OVERVIEW OF HOUSING UNIT INVENTORIES AND APPLICATIONS

in

DETAILED HOUSEHOLD AND HOUSING UNIT CHARACTERISTICS:

ALPHA RELEASE OF HOUSING UNIT INVENTORIES

NATHANAEL ROSENHEIM

College of Architecture

Hazard Reduction & Recovery Center

Department of Landscape Architecture & Urban Planning

Texas A&M University – College Station

Suggested Citation: Rosenheim, N. (2021). "Detailed Household and Housing Unit Characteristics: Alpha Release of Housing Unit Inventories." DesignSafe-CI. [check for doi] https://www.designsafe-ci.org/data/browser/public/designsafe.storage.published/PRJ-2961]

EXECUTIVE SUMMARY

People are the most important part of community resilience planning. While publicly available population data provide significant details about people these data are aggregated in spatially large areal units. Engineering models for building and infrastructure damage tend to be highly granular, which creates a mismatch with population data. The housing unit inventory method provides a way to overcome this limitation. The method transforms aggregated areal unit data into disaggregated housing unit data that includes occupied and vacant housing unit characteristics. Applications use the housing unit allocation method to assign the housing unit inventory to structures through a reproducible and randomized process. Housing units, in turn, are linked to specific residential structures, which are linked to infrastructure. These fully integrated data enable researchers to design "people-first" community resilience models. This project archives example data and replication files for six communities. The examples demonstrate the scalability of the methodology and the applications to different hazards. The applications of the method allow for detailed statistics on how modeled hazards impact populations and population subgroups. The archive provides an Alpha Release of the data, codebooks, replication code and applications. An Alpha Release indicates that the features have not been locked down and that the archive includes the exploratory phase of the methodology. Applications of the method are reproducible in the open-source platform called Interdependent Networked Community Resilience Modeling Environment (IN-CORE).

FUNDING ACKNOWLEDGMENTS

Funding for this study was provided as part of a cooperative agreement between the US National Institute of Standards and Technology and Colorado State University (NIST Financial Assistance Award Numbers: 70NANB15H044 and 70NANB20H008). The views expressed are those of the authors and may not represent the official position of the National Institute of Standards and Technology or the US Department of Commerce.

The material related to Santa Rose County, CA was based upon work supported by the National Science Foundation under Grant Nos. 1638273. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

KEY TERMS

American Community Survey (ACS): "The American Community Survey (ACS) is a nationwide survey designed to provide communities a fresh look at how they are changing. The ACS replaced the decennial census long form in 2010 and thereafter by collecting long form type information throughout the decade rather than only once every 10 years." (U.S. Census Bureau Glossary: https://www.census.gov/glossary/#term_AmericanCommunitySurveyACS).

Decennial census: "The census of population and housing, taken by the Census Bureau in years ending in 0 (zero)." (U.S. Census Bureau Glossary: https://www.census.gov/glossary/#term_Decennialcensus).

Family Household: "A family includes a householder and one or more people living in the same household who are related to the householder by birth, marriage, or adoption. All people in a household who are related to the householder are regarded as members of his or her family." (U.S. Census Bureau Glossary: https://www.census.gov/glossary/#term_Familyhousehold).

Household: "A household includes all the people who occupy a housing unit (such as a house or apartment) as their usual place of residence. A household includes the related family members and all the unrelated people, if any, such as lodgers, foster children, wards, or employees who share the housing unit. A person living alone in a housing unit, or a group of unrelated people sharing a housing unit such as partners or roomers, is also counted as a household. The count of households excludes group quarters." (U.S. Census Bureau Glossary: https://www.census.gov/glossary/#term_Household).

Householder: "The person, or one of the people, in whose name the home is owned, being bought, or rented." (U.S. Census Bureau Glossary: https://www.census.gov/glossary/#term_Householder).

Housing Unit: "A house, an apartment, a mobile home or trailer, a group of rooms, or a single room occupied as separate living quarters, or if vacant, intended for occupancy as separate living quarters. Separate living quarters are those in which the occupants live separately from any other individuals in the building and which have direct access from outside the building or through a common hall." (U.S. Census Bureau Glossary: https://www.census.gov/glossary/#term_Housingunit).

Nonfamily Household: "A nonfamily household consists of a householder living alone (a one-person household) or where the householder shares the home only with people to whom he/she is not related (e.g., a roomate)." (U.S. Census Bureau Glossary: https://www.census.gov/glossary/#term_Nonfamilyhousehold).

Census block: "A statistical area bounded by visible features, such as streets, roads, streams, and railroad tracks, and by nonvisible boundaries, such as selected property lines and city, township, school districts, and county boundaries. A block is the smallest geographic unit for which the Census Bureau tabulates decennial census data… Over 11 million blocks were identified for the 2010 Census." (U.S. Census Bureau Glossary: https://www.census.gov/glossary/#term_Censusblock).

Tenure: "Refers to the distinction between owner-occupied and renter-occupied housing units." (U.S. Census Bureau Glossary: https://www.census.gov/glossary/#term_Tenure).

TABLE OF CONTENTS

1. OVERVIEW OF HOUSING UNIT INVENTORY APPROACH

All text in this section is from:

Rosenheim, N., Guidotti, R., Gardoni, P., & Peacock, W. G. (2019). Integration of detailed household and housing unit characteristic data with critical infrastructure for post-hazard resilience modeling. *Sustainable and Resilient Infrastructure*, 1-17. https://doi.org/10.1080/23789689.2019.1681821

Please cite the original article when referencing this methodology.

The goal of the methodology presented in Rosenheim et al. (2019) is to link social units, such as households with specific attributes, to infrastructure systems. Together the integrated data will enhance models of community resilience. The critical link in this process is the housing unit, which is defined by the U.S. Census as "a house, an apartment, a mobile home or trailer, a group of rooms, or a single room occupied as a separate living quarters, or if vacant, intended for occupancy as a separate living quarters." (U.S. Census Bureau, n.d.). The proposed methodology transforms (namely transposes, expands, and appends) the aggregated household and housing unit areal data to individual housing units. Each housing unit has specific characteristics and is associated with a particular type of housing unit that can be linked to the building inventory via a random process. Both the specific set of characteristics associated with households and the nature of the randomization process will depend upon the data needs associated with the modeling the researchers are interested in undertaking and upon potential data limitations of the census and administrative data available to construct respective inventories. This paper illustrates an application based on the housing unit characteristics (tenure, size and vacancy) and a naïve Monte Carlo Simulation (MCS) process. We use Stata 15.1 for the MCS and the statistical analysis (StataCorp, 2017). As an example, this application considers one infrastructure system, a potable water system, although the methodology could be extended to multiple systems. The following section describes the data base structure for the proposed housing unit inventory approach.

