Knowledge-Informed Data-Driven Modeling of Coupled Human-Built–Natural Systems: The Case of Hurricane-Induced Debris

Catalina González-Dueñas[1]; Mitchell M. Meads[2]; Jamie E. Padgett, M.ASCE[3]; and Wesley E. Highfield[4]

Abstract: Debris is one of the most challenging cascading effects posed by hurricane events, causing large financial and logistical burdens to coastal communities. Moreover, disaster debris can have cascading consequences on the safety and functionality of infrastructure, inhibit community recovery efforts, and lead to public health concerns. However, existing debris predictive models have shown an unsatisfactory performance, with errors up to 50% in debris estimates, and only consider a limited set of predictive variables. Given the importance of debris in coastal community resilience, this study leveraged a convergent research strategy to propose a knowledge-informed data-driven methodology by developing a comprehensive database that expands across human-built–natural systems to inform a probabilistic data-driven model of debris volume. A Gaussian process model was used to generate the debris volume model from a debris removal database for Hurricane Ike and the human-built–natural systems predictors. Moreover, different spatial resolutions (500, 250, and 125 m) were tested to analyze their effect on the model performance. Results showed that the low-resolution (500 m) and the intermediate-resolution (250 m) models have the best performance with a normalized root-mean squared error (RMSE) of 0.49 and 0.50, respectively. These two models were then used to explore the relative variable importance of the predictive variables in the model in order to get insights on the drivers of the debris process and propose more flexible lower-dimensionality models. The influence of different predictors and the trade-offs of resolution and model performance were also discussed before demonstrating application of the model for a synthetic storm in the Galveston, Texas, region. The proposed methodology and the probabilistic estimates of debris quantities are key to develop comprehensive risk estimates of storm impacts on coastal communities and can support informed decision making and mitigation planning strategies. **DOI: 10.1061/NHREFO.NHENG-1705.**

Introduction

The debris generated in the aftermath of hurricane events is a prime example of cascading consequences that presents significant challenges for the resilience of coastal communities now and into the future. Considering cascading effects is key for the performance evaluation of the built and natural environment and the subsequent consequences on communities. These consequences can range from impaired accessibility to critical facilities due to damages to infrastructure systems caused by debris loading or accumulation (Gonzalez Duenas et al. 2019; Green et al. 2017) to health hazards due to improper disposal or prolonged exposure to toxic debris (Luther and Science Resources and Industry Division 2006; Reible et al. 2006). Moreover, debris collection and disposal activities can account for up to 30% of disaster recovery funds in the aftermath of a hurricane event (FEMA 2019), causing a large financial distress to coastal communities that struggle to recover from the impact of the storm.

The multiple hazards with potential to generate debris during a hurricane event and their interaction with a dynamic coastal landscape make hurricane-induced debris behavior a very complex phenomenon. This complexity derives from the interconnected nature of human-built–natural systems in coastal regions and the multihazard effects of the storm in these systems, which makes debris management strategies a challenging task. Thus, to properly model the debris generation and spreading process, it is necessary to understand the different factors involved, their interactions, and their effect on the process. However, physics-based debris modeling from coastal multihazard storms, and in particular waterborne debris modeling, is at present an emerging field (Kameshwar et al. 2021; Park and Cox 2019; Stolle et al. 2018), for which data-driven models have been used as a promising alternative to forecast debris accumulation and spreading in coastal regions (Marchesini et al. 2021).

For instance, Escobedo et al. (2009) proposed a regression model to estimate the amount of tree debris and damage from hurricane events leveraging data from the 2004 and 2005 hurricane season in Florida. The Hazards US (HAZUS) method (FEMA 2012) estimates hurricane-induced debris quantities using damage-related measures of the built environment and trees, making it one of the most widespread methodologies for debris management. Nevertheless, current debris predictive models have shown an unsatisfactory performance, with errors up to 30% to 50% on average in debris estimates (H-GAC 2011; USEPA 2008) and use a limited set of predictors, usually focusing on a specific characteristic (Marchesini et al. 2021) such as the type of debris [e.g., vegetative debris (Thompson et al. 2011)] or wind-field characteristics

[1]Graduate Research Assistant, Dept. of Civil and Environmental Engineering, Rice Univ., Houston, TX 77005. Email: cdg7@rice.edu

[2]Graduate Research Assistant, Dept. of Marine and Coastal Environmental Science, Texas A&M Univ. at Galveston, Galveston, TX 77553. Email: mitchell.meads@tamu.edu

[3]Stanley C. Moore Professor, Dept. of Civil and Environmental Engineering, Rice Univ., Houston, TX 77005 (corresponding author). ORCID: https://orcid.org/0000-0002-7484-2871. Email: jamie.padgett@rice.edu

[4]Professor, Dept. of Marine and Coastal Environmental Science, Texas A&M Univ. at Galveston, Galveston, TX 77553. Email: highfiew@tamug.edu

Note. This manuscript was submitted on July 18, 2022; approved on January 6, 2023; published online on March 7, 2023. Discussion period open until August 7, 2023; separate discussions must be submitted for individual papers. This paper is part of the ***Natural Hazards Review***, © ASCE, ISSN 1527-6988.

(Escobedo et al. 2009; FEMA 2012; Umpierre and Margoles 2005; USACE 2017), failing to capture the multidimensionality of hurricane-induced debris behavior.

Knowledge-informed data science strategies can be leveraged to select and interpret relevant features in the debris process using domain knowledge. Domain knowledge can be used to inform machine learning models in several ways, including model selection, regularization of the model, and feature engineering strategies (Hoffer et al. 2022). Feature engineering, in particular, leverages domain knowledge to define, select, and transform the predictive variables of the regression model to improve its predictive capacity and interpretability (Hoffer et al. 2022). Moreover, feature engineering strategies can also be used to perform feature selection, identifying important variables in the model that can be used to propose a more flexible model by reducing the number of predictors needed to inform it.

In this regard, convergent research approaches offer an ideal baseline for knowledge-informed data-driven models, by facilitating multidisciplinary knowledge sharing and coproduction, which can be directed not only to enhance model performance but also to define model characteristics needed to address a certain problem (Peek et al. 2020). In the case of debris management, projections of probabilistic estimates of debris can help stakeholders by supporting risk-informed decision making.

The objective of this study is to derive a knowledge-informed data-driven model that can capture the multidimensional aspects involved in the debris generation process and that can provide confidence levels on the debris estimations. As a proof of concept, a comprehensive database of potential factors affecting the debris process was collected in Galveston Island, Texas, under Hurricane Ike conditions. This database expands on multihazard storm and land-cover parameters (González-Dueñas et al. 2022) by including factors related to human-built–natural systems.

