Supplementary Materials: ODD+D protocol

This document describes the model, agents, and household decision-making in our model following the ODD+D protocol [1].

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:

<u>Household agents</u>: 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 [2,3].

<u>Geospatial patches</u>: 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?

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

<u>Geospatial patches</u> 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:

<u>Population dynamics mechanism</u>: 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 [2,3].

<u>Physical hazard mechanism</u>: 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 [4,5] 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 [45], 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.

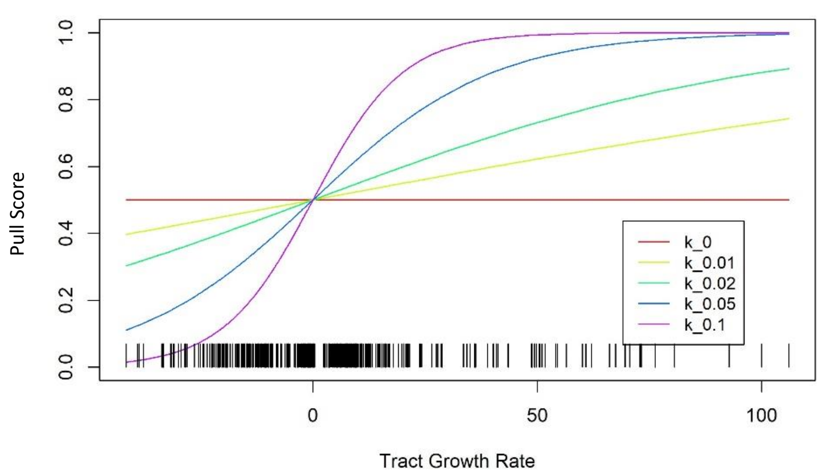

Figure S1. Census tract growth rates are mapped to baseline pull scores through a logistic function with k_{pull} parameter

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.

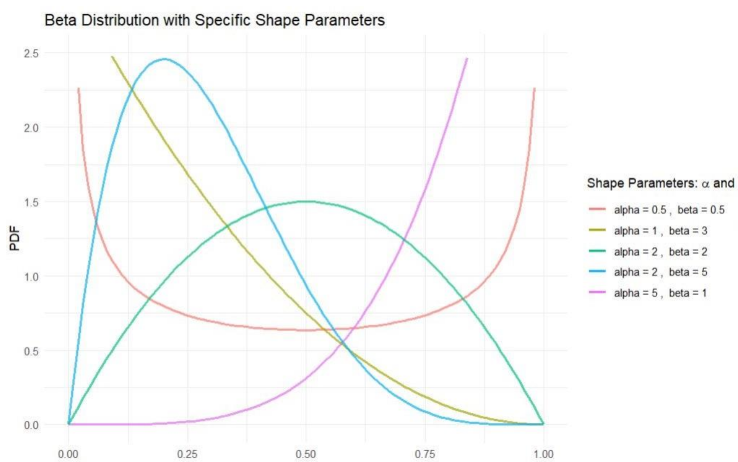

Figure S2. Beta distribution. Shape parameters examples are illustrated

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 [6].

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 [2,3]. 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 [4,5] 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 [7], 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 analysing 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 [2,3] 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 initialisation 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 [2,3] 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 [8–10]. 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

1. Müller, B.; Bohn, F.; Dreßler, G.; Groeneveld, J.; Klassert, C.; Martin, R.; Schlüter, M.; Schulze, J.; Weise, H.; Schwarz, N. Describing human decisions in agent-based models – ODD + D, an extension of the ODD protocol. *Environ Model Softw*. 2013, *48,* 37–48.

2. Hauer, M. Center For International Earth Science Information Network-CIESIN-Columbia University. In Georeferenced U.S. County-Level Population Projections, Total and by Sex, Race and Age, Based on the SSPs, 2020–2100; NASA Socioeconomic Data and Applications Center (SEDAC): Palisades, NY, USA, 2020. Available online: https://sedac.ciesin.columbia.edu/data/set/popdynamics-us-county-level-pop-projections-sex-race-age-ssp-2020-2100 (accessed on 23 November 2022).
3. Hauer, M.E. Population projections for U.S. counties by age, sex, and race controlled to shared socioeconomic pathway. Sci. Data 2019, 6, 190005.
4. 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. North Atlantic Coast Comprehensive Study (NACCS) Coastal Storm Model Simulations: Waves and Water Levels; Rep. No. ERDC-CHL TR-15-14; US Army Engineer Research and Development Center, Coastal and Hydraulics Laboratory: 2015.
5. Nadal-Caraballo, N.C.; Melby, J.A.; Gonzalez, V.M. Statistical Analysis of Historical Extreme Water Levels for the U.S. North Atlantic Coast Using Monte Carlo Life-Cycle Simulation. J. Coast. Res. 2015, 32, 35.
6. Greenwood, M.J. Research on Internal Migration in the United States: A Survey. J. Econ. Lit. 1975, 13, 397–433.
7. Frost, R. Are Americans Stuck in Place? Declining Residential Mobility in the US. 28 April 2020. Available online: https://policycommons.net/artifacts/2125811/are-americans-stuck-in-place-declining-residential-mobility-in-the-us/2881109/ (accessed on 23 November 2022).
8. Dorigo, G.; Tobler, W. Push-Pull Migration Laws. Ann. Assoc. Am. Geogr. 1983, 73, 1–17.
9. Hunter, L.M. Migration and Environmental Hazards. Popul. Environ. 2005, 26, 273–302.
10. Pan, G. The Push-Pull Theory and Motivations of Jewish Refugees. In A Study of Jewish Refugees in China (1933–1945): History, Theories and the Chinese Pattern; Pan, G., Ed.; Springer: Singapore, 2019; pp. 123–131. https://doi.org/10.1007/978-981-13-9483-6_9.