1.1 Overview of Database Structure

Four inventories are needed to link the housing unit characteristic data with critical infrastructure data. Figure 1 outlines the database structure and relationships between them. Figure 1 provides an entity-relationship model, with “crow’s feet” notation, that separates information about the entities and their relationships to each other (Chen, 1976; Hitchman, 2002). The entity relationship diagram provides a standard format to communicate how relational databases need to be structured in order for the proposed methodology to function. Each of the four inventories is represented by an entity box with a list of attributes. Attributes that uniquely identify the unit of analysis are highlighted as Primary Keys (PK). As an example, the *Housing Unit Inventory* has a PK called *Housing Unit ID* which uniquely identifies each housing unit in the dataset. The entity box also lists three attributes that describe the characteristics of the housing unit and occupying household. The attributes include the number of people in each housing unit, tenure status, and vacancy type. Each inventory is a datafile where each row is a unique unit of analysis and each column has attribute information. Figure 1 identifies the dependencies or key attributes that link the housing unit characteristics to address points, which are linked to buildings and ultimately to critical infrastructure.

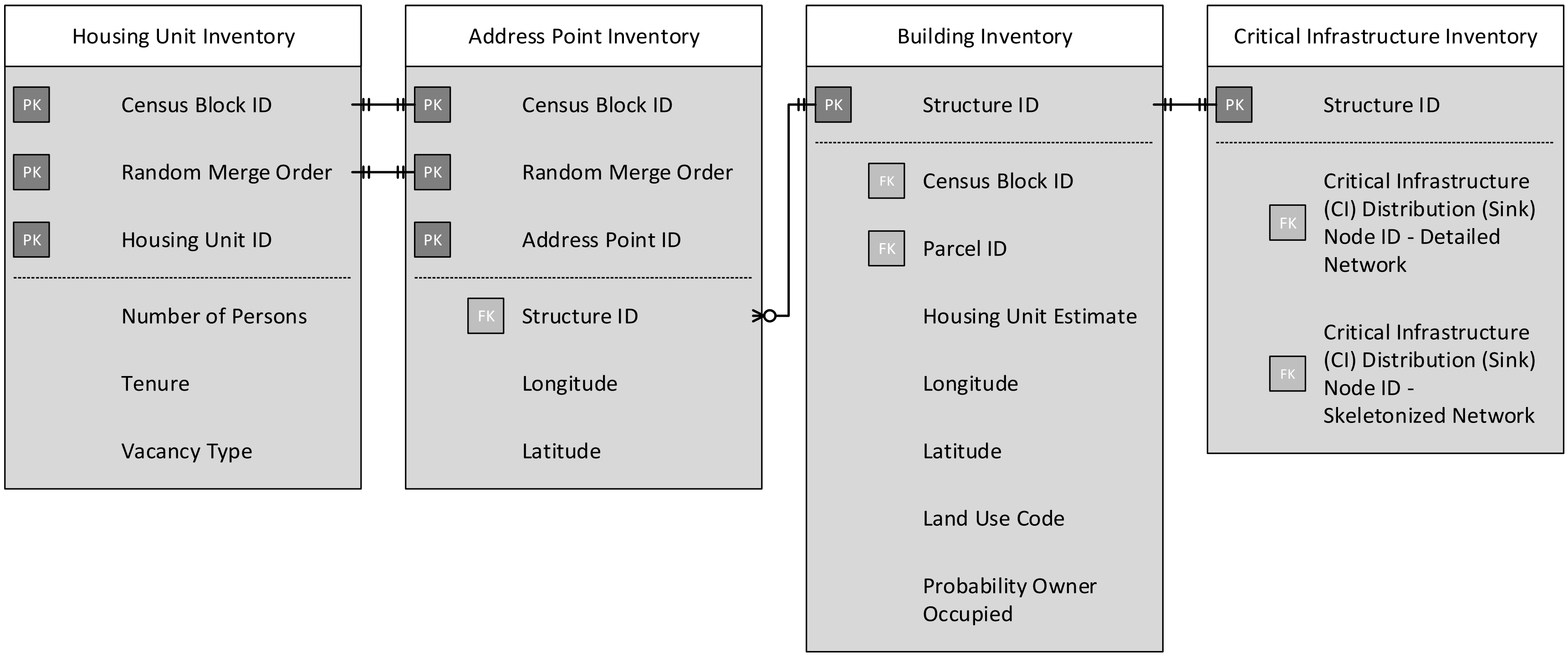

Figure 1.1. Database structure and relationships between *Housing Unit*, *Address Point*, *Building* and *Critical Infrastructure Inventories*.

The first inventory is the *Housing Unit (HU) Inventory*, where each observation corresponds to a vacant or occupied housing unit along with details of size, tenure, and vacancy type attributes based on occupancy status. Details on the creation of this inventory are discussed in Section 4.3. The second inventory is the *Address Point (AP) Inventory*, where each observation corresponds to a specific address where one and only one HU might be located. Buildings may have zero, one or more APs; however, each AP is assigned to only one building/structure. For example, a detached garage would have no AP, a single-family or mobile home would have one AP, and an apartment building with multiple HUs would have APs for each HU. Each AP includes a *Structure ID* code.

The third inventory is the *Building Inventory* (which includes all residential structures), where each observation corresponds to a single structure at a known location. Each building is assigned a unique *Structure ID* code as well, which can be linked to one or more APs and critical infrastructure nodes. The *Building Inventory* could be based on a community's parcel data under the assumption that each parcel has one building. For example, in communities comprised mainly by single-family homes the majority of parcels will have one building, with one HU and, hence, one AP. However, the assumption that one parcel equals one building is often not valid, because apartment complexes, mixed-use developments, and mobile home parks are often located on a single parcel.

