Natural Hazards
https://doi.org/10.1007/s11069-023-06132-5

ORIGINAL PAPER

Single-family housing inventory projection method for natural hazard risk modeling applications

Caroline J. Williams[1] · Rachel A. Davidson[1] · Linda K. Nozick[2] · Meghan Millea[3] · Jamie L. Kruse[3] · Joseph E. Trainor[4]

Received: 16 March 2023 / Accepted: 2 August 2023

Abstract

Today's regional natural hazards loss models rarely incorporate changes in a region's built environment over time, and thus likely misestimate a region's natural hazard risk. Of the existing natural hazard loss models that incorporate changes in the built environment, none are developed at an adequately granular spatiotemporal scale that is appropriate for regional (multi-county) natural hazards loss modeling. This work presents the new Housing Inventory Projection (HIP) method for estimating changes in a region's housing inventory for natural hazards loss modeling purposes. The method includes two modules: (1) the Regional Annual County-Level Housing module, which estimates the annual number of housing units per county over a multi-county region and multi-decadal projection period, and (2) the Single-family Location Estimation module, which estimates the likely location of future single-family housing units across a subcounty grid space. While the HIP method can be applied over a range of spatiotemporal scales, we present a case study that estimates the number of single-family houses per 1 km^2 grid cell in the state of North Carolina for each year from 2020 to 2049. We then used these projections to estimate how a future housing stock would experience a Hurricane Florence-type event. Future housing projections suggest that between 2020 and 2049, nearly 2900 new houses will be built, each year, in areas that experienced at least two feet of flooding following Hurricane Florence.

Keywords Housing · Building inventory · Projections · Hurricane

✉ Rachel A. Davidson
rdavidso@udel.edu

[1] Department of Civil and Environmental Engineering, University of Delaware, Newark, DE 19716, USA

[2] School of Civil and Environmental Engineering, Cornell University, Ithaca, NY 14850, USA

[3] Department of Economics, East Carolina University, Greenville, NC 27858, USA

[4] Biden School of Public Policy and Administration, University of Delaware, Newark, DE 19716, USA

Published online: 17 August 2023

1 Introduction

Regional probabilistic natural hazards loss models estimate expected annualized losses over a long time period, but typically, do not incorporate changes in the built environment which will certainly take place during that period. By not incorporating the spatiotemporal changes in a region's built environment, today's natural hazard loss models misrepresent reality and likely underestimate a region's natural hazard risk (Ferguson and Ashley 2017; Wagenaar et al. 2020). As an example, if a hurricane risk assessment had been conducted in New Hanover County (Wilmington), North Carolina following Hurricane Floyd in 1999 based on the building inventory at the time, when there were 159,000 residents living in 77,000 housing units, it would have underestimated the losses that occurred in Hurricane Florence just 19 years later in 2018, by which time there were 224,000 residents living in 109,000 housing units. That 42% growth in housing units would have substantial implications for the hurricane losses. In fact, variation across land development scenarios alone has been found to account for larger variability in natural hazard risk than variation in climate change scenarios (Bryant and Westerling 2014; Wing et al. 2022; Bozzolan et al. 2023).

For the goal of estimating changes in regional housing exposure to natural hazards at a granular scale, we developed a new Housing Inventory Projection (HIP) method and applied the method to estimate the number of single-family houses per 1 km^2 grid cell in the state of North Carolina for each year from 2020 to 2049. We then used these projections to suggest how a future housing stock would experience a Hurricane Florence-type event and thus estimated the expected changes in hurricane risk due to new single-family housing construction. While we present this specific application using a case study in North Carolina for ease of presentation, a user can apply the modeling method for any multi-county region in the USA for different time periods and different area units if desired, as described in Sect. 5. In addition to helping to understand changing regional exposure to natural hazard risk, the HIP method can be useful for a variety of applications, such as economic development or planning, conservation investigations, or real estate modeling.

The HIP method focuses on single-family housing in particular because it represents approximately 60% of the housing stock in the United States (US) (US Census Bureau 2022) and a large portion of wealth for many households, and is vulnerable to hazards. Across the US, single-family houses represented 84% of National Flood Insurance Program (NFIP) claims between 2012 and 2022. Following Hurricane Florence's arrival in North Carolina in 2018, flood insurance claims in single-family residences alone totaled to US$543 million (FEMA 2022).

Following a review of related literature on regional landscape and urbanization modeling methods as well as population and housing change modeling methods in Sect. 2, a description of the steps involved in the HIP method is provided in Sect. 3. The HIP method is then implemented in Sect. 4, providing estimates of the expected annual housing location projections in North Carolina between 2020 and 2049 as well as expected implications for hurricane impacts. Section 5 provides a brief description on how a user could apply the HIP method for alternative applications. The paper concludes with a summary of the key findings and future work.

2 Literature review

As a community develops, its exposure to potential harm from natural hazards increases. This phenomenon is known as the "expanding bullseye effect" when the footprint of a metropolitan region's urban, suburban, and exurban area expands, thus increasing the "bullseye" for natural hazard impact (Strader and Ashley 2015).

Previous research that combined changing exposure models with regional (multi-county) natural hazard risk models largely captures spatiotemporal exposure changes via land cover, population, or housing changes. While each method offers its own advantages, those that result in estimates of the number and location of buildings are most useful for natural hazards risk research because regional natural hazard loss models use building attributes and structural engineering to estimate physical damage to buildings followed by an aggregation of the resulting economic losses (Wang et al. 2020; Federal Emergency Management Agency 2021; Center for Risk-Based Community Resilience Planning 2021; McKenna et al. 2022). Thus, in this review, we focus on those models that produce as output estimates of a region's housing inventory (rather than population or land use class) over space and time.

Further, for use in natural hazard risk modeling, it is desirable for the results to represent a spatial resolution sufficient to incorporate the spatial variation of localized hazard impacts such as flooding (e.g., subcounty), and a spatial extent sufficient to include the entire region affected by a hazard event (e.g., the many counties affected by a hurricane). Additionally, natural hazard events are infrequent, so the time horizon should be sufficient to capture a period during which an event may occur, and greater temporal resolution can be beneficial in allowing nonlinear annual effects to be incorporated in the model and offering flexibility in the modeling framework where different risk mitigation policies (such as land use policy or building code interventions) could be implemented and evaluated at different years in the interested projection periods.

Sections 2.1 and 2.2.1 summarize land use-land cover models and population projections that, though they offer useful insights, are not directly usable for natural hazard loss models because they do not result in building inventory projections. Section 2.2.2 summarizes models that are closest to the HIP model presented herein and that result in housing inventory projections. Those few models differ from HIP in the particulars of the data and modeling processes used.

2.1 Changing landscape and urbanization models

Methods for modeling changing landscapes across large areas are rooted in the land use–land cover (LULC) change literature (Daniel et al. 2016; Sleeter et al. 2017). Many LULC models are used to predict urbanization changes over long periods of time, usually at decadal time intervals. The units of analysis for LULC models are typically at 1 km^2 or less and can span city to national scales. The structures of LULC models vary depending on the application, including logistic regression-type models that predict urban versus non-urban areas over time (Song et al. 2018), classification-type models that identify the LULC category (urban, forest, agriculture, etc.) per area unit over time (Sleeter et al. 2017), and regression-type models that predict quantifiable changes per area unit such as the percent or urban land in an area unit (Gao and O'Neill 2019). Overall, LULC models for large spatial scales tend to utilize linear regression, cellular automata (CA) methods,

machine learning methods, or a combination thereof (National Research Council et al. 2014; Cremen et al. 2022). Aburas et al. (2019), Briassoulis (2019), Musa et al. (2017), and Verburg et al. (2004) provide reviews of different CA and machine learning methods for LULC modeling, as well as commonly used modeling parameters. Cellular automata models are usually set up as classification-type problems and develop a set of transition rules and transition probabilities where each area unit (often a grid cell) can change states based on neighboring characteristics. Machine learning models can be more flexible and designed as logistic, classification, or regression problems. Satellite imagery has been used with convolutional neural networks (CNNs) to classify changes in land cover types per pixel at a given time, and with recurrent neural networks (RNNs) to estimate time series landscape changes (Aburas et al. 2019; Briassoulis 2019; Musa et al. 2017; Verburg et al. 2004).

Studies using urbanization or land cover change models to estimate future regional impacts from natural hazards such as flooding (Song et al. 2018; Ford et al. 2019; Rifat and Liu 2022) or tsunamis (Sleeter et al. 2017) are becoming easier to calibrate using improved granular satellite imagery and advanced modeling methods; however, the land cover change models do not explicitly estimate the number and locations of a region's buildings. With this, a generalized assumption of constant population or housing density per land cover or land use type over time is typically required to estimate the amount of damage or economic loss a region is expected following natural hazard events. Overall, while the LULC literature offers valuable insights on different approaches for modeling spatiotemporal landscape changes across a region, directly modeling the expected changes in a region's building inventory is needed to better estimate future natural hazard impacts.

2.2 Population and housing change models

Methods for modeling the spatiotemporal changes of people or housing units vary depending on the spatial extent, temporal projection, and spatiotemporal granularity required for the analysis, as well as the available of data. For example, housing economics and real estate models are often developed using either supply-and-demand land value models or computable general equilibrium (CGE) models (Cho et al. 2005; Ustaoglu and Lavalle 2017; Ali et al. 2020). However, these methods largely only use economic or demographic trends to predict the number of housing units over a large area (county-level or larger) and do not include methods for estimating the subcounty location of new housing units. For community-level housing estimates, agent-based are often used to estimate where new houses will be built following a set of modeling rules and agent decision-making among relevant stakeholder groups (e.g., developers, home owners, renters) (Parker and Filatova 2008; Magliocca et al. 2011; Filatova 2015; Sanderson et al. 2022). Haer et al. (2020) is a unique study in which an agent-based modeling approach was used to estimate future flood risk across the entire European Union (EU) over a $30'' \times 30''$ grid space under different population growth and migration scenarios governed by government and household decision-making. However, this agent-based model assumes population change throughout the EU follows the same growth pattern as experienced in the Netherlands between 1960 and 2000 and relies on subjective agent decision-making modeling assumptions. While agent-based models can be valuable for understanding emerging dynamics following a series of agent decision-making, the detailed modeling approach is not necessarily required to simply estimate the expected changes in population or housing over a large region over a long time horizon.

