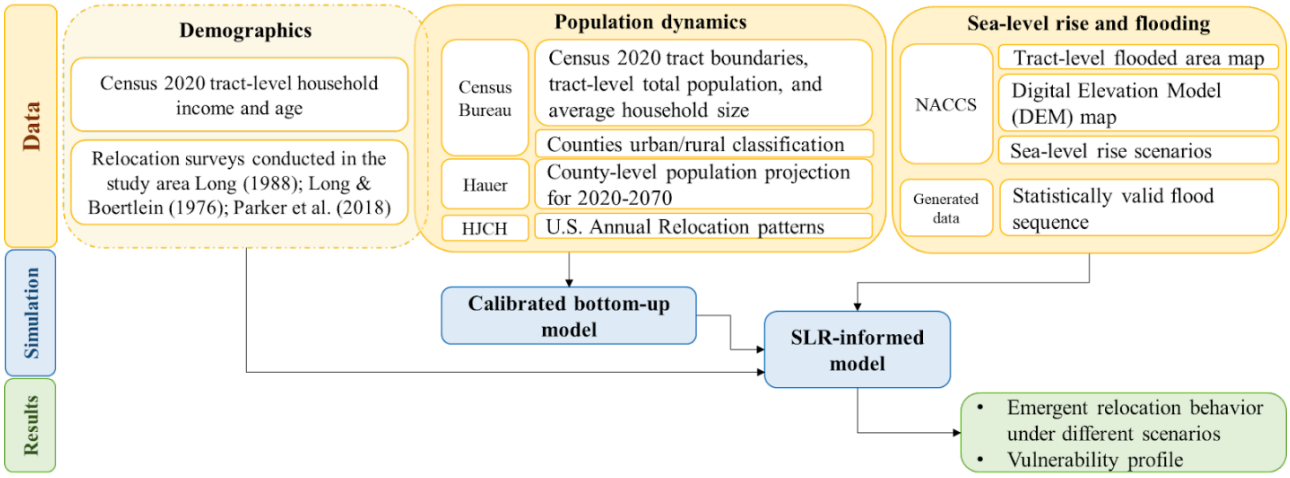


Figure 3-2. Model framework

3.2.2.1. Agents' Demographic Attributes

The ABM used in this study conceptualizes each agent as representing a number of households based on average household size in each census tract (U.S. Census Bureau, 2020). In the model initialization step, each agent is assigned the attributes of age, income, and location's urban/rural classification. Census bureau data is used to determine

the distribution of agents across age and income categories within each census tract Agent ages were then classified as either working age (19 to 59 years old) or retired (over 60 years old). Agents' age attribute in the model corresponds to the age of an adult(s) heading the household, as it is assumed that all households are led by at least one adult. To simulate differences in income levels, the model defines three income categories: low income (less than $50,000 annual income), medium income (between $50,000 and $99,999), and high income ($100,000 and above). While these coarse groupings may mask more detailed differences between groups (e.g., young adults versus middle age adults with families), this level of aggregation was chosen to be consistent with the available survey data relocation preferences. Table 1 shows the percentage of agents in each category across the whole study area, as well as statistics summarizing variations across census tracts.

To simulate aging, the model employs a mechanism wherein, at each time step (tick), a percentage of agents randomly die, with a higher likelihood of death for older agents. The probabilities of death by age are derived from the 2020 period life table (Social Security Administration, 2020), with the average probability taken since the model does not distinguish between male and female agents. This stochastic approach to agent death aims to realistically yet simply account for the increased mortality risk associated with older age. As living agents progress in time, their ages are incremented at each tick and assigned with the corresponding age group. Additionally, new agents are "born" into the simulation to compensate for the agents that have died and also, to adjust the total population according to projected changes in the study area's population, which align with total projections for the study area as referenced from Hauer & CIESIN (2020)

and Hauer (2019). If the study area population adjustment is negative for a given year, a random selection of agents is “removed” from the total population during that simulation time step. In the case of positive population change, new agents are assigned the age 19 to account for children becoming adults and forming households. This assumption is applied to address the lack of explicit representation of individuals under 19 years old in model initialization and avoid the final population at the end of the time period skewing older than projections suggest. However, it does not account for differing ages of agents who migrate from outside of the study area.

Table 3-1. Percentage of agents in each category across the whole study area, and its variation across census tracts

Characteristics	Category	Study area	Census tracts	
			Mean	Standard deviation
Age	Adult working	68.35	71.41	11.73
	Adult retired	31.65	28.59	11.73
Income	Low	36.85	38.05	18.02
	Medium	32.88	32.35	9.17
	High	30.26	29.6	17.71
Average household size		2.52	2.52	0.4

3.2.2.2. Relocation Decision-making

The ABM described by Nourali et al. (2024) simulates migration decisions by “push” factors, prompting them to leave their current location, and "pull” factors which determine the destination where agents will relocate to. This study considers inherent desirability as well as flooding as the drivers of these push-pull forces affecting agents’ decisions. When contemplating migration, agents assess their current location based on both push and pull factors to determine their preference for moving. If they decide to migrate, their subsequent evaluation of alternative locations is again based on push-pull

effects of these drivers to select their destination. Migration decision is made by threshold approach, as it is typically resorted to only when a specific threshold is surpassed, beyond which other on-site adaptation measures to climate-related hazards prove insufficient or progressively less effective (R. McLeman, 2018b). We modified the framework presented in Nourali et al. (2024) in several ways to explicitly differentiate behavior across demographic groups.

The calibration of bottom-up baseline framework closely resembles that of Nourali et al. (2024) to replicate top-bottom data driven population projections proposed by Hauer & CIESIN (2020) and Hauer (2019), without considering the effect of flooding on population displacement. However, we adjust this approach to improve the model's representation of counties experiencing population decline by incorporating a new parameter *b* into the logistic function governing the push-pull score. This parameter allows for greater flexibility in modeling negative-growth scenarios by controlling the bottom boundary of the logistic function. The revised push/pull score equation here takes the form of Equation (3-1):

$$P_i = b + \frac{1 - b}{1 + exp\left(-k_{pull} \times G_i\right)} \qquad (3-1)$$

where P_i is push pull score for census tract i, G_i is population growth or decline rate of census tract i, k_{pull} is the parameter of logistic equation, and *b* is the new parameter introduced to improve the model's performance in counties with declining populations. Parameter tuning was performed by running 144 simulations in a full factorial arrangement of these five global parameters. The calibrated set meets minimum relative root mean square error (Relative RMSE) between simulated population by the baseline

model and population projections, while providing an annual move consistent with nation-wide statistics on local and regional scale mobility (HJCHS; Frost, 2020). Table 3-2 summarizes final results of the calibration process for the study area.

Table 3-2. Calibration parameters, ranges, and calibrated values

Parameter	Minimum value	Maximum value	Increment	Calibrated value
k_{pull}	0.0004	0.0008	0.0002	0.0006
δ	0.2	0.6	0.2	0.2
ε	10	16	2	14
MT	0.25	0.40	0.05	0.35
b	0	0.4	0.1	0.1

Inherent diversity in agents' migration decision-making processes is incorporated through agents' migration threshold. Investigations into migration behavior in the U.S. suggest that the highest likelihood of moving is among lower income, younger (Parker et al., 2018), and urban residents (Chalabi, 2015). This is consistent with relocation preferences in our study area obtained via household semi-structured surveys administered by (Bukvic & Zobel, 2023). These surveys explored quantitative and qualitative considerations that would prompt homeowners to consider permanent relocation and their preferences for relocation destinations. From the survey results, we specifically investigated responses to the question "would you consider relocating elsewhere due to coastal flooding?". We calculated an overall "movement score" for each demographic group, based on weighted responses to the survey question. The resulting values, listed in Table 3-3, indicated varying likelihoods of considering relocation due to coastal flooding across different demographic groups. Rural residents were less inclined to consider moving compared to urban residents. Younger individuals were more likely to express willingness to relocate compared to older individuals. Additionally,

respondents with lower incomes were more inclined to consider relocation than those with higher incomes.

Table 3-3. Likelihood of each response across different age, income, and setting groupings

Response	N	Inclination to move	Percentage of respondents	Residential setting		Age group		Income level		
				Urban	Rural	Working	Retired	Low	Medium	High
No	15	0	0.146	0.116	0.206	0.152	0.140	0.125	0.167	0.136
Yes	38	1	0.369	0.377	0.353	0.435	0.316	0.438	0.467	0.341
Maybe in future	50	0.5	0.485	0.507	0.441	0.413	0.544	0.438	0.367	0.523
Movement score			0.612	0.630	0.574	0.641	0.588	0.656	0.650	0.602

We integrate diversity in migration behavior stemming from demographic characteristics into our model so that the migration threshold is elevated (increasing the probability of relocation) for individuals characterized by low income, urban residence, and youth, whereas it is lowered for those with high income, rural dwelling, and advanced age. Considerations and details of these changes in the move threshold, specific to the model developed for the study area, can be found in the section 3.2.2.3.

3.2.2.3. Demographic diversity in relocation

To address the key question of how diversity in relocation decisions influences municipal-scale change in the context of SLR, we compared model outcomes under two scenarios: one with homogenous rules for stochastic relocation behavior for all agents and another where the parameters of the stochastic relocation rules vary based on the agent's demographic class. In the homogenous behavior simulations, all agents share a single move threshold when considering migration, while in the heterogeneous relocation behavior simulations, the move threshold was determined as a function of agent age, income, and urban-rural classification, adjusted to be higher (resulting in greater

likelihood of moving) for agents that are low income, urban, and younger, while making the threshold lower for high income, rural, and older agents. To maintain the overall number of annual moves the move thresholds were adjusted inversely proportional to the number of agents in a given class. Accordingly, the move threshold was adjusted based on both demographic characteristics and a score which reflected the relative population size of each class, where the score of a class increased for low income, urban, and working age agents, while it decreased for high income, rural, and advanced age agents. The negativity or positivity of a score value reflects the increase or decrease in the relocation probability of the class from the baseline move threshold due to its demographic characteristics. All these considerations were applied to heterogenous move threshold through Equation (3-3):

$$MT_c^{Adjusted} = MT^{Baseline} + \frac{I_c \times SF}{N_c/N_T}\ ; \qquad \forall c \in Class\{1,2,\dots,12\} \qquad (3-3)$$

where $MT_c^{Adjusted}$ is the adjusted move threshold of class c, $MT^{Baseline}$ is a constant move threshold considered for all classes, I_c is the score of class c, SF is a scaling factor governing the magnitude of move threshold changes, and $N_T\ and\ N_c$ are the total number of agents in the study area and number of agents in class c in year 2020 respectively.

Due to uneven population distribution among classes, Equation (3-3) resulted in disproportionately large adjustment values for very low populated classes. To address this, move thresholds for classes with population less than 2% of total number of agents in the study area were adjusted manually. For our specific study area, using a scaling factor of 0.0008 and calibrated value of homogenous move threshold yielded an average

annual move percentage exceeding the general statistics of annual moves in the U.S. To fix this issue, the base move threshold in heterogeneous simulations was reduced to 0.34. This resulted in an annual percent moved of 12.6% for the study area as a whole, with the annual movement percentages for each demographic class as listed in Table 3-4.

Table 3-4. Move threshold adjustment across different social classes, defined based on three income levels of high (H), medium (M), low (L), two age groups of working (W) and retired (R), two settings of urban (U) and rural (RR).

Class no.	Income	Age	Setting	Percentage of agents	Move threshold		Average percent annually moved
					Static (Calibrated)	Variable (Adjusted)	
1	H	W	U	21.1	0.34	0.34	0.109
2	L	W	U	24.4	0.34	0.3498	0.122
3	M	W	U	22.8	0.34	0.347	0.141
4	H	R	U	7.8	0.34	0.3196	0.144
5	L	R	U	8.2	0.34	0.3497	0.167
6	M	R	U	7.7	0.34	0.34	0.142
7	H	W	RR	1.2	0.34	0.32	0.116
8	L	W	RR	2.0	0.34	0.3599	0.156
9	M	W	RR	1.7	0.34	0.3165	0.118
10	H	R	RR	0.7	0.34	0.3	0.122
11	L	R	RR	1.3	0.34	0.33	0.162
12	M	R	RR	1.0	0.34	0.32	0.153

Given that the destination selection is typically influenced by the economic circumstances of households, our weighted random approach is designed to increase the likelihood of agents being moved to census tracts with a majority of households sharing a similar economic level. To simulate this behavior, a random number is generated. If the random number is less than 0.6, the agent will move to a census tract with the same income class; Otherwise, the agent will move to a census tract with a different income class.

3.2.2.4. Flood Simulations

We utilized the statistical coastal flood hazard data from the U.S. Army Corps of Engineers' 2015 North Atlantic Coast Comprehensive Study (NACCS; Cialone et al., 2015; Nadal-Caraballo et al., 2015) to depict potential flood events throughout the simulation period. This dataset encompassed Nor'easters, tropical storms, hurricanes, and the influence of astronomical tides. These statistical values, which are based on different SLR projections and return periods ranging from 1 to 10,000 years, align with the probabilistic surge hazard methodology adopted by FEMA for establishing Flood Insurance Rate Maps. To estimate the extent of flooding in each census tract, we utilized the NACCS statistical values for three SLR projection scenarios, including a low (global SLR of 30 cm by 2100), medium (global SLR of 100 cm by 2100) and high (global SLR of 200 cm by 2100) scenario. These correspond to the local SLR scenarios of 45, 64, and 89 cm by 2070, as illustrated in Figure 3-3. To generate a time series of annual flood impacts based on the storm return period information presented in the NACCS datasets, we first determined the percentage of each census tract inundated for floods with return periods of 1, 2, 5, 10, 20, 50, 100, 200, and 500 years. By linearly interpolating these inundation percentages, we obtained estimates for storms with return periods between the aforementioned nine values.

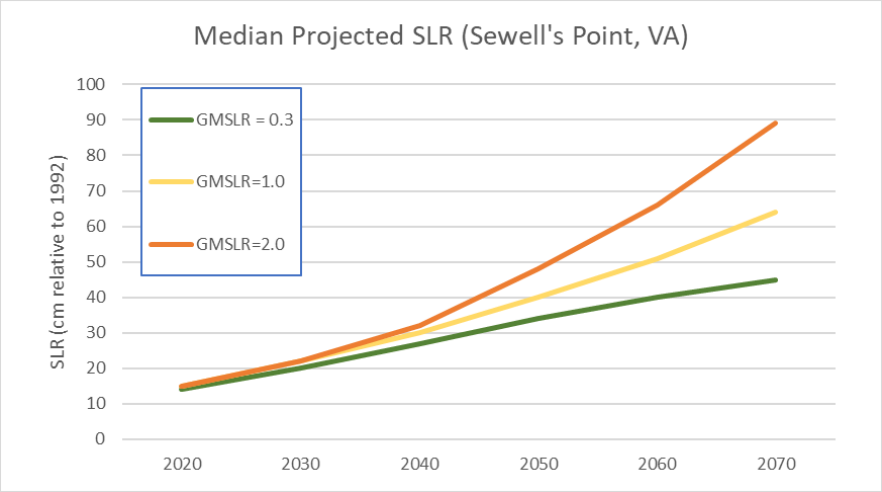


Figure 3-3. SLR projection scenarios. Scenarios refer to each of the 3 Global mean sea level scenarios (identified by the rise amounts in meters by 2100. 0.3 m, 1 m, and 2 m)

In order to generate long-term sequences of flood incidents within the region, we assumed a single occurrence of flooding per year and generated 300 sequences of fifty random numbers ranging from 0 to 1 from a uniform distribution. The generated numbers were inverted to derive sequences denoting flood return periods. For instance, a random value of 0.5 corresponded to a flood event transpiring every two years, while a value of 0.1 indicated a flood recurring every ten years. we generated. From this ensemble of statistically feasible 50-year flood sequences, we searched for sequences that depicted different conceptual long-term flood occurrence possibilities associated with two storylines. Storyline 1 assumes frequent small floods (with a return period of 2-10 years) throughout the 50-year simulation. Storyline 2 assumes frequent small floods as in storyline 1, but with an additional severe storm early in the simulation period. Figure 3-4 demonstrates log-transformed return periods of flooding sequences in both storylines. We then used NACCS to estimate a percentage of each census tract inundated within the study area under these two flood scenarios. This approach assumes a single flood event per year, which is a simplification that does not account for the effect of interannual

flooding fluctuations and multiple flood events in a year. However, it does provide a means to assess long-term trajectories over the course of our 50-year study period in support of the research objectives described above.

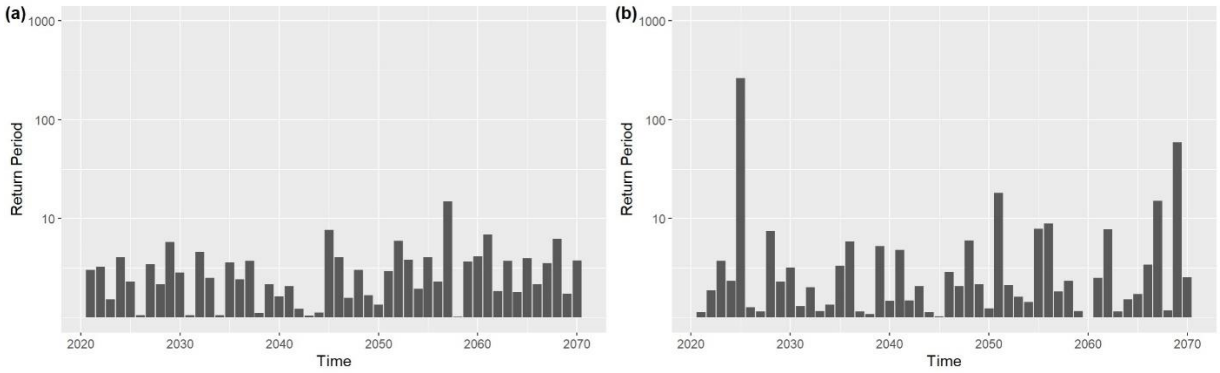


Figure 3-4. Conceptual flood storylines. (a) minor flooding, (b) minor flooding + one early severe flooding event.

Integration of flooding impact with general relocation patterns was done through conceptualizing exposure to flooding as a push factor, where the occurrence of flooding adjusts the distribution from which the agents' random satisfaction scores are drawn. In previous research (Nourali et al., 2024), a decision rule was introduced for agent relocation in response to flooding events, with a constant weight parameter W to determine the influence of flooding on agent satisfaction, (Equation 3-2):

$$P^*_i = P_i - W \times F_i \qquad (3-2)$$

where F_i and P^*_i are respectively the percentage of census tract i that is inundated in a given year and its flood-influenced push-pull score. The weight W represents the relative importance of flooding in comparison with the census tract's baseline desirability. Importantly, the values of W need to be constrained to a relatively small range to prevent negative P^*_i values, ensuring a meaningful representation of satisfaction. As a general

approach, suitable range of W can be determined through looking at statistics of distributions of P_i and F_i across the study area, ensuring that upper bound of W does not result in too small P^*_i values to preserve the underlying beta distribution. For our specific study area, we set W to the values of 0.1 and 0.3 to assess two different scenarios where flooding plays an increasingly important role in agent decision making.

3.2.3. Experimental Design and Evaluation

Our experimental design included three sea-level rise projection scenarios and two flood storylines. These flood scenarios were examined under two move threshold assumptions (homogenous versus heterogenous) for agents' push-pull score evaluation and two flood weights (W) to evaluate how the significance of flooding for agents impacts migration decisions. Addressing the influence of random processes within our simulation framework, we implemented 30 stochastic iterations for each of these 24 combinations. We evaluated the profile by age and income factors in flood-prone census tracts, which were assumed to be the 10% of census tracts (total of 48 tracts) that experience the greatest percentage of inundated area averaged across the simulation period. Figure 3-5 shows the spatial distribution of these census tracts.

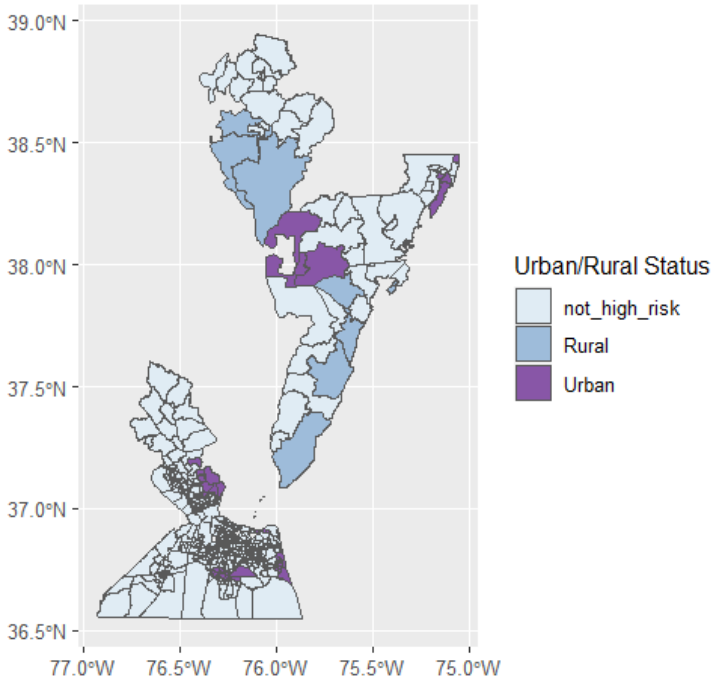


Figure 3-5. Spatial distribution of highly prone census tracts to flooding, colored based on urbanization level.

The comparison of relocation response to each of these flood scenarios and model assumptions was done through looking at different spatial and temporal outcomes. Spatial assessment of vulnerable groups focused on distribution of percentage of these groups across flood prone census tracts in year 2070. Socioeconomic status can significantly influence social vulnerability, with wealthier households better equipped to prepare for and recover from disasters (Koks et al., 2015). In this study, census tract population percentage falling in each category of annual household income (in 2020 inflation-adjusted USD) serves as a proxy for wealth in that census tract. Identifying socially vulnerable groups, such as the elderly and low-income households can inform policymaking to develop effective adaptation strategies in flood-prone areas. For instance, elderly individuals may face challenges in rapid evacuation during flood events,

which may potentially increase the burden of care on others (Cutter et al., 2000; Hewitt, 2014). Furthermore, elderly households may experience mobility constraints, hindering their while low-income households may struggle to afford flood-proofing measures for their homes.

Kernel density estimation (KDE) was used as a non-parametric method to estimate the probability density function (PDF) of percent low-income or percent retired based on the model results in 48 census tracts with most severe flooding experience. Equation (3-4) shows the kernel density estimate at vulnerable percent x:

$$f(x) = \left(\frac{1}{n \times h}\right) \times \sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right) \qquad (3-4)$$

where $f(x)$ is kernel density estimate, X_i represents the vulnerable group percentage data for flood-prone census tract i in year 2070, n is the total number of flood-prone census tracts, the bandwidth h controls the width of the kernel and influences the smoothness of the estimated density. K represents the scaled kernel function centered at each data point X_i, with the scaling factor $\frac{x-X_i}{h}$. We used Gaussian kernel function for distribution estimation.

The temporal assessment considered average number of moves per year among potentially vulnerable groups (older adults and low-income individuals) through the simulation time horizon. Temporal dynamics of mobility cam reveal possible repeated relocations, which can impose significant burdens on individuals, including financial costs and disconnection from local social networks. Equation 3-5 demonstrates how this migration flux for each vulnerable group.

$$\bar{m}_t = \sum_{p=1}^{P_t} \frac{m_t}{P_t} \qquad (3-5)$$

$\bar{m}_t$ is average number of moves of a potentially vulnerable group, m_t is number of vulnerable agents moving, and P_t is vulnerable agents' population in year t.