Moreover, composite predictors were also tested, such as the ones related to the fragility of the built environment, and second-order storm intensity parameters, such as momentum flux. The derived set of predictors leveraged physics-informed variables and engineering knowledge, setting the basis to identify potential drivers of the debris generation process. This comprehensive database was then used to develop probabilistic estimates of debris volume in the study area, leveraging a Gaussian process model (Rasmussen and Williams 2006). This machine learning technique provides not only the mean estimates of debris volume but also provides the confidence intervals of the prediction. Moreover, Gaussian processes maintain spatial correlations, which is of extreme importance while analyzing debris spread at a regional scale. The effects of spatial resolution were also investigated by performing a sensitivity analysis across three different grid cell sizes: 500, 250, and 125 m.

In the next section, the methodological approach of this study is presented, followed by the description of the different sets of predictors used to inform the regression model. Then, the debris volume predictive model is introduced, starting with the general definition of the Gaussian process model. In this section, the sensitivity of the model performance with respect to the spatial resolution is also explored, along with the analysis of the relative importance of the different predictors in the model. The variable importance analysis is then leveraged to propose more flexible models with a reduced set of predictors. An application example of the model in Galveston Island using present conditions (Galveston Central Appraisal District 2020) and a scenario 500-year storm is then introduced, followed by a discussion of the applicability of the model and potential policy implications. This paper finishes with a summary and general conclusions of the study.

Methodology

This study aims to identify potential drivers of the debris generation and spreading process and provide probabilistic estimates of debris quantities over coastal regions by modeling the coupled effect of human-built–natural systems. The methodology proposed in this study is presented in Fig. 1. The methodology starts by defining the area of study, which was divided into subregions using a grid in order to provide spatial-varying estimations of debris and facilitate geospatial analysis tasks. In this study, square grid cells were used to subdivide the region of analysis.

The second step of the methodology consists of the identification and characterization of relevant predictive features to effectively inform the debris regression model. A convergent research approach was followed to build a comprehensive database with parameters related to the multihazard storm effects and built, natural, and human systems, and to propose features at the intersection of these systems. To complete the database, a debris removal data set was used to inform the response variable—the debris volume. A detailed description of the debris volume database and the predictive parameters, their relevance, and characterization is presented in the following section.

Once the database was complete, geographic information system (GIS) software was used to couple the data in each cell of the grid. The preprocessing of the database was pursued in the third step of the methodology. In this step, feature selection techniques were leveraged to discard rows of data (each row represents a grid cell with their respective predictors and debris volume) that have missing information and features that do not present significant variation across the area (i.e., variables with low variance).

In the fourth step of the methodology, the processed database was used to train a Gaussian regression model. Gaussian processes are nonparametric models that are flexible enough to model different types of data, maintain spatial correlations, and have the advantage of providing probabilistic estimates of the response variable (Gelfand and Schliep 2016; Rasmussen and Williams 2006). With the trained model, the relative importance of each one of the predictors was computed and used to select and analyze the variables with a higher influence on the debris predictive model. Finally, these variables were used to train a lower-dimension model to facilitate its implementation and analyze the effect of feature selection on the performance of the model.

Knowledge-Informed Data Science: Modeling Human-Built–Natural Systems Interactions

In the debris process, the interactions of the storm with human-built–natural systems depend both on the spatial characteristics of the region and the time in which individual component or system fails. As the storm moves along the region, system components such as houses and trees will react to both the loads imposed by the storm at that time instant, as well as the loads of any component dragged by the storm (either by water or wind). These loads might be direct loads, as in the case of an impact (Gonzalez Duenas et al. 2019; Stolle et al. 2020), or indirect loads, such as the ones imposed on bridge foundations by accumulation of debris (i.e., damming loads) (Mauti et al. 2020).

Spatial characteristics such as the urban form of a region and the topography can also influence the loading and debris spreading process. For instance, during hurricane events, gaps between houses can become channels in which the water flows, augmenting the flow velocity and leading to problems such as scour in adjacent structures (FEMA 2009). The diversity and coverage of natural systems are also directly related to development patterns, which can ultimately