The fourth inventory is the *Critical Infrastructure (CI) Inventory*. This inventory differs from the others in the sense that it is a relationship table where the unit of analysis remains an individual building with a *Distribution Node ID* code assigned to the structure. Each building should be assigned to a single distribution node but multiple buildings could be assigned to a single distribution node, depending on the nature of the infrastructure and how detailed the infrastructure system is modelled. The link to one building ensures that each HU is linked to a distribution node, which will make it possible to measure changes in demand or changes in supply due to infrastructure disruption. CI systems can be modelled at different levels of granularity and multiple scales of network resolution. The number of structures associated with a distribution node will vary depending on whether detailed or skeletonized are the models of infrastructure employed (Guidotti and Gardoni, 2018).

The database structure illustrated in Figure 1 also identifies unique variables or keys, referred to as Primary Keys (PKs) and Foreign Keys (FKs). Both keys are unique, non-missing values that help identify and link specific observations in each inventory, where a PK from one inventory is termed as FK when incorporated into another inventory. Keys allow the linkage across inventories and facilitate modularity. Relational databases allow inventories to be independently updated as new housing unit, building, or critical infrastructure data become available. As suggested in Figure 1, *Structure ID* is the key link between the *AP*, *Building* and *CI Inventories*. For the purposes of this paper, the links between these three inventories are treated as deterministic, although not without error. In other words, it is assumed that there will be sufficient data to identify residential structures from non-residential structures (as well as their associated APs) and to ascertain how infrastructure components are linked to buildings (both residential and non-residential). The degree to which this assumption holds will be variable across, as well as within, communities, and several approaches might otherwise be employed to link these inventories. The paper focuses here on the critical issue of a probabilistic match between the *HU* and *AP Inventories*.

As illustrated in Figure 1, the link between the *HU* and *AP Inventories* requires the use of multiple PKs. Both inventories have a *Census Block ID* which will directly guide which observations in each dataset should be linked within relatively small geographic areal units defined by the U.S. Census. In the *HU Inventory*, the variable *HU ID* uniquely identifies each observation while in the *AP Inventory*, as discussed above, the *AP ID* identifies each unique physical location within specific residential buildings where a HU could be located. Additionally, unique identifiers termed *Random Merger Order* (RMOs) will be generated in both inventories. The *Census Block IDs* and *RMOs* will be used to assign the detailed household level data to specific APs, which process of assignment ultimately links all inventories together. Before clearly articulating the probabilistic process that links the *HU* and *AP Inventories* the next section provides a detailed discussion of how the HU inventory data was generated from census block data.

1.2 From Areal Census Units to Housing Unit Inventory

Figure 2 provides an illustration of the basic steps in the transformation process of moving from standard block level, decennial census data to inventory data capturing vacant housing unit and occupied housing unit data with integrated household characteristics. The original datasets, shown on the left side of Figure 2, with upper and lower panels, have observations (rows) representing each block with data on counts of occupied HUs by household size and tenure status (top panel left) and vacant HUs by vacancy type (lower panel, left). These tables provide a full accounting of the households, tenure status, occupied and vacant HUs, and the number of people that live in occupied housing units in each block at the time of the decennial census.

Original Data Format

Households by Size and Tenure

Block ID	1 Person Household Owner	2+ Person Household Owner
Block 1	2	3
Block 2	0	5
Block 3	0	0

HUs by Vacant Unit Type

Block ID	For Sale	For Rent
Block 1	0	1
Block 2	2	2
Block 3	0	0

Transposed Data Format

Households by Size and Tenure

Block ID	Household Type	Number of HUs
Block 1	1 Person Household Owner	2
Block 1	2+ Person Household Owner	3
Block 2	1 Person Household Owner	0
Block 2	2+ Person Household Owner	5
Block 3	1 Person Household Owner	0
Block 3	2+ Person Household Owner	0

HUs by Vacant Unit Type

Block ID	Vacant Unit Type	Number of HUs
Block 1	For Sale	0
Block 1	For Rent	1
Block 2	For Sale	2
Block 2	For Rent	2
Block 3	For Sale	0
Block 3	For Rent	0

Expanded and Appended Data Format

HU Inventory

Block ID	HU ID	Number of Persons	Tenure	Vacancy Type
Block 1	B1H1	1	Owner	
Block 1	B1H2	1	Owner	
Block 1	B1H3	2	Owner	
Block 1	B1H4	2	Owner	
Block 1	B1H5	2	Owner	
Block 2	B2H1	2	Owner	
Block 2	B2H2	2	Owner	
Block 2	B2H3	2	Owner	
Block 2	B2H4	2	Owner	
Block 2	B2H5	2	Owner	
Block 1	B1H6	0		For Rent
Block 2	B2H6	0		For Sale
Block 2	B2H7	0		For Sale
Block 2	B2H8	0		For Rent
Block 2	B2H9	0		For Rent

Figure 1.2. Data formats and process flow required to generate a *Housing Unit Inventory*.

The next step (center panels, Figure 2) is to transpose the original block data generating two tables for each block: 1) household types defined by number of persons and tenure and 2) vacant housing units by expected use. The simplified original data tables have one observation for each block while the transposed tables have multiple observations for each block, the number of observations depending on the number of variable types for households and HUs. For example, the occupied housing unit table (top panel, left column) has 3 blocks with two household type possibilities, generating a transposed table with two observations for each block capturing detailed household types and the expected number of these types in occupied housing units located in each block. Similar transposition patterns are evident for vacant HUs.

In this paper's application, the focus is on occupancy status, tenure, and household size. However, depending on the research goals, multiple household and housing unit characteristics might be employed. Census Block files provide data on seven types of household size (1 through 7+), for owners and for renters, across racial and ethnic categories. Census tables provide 196 mutually exclusive household types (combinations) where each household can be identified by size, race, ethnicity, and tenure. Additionally, Census data can also provide up to seven types of vacancies (for sale, for rent, seasonal, etc.). It is important to notice that households with more than 7 members are assumed to only have 7 members, which typically result in a small reduction on the size of the community.