Another common method for estimating a region's change in population or housing is a two-step process that combines population projections for larger area units with a downscaling process to allocate the new people or housing units to smaller area units over time (Davidson and Rivera 2003; Theobald 2005; Mann et al. 2014; Hauer et al. 2016; US EPA 2023). The two-step method offers flexibility to estimate the changes across a regional study area where the larger area unit estimates over long projection periods are based on different variables than those used to locate buildings in smaller area units. This is important because it allows larger trends (e.g., economic, demographic) to be captured in the long-term projections while also capturing detailed landscape-based attributes that influence building location in localized areas.

2.2.1 Two-step population projections

Some studies have combined regional population or demographic projections for larger area units (countries) with historical growth rates in smaller area units (e.g., counties, 1/8° grid cells) to produce spatially granular estimates of demographic changes (Jones and O'Neill 2016; Hauer 2019). However, while these granular population projections can be valuable, they ultimately require a headship (number of people per house) assumption to estimate housing units and therefore do not directly align with common regional natural hazard risk models.

2.2.2 Two-step building projections

A number of studies have used the two-step approach to estimate the spatial distribution of housing units in particular by using county-level population projections. This includes Davidson and Rivera (2003), who utilized county-level population projections, headship rates, and census-tract-level housing data to predict the number, location, and types of housing units per census tract for 15 counties in North and South Carolina over 5-years intervals between 2000 and 2020. Jain and Davidson (2007a, b) then used these census tract-level estimates in conjunction with a hurricane loss model to assess the expected change in risk over time. The Spatially-Explicit Regional Growth Model (SERGoM) also combined county-level population projections and headship rates with historic housing block and block group census data, land data, and road network data to estimate the spatial distribution of housing units over a 100 m^2 grid space at decadal intervals for the entire conterminous US (CONUS). Census block data were transformed to a uniform grid space using simple dasymetric mapping techniques, which assumes a homogeneous distribution of housing units across the blocks. Future changes in housing unit densities per grid cell were calculated by (1) applying average growth rates between decades for 16 classification types formed from four housing density levels and four road density levels, and (2) estimating the travel time along major roads from a cell to an urban area. The previous decade's growth rate and urban travel time combine to form the allocation weights that were used to distribute county-level population estimates to housing densities per grid cell via county-specific headship rates in 2000 (Theobald 2005). The Integrated Climate and Land-Use Scenarios (ICLUS) then extended SERGoM to predict decadal changes in both population and housing units over the 100 m^2 grid space by sourcing population projections under five different climate change scenarios to produce five housing projection scenarios in each decade between 2010 and 2100 (US EPA 2023). The ICLUS population projections

were used by Wing et al. (2022) to estimate expected changes in flood risk across the CONUS while the housing projections were applied by Freeman and Ashley (2017) to estimate regional hurricane risk and Strader et al. (2015) to estimate the change in volcano risk in the Pacific Northwest. Mann et al. (2014) evaluated wildfire risk in California by decade 2000–2050 using spatially lagged fixed effect panel regression models to estimate housing density in each split block group and decade, and adjusting the forecasts based on county-level decadal population forecasts.

2.3 Contribution

This paper builds on and adds to this growing body of literature by developing a method that explicitly estimates the number and location of new housing units to subcounty area units over annual increments in the future across a large, multi-county region. This output makes the model directly usable by natural hazard loss estimation models, which are building-based. The new HIP method differs from the relatively few examples with similar output (i.e., those in 2.2.2) in the specifics of the input data, methods, and assumptions used. All previous methods have projected changes over time using decadal population data. By contrast, HIP uses annual housing data directly, avoiding the need for assumptions about headship rates and facilitating the ability to capture nonlinear changes over time. In allocating the housing units within a county, HIP is the first attempt to fit a model to annual, coordinate-based building data (rather than decadal block group data), and to compare multiple possible predictor variables and statistical and machine learning model types to find the best one (Sect. 3.3.1, Online Resource 1 Section S1.1). Davidson and Rivera (2003) assumed the distribution of households within a county remains constant at the average of the two previous decades. Theobald (2005) assumed the number of new housing units in each 100-m grid cell is based on the average growth rate in the previous decade, computed separately for each of 16 classes defined based on density, and accessibility. Mann et al. (2014), the only one that fits a model to data, uses a different set of predictors and type of model, and fits it to decadal data.

3 Housing inventory projection (HIP) method

The goal of this study is to estimate the spatiotemporal changes of single-family housing units over a large area for multiple decades for the purposes of natural hazards risk modeling. To do so, we developed the Housing Inventory Projection (HIP) method, which includes two modules: (1) the Regional Annual County-Level Housing (REACH) module, developed in Williams et al. (2022), and (2) the Single-family Location Estimation (SLE) module, introduced here (Fig. 1). Using past housing trends, demographic trends, and land attributes, the REACH module captures regional trends by estimating the number of housing units per area unit, A_1, over a study area, S_1, per time step, Δt over the projection period, T. These area unit A_1-level estimates are then used as input for the SLE module, which estimates the number of new single-family housing units for each area unit A_2 (where A_2 is smaller subunit of A_1) using a single regression model based on past coordinate-based housing data and a cell capacity check.

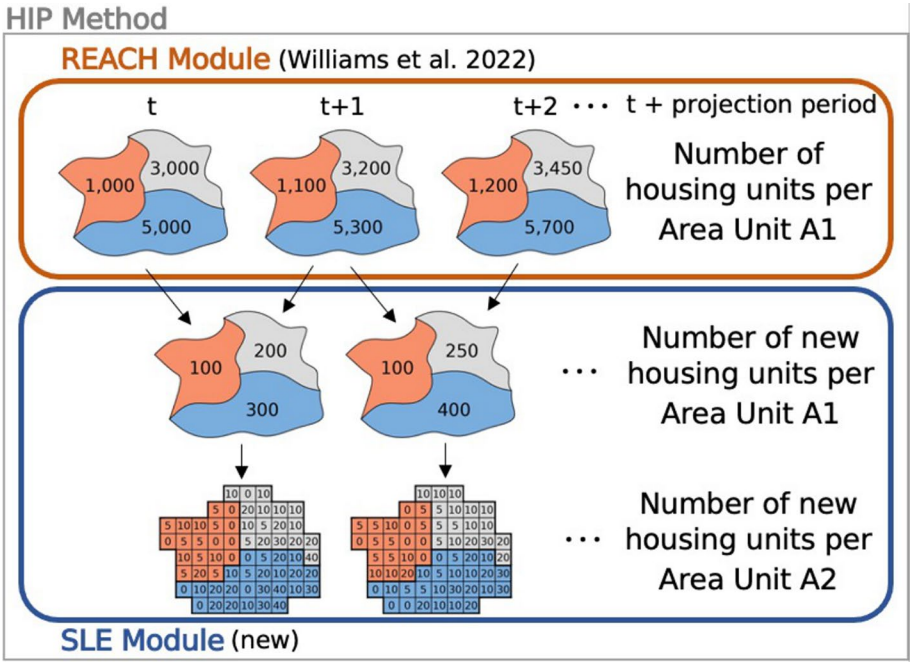

Fig. 1 HIP method overview schematic. The HIP method is comprised of the existing REACH module (Williams et al. 2022) in conjunction with the newly-developed SLE module. The SLE module is run separately for each time step, Δt

3.1 Key parameters

Six key parameters must be defined at the start: (1) projection period, T, (2) time step, Δt, (3) study area for the REACH module, S_1, (4) large area unit, A_1, (5) study area for the SLE module, S_2, which must be the same as or a subset of S_1, and (6) small area unit, A_2. The choices for these parameters are driven by the intended use of the analysis, data availability, and computational and modeling concerns. For this study, the time step is $\Delta t = 1$ year and projection period is $T = 30$ years, which is the typical span of a home mortgage and a reasonable time horizon for policy and planning applications. The 12 southeastern, hurricane-prone coastal states from Texas to Delaware make up S_1, the study area for the REACH module, and North Carolina was selected as S_2, the study area for the SLE module. The large area unit A_1 is a county and the small area unit A_2 is a 1 km^2 square grid cell. The data used to develop the REACH Module is publicly-available county-level data and the county-level A_1 area unit size is a reasonable scale for capturing regional housing dynamics. The 1 km^2 grid cell was chosen as A_2 to capture the neighborhood housing dynamics at sufficiently high resolution. A 1 km^2 gird space is an appropriate resolution for natural hazard risk modeling while also making the analysis computationally tractable. That is, in the analysis presented in this paper, the REACH module is used to develop 30-years projections of the total number of housing units for each of the 1000 counties in study

region S_1 in annual time steps. Focusing on North Carolina, the SLE module is then applied to allocate those annual county-level houses to each of the 130,983 one km^2 grid cells in the state. For ease of description, Sects. 3.2 and 3.3 present the modules assuming these parameter settings. Section 5 describes how to apply the HIP model with other parameter settings.

3.2 Regional annual county-level housing module

The existing REACH module of the HIP method estimates county-level housing demand over a long timeframe by utilizing housing, population, demographic, and land data available at the county level. As described in Williams et al. (2022), the REACH module utilizes a Long Short-Term Memory (LSTM) recurrent neural network model to predict the expected annual number of housing units per county based on 13 publicly-available variables including: population, population density, number of housing units, housing unit density, annual percent change in number of housing units, percentage of vacant housing units, percentage of owner-occupied housing units, average household size, percentage of non-white population, percentage of population with high school degree, percentage of population with college degree, percentage of non-buildable land area, and the distance to the coastline. Specifically, in Williams et al. (2022), the LSTM model was fitted for an $S_1 = 1000$-county region in the southeastern US, for $T = 10$-, 20-, and 30-year time horizons. The analysis in this paper directly used the $T = 30$ projections, resulting in housing projections, h_{ct}, for each county, c (where c equals A_1), in the North Carolina study area, S_2, and year, t.

3.3 Single-family location estimation module

The Single-family Location Estimation (SLE) module is a new method that obtains county-level estimates, h_{ct}, produced in the REACH module and allocates the new single-family housing units into the smaller area units ($A_2 = 1$ km^2 cells) across North Carolina ($S_2 =$ North Carolina). The module includes four steps for each year in the projection period (Fig. 2): (1) run Allocation Model to obtain grid cell weights, (2) assign housing units to grid cells, (3) check capacity of each grid cell, and (4) recalculate supplementary variables. All data used in the SLE module were sourced from publicly-available datasets, except for the housing-related variables which were obtained from the Zillow Transaction and Assessor Dataset (ZTRAX) (Zillow 2021). Data required to apply the method are discussed in Sect. 5.