3.3. Results

3.3.1. Spatial analysis

To characterize how heterogeneity in relocation decision making impacts the population of potentially vulnerable individuals in flood-prone areas, we first assessed the percentage of agents in flood-prone census tracts (defined as the 10% of census tracts that experience the greatest percent of area inundated averaged across the simulation period) who were classified as either low income or retirement age. Table 3-5 presents the mean low-income population percentage averaged across all flood-prone census tracts in year 2070. The results presented in this section are averaged over 30 stochastic iterations. Table 3-5 also includes the kurtosis in the distribution of low-income percentages across census tracts to indicate how modeling assumptions impact the occurrence of outlier census tracts with especially high or low populations of low-income individuals.

Table 3-5. Statistics of average distributions of percent low income in highly flood-prone census tracts under different flood scenarios and model assumption in year 2070

Statistic	Flood scenarios		Model variables			
	SLR	Storyline	W 0.1		W 0.3	
			MT-static	MT-var	MT-static	MT-var
Mean Percentage	030	1	38.7	37.5	38.6	37.3
		2	38.4	38.0	39.0	37.6
	100	1	39.2	37.8	38.6	38.2
		2	38.9	37.9	39.2	37.5
	200	1	38.3	38.2	37.8	37.6
		2	38.5	37.7	39.0	38.0
Kurtosis	030	1	1.59	2.03	1.61	1.87
		2	1.46	1.70	1.42	1.84
	100	1	1.69	1.93	1.67	1.84
		2	1.52	1.79	1.66	1.81
	200	1	1.60	1.62	1.62	1.95
		2	1.63	2.12	1.64	1.88

The mean values in Table 3-5 indicate that incorporating a variable move threshold results in lower percentages of low-income households in flood-prone areas across all SLR scenarios, flood storylines, and values of *W*. When flooding plays a greater role in agent decision making as represented by a higher value of *W*, the mean percentage of low-income agents increases under the second flood storyline that experiences an extreme flood, compared to the first storyline which only includes minor flooding.

Comparing kurtosis values indicates that incorporating a variable move threshold results in the distribution of low-income agents becoming more tailed. Higher kurtosis values indicate a greater presence of values that are extremely high or low compared to the mean, that is, how often outlying percent low-income values occur across flood-prone census tracts.

The distribution of low-income percentages across flood-prone census tracts (as the example depicted in the Figure 3-6) exhibits a bimodal pattern. Many of the flood-prone census tracts are either characterized by a relatively low proportion of low-income residents (25 – 35%) or a relatively high proportion (40-50%) of low-income residents. Transitioning from a static move threshold (depicted in blue) to demographically varying movement (illustrated in red), the distribution becomes skewed and compressed, exhibiting heightened peaks. This pattern was consistent across multiple flooding storylines, SLR scenarios, and flood importance (*W*) values.

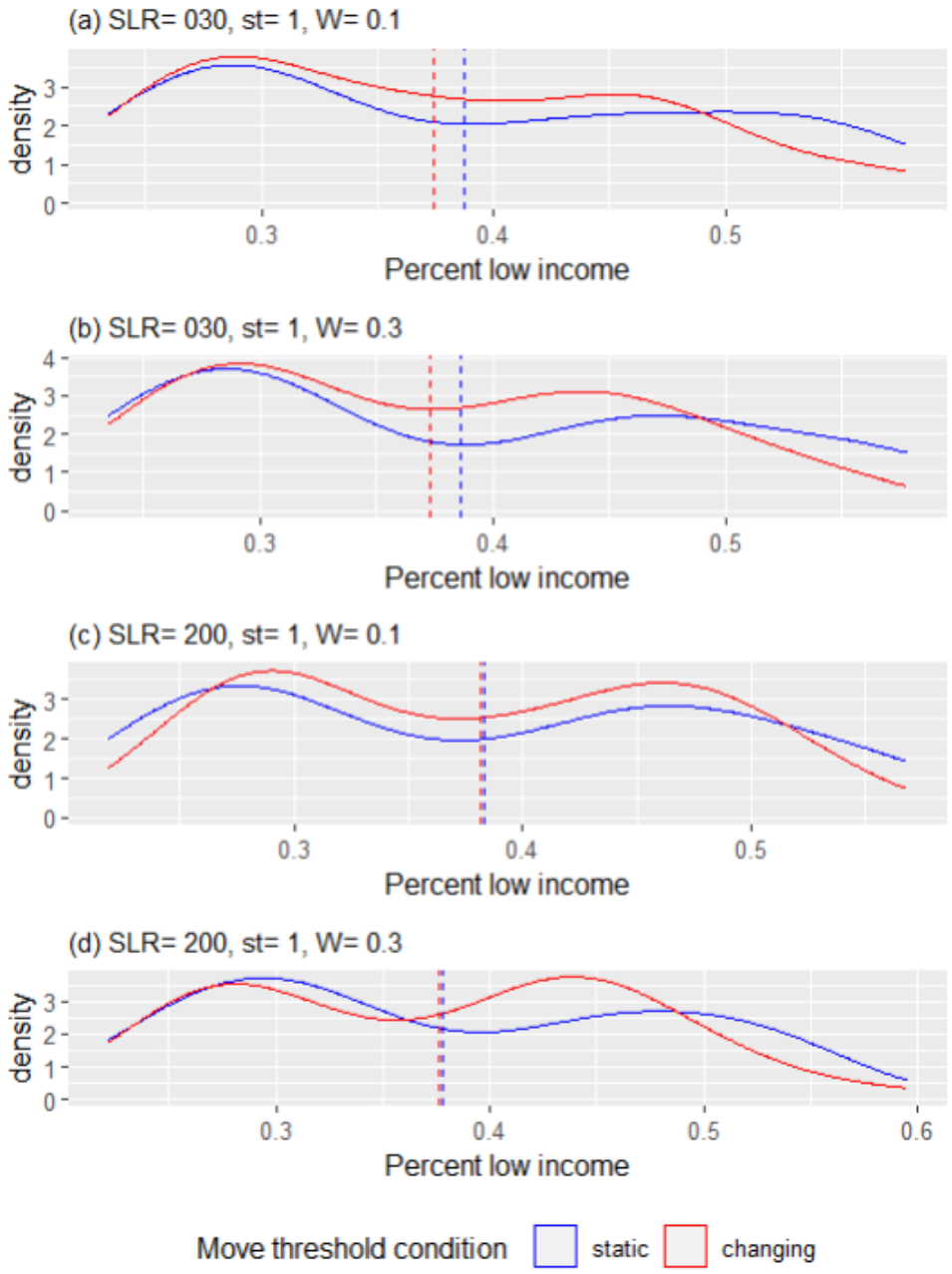


Figure 3-6. Kernel distribution of percent low-income in highly flood-prone census tracts under storyline 1 and scenarios: (a) SLR 30, flood weight 0.1, (b) SLR 30, flood weight 0.3 (c) SLR 200, flood weight 0.1, (d) SLR 200, flood weight 0.3

Table 3-6 presents the mean percentage of agents in flood-prone census tracts who were in the retirement age at the end of simulations (year 2070). Distributions of the percentage of retirement age agents presented different behavior than low-income agents. Incorporating a variable move threshold in most SLR scenarios and flood storylines

results in an increase in the percent retired in flood-prone areas compared to a static threshold. The effect of flood weight on the mean results is more consistent in this set of distributions, showing a lower percentage of retirement age agents when flood exposure is weighted more heavily (W = 0.3) in agents' relocation decisions with both homogenous and heterogenous move thresholds.

The kurtosis values of percent retired distributions are mostly greater than 3, indicating that they have heavier tails compared to the percent low-income distributions. Under varying move threshold, the percent retired distribution becomes more tailed.

In the example distribution of percent retired depicted in Figure 3-7, when considering relocation diversity, a higher peak and more variability are observed.

Table 3-6. Statistics of average distributions of percent retired in highly flood-prone census tracts under different flood scenarios and model assumption in year 2070

Statistic	Flood scenarios		Model variables			
	SLR	Storyline	W 0.1		W 0.3	
			MT-static	MT-var	MT-static	MT-var
Mean Percentage	030	1	35.6	36.3	35.0	35.6
		2	35.9	35.6	35.4	35.4
	100	1	35.3	36.1	34.6	35.2
		2	35.8	36.1	35.0	35.9
	200	1	35.9	36.1	34.9	35.2
		2	35.7	35.6	35.1	34.9
Kurtosis	030	1	3.75	4.72	3.42	3.36
		2	3.49	4.05	3.65	3.58
	100	1	3.79	3.62	3.29	4.45
		2	2.78	3.37	2.91	3.55
	200	1	3.13	3.37	3.20	2.81
		2	2.76	3.35	2.83	3.81

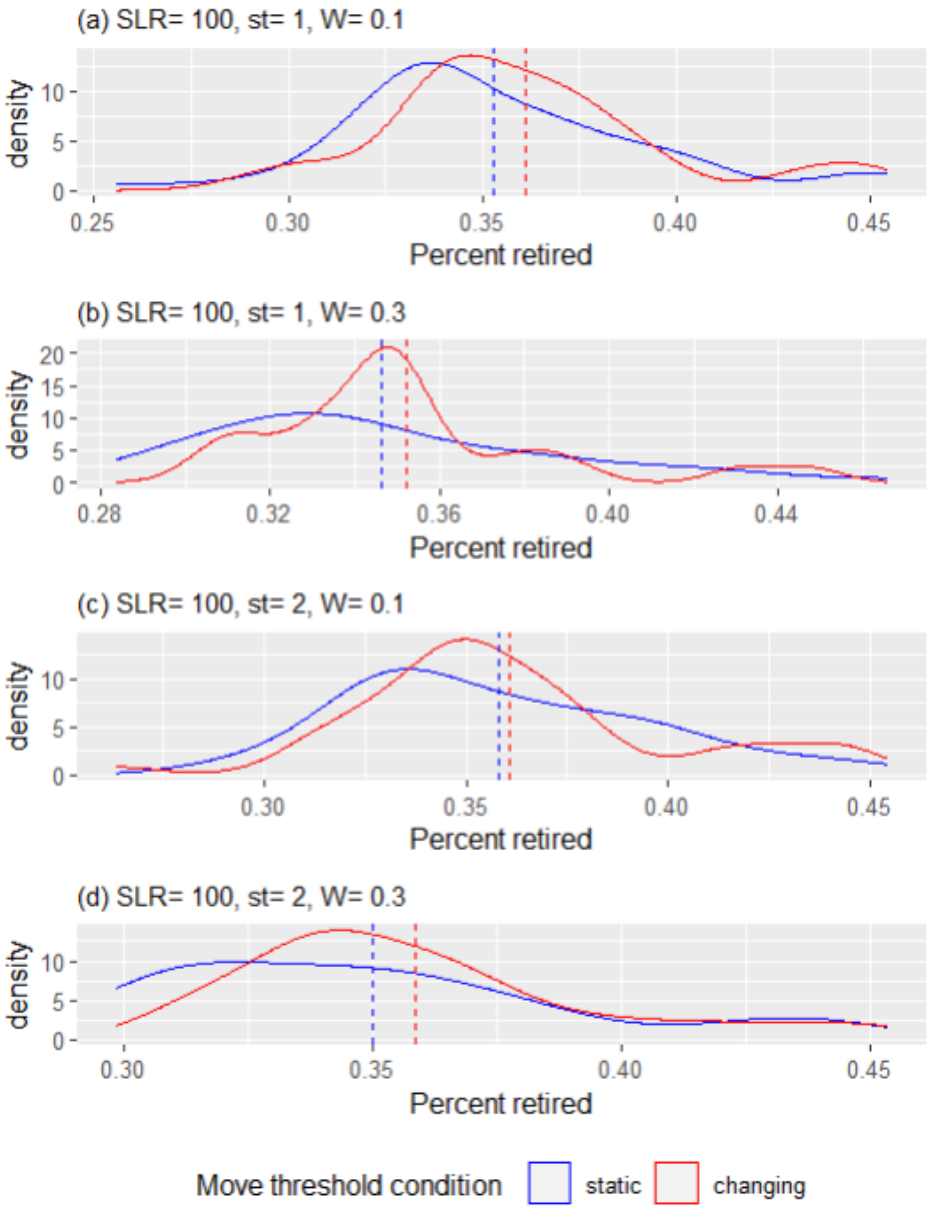


Figure 3-7. Kernel distribution of percent retired in highly flood-prone census tracts under SLR 100 and scenarios: (a) storyline 1, flood weight 0.1, (b) storyline 2, flood weight 0.3 (c) storyline 1, flood weight 0.1, (d) storyline 2, flood weight 0.3

3.3.2. Temporal analysis

To investigate the impact of flood scenarios and model structure on relocation patterns over time, our initial step examined the total population relocated annually across the study area. As depicted in Figure 3-8, introducing the variable move threshold results in fewer moves, on average, compared to a static move threshold. However, the

temporal patterns of moves across the simulation period is consistent across both static and variable move thresholds. Assuming greater importance of flood influence on movement decisions ($W = 0.3$) results in a higher average number of relocations and increased year-to-year variability in relocation patterns. This weighting of flood impact demonstrates a more pronounced effect on simulation outcomes compared to differences across flood storyline or sea-level rise scenarios.

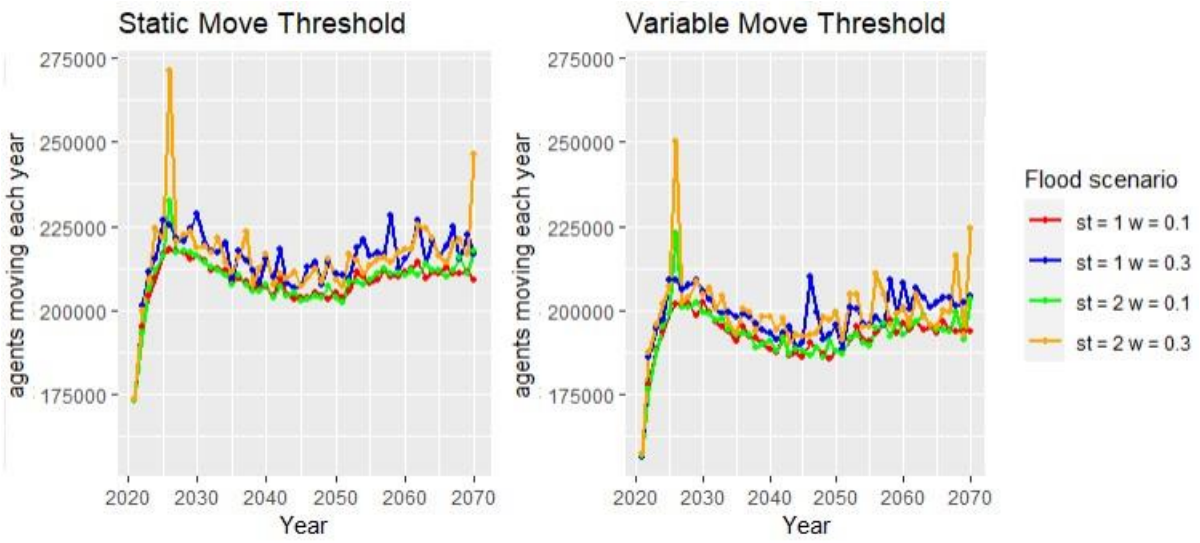


Figure 3-8. Time-series of population moved under SLR 200 scenario across the study area under static and variable move threshold assumptions for storylines (st) 1 and 2, and values of W equal to 0.1 and 0.3.

Looking further into the migration flux (the number of moves per year) across different social groups, we observe a higher percentage of movement among low-income individuals and a lower percentage among retirees, even when a static move threshold is used (see Figure 3-9-a and -b). This is likely reflective of the baseline desirability scores of different areas, e.g. high-income agents live in highly desirable places and are less probable to move. When we consider the diversity in relocation behavior, these

disparities become more pronounced, as individuals in the low-income and employed categories are assumed to exhibit more willingness to relocate.

Additionally, when comparing settings, a higher percentage of rural residents relocate, as these areas are typically less desirable, prompting residents to seek opportunities in more urbanized, developed regions. However, when factoring in variable move thresholds, the discrepancy between the percentage of rural and urban moves diminishes, as rural residents are assumed to demonstrate a lower willingness to relocate (Figure 3-9-c).

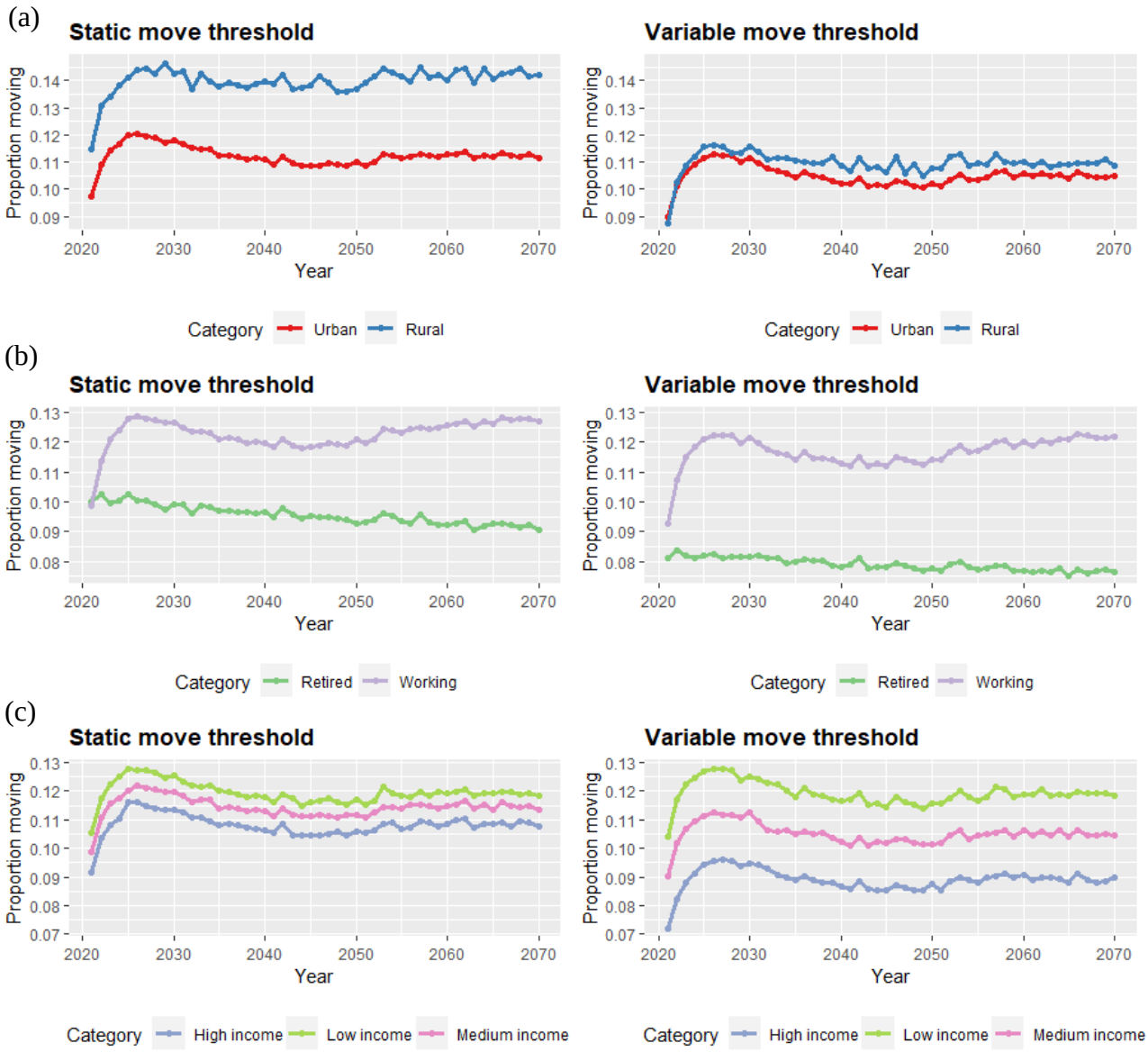


Figure 3-9. Each subpopulation's proportion moved per year under SLR 30, storyline 1, and flood weight 0.1 by (a) income; (b) age; (c) residential setting

3.4. Discussion

Our study contributes to the expanding literature on relocation behavior in flood-prone regions by utilizing a stochastic ABM to examine the interaction between socioeconomic factors and relocation dynamics in response to rising sea levels. We aimed to investigate the effects of SLR and recurrent flooding on vulnerable population

subgroups, an area of increasing interest within climate-induced relocation research (Piguet, 2022).

Existing research focusing on housing market (e.g. Koning & Filatova, 2020; Pryce et al., 2011) suggest that affluent areas gradually lose their appeal due to frequent flooding. This prompts high-income households to relocate to safer regions, leaving vulnerable communities increasingly concentrated in flood-prone zones. Climate gentrification exacerbates this trend, driven by investments in low-hazard properties, thereby accelerating the deterioration of flood-prone neighborhoods. However, our model results demonstrate distinct relocation patterns among different social groups in most exposed regions across the study area. Incorporating diversity in relocation behaviors by age groups and income levels results in fewer low-income households in flood prone areas but higher elderly people compared to the homogenous approach. Initial percentage of low-income households in highly flood-prone census tracts (U.S. Census Bureau, 2021) follows a right-skewed unimodal distribution, indicating a prevalence of affluent areas. Yet, as flooding scenarios unfold in our simulation, this distribution shifts towards a bimodal pattern, segregating census tracts into distinct socioeconomic clusters by the end of simulation. These findings align with the idea of trapped low-income or retired population, but diverge from high income groups' out-migration patterns in the research by Koning & Filatova (2020). This nuanced difference suggests that migration patterns are not solely dictated by economic factors, but are also shaped by individuals' preferences and social dynamics across different groups and coastal communities, extending beyond utility maximization. As previously shown by Paul & Routray (2011), there can be substantial variation in terms of capacity to cope with flood events

depending on households' socioeconomic status, such as income, that can impact their preferences. Additionally, according to a spatial assessment by Bukvic et al. (2022), stronger place attachment driven by attractive amenities, such as public open space and access to water-based economic opportunities is commonly found in affluent and high-income neighborhoods, and influences their decisions regarding relocation in response to flooding. Understanding the underlying drivers of these spatial disparities is essential for designing targeted adaptation and mitigation strategies.

From the perspective of physical hazards, we also found that higher degrees of SLR, particularly when coupled with the second flood storyline, increase the exposure of low-income populations to flooding, potential drivers of which should be investigated further. Despite these contributions, it is important to acknowledge the limitations of our study.

Currently, relocation decisions in our model are only based on agents' immediate exposure to flooding, overlooking the potential effect of individuals' prior experiences on relocation decision making, their move threshold, and destination selection. Human-environment dynamics are influenced by factors such as imperfect information, bounded rationality, and the adaptive nature of human decision-making (Zhuo & Han, 2020). Future research can explore how variable move thresholds and the incorporation of agents' "memories" of flooding influence the relocation experiences of vulnerable groups over time. Adding the effect of social networks, information sharing, and socio-cultural ties is another level of complexity that can hugely impact the emerging relocation patterns. In the specific context of ABMs for SLR-induced migration, building on ideas such as the effect of family ties in potential destinations (Hassani-Mahmooei & Parris,

2012) on migration probability can be an starting point to refine our modeling framework for a more robust impact and vulnerability assessment.

Additionally, we observed that flood weight has a greater impact on relocation behavior compared to predefined flood storylines. In the current version of the model, flood weight is defined as a model assumption under different values of which we examined the output. This highlights the importance of additional empirical research that can quantify the relative impact of environmental stressors such as flooding in driving relocation decisions specifically, and adaptive actions more broadly. Acknowledging challenges for the robust adaptation strategies design presented by the evolving, deep uncertain nature of flood disasters and their implications for communities (Keller et al., 2008), connecting this model variable with survey-extracted evidence for different subpopulations could be an important area for additional research.