Once transposed, the data are further processed as reflected in the final panel on the right side of Figure 2. For the household by size and tenure data table (top middle panel), all zero (0) count data are dropped, while household size and tenure combinations with counts greater than zero (0) in each block are expanded by their respective counts. The process of expansion makes the unit of analysis a single occupied HU with information on the number of persons and tenure of the household. Similarly, the vacant housing unit types data with counts equal to zero (0) are dropped and those with counts greater than zero (0) are expanded for each block and appended to the evolving housing unit inventory data. The right most panel of Figure 2, shows the new *HU Inventory* data variables generated to identify detailed housing unit characteristics. The final HU inventory retains all of the information of the original data formatted to facilitate linking with APs assigned to residential structures which have been linked to *CI Inventories*.

1.3 The Probabilistic Allocation of Detailed Household and Housing Unit Characteristics

There are a variety of approaches that might be utilized to link the derived housing unit inventory to APs and subsequently to critical infrastructure. For example, HUs can be developed specifying not only household size, tenure status and vacancy, but also disposition to specific types of structures (single family detached, multi-family with varying numbers of HUs, etc.). If building inventory data, based on parcel data and other forms of administrative data, provide details related to these factors, it may be possible to: i) deterministically assign HUs to specific types of buildings with limited error; ii) to utilize partial deterministic assignment in combination with random assignment of specific types of HUs within certain types of buildings; or iii) to utilize complete random assignment. In this paper's application, HUs are assigned to APs probabilistically using parcel data to predict the probability that a building will be owner-occupied. A naïve Monte Carlo Simulation (MCS) process then matches HUs to APs within each census block.

As illustrated in Figure 1, the individual HUs and APs in both inventories have unique PKs (*HU IDs* and *AP IDs* respectively). To ensure reproducibility of results, a different random seed was set for each MCS iteration. For each iteration of the MCS the *HU* and *AP Inventories* are initially ordered by sorting on the unique PKs before using a pseudorandom-number generator to generate *Random Merge Order* (*RMOs*) PKs in each inventory. These *RMOs*, along with the *Census Block IDs*, are used to merge the inventories. As noted above, the number of HUs should be fewer than or equal to the number of APs within census blocks. In cases where there are too few APs, either due to error in the *AP Inventory* or potentially to a block that has lost buildings after the decennial census, the linked datasets may not account for all housing units and may incorrectly represent the community.

2. CASE STUDY 1: SEASIDE, OR

The first example application of the housing unit inventory and housing unit allocation methodology was Seaside, Oregon (Guidotti, Gardoni, & Rosenheim 2019; Rosenheim, Guidotti, Gardoni, & Peacock, 2019). The motivation for the method arose from a need to estimate the demand on the water network before and after an earthquake.

2.1 Community Description

All text in this section is from:

Rosenheim, N., Guidotti, R., Gardoni, P., & Peacock, W. G. (2019). Integration of detailed household and housing unit characteristic data with critical infrastructure for post-hazard resilience modeling. *Sustainable and Resilient Infrastructure*, 1-17. https://doi.org/10.1080/23789689.2019.1681821

Please cite the original article when referencing text in this section.

The Seaside Housing Unit Inventory is based on data files from the 2010 Decennial Census that include housing unit characteristics summarized to the block level (U.S. Census Bureau, 2011, 2012). The U.S. Census Bureau (2011) summary files contain all levels of geographies for an entire state, and to reduce the file size the initial HU Inventory is generated for Clatsop County (FIPS code 41007), which contains Seaside. For example, the political boundaries of Seaside contain 303 Census Blocks; of these, 207 have at least 1 person living in the block, but, as will be discussed in the next section, the Building Inventory includes 24 blocks that are outside of the political boundaries of Seaside. Since researchers will often face these minor inconsistencies in boundaries among inventories, it will often be easier to work with county level census files to enhance matching between the Building and CI Inventories that may extend beyond the political boundaries of a city, Seaside in this case. Final adjustments to consistent boundaries can be resolved after the linking process.

For the 207 Seaside blocks with population, the average block had a population of 31 people with the largest block having a population of 458. The process described in section 2.2

transforms the block level data from areal unit counts into a housing unit level inventory with 21,579 observations for the county and 4,638 observations for the city. Table 2.1 provides a summary of the HU Inventory for Seaside by household size, tenure, and vacancy.

Table 2.1 Housing Unit Inventory comparison of HU counts by household size, tenure and vacancy, compared with 2010 Census total population in occupied HUs: Seaside, OR.

Detailed HU Characteristics	Owner-Occupied	Renter-Occupied	HU Count
1-person household	431	715	1,146
2-person household	568	451	1,019
3-person household	150	191	341
4-person household	98	155	253
5-person household	39	88	127
6-person household	19	34	53
7-or-more-person household	12	18	30
Vacant			1,669
Total HUs:	1,317	1,652	4,638
Total Population in HU Inventory:	2,802	3,580	6,382
Total Population in Occupied HU:	2,812	3,598	6,410

Sources: (U.S. Census Bureau, 2010a, 2010b, 2010c)

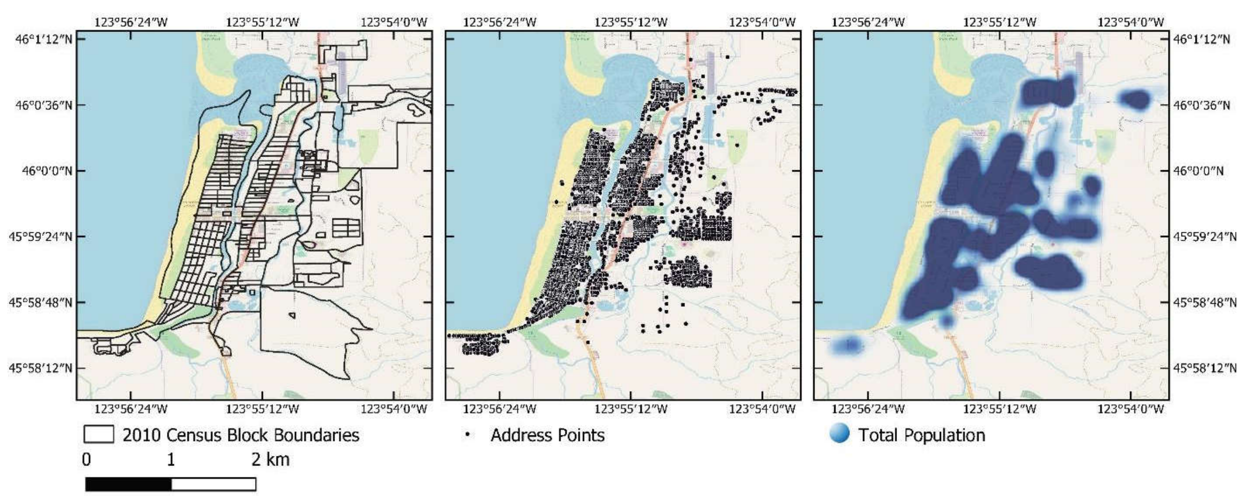

Figure 2.1 Seaside, OR overview maps: 2010 Census Blocks (left), Address Points (center), Allocated HU Heat Map (right).