3.3.1 SLE step 1: run allocation model

First, a newly-developed "Allocation Model" predicts the spatial distribution of new single-family housing units across the grid space for a given year. The Allocation Model predicts the number of new housing units, y_i, in each cell i based on a set of explanatory variables, x_j. The Allocation Model is an ordinary least squares (OLS) regression model expressed as $y_i = \beta_0 + \sum_j \vec{\beta_j} \vec{x_{ij}} + \varepsilon_j$, where β_0 is the intercept, $\vec{\beta_j}$ is the vector of coefficients for each variable, $\vec{x_{ij}}$ is the vector of all explanatory variables for all cells, and ε_j is the error term for each variable. Since the goal of the study is to predict new housing construction

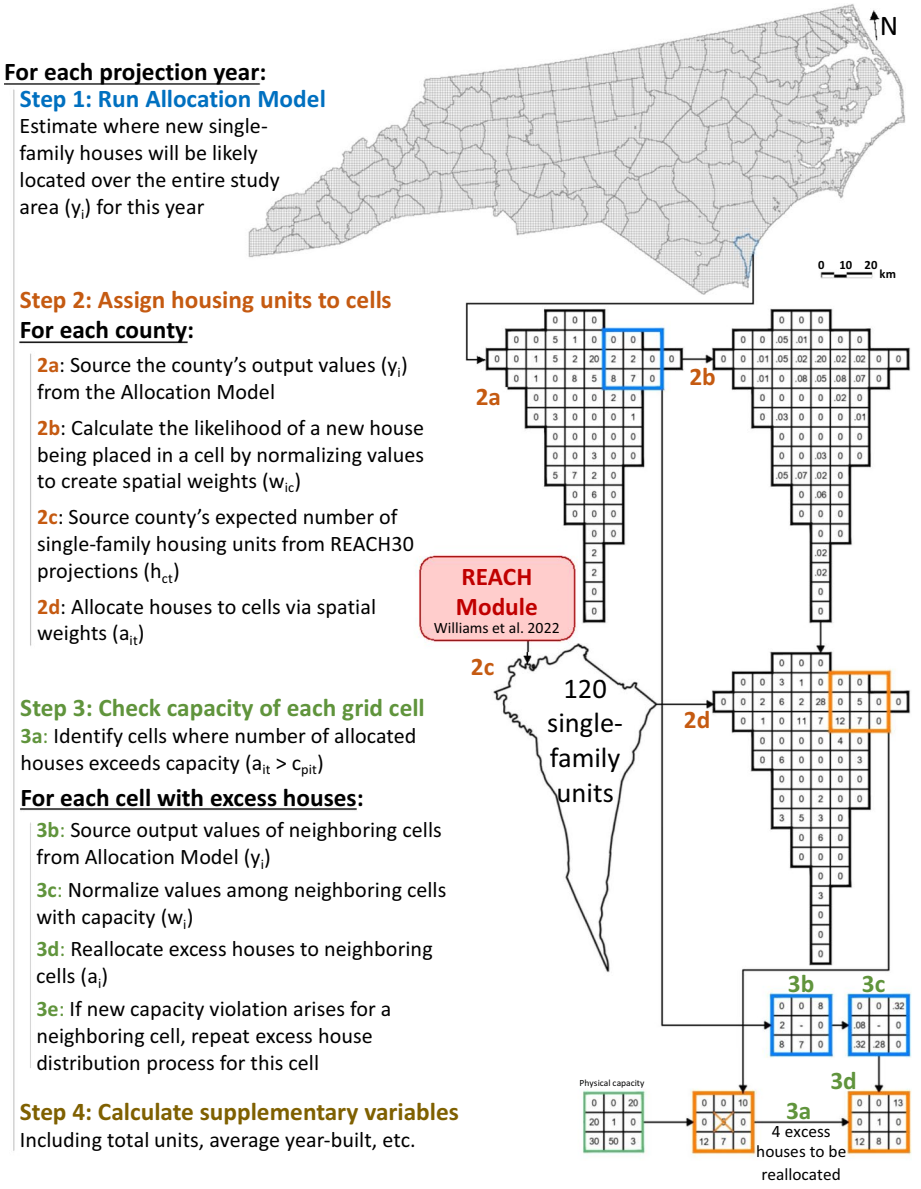

Fig. 2 SLE module overview. The newly-developed SLE module sources the expected number of new houses per county from the existing REACH module (Williams et al. 2022) and distributes the new housing units over the grid space. Note that values presented are for illustration purposes only and displayed cell sizes scale to $3km^2$, not $1km^2$, to better illustrate the SLE module

over a 30-year forecasted period between 2020 and 2049, the Allocation Model was calibrated over 30 prediction years from 1990 to 2019. With 130,983 cells in the study area, and 30 years of response variable data, 3,929,490 observations were used for each model regression.

Table 1 Tested OLS model configurations

Set name	Attribute	Set
S1	Explanatory variables standardized[a]	Yes
		No
V1	Previous years of new single-family housing data	1 year
		1–2 years
		1–3 years
		1–4 years
		1–5 years
V2	Non-distance-related variable sets	None
		Mean year-built
V3	Distance-related variable sets	None
		Distance to urbanized area, distance to US road, distance to coast
V4	Number of layers for new single-family housing data	None
		1 layer
		1–2 layers
		1–3 layers
V5	Number of layers for total single-family housing data	None
		1 layer
		1–2 layers
		1–3 layers
V6	Number of layers for non-distance-related data	None
		1 layer (only if V2 is not None)[b]
G1	Tile group size for training/testing split	1 cell
		3 × 3 cells
		5 × 5 cells
		10 × 10 cells
		20 × 20 cells

[a]"Standardized" means explanatory variables were centered and scaled on zero by subtracting the mean value of the set and dividing by the standard deviation of the set ($z_{ij} = \left(x_{ij} - \mu_j\right)/\sigma_j$

[b]The adjacent average value was not calculated for set V6 if V2 was set to "None" because there would be no values to average over. Therefore, 4800 tests were compared (not 6400 tests).

4800 different modeling configurations were tested to develop the Allocation Model (all combinations of sets in Table 1). In Table 1, "layers" refers to the number of adjacent cells (i.e. number of layers) considered to calculate average values of a given variable and capture proximal effects (see Section S1.1.1 in Online Resource 1 for details). Additional explanatory variables that are common in the LULC and housing change modeling literature, such as the distance to an urban area or distance to the coast, were considered, but were not found to enhance the Allocation Model's performance. A complete description of the data processing and variable selection is available in Online Resource 1 (Section S1.1).

For each test, a cell is assigned to a tile group, which is a square grouping of cells ranging from a size of 1 cell by 1 cell to 20 cells by 20 cells that spans the entire study area. The maximum tile group size of 20 × 20 was chosen because it covers approximately the same area as North Carolina capital city of Raleigh (388 km^2). Tile groups were

assigned to sort a cell either into a training set or testing set. For each tile group size, tiles were randomly assigned to either training or testing sets using an 80/20 split. The tile group training and testing split method allows an investigation of model performance across different subregion (tiles) sizes. The tile split method also allows a cross-validation comparison across sets S1 and V1-V6 for each training/testing division of sample in set G1 to understand whether training or testing errors are affected by the spatial location of cells.

The best-performing OLS modeling configuration was assessed using four error metrics. To ensure that the total number of predicted houses over the entire study area in a given year aligned with actual observations, the first error metric, ε_1, calculates the absolute value of the summed residuals of the total number of new single-family housing units in the study area for each year in the testing set, corresponding to Eq. (1), where y_{it} and $\hat{y}_{it}$ are the observed and predicted number of single-family houses per cell, i, in year, t, respectively, for all cells, I, and all years, T.

$$\varepsilon_1 = \left| \sum_t \left(\sum_i \hat{y}_{it} - \sum_i y_{it} \right) \right| \tag{1}$$

To ensure good predictions of new single-family housing units per cell across the entire study area, the second error metric, ε_2, calculates the mean of the absolute value of the residuals for each observation in the study area in the testing set (Eq. 2).

$$\varepsilon_2 = \frac{\sum_t \sum_i \left| \hat{y}_{it} - y_{it} \right|}{I * T} \tag{2}$$

Lastly, the Allocation Model will ultimately be applied to a natural hazards loss model to estimate future expected losses caused by new single-family housing development. As most new single-family housing development is built in urban and suburban areas, it is important to evaluate the accuracy of the predicted values in urban and suburban areas to ensure the natural hazard loss model results are realistic. Therefore, the third and fourth error metrics (ε_3 and ε_4) used for model selection calculates the mean of the absolute value of the residuals for all cells in urban or suburban areas in the testing set, I_u, (Eq. 3), as well as the mean of the absolute value of the percent residual difference for cells in urban or suburban areas in the testing set (Eq. 4). Cells located in urban or suburban areas are located within a 10 km radius of the Census-designated urban area boundary, which is a reasonable boundary based on the number of total and new single-family housing units between 2012 and 2019, as depicted

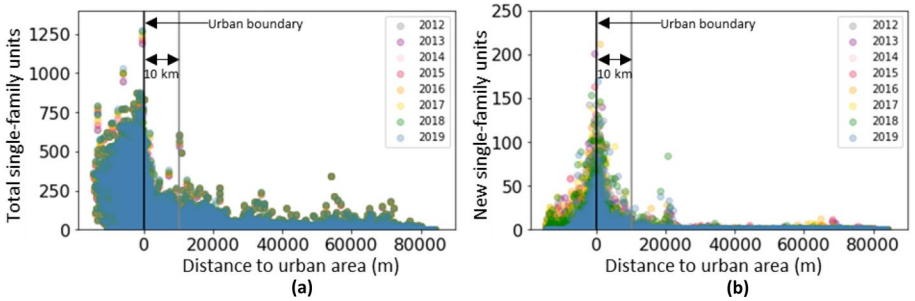

Fig. 3 **a** Number of total single-family housing units versus distance to urban area **b** Number of new single-family housing units versus distance to urban area

in Fig. 3. Graphical comparisons of the error metrics among the 4800 tests are available in Online Resource Section S1.1.3.

$$\varepsilon_3 = \frac{\sum_t \sum_{i_u} \left| \hat{y}_{i_u t} - y_{i_u t} \right|}{I_u * T} \tag{3}$$

$$\varepsilon_4 = \frac{\sum_t \sum_{i_u} \left| 100 * \frac{(\hat{y}_{i_u t} - y_{i_u t})}{y_{i_u t}} \right|}{I_u * T} \tag{4}$$