Our current spatial assessment is based upon the distribution of vulnerable groups in flood prone census tracts, which can be a great feed for policymaking to adjust adaptation measures for specific social groups. For instance, areas with a high proportion of elderly residents may benefit more from structural flood protection infrastructure like dikes or elevated buildings, rather than reliance on individual risk mitigation actions. Similarly, low-income individuals may require financial assistance, such as low-interest loans or subsidies, to make their homes flood-resistant. The spatial analysis presented in this study can be further enhanced by providing climate vulnerability mapping methodologies and hotspot analysis in vulnerable areas. As stated by (Piguet, 2022), an existing gap in hotspot analyses lies in the oversight of population projections, especially in disadvantaged neighborhoods within urban zones where population growth is

occurring rather than outmigration. Our framework, equipped with a built-in population projection mechanism, could serve as a valuable tool to address this gap.

Lastly, our model considers relocation as the only response available to act on. The interplay between relocation with other coping mechanisms can affect vulnerability profile and resiliency to flooding. Considering adaptation measures, including buyout and building hardening, by agents alongside their relocation options, as well as differing availability of resources for these options across subpopulations, can provide insights into the interacting effects of multiple adaptive actions in response to SLR and flooding.

3.5. Conclusion

This study represents a stochastic ABM for relocation behavior simulation under different SLR projection scenarios, with a specific focus on how differing relocation decision rules across potentially-vulnerable groups impact broad-scale patterns on flood exposure. We explored relocation response to repeated yearly flooding for 16 counties in coastal Virginia and Maryland, including both urban and rural settings. We investigated the effect of heterogeneity in relocation behavior impacts the population of potentially vulnerable individuals in flood-prone areas compared to homogenous decisions. We also analyzed the temporal changes in emergent relocation patterns among the total population as well as different socio-economic subpopulations, considering various flooding scenarios and model variables. Our spatiotemporal simulation framework lays the groundwork for assessments of diverse populations' vulnerability to SLR for municipality-level adaptation planning. The results reveal distinct relocation patterns among low-income and retirees as well as economic clusters within the highly flood-

prone census tracts at the end of simulations. These findings suggest multiple directions for refinement in our modeling approach in future. Specifically, we recognize the oversight of crucial factors such as individuals' prior experiences and social networks in relocation decision-making, as well as possible insights from relocation surveys regarding the importance of flooding for different subgroups. Our spatial analysis can be enhanced by vulnerability mapping and hotspot assessments.

4. Quantifying the Impact of Extreme Weather on Agricultural Losses in the Delmarva Peninsula using Multi-Step Machine Learning and Financial Crop Loss Data

Zahra Nourali, Julie E Shortridge

Abstract

The increasing occurrence of extreme weather events due to climate change presents significant challenges for agricultural production, causing substantial losses and disrupting farming systems. Existing research has predominantly focused on changes in yield for major crops, providing limited insights into overall losses and impacts on diverse regional agricultural systems. To bridge these gaps, this study employs machine learning techniques to develop predictive models that identify and quantify the climatic factors contributing to agricultural losses based on financial crop loss data and crop insurance payouts. These records are combined with historical meteorological data to assess the impact of extreme weather events on agriculture in the Delmarva Peninsula in the Eastern United States from 1980 to 2018. To address irregularly distributed loss data, we compare various multistep modeling configurations that leverage linear, random forest, and support vector machine approaches. The classification and estimation accuracy of each approach are compared using a repeated hold-out cross-validation analysis. Results indicate that machine learning methods, particularly random forest, outperform both statistical approaches and our null baseline model, demonstrating superior generalizability in agricultural damage estimation. Multistep configurations that address irregular data distributions are shown to have a significant influence on models'

capacity to detect damage occurrence. The study reveals a preference for simpler modeling approaches that minimize variance in handling unseen data, as well as the importance of accounting for seasonal patterns, spatial groupings, and persistent weather phenomena in accurately estimating agricultural losses.

Keywords: Extreme weather, Crop damage, Machine learning, Predictive modeling, Zero-inflated

4.1. Introduction

Increasing temperatures and changes in the magnitude, intensity, and frequency of precipitation can have profound implications for agriculture. Elevated temperatures and prolonged drought conditions heighten the risk of substantial crop damage (Reyes & Elias, 2019), as do intensified precipitation, flooding, and excessive moisture (Rosenzweig et al., 2002). For example, an evaluation of crop yield and climate between 1968 and 2013 in the U.S. Great Plains revealed that weather conditions accounted for 25% of variability in crop yields (Kukal & Irmak, 2018). The escalating frequency of disasters also impacts agriculture, with droughts, floods, and storms as the predominant contributors to agricultural losses (FAO, 2021). These damages have prompted extensive research into the impacts of climate change on crop yields using biophysical crop models (e.g., Rosenzweig et al., 2013), statistical models that regress crop yields against meteorological data (e.g., Lobell & Burke, 2010), and methods that integrate the two (Roberts et al., 2013, 2017; Urban et al., 2015). Although empirical studies show consistent patterns of decline in agricultural production due to extreme weather events (Auffhammer et al., 2012; Davis et al., 2019; Fishman, 2016; Lobell et al., 2011; Troy et al., 2015), these studies typically use yield data that represents marginal changes in yield across different time periods and

locations but does not account for complete crop losses that can occur from extreme events. Existing research also focuses most extensively on impacts to major cereal crops over very large (national to global) scales; this can mask regional scale patterns and provide little insight to support adaptation in regions with highly diverse or specialized agricultural systems.

Financial crop loss data, including crop insurance payouts for losses stemming from extreme weather, can capture a broader set of impacts compared to yield data alone. Prior research has leveraged financial crop loss data and meteorological records to measure the economic consequences of climate change for agriculture (Botzen et al., 2010; Diffenbaugh et al., 2021; Li et al., 2019) and estimate environmental change through economic losses and insurance compensation (Nannos et al., 2013). Multiple studies have leveraged statistical models that integrate crop loss and climate data, for instance to develop climate-driven damage projections (Gerstengarbe et al., 2013) and risk assessments using insurance and revenue data (Falco et al., 2014; E. D. Perry et al., 2020); more recent studies have leveraged satellite data to explain and predict agricultural losses (Benami et al., 2021; Shirzaei et al., 2021). Siegle (2023) integrated classification and regression methods to measure the severity of insured losses based on historical commodity yield statistics and climate data, using a two-step structure to generate baseline expectations for indemnification for specialty crop insurance. However, these studies typically employ classical statistical approaches that have limited capabilities in replicating nonlinearities and interactions between explanatory factors and addressing irregularly distributed response data.

Machine learning techniques can be leveraged to describe complex interactions among different variables explaining crop losses, but multiple reviews argue that they remain underutilized in agricultural assessments (Kamilaris & Prenafeta-Boldú, 2018; Liakos et al., 2018; Lyubchich et al., 2019). Previous research that leverages machine learning to understand relationships between meteorological conditions and agriculture has primarily focused on crop yield prediction. For example, Kuwata & Shibasaki (2015, 2016) integrate gridded weather remote sensing data and deep neural networks to estimate corn yields. Pantazi et al. (2016) employ supervised Kohonen networks, counter-propagation artificial networks, and XY-fusion networks to predict wheat yield, utilizing NDVI, soil moisture content, pH, and nutrient concentrations. Kung et al. (2016) suggest combining multiple neural networks in an ensemble to enhance predictions of tomato yields based on various meteorological factors. Beillouin et al. (2020) investigate the impacts of climatic conditions on the yields of nine crop species in different parts of Europe. Collectively, this work has demonstrated how machine learning can be leveraged for yield prediction for different crops and locations, but has not yet been applied to loss data that can encompass impacts across a diverse range of agricultural production systems.

This study employs machine learning models to predict climate-induced agricultural losses in the Delmarva peninsula, an agriculturally-important coastal region of the Eastern U.S. We utilize a comprehensive dataset combining county-level hazard-loss data from the Spatial Hazard Events and Losses Database for the United States (SHELDUS; CEMHS, 2019) with historical time series of daily weather data sourced from Parameter-elevation Regressions on Independent Slopes Model (PRISM; Daly et al., 2008) spanning the period 1981-2018. To address the zero inflated and heavy-tailed distribution

of loss data, we compared several different model formulations that combined stratified sampling, damage classification, and loss estimation. Machine learning approaches of random forests (RF) and support vector machines (SVM) were applied using these multiple model formulations and compared to a Generalized Linear Mixed Model (GLMM) and a null model based on county and monthly averages of losses. A repeated holdout cross validation demonstrated the predictive benefit of multi-step machine learning formulations, and partial dependence plots were used to demonstrate the predictor variables with the greatest influence on model predictions.

4.2. Methods

To stablish a foundation for hazard-damage analysis, county-level financial loss data was integrated with gridded meteorological data from 1980 to 2018. We compared multiple regression and machine learning methods in their ability to replicate observed relationships between these datasets, and multiple model formulations were compared to assess their ability in replicated the irregular distribution of observed loss data. Model performance was compared using repeated holdout cross validation to assess classification and regression accuracy.

4.2.1. Study Area

Our study area is the Delmarva peninsula, encompassing the entire state of Delaware as well as parts of Maryland and Virginia, located within the Mid-Atlantic Coastal Plain of North America. Covering an area of 14,230 km^2, the peninsula is bordered by the Chesapeake Bay to the west and the Delaware Bay and Atlantic Ocean to the east (Figure 4-1). It has a humid subtropical climate and flat topography, with the highest

elevation reaching 31 m above sea level, and a predominantly rural character with agriculture as the primary land use (Ciavola et al., 2014). Agriculture in Delmarva is diverse, including extensive row crop (corn, soybeans, etc.), vegetable, nursery, and poultry production. Due to the unpredictable nature of precipitation during the growing season and low water holding capacity of soils, irrigated agriculture has increased in recent decades. Across the 14 counties comprising the peninsula, irrigated farmland increased by 54.5% between 2002 and 2017 with 11.5% of cropland area under irrigation (Brinson, 2023; U.S. Department of Agriculture, 2018).

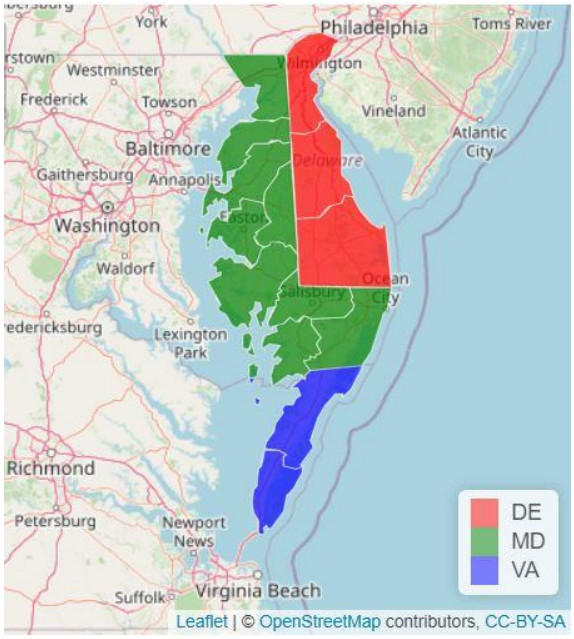


Figure 4-1. Delmarva peninsula

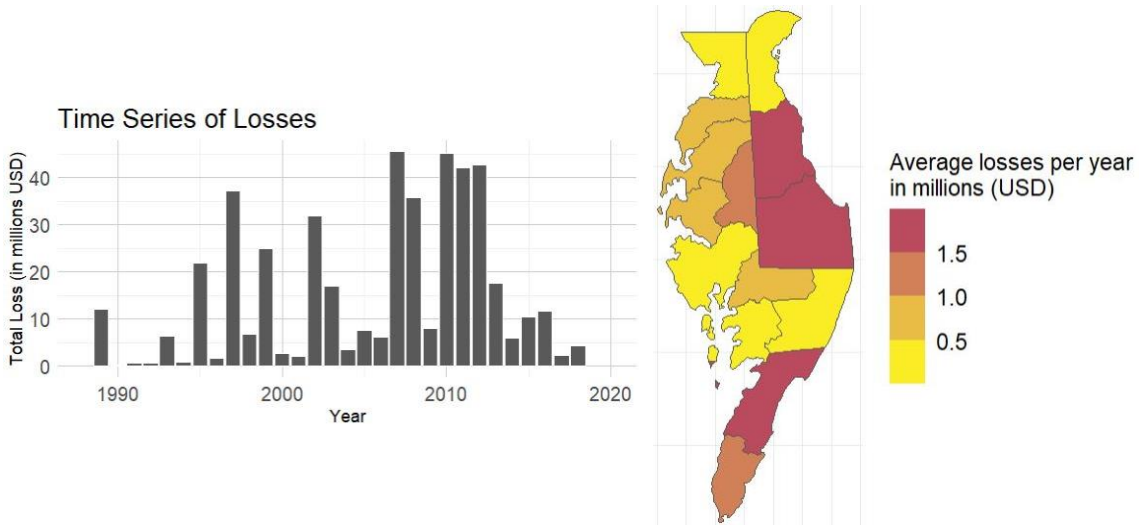
4.2.2. Data Sources and Processing

Financial loss data was obtained from SHELDUS (CEMHS, 2019), a county-level hazard dataset for the U.S. This dataset contains records detailing crop losses (both direct and insured indemnity payments by the USDA) linked to 18 weather phenomena such as

drought, flooding, extreme heat, hurricanes, severe storms, and high winds. The information includes the geocoded location of the loss and the start and end dates of the climatic event causing the loss, and is sourced from USDA Risk Management Agency (*RMA*) indemnity records and NOAA Storm Data and Unusual Weather Phenomena (*NOAA Climate.Gov*) Reports. The spatial distribution and temporal variability of extreme weather induced crop losses within Delmarva counties between 1980 and 2018 (SHELDUS; CEMHS, 2019) is demonstrated in Figure 4-2. The initial list of 18 hazard types included in SHELDUS data were consolidated into two overarching categories of drought (comprising both drought and heat waves) and storm (encompassing all events involving flooding, hail, hurricanes, lightning, thunderstorms, severe storms with wind, tropical storms, and winter weather). This reduction aimed to ensure a sufficient number of records for utilization in both the training and testing phases of the estimation models, and to avoid instances where similar climatic events (e.g., thunderstorms and severe storms) were classified differently.

SHELDUS database contains information on the direct losses caused by the event (property and crop losses, injuries, and fatalities) as well as insured crop losses (indemnity payments by the USDA) in adjusted U.S. dollars. In the uninsured crop damage data, some records with low spatial detail were identified, where losses associated with large-scale statewide events were evenly distributed across affected counties. These records were consequently excluded. For insured crop damage data, multiple loss records for single events within a single county were consolidated into a total value for that county. Additionally, records with missing values in terms of date information were deleted. To mitigate the zero-inflation issue, SHELDUS daily total loss values were aggregated

monthly. The insured and uninsured losses were combined to be considered as the total crop damage. All loss data were aggregated to monthly sums for each county. To improve models' estimation capacity, log-transformed values of financial losses were used in all formulations.



(a) (b)

Figure 4-2. Distribution of financial losses from 1980 to 2018: (a) total losses across the whole peninsula (b) average losses per year by county

Daily temperature and precipitation records from 1981 through 2018 were obtained from PRISM (Daly et al., 2008), a spatial dataset that provides historical observations of climatic variables across the United States, employing interpolation models for values between weather stations. PRISM raster data was cropped to the Delmarva peninsula and aggregated to county averages that excluded non-agricultural areas using the National Land Cover Database classifications from 2001 (Homer et al., 2015). This resulted in a time series of daily maximum temperature and precipitation for each county. Daily values were aggregated into several monthly measures to capture average conditions and more extreme weather events. These included total monthly

rainfall, average monthly temperature, maximum 5-day rainfall in the month, the maximum consecutive dry days in a month, and the number of days with temperature above 30°C (Table 4-1).

These weather variables were transformed into anomalies to capture changes relative to average conditions rather than absolute values, and increase the sensitivity of models to subtle shifts in weather patterns that might be obscured when working with absolute values that exhibit significant seasonal variations. Each variable anomaly in county c and month $(x_{c,t}^{AN})$ is calculated by subtracting the long-term average for that month and county ($\bar{x}_{c,m}$) from the observed value and dividing this number by the long-term standard deviation ($\sigma(x_{c,m})$) of meteorological variable as in Equation (4-1) (Shortridge et al., 2016).

$$x_{c,m}^{AN} = \frac{x_{c,m} - \bar{x}_{c,m}}{\sigma(x_{c,m})} \qquad (4-1)$$

Furthermore, 1- and 2-month lagged values were added to the explanatory variable set to account for possible temporal persistence where prolonged conditions ultimately lead to agricultural losses.

Table 4-1. Climatic variables used to describe weather conditions in the modeling approach

Explanatory Variable	Unit	Definition
$\boldsymbol{ppt_m}$	in	monthly rainfall - the sum of daily rainfall values
$\boldsymbol{t_{max}}$	°F	Maximum monthly temperature for every month - mean of daily values
$\boldsymbol{ppt_{5dmax}}$	in	Maximum 5-day rainfall. Calculated as a 5-day moving sum, this approach avoids splitting consecutive days with rainfall, providing a more accurate representation of cumulative rainfall over a 5-day period within a 30-day window.
$\boldsymbol{ppt_{ndry}}$	in	Rain shortage measure. Typically defined as the number of consecutive days with less than 2mm of rainfall, this measure helps capture extended dry periods, even if a heavy rain event occurs later in the month.
$\boldsymbol{t_{max30}}$	°F	Number of days with temperature more than 30°C, identifying heatwaves.
$\boldsymbol{variable_{lag,i}}$		Lag data – weather variable value of previous *i* month(s)

4.2.3 Model structures

One challenge in developing empirical models of damage data is that the data is imbalanced due to the majority of response variable values being equal to zero, and non-zero values exhibit a heavy-tailed distribution due to a small number of very large response values. Strategies for handling imbalanced datasets are either data-based or algorithm-based. Data-based methods aim to modify the class distribution using approaches such as stratified sampling or synthesizing data. Algorithm-based methods try to reinforce learning of the minority class. Multi-step hybrid modeling configuration is a potential algorithmic method, based on the notion that feeding an imbalanced data to a single model rarely results in accurate estimations (Abraham & Tan, 2009; Diaz & Joseph, 2019). A hybrid model breaks down the problem into multiple steps focusing on different aspects of the data and employs specific techniques to handle imbalances. In this study, we compared three model structures aimed at addressing the imbalance and heavy-tailed distribution of loss data:

I. *2-step model*: In this hybrid configuration (Guikema & Quiring, 2012), each observation is initially used in fitting a binary classification model to identify the occurrence/absence damage (nonzero/zero classes). Once nonzero damage incidents are predicted in the initial stage, a regression model is fit to the nonzero class data to estimate the associated crop losses.

II. *2-step stratified model*: This structure adds stratified sampling to the 2-step model to obtain a more balanced distribution of response variable for the classification step (Shashaani et al., 2018). Specifically, k sub-datasets are created through random sampling without replacement of the majority class (zeroes) and combining each with the whole minority class (nonzeroes). A classification model is fit to each subset, and the overall probability is taken as the average across all subset classifiers. The number of sub-datasets (k) is obtained using Equation (4-2) (Shashaani et al., 2018):

$$k = min\left\{2, \left\lceil \frac{\sum_{i=1}^{N} 1\{y(\boldsymbol{X}_i)=0\}}{\sum_{i=1}^{N} 1\{y(\boldsymbol{X}_i)>0\}} \right\rceil\right\} \qquad (4-2)$$

ere $y(\boldsymbol{X}_i)$ is the response value for the vector of climatic variables $\boldsymbol{X}_i$ in the record i and N is the total number of records. The fraction within the curly brackets calculates the ratio of the majority class to the minority class, ensuring that in each sub-dataset, the proportion of non-zero data points and zeros does not exceed 50%. Each data point is labeled by hard voting, that predicts the class label $\hat{y}$ by majority voting of each classifier C_k:

$$\hat{y}(\boldsymbol{X}_i) = mode\{C_1(\boldsymbol{X}_i) + C_2(\boldsymbol{X}_i) + \cdots + C_k(\boldsymbol{X}_i)\} \qquad (4-3)$$

We considered this method to vote for class 1 if at least half of k binary classifiers label the record i as nonzero. The rest of the process is same as 2-step model,

where a regression model is fitted on nonzero response. Our cleaned dataset comprised 6356 data records, with 1385 instances of non-zero response variables for drought events and 1609 instances for storm events. Given this distribution, we chose to set the parameter k to be 4 for both hazard types and set T to 2.

III. *3-step stratified model*: This structure is based on Shashaani et al. (2018) which introduces an intermediate stage before fitting regression models to nonzero responses in an effort to address high skewness and kurtosis in response data. The nonzero response variables are categorized into multiple classes of equal size and different regression models are applied to each class. We set the number of loss classes (m) to a value to 2 based on available data for each hazard type; in future work, this parameter could be tuned for improved model accuracy. Because the county-level spatial resolution made it impractical to do the stratification step using a clustering approach used by Shashaani et al. (2018), we used random under-sampling of the majority class (stratified sampling described in structure II) to create balanced subsets. The three steps are illustrated in Figure 4-3.

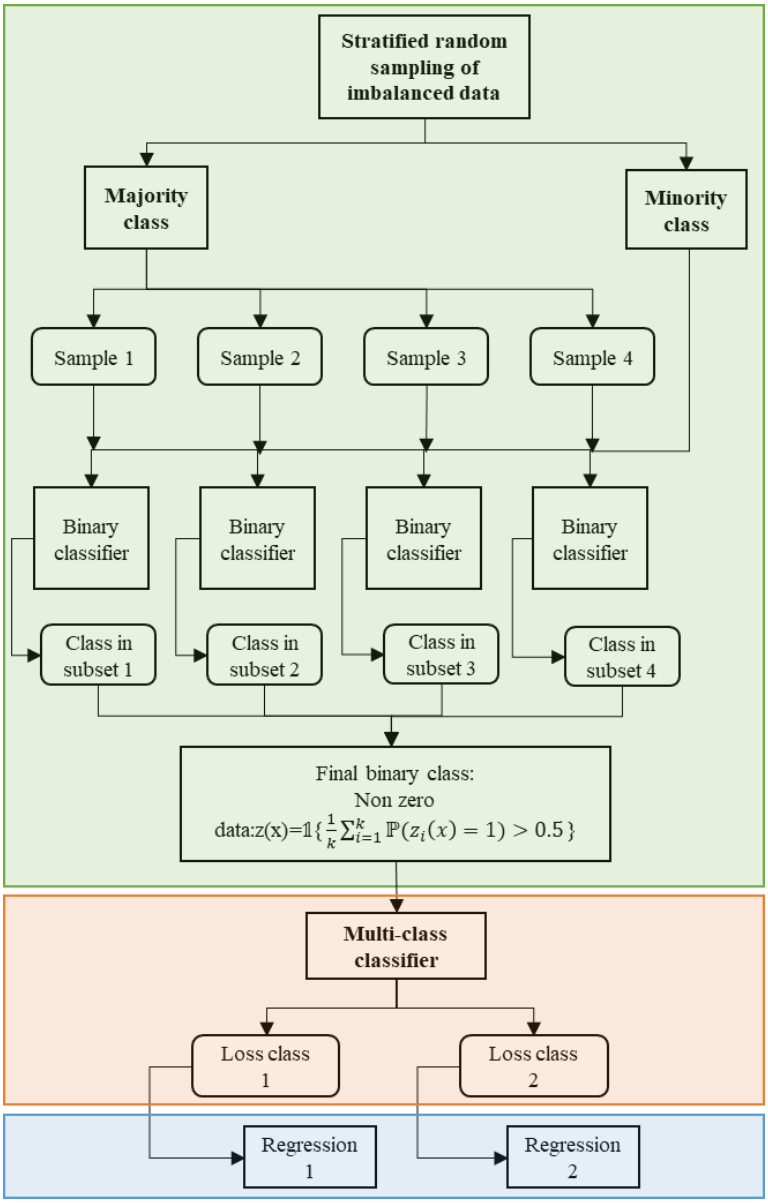


Figure 4-3. Three-step model based on Shashaani et al. (2018)

Summarizing the approach discussed in section 4.2.2, we used the general formulation in Equation 4 for the regression step of the empirical models.

$$Log\big(y(\boldsymbol{X}_i)\big) = f\big(\boldsymbol{X}_i^{AN} + 1|c_i + 1|m_i\big) \qquad (4-4)$$

where $y(\boldsymbol{X}_i)$ is the crop loss of the record i in county c_i and month m_i, and $\boldsymbol{X}_i^{AN}$ represents the vector of explanatory climate variables anomalies listed in Table 1.