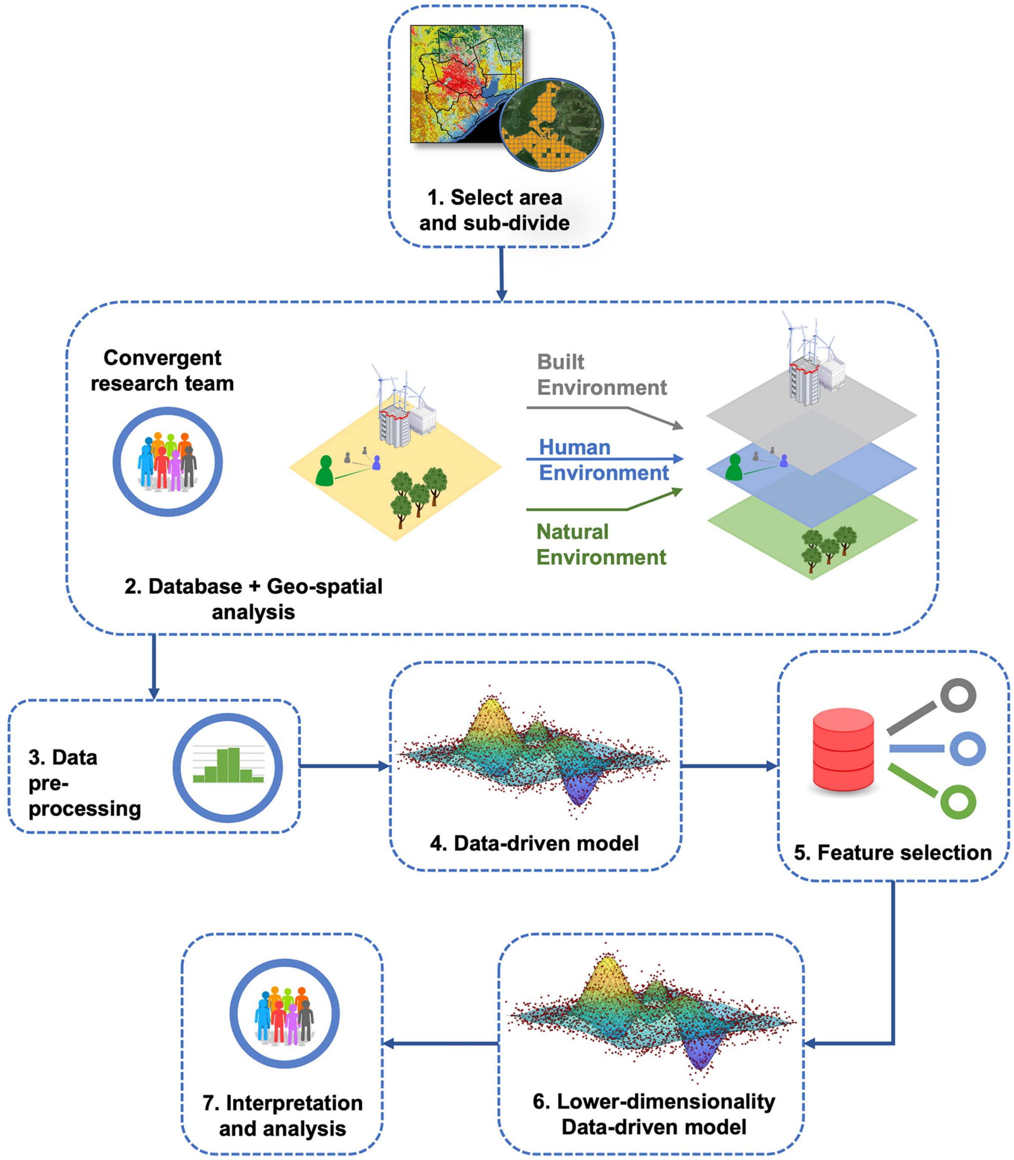

Fig. 1. Knowledge-informed data-driven methodology. (Base maps from Esri, DigitalGlobe, GeoEye, i-cubed, USDA FSA, USGS, AEX, Getmapping, Aerogrid, IGN, IGP, swisstopo, and the GIS User Community.)

influence storm debris presence and location (Beatley 2012). For example, woody debris from vegetated and forested areas can be transported through existing stream channels and result in large accumulations of natural environment–derived debris (West et al. 2011). Human systems parameters such as socioeconomic variables are also related to hazard damage patterns and are interdependent of built and natural system patterns (Santi et al. 2011). For instance, low-income and vulnerable populations may reside in homes that were constructed under older building code requirements or in mobile homes, which are at higher risk of suffering substantial damage in a storm event and potentially become direct sources of debris.

In this study, convergent engineering knowledge was leveraged to identify and characterize human-built–natural system's variables with potential effects on the debris generation and spreading process during hurricane events. Convergent engineering knowledge is defined herein as the process of knowledge coproduction that enables convergent research. This study adopted the definition of convergent research proposed by Peek et al. (2020), in which convergent studies are understood as a dynamic process in which experts from different disciplines come together to identify, characterize, and pose solutions and actions to unravel complex problems in the context of natural disasters. The different variables used in this study as predictors are given in Table 1.

Debris Collection Database

Hurricane Ike made landfall in Galveston Island, Texas, in 2008 and was predominately characterized as a storm surge and wave event, leading to a vast destruction in the Houston–Galveston area (FEMA 2009; Stearns and Padgett 2012). The storm also produced large quantities of debris in the region, and its removal accounted for approximately $752 million in recovery expenditures (FEMA 2019; González-Dueñas et al. 2022). In this study, a debris removal database from Hurricane Ike was used to inform the response variable of the model, the debris volume. The database, provided by the Houston-Galveston Area Council (H-GAC 2022) and the engineering firm Tetra Tech (2022), is part of the FEMA Public Assistance Project Worksheets (FEMA 2019). The worksheets are needed to receive support for debris removal activities from FEMA, and

Table 1. Human-built–natural environment predictive variables with the mean and coefficient of variation of the intermediate-resolution data (250 m grid cell size)

No.	Type of predictive variable	Variable description	Mean	COV (%)	Units
1	Multihazard storm parameters	Surge depth	2.7	52	m
2		Bathymetry	−0.01	23,867	m
3		Significant wave height	1.2	65	m
4		Wave period	8.0	74	s
5		Wave direction	249.8	37	Degrees
6		Water velocity in the x-direction	0.46	84	m/s
7		Water velocity in the y-direction	0.37	62	m/s
8		Wind speed	37.3	4	m/s
9		Wind direction	228.6	14	Degrees
10		Wind steadiness	0.17	38	—
11		Wind duration	3.2	39	h
12		Momentum flux	2.0	239	N/m^2
13	Built-systems parameters	Number of buildings	3.9	311	—
14		Total building footprint area	688.3	300	m^2
15		Number of accessory structures	12.6	298	—
16		Total area of accessory structures	397.7	304	m^2
17		Number of mobile homes	0.81	329	—
18		Weighted probability of failure 1	56.0	632	m^2
19		Weighted probability of failure 2	72.4	629	m^2
20		Development total	22.1	162	%
21		Development open	3.8	284	%
22		Development low	5.5	205	%
23		Development medium	7.7	222	%
24		Development high	5.0	280	%
25		Number of households	460.2	64	—
26		Number of housing units	1,021.3	73	—
27		Occupied housing units	460.2	64	—
28		Vacant housing units	561.1	94	—
29		Urban lag	0.22	141	%
30		Road density	0.0027	188	%
31		Average distance to the seawall	8,292.0	100	m
32		Average angle to the seawall	−5.2	1,208	Degrees
33	Natural-systems parameters	Open water	49.9	91	%
34		Barren land	4.8	342	%
35		Herbaceous	3.9	311	%
36		Woody wetlands	0.69	737	%
37		Emergent wetlands	18.5	163	%
38		Wetlands	19.2	162	%
39		Vegetation	23.3	149	%
40		Minimum elevation	−1.8	191	m
41		Maximum elevation	0.87	347	m
42		Average elevation	−0.52	567	m
43		Average distance to the shoreline	581.1	104	m
44		Average angle to the shoreline	21.4	487	Degrees
45		Vegetation lag	0.23	125	%
46		Wetland lag	0.19	131	%
47	Human-systems parameters	Total population	1,025.7	54	—
48		Population density	774.9	253	—
49		Median household income	42,555.9	47	$
50		Median family income	80,826.6	63	$
51		Average family income	74,839.3	45	$
52		Percentage renters	0.28	67	%

Note: COV = coefficient of variation.

contain information related to debris quantities, type, pick-up address, and disposal site.