Figure 2.1 provides an illustration of the integration of the *HU* and *AP Inventories* as well as an overview of Seaside. The left-side map provides the outlines of the 2010 Census Block boundaries that have APs. The center map shows the location of the 5,988 APs for the community. The right-side map illustrates the resulting population heat map capturing the number of people located in neighborhoods within Seaside based on one probabilistic allocation of the *HU* and *AP Inventories*.

2.2 Verification and Validation

All text in this section is from:

Rosenheim, N., Guidotti, R., Gardoni, P., & Peacock, W. G. (2019). Integration of detailed household and housing unit characteristic data with critical infrastructure for post-hazard resilience modeling. *Sustainable and Resilient Infrastructure*, 1-17. https://doi.org/10.1080/23789689.2019.1681821

Please cite the original article when referencing text in this section.

The HU Inventory provides a representative count of the community population, but it does have some error. For the HU Inventory, the primary source of error is from family size. The block level data reports family sizes from 1 to 6 and 7 or more; for households with more than 7 members, the inventory simply assumes that there are 7 people in the household. As shown in Table 2.1, the HU Inventory thus estimates the total population to be 6,382, an undercount of 28 people compared with the total population in occupied housing units from the 2010 Census. A secondary source of error comes from not including group quarters as a type of housing. Seaside has 47 people that lived in Group Quarters, 19 in nursing facilities and 28 in other noninstitutional facilities (U.S. Census Bureau, 2010d). Comparing the total HU Inventory population estimate from Table 1 of 6,382 people with the 2010 census total population count of 6,457 reveals an undercount of 75 people or 1.2%. Despite this known error, the HU Inventory provides detailed household and housing unit characteristics that are representative of the 2010 Census. Each occupied HU in the *HU Inventory* was assigned a tenure status, either owner or renter. Each vacant housing unit was assigned a vacancy type which includes housing units that were for rent, for sale or for seasonal use. It is important to recognize that the decennial census was conducted on April 1, 2010, and therefore counts reflect the community during the off-season and not during the height of the summer tourist season.

To summarize the sources of error generated by the integration of Census, parcel, and infrastructure data, Table 2.2 provides an accounting of how the errors impact the count of HUs and hence the estimate of the community's total population. For Seaside, differences between the political boundary and the location of buildings and infrastructure associated with the community introduced an additional 112 HUs and 238 people after the inventories were integrated. The community lost 122 HUs and 180 people due to errors in the matching of APs to buildings or buildings to blocks. These two sources of error combine to generate a final integrated dataset that closely approximating the Census figures with 10 fewer [4,628 vs 4,638] total HUs (16 fewer [1,653 vs 1,669] vacant and 6 more [2,975 vs 2,969] occupied HUs) and 17 fewer [6,440 vs 6,457] people. A third source of error that does not impact HU count or population size comes from HUs that are missing CI data. For the integrated data, 74 HUs and an estimated 96 persons were not connected to a water demand tributary within the political boundary of Seaside.

Table 2.2 Summary of sources of error from integration of *Housing Unit*, *Address Point*, *Building* and *Critical Infrastructure Inventories*. Seaside, OR.

Detailed HU Characteristics	Outside Political Boundary	Missing AP/Building Data	Missing CI Data
1-person household	18	38	12
2-person household	46	32	17
3-person household	16	9	4
4-person household	8	6	4
5-person household	3	4	3
6-person household	2	0	0
7-or-more-person household	3	1	1
Vacant	16	32	33
Total Housing Units:	+112	-122	74
Total Population Impacted by Error:	+238	-180	96

2.3 Census references for verification

U.S. Census Bureau. (2010a). Table H5: Vacancy status. Universe: Vacant Housing Units, 2010 Census Summary File 1. Retrieved from https://data.census.gov/cedsci/table?g=1600000US4165950&tid=DECENNIALSF12010.H5

U.S. Census Bureau. (2010b). Table H11: Total population in occupied housing units by tenure. Universe: Population in occupied housing units, 2010 Census Summary File 1. Retrieved from https://data.census.gov/cedsci/table?g=1600000US4165950&tid=DECENNIALSF12010.H11

U.S. Census Bureau. (2010c). Table H16: Tenure by household size. Universe: Occupied housing units, 2010 Census Summary File 1. Retrieved from https://data.census.gov/cedsci/table?g=1600000US4165950&tid=DECENNIALSF12010.H16

U.S. Census Bureau. (2010d). Table P42: Group quarters population by group quarters type. Universe: Population in group quarters, 2010 Census Summary File 1. Retrieved from https://data.census.gov/cedsci/table?g=1600000US4165950&tid=DECENNIALSF12010.P42

2.4 Lessons Learned from Seaside, Oregon

The integration of social science and engineering data opens up unprecedented opportunities for coupling engineering and social science modeling and improvements to post-hazard resilience models. As presented in this paper, Census data, which provides detailed household and housing unit characteristics, can be transformed and probabilistically allocated to individual buildings. Additionally, critical infrastructure can be mapped to buildings to accurately capture the demand on the system. The allocation of Census data to buildings helps to determine the baseline demand within critical infrastructure's tributary areas. The database structure presented in this paper constitutes an example of how datasets from different disciplines can be effectively structured so that data can be integrated.