The range of values among the four error metrics are orders of magnitude different and cannot be compared as-is with equal weighting. Therefore, each of the four error metrics were each scaled to values between 1 and 100 and then summed to calculate the selection metric, E, for each test (Eq. 5). The variable combination with the minimum average selection metric was determined to be the "best" variable combination. Of the five tests with the best variable combination, the tile size with the lowest selection metric was chosen as the "best" training/testing split tile group size. The best combination of variables, x_j, is available in Table 2.

$$E = \sum_n^4 \left(\left(\frac{\varepsilon_n - \varepsilon_{n,\min}}{\varepsilon_{n,\max} - \varepsilon_{n,\min}} \right) * (100 - 1) \right) + 1 \tag{5}$$

More complex modeling methods were also compared with the OLS method to understand whether increased modeling complexity leads to improved model performance for the problem at-hand. The best variable combination and tile group size found from the series of OLS tests performed were used in a robust linear (RLM) model, gradient-boosted regression (GBR) tree model, and artificial neural network (NN) model and values for the four error metrics were compared. Robust linear models can be used as an alternative method for regression modeling where there are large outliers in the response variable (as in heavily-urbanized cells). The machine learning GBR and NN methods are common modeling methods that are often accompanied with complex datasets and can lead to

Table 2 Allocation model explanatory variables

j	Attribute
1	New single-family housing units per cell in time t – 1
2	New single-family housing units per cell in time t – 2
3	Total single-family housing units per cell in time t – 1
4	Average year-built value of single-family housing units per cell in time t – 1
5[a]	Average number of new single-family housing units within 1 km of a given cell in time t – 1
6[a]	Average number of new single-family housing units within 1 km of a given cell in time t – 2
7[a]	Average number of total single-family housing units within 1 km of a given cell in time t – 1
8[a]	Average year-built value of single-family housing units within 1 km of a given cell in time t – 1

[a]See Online Resource 1 (Section S1.1.1) for a detailed description for the calculation method of average adjacent values

Table 3 Errors for each compared model

Model	ε_1	ε_2	ε_3	ε_4
OLS	2175	0.3569	0.3500	78.12
RLM	150,289	0.3159	0.3256	83.06
GBR	4880	0.3613	0.3556	79.52
NN	63,998	0.3131	0.3078	88.92

improved model performance as compared to a simple OLS method. Online Resource 1 (Section S1.1.3) provides a list of the hyperparameters and modeling packages used.

A comparison of the four error metrics for the OLS, RLM, GBR, and NN methods is available in Table 3. While the NN model performed best for metrics ε_2 and ε_3, summed predicted values per year across the study area are much different than the observed values, as indicated by the high ε_1 score. This is likely due to underprediction across outlier values. The RLM model, which discounts outlier values, performed similarly to the NN model. The GBR model performed similarly as the OLS model, although had slightly worse results. Overall, for the task of predicting the location of new single-family housing units over a 30-year time period in the future, a simple and easy to understand modeling method is preferred over a complex and black-box machine learning algorithm when possible Researchers such as Wagenaar et al. (2020) have argued that machine learning methods are not always appropriate for predicting local urban change and simpler modeling methods should be utilized when possible. Therefore, the OLS model was chosen as the "Allocation Model" henceforth.

The Allocation Model reasonably predicts the location of new single-family housing units across a 1 km^2 grid space, where 93.7% of cells across all calibration years had predicted values within one house of the observed value and 97.2% were within two houses of the observed value. The residuals of the Allocation Model, $r_i = \hat{y}_i - y_i$, are centered on

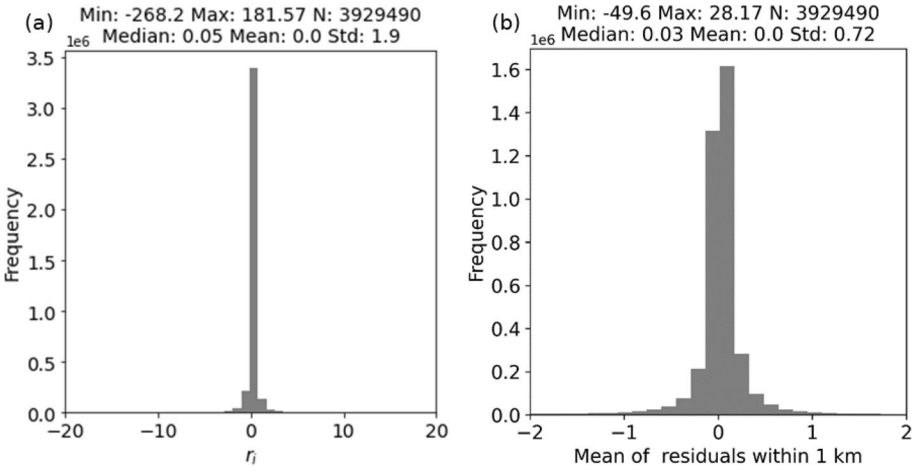

Fig. 4 Histograms of Allocation Model residuals for all cells in all calibration years: **a** residuals r_i, and **b** average residuals among cells within a 1 km radius of cell i (a 3 km by 3 km area), for all cells. Note that for clarity, the x-axes are truncated; 0.16% and 1.66% of observations are outside the x-axis bounds in **a** and **b**, respectively

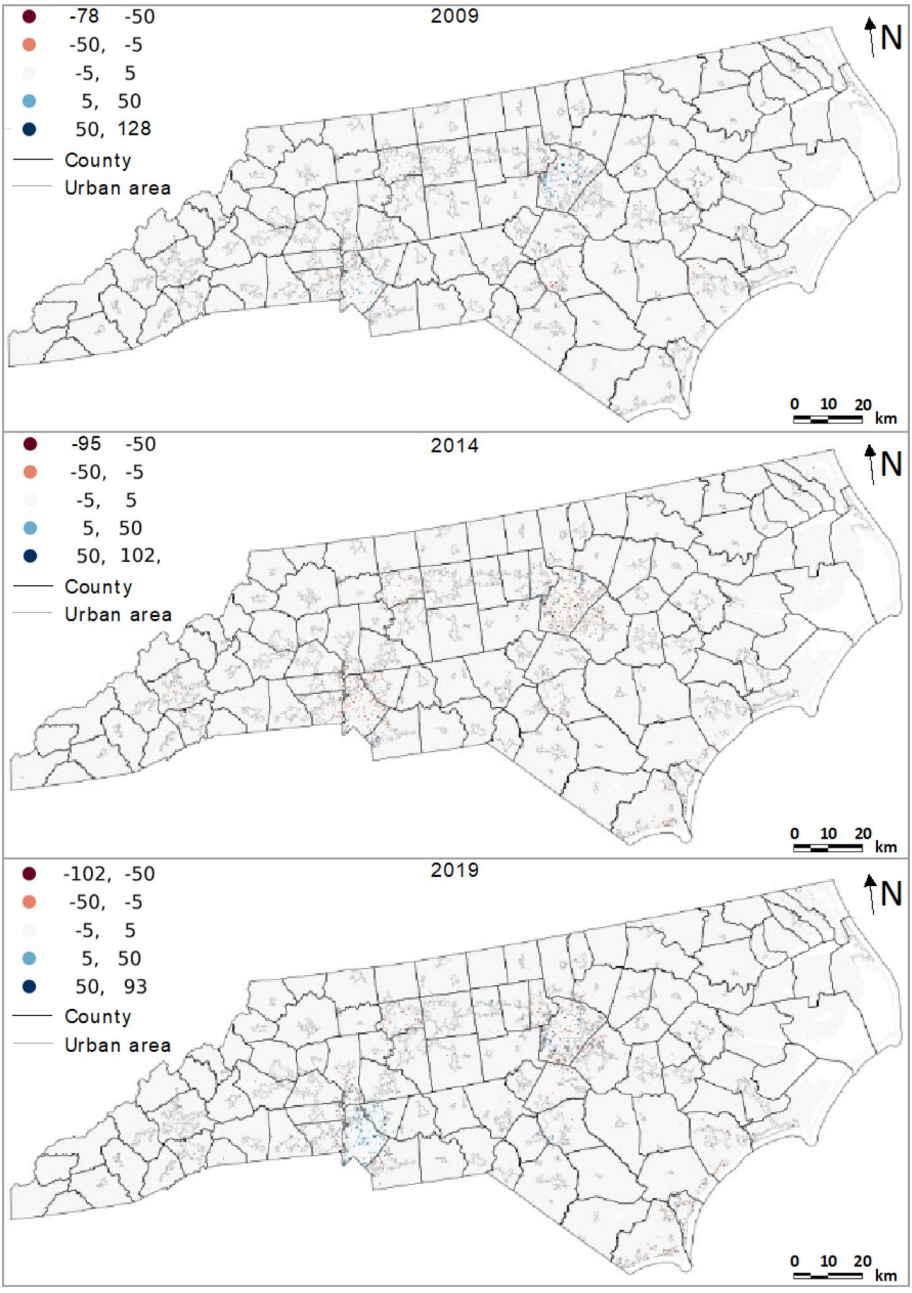

Fig. 5 Spatial distribution of Allocation Model residual values for 2009, 2014, and 2019 prediction years across the study area. Blue cells represent overprediction while red cells represent under prediction

zero (Fig. 4a) and model performance is similar across the tested years (Fig. 5), where the errors are largest near urban boundaries. Additionally, Fig. 6a compares the predicted versus observed spatial distributions of weights w_{ic}, for all cells i belonging to county c,