To geocode the database, a data processing workflow was developed leveraging Google APIs, Jupyter notebooks, and the DesignSafe CI platform (Dukes 2019; Molina et al. 2019; Rathje et al. 2017). In order to facilitate the data aggregation process, Jupyter notebooks were also developed to automate geospatial analysis tasks (e.g., visualization, projection) in the GIS software. More details on the workflow have been given by Dukes (2019). After the preprocessing of the debris volume database, 24,685 unique debris pick-up geographic locations were identified in Galveston Island, along with their respective debris volume.

Multihazard Storm Parameters

Hurricanes are a combination of multiple hazards such as wind, storm surge, and heavy rainfall. These hazards create different loading conditions on the built and natural environments, interacting with them to create different types of debris. These debris are in turn

Nat. Hazards Rev., 2023, 24(2): 04023015

transported by the wind and/or storm surge, which leads to cascading damages and their spreading over the region. In this study, these effects were captured through the multihazard intensity parameters of the storm. The wind field intensity parameters were defined using risk management solutions (RMS) HWind data (Powell et al. 1998), and the storm surge intensity parameters (e.g., surge depth, wave height, and water velocity) were obtained from advanced circulation model (ADCIRC) + simulating waves nearshore (SWAN) simulations of Hurricane Ike (CGH 2017).

Moreover, the momentum flux, a composite intensity parameter, was also tested in the context of hurricane-induced debris. This intensity parameter is a function of the surge depth and flow velocity, and has been correlated to the hydrodynamic forces imposed on structures during tsunami events and subsequent damage (Charvet et al. 2017; Park et al. 2013; Song et al. 2017). However, its effect on hurricane-induced hydrodynamic loading and debris is yet to be explored. At the grid cell level, the intensity parameters were computed as the average value in the grid cell.

Built-Systems Parameters

Predictors associated with the built environment were included by considering the spatial patterns of the urban landscape, its characteristics, and its potential vulnerability to hurricane hazard. The percentages of developed areas (developed open, low, medium, high, and total) within each grid cell were calculated using the 2008 National Land Cover Data Set (NLCD) at a 30-m resolution (US Geological Service 2008). To capture potential shielding effects related to the built environment, the urban lag and the relative angle and distance from the centroid of the grid cell to Galveston Island's 12.47-km seawall were computed. Urban lag variables were calculated using a spatial weights matrix of adjacent cells' percentage of developed land cover (GeoDa 2018). These variables quantified proximity relationships between grid cells and other important aspects of the built environment.

Road networks are another important component of the built environment and help determine development patterns by providing connecting infrastructure. Therefore, the road density in each grid cell was calculated as the total length of roads divided by the total area of the grid cell. The road network spatial data sets were obtained from the Texas Department of Transportation (TxDOT 2022).

The likelihood of damage or total failure of different structures or structural components not only adds to the composition of the debris, but also leads the cascading failure of other systems impacted by the generated construction debris. Thus, to approximately capture the effect of the physical vulnerability of the built environment on debris behavior, two variations of the constructed weighted failure probability of a grid cell φ were proposed

$$\varphi_1 = \sum_{i=1}^{n} A_i \cdot P_{f_i} \tag{1}$$

$$\varphi_2 = \sum_{i=1}^{n} A_i \cdot P_{f_i} \cdot N_{f_i} \tag{2}$$

where A_i = footprint area of the ith building; P_{f_i} = structural probability of failure of the ith building; N_{f_i} = number of floors of the ith building; and n = total number of buildings in a grid cell.

To define these variables, a housing unit inventory developed for Galveston Island was used to characterize the built environment in the region (Fereshtehnejad et al. 2021). This database contains information on more than 14,000 housing units and features such as geographical coordinates, type of foundation, year of construction, and elevation of the house with respect to the ground. To incorporate the footprint area, the database was coupled with the 2018 building footprint data for Galveston Island (H-GAC 2018). The empirical fragility model (Variant 5) proposed by Tomiczek et al. (2014) for wood-framed residences in the Galveston area following Hurricane Ike was used in this study to predict the probability of failure of the building stock.

Moreover, to explore the influence of different built systems components on the debris process, variables related to the number of residential houses, the number of accessory structures (i.e., structures in a parcel that are not for living purposes), and their associated areas were computed at the grid cell level. Examples of accessory structures include greenhouses, detached garages, and storage sheds, among others. The number of mobile homes per grid cell was also included because they can become moving debris in the event of a storm or hurricane event. The information related to the accessory structures (number and respective area) and the number of mobile homes was obtained by spatially joining tax appraisal data of Galveston County to the building database (Galveston Central Appraisal District 2020). Only the houses and accessory structures constructed in or before 2008 were considered in this study.

To complete the built systems parameters within the Galveston Island landscape, the total number of households, housing units, occupied housing units, and vacant housing units were computed based on the 2009 American Community Survey (ACS) 5-year population estimates at the census block group resolution (US Census Bureau 2009). For each grid cell, the block group that covered the most area in the cell was assumed to represent its statistics.

Natural-Systems Parameters

Spatial patterns of the natural environment capture natural components within the landscape of the study area. Spatial coverage and diversity measurements were made for each grid cell as a percentage of coverage and were computed using the 2008 NLCD data (US Geological Service 2008). The percentage of coverage within each grid cell for each of the natural environment land-cover classes was quantified and then reviewed for sufficient statistical variation to include in the debris model. Due to a low variance, classifications of shrub/scrub, hay/pasture, cultivated crops, mixed forests, and deciduous forests were omitted. Aggregated sums of percentage coverage within each grid cell were also computed for all vegetative land-cover classes and all wetland land-cover classes. These aggregated sum values were included in an attempt to capture additive effects that the natural environment could potentially have on storm debris presence.

Furthermore, the vegetation lag and wetland lag variables were used to calculate averages of adjacent cells' vegetation and wetland coverages per grid cell using the same queen-based adjacency matrix methodology used to create the urban lag variables (GeoDa 2018). Beyond the quantification of natural environment land-cover classes, relative shoreline proximity variables and elevation statistics were included based on National Oceanic and Atmospheric Administration (NOAA) digital elevation models (NOAA 2007, 2022). Shoreline proximity variables were computed by measuring the relative angle and distance from the centroid of each grid cell to the nearest shoreline. The shoreline proximity, vegetation lag, and wetland lag variables were included to explore quantitative natural shielding measures. Finally, the minimum, maximum, and mean elevation values were calculated for each grid cell to include topographic effects.