The housing unit allocation method made it possible to predict the number of people served at each critical infrastructure demand node (Guidotti, Gardoni, & Rosenheim 2019). The flexibility of the housing unit allocation methodology made it possible to predict household dislocation based on building damage and then estimate the change in demand for water, assuming that a dislocated household would either leave the community (external dislocation) or relocate to an assembly point (internal dislocation). Since the housing unit allocation method used disaggregated housing unit level estimates of population size, the critical infrastructure network could be very detailed (one node for each building) or skeletonized, with groups of buildings aggregated. Previous methodologies that used aggregated spatial boundaries (such as census tracts) made it difficult to link critical infrastructure to population data, since the skeletonized critical infrastructure network did not match to the arbitrary census geography. The application of the housing unit allocation simplified the process of estimating demand.

3. CASE STUDY 2: SHELBY COUNTY, TN

All text in this section is from:

Roohi, M., van de Lindt, J. W., Rosenheim, N., Hu, Y., & Cutler, H. (2020). Implication of building inventory accuracy on physical and socio-economic resilience metrics for informed decision-making in natural hazards. *Structure and Infrastructure Engineering*, 1-21. https://doi.org/10.1080/15732479.2020.1845753

Please cite the original article when referencing text in this section.

Roohi et al. (2020) highlights the integration of building data and housing unit inventories to predict population dislocation after an earthquake. This case study demonstrates that the housing unit inventory and allocation method is scalable to a large densely populated urban area. Additionally, this case study shows the potential benefits of combining engineering and social science data to help identify areas to improve engineering data and limit impacts of missing data on building attributes. For the Shelby County, TN the housing unit inventory allow for comparisons of results by tenure status.

3.1 Community Description

In 2010, Shelby County, TN, had a population of 909,315 people living in 350,971 occupied housing units (US Census Bureau, 2010a; US Census Bureau, 2010b). Housing unit and household characteristics were allocated to the building inventory described above (Rosenheim et al., 2019).

3.2 Verification and Validation

The resulting model population included 549,647 people living in owner-occupied structures and 308,165 people living in renter-occupied structures. The modeled population had 5.6% fewer people than the estimated population due to missing buildings in the inventory. The validation process revealed that the missing data disproportionately impacted renters with over 43,000 renters not allocated to structures.

3.3 Census references for verification

U.S. Census Bureau. (2010a). Table H3: Occupancy status. Universe: Housing Units, 2010 Census Summary File 1. Retrieved from https://data.census.gov/cedsci/table?g=0500000US47157&tid=DECENNIALSF12010.H3

U.S. Census Bureau. (2010b). Table H11: Total population in occupied housing units by tenure. Universe: Population in occupied housing units, 2010 Census Summary File 1. Retrieved from https://data.census.gov/cedsci/table?g=0500000US47157&tid=DECENNIALSF12010.H11

3.4 Lessons Learned

When allocated household characteristics were compared to building attributes, it was found that renters are more likely to live in buildings with more than one story, and the buildings tend to be older. Missing data impacted renter-occupied structures more than owner-occupied structures. For example, if all buildings are assumed to be single-family, renter-occupied households would have a 68% error in building data, while owner-occupied households would have a 35% error. The error would also be the largest for vacant housing units at 70%. Overall, the use of the housing unit inventory and housing unit allocation reduced error by 50% for the population dislocation model. The allocation of housing unit data to the building data can help identify gaps in the building inventory and supplement missing building data.

This case study supports the integration of engineering and social science models and data for community resilience planning. This case study highlights that even as building data may have greater uncertainty and limited information, the addition of US Census Data, housing unit inventories based on the Decennial Census…. helps to mitigate data that many communities may not have the capacity to collect. Within the context of the United States, the US Census Data is available for all communities in a standardized format, which should make models for community resilience more robust and generalizable.

4. CASE STUDY 3: JOPLIN, MO

All text in this section is from:

Wang, L., van de Lindt, J.W., Rosenheim, N., Cutler, H., Hartman, B., Lee, J.S., & Calderon D. (2021). Effect of Building Wind-Retrofit Strategies on Social and Economic Community-Level Resilience Metrics. *Journal of Infrastructure Systems*.

Please cite the original article when referencing text in this section.

Wang et al. (2021) highlights the integration of building data and housing unit inventories to predict population dislocation after a tornado with damage to the electric power network. This case study demonstrates that the housing unit inventory and allocation method can be applied to different hazards types and critical infrastructure.

4.1 Community Description

Table 4.1. Allocated housing unit characteristics for Joplin, MO in 2010

Detailed HU Characteristics	Owner-occupied	Renter-occupied	Housing unit count
1-person household	3,135	3,788	6,923
2-person household	4,398	2,564	6,963
3-person household	1,658	1,414	3,072
4-person household	1,255	944	2,199
5-person household	566	433	1,000
6-person household	206	194	403
7-or-more-person household	126	98	224
Group Quarters	-	-	22
Vacant	-	-	2,455
Total number	11,344	9,435	23,261
Total population	26,873	20,949	49,810

Joplin, Missouri spans part of both Jasper and Newton County, with the majority being located in Jasper County. Based on the 2010 Decennial Census (U.S. Census Bureau 2010a, 2010b, 2010c), the city of Joplin had 23,322 housing units (11,389 owner-occupied, 9,471 renter-occupied, and 2,462 vacant). The household size was distributed from 1-person to 7-or-more persons, and the total population count of the community was 50,150, with 27,076 people living in owner-occupied housing units and 21,086 people living in renter-occupied housing units. Table 4.1 provides a summary of the allocated housing units and population characteristics.

4.2 Verification and Validation

Notice that the estimated housing units was 23,261 in the building inventory, compared to 23,322 based on the 2010 Census data, an error of 0.26% (U.S. Census Bureau 2010a; 2010b; 2010c). The average household size in Joplin was estimated to be 2.31 people (lower than the state average of 2.45), with owner-occupied households being slightly larger than renter-occupied households (2.38 vs. 2.23) (U.S. Census Bureau 2010d). The average allocated household size was 2.30; 2.37 for owner-occupied housing units and 2.22 for renter-occupied housing units, an error of 0.47%. Table 5.1 verifies that the population data in the model accurately reflects Joplin's estimated household size distribution by tenure status. These allocated data are important parameters to estimate population dislocation and the residents who were served by electric power following the tornado hazard.