4.2.4. Machine learning and statistical approaches

Within each model formulation, we compared three statistical and machine learning modeling approaches to determine how model structure impacts accuracy in capturing the impacts of extreme weather events. The three methods included were:

I. Generalized Linear Mixed Model (GLMM): Extending the linear mixed model to accommodate both fixed and random effects, GLMM captures random variations across different groups (in our case, counties and months). GLMMs manage non-normal data through link functions and distributions from the exponential family (e.g., normal, Poisson, or binomial distributions) (Bolker, 2015). We applied GLMM from binomial family to the classification step of our problem and the Gaussian family to the regression step using 'lme4' package in R (Bates, 2010).

II. Random Forest (RF): A non-parametric ensemble learning method that uses the results of multiple regression or classification trees and outputs the mode of the classes for classification problems or the mean prediction for regression problems (Breiman, 2001). These trees are trained on distinct bootstrapped resamples of the data, with each tree fitted using a randomly selected subset of predictor variables. This approach reduces correlation between trees to minimize prediction variance (Breiman, 2001). The implementation of Random Forest models utilized the 'randomForest' package in R (Liaw & Wiener, 2002).

III. Support Vector Machine (SVM): A supervised machine learning algorithm used for both classification and regression tasks. SVM operates by identifying

the optimal hyperplane that effectively separates data into different classes or predicts a continuous outcome, maximizing the margin between classes (Hearst et al., 1998) to capture complex boundaries and non-linear relationships (Cortes & Vapnik, 1995). SVM was applied to the present problem using the 'e1071' package in R (Chang & Lin, 2011).

4.2.5. Model Evaluation

4.2.5.1. Baseline Null Model

For the purposes of comparison, we used a baseline null model that randomly identified and estimated losses based on county and monthly averages. To develop these null model predictions, we first calculate the percentage of non-zero loss values due to either drought or storm for every unique combination of month and county, as well as the mean value of loss that occurs in that county and month. The first step is a random selection from observations of each county and month, the size of which corresponds to the observed frequency in that month and county, to simulate instances where damage occurs. Equation 4-5 shows damage frequency for each county c in month m ($DF_{m,c}$).

$$\overline{DF}_{m,c} = \frac{1}{Y}\sum_{t=1}^{Y}\left\{y_{m,c_t} > 0\right\}; \qquad \forall m \in \{1,2,\dots,12\};\ \forall c \in \{1,2,\dots,14\} \quad (4-5)$$

where m, c are month and county indices, Y is the number of simulation years and y_{m,c_t} is the damage response to weather conditions in a combination of month and county in year t.

Each observation is assigned a random value sampled from a uniform distribution between 0 and 1; if this value is greater than the damage frequency for that month and

county, the observation is assigned a damage value of zero; otherwise, it is assigned a value of equal to the mean for that specific county and month (Equation 4-6).

$$\hat{y}_{m,c_t} = \begin{cases} \frac{1}{Y}\sum_{t=1}^{Y} y_{m,c_t} & ; u_{m,c_t} \leq \overline{DF}_{m,c} \\ 0 & ; o.w. \end{cases} \quad (4-6)$$

This creates a representative sampling of real-world distribution of damage instances across temporal and spatial dimensions and provides a reference point against which we can compare the additional predictive value gained through the use of empirical models that incorporate climatic predictors and more sophisticated structures.

4.2.5.2. Cross-validation design and error metrics

Our experimental comparison consists of the baseline null model and 9 different combinations of three model structures and three machine learning or statistical approaches. The out-of-sample classification and regression accuracy for each approach was evaluated using a 50-fold holdout cross validation, where each holdout iteration entailed randomly selecting 75% of the data for model training and 25% for model testing. Four error metrics were averaged across the testing dataset observations for each holdout iteration, and these metrics were compared to assess the predictive accuracy of different model structures and forms.

The first two error metrics assess classification performance. Accuracy provides an overall measure of correctness based on the number of true positives (*TP*), true negatives (*TN*), false positives (*FP*), and false negatives (*FN*), and recall assesses the occurrence of false negatives.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (4-7)$$

$$Recall = \frac{TP}{TP + FN} \qquad (4-8)$$

To quantify regression error, we used an adjusted Symmetric Mean Absolute Percentage Error (sMAPE) measure to provide a metric of regression error relative to the magnitude of the observed loss that could be applied even when observed loss values are zero. sMAPE normalizes the relative errors by dividing by both actual and predicted values, with a range between 0% and 100% (Paprotny et al., 2020). The symmetric nature of sMAPE ensures that both small and large values contribute equally, offering a more comprehensive measure of regression accuracy across the full range of damage values. We revised previously used applications of sMAPE to include a weighted term to penalize false negatives, i.e., when the true value of the response y(X_i) is nonzero, but its estimated value $\hat{y}(X_i)$ is zero:

$$WsMAPE = \frac{100}{N_{test}} \sum_{i=1}^{n} w^{\{FN(i)\}} \frac{|y(\boldsymbol{X}_i) - \hat{y}(\boldsymbol{X}_i)|}{|y(\boldsymbol{X}_i)| + |\hat{y}(\boldsymbol{X}_i)|} \qquad ; w \in \{1,10\} \qquad (4-9)$$

where N_{test} is the number of data points in the test set, w is the weight associated to the error of false negative (FN) prediction cost relative to regular error, and $FN(i)$ is a binary variable that is 1 when $y(\boldsymbol{X}_i) > 0 \; AND \; \hat{y}(\boldsymbol{X}_i) = 0$ and zero otherwise. The coefficient w in the WsMAPE formula ensures that the magnifying factor can be applied to the absolute error in cases where the false negative results are observed, and thus applies an additional penalty if models regularly miss loss occurrences. In the results, this metric is reported with w values of both 1 and 10, providing insights into the models' performance under two conditions: one where the error metric treats all errors uniformly, and another where it specifically penalizes false negatives by a factor of 10.

4.2.5.3. Variable importance analysis

To identify the explanatory variables that most directly impacted the likelihood and magnitude of agricultural losses, variable importance analysis was performed on the model with best overall performance. Mean decrease in Gini and increase in node purity were used to characterize variables' influence on the classification and regression steps respectively. The Gini Index is a measure of impurity that quantifies how often a randomly chosen element would be incorrectly classified, and presents a measure of the contribution of each variable to the overall reduction in classification impurity across all classifiers. The increase in node purity quantifies how well a variable helps in reducing the variance in the predicted values. For both metrics, higher values signify greater importance for classification or regression tasks.

4.3. Results

4.3.1. Model Training Performance

The random forest model, implemented with stratified sampling (2-step-strat and 3-step model) during the classification step, yielded highest accuracy and recall overall on the training data (Table 4-2). Regression performance results indicate that, for both hazard types, the Null Model exhibits highest sMAPE values, indicating limited predictive capability without a structured model. The GLMM demonstrates a notable reduction in sMAPE compared to the Null Model, suggesting improved accuracy with a structured approach. RF generally outperforms both Null and GLMM models, showing lower sMAPE values, but the 3-step GLMM achieved a lower estimation error compared to 3-step RF. SVM also exhibits competitive performance, with lower WsMAPE

compared to all the algorithms above. The 3-step SVM model has the best performance overall. The efficacy of the machine learning approaches in capturing complex patterns in estimation is boosted by the algorithmic configuration. Note that all the models are trained in the estimation step by nonzero observed data, resulting in no false negatives and similar error values for different weights in the error metric.

Table 4-2. Performance of different models in the training phase

Step	Hazard	Method	Evaluation Metric	Model Structure		
				2-step	2-step-strat	3-step
Classification	Drought	Null	Accuracy	0.754		
		GLMM		0.832	0.832	0.832
		RF		0.872	0.872	0.872
		SVM		0.855	0.855	0.855
		Null	Recall	0.436		
		GLMM		0.445	0.445	0.445
		RF		0.570	0.570	0.570
		SVM		0.408	0.408	0.408
	Storm	Null	Accuracy	0.648		
		GLMM		0.767	0.767	0.767
		RF		0.836	0.836	0.836
		SVM		0.793	0.793	0.793
		Null	Recall	0.308		
		GLMM		0.237	0.237	0.237
		RF		0.461	0.461	0.461
		SVM		0.205	0.205	0.205
Estimation	Drought	Null	$sMAPE$	30.353		
		GLMM		11.902	11.902	8.789
		RF		11.362	11.355	9.056
		SVM		9.831	9.831	7.425
	Storm	Null	$sMAPE$	39.682		
		GLMM		13.079	13.079	8.972
		RF		12.428	12.401	9.180
		SVM		11.151	11.151	7.650

Figure 4-4 depicts scatterplots of the observed versus predicted damage values, demonstrating how different methods and structures have learned the relationship between hazard losses and weather variables from the observed data.

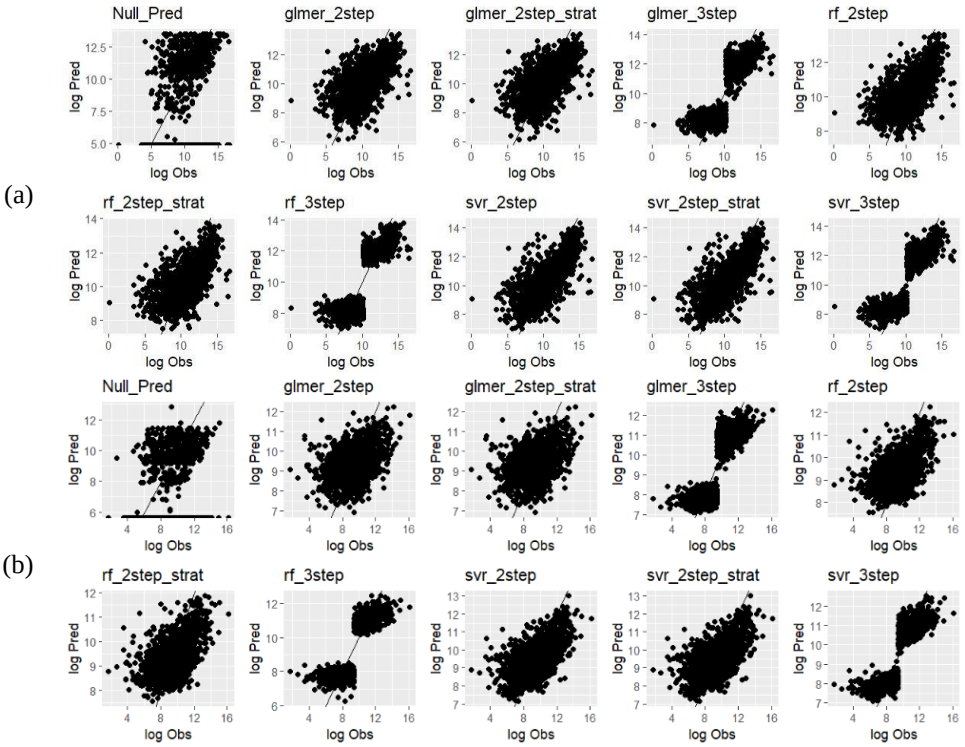


Figure 4-4. Models' training performance demonstrated by prediction versus observation (both log-transformed) scatter plots for nonzero loss observations under (a) drought hazard (b) storm hazard.

4.3.2. Cross-validation Performance

A 50-fold repeated holdout cross validation was used to assess the classification and regression performance of different modeling structures (e.g., 2-step, 2-step stratified, and 3-step) and formulations (RF, SVM, and GLMM) on unseen testing data. Classification performance was assessed using the accuracy and recall metrics, which provide estimates of overall classification and the avoidance of false negatives, respectively. The classification performance for each model formulation is presented in Table 4-3. Bold values indicate the structure and formulation that resulted in the highest performance for each hazard and metric. Note that classification metric values for the 2-

step-stratified and 3-step models are equal because the classification step in each of these approaches was identical.

Across both hazards and classification metrics, the best performance was achieved by the random forest model with stratified sampling (the 2-step stratified and 3-step structures). Model structure has a greater impact on classification performance compared to its formulation. Addressing data imbalance issues by combining stratified sampling with any formulation outperforms the null model in terms of classification accuracy and recall, with the latter exhibiting a more significant improvement. However, the same cannot be said for the two-step structure. RF is the only method that shows higher accuracy and recall compared to the null under both hazards when used in the 2-step structure. GLMM obtains a higher accuracy than the null model but does not always have a better recall, suggesting a limitation in detecting nonzero damages. Comparing results across hazard types, the classification performance metrics using the 2-step RF model with stratified sampling are fairly comparable, but the improvement relative to the null model is greatest for the storm hazard.

Table 4-3. Classification performance of different model formulations and structures in the cross-validation.

Hazard	WsMAPE	Method	Model Structure		
			2-step	2-step-strat	3-step
Drought	Accuracy	Null	0.737		
		GLMM	0.828	0.961	0.961
		RF	0.867	**0.974**	**0.974**
		SVM	0.832	0.966	0.966
	Recall	Null	0.401		
		GLMM	0.438	0.821	0.821
		RF	0.544	**0.880**	**0.880**
		SVM	0.319	0.844	0.844
Storm	Accuracy	Null	0.648		
		GLMM	0.764	0.947	0.947
		RF	0.828	**0.974**	**0.974**
		SVM	0.776	0.961	0.961

Recall	Null	0.308		
	GLMM	0.234	0.791	0.791
	RF	0.425	**0.895**	**0.895**
	SVM	0.139	0.844	0.844

Regression performance based on the sMAPE and weighted sMAPE measures are shown in Figure 4-5. Based on the unweighted sMAPE measure, all advanced model structures and formulations consistently outperform the null model. However, model performance based on the weighted WsMAPE measure, which penalizes false negatives, was significantly more variable. Based on this metric, some model formulations resulted in greater error than the null model. However, the simpler 2-step model structure using RF resulted in the lowest WsMAPE values for both the storm and drought hazards. This suggests that the added complexity of stratified sampling or fitting multiple regression models may not necessarily lead to improved out-of-sample predictive performance. The choice of model structure (2-step, 2-step-strat, 3-step) becomes more critical under the increased emphasis on false negatives. Some structures may inherently lend themselves better to handling extreme events and the associated consequences.

Taking a closer look at differences in the performance of the machine learning methods, RF consistently outperforms SVM in cross validation, suggesting its effectiveness in capturing complex patterns associated with both hazards. RF's ensemble approach allows it to leverage diverse decision trees, each trained on a subset of the data. Accordingly, it tends to generalize well to unseen data, even if individual trees may overfit the training data. SVM might have been more prone to overfitting during training, resulting in better performance on the training set (As shown in Table 4-2) but reduced generalization to new data.

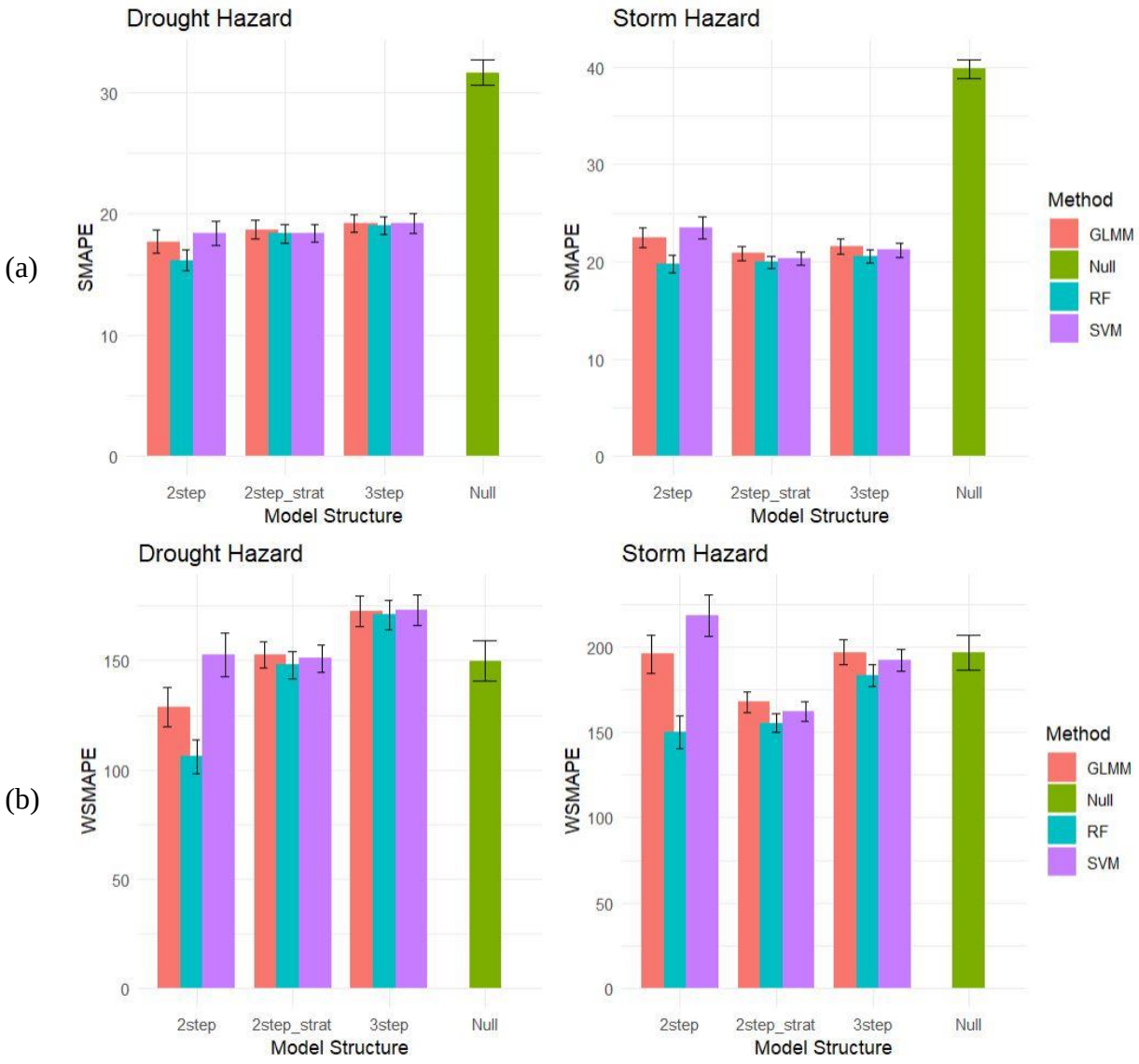


Figure 4-5. Regression performance of different models in cross-validation demonstrated by mean value of WsMAPE metric and associated standard deviation shown by error bars (a) w=1, regular sMAPE. (b) w=10, penalizing false negatives 10 times.

Considering the two pieces together, incorporating stratified sampling has yielded a significant decrease in false negatives and proven effective in mitigating classification errors, as highlighted by recall values. However, the associated regression errors are notably larger, despite relatively similar model structures. While the 2-step model exhibits a poor classification performance in terms of recall metric, WsMAPE value obtained from this model structure for each method is less than WsMAPE values

obtained from other structure alternative. To further investigate the reasons behind this difference in classification and regression performance when dealing with unseen data while both metrics target false negatives, we looked at the performance errors obtained from cross-validation process for drought hazard, averaged over 50 holdout iterations (Table 4-4).

Table 4-4. Average error values from 50-fold holdout cross-validation for drought hazard

Phase	Metric	Model									
		Null	GLMM			RF			SVM		
		2-step	2-step	2-step stratified	3-step stratified	2-step	2-step stratified	3-step stratified	2-step	2-step stratified	3-step stratified
Train	TP	138.14	464	864.12	863.6	562.54	910.72	911.12	378.68	898.62	899.42
	FN	209.86	573	172.88	173.4	474.46	126.28	125.88	658.32	138.38	137.58
	FP	205.9	219.32	0	0	160.08	0	0	64.3	0	0
	TN	1035.1	3510.7	3730	3730	3570	3730	3730	3665.7	3730	3730
Test	TP	138.14	152.2	285.8	285.82	189.12	306.22	306.54	111	293.24	293.18
	FN	209.86	195.8	62.2	62.18	158.88	41.78	41.46	237	54.76	54.82
	FP	205.9	78.3	0	0	52.52	0	0	29.6	0	0
	TN	1035.1	1162.7	1241	1241	1188.5	1241	1241	1211.4	1241	1241
	sMAPE	31.50	17.71	18.70	19.20	16.16	18.34	19.03	18.38	18.39	19.24
	Sd sMAPE	0.99	0.95	0.74	0.75	0.83	0.76	0.76	1	0.77	0.78
	WsMAPE	150.38	128.63	152.78	172.61	106.17	147.69	170.66	152.64	151.01	173.33
	Sd WsMAPE	8.41	8.94	5.84	6.70	7.46	6.18	6.73	10.06	6.33	6.91
	TP error	323500	343800	178200	218200	285100	174500	226400	379600	167100	228000
	FN error	215100	140800	271200	241900	143700	273400	236800	144400	270600	237800
	FP error	522	0	0	0	0	0	0	0	0	0

These performance errors suggest that while the number of false negatives classified by each method combined with the 2-step structure is higher compared to its other structures, these false negatives are linked to lower magnitudes of losses. Consequently, this leads to a lower WsMAPE value in the regression step.

4.3.3. Variable importance results

Variable importance analysis was performed on the 2-step RF model, which exhibited the best overall performance on the classification and regression steps for both drought and storm hazards.

Table 4-5 summarizes the values of these importance metrics for both hazard types, indicating the four variables with the greatest values in bold. This analysis emphasizes the role of geographical and seasonal factors (county and month) in predicting losses for both hazard types and the advantage of modeling structures that can account for hierarchical or clustered observations across different locations or time periods. Monthly rainfall, in terms of both total amounts and the occurrence of extreme rainfall events characterized by the five-day rainfall maximum, plays a crucial role the likelihood and magnitude of storm damage. Drought damage is influenced by a combination of rainfall and temperature explanatory variables. The relatively high values of variable importance for lagged weather variables underscores the importance of past weather conditions in predicting current crop losses and enhancing model accuracy.

Table 4-5. Variable importance metrics for different explanatory variables used in classification and regression steps of the 2-step RF model applied to storm and drought hazard data

Weather variable (Anomaly)	Lag	Storm		Drought	
		Classification Mean Decrease Gini	Regression Increase in Node Purity	Classification Mean Decrease Gini	Regression Increase in Node Purity
Max temperature	0	95.6	213	93.3	**369**
Precipitation	0	**120**	**377**	85.7	**409**
High heat days	0	99.1	181	**141**	303
Rain shortage days	0	105	220	79.6	274
Max 5-day rainfall	0	**113**	260	85.4	286
County	-	**177**	**601**	**145**	**623**
Month	-	**158**	**444**	**234**	**854**
Max temperature	1	98.2	221	82.8	248
Precipitation	1	112	**263**	68.9	264
High heat days	1	73.1	161	**104**	346
Rain shortage days	1	110	214	72.7	228
Max 5-day rainfall	1	105	173	70.2	247
Max temperature	2	99.3	214	80.7	255
Precipitation	2	99.4	176	72.0	218
High heat days	2	67.8	142	87.0	206
Rain shortage days	2	92.6	186	73.1	228
Max 5-day rainfall	2	98.5	198	68.1	239

Figure 4-6 presents partial dependence plots (PDPs) that demonstrate the marginal effect of the top important continuous variables in classification and regression steps for both hazard types. PDPs show how the average predicted value changes when the specified feature(s) vary across their marginal distribution (Friedman, 2001).

The values represented as $\hat{y}$ in the classification model's partial dependence plots (PDP) represent the predicted probability of belonging to the positive class (damage occurrence). In the regression step, $\hat{y}$ values represent the predicted values of the continuous response variable on the log scale. All x-axis values on the partial dependence plots for climatic variables are for anomaly values; for example, a value of 1 corresponds to a value that is one standard deviation greater than the mean for that specific variable.