Human-Systems Parameters

Whereas the built and natural environments may constitute the physical landscape of a coastal socioecological system, the human environment is inherently and directly affected by natural hazards and hurricane-induced debris. Furthermore, socioeconomic parameters of the human environment can influence the level of damages and impact from other natural hazards (Cutter et al. 2012). Therefore, a suite of relevant demographic and socioeconomic variables was included in the analysis (Table 1). The variables were collected at the census block group resolution from the 2009 ACS 5-year estimate, which represent the closest population statistics to our study's period (i.e., 2008) at the smallest spatial resolution available (US Census Bureau 2009).

Geospatial Analysis

With all the predictive variables defined, a geospatial analysis was conducted to aggregate and join the data based on its geographical location. As mentioned in the "Methodology" section, Galveston Island was divided into subregions using a grid to couple the different predictors and the response variable (i.e., debris volume). Three different models with varying spatial resolution were used in this study to analyze the effect of the spatial aggregation in the results: a low-resolution model with square grid cells of 500 m, an intermediate-resolution model with 250 m cells, and a high-resolution model with 125-m cells. For each model, the debris volume at each grid cell was computed as the total sum of the volume of all the debris pick-up points lying in the grid cell. The human-built–natural systems predictive variables at the grid cell level were estimated as described in the previous section.

Debris Volume Predictive Model

Gaussian Process Regression

In general, a Gaussian process can be defined as a collection of random variables whose joint probability density function follows a Gaussian distribution, and therefore is completely defined by its mean and covariance functions (Rasmussen and Williams 2006). For the general case of a regression model with noisy observations, we can define

$$y = f(\boldsymbol{x}) + \varepsilon \tag{3}$$

where y = observed target value (i.e., dependent variable); f = mapping function; $f(\cdot)$ = function value; $\boldsymbol{x}$ is the input vector (i.e., vector of explanatory variables); and $\varepsilon \sim N(0, \sigma^2_{\text{noise}})$ = zero-mean Gaussian noise with variance σ^2_{noise}. Thus, we can define a Gaussian process $f(\boldsymbol{x})$ as follows:

$$f(\boldsymbol{x}) \sim \text{GP}(m(\boldsymbol{x}), k(\boldsymbol{x}, \boldsymbol{x}')) \tag{4}$$

where $m(\boldsymbol{x})$ = mean function; and $k(\boldsymbol{x}, \boldsymbol{x}')$ = covariance function (i.e., covariance between $f(\boldsymbol{x})$ and $f(\boldsymbol{x}')$). For a training set D with n observations, $D = \{(\boldsymbol{x}_i, y_i) | i = 1, \ldots, n\}$, where $\boldsymbol{x}_i$ is the input vector of the ith observation (i.e., predictors) and y_i represents the dependent variable of the ith observation (e.g., one observation of debris volume), the prior distribution of the vector of dependent variables can be represented

$$\boldsymbol{y} \sim N(\boldsymbol{m}(X), K(X, X) + \sigma^2_{\text{noise}} I) \tag{5}$$

where $\boldsymbol{y}$ = vector of dependent variables; $X = [\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n]^{\text{T}}$ = matrix of explanatory variables; $K(X, X)$ is the $n \times n$ covariance matrix ($K_{i,j} = k(\boldsymbol{x}_i, \boldsymbol{x}_j)$); and I = identity matrix of size $n \times n$. Then, to make an inference on a test set X^*, a random Gaussian vector f^* is defined, and the prior joint distribution is computed

$$\begin{bmatrix} \boldsymbol{y} \\ \boldsymbol{f}^* \end{bmatrix} \sim \left(\begin{bmatrix} \boldsymbol{m}(X) \\ \boldsymbol{m}(X^*) \end{bmatrix}, \begin{bmatrix} K(X, X) + \sigma^2_{\text{noise}} I & K(X, X^*) \\ K(X^*, X) & K(X^*, X^*) \end{bmatrix} \right) \tag{6}$$

Finally, the prediction equations can be written (Rasmussen and Williams 2006)

$$\boldsymbol{f}^* | X, \boldsymbol{y}, X^* \sim N(\overline{\boldsymbol{f}^*}, \text{cov}(\boldsymbol{f}^*)) \tag{7}$$

$$\overline{\boldsymbol{f}^*} = K(X^*, X)[K(X, X) + \sigma^2_{\text{noise}} I]^{-1}(\boldsymbol{y} - \boldsymbol{m}(X)) + \boldsymbol{m}(X^*) \tag{8}$$

$$\text{cov}(\boldsymbol{f}^*) = K(X^*, X^*) - K(X^*, X)[K(X, X) + \sigma^2_{\text{noise}} I]^{-1} K(X, X^*) \tag{9}$$

More details on Gaussian process regression have been given by Rasmussen and Williams (2006). In Gaussian process regression, special consideration should be given to the selection of the mean and covariance functions. The mean function can be expressed as $\boldsymbol{m}(X) = \boldsymbol{h}(X)\boldsymbol{\beta}$, where $\boldsymbol{h}(X)$ is the basis function whose form depends on the type of mean function (zero-mean, constant mean, linear mean, or quadratic mean), and $\boldsymbol{\beta}$ is the respective coefficients. The mean function can make a model more interpretable and also has implications on predictions that are out of the ranges of the training data (DeltaIV 2017).

The covariance function $k(\boldsymbol{x}, \boldsymbol{x}')$, also known as the kernel function, is parameterized by a set of parameters $\boldsymbol{\theta}$ and determines the similarity between two inputs (Guo 2021). This is important, because a Gaussian process will give similar predictions for inputs that are similar, thereby maintaining spatial correlations (i.e., correlation of nearby points). Typical kernel functions include the exponential quadratic kernel, rational quadratic kernel, and periodic kernel, among others (Guo 2021; Jia et al. 2019; Rasmussen and Williams 2006). Alternatively, custom kernels can be formulated by combining different kernel functions. During the training phase, the parameters $\boldsymbol{\theta}$, $\boldsymbol{\beta}$, and σ^2_{noise} were optimized.

Debris Volume Predictive Model

The final data set comprised 52 predictors and 1 response variable denoting the total volume of debris per grid cell. From the 52 predictive variables, 12 were multihazard storm parameters, 16 built-environment parameters, 14 natural-environment parameters, and 10 human-environment parameters. There were in total 837 observations (i.e., grid cells) for the low-resolution model, 3,565 for the intermediate-resolution model, and 13,790 for the high-resolution model. In each case, the data were split into 50% training and 50% test data for the subsequent regression analysis. This was done to augment the test set and to try to avoid any geographical and data availability bias.