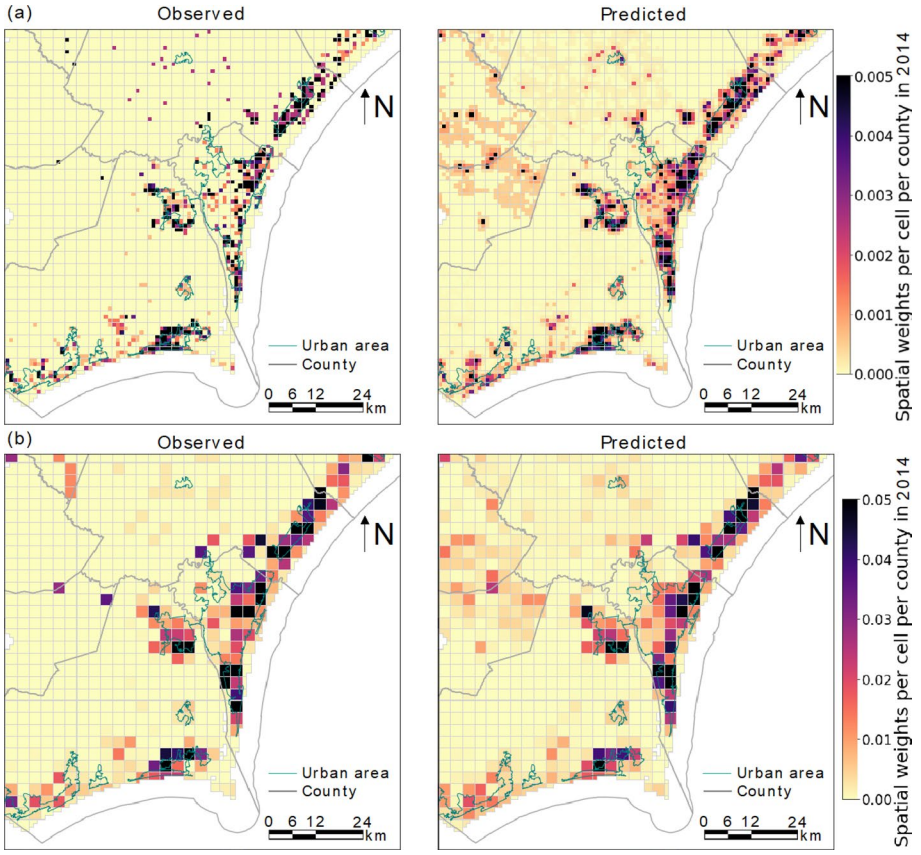

Fig. 6 Observed (w_{ic}) versus predicted ($\hat{w}_{ic}$) spatial distribution of weights in 2014 near Wilmington, NC using the Allocation Model for **a** 1 km^2 grid cells and **b** 9 km.^{2} grid cells. Similar figures for other localized regions in the study area are provided in Section S1.1.4 of Online Resource 1

($w_{ic} = y_{ic} / \sum_c y_i$) for the New Hanover, NC area in 2014, highlighting aligned spatial trends between the observed and predicted values. Additional comparisons of the observed versus predicted model performance, including the weights for six other metropolitan areas over multiple years, are available in Online Resource 1 (Section S1.1.4).

Grid-cell-based errors (Fig. 4a) do not capture the spatial relationships of the grid cells. To further examine the spatial error pattern, Fig. 4b provides residuals averaged with their neighbors. Specifically, in Fig. 4b, each observation represents, for cell i, the average of the residuals of that cell and its 8 neighbors. In Fig. 6b, predicted and observed values in each 3 by 3 km tile are summed and the associated county-level weights are calculated. Figures 4b and 6b suggest that errors in grid cell weights are often corrected when considering neighboring cells. Overall, Step 1 of the SLE module approximates the expected spatial distribution of new single-family housing units across the entire study area for a given year based on the spatial distribution of housing in the previous two years.

3.3.2 SLE step 2: assign housing units to cells

In Step 2 of the SLE module, the housing projections, h_{ct}, from the REACH module are distributed to 1 km^2 grid cells each year using outputs from the Allocation Model. First in Step 2, output values from the Allocation Model are normalized so the values in all cells in a given county sum to one. This weight value, w_{ic}, represents the expected probability of a new single-family housing unit being placed in a given cell in the county. The location of the new housing units expected in each county is then simulated by assigning houses to a given cell using a spatially-weighted random distribution process. Rather than multiplying h_{ct} by w_{ic} to calculate the allocated number of houses per cell per year, a_{it}, as a fractional value, we chose to allocate whole houses to cells through a spatially-weighted random distribution process which simulates naturally-occurring random effects of housing development. For example, the simulation method can estimate development in rural areas where the probability of new development is low, yet development still occurs on occasion in sparse spatial patterns that is difficult to predict precisely. This process repeats for each county until all projected housing units are assigned to a grid cell.

3.3.3 SLE step 3: check capacity of each grid cell

The third step of the SLE module checks whether the number of newly-allocated housing units per cell exceeds the cell's physical capacity. The physical capacity, c_{pit}, identifies the theoretical maximum number of new single-family housing units that could fit within a cell for each year, given the land conditions in 2019. The physical capacity is first determined by calculating the area within each cell that is covered by non-buildable area (NC Dept. of Information Technology 2020; USGS 2020) or is already developed (Dewitz 2021). The remaining area is theoretically developable and physically available for new housing units. The developable area is then divided by the estimated land area required for a new single-family housing unit to determine the physical capacity of the cell. The estimated land area is derived from the median lot size for single-family housing units in North Carolina between 2010 and 2019 (14,375 ft^2), as well as the expected landscape and sidewalk buffer (10 feet) and road width (11 feet) (NCDOT 2022) adjacent to the new house, totaling to a land area of 16,893 ft^2 for each new house.

The cell capacities are then compared with the allocated new housing units derived from Step 2 to identify cells with excess houses. For each cell, where the allocated houses exceed the capacity, excess cells are reallocated to neighboring cells using the following procedure. First, cells that have capacity within one neighboring layer of the cell in question are identified and the associated output values from the Allocation Model are sourced. If all neighboring cells do not have positive outputs from the Allocation Model, then Allocation Model outputs for an additional layer are obtained. These values are then normalized to sum to one to create spatial weights among the set of neighboring cells with capacity. Similar to the allocation process in Step 2, excess houses for a given cell are then randomly assigned among the neighboring cells using the spatial weighting, followed by a capacity check calculation. The reallocation process restarts until there are no capacity violations and all excess houses are allocated to cells. The output of Step 3 provides the final number of new single-family housing units estimated to be located in each cell in the study area in the given projection year.

3.3.4 SLE step 4: calculate supplementary variables

Finally, once all new housing units have been assigned to a cell for a given year, all explanatory variables needed for the Allocation Model are calculated so they can be applied in the next year (Table 2). This includes the total number of housing units per cell in the year, the mean year-built, as well as associated adjacency averages. Physical capacities are also recalculated to reflect the new spatial distribution of housing availability.

Steps 1 through 4 repeat each year until all projected single-family housing units are assigned to a grid cell for each projection year.

3.4 Assumptions and limitations

Like all forecasting methods, the HIP method includes some important assumptions and limitations. First, the 2019 proportion of housing units that are single-family in each county is assumed to remain constant across the entire projection period. This is because the REACH30 model predicts the total number of housing units per county per year, but does not distinguish housing units by construction type (e.g., single-family, multi-family, manufactured house). However, we believe this is a reasonable assumption because the mean and standard deviation of the proportion of single-family houses to total housing units per county between 2009 and 2019 in North Carolina remained essentially constant (mean: [0.673, 0.676]; standard deviation: [0.066, 0.068]). Second, we assume the relationship between explanatory variables, x_{ij}, and response variable, y_i, in the OLS Allocation Model, based on 1990–2019 data, remains constant for all future years (i.e., coefficients, β_j, do not change over time). Additionally, demolition is not considered in this analysis due to a primary interest in net new construction and the lack of data availability for the entire state over the 30-year calibration period. Furthermore, all new single-family houses are assumed to occupy the same amount of land area (based on the median single-family lot size between 2010 and 2019, as well as adjacent road requirements), however there is both spatial and temporal variation in the land required for single-family construction. Lot sizes tend to be smaller in urban areas than rural, and have been trending downward (NAHB 2022). Structural and property attributes of new housing units were also not inferred for new housing units but could be applied in the future using an approach similar to Davidson and Rivera (2003) and Jain and Davidson (2007a; b). Lastly, uncertainty in the housing estimates is not explicitly captured in the HIP method and could be a fruitful area for future work, however, the housing estimates do align with neighborhood housing development patterns, as exhibited by Figs. 4b and 6b.

4 Housing inventory projection (HIP) implementation

4.1 Estimated spatial distribution of future single-family housing units

Applying the HIP method, we produced annual estimates of new expected single-family housing units for each cell in the study area from 2020 to 2049. Figure 7 provides a distribution of the number of new single-family housing units per cell over the 30-year projected period. Over this projection period, no new single-family housing construction is expected in 33.97% of cells and 46.6% of the cells are expected to have one to twenty

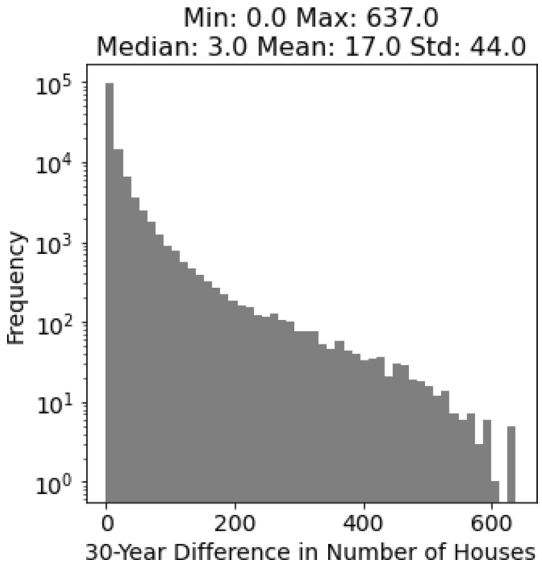

Fig. 7 Distribution of the expected number of new single-family housing units per cell from 2019 to 2049 using the HIP method. Note that the y-axis is a $\log_{10}$ scale