4.3 Census references for verification

U.S. Census Bureau. (2010a). Table H5: Vacancy status. Universe: Vacant Housing Units, 2010 Census Summary File 1. Retrieved from https://data.census.gov/cedsci/table?g=1600000US2937592&tid=DECENNIALSF12010.H5

U.S. Census Bureau. (2010b). Table H11: Total population in occupied housing units by tenure. Universe: Population in occupied housing units, 2010 Census Summary File 1. Retrieved from https://data.census.gov/cedsci/table?g=1600000US2937592&tid=DECENNIALSF12010.H11

U.S. Census Bureau. (2010c). Table H16: Tenure by household size. Universe: Occupied housing units, 2010 Census Summary File 1. Retrieved from https://data.census.gov/cedsci/table?g=1600000US2937592&tid=DECENNIALSF12010.H16

U.S. Census Bureau. (2010d). Table H12: Average household size of occupied housing units by tenure. Universe: Occupied housing units, 2010 Census Summary File 1. Retrieved from https://data.census.gov/cedsci/table?g=1600000US2937592&tid=DECENNIALSF12010.H12

4.4 Lessons Learned

The allocation of detailed household characteristics helped to identify the impacts of different retrofit strategies by linking building damage results and the population dislocation model. Wang et al. (2021) was the first time that "the effect of retrofit strategies for tornado loading [were] quantified in terms of their effect on socio-economic metrics. The ability to quantify these effects to examine different retrofits (or policies) at the community level can help support community resilience planning."

5. CASE STUDY 4: GALVESTON, TX

All text in this section is from:

Fereshtehnejad, E., Gidaris, I., Rosenheim, N., Tomiczek, T., Padgett, J., Cox, D., Van Zandt, S., & Peacock, W. G. (2021). Probabilistic risk assessment of coupled natural-physical-social systems: the cascading impact of hurricane-induced damages to civil infrastructure in Galveston, Texas. *Natural Hazards Review.* https://doi.org/10.1061/(ASCE)NH.1527-6996.0000459

Please cite the original article when referencing text in this section.

Fereshtehnejad et al. (2021) highlights the integration of building data and housing unit inventories to predict population evacuation before a hurricane with damage to the transportation network. This case study demonstrates that the housing unit inventory and allocation method can be applied to different hazards types and critical infrastructure. Additionally, this case study links the housing unit inventory with household level field study data. This example helps demonstrate benefits of using household level data to generate social science models based on household level field study data. The Galveston housing unit inventory adds a predicted household level income to each household.

5.1 Community Description

Galveston Island is a barrier island located southeast of Houston, TX. The island's population is racially and ethnically diverse, with a wide income distribution. Table 3 summarizes housing unit, population, and socio-economic characteristics of Galveston Island based on data from the 2010 U.S. Census. At the time of the 2010 Census, the island population was 48,726, with the following racial/ethnic composition: 31% Hispanic, 46% non-Hispanic White and 18% non-Hispanic Black. The median household income on the island was $37,770, significantly lower than the median income for the United States ($53,046), Texas ($51,563), and the surrounding county ($61,555) (U.S. Census, 2012a). Race and ethnicity were significantly related to median household income: non-Hispanic White households had a median income of $50,331, compared to Hispanic households with a median income of $35,593 and non-Hispanic Black households with a median income of $18,866 (U.S. Census Bureau, 2012a). Income has a significant relationship with the decision and ability to evacuate, with lower income households having a lower probability of evacuating. Therefore, the ability to include household income in evacuation models is important to accurately reflect the variability in evacuation response due to a diverse household population on Galveston Island.

Table 5.1. Summary of housing unit, population and socio-economic characteristics of Galveston, Island, 2010.

Detailed HU Characteristics	Housing Unit Count	Population	Median Income
White alone, Not Hispanic	11,064	22,370	$50,331
Black or African American alone, Not Hispanic	4,641	8,909	$18,866
Other Race, Not Hispanic		2,450	
Hispanic or Latino	4,730	14,997	$35,593
Vacant	13,107	0	
Total:	33,542	48,726	$37,770

Fereshtehnejad et al. (2021) expanded the housing unit inventory approach by not only predicting vacant or occupied housing units but, when occupied, the household characteristics, including race/ethnicity, tenure status, and income. The distribution of income across Galveston Island based on race and ethnicity provides enhanced modeling capabilities. While the median income for the island was $37,770 in 2012, more than 70% of Black or African American households lived below the median income, and more than 60% of non-Hispanic White households lived above the median income.

To generate this more refined housing unit inventory, the 2012 American Community Survey (ACS) for Galveston is employed. The 2012 ACS covers a five-year period from 2008-2012, which includes the period in which Hurricane Ike struck the island and is the only source of income distribution data that can be linked to the 2010 Census. These data provide estimates of counts of households within 16 income categories ranging from less than $10,000 to $200,000 or more, broken down by the head of household's race/ethnicity. These data were disaggregated and associated with occupied housing units within each of the 23 census tracts associated with Galveston Island. The result was the generation of vacant and occupied housing unit inventory associated with residential units throughout Galveston Island.

5.2 Verification and Validation

The housing unit inventory for Galveston Island provides a representative model of the population based on race, ethnicity and household income. Table 6.2 summarizes the differences between the housing unit inventory and the 2010 Census (U.S. Census Bureau, 2010). Overall, the housing unit inventory is accurate to within 8% of the total population, with larger sources of error found in estimates of minority households. Race and ethnicity provide the bases for predicting household income.

Table 6.2. Comparison of population estimates between 2010 Census and Housing Unit Inventory for Galveston Island.

Population by Race, Ethnicity	2010 Census	Housing Unit Inventory	% Difference
White, Not Hispanic	22,370	21,220	-5%
Black or African American, Not Hispanic	8,909	8,302	-7%
Other, Not Hispanic	2,450	1,698	-31%
Hispanic or Latino	14,997	13,424	-10%
Total:	48,726	44,644	-8%

Table 6.3. Comparison of median income estimates between 2012 ACS and Housing Unit Inventory for Galveston Island.