The python package GPflow version 2.7.0 (Matthews et al. 2017) was used to build the Gaussian process (GP) model. Different mean functions, including zero-mean, constant mean, linear mean, and quadratic mean were tested along with different kernel functions such as rational quadratic kernel, exponential kernel, radial basis function kernel, and Matérn kernel. The parameters $\boldsymbol{\theta}$, $\boldsymbol{\beta}$, and σ^2_{noise} were optimized by adopting the maximum likelihood estimation method. In GPflow the loss function (i.e., negative of the log marginal likelihood) is minimized using the limited-memory Broyden-Fletcher-Goldfarb-Shanno with bound constraints (L-BFGS-B) algorithm (Zhu et al. 1997). The normalized RMSE was used to compare the performance of the three varying-resolution models. The performance results and the optimum model definition for the three models are presented in Tables 2, S1, and S2, respectively. The rational quadratic kernel provided the best results for the high-resolution (125 m) and intermediate-resolution (250 m) models, whereas the exponential kernel was selected for the low-resolution

Nat. Hazards Rev., 2023, 24(2): 04023015

Table 2. Gaussian-process model performance in terms of normalized RMSE

No. of variables	Model resolution	Normalized RMSE
52	Low (500 m)	0.49
	Intermediate (250 m)	0.50
	High (125 m)	0.69
30	Low (500 m)	0.51
	Intermediate (250 m)	0.54

(500 m) model. The kernel parameters $\boldsymbol{\theta}_{\mathbf{Exp}}$ for the exponential kernel include the signal variance $\sigma^2_{\mathrm{signal}}$ and the separate length scales l_m of each predictor $m = 1, 2, \ldots, d$. The rational quadratic kernel parameters $\boldsymbol{\theta}_{\mathbf{RQ}}$ include the signal variance and separate length scales, as well as the positive-valued scalar mixture parameter α.

Results

The performances of the low- and intermediate-resolution models were comparable, with a percent difference of approximately 3%, whereas the high-resolution model showed a lower performance in terms of normalized RMSE when compared with the other two models. For instance, there was an approximate 43% and 38% increase in the normalized RMSE value of the high-resolution model with respect to the low- and intermediate-resolution models, respectively. Moreover, to analyze the spatial characteristics of the predictive model, Fig. 2 shows the comparison between the collected (i.e., actual debris volume) and predicted debris volumes in the test set for the three models. The models were able to capture the spatial distribution of debris in the region quite accurately when compared with the actual debris data. Moreover, the models were able to identify regions with high debris accumulation, which is key for poststorm response and for debris-mitigation plans. Also, the areas with the highest levels of debris volume in the region were constructed areas, which indicates a relationship between debris accumulation and urban development.

In addition to the spatially distributed debris estimations, estimates of the total debris volume are helpful when planning and budgeting disaster debris policies and removal contracts (Luther and Resources and Industry Division 2006). For this purpose, the percent error between the total predicted debris volume and the actual debris volume in the test data set was computed for the three models (Table 3). The high-resolution model presented the highest percent error (10%) with an increase of 9% and 10% on average when compared with the intermediate- and low-resolution models, respectively. The intermediate- and low-resolution models showed the best predictive performance and had a very good agreement of estimated total debris volume in the test data set, showing the promise of implementing the proposed models when planning hurricane-induced debris removal strategies in coastal regions.

Variable Importance

In order to gain insights on the influence of the different predictive variables in the debris generation and spreading process, automatic relevance determination (ARD) was used to perform feature selection by evaluating the relative importance of the variables in the regression model (Paananen et al. 2020). The relative importance of each predictive variable was computed as the inverse of the length scale l_m. A low value of $1/l_m$ for the dth feature implies that the prediction of the target value (i.e., debris volume) does not vary significantly when moving across the dth dimension. This allowed us to rank the different variables in terms of their predictive capacity and select the ones that contribute the most to the regression model using a predetermined threshold.

The classification of the variables by predictive power also allows feature selection to be implemented by proposing alternative models with a reduced number of predictors. Models with a lower dimensionality are not only more interpretable, but also easier to implement for other scenarios or in other regions. The intermediate- and low-resolution models were selected for further analysis and leveraged to propose more flexible models using feature selection strategies because they presented a better performance. In this study, a qualitative threshold of 30 variables was adopted for both models to promote the flexibility of its application while maintaining the sufficient range of information needed for the good performance of the regression model. In the future, a one-by-one analysis will be performed to find the optimum number of parameters based on their relative importance.

Figs. 3 and 4 show the relative importance of the predictive variables in the intermediate-resolution (250 m) and low-resolution (500 m) models, respectively. In general, the built-systems parameters showed the highest influence in the intermediate-resolution model when compared with the other sets of parameters, whereas different multihazard storm, built- and natural-environment parameters showed relevance in the low-resolution model. In the low-resolution model, features that capture general characteristics of the built environment, such as the total development and the total area of accessory structures, had a relatively higher influence in the model, whereas at the intermediate-resolution of 250 m, more specific structure-dependent features such as the weighted probability of failure and the number of buildings and accessory structures started to show importance in the regression model.

The developed-land categories of total development and developed open space (less than 20% impervious) showed importance in both models, and the urban lag—a parameter that quantifies urban land coverage in adjacent grid cells—showed a high relative importance in the intermediate-resolution model. In both models, the surge, wave height, and water velocity appeared as important parameters, emphasizing the significance of considering the multihazard effects of the storm when analyzing debris behavior in coastal regions. It is key to highlight the importance of second-order features such as the damage-related measures and the momentum flux in the intermediate-resolution model.

The weighted probability of failure appeared as a one of the variables with the highest relative importance in the model, showing the link between structural behavior and capacity in the debris-generation process. Moreover, the momentum flux, which has been observed to have an influence on the damage during tsunami events, also showed importance within the multihazard storm parameters, demonstrating the relevance of considering coupled storm effects on debris behavior studies.