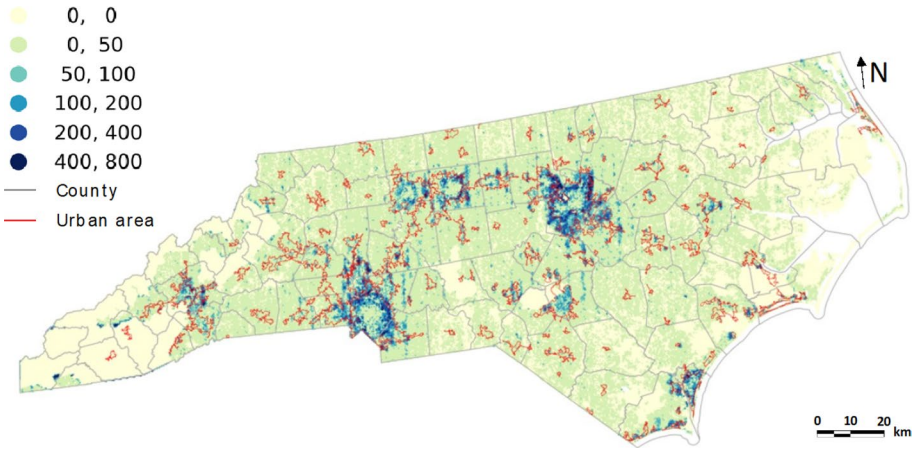

Fig. 8 Projected number of new single-family housing units across the study area from 2019 to 2049 using the HIP method

new single-family housing units. The maximum number of new housing units over the 30-year period is 637 units, which equates to a new housing subdivision occupying the entire cell area. Over the entire study area, the largest changes in new housing development are expected to occur on the peripheries of the urban areas. (Figs. 8 and 9). Figure 8 provides the expected number of new houses per cell over the 30-year projection period, while Fig. 9 displays the expected percent change in new housing units per cell over the 30-year period. Figure 10 shows an example of a magnified comparison of the spatial distribution of total single-family housing units in 2019 versus 2049 in the Wilmington, NC area, and additional zoom-in comparison examples are available in Section S2 of Online Resource 1. Single-family housing construction is expected to sprawl outwards from urban

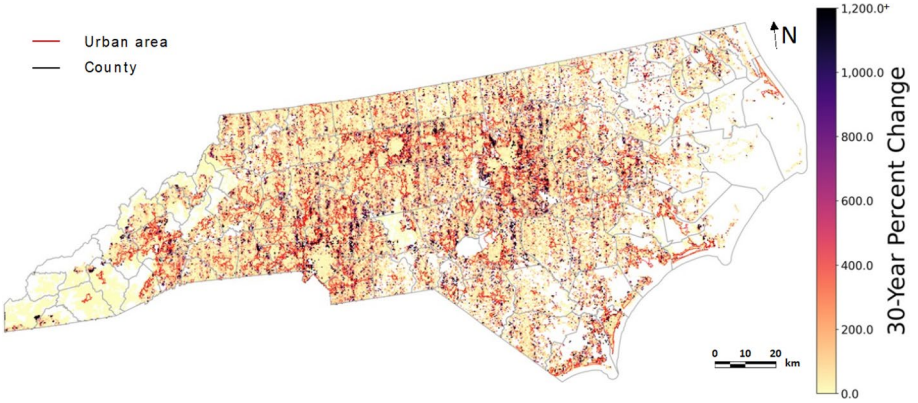

Fig. 9 Projected percent change in the total number of new single-family housing units across the study area from 2019 to 2049 using the HIP method. Note that cells with zero houses in 2019 have infinite or indeterminant percent change value and are colored white in the figure

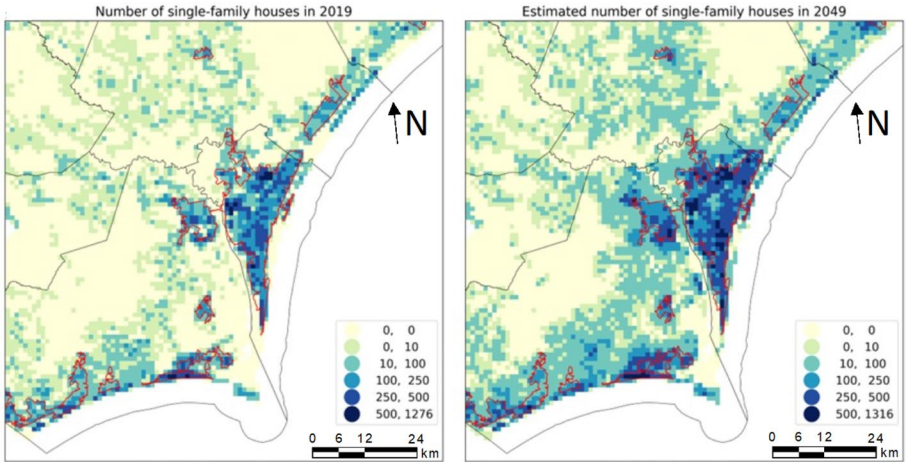

Fig. 10 Spatial distribution of the number of single-family housing units in 2019 versus the expected number of single-family housing units in 2049 in the Wilmington, NC area using the HIP method. Each cell depicted is 1 km^2 for scale

boundaries, assuming future housing construction patterns follow past behavior. Tabulations of the expected new and total single-family housing units per cell per year are available in a publicly-accessible dataset (Williams and Davidson 2023).

4.2 Implications for hurricane impacts

Modeling the expected spatial and temporal changes in a region's housing inventory is not only helpful for understanding where housing development pressures are expected in the future, but also to estimate how a region's natural hazards risk is expected to change over time. North Carolina is a state with significant hurricane risk; 12 tropical

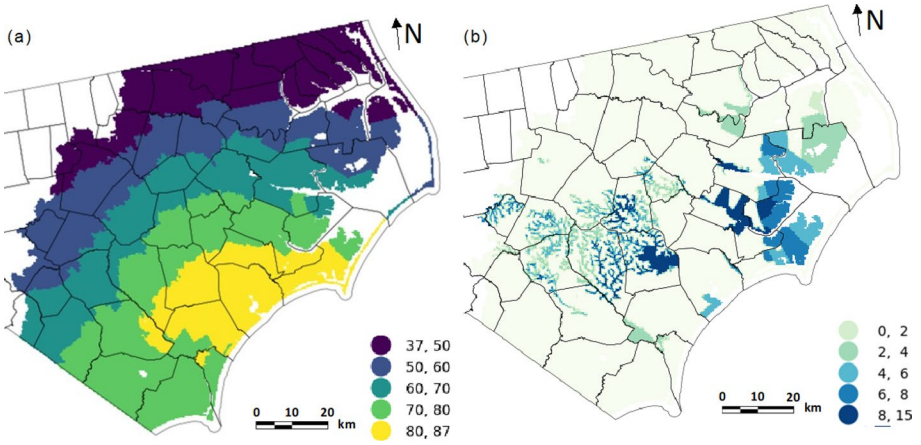

Fig. 11 Hurricane Florence **a** 1 min sustained wind speed (mph), **b** maximum flood depth (ft) per cell, per cell in the eastern half of North Carolina (Yang et al. 2022)

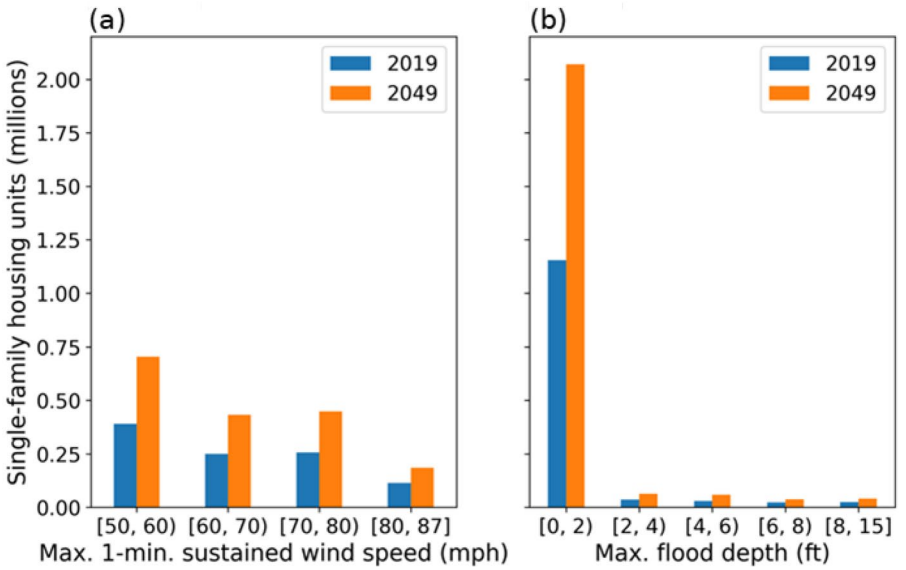

Fig. 12 Projected total number of single-family housing units in 2019 versus 2049 per **a** wind speed (mph), **b** flood depth (ft)

cyclones have made landfall in North Carolina since 2003 (NOAA 2022). On September 14, 2018, Hurricane Florence made landfall in Wrightsville Beach, NC as a slow-moving Category 1 hurricane and produced record-breaking rainfall over the following two days (NOAA 2018). It was the costliest hurricane in North Carolina's history, with total losses of US$28.3 billion (2022 Consumer Price Index adjusted cost), including both

insured and uninsured losses for agriculture, individual insurance claims payouts, and disaster relief money from the federal government (NOAA 2023). The approximately 1.2 million single-family homes in the eastern half of North Carolina in 2018 experienced 1-min sustained wind speeds as high as 87 mph and flood depths as high as 15 feet. Figure 11 depicts the spatial distribution of maximum wind speeds and maximum flood depths in the eastern half of North Carolina (Yang et al. 2022).

Using the HIP method to project the location of new single-family housing units in 2049, and assuming the same 2018 distributed hazard and vulnerability profiles of newly built homes, a hurricane similar to Hurricane Florence hitting North Carolina in 2049 would affect an additional 1 million single-family homes, an 80.6% increase from the number of single-family housing units in the region in 2018. If future housing construction patterns follow past behavior, the number of new housing units will increase across all hurricane wind speeds and flood depths under a Hurricane Florence-type scenario (Fig. 12). Lastly, given the significant amount of rainfall caused by Hurricane Florence, it is important to evaluate whether new housing units are expected to be built in areas that experienced high (> 2 ft.) of flooding following Hurricane Florence. Focusing on new houses in areas that experienced > 2 ft flooding in Hurricane Florence, Fig. 13 suggests that there is a higher expected rate of new construction in areas that experienced two to six feet of flooding, as compared to the expected new construction rate for areas that experienced greater than six feet of flooding. In the areas that experienced two to six feet of flooding following Hurricane Florence, we can expect approximately 1,900 new houses constructed each year between 2020 and 2049 (57,092 total). Crucially, the expected percent change in new construction in areas that experienced two to six feet of flooding during Hurricane Florence is higher than the construction rate over the entire study area between 2019 and 2049 (87.5% and 79.2%, respectively). The area that experienced greater than six feet of flooding can expect approximately 1,000 new houses built each year between 2020 and 2049 (29,699 total, 62.2% change). In these scenarios, nearly 2,900 additional households are in harm's way each year.