Examining the outcomes of drought hazard (panel a) for the classification model indicates a notably elevated likelihood of experiencing a nonzero crop loss during warmer months and when the number of days exceeding 30°C in the current or previous month is higher than average. The PDPs of the regression model (panel b) demonstrate that the most substantial losses occur during dry summers with rainfall less than the average and high temperatures.

As demonstrated in Figure 4-6, panel b, storm events leading to significant crop losses are more likely to occur with high precipitation values, in terms of total precipitation in the current month and one moth prior and greater values of five-day maximum rainfall. A noteworthy threshold effect is observed in the precipitation partial dependence plots (PDPs) of the regression model. This suggests that the vulnerability of agriculture in the peninsula to storm events is not limited to extremely rare and extreme occurrences of storms. It also demonstrates the importance of using nonlinear model formulations that can capture these relationships.

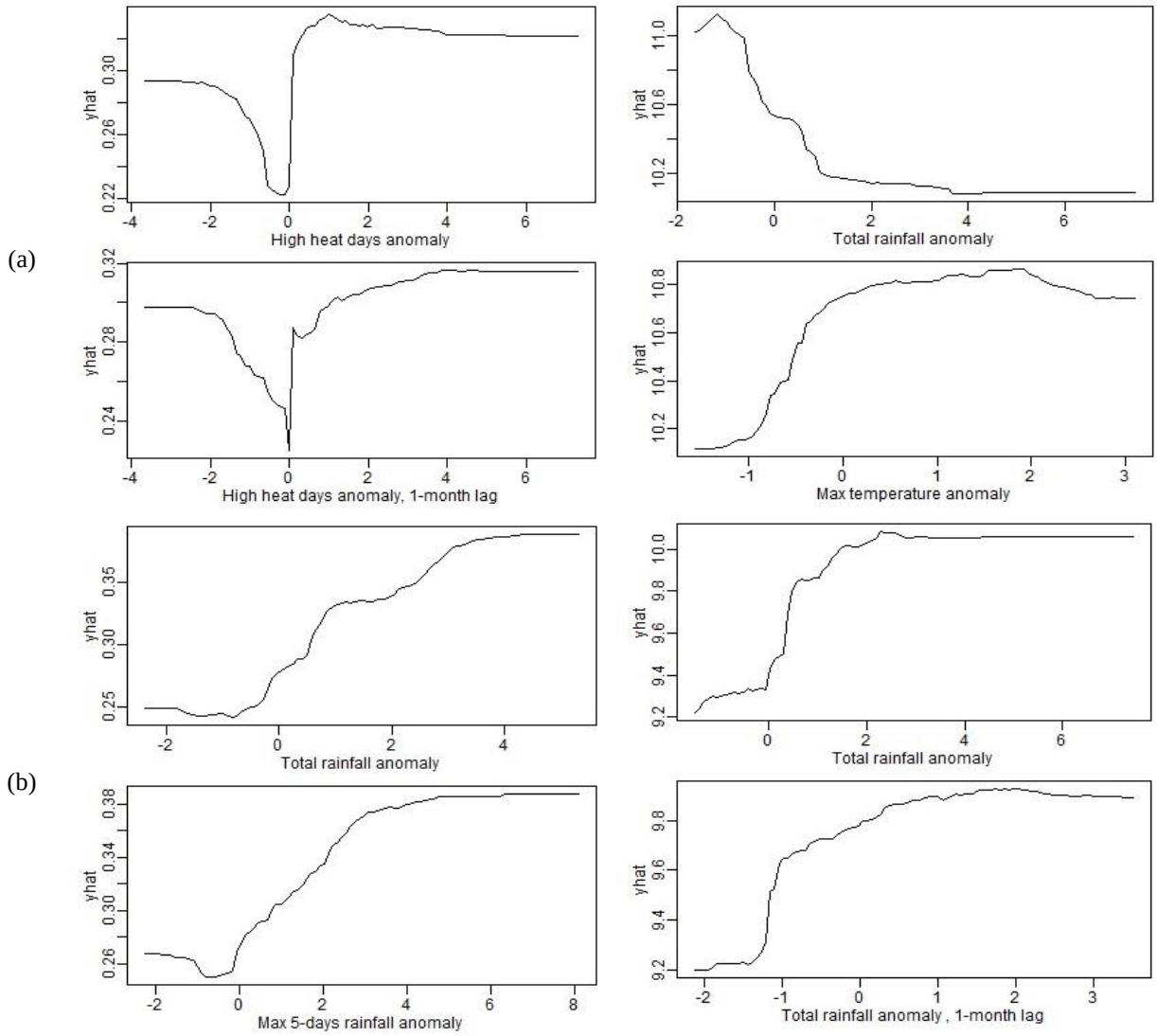


Figure 4-6. Partial dependence of continuous important variables in different steps of classification (left) and regression (right) for (a) drought hazard (b) storm hazard.

Spatial and temporal variations under drought hazard is illustrated in Figure 4-7. Classification results (panel a) demonstrate that the likelihood of damage is also highest in the three counties of Delaware. The PDPs of the regression model (panel b) demonstrate that the most substantial losses occur during dry summers with rainfall less than the average and high temperatures. This encompasses not only Delaware but also the southern portion of the peninsula.

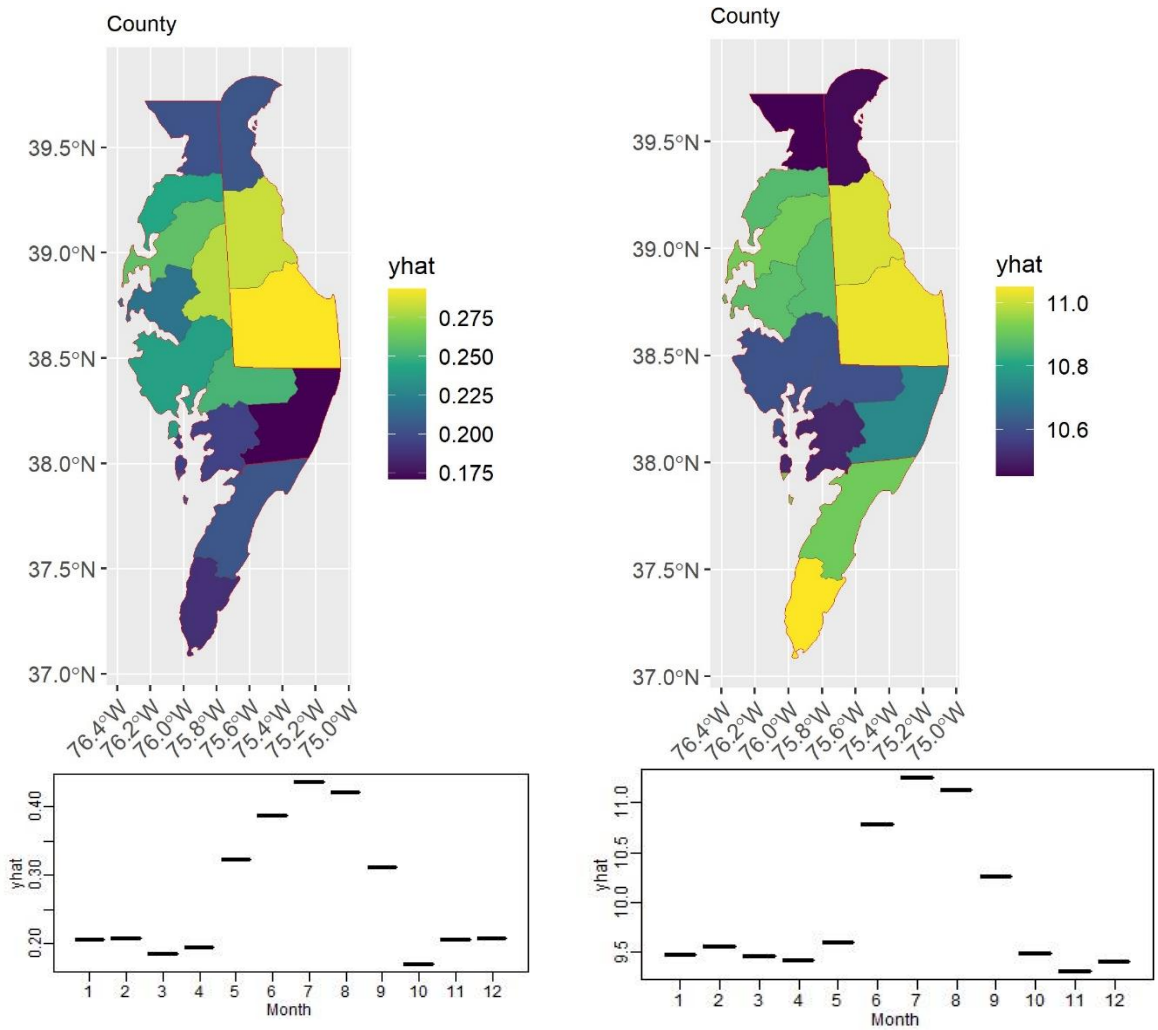
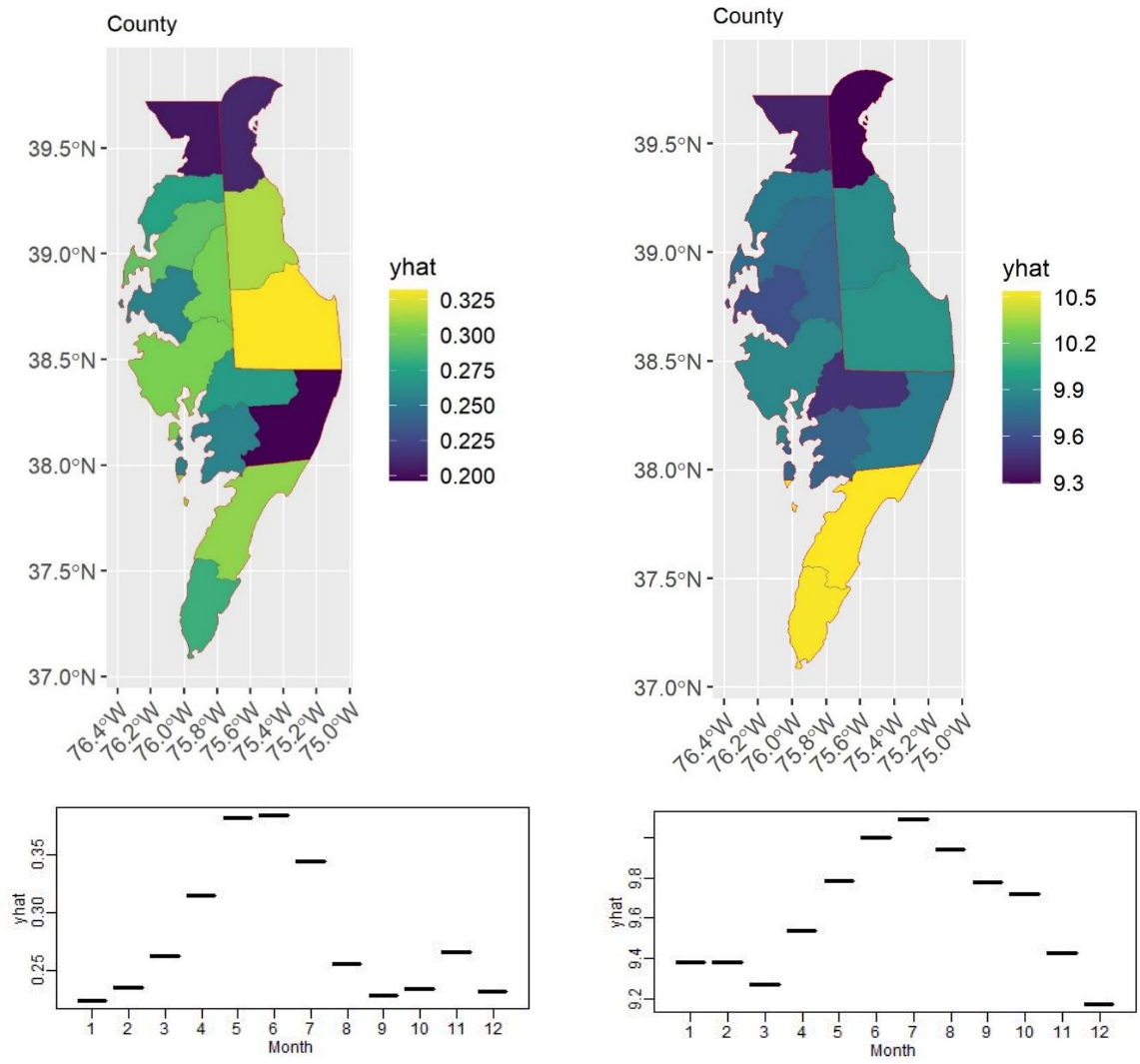


Figure 4-7. Partial dependence of county and month variables in drought hazard assessment (a) classification step (b) regression step.

Storm events (Figure 4-8) leading to significant crop losses are more likely to occur in the summer months. The geographic patterns demonstrate that while these storm-induced losses are infrequent in the Virginia counties, their magnitudes are high when they do occur compared to the rest of the peninsula. High precipitation values, in terms of total precipitation in the current month and one moth prior and greater values of five-day maximum rainfall, are all associated with a higher likelihood and magnitude of damage. A noteworthy threshold effect is observed in the precipitation partial dependence

plots (PDPs) of the regression model. This suggests that the vulnerability of agriculture in the peninsula to storm events is not limited to extremely rare and extreme occurrences of storms.



(a) (b)

Figure 4-8. Partial dependence of county and month variables in storm hazard assessment (a) classification step (b) regression step

4.4. Discussion

The primary goal of this study was to assess the capability of different machine learning approaches and model structures in replicating observed crop damage caused by extreme weather conditions. We used financial loss as a response variable to encompass a broad spectrum of weather-induced losses, particularly those hard to capture by traditional consideration of marginal changes in yield. Machine learning methodologies introduce more flexibility by easing assumptions about distributions and forms of relationships and offer solutions for addressing unconventional characteristics of financial response variable (Lyubchich et al., 2019).

Compared to a null model of losses based on county and monthly average losses and damage frequencies, random forest improved crop damage classification accuracy and regression errors for out-of-sample losses under both hazard types. The significant improvement in classification accuracy through stratified sampling suggests addressing data-related issues as a fundamental initial step for sufficient model performance. Regression outcomes revealed a discernible preference for simpler modeling approaches with greater out-of-sample accuracy, particularly in accurately identifying the occurrence of extreme damages.

The insights obtained from our results are accompanied by several limitations that should be acknowledged in our performance interpretation. First, the focus of this study was placed on discerning the broader effects of formulations and algorithmic structures on the overall performance of hazard-damage modeling, and didn’t proceed to fine-tuning of individual approach parameters. Moving forward, this limitation could be addressed by future research that employs detailed optimization of model parameters to further

enhance performance. Second, this study used several widely available explanatory variables, but incorporation of others (such as wind speed for storm events) could potentially improve the description of extreme weather conditions. However, even conditioned on temperature and precipitation only, machine learning approaches demonstrate promising performance.

A key finding of this research is that constructing multi-step machine learning models for loss data appears to necessitate a tradeoff between predicting the occurrence of damage and quantifying its magnitude. An evident area for future investigation lies in refining these distinct components and discerning the circumstances under which stakeholders are primarily concerned with the mere presence of damage versus those wherein the magnitude of damage is of greater concern, or essentially both. This tradeoff between classification and regression performance of multi-step models has not been investigated by prior studies for other types of damage data, such as hurricane-induced power outages (Shashaani et al., 2018), which introduces a valuable avenue for the future research in other domains characterized by extreme distributions. Last, it is acknowledged that crop yield is impacted not only by seasonal agroclimatic conditions but also by the strategic choices made by farmers in advance, including decisions on the types of crops to cultivate (Parry, 1986). Recognizing that farmers are inclined to adjust their practices in response to changing conditions, the consideration of what Smit et al. (1996) distinguish as response (a decision to change a crop or management system) from effects (change in yield given a chosen crop and management system) in the context of climate change adaptation emerges as an area for future research. Incorporating explanatory variables that describe common agricultural practices in a location could be

particularly valuable in demonstrating which approaches can best address impacts from extreme weather. Identifying the conditions with the greatest impact on losses through variable importance analysis can support in identifying strategies that address climate variability (Khan et al., 2021). and optimize crop yield (E. Han et al., 2017) in the face of increasingly extreme weather conditions.

4.5. Conclusion

This research compared multiple statistical and machine learning modeling approaches to assess the degree to which they can predict agricultural losses and identify what weather conditions are associated with the greatest financial losses. By using financial losses as a response variable instead of marginal differences in crop yield, we are able to model largest damages stemming from extreme weather events and simulate impacts across a highly diverse agricultural system that includes multiple types of production. Our results demonstrate how machine learning approaches like random forest can be combined with multi-step modeling formulations to enhance predictive ability in simulating irregularly distributed loss data, even using small number of explanatory variables. Regional characteristics and seasonal variations in crop damage modeling are particularly important in crop damage estimation, and the inclusion of explanatory variables that represent average and extreme conditions as well as lagged conditions from periods leading up to damage were key variables required for accurate predictions. Accurate models that can simulate diverse agricultural losses in the face of increasingly extreme climate conditions provides an important complement to crop-specific studies focused on changes in yields, ultimately leading to a more comprehensive understanding of the risks that climate change poses to food production.

5. Conclusion

5.1. Summary of findings

This dissertation aims to evaluate the impacts of climatic hazards on coastal communities and their socio-economic responses to different types of climatic events, varying in terms of duration, intensity, and hazard type, using complex systems methodologies. Through the application of agent-based modeling and machine learning techniques, this study delves into two specific socio-ecological interactions: the relationship between sea-level rise and repetitive flooding and human relocation, and the correlation between extreme weather events and agricultural losses. The significance of this research lies in its recommendations regarding the consideration of spatial disparities and the importance of conducting region-specific, small-scale studies, as well as acknowledging the heterogeneity in response patterns. Moreover, this work emphasizes the necessity for a forward-looking approach in model development, appropriate scoping of models, and careful selection of response variables to better capture the diverse impacts of both long-term, chronic changes and short-term, acute events. Furthermore, attention is drawn to how certain technical aspects of the model structure can be refined to produce results that are more conducive to informing policy design. The study can be also of interest to the broader community focusing on SES, as many of the practical and modeling challenges are likely to be applicable to other socio-environmental problems.

5.2. Synthesizing across modeling approaches

Through our exploration, we demonstrated the complexity of incorporating human aspects into socio-ecological simulations while leveraging computational

approaches to gain deeper insights. The methods we employed exhibited promising potential in navigating this complexity.

In comparing the two approaches utilized in this study—agent-based modeling (ABM) and machine learning—we discerned distinct strengths. ABMs serve as proper ground for exploratory what-if scenario analyses of complex emergent phenomena. We conceptualize relationships and integrate data inputs and simulate the system under specific scenarios to see simulation outputs that can't simply be perceived by reductive analysis of the system's numerous interactive components and loops at different scales and levels. Conversely, machine learning proved adept at tackling problems where inputs and outputs are known, yet the underlying relationships are too complex to determine, thus relying on blackbox algorithms to understand these relationships (Figure 5-1).

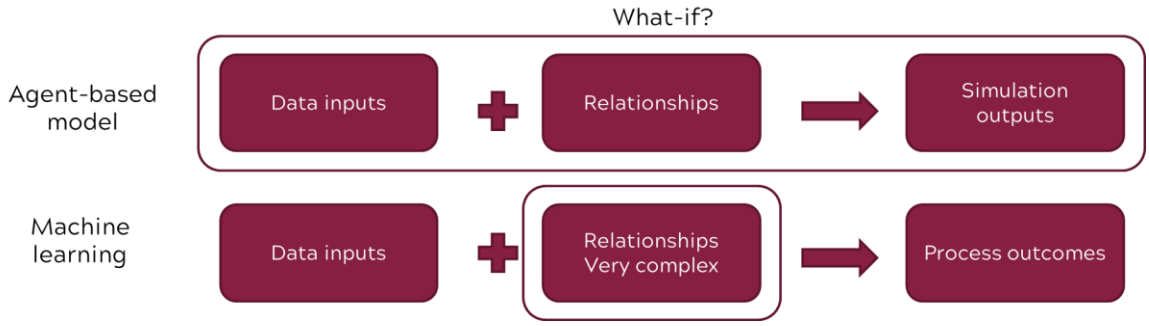


Figure 5-1. Schematic comparison between agent-based approach and machine-learning

5.3. Challenges and future needs

Acknowledging the terrain of complexity, we recognize the substantial data requirements inherent in complex systems approaches to accurately capture relationships. Moreover, we acknowledge the communication and validation challenges, particularly pronounced in ABMs. Bridging the gap between theoretical frameworks and empirical

research on individual-level responses to climate change impacts demands special attention.

5.4. Takeaways

From this dissertation, we conclude that SES are multifaceted and require careful attention to the numerous components and interactions involved. Through our analyses, we demonstrate that to quantitatively assess complex SES, we are equipped with resources: rich data sources and computational modeling methods. This type of research helps us to be prepared for and adapt to challenges posed by climate change. By embracing complexity and using computational tools, we embark on a journey towards a more resilient and sustainable future.

Appendix. ODD+D protocol

This document describes the model, agents, and household decision-making in our model following the ODD+D protocol (Müller et al., 2013).

I. Overview

I.i Purpose

I.i.a What is the purpose of the study?

Our forward-looking stochastic ABM framework aims to assess migration response to flooding from rising sea levels at the municipal to regional level across rural and urban coastal areas. Providing a more comprehensive representation, this model allows us to incorporate top-bottom data-driven population projections and fine-tune model parameters in a futuristic way. The study area chosen to apply the model on consists of 481 census tracts within 16 counties located in coastal Virginia and Maryland, United States.

I.i.b For whom is the model designed?

This model is designed for scientists and municipal disaster resilience planners to increase their understanding of local household migration dynamics under scenarios of flooding and sea-level rise.

I.ii. Entities, state variables, and scales

I.ii.a What kinds of entities are in the model?

This agent-based model (ABM) includes the following entities:

Household agents: spatially explicit entities that are initialized in the model based on census 2020 tract level population data, geographical borders, and an average

household size specific to the case study area. Agents may undertake no action, implement adaptation measures, move to another tract within the same county, or migrate to another out-of-county census tract. In this model, the abstraction level of 200 individuals per agent was considered to balance computational efficiency and statistical representation, resulting in an initial population of 2602 household agents in timestep t = 0. The total number of agents varies over time according to county-level population projections suggested by Hauer and CIESIN (M. Hauer & Center For International Earth Science Information Network-CIESIN-Columbia University, 2020; M. E. Hauer, 2019).

Geospatial patches: Spatially-static entities that represent the geographic map of the study area, and some number of which fall within each census tract's borders based on the defined resolution in the model. An ID allocated to each patch defines its belonging to a tract, meaning that all patches falling within the borders of a tract have the same IDs. Since the finest spatial resolution in the model is census tracts, household agents' creation, movement, and removal processes are tract-level (using tract IDs) and their location within the patches of each tract is selected randomly, as is not meaningful and interpretable to the model.

I.ii.b By what attributes (i.e. state variables and parameters) are these entities characterized?

Household agents possess attributes such as spatial location, satisfaction level, migration threshold based on household-level census data.

Geospatial patches keep census tract information regarding residential setting (urban/rural), initial population, household size, population growth rate (derived from

data-driven county-level population projections), and percent of census tract area inundated.

I.ii.c. What are the exogenous factors / drivers of the model?

There are two exogenous mechanisms deriving migration decisions in the model:

Population dynamics mechanism: Before simulating the effect of flooding, we developed and calibrated a bottom-up baseline model to replicate best available population projections that are driven from historical data (M. Hauer & Center For International Earth Science Information Network-CIESIN-Columbia University, 2020; M. E. Hauer, 2019).