Within the natural-environment parameters, an important similarity between the two models was the shared importance of open water as a predictor of debris volume. This finding suggests that water can be a method of transport for debris from a storm, especially when the storm has significant storm surge. Natural systems variables displayed less relative importance on debris volume in the intermediate-resolution model when compared with the low-resolution model. A potential reasoning for this discrepancy could be related to the spatial resolution of the natural-environment data sets. The larger grid cells in the low-resolution model allowed for more spatial variation in vegetative land coverages within each grid and provided a more robust exploration of the relationship between debris volume and the natural environment. Also, the relative importance of the distance/angle to the seawall and shoreline in both models suggests

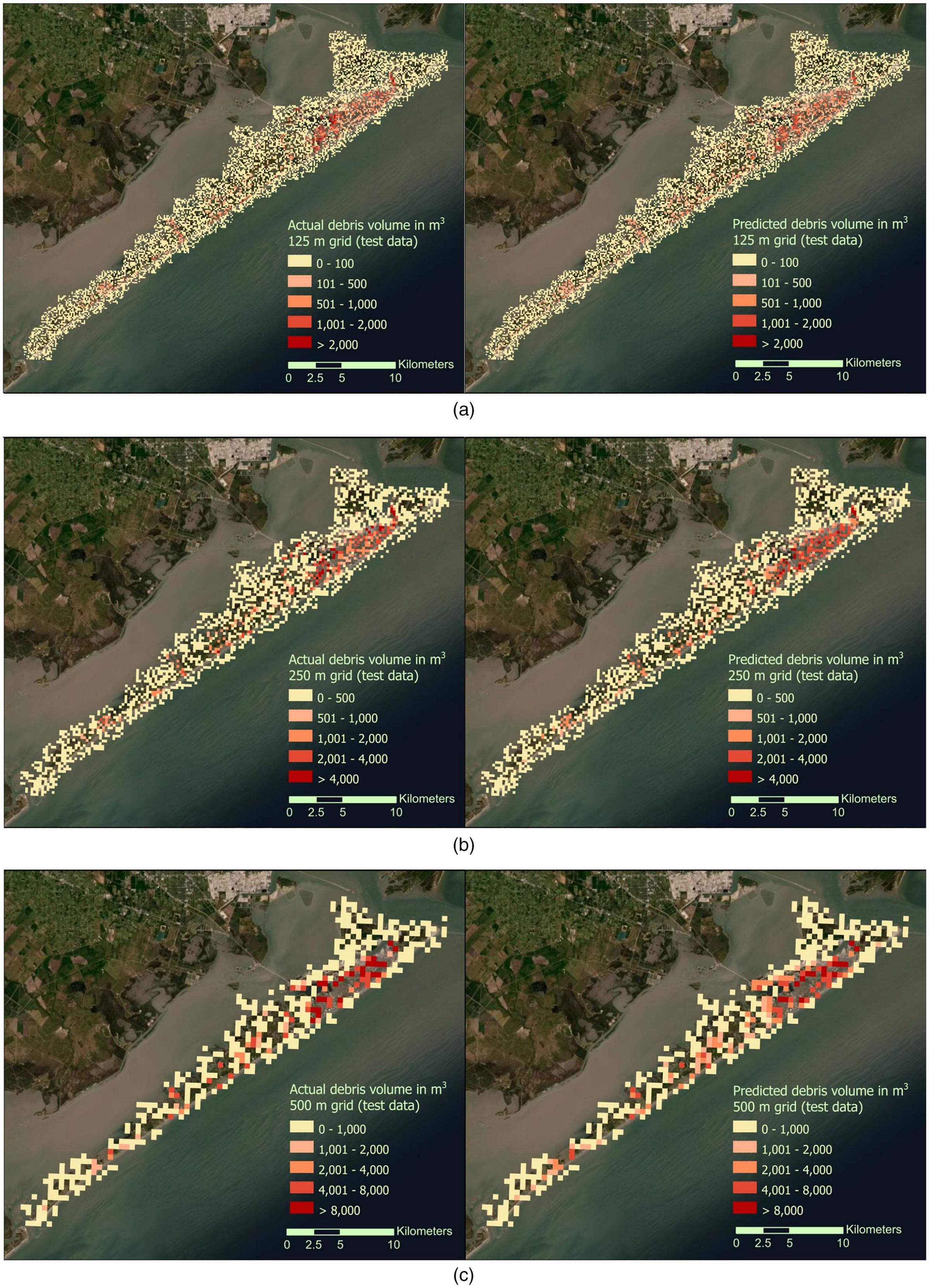

Fig. 2. Comparison of the actual debris volume collected in the aftermath of Hurricane Ike with the predicted debris volume at a spatial resolution of (a) 125 m (high-resolution model); (b) 250 m (intermediate-resolution model); and (c) 500 m (low-resolution model). (Base maps from Esri, DigitalGlobe, GeoEye, i-cubed, USDA FSA, USGS, AEX, Getmapping, Aerogrid, IGN, IGP, swisstopo, and the GIS User Community.)

Table 3. Percent error in the aggregated debris volume estimation for the test data

No. of variables	Model resolution	Error (%)
52	Low (500 m)	0.23
	Intermediate (250 m)	1.39
	High (125 m)	10.11
30	Low (500 m)	2.96
	Intermediate (250 m)	0.74

an important relationship between hurricane-induced debris and potential shielding parameters.

The human-systems parameters did not consistently appear as important variables in the models, possibly due to data resolution obstacles and a relative lack of spatial variation. However, population density, which also correlates with the importance trend of development and highly constructed areas, appeared as an important variable in the two models. Moreover, the percentage of renter-owned households and median household income displayed