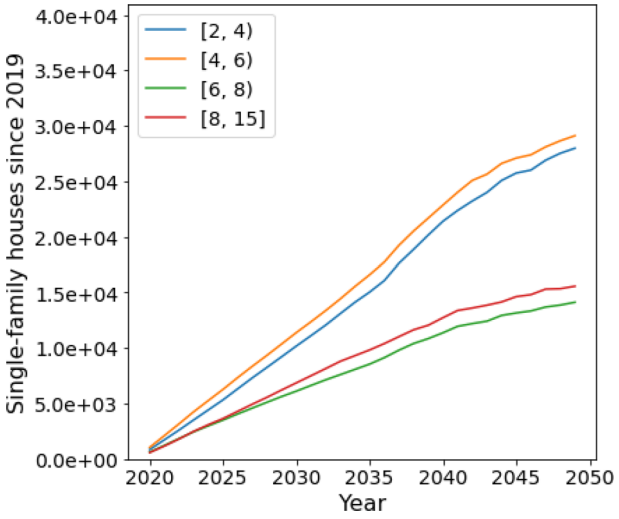

Fig. 13 Projected number of new single-family housing units per year per flood depth (ft) since 2019

5 Applicability of housing inventory projection method

In Sects. 3 and 4, the HIP method was applied to produce a 30-year projection of annual single-family housing values in 1 km^2 grid cells in North Carolina. The method is intended to be usable by others, however, with other key parameter values. Overall, county-level area units are recommended for A_1 in the REACH module because housing demand projections rely on long-term sociodemographic data, among other factors, which are available at the county-level within the US. A_2 area units should be equally-sized grid cells, as opposed to irregular-shaped area units such as block groups, to minimize spatial bias in the Allocation Model calibration (i.e., a skewed housing distribution based on differing block group sizes). The following provides a description on how the HIP method can be applied for alternative uses.

To apply the HIP method in an area within the $S_1 = 1000$-county southeastern US region for a period of up to 30 years, one could directly use the REACH module's h_{ct} results, which are publicly-available in Williams and Davidson (2022) and supported by the methods presented in Williams et al. (2022). A user could then apply the SLE to obtain the final results. Outside the 1,000-county southeast region, one would have to obtain data for the 13 variables used in the REACH module, which are publicly available within the US, then follow the data processing, model testing, and model selection methods explained in Williams et al. (2022) using the python notebooks in Williams and Davidson (2022) to obtain the annual county-level housing estimates, h_{ct}, over a projection period up to 30 years. If one needs annual county-level housing estimates for a projection period longer than 30 years, they could reference the methods used in Williams et al. (2022), but may be limited by data availability for model calibration. Alternatively, since HIP is modular, one could obtain county-level housing estimates for each projection year from another source instead of the REACH module, and then apply the SLE module. Potential alternative sources of county-level estimates could include county-level population projections provided by Hauer (2019) or Striessnig et al. (2019) and supplemented by county-level headship data.

Similar to the available variations of the REACH module application, one could also alter the parameters used in the SLE module. To predict the location of new single-family housing units over a different A_2 area unit size, one would need to recalibrate the Allocation Model and obtain past single-family housing data and supplementary variable data (e.g., distance to coast) at the selected time step, Δt, and area unit, A_2, over a grid space spanning the selected study area S_2. The user should obtain historical data at least as long as the desired projection period for model calibration. To obtain the historical data at the desired A_2 size, one should obtain housing coordinates and year-built data, which are available in privately-sourced datasets such as ZTRAX (Zillow 2021). The National Structures Inventory (NSI) provides coordinates of structures (including houses) in their publicly-available dataset, but year-built data is only available in the private NSI dataset (USACE 2022), as of the time of this publication. Non-housing related variables used in the SLE module, such as the non-buildable area data, are publicly-available and can be configured to the selected A_2 size. The python notebooks and supporting data for this project are supplied in Williams and Davidson (2023). The user manual provided in Online Resource 1 (Sect. S2), provide instructions for data processing, data review, model calibration, and housing projections.

6 Conclusions

Our study is motivated by the need to estimate regional changes in expected losses due to building damages caused by natural hazard impacts over space and time. Therefore, we focus on the granular spatiotemporal changes of building exposure, specifically housing exposure, to natural hazards, rather than other types of exposure such as population or land use. This innovative work develops a method for estimating the spatial distribution of single-family housing units with high spatial and temporal resolution (i.e., over a 1 km^2 grid space each year) for a large spatial and temporal scope (i.e., over a multi-decadal time horizon across an entire state). The presented HIP method combines direct county-level annual housing projections via the REACH30 developed in Williams et al. (2022), with the newly-developed SLE module which uses housing location, year-built, and lot size data from ZTRAX (Zillow 2021), in addition to publicly-available land cover and protected area data (NC Dept. of Information Technology 2020; USGS 2020; Dewitz 2021), to locate new houses within grid cells.

The HIP case study presented also provides a dataset that has been made available for researchers interested in evaluating the expected spatiotemporal changes of single-family housing units in North Carolina between 2020 and 2049. Such data has the potential to support government agencies, insurers, mortgage lenders, economists, researchers, and practitioners, in understanding the expected changes in North Carolina's housing portfolio and associated long-term effects. Additionally, the data processing, modeling, and analysis used in this study are publicly-available through a set of python notebooks that are supported by a detailed user manual (Sect. S2 of Online Resource 1) and data set (Williams and Davidson 2023). The creation of the HIP method with the associated python notebooks and dataset supports the goals identified within the natural hazards community of creating and sharing useable housing datasets and methods that can accompany flexible computational workflows for regional natural hazards modeling (Zsarnóczay et al. 2023).

Finally, the presented work can be expanded in several ways to advance future natural hazards research. For example, the presented HIP method does not infer structural or property attributes of new single-family housing units. With this, future research can quantify expected hurricane damages and losses due to new housing construction by inferring structural and property characteristics to the new housing units and estimate component-level damages over the projection period. Similarly, a range of spatially-explicit building code scenarios could be tested to evaluate expected regional hurricane losses when enhanced building codes are applied to new construction. Previous researchers have also identified the need for improved policies that direct new construction in safe areas, specifying that development in areas with high risk of natural hazard will continue if not restricted (Strader et al. 2018; Wing et al. 2022; Rifat and Liu 2022). While the presented HIP method does not yet incorporate a method for comparing different land use policy scenario, another extension could allow variation in lot sizes or a change cell capacities to create a variety of pseudo-zoning codes that reduce housing construction in high-risk areas to understand the associated changes in expected losses. Additionally, as the SLE module of the HIP method only predicts the location of new single-family housing units, the presented method can be expanded to predict locations of all housing types (e.g., multi-family, manufactured house) and/or building types. Finally, as the HIP method was developed for the case study of North Carolina using data prior to 2020, future work could apply the HIP method to other regions and/or to post-COVID-19 pandemic conditions to compare trends in regional housing development.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s11069-023-06132-5.

Acknowledgements This material is based on work supported by the National Science Foundation (award no. 1830511). The statements, findings, and conclusions are those of the authors and do not necessarily reflect the views of the National Science Foundation. Data is provided by Zillow through the Zillow Transaction and Assessment Dataset (ZTRAX). More information on accessing the data can be found at http://www.zillow.com/ztrax. The results and opinions are those of the author(s) and do not reflect the position of Zillow Group.

Author contributions All authors contributed to the conception and design of the presented work. CW was involved in the data curation, formal analysis, software implementation, and preparation of the visualizations. Rachel Davidson supervised the project at large. CW and RD were involved in the investigation, methodology, validation, and the original preparation and writing of the draft. LN, JT, MM, and JK were involved in manuscript editing. RD, LN, JT, MM, and JK were involved in conceptualization, funding acquisition, project administration, and the supplying of resources for the project. All authors assisted in the final review and editing of the text.

Funding This material is based on work supported by the National Science Foundation (award no. 1830511). The statements, findings, and conclusions are those of the authors and do not necessarily reflect the views of the National Science Foundation.

Data availability Housing data used in this study is provided by Zillow through the Zillow Transaction and Assessment Dataset (ZTRAX). More information on accessing the data can be found at http://www.zillow.com/ztrax. The results and opinions are those of the authors and do not reflect the position of Zillow. Aggregated data consumed and all data produced is available on the DesignSafe Data Depot (project no. PRJ-4083), and supporting documentation is available in the Online Resource 1 (Williams and Davidson 2023).

Code availability All code is available on the DesignSafe Data Depot (project no. PRJ-4083), and supporting documentation is available in Online Resource 1 (Williams and Davidson 2023).

Material availability Supplemental information is provided in Online Resource 1. Section S1 of Online Resource 1 provides supporting information for the SLE module. Section S2 of Online Resource 1 provides technical documentation and description of the data and code available.

Declarations

Conflict of interest The authors have no relevant financial or non-financial interests to disclose.

References

Aburas MM, Ahamad MSS, Omar NQ (2019) Spatio-temporal simulation and prediction of land-use change using conventional and machine learning models: a review. Environ Monit Assess 191:205. https://doi.org/10.1007/s10661-019-7330-6

Ali GG, El-Adaway IH, Dagli C (2020) A system dynamics approach for study of population growth and the residential housing market in the US. Proc Comput Sci 168:154–160. https://doi.org/10.1016/j.procs.2020.02.281

Bozzolan E, Holcombe EA, Pianosi F et al (2023) A mechanistic approach to include climate change and unplanned urban sprawl in landslide susceptibility maps. Sci Total Environ 858:159412. https://doi.org/10.1016/j.scitotenv.2022.159412

Briassoulis H (2019) Analysis of land use change: theoretical and modeling approaches

Bryant BP, Westerling AL (2014) Scenarios for future wildfire risk in California: links between changing demography, land use, climate, and wildfire. Environmetrics 25:454–471. https://doi.org/10.1002/env.2280

Center for Risk-Based Community Resilience Planning (2021) IN-CORE Manual

Cho S-H, English BC, Roberts RK (2005) Spatial analysis of housing growth. Rev Reg Stud 35:311–335

Cremen G, Galasso C, McCloskey J (2022) Modelling and quantifying tomorrow's risks from natural hazards. Sci Total Environ 817:152552. https://doi.org/10.1016/j.scitotenv.2021.152552

Daniel CJ, Frid L, Sleeter BM, Fortin M-J (2016) State-and-transition simulation models: a framework for forecasting landscape change. Methods Ecol Evol 7:1413–1423. https://doi.org/10.1111/2041-210X.12597