Population by Race, Ethnicity	2012 ACS	Housing Unit Inventory	% Difference
White, Not Hispanic	$50,331	$51,419	2%
Black or African American	$18,866	$19,427	3%
Hispanic or Latino	$35,593	$32,934	-8%
Total Population	$37,770	$38,746	3%

When median income is used as a benchmark, the housing unit inventory with household income provides a reasonable approximation of estimates of the socio-economic characteristics of

Galveston Island. As shown in Table 6.3, the housing unit inventory matches median income of the entire island to within 3% for the total household median income, overestimating the median household income for Black households, and underestimating median income for Hispanic or Latino households (U.S. Census Bureau, 2012b). Having developed a full household inventory for the island's households, the next step was to develop an evacuation model whereby evacuating and non-evacuating households could be predicted.

5.3 Census references for verification

U.S. Census Bureau. (2010a). Table P2 Hispanic or Latino, and not Hispanic or Latino by Race, 2010 Census Summary File 1. Retrieved from https://data.census.gov/cedsci/table?g=1600000US4828068,4837252&tid=DECENNIALSF12010.P2

U.S. Census Bureau. (2012). B19013H, B19013B, B19013I: Median Household Income in the Past 12 Months (in 2012 Inflation-Adjusted Dollars) (White Alone, Not Hispanic or Latino Householder), (Black or African American Alone Householder), (Hispanic or Latino Householder), 2008-2012 American Community Survey 5-Year Estimates. Retrieved from https://data.census.gov/cedsci/table?g=1600000US4828068&tid= ACSDT5Y2012.B19013[H/B/I]

5.4 Lessons Learned

The housing unit inventories and allocation method help link community resilience planning models with models based on post hazard field study surveys. The housing unit inventories based on the decennial census can be expanded by adding more details household characteristics, like household income. The household data helps to preserve the intersectionality of race and class. The data validation and verification processes help to identify that error disproportionately impacts minority populations.

6. ABOUT THE RESEARCHER

Nathanael Rosenheim is an Associate Research Scientist for the Hazard Reduction and Recovery Center in the Department of Landscape Architecture and Urban Planning at Texas A&M University. Rosenheim's areas of interest are spatial modeling, data science, community development, and food system planning. His recent research uses public demographic and economic data to improve fact-based community planning for hazard mitigation and recovery planning. Rosenheim served as the Principal Investigator for this research project.

7. REFERENCES

Fereshtehnejad, E., Gidaris, I., Rosenheim, N., Tomiczek, T., Padgett, J., Cox, D., Van Zandt, S., & Peacock, W. G. (2021). Probabilistic risk assessment of coupled natural-physical-social systems: the cascading impact of hurricane-induced damages to civil infrastructure in Galveston, Texas. *Natural Hazards Review*. https://doi.org/10.1061/(ASCE)NH.1527-6996.0000459

Guidotti, R., Gardoni, P., & Rosenheim, N. (2019). Integration of physical infrastructure and social systems in communities' reliability and resilience analysis. *Reliability Engineering & System Safety*, 185, 476-492. https://doi.org/10.1016/j.ress.2019.01.008

National Science Foundation. (2016d). *Award abstract #1638273: CRISP type 2/collaborative research: Scalable decision model to achieve local and regional resilience of interdependent critical infrastructure systems and communities*. Retrieved from https://www.nsf.gov/awardsearch/showAward?AWD_ID=1638273

Nozhati, S., Rosenheim, N., Ellingwood, B. R., Mahmoud, H., & Perez, M. (2019). Probabilistic framework for evaluating food security of households in the aftermath of a disaster. *Structure and Infrastructure Engineering*, 15(8), 1060-1074. https://doi.org/10.1080/15732479.2019.1584824

Roohi, M., van de Lindt, J. W., Rosenheim, N., Hu, Y., & Cutler, H. (2020). Implication of building inventory accuracy on physical and socio-economic resilience metrics for informed decision-making in natural hazards. *Structure and Infrastructure Engineering*, 1-21. https://doi.org/10.1080/15732479.2020.1845753

Rosenheim, N., Guidotti, R., Gardoni, P., & Peacock, W. G. (2019). Integration of detailed household and housing unit characteristic data with critical infrastructure for post-hazard resilience modeling. *Sustainable and Resilient Infrastructure*, 1-17. https://doi.org/10.1080/23789689.2019.1681821

StataCorp. (2017). Stata Statistical Software: Release 15. College Station, TX: StataCorp LLC.

U.S. Census Bureau. (n.d.). Glossary. Retrieved from https://www.census.gov/glossary

U.S. Census Bureau. (2011). 2010 Census Summary File 1—(Oregon) [machine-readable data files]. Retrieved from https://www2.census.gov/census_2010/04-Summary_File_1/Oregon/

U.S. Census Bureau. (2012). TECHNICAL DOCUMENTATION: 2010 Census Summary File 1—Technical Documentation. Retrieved from https://www.census.gov/prod/cen2010/doc/sf1.pdf.

van De Lindt, J., Peacock, W., Mitrani-Reiser, J., Rosenheim, N., Deniz, D., Dillard, M., Tomiczek, T., Koliou, M., Graettinger, A., Crawford, P., Harrison, K., Barbosa, A., Tobin, J., Helgeson, J., Peek, L., Memari, M., Sutley, E., Hamideh, S., Gu, D., Cauffman, S. and Fung, J. (2018), The Lumberton, North Carolina Flood of 2016: A Community Resilience Focused Technical Investigation, Special Publication (NIST SP), National Institute of Standards and Technology, Gaithersburg, MD, [online], https://doi.org/10.6028/NIST.SP.1230

Van De Lindt, J., Peacock, W., Mitrani-Reiser, J., Rosenheim, N., Deniz, D., Dillard, M., Tomiczek, T., Koliou, M., Greattinger, A., Crawford, P., Harrison, K., Barbosa, A., Tobin, J., Helgeson, J., Peek, L., Memari, M., Sutley, E., Hamideh, S., Gu, D., Cauffman, S., & Fung, J. (2020). Community resilience-focused technical investigation of the 2016 Lumberton, North Carolina, flood: An interdisciplinary approach. *Natural Hazards Review*, 21(3), 04020029. https://doi.org/10.1061/(ASCE)NH.1527-6996.0000387

Wang, L., van de Lindt, J.W., Rosenheim, N., Cutler, H., Hartman, B., Lee, J.S., & Calderon D. (2021). Effect of Building Wind-Retrofit Strategies on Social and Economic Community-Level Resilience Metrics. *Journal of Infrastructure Systems*.