Physical hazard mechanism: Sea-level rise and coastal flooding is a push factor in our model that may force people to migrate. We used on U.S. Army Corps of Engineers' 2015 North Atlantic Coast Comprehensive Study (Cialone et al., 2015; Nadal-Caraballo et al., 2015) statistical coastal flood hazard data. We converted flooded area maps to a percentage of census tracts inundated and interpolated this data to get the percent inundated for more return periods. To feed this flood data into the model, a statistically valid 50-year sequence of flood intensity was required. We generated sequences of random values from uniform distribution and inverted them to obtain sequences of return periods. In order to capture the migration response under different flooding scenarios, four flood storylines were designed, the approach of which was a what-if scenario rather than the whole distribution of storms return periods. The description of the storylines is as follows:

- Storyline 1: Frequent small floods (with the return period of 2-10 years)

- Storyline 2: Frequent small floods and one severe storm (with a return period of 100 years or more) occurring early in the horizon (within the first 15 years)
- Storyline 3: Frequent small floods, one large storm (with a return period of 10-100 years), and two severe storms (with a return period of 100 years or more) occurring late in the horizon (within the last 15 years)
- Storyline 4: Frequent nuisance flooding, one large storms (with a return period of 10-100 years), and three severe storms (with a return period of 100 years or more).

I.ii.d. If applicable, how is space included in the model?

Space is included as census tracts where households are located and move, and flooding events occur.

I.ii.e What are the temporal and spatial resolutions and extents of the model?

The model runs at census tract spatial scale and can be implemented at a different location. It simulates household decisions in yearly timesteps spanning 50 years between 2021 and 2070. Inundation maps are available at census tract level under current sea-level rise scenario which are combined with our statically valid flood storylines generated for 2021-2070 time horizon.

I.iii. Process overview and scheduling

I.iii.a. What entity does what, and in what order?

In each time step, which represents one year, the corresponding inundation data is updated. Based on the push-pull score of each census tract at that specific time step, a beta distribution representing the probability of satisfaction scores for residing agents is created for that census tract. Agents' satisfaction scores are then randomly sampled from the distribution of their current location. Agents compare their satisfaction with a

migration threshold. Agents whose satisfaction falls below the migration threshold decide to move.

To align with nationwide statistics on local and regional-scale mobility developed by the HJCHS (Frost, 2020), we randomly select 65% of these 'dissatisfied' agents to move within their county but to a different census tract, and the remaining agents move to census tracts in counties different from their current county. In their destination decision, agents stochastically select their destination in a way that census tracts with higher push-pull scores in this time step are more likely to be selected by migrants.

The model does not explicitly simulate birth, death, aging processes, or in- and out-migration flows. Instead, we account for all these changes in the study area population by considering growth or decline in population projections for the simulation year. If the population across the study area increases, the required number of agents is created and randomly distributed to census tracts, which are more probable to be assigned for the new agents if their pull score (inherent desirability) is higher. Otherwise, excess agents are randomly selected for removal based on their dissatisfaction; the more dissatisfied an agent, the more likely they are to be eliminated.

II. Design Concepts

II.i Theoretical and Empirical Background

II.i.a Which general concepts, theories or hypotheses are underlying the model's design at the system level or at the level(s) of the submodel(s) (apart from the decision model)? What is the link to complexity and the purpose of the model?

This study proposes a bottom-up framework for the long-term projection of municipal-scale migration due to coastal flood events. Considering heterogeneity in

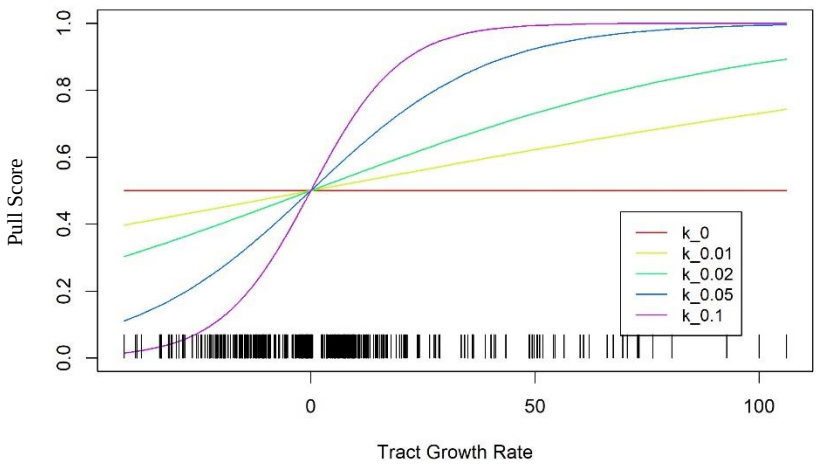
household level migration decisions and municipal scale adaptation policy making as well as the great variance of natural hazard across space, we adopt a stochastic forward-looking bottom-up approach to explore two main questions:

- Can bottom-up decision rules replicate top-bottom data-driven projections?
- How does the introduction of different flooding assumption result in outcomes that diverge from the projections?

II.i.b On what assumptions is/are the agents' decision model(s) based?

The decision-making process for households in our model is grounded in the push-pull theory of migration, which posits that certain features either attract or repel individuals from specific locations. In our model, push-pull effects are derived from a combination of baseline desirability and the flooded area (percentage) of each census tract.

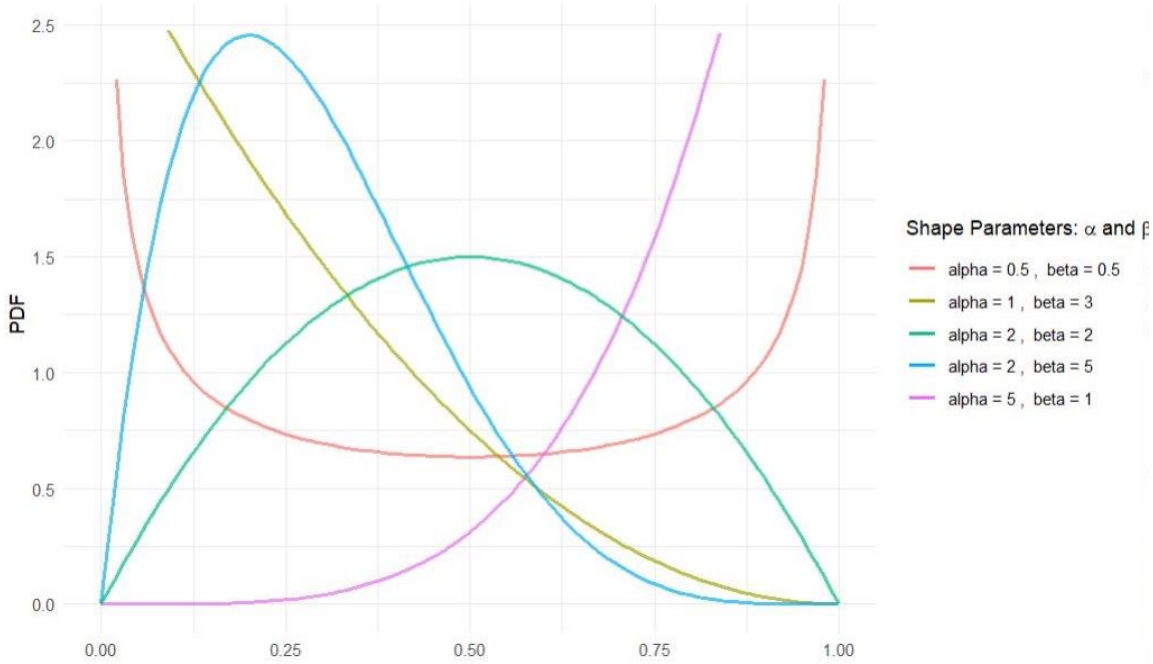
The baseline desirability, represented by the pull score, for each census tract is determined using a logistic function. This function is based on the annual linear growth/decline rate obtained from top-bottom population projections. The underlying concept is that lower growth rates correspond to greater dissatisfaction among agents, making them more likely to move. Conversely, higher growth rates make a location more attractive as a destination for migration. The strength of the baseline pull effect is controlled by the parameter k_{pull} in the logistic function. A k_{pull}value of zero implies no weighting, resulting in random movement between census tracts regardless of their projected growth rate. As the k_{pull}value increases, the weighting becomes stronger, leading more agents to move to high-growth tracts.



Appendix Figure 1. Census tract growth rates are mapped to baseline pull scores through a logistic function with k_{pull}parameter

Agents in our model make migration decisions based on their satisfaction with their current location (whether to migrate or not) and with a potential destination (where to relocate). We assume that the satisfaction of agents within a census tract is a random variable governed by a beta distribution, and the parameters of this distribution are adjusted based on the push-pull score of that census tract.

The beta distribution is limited to the range (0, 1) and is characterized by two shape parameters, α and β. Opting to represent agents' satisfaction scores with a beta distribution was a deliberate choice, as its range corresponds with the pull scores, and the distribution offers significant versatility in depicting various shapes, as illustrated in Figure 2. This adaptability was preferred since we lacked predefined assumptions about the distribution of satisfaction levels within the study area and aimed to calibrate it with a high degree of flexibility.



Appendix Figure 2. Beta distribution. Shape parameters examples are illustrated.

As an initial and straightforward decision rule structure, we assumed that agents would decide whether to move or stay in a census tract based on a weighted average of the pull-push score from the baseline model and the severity of flooding they experience. This weight represents the significance of the flooding factor in decision-making. Various values were tested to observe how the migration response would deviate from top-bottom projections as agents placed more emphasis on the flooding aspect of the decision-making process.

II.i.c Why is/are certain decision model(s) chosen?

The push-pull theory of migration enables us to aggregate various migration drivers into a single metric that represents factors influencing an individual's inclination and ability to relocate. These factors encompass economic considerations, job opportunities, information about alternative localities, demographic characteristics (such

as age, education level, and parenthood), and the psychic cost of migration (Greenwood, 1975).

Given the inherent challenges in directly measuring and simulating individual behaviors, perceptions, preferences, and values, we embrace a stochastic approach. This method involves sampling a random satisfaction value from a known distribution, recognizing that individuals will exhibit diverse levels of satisfaction that are inherently unpredictable.

To maximize the model's alignment with historically-derived top-bottom projections, we leverage the most flexible distributions for various components of our stochastic approach, allowing the model to calibrate itself effectively.

II.i.d If the model/submodel (e.g. the decision model) is based on empirical data, where do the data come from?

The population dynamics system uses Georeferenced U.S. County-Level Population Projections, Total and by Sex, Race and Age, Based on the SSPs, 2020-2100 by Hauer M, Center For International Earth Science Information Network-CIESIN-Columbia University (M. Hauer & Center For International Earth Science Information Network-CIESIN-Columbia University, 2020; M. E. Hauer, 2019). This dataset is available on the internet since 2020 on NASA Socioeconomic Data and Applications Center (SEDAC) website.

The flood hazard system leverages U.S. Army Corps of Engineers' 2015 North Atlantic Coast Comprehensive Study (Cialone et al., 2015; Nadal-Caraballo et al., 2015) statistical coastal flood hazard data. The NACCS statistics represent the combined flood hazards from Nor'easters, tropical storms, and hurricanes and include the influence of

astronomical tides. The NACCS statistical values are reported for present-day sea level at return periods (one over annual exceedance probability) ranging from 1 to 10,000 years. The NACCS probabilistic surge hazard methodology is consistent with the methodology now adopted by FEMA for establishing Flood Insurance Rate Maps.

II.i.e At which level of aggregation were the data available?

Population projection data was available at county level. Inundated area maps developed based on NACCS were available at census tract level for the study area.

II.ii Individual Decision-Making

II.ii.a What are the subjects and objects of the decision-making? On which level of aggregation is decision-making modelled? Are multiple levels of decision making included?

Households make individual decisions to migrate to other places. Only a single level of decision making is included.

II.ii.b What is the basic rationality behind agent decision-making in the model? Do agents pursue an explicit objective or have other success criteria?

Household agents aim to reside in a census tract where their satisfaction from the push-pull effect (inherent desirability and attraction of the area as well as its exposure to flooding) at least meets a threshold. If their current location does not provide this level of satisfaction, they can reach this goal by migrating to another place.

II.ii.c How do agents make their decisions?

Agents evaluate their satisfaction against a predefined calibrated migration threshold. Agents with satisfaction levels below this migration threshold opt to relocate. Whether categorized as within-county or out-of-county migrants, agents stochastically

choose their destination from a weighted list of census tracts, employing a push-pull score-based weighted random draw. This implies that census tracts with higher push-pull scores are more likely to be chosen as the destination for a migrating agent.

II.ii.d Do the agents adapt their behaviour to changing endogenous and exogenous state variables? And if yes, how?

The only exogenous variable driving household decisions are increasing flood extent under sea level rise. We assume that households know the inundated percentage of census tract area in the current timestep. Households thus adapt their behavior to changing coastal flood extent.

II.ii.e Do social norms or cultural values play a role in the decision-making process?

No.

II.ii.f Do spatial aspects play a role in the decision process?

Households determine their satisfaction by considering the inundated percentage in a specific time step and the inherent desirability of their current location. The scale of their relocation distance is defined by a random process that aligns with the statistics of U.S. annual displacements (Frost, 2020), leading to either within-county or out-of-county moves.

II.ii.g Do temporal aspects play a role in the decision process?

Population change varies over time, based on a linear growth rate pulled from top-bottom population projections. Inundation percentage changes over time based on current sea-level rise scenario and NACCS statistics, affecting migration decisions.

II.ii.h To which extent and how is uncertainty included in the agents' decision rules?

Agents' satisfaction from their current location is a stochastic variable.

Flood storylines are generated stochastically, accounting for uncertainties in flooding through time. So, push-pull scores of census tracts are impacted by the uncertainty of flood impacts.

Accordingly, agents' migration decision which is associated with their satisfaction score and their destination decisions are stochastic in their different components.

II.iii Learning

II.iii.a Is individual learning included in the decision process? How do individuals change their decision rules over time as consequence of their experience?

No.

II.iii.b Is collective learning implemented in the model?

No.

II.iv Individual Sensing

II.iv.a What endogenous and exogenous state variables are individuals assumed to sense and consider in their decisions? Is the sensing process erroneous?

Households perceive changes in coastal flood hazard and the inherent desirability of both their current location and potential destinations when considering migration to other census tracts. This sensing process is not prone to errors, but it lacks influence from past experiences, as there is no learning component incorporated into their decision model.

II.iv.b What state variables of which other individuals can an individual perceive? Is the sensing process erroneous?

Households do not sense the state variables of other individuals.

II.iv.c What is the spatial scale of sensing?

Households sense the flood extent and baseline desirability of their current census tract and other census tracts within the county or study area, depending on the scale of their intended relocation distance.

II.iv.d Are the mechanisms by which agents obtain information modelled explicitly, or are individuals simply assumed to know these variables?

No.

II.iv.e Are the costs for cognition and the costs for gathering information explicitly included in the model?

No.

II.v Individual Prediction

II.v.a Which data do the agents use to predict future conditions?

Agents don't predict future conditions in the current version of the model.

II.v.b What internal models are agents assumed to use to estimate future conditions or consequences of their decisions?

NA.

II.v.c Might agents be erroneous in the prediction process, and how is it implemented?

NA.

II.vi Interaction

II.vi.a Are interactions among agents and entities assumed as direct or indirect?

Interactions between agents and their environment (flood hazard) are modeled directly. The population size in a census tract influences the number of households relocating to that area, attributed to the positive impact of a larger population on baseline

desirability—an indicator of the census tract's development level. However, the current version of the model does not account for interactions among agents themselves.

II.vi.b On what do the interactions depend?

Interactions depend on the spatial location and its conditions in a timestep.

II.vi.c If the interactions involve communication, how are such communications represented?

Interactions currently do not involve communication.

II.vi.d If a coordination network exists, how does it affect the agent behaviour? Is the structure of the network imposed or emergent?

No coordination network exists.

II.vii Collectives

II.vii.a Do the individuals form or belong to aggregations that affect and are affected by the individuals? Are these aggregations imposed by the modeller or do they emerge during the simulation?

There is no aggregation involved

II.vii.b How are collectives represented?

NA

II.viii Heterogeneity

II.viii.a Are the agents heterogeneous? If yes, which state variables and/or processes differ between the agents?

Household agents are heterogeneous in state variables. Satisfaction of agents residing in the same census tract is a stochastic variable randomly pulled from census tract's beta distribution. Agents' location in the study area is also heterogenous since

satisfaction distribution varies based on baseline desirability and flooding scores, resulting in exposure to different flood extents and census tract's development level.

II.viii.b Are the agents heterogeneous in their decision-making? If yes, which decision models or decision objects differ between the agents?

Agents are heterogeneous in their decision-making. The decision to migrate from or stay in the current is a function of agent's satisfaction. The decision to where to migrate involves relocation distance category and weighted random draw.

II.ix Stochasticity

II.ix.a What processes (including initialisation) are modelled by assuming they are random or partly random?

Satisfaction score of agents is defined as a random variable. Satisfaction distributions vary by attributes of census tracts. Within-county or out-of-county scale of relocation is defined randomly but subject to HJCHS statistics. Migration destination is a weighted random draw.

The annual change in study area population involves random processes. Agents created or eliminated based on the timestep's population growth rate are randomly assigned using a weighted random draw from push-pull score probabilities. To create agents in the case of population increase, tracts with higher push-pull score are more probable to gained these new agents. If population decreases, tracts with lower push-pull score are more probable to lose agents.

II.x Observation

II.x.a What data are collected from the ABM for testing, understanding and analyzing it, and how and when are they collected?

The total population and percentage of agents moved for each census tract are tracked and exported after each timestep.

Calibrated model parameters were pulled from calibration step to be fed to the flood-informed model. As the Calibrated parameter set, we selected the combination which resulted in the lowest relative root mean square error (Relative RMSE) between modeled and projected population (M. Hauer & Center For International Earth Science Information Network-CIESIN-Columbia University, 2020; M. E. Hauer, 2019) in each simulation year that also had an annual movement percentage within 2% of the HJCHS estimate of 10.7%.

The performance of the flood-informed ABM was compared with the baseline model by calculating the deviation of each stochastic iteration under weight of W from the average of stochastic iterations in the baseline scenario. The root mean square of relative deviations from the baseline was then used to characterize spatial and temporal variations. In order to measure temporal variability across the study area, we aggregated the relative deviations over all census tracts in each simulated year under each storyline. We also defined a measure for relative deviation in each census tract in year 2070. This measure was used to show spatial variation of deviation in the study area in year 2070 for example flood weights.

III. Details

III.i Implementation Details

III.i.a How has the model been implemented?

The model is implemented in Netlogo 6.2

III.i.b Is the model accessible, and if so where?

The model code and data is made publicly available on OSF via: https://osf.io/6r5ab/?view_only=aab370279b9d4c5bbd44b33238bf25c3.

III.ii Initialisation

III.ii.a What is the initial state of the model world, i.e. at time t=0 of a simulation run?

In this section we describe to model application to coastal Virginia and Maryland. Data sources may vary based on the case study location; however, the procedure remains roughly the same.

Census tract boundary shapefiles are pulled from the Census Bureau. For each census tract, households are created based on census tract population divided by 200, i.e. each agent represents 200 individuals of the real population. At t=0, a total of 2602 household agents are created in the study area.

Each census tract is assigned with a push-pull score.

In the flood-informed model, parameters are set to the calibrated values that obtained from calibration step

III.ii.b Is the initialization always the same, or is it allowed to vary among simulations?

The initialization is always the same.

III.ii.c Are the initial values chosen arbitrarily or based on data?

Initial values are based on data.

III.iii Input Data

III.iii.a Does the model use input from external sources such as data files or other models to represent processes that change over time?

The model makes use of census tract shapefile map and population data, top-bottom population projections available in the literature (M. Hauer & Center For International Earth Science Information Network-CIESIN-Columbia University, 2020; M. E. Hauer, 2019) and inundation maps developed for the study area based on NACCS statistics.

III.iv Submodels

III.iv.a What, in detail, are the submodels that represent the processes listed in 'Process overview and scheduling'?

The population dynamics system uses the push-pull theory of migration to replicate the best available population projections using bottom-up heterogenous migration decision rules.

The flood hazard system incorporates the flooding information he NACCS statistical values are reported for present-day sea level at return periods (one over annual exceedance probability) ranging from 1 to 10,000 years. The NACCS probabilistic surge hazard methodology is consistent with the methodology now adopted by FEMA for establishing Flood Insurance Rate Maps.

The logic behind our migration decision submodel has been described above and associated details are discussed and formulated in the Methods section of the manuscript.

III.iv.b What are the model parameters, their dimensions and reference values?

Model parameters are fully described in Table 1 of the manuscript.

III.iv.c How were the submodels designed or chosen, and how were they parameterised and then tested?

The push-pull theory of migration is well established modeling framework of simulation migration flows (Dorigo & Tobler, 1983; Hunter, 2005; Pan, 2019). In this model application of push-pull migration theory is calibrated on available population projections and applied to simulate annually continued migration in response to flooding through the year 2070. Our decision model uses different stochastic processes and simple heuristic rules to account for intricate human environmental interactions and allows for exploring the potential changes in population displacement as a consequence of flooding due to sea-level rise.

References

Abraham, Z., & Tan, P.-N. (2009). A Semi-supervised Framework for Simultaneous Classification and Regression of Zero-Inflated Time Series Data with Application to Precipitation Prediction. *2009 IEEE International Conference on Data Mining Workshops*, 644–649. https://doi.org/10.1109/ICDMW.2009.80

Adade, R., Jaiye, D., Klutse, N. A. B., & Okhimamhe, A. A. (2023). Adapting to Changing Climate: Understanding Coastal Rural Residents' Relocation Intention in Response to Sea Level Rise. *Climate*, *11*(5), Article 5. https://doi.org/10.3390/cli11050110

An, L. (2012). Modeling human decisions in coupled human and natural systems: Review of agent-based models. *Ecological Modelling*, *229*, 25–36. https://doi.org/10.1016/j.ecolmodel.2011.07.010

Aral, M. M. (2020). Knowledge based analysis of continental population and migration dynamics. *Technological Forecasting and Social Change*, *151*, 119848. https://doi.org/10.1016/j.techfore.2019.119848

Atkinson, L. P., Ezer, T., & Smith, E. (2012a). Sea Level Rise and Flooding Risk in Virginia. *Sea Grant Law and Policy Journal*, *5*, 3.

Atkinson, L. P., Ezer, T., & Smith, E. (2012b). Sea Level Rise and Flooding Risk in Virginia. *Sea Grant Law and Policy Journal*, *5*, 3.

Auffhammer, M., Ramanathan, V., & Vincent, J. R. (2012). Climate change, the monsoon, and rice yield in India. *Climatic Change*, *111*(2), 411–424. https://doi.org/10.1007/s10584-011-0208-4

Barnsley, M. J. (2007). *Environmental Modeling: A Practical Introduction*. CRC Press.