Davidson RA, Rivera MC (2003) Projecting building inventory changes and the effect on hurricane risk. J Urban Plan Dev 129:211–230. https://doi.org/10.1061/(ASCE)0733-9488(2003)129:4(211)
Dewitz J (2021) National land cover database (NLCD) U.S. Geological Survey data release
Federal Emergency Management Agency (2021) Hazus–MH 4.2: Hurricane Model Technical Manual
FEMA (2022) OpenFEMA Dataset: FIMA NFIP Redacted Claims - v1. In: Federal Emergency Management Agency. https://www.fema.gov/openfema-data-page/fima-nfip-redacted-claims-v1. Accessed 25 Jan 2023
Ferguson AP, Ashley WS (2017) Spatiotemporal analysis of residential flood exposure in the Atlanta, Georgia metropolitan area. Nat Hazards 87:989–1016. https://doi.org/10.1007/s11069-017-2806-6
Filatova T (2015) Empirical agent-based land market: Integrating adaptive economic behavior in urban land-use models. Comput Environ Urban Syst 54:397–413. https://doi.org/10.1016/j.compenvurbsys.2014.06.007
Ford A, Barr S, Dawson R et al (2019) A multi-scale urban integrated assessment framework for climate change studies: a flooding application. Comput Environ Urban Syst 75:229–243. https://doi.org/10.1016/j.compenvurbsys.2019.02.005
Freeman AC, Ashley WS (2017) Changes in the US hurricane disaster landscape: the relationship between risk and exposure. Nat Hazards 88:659–682. https://doi.org/10.1007/s11069-017-2885-4
Gao J, O'Neill BC (2019) Data-driven spatial modeling of global long-term urban land development: The SELECT model. Environ Model Softw 119:458–471. https://doi.org/10.1016/j.envsoft.2019.06.015
Hauer ME (2019) Population projections for U.S. counties by age, sex, and race controlled to shared socioeconomic pathway. Sci Data 6:1–15. https://doi.org/10.1038/sdata.2019.5
Hauer ME, Evans JM, Mishra DR (2016) Millions projected to be at risk from sea-level rise in the continental United States. Nat Clim Change 6:691–695. https://doi.org/10.1038/nclimate2961
Jain VK, Davidson RA (2007a) Forecasting changes in the hurricane wind vulnerability of a regional inventory of wood-frame houses. J Infrastruct Syst 13:31–42. https://doi.org/10.1061/(ASCE)1076-0342(2007)13:1(31)
Jain VK, Davidson RA (2007b) Application of a regional hurricane wind risk forecasting model for wood-frame houses. Risk Anal 27:45–58. https://doi.org/10.1111/j.1539-6924.2006.00858.x
Jones B, O'Neill BC (2016) Spatially explicit global population scenarios consistent with the shared socioeconomic pathways. Environ Res Lett 11:084003. https://doi.org/10.1088/1748-9326/11/8/084003
Magliocca N, Safirova E, McConnell V, Walls M (2011) An economic agent-based model of coupled housing and land markets (CHALMS). Comput Environ Urban Syst 35:183–191. https://doi.org/10.1016/j.compenvurbsys.2011.01.002
Mann ML, Berck P, Moritz MA et al (2014) Modeling residential development in California from 2000 to 2050: integrating wildfire risk, wildland and agricultural encroachment. Land Use Policy 41:438–452. https://doi.org/10.1016/j.landusepol.2014.06.020
McKenna F, Gavrilovic S, Zsarnoczay A, et al (2022) NHERI-SimCenter/R2DTool: Version 2.0.0
Musa SI, Hashim M, Reba MNM (2017) A review of geospatial-based urban growth models and modelling initiatives. Geocarto Int 32:813–833. https://doi.org/10.1080/10106049.2016.1213891
NAHB (2022) Share of Smaller Lots Record High Amid Pandemic. In: NAHB. https://eyeonhousing.org/2022/09/share-of-smaller-lots-record-high-amid-pandemic/. Accessed 23 Jan 2023
National Research Council, National Research Council (U.S.), National Research Council (U.S.), National Research Council (U.S.) (eds) (2014) Advancing land change modeling: opportunities and research requirements. National Academies Press, Washington, D.C
NC Dept. of Information Technology (2020) NC OneMap
NCDOT (2022) Roadway Design Manual. https://connect.ncdot.gov/projects/Roadway/Pages/RDM.aspx. Accessed 3 Jan 2023
NOAA (2022) Continental United States Hurricane Impacts/Landfalls 1851–2021. https://www.aoml.noaa.gov/hrd/hurdat/All_U.S._Hurricanes.html. Accessed 13 Jan 2023
NOAA (2018) Hurricane Florence: September 14, 2018. https://www.weather.gov/ilm/HurricaneFlorence. Accessed 13 Jan 2023
NOAA (2023) Costliest U.S. Tropical Cyclones. https://www.ncei.noaa.gov/access/billions/dcmi.pdf. Accessed 13 Jan 2023
Parker DC, Filatova T (2008) A conceptual design for a bilateral agent-based land market with heterogeneous economic agents. Comput Environ Urban Syst 32:454–463. https://doi.org/10.1016/j.compenvurbsys.2008.09.012
Rifat SAA, Liu W (2022) Predicting future urban growth scenarios and potential urban flood exposure using artificial neural network-Markov chain model in miami metropolitan area. Land Use Policy 114:105994. https://doi.org/10.1016/j.landusepol.2022.105994
Sanderson DR, Cox DT, Amini M, Barbosa AR (2022) Coupled urban change and natural hazard consequence model for community resilience planning. Earth's Future 10:e2022EF003059. https://doi.org/10.1029/2022EF003059

Sleeter BM, Wood NJ, Soulard CE, Wilson TS (2017) Projecting community changes in hazard exposure to support long-term risk reduction: a case study of tsunami hazards in the U.S. Pacific Northwest. Int J Disaster Risk Reduct 22:10–22. https://doi.org/10.1016/j.ijdrr.2017.02.015

Song J, Fu X, Wang R et al (2018) Does planned retreat matter? Investigating land use change under the impacts of flooding induced by sea level rise. Mitig Adapt Strateg Glob Change 23:703–733. https://doi.org/10.1007/s11027-017-9756-x

Strader SM, Ashley W, Walker J (2015) Changes in volcanic hazard exposure in the Northwest USA from 1940 to 2100. Nat Hazards 77:1365–1392. https://doi.org/10.1007/s11069-015-1658-1

Strader SM, Ashley WS (2015) The expanding bull's-eye effect. Weatherwise 68:23–29. https://doi.org/10.1080/00431672.2015.1067108

Strader SM, Ashley WS, Pingel TJ, Krmenec AJ (2018) How land use alters the tornado disaster landscape. Appl Geogr 94:18–29. https://doi.org/10.1016/j.apgeog.2018.03.005

Striessnig E, Gao J, O'Neill BC, Jiang L (2019) Empirically based spatial projections of US population age structure consistent with the shared socioeconomic pathways. Environ Res Lett 14:114038. https://doi.org/10.1088/1748-9326/ab4a3a

Theobald D (2005) Landscape patterns of exurban growth in the USA from 1980 to 2020. Ecol Soc. https://doi.org/10.5751/ES-01390-100132

US Census Bureau (2022). In: American Community Survey: B25024 Units in Structure. https://data.census.gov/table?t=Units+and+Stories+in+Structure&tid=ACSDT5Y2021.B25024

US EPA (2023) About ICLUS. https://www.epa.gov/gcx/about-iclus. Accessed 13 Jan 2023

USACE (2022) Technical Documentation. In: NSI Technical References. https://www.hec.usace.army.mil/confluence/nsi/technicalreferences/latest/technical-documentation. Accessed 11 Feb 2023

USGS (2020) Protected Areas Database of the United States (PAD-US) 2.1. https://doi.org/10.5066/P92QM3NT. Accessed 15 Jul 2020

Ustaoglu E, Lavalle C (2017) Examining lag effects between industrial land development and regional economic changes: The Netherlands experience. PLoS ONE 12:e0183285. https://doi.org/10.1371/journal.pone.0183285

Verburg PH, Schot PP, Dijst MJ, Veldkamp A (2004) Land use change modelling: current practice and research priorities. GeoJournal 61:309–324. https://doi.org/10.1007/s10708-004-4946-y

Wagenaar D, Curran A, Balbi M et al (2020) Invited perspectives: How machine learning will change flood risk and impact assessment. Nat Hazard 20:1149–1161. https://doi.org/10.5194/nhess-20-1149-2020

Wang D, Davidson RA, Nozick LK et al (2020) Computational framework to support government policy-making for hurricane risk management. Nat Hazards Rev 21:04019012. https://doi.org/10.1061/(ASCE)NH.1527-6996.0000348

Williams C, Davidson R (2022) Regional county-level housing inventory predictions and the effects on hurricane risk using long-short term memory (LSTM) methods and applied to the southeastern United States (US). Data Publication PRJ-3303 DesignSafe-CI: https://doi.org/10.17603/ds2-vd28-pe79

Williams C, Davidson R (2023) Housing Inventory Projection (HIP) method. In: Projection method for predicting spatiotemporal changes of a regional single-family housing inventory for hurricane risk modeling applications, vol 1. https://doi.org/10.17603/ds2-n1w2-e050

Williams CJ, Davidson RA, Nozick LK et al (2022) Regional county-level housing inventory predictions and the effects on hurricane risk. Nat Hazard 22:1055–1072. https://doi.org/10.5194/nhess-22-1055-2022

Wing OEJ, Lehman W, Bates PD et al (2022) Inequitable patterns of US flood risk in the anthropocene. Nat Clim Chang 12:156–162. https://doi.org/10.1038/s41558-021-01265-6

Yang K, Davidson RA, Blanton B et al (2022) Evaluation of hurricane evacuation order plans: hurricane florence case study. Nat Hazard Rev 23:05022010. https://doi.org/10.1061/(ASCE)NH.1527-6996.0000589

Zillow (2021) ZTRAX: Zillow Transaction and Assessor Dataset, 2021-Q3. http://www.zillow.com/ztrax/. Accessed 13 Mar 2022

Zsarnóczay A, Deierlein GG, Williams CJ et al (2023) Community perspectives on simulation and data needs for the study of natural hazard impacts and recovery. Nat Hazard Rev 24:04022042. https://doi.org/10.1061/NHREFO.NHENG-1551