Bates, D. C., & Rudel, T. K. (2004). Climbing the "Agricultural Ladder": Social Mobility and Motivations for Migration in an Ecuadorian Colonist Community*. *Rural Sociology*, *69*(1), 59–75. https://doi.org/10.1526/003601104322919900

Beillouin, D., Schauberger, B., Bastos, A., Ciais, P., & Makowski, D. (2020). Impact of extreme weather conditions on European crop production in 2018. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *375*(1810), 20190510. https://doi.org/10.1098/rstb.2019.0510

Benami, E., Jin, Z., Carter, M. R., Ghosh, A., Hijmans, R. J., Hobbs, A., Kenduiywo, B., & Lobell, D. B. (2021). Uniting remote sensing, crop modelling and economics for agricultural risk management. *Nature Reviews Earth & Environment*, *2*(2), 140–159. https://doi.org/10.1038/s43017-020-00122-y

Bilsborrow, R. E., & Henry, S. J. F. (2012). The use of survey data to study migration–environment relationships in developing countries: Alternative approaches to data collection. *Population and Environment*, *34*(1), 113–141. https://doi.org/10.1007/s11111-012-0177-1

Black, R., Adger, W. N., Arnell, N. W., Dercon, S., Geddes, A., & Thomas, D. (2011). The effect of environmental change on human migration. *Global Environmental Change*, *21*, S3–S11. https://doi.org/10.1016/j.gloenvcha.2011.10.001

Bohra-Mishra, P., Oppenheimer, M., & Hsiang, S. M. (2014). Nonlinear permanent migration response to climatic variations but minimal response to disasters. *Proceedings of the National Academy of Sciences*, *111*(27), 9780–9785. https://doi.org/10.1073/pnas.1317166111

Bolker, B. M. (2015). Linear and generalized linear mixed models. In G. A. Fox, S. Negrete-Yankelevich, & V. J. Sosa (Eds.), *Ecological Statistics* (1st ed., pp. 309–333). Oxford University PressOxford. https://doi.org/10.1093/acprof:oso/9780199672547.003.0014

Bonabeau, E. (2002). Agent-based modeling: Methods and techniques for simulating human systems. *Proceedings of the National Academy of Sciences*, *99*(suppl_3), 7280–7287. https://doi.org/10.1073/pnas.082080899

Boon, J., Brubaker, J., & Forrest, D. (2010). Chesapeake Bay Land Subsidence and Sea Level Change: An evaluation of past and present trends and future outlook. *Reports*. https://doi.org/10.21220/V58X4P

Botzen, W. J. W., Bouwer, L. M., & van den Bergh, J. C. J. M. (2010). Climate change and hailstorm damage: Empirical evidence and implications for agriculture and insurance. *Resource and Energy Economics*, *32*(3), 341–362. https://doi.org/10.1016/j.reseneeco.2009.10.004

Bousquet, F., & Le Page, C. (2004). Multi-agent simulations and ecosystem management: A review. *Ecological Modelling*, *176*(3), 313–332. https://doi.org/10.1016/j.ecolmodel.2004.01.011

Breiman, L. (2001). Random Forests. *Machine Learning*, *45*(1), 5–32. https://doi.org/10.1023/A:1010933404324

Brinson, K. R. (2023). *Exploring Hydroclimatic Variability for Agriculture and Water Resource Applications on the Delmarva Peninsula* [Ph.D., University of Delaware]. https://www.proquest.com/docview/2867896762/abstract/92604451AEDE406EPQ/1

Bukvic, A. (2015). Integrated framework for the Relocation Potential Assessment of Coastal Communities (RPACC): Application to Hurricane Sandy-affected areas. *Environment Systems and Decisions*, *35*(2), 264–278. https://doi.org/10.1007/s10669-015-9546-5

Bukvic, A., & Barnett, S. (2023). Drivers of flood-induced relocation among coastal urban residents: Insight from the US east coast. *Journal of Environmental Management*, *325*, 116429. https://doi.org/10.1016/j.jenvman.2022.116429

Bukvic, A., & Owen, G. (2017). Attitudes towards relocation following Hurricane Sandy: Should we stay or should we go? *Disasters*, *41*(1), 101–123. https://doi.org/10.1111/disa.12186

Bukvic, A., Whittemore, A., Gonzales, J., & Wilhelmi, O. (2022). Understanding relocation in flood-prone coastal communities through the lens of place attachment. *Applied Geography*, *146*, 102758. https://doi.org/10.1016/j.apgeog.2022.102758

Bukvic, A., & Zobel, C. W. (2023). Drivers of Slow-Onset Displacement in the Coastal Mid-Atlantic Region and Preferences for Receiving Locations. In M. Hamza, D. Amaratunga, R. Haigh, C. Malalgoda, C. Jayakody, & A. Senanayake (Eds.), *Rebuilding Communities After Displacement: Sustainable and Resilience Approaches* (pp. 121–143). Springer International Publishing. https://doi.org/10.1007/978-3-031-21414-1_6

Burger, A., Oz, T., Crooks, A., & Kennedy, W. G. (2017). Generation of Realistic Mega-City Populations and Social Networks for Agent-Based Modeling. *Proceedings of the 2017 International Conference of The Computational Social Science Society of the Americas*, 1–7. https://doi.org/10.1145/3145574.3145593

Cai, R., & Oppenheimer, M. (Eds.). (2013). *An Agent-Based Model of Climate-Induced Agricultural Labor Migration*. https://doi.org/10.22004/ag.econ.150972

Cairns, G., Ahmed, I., Mullett, J., & Wright, G. (2013). Scenario method and stakeholder engagement: Critical reflections on a climate change scenarios case study. *Technological Forecasting and Social Change*, *80*(1), 1–10. https://doi.org/10.1016/j.techfore.2012.08.005

CEMHS. (2019). *SHELDUS™ | Spatial Hazard Events and Losses Database for the United States* (18.1) [Online Database]. https://sheldus.asu.edu/SHELDUS/

Chalabi, M. (2015, January 29). How Many Times Does The Average Person Move? *FiveThirtyEight*. https://fivethirtyeight.com/features/how-many-times-the-average-person-moves/

Chang, C.-C., & Lin, C.-J. (2011). LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology*, *2*(3), 27:1-27:27. https://doi.org/10.1145/1961189.1961199

Charney, A. H. (1993). Migration and the Public Sector: A Survey. *Regional Studies*, *27*(4), 313–326. https://doi.org/10.1080/00343409312331347585

Chen, Y., Wang, C., Du, X., Shen, Y., & Hu, B. (2023). An agent-based simulation framework for developing the optimal rescue plan for older adults during the emergency evacuation. *Simulation Modelling Practice and Theory*, *128*, 102797. https://doi.org/10.1016/j.simpat.2023.102797

Chumky, T., Basu, M., Onitsuka, K., Parvin, G. A., & Hoshino, S. (2022). Disaster-induced migration types and patterns, drivers, and impact: A union-level study in Bangladesh. *World Development Sustainability*, *1*, 100013. https://doi.org/10.1016/j.wds.2022.100013

Cialone, M. A., Massey, T. C., Anderson, M. E., Grzegorzewski, A. S., Jensen, R.E., Cialone, A., Mark, D. J., Pevey, K. C., Gunkel, B. L., & McAlpin, T. O. (2015). "North Atlantic Coast Comprehensive Study (NACCS) Coastal Storm Model Simulations: Waves and Water Levels" Rep. No. ERDC-CHL TR-15-14. *Engineer Research and Development Center, Vicksburg, MS Coastal and Hydraulics Lab.*

Ciavola, S. J., Jantz, C. A., Reilly, J., & Moglen, G. E. (2014). Forecast Changes in Runoff Quality and Quantity from Urbanization in the DelMarVa Peninsula. *Journal of Hydrologic Engineering*, *19*(1), 1–9. https://doi.org/10.1061/(ASCE)HE.1943-5584.0000773

Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning*, *20*(3), 273–297. https://doi.org/10.1007/BF00994018

Crooks, A., Castle, C., & Batty, M. (2008). Key challenges in agent-based modelling for geo-spatial simulation. *Computers, Environment and Urban Systems*, *32*(6), 417–430. https://doi.org/10.1016/j.compenvurbsys.2008.09.004

Cutter, S. L., Mitchell, J. T., & Scott, M. S. (2000). Revealing the Vulnerability of People and Places: A Case Study of Georgetown County, South Carolina. *Annals of the Association of American Geographers*, *90*(4), 713–737. https://doi.org/10.1111/0004-5608.00219

Daly, C., Halbleib, M., Smith, J. I., Gibson, W. P., Doggett, M. K., Taylor, G. H., Curtis, J., & Pasteris, P. P. (2008). Physiographically sensitive mapping of climatological temperature and precipitation across the conterminous United States. *International Journal of Climatology*, *28*(15), 2031–2064. https://doi.org/10.1002/joc.1688

Davis, K. F., Bhattachan, A., D'Odorico, P., & Suweis, S. (2018). A universal model for predicting human migration under climate change: Examining future sea level rise in Bangladesh. *Environmental Research Letters*, *13*(6), 064030. https://doi.org/10.1088/1748-9326/aac4d4

Davis, K. F., Chhatre, A., Rao, N. D., Singh, D., & DeFries, R. (2019). Sensitivity of grain yields to historical climate variability in India. *Environmental Research Letters*, *14*(6), 064013. https://doi.org/10.1088/1748-9326/ab22db

Dawson, R. J., Peppe, R., & Wang, M. (2011). An agent-based model for risk-based flood incident management. *Natural Hazards*, *59*(1), 167–189. https://doi.org/10.1007/s11069-011-9745-4

De Cian, E., Dasgupta, S., Hof, A. F., van Sluisveld, M. A. E., Köhler, J., Pfluger, B., & van Vuuren, D. P. (2020). Actors, decision-making, and institutions in quantitative system modelling. *Technological Forecasting and Social Change*, *151*, 119480. https://doi.org/10.1016/j.techfore.2018.10.004

de Sherbinin, A., VanWey, L. K., McSweeney, K., Aggarwal, R., Barbieri, A., Henry, S., Hunter, L. M., Twine, W., & Walker, R. (2008). Rural household demographics, livelihoods and the environment. *Global Environmental Change*, *18*(1), 38–53. https://doi.org/10.1016/j.gloenvcha.2007.05.005

Diaz, J., & Joseph, M. B. (2019). Predicting property damage from tornadoes with zero-inflated neural networks. *Weather and Climate Extremes*, *25*, 100216. https://doi.org/10.1016/j.wace.2019.100216

Diffenbaugh, N. S., Davenport, F. V., & Burke, M. (2021). Historical warming has increased U.S. crop insurance losses. *Environmental Research Letters*, *16*(8), 084025. https://doi.org/10.1088/1748-9326/ac1223

D.M. Bates. (2010). lme4: Mixed-effects modeling with R. *Springer*, 470–474.
Döös, B. R. (1994). Environmental Degradation, Global Food Production, and Risk for Large-Scale Migrations. *Ambio*, *23*(2), 124–130.
Dorigo, G., & Tobler, W. (1983). Push-Pull Migration Laws. *Annals of the Association of American Geographers*, *73*(1), 1–17. https://doi.org/10.1111/j.1467-8306.1983.tb01392.x
Eggleston, J. E., & Pope, J. (2013). *Land Subsidence and Relative Sea-Level Rise in the Southern Chesapeake Bay: USGS Circular 1392*. https://pubs.usgs.gov/circ/1392/
Elsawah, S., Pierce, S. A., Hamilton, S. H., van Delden, H., Haase, D., Elmahdi, A., & Jakeman, A. J. (2017). An overview of the system dynamics process for integrated modelling of socio-ecological systems: Lessons on good modelling practice from five case studies. *Environmental Modelling & Software*, *93*, 127–145. https://doi.org/10.1016/j.envsoft.2017.03.001
Engelhart, S. E., Peltier, W. R., & Horton, B. P. (2011). Holocene relative sea-level changes and glacial isostatic adjustment of the U.S. Atlantic coast. *Geology*, *39*(8), 751–754. https://doi.org/10.1130/G31857.1
Ettema, D. (2011). A multi-agent model of urban processes: Modelling relocation processes and price setting in housing markets. *Computers, Environment and Urban Systems*, *35*(1), 1–11. https://doi.org/10.1016/j.compenvurbsys.2010.06.005
Ezer, T., & Atkinson, L. P. (2014). Accelerated flooding along the U.S. East Coast: On the impact of sea-level rise, tides, storms, the Gulf Stream, and the North Atlantic Oscillations. *Earth's Future*, *2*(8), 362–382. https://doi.org/10.1002/2014EF000252
Falco, S. D., Adinolfi, F., Bozzola, M., & Capitanio, F. (2014). Crop Insurance as a Strategy for Adapting to Climate Change. *Journal of Agricultural Economics*, *65*(2), 485–504. https://doi.org/10.1111/1477-9552.12053
FAO. (2021). *The impact of disasters and crises on agriculture and food security: 2021*. FAO. https://doi.org/10.4060/cb3673en
Feng, S., Krueger, A. B., & Oppenheimer, M. (2010). Linkages among climate change, crop yields and Mexico–US cross-border migration. *Proceedings of the National Academy of Sciences*, *107*(32), 14257–14262. https://doi.org/10.1073/pnas.1002632107
Filatova, T., Verburg, P. H., Parker, D. C., & Stannard, C. A. (2013). Spatial agent-based models for socio-ecological systems: Challenges and prospects. *Environmental Modelling & Software*, *45*, 1–7. https://doi.org/10.1016/j.envsoft.2013.03.017
First Street Foundation. (2020). *The first national flood risk assessment: Defining America's growing risk*. FirstStreet. https://firststreet.org/press/2020-first-street-foundation-flood-model-launch/
Fishman, R. (2016). More uneven distributions overturn benefits of higher precipitation for crop yields. *Environmental Research Letters*, *11*(2), 024004. https://doi.org/10.1088/1748-9326/11/2/024004
Friedman, J. H. (2001). Greedy Function Approximation: A Gradient Boosting Machine. *The Annals of Statistics*, *29*(5), 1189–1232.
Frost, R. (2020). *Are Americans Stuck in Place? Declining Residential Mobility in the US*. https://policycommons.net/artifacts/2125811/are-americans-stuck-in-place-declining-residential-mobility-in-the-us/2881109/
Gao, J. (2017). *Downscaling Global Spatial Population Projections from 1/8-degree to 1-km Grid Cells*. NCAR Library. https://doi.org/10.5065/D60Z721H
Gerstengarbe, F.-W., Werner, P. C., Österle, H., & Burghoff, O. (2013). Winter storm- and summer thunderstorm-related loss events with regard to climate change in Germany. *Theoretical and Applied Climatology*, *114*(3), 715–724. https://doi.org/10.1007/s00704-013-0843-y
Gray, C. L., & Mueller, V. (2012). Natural disasters and population mobility in Bangladesh. *Proceedings of the National Academy of Sciences*, *109*(16), 6000–6005. https://doi.org/10.1073/pnas.1115944109

Gray, C., & Wise, E. (2016). Country-specific effects of climate variability on human migration. *Climatic Change*, *135*(3), 555–568. https://doi.org/10.1007/s10584-015-1592-y
Greenwood, M. J. (1975). Research on Internal Migration in the United States: A Survey. *Journal of Economic Literature*, *13*(2), 397–433.
Guikema, S. D., & Quiring, S. M. (2012). Hybrid data mining-regression for infrastructure risk assessment based on zero-inflated data. *Reliability Engineering & System Safety*, *99*, 178–182. https://doi.org/10.1016/j.ress.2011.10.012
Habete, D., & Ferreira, C. M. (2017). Potential Impacts of Sea-Level Rise and Land-Use Change on Special Flood Hazard Areas and Associated Risks. *Natural Hazards Review*, *18*(4), 04017017. https://doi.org/10.1061/(ASCE)NH.1527-6996.0000262
Hailegiorgis, A., Crooks, A., & Cioffi-Revilla, C. (2018). An Agent-Based Model of Rural Households' Adaptation to Climate Change. *Journal of Artificial Societies and Social Simulation*, *21*(4), 4. https://doi.org/10.18564/jasss.3812
Hamilton, S. H., ElSawah, S., Guillaume, J. H. A., Jakeman, A. J., & Pierce, S. A. (2015). Integrated assessment and modelling: Overview and synthesis of salient dimensions. *Environmental Modelling & Software*, *64*, 215–229. https://doi.org/10.1016/j.envsoft.2014.12.005
Han, E., Ines, A. V. M., & Baethgen, W. E. (2017). Climate-Agriculture-Modeling and Decision Tool (CAMDT): A software framework for climate risk management in agriculture. *Environmental Modelling & Software*, *95*, 102–114. https://doi.org/10.1016/j.envsoft.2017.06.024
Han, Y., Ash, K., Mao, L., & Peng, Z.-R. (2020). An agent-based model for community flood adaptation under uncertain sea-level rise. *Climatic Change*, *162*(4), 2257–2276. https://doi.org/10.1007/s10584-020-02802-6
Han, Y., & Peng, Z. (2019). The integration of local government, residents, and insurance in coastal adaptation: An agent-based modeling approach. *Computers, Environment and Urban Systems*, *76*, 69–79. https://doi.org/10.1016/j.compenvurbsys.2019.04.001
Hassani-Mahmooei, B., & Parris, B. W. (2012). Climate change and internal migration patterns in Bangladesh: An agent-based model. *Environment and Development Economics*, *17*(6), 763–780. https://doi.org/10.1017/S1355770X12000290
Hauer, M., & Center For International Earth Science Information Network-CIESIN-Columbia University. (2020). *Georeferenced U.S. County-Level Population Projections, Total and by Sex, Race and Age, Based on the SSPs, 2020-2100* [dataset]. Palisades, NY: NASA Socioeconomic Data and Applications Center (SEDAC). https://doi.org/10.7927/DV72-S254
Hauer, M. E. (2017). Migration induced by sea-level rise could reshape the US population landscape. *Nature Climate Change*, *7*(5), 321–325. https://doi.org/10.1038/nclimate3271
Hauer, M. E. (2019). Population projections for U.S. counties by age, sex, and race controlled to shared socioeconomic pathway. *Scientific Data*, *6*(1), 190005. https://doi.org/10.1038/sdata.2019.5
Hauer, M. E., Evans, J. M., & Mishra, D. R. (2016). Millions projected to be at risk from sea-level rise in the continental United States. *Nature Climate Change*, *6*(7), 691–695. https://doi.org/10.1038/nclimate2961
Hearst, M. A., Dumais, S. T., Osuna, E., Platt, J., & Scholkopf, B. (1998). Support vector machines. *IEEE Intelligent Systems and Their Applications*, *13*(4), 18–28. https://doi.org/10.1109/5254.708428
Hemmati, M., Mahmoud, H. N., Ellingwood, B. R., & Crooks, A. T. (2021). Shaping urbanization to achieve communities resilient to floods. *Environmental Research Letters*, *16*(9), 094033. https://doi.org/10.1088/1748-9326/ac1e3c
Hewitt, K. (2014). *Regions of Risk: A Geographical Introduction to Disasters*. Routledge. https://doi.org/10.4324/9781315844206
Homer, C., Dewitz, J., Yang, L., Jin, S., Danielson, P., Xian, G., Coulston, J., Herold, N., Wickham, J., & Megown, K. (2015). Completion of the 2011 National Land Cover Database for the

Conterminous United States – Representing a Decade of Land Cover Change Information. *Photogrammetric Engineering & Remote Sensing*, *81*(5), 345–354. https://doi.org/10.1016/S0099-1112(15)30100-2
Hugo, G. (1996). Environmental Concerns and International Migration. *International Migration Review*, *30*(1), 105–131. https://doi.org/10.1177/019791839603000110
Hunter, L. M. (2005). Migration and Environmental Hazards. *Population and Environment*, *26*(4), 273–302. https://doi.org/10.1007/s11111-005-3343-x
Hunter, L. M., Luna, J. K., & Norton, R. M. (2015). The Environmental Dimensions of Migration. *Annual Review of Sociology*, *41*, 377–397. https://doi.org/10.1146/annurev-soc-073014-112223
Intergovernmental Panel on Climate Change (IPCC). (2022). *The Ocean and Cryosphere in a Changing Climate: Special Report of the Intergovernmental Panel on Climate Change* (1st ed.). Cambridge University Press. https://doi.org/10.1017/9781009157964
Jakeman, A. J., Letcher, R. A., & Norton, J. P. (2006). Ten iterative steps in development and evaluation of environmental models. *Environmental Modelling & Software*, *21*(5), 602–614. https://doi.org/10.1016/j.envsoft.2006.01.004
Javeline, D., & Kijewski-Correa, T. (2019). Coastal homeowners in a changing climate. *Climatic Change*, *152*(2), 259–274. https://doi.org/10.1007/s10584-018-2257-4
John A Hall, Stephen Gill, Jayantha Obeysekera, William Sweet, Kevin Knuuti, & Marburger,John. (2016). *Regional Sea Level Scenarios for Coastal Risk Management: Managing the Uncertainty of Future Sea Level Change and Extreme Water Levels for Department of Defense Coastal Sites Worldwide* (p. 224). STRATEGIC ENVIRONMENTAL RESEARCH AND DEVELOPMENT PROGRAM ALEXANDRIA VA ALEXANDRIA, U.S. Department of Defense. https://apps.dtic.mil/sti/citations/tr/AD1013613
Jones, B., O'Neill, B. C., & Gao, J. (2020). *Global 1-km Downscaled Population Base Year and Projection Grids for the Shared Socioeconomic Pathways (SSPs), Revision 01* [dataset]. Palisades, NY: NASA Socioeconomic Data and Applications Center (SEDAC). https://doi.org/10.7927/Q7Z9-9R69
Joo, J., Kim, N., Wysk, R. A., Rothrock, L., Son, Y.-J., Oh, Y., & Lee, S. (2013). Agent-based simulation of affordance-based human behaviors in emergency evacuation. *Simulation Modelling Practice and Theory*, *32*, 99–115. https://doi.org/10.1016/j.simpat.2012.12.007
Kamilaris, A., & Prenafeta-Boldú, F. X. (2018). Deep learning in agriculture: A survey. *Computers and Electronics in Agriculture*, *147*, 70–90. https://doi.org/10.1016/j.compag.2018.02.016
Karanci, A., Berglund, E., & Overton, M. (2017). An Agent-based Model to Evaluate Housing Dynamics of Coastal Communities Facing Storms and Sea Level Rise. *Coastal Engineering Proceedings*, *35*, 23. https://doi.org/10.9753/icce.v35.management.23
Karcagi Kováts A., & Katona Kovács J. (2012). Factors of population decline in rural areas and answers given in EU member states' strategies. *Studies in Agricultural Economics*, *114*(1), Article 1.
Keller, K., Yohe, G., & Schlesinger, M. (2008). Managing the risks of climate thresholds: Uncertainties and information needs. *Climatic Change*, *91*(1–2), 5–10. https://doi.org/10.1007/s10584-006-9114-6
Kelly (Letcher), R. A., Jakeman, A. J., Barreteau, O., Borsuk, M. E., ElSawah, S., Hamilton, S. H., Henriksen, H. J., Kuikka, S., Maier, H. R., Rizzoli, A. E., van Delden, H., & Voinov, A. A. (2013). Selecting among five common modelling approaches for integrated environmental assessment and management. *Environmental Modelling & Software*, *47*, 159–181. https://doi.org/10.1016/j.envsoft.2013.05.005
Khan, N. A., Qiao, J., Abid, M., & Gao, Q. (2021). Understanding farm-level cognition of and autonomous adaptation to climate variability and associated factors: Evidence from the rice-growing zone of Pakistan. *Land Use Policy*, *105*, 105427. https://doi.org/10.1016/j.landusepol.2021.105427

Khansari, N., Silverman, B. G., Du, Q., Waldt, J. B., Braham, W. W., & Lee, J. M. (2017). An agent-based decision tool to explore urban climate & smart city possibilities. *2017 Annual IEEE International Systems Conference (SysCon)*, 1–6. https://doi.org/10.1109/SYSCON.2017.7934788

Kleinosky, L. R., Yarnal, B., & Fisher, A. (2007). Vulnerability of Hampton Roads, Virginia to Storm-Surge Flooding and Sea-Level Rise. *Natural Hazards*, *40*(1), 43–70. https://doi.org/10.1007/s11069-006-0004-z

Kniveton, D. R., Smith, C. D., & Black, R. (2012). Emerging migration flows in a changing climate in dryland Africa. *Nature Climate Change*, *2*(6), 444–447. https://doi.org/10.1038/nclimate1447

Kniveton, D., Smith, C., & Wood, S. (2011). Agent-based model simulations of future changes in migration flows for Burkina Faso. *Global Environmental Change*, *21*, S34–S40. https://doi.org/10.1016/j.gloenvcha.2011.09.006

Koks, E. E., Jongman, B., Husby, T. G., & Botzen, W. J. W. (2015). Combining hazard, exposure and social vulnerability to provide lessons for flood risk management. *Environmental Science & Policy*, *47*, 42–52. https://doi.org/10.1016/j.envsci.2014.10.013

Koning, K. de, & Filatova, T. (2020). Repetitive floods intensify outmigration and climate gentrification in coastal cities. *Environmental Research Letters*, *15*(3), 034008. https://doi.org/10.1088/1748-9326/ab6668

Kraybill, D. S., & Lobao, L. (Eds.). (2001). *The Emerging Roles of County Governments in Rural America: Findings from a Recent National Survey*. https://doi.org/10.22004/ag.econ.20697

Kukal, M. S., & Irmak, S. (2018). Climate-Driven Crop Yield and Yield Variability and Climate Change Impacts on the U.S. Great Plains Agricultural Production. *Scientific Reports*, *8*(1), Article 1. https://doi.org/10.1038/s41598-018-21848-2

Kung, H.-Y., Kuo, T.-H., Chen, C.-H., & Tsai, P.-Y. (2016). Accuracy Analysis Mechanism for Agriculture Data Using the Ensemble Neural Network Method. *Sustainability*, *8*(8), 735. https://doi.org/10.3390/su8080735

Kuwata, K., & Shibasaki, R. (2015). Estimating crop yields with deep learning and remotely sensed data. *2015 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, 858–861. https://doi.org/10.1109/IGARSS.2015.7325900

Kuwata, K., & Shibasaki, R. (2016). ESTIMATING CORN YIELD IN THE UNITED STATES WITH MODIS EVI AND MACHINE LEARNING METHODS. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, *III–8*, 131–136. https://doi.org/10.5194/isprs-annals-III-8-131-2016

Kwakkel, J. H., & Haasnoot, M. (2019). Supporting DMDU: A Taxonomy of Approaches and Tools. In V. A. W. J. Marchau, W. E. Walker, P. J. T. M. Bloemen, & S. W. Popper (Eds.), *Decision Making under Deep Uncertainty: From Theory to Practice* (pp. 355–374). Springer International Publishing. https://doi.org/10.1007/978-3-030-05252-2_15

Lal, P., Alavalapati, J. R. R., & Mercer, E. (2011). Socio-economic impacts of climate change on rural United States. *Mitigation and Adaptation Strategies for Global Change 16(7):819-844.*, *16*(7), 819–844. https://doi.org/10.1007/s11027-011-9295-9

Lawyer, C., An, L., & Goharian, E. (2023). A Review of Climate Adaptation Impacts and Strategies in Coastal Communities: From Agent-Based Modeling towards a System of Systems Approach. *Water*, *15*(14), Article 14. https://doi.org/10.3390/w15142635

Lee, E. S. (1966). A theory of migration. *Demography*, *3*(1), 47–57. https://doi.org/10.2307/2060063

Lee, S., Jain, S., Ginsbach, K., & Son, Y.-J. (2021). Dynamic-data-driven agent-based modeling for the prediction of evacuation behavior during hurricanes. *Simulation Modelling Practice and Theory*, *106*, 102193. https://doi.org/10.1016/j.simpat.2020.102193

Li, Y., Guan, K., Schnitkey, G. D., DeLucia, E., & Peng, B. (2019). Excessive rainfall leads to maize yield loss of a comparable magnitude to extreme drought in the United States. *Global Change Biology*, *25*(7), 2325–2337. https://doi.org/10.1111/gcb.14628
Liakos, K., Busato, P., Moshou, D., Pearson, S., & Bochtis, D. (2018). Machine Learning in Agriculture: A Review. *Sensors*, *18*(8), 2674. https://doi.org/10.3390/s18082674
Liaw, A., & Wiener, M. (2002). *Classification and Regression by randomForest*. *2*.
Lin, L., Carley, K. M., & Cheng, S.-F. (2016). An agent-based approach to human migration movement. *2016 Winter Simulation Conference (WSC)*, 3510–3520. https://doi.org/10.1109/WSC.2016.7822380
Lobell, D. B., & Burke, M. B. (2010). On the use of statistical models to predict crop yield responses to climate change. *Agricultural and Forest Meteorology*, *150*(11), 1443–1452. https://doi.org/10.1016/j.agrformet.2010.07.008
Lobell, D. B., Schlenker, W., & Costa-Roberts, J. (2011). Climate Trends and Global Crop Production Since 1980. *Science*, *333*(6042), 616–620. https://doi.org/10.1126/science.1204531
Lyubchich, V., Newlands, N. K., Ghahari, A., Mahdi, T., & Gel, Y. R. (2019). Insurance risk assessment in the face of climate change: Integrating data science and statistics. *WIREs Computational Statistics*, *11*(4), e1462. https://doi.org/10.1002/wics.1462
M. L. Parry. (1986). *Some implications of climatic change for human development*. 378–406.
MacKellar, F. L., Lutz, W., McMichael, A. J., & Suhrke, A. (1998). *Population and climate change* (S. Rayner & E. L. Malone, Eds.). Battelle Press. https://iiasa.dev.local/
Magadza, C. H. D. (2000). Climate Change Impacts and Human Settlements in Africa: Prospects for Adaptation. *Environmental Monitoring and Assessment*, *61*(1), 193–205. https://doi.org/10.1023/A:1006355210516
Marchiori, L., Maystadt, J.-F., & Schumacher, I. (2012). The impact of weather anomalies on migration in sub-Saharan Africa. *Journal of Environmental Economics and Management*, *63*(3), 355–374. https://doi.org/10.1016/j.jeem.2012.02.001
Marino, E. (2018). Adaptation privilege and Voluntary Buyouts: Perspectives on ethnocentrism in sea level rise relocation and retreat policies in the US. *Global Environmental Change*, *49*, 10–13. https://doi.org/10.1016/j.gloenvcha.2018.01.002
Massey, D. S. (2015). A Missing Element in Migration Theories. *Migration Letters*, *12*(3), 279–299.
Mastrorillo, M., Licker, R., Bohra-Mishra, P., Fagiolo, G., D. Estes, L., & Oppenheimer, M. (2016). The influence of climate variability on internal migration flows in South Africa. *Global Environmental Change*, *39*, 155–169. https://doi.org/10.1016/j.gloenvcha.2016.04.014
McLeman, R. (2018a). Thresholds in climate migration. *Population and Environment*, *39*(4), 319–338. https://doi.org/10.1007/s11111-017-0290-2
McLeman, R. (2018b). Thresholds in climate migration. *Population and Environment*, *39*(4), 319–338. https://doi.org/10.1007/s11111-017-0290-2
McLeman, R. A. (2014). *Climate and Human Migration: Past Experiences, Future Challenges*. Cambridge University Press.
McLeman, R., & Smit, B. (2006). Migration as an Adaptation to Climate Change. *Climatic Change*, *76*(1), 31–53. https://doi.org/10.1007/s10584-005-9000-7
McMichael, C., Dasgupta, S., Ayeb-Karlsson, S., & Kelman, I. (2020). A review of estimating population exposure to sea-level rise and the relevance for migration. *Environmental Research Letters*, *15*(12), 123005. https://doi.org/10.1088/1748-9326/abb398
Mehdizadeh, M., Nordfjaern, T., & Klöckner, C. A. (2022). A systematic review of the agent-based modelling/simulation paradigm in mobility transition. *Technological Forecasting and Social Change*, *184*, 122011. https://doi.org/10.1016/j.techfore.2022.122011

Meze-Hausken, E. (2000). Migration caused by climate change: How vulnerable are people inn dryland areas? *Mitigation and Adaptation Strategies for Global Change*, *5*(4), 379–406. https://doi.org/10.1023/A:1026570529614

Mills, A. K., Ruggiero, P., Bolte, J. P., Serafin, K. A., & Lipiec, E. (2021). Quantifying Uncertainty in Exposure to Coastal Hazards Associated with Both Climate Change and Adaptation Strategies: A U.S. Pacific Northwest Alternative Coastal Futures Analysis. *Water*, *13*(4), Article 4. https://doi.org/10.3390/w13040545

Mitchell, M., Herschner, C. H., Herman, J. D., Schatt, D. E., Eggington, E., & Center For Costal Resources Management, Virginia Institute Of Marine Scienc. (2013). *Recurrent flooding study for Tidewater Virginia*. https://doi.org/10.21220/V5TG79

Moallemi, E. A., Kwakkel, J., de Haan, F. J., & Bryan, B. A. (2020). Exploratory modeling for analyzing coupled human-natural systems under uncertainty. *Global Environmental Change*, *65*, 102186. https://doi.org/10.1016/j.gloenvcha.2020.102186

Moench, M. (2010). Responding to climate and other change processes in complex contexts: Challenges facing development of adaptive policy frameworks in the Ganga Basin. *Technological Forecasting and Social Change*, *77*(6), 975–986. https://doi.org/10.1016/j.techfore.2009.11.006

Mueller, V., Gray, C., & Kosec, K. (2014). Heat stress increases long-term human migration in rural Pakistan. *Nature Climate Change*, *4*(3), Article 3. https://doi.org/10.1038/nclimate2103

Müller, B., Bohn, F., Dreßler, G., Groeneveld, J., Klassert, C., Martin, R., Schlüter, M., Schulze, J., Weise, H., & Schwarz, N. (2013). Describing human decisions in agent-based models – ODD + D, an extension of the ODD protocol. *Environmental Modelling & Software*, *48*, 37–48. https://doi.org/10.1016/j.envsoft.2013.06.003

Mycoo, M. (2015). Communicating climate change in rural coastal communities. *International Journal of Climate Change Strategies and Management*, *7*(1), 58–75. https://doi.org/10.1108/IJCCSM-04-2013-0042

Myers, N. (2002). Environmental refugees: A growing phenomenon of the 21st century. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, *357*(1420), 609–613. https://doi.org/10.1098/rstb.2001.0953

Nadal-Caraballo, N. C., Melby, J. A., & Gonzalez, V. M. (2015). Statistical Analysis of Historical Extreme Water Levels for the U.S. North Atlantic Coast Using Monte Carlo Life-Cycle Simulation. *Journal of Coastal Research*, *32*(1), 35–45. https://doi.org/10.2112/JCOASTRES-D-15-00031.1

Nannos, N., Bersimis, S., & Georgakellos, D. (2013). Evaluating climate change in Greece through the insurance compensations of the rural production damages. *Global and Planetary Change*, *102*, 51–66. https://doi.org/10.1016/j.gloplacha.2013.01.006

Neupane, K. W., Rubinyi, L., Sivappha, T., & Wang, Y. (2016). *Climate Migrants and Urban Adaptation in India and China*.

Nguyen, H. K., Chiong, R., Chica, M., & Middleton, R. H. (2018). Agent-based Modeling of Inter-provincial Migration in the Mekong Delta, Vietnam: A Data Analytics Approach. *2018 IEEE Conference on Big Data and Analytics (ICBDA)*, 27–32. https://doi.org/10.1109/ICBDAA.2018.8629751

Nicholls, R., Wong, P., Burkett, V., Codignotto, J., Hay, J., McLean, R., Ragoonaden, S., Woodroffe, C., Abuodha, P., Arblaster, J., Brown, B., Forbes, D., Hall, J., Kovats, S., Lowe, J., McInnes, K., Moser, S., Rupp-Armstrong, S., & Saito, Y. (2007). Coastal systems and low-lying areas. *Faculty of Science - Papers (Archive)*. https://ro.uow.edu.au/scipapers/164

NOAA Climate.gov. (n.d.). Retrieved May 4, 2024, from http://www.climate.gov/maps-data/dataset/storms-and-unusual-weather-phenomena-descriptions

Nourali, Z., Shortridge, J. E., Bukvic, A., Shao, Y., & Irish, J. L. (2024). Simulation of Flood-Induced Human Migration at the Municipal Scale: A Stochastic Agent-Based Model of Relocation Response to Coastal Flooding. *Water*, *16*(2), Article 2. https://doi.org/10.3390/w16020263

Pan, G. (2019). The Push-Pull Theory and Motivations of Jewish Refugees. In G. Pan (Ed.), *A Study of Jewish Refugees in China (1933–1945): History, Theories and the Chinese Pattern* (pp. 123–131). Springer. https://doi.org/10.1007/978-981-13-9483-6_9

Pantazi, X. E., Moshou, D., Alexandridis, T., Whetton, R. L., & Mouazen, A. M. (2016). Wheat yield prediction using machine learning and advanced sensing techniques. *Computers and Electronics in Agriculture*, *121*, 57–65. https://doi.org/10.1016/j.compag.2015.11.018

Paprotny, D., Kreibich, H., Morales-Nápoles, O., Castellarin, A., Carisi, F., & Schröter, K. (2020). Exposure and vulnerability estimation for modelling flood losses to commercial assets in Europe. *Science of The Total Environment*, *737*, 140011. https://doi.org/10.1016/j.scitotenv.2020.140011

Parker, K., Horowitz, J., Brown, A., Fry, R., Cohn, D., & Igielnik, R. (2018). *What unites and divides urban, suburban and rural communities* (United States of America) [Report]. Pew Research Center. https://apo.org.au/node/173886

Parris, A. S., Bromirski, P., Burkett, V., Cayan, D. R., Culver, M. E., 1963-, Hall, J., Horton, R. M., Knuuti, K., Moss, R. H., Obeysekera, J., Sallenger, A. H., & Weiss, J. (2012). *Global sea level rise scenarios for the United States National Climate Assessment* (noaa:11124). https://repository.library.noaa.gov/view/noaa/11124

Paul, S. K., & Routray, J. K. (2011). Household response to cyclone and induced surge in coastal Bangladesh: Coping strategies and explanatory variables. *Natural Hazards*, *57*(2), 477–499. https://doi.org/10.1007/s11069-010-9631-5

Perry, E. D., Yu, J., & Tack, J. (2020). Using insurance data to quantify the multidimensional impacts of warming temperatures on yield risk. *Nature Communications*, *11*(1), 4542. https://doi.org/10.1038/s41467-020-17707-2

Perry, R. W., & Lindell, M. K. (1997). Principles for Managing Community Relocation as a Hazard Mitigation Measure. *Journal of Contingencies and Crisis Management*, *5*(1), 49–59. https://doi.org/10.1111/1468-5973.00036

Piguet, E. (2010). Linking climate change, environmental degradation, and migration: A methodological overview. *WIREs Climate Change*, *1*(4), 517–524. https://doi.org/10.1002/wcc.54

Piguet, E. (2022). Linking climate change, environmental degradation, and migration: An update after 10 years. *WIREs Climate Change*, *13*(1), e746. https://doi.org/10.1002/wcc.746

Plantinga, A. J., Détang-Dessendre, C., Hunt, G. L., & Piguet, V. (2013). Housing prices and inter-urban migration. *Regional Science and Urban Economics*, *43*(2), 296–306. https://doi.org/10.1016/j.regsciurbeco.2012.07.009

Pryce, G., Chen, Y., & Galster, G. (2011). The Impact of Floods on House Prices: An Imperfect Information Approach with Myopia and Amnesia. *Housing Studies*, *26*(2), 259–279. https://doi.org/10.1080/02673037.2011.542086

Ratcliffe, M., Burd, C., Holder, K., & Fields, A. (2016). *Defining Rural at the U.S. Census Bureau*. https://doi.org/10.13140/RG.2.2.16410.64969

Reyes, J. J., & Elias, E. (2019). Spatio-temporal variation of crop loss in the United States from 2001 to 2016. *Environmental Research Letters*, *14*(7), 074017. https://doi.org/10.1088/1748-9326/ab1ac9

RMA. (n.d.). U.S. Department of Agriculture Risk Management Agency. Retrieved May 4, 2024, from https://www.rma.usda.gov/

Roberts, M. J., Braun, N. O., Sinclair, T. R., Lobell, D. B., & Schlenker, W. (2017). Comparing and combining process-based crop models and statistical models with some implications for climate change. *Environmental Research Letters*, *12*(9), 095010. https://doi.org/10.1088/1748-9326/aa7f33

Roberts, M. J., Schlenker, W., & Eyer, J. (2013). Agronomic Weather Measures in Econometric Models of Crop Yield with Implications for Climate Change. *American Journal of Agricultural Economics*, *95*(2), 236–243. https://doi.org/10.1093/ajae/aas047

Robinson, C., Dilkina, B., & Moreno-Cruz, J. (2020). Modeling migration patterns in the USA under sea level rise. *PLOS ONE*, *15*(1), e0227436. https://doi.org/10.1371/journal.pone.0227436

Rosenzweig, C., Jones, J. W., Hatfield, J. L., Ruane, A. C., Boote, K. J., Thorburn, P., Antle, J. M., Nelson, G. C., Porter, C., Janssen, S., Asseng, S., Basso, B., Ewert, F., Wallach, D., Baigorria, G., & Winter, J. M. (2013). The Agricultural Model Intercomparison and Improvement Project (AgMIP): Protocols and pilot studies. *Agricultural and Forest Meteorology*, *170*, 166–182. https://doi.org/10.1016/j.agrformet.2012.09.011

Rosenzweig, C., Tubiello, F. N., Goldberg, R., Mills, E., & Bloomfield, J. (2002). Increased crop damage in the US from excess precipitation under climate change. *Global Environmental Change*, *12*(3), 197–202. https://doi.org/10.1016/S0959-3780(02)00008-0

Shashaani, S., Guikema, S. D., Zhai, C., Pino, J. V., & Quiring, S. M. (2018). Multi-Stage Prediction for Zero-Inflated Hurricane Induced Power Outages. *IEEE Access*, *6*, 62432–62449. https://doi.org/10.1109/ACCESS.2018.2877078

Shirzaei, M., Khoshmanesh, M., Ojha, C., Werth, S., Kerner, H., Carlson, G., Sherpa, S. F., Zhai, G., & Lee, J.-C. (2021). Persistent impact of spring floods on crop loss in U.S. Midwest. *Weather and Climate Extremes*, *34*, 100392. https://doi.org/10.1016/j.wace.2021.100392

Shortridge, J. E., Guikema, S. D., & Zaitchik, B. F. (2016). Machine learning methods for empirical streamflow simulation: A comparison of model accuracy, interpretability, and uncertainty in seasonal watersheds. *Hydrology and Earth System Sciences*, *20*(7), 2611–2628. https://doi.org/10.5194/hess-20-2611-2016

Siegle, J. (2023). *The Capitalization Effect of Designating Scenic Rivers, the Indemnification Effects of Successive Droughts, and Using Machine Learning to Price Specialty Crop Insurance* [The Ohio State University].
https://etd.ohiolink.edu/acprod/odb_etd/etd/r/1501/10?clear=10&p10_accession_num=osu1670493499922952

Silva, A. F. R., & Eleutério, J. C. (2023). Effectiveness of a Dam-Breach Flood Alert in Mitigating Life Losses: A Spatiotemporal Sectorisation Analysis in a High-Density Urban Area in Brazil. *Water*, *15*(19), Article 19. https://doi.org/10.3390/w15193433

Silveira, J. J., Espíndola, A. L., & Penna, T. J. P. (2006). Agent-based model to rural–urban migration analysis. *Physica A: Statistical Mechanics and Its Applications*, *364*, 445–456. https://doi.org/10.1016/j.physa.2005.08.055

Singh, N. P., Anand, B., Singh, S., & Khan, A. (2019). Mainstreaming climate adaptation in Indian rural developmental agenda: A micro-macro convergence. *Climate Risk Management*, *24*, 30–41. https://doi.org/10.1016/j.crm.2019.04.003

Smit, B., McNabb, D., & Smithers, J. (1996). Agricultural adaptation to climatic variation. *Climatic Change*, *33*(1), 7–29. https://doi.org/10.1007/BF00140511

Smith, C., Wood, S., & Kniveton, D. (2010). Agent Based Modelling of Migration Decision-Making. *Proceedings of the European Workshop on Multi-Agent Systems (EUMAS)*, 15.

Sornette, D. (2007). *Probability Distributions in Complex Systems* (arXiv:0707.2194). arXiv. https://doi.org/10.48550/arXiv.0707.2194

Stephenson, D. B. (2008). Definition, diagnosis, and origin of extreme weather and climate events. In H. F. Diaz & R. J. Murnane (Eds.), *Climate Extremes and Society* (1st ed., pp. 11–23). Cambridge University Press. https://doi.org/10.1017/CBO9780511535840.004

Sweet, W. V., Obeysekera, J. T. B., Marra, J. J., & Dusek, G. (2018). *Patterns and projections of high tide flooding along the U.S. coastline using a common impact threshold.* [PDF]. https://doi.org/10.7289/V5/TR-NOS-COOPS-086

Sweet, W. V., & Park, J. (2014). From the extreme to the mean: Acceleration and tipping points of coastal inundation from sea level rise. *Earth's Future*, *2*(12), 579–600. https://doi.org/10.1002/2014EF000272
Thiede, B. C., & Gray, C. L. (2017). Heterogeneous climate effects on human migration in Indonesia. *Population and Environment*, *39*(2), 147–172. https://doi.org/10.1007/s11111-016-0265-8
Thober, J., Schwarz, N., & Hermans, K. (2018). Agent-based modeling of environment-migration linkages: A review. *Ecology and Society*, *23*(2), art41. https://doi.org/10.5751/ES-10200-230241
Tisue, S., & Wilensky, U. (2004). NetLogo: A Simple Environment for Modeling Complexity. *International Conference on Complex Systems*, *21*, 16–21.
Tompkins, F., & Deconcini, C. (2014). *Sea-Level Rise and Its Impact on Virginia* (p. 5). World Resources Institute.
Trinh, T. T., & Munro, A. (2023). Integrating a choice experiment into an agent-based model to simulate climate-change induced migration: The case of the Mekong River Delta, Vietnam. *Journal of Choice Modelling*, *48*, 100428. https://doi.org/10.1016/j.jocm.2023.100428
Troy, T. J., Kipgen, C., & Pal, I. (2015). The impact of climate extremes and irrigation on US crop yields. *Environmental Research Letters*, *10*(5), 054013. https://doi.org/10.1088/1748-9326/10/5/054013
Urban, D. W., Sheffield, J., & Lobell, D. B. (2015). The impacts of future climate and carbon dioxide changes on the average and variability of US maize yields under two emission scenarios. *Environmental Research Letters*, *10*(4), 045003. https://doi.org/10.1088/1748-9326/10/4/045003
U.S. Census Bureau. (2021). [dataset]. U.S. Census Bureau. https://www.census.gov/en.html
U.S. Department of Agriculture. (2018). *2018 Census of Agriculture: Irrigation and Water Management Survey*.
https://www.nass.usda.gov/Publications/AgCensus/2017/Online_Resources/Farm_and_Ranch_Irrigation_Survey/
Walsh, S. J., Malanson, G. P., Entwisle, B., Rindfuss, R. R., Mucha, P. J., Heumann, B. W., McDaniel, P. M., Frizzelle, B. G., Verdery, A. M., Williams, N. E., Yao, X., & Ding, D. (2013). Design of an agent-based model to examine population–environment interactions in Nang Rong District, Thailand. *Applied Geography*, *39*, 183–198. https://doi.org/10.1016/j.apgeog.2012.12.010
Yoon, J., Wan, H., Daniel, B., Srikrishnan, V., & Judi, D. (2023). Structural model choices regularly overshadow parametric uncertainty in agent-based simulations of household flood risk outcomes. *Computers, Environment and Urban Systems*, *103*, 101979. https://doi.org/10.1016/j.compenvurbsys.2023.101979
Zhang, R., Liu, D., Du, E., Xiong, L., Chen, J., & Chen, H. (2024). An agent-based model to simulate human responses to flash flood warnings for improving evacuation performance. *Journal of Hydrology*, *628*, 130452. https://doi.org/10.1016/j.jhydrol.2023.130452
Zhuo, L., & Han, D. (2020). Agent-based modelling and flood risk management: A compendious literature review. *Journal of Hydrology*, *591*, 125600. https://doi.org/10.1016/j.jhydrol.2020.125600
Žurovec, O., Čadro, S., & Sitaula, B. K. (2017). Quantitative Assessment of Vulnerability to Climate Change in Rural Municipalities of Bosnia and Herzegovina. *Sustainability*, *9*(7), Article 7. https://doi.org/10.3390/su9071208