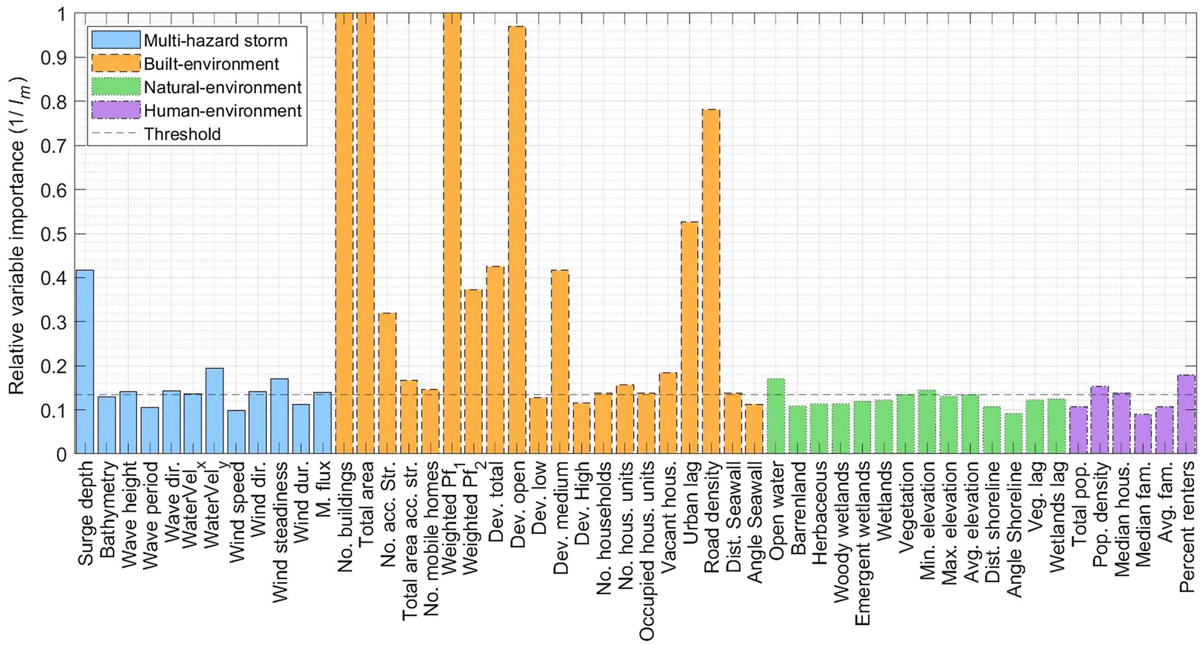

Fig. 3. Relative variable importance for the intermediate-resolution (250 m) model.

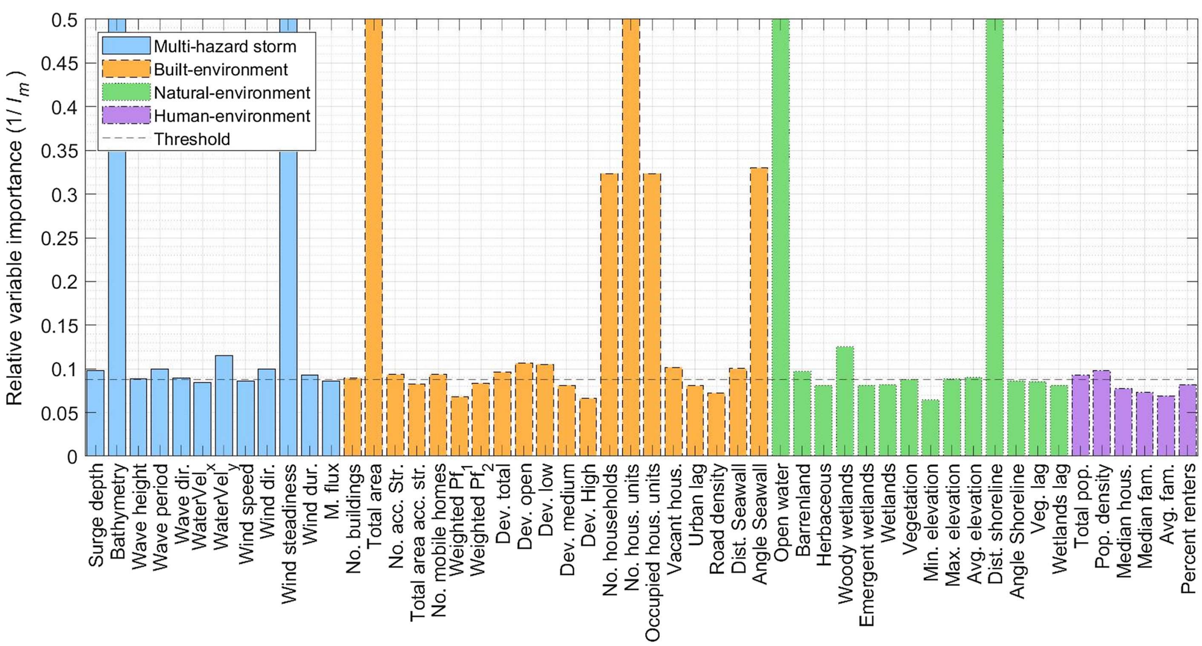

Fig. 4. Relative variable importance for the low-resolution (500 m) model.

importance in the intermediate-resolution model. One potential explanation could be that renters and lower-income population are more likely to reside in structures of older construction built under less effective building codes, ultimately resulting in more debris.

Feature Selection

As previously mentioned, the 30 most significant variables were selected to propose two new lower-dimensionality models at the intermediate and low resolutions. Table S3 presents the selected variables for each model. In order to compare the impact of excluding features in the model, the performance of the lower-dimensionality models was compared with the original models (i.e., models with the full set of predictors). Both models maintained a high accuracy (less than 3% error) when estimating the total volume of debris in the region (Table 3). Comparisons of the intermediate-resolution model with the HAZUS model (USACE 2017) showed an improvement of approximately 81% when compared with the actual debris removed in Galveston Island in the aftermath of Hurricane Ike.

When analyzing the change in performance, the low-resolution resolution model performed better than the intermediate-resolution one (Table 2). There was an approximate change in performance of 6% and 8% for the low- and intermediate-resolution models, respectively, with respect to the models with the full set of predictors. This shows that removing some of the features does not significantly hinder the performance of the regression model. Tables S1 and S4 provide the parameters of the lower-dimensionality models.

In this regard, it is important to analyze the potential trade-offs among model resolution, resources available (e.g., data, computational resources, and time), and performance requirements, as well as their implication in policy development. When proposing debris management and predisaster debris planning strategies, it is important for local governments and stakeholders to consider the availability of data for the region and the key factors driving the decision process. For instance, if one of the main objectives is to reduce any connectivity issues between the local community and critical facilities in the aftermath of the storm, a higher resolution model will provide more information on the state of arterial roads. On the other hand, if overall estimates of debris are required to determine if existing landfills provide the necessary infrastructure to collect the debris generated by the storm or if temporary disposal sites are necessary, a compromise can be made between spatial resolution and the confidence on the estimates.

Moreover, to enhance the resilience of coastal communities exposed to hurricane events, it is necessary to consider the couple coupled effects of human development patterns and land-use policies on storm induced debris. As seen in the variable importance analysis of the intermediate- and low-resolution models, different types of urban development have an influence on predicted debris volume. Thus, the effect of development patterns on debris volume should be considered when proposing future land-use policies. Moreover, given the influence of the weighted probability of failure and the number of mobile homes, it is key to consider the retrofitting of non-code-compliant structures and guidelines for mobile-home owners during storm events. The relationship of variables related to housing quality with the socioeconomic characteristics of the region should be also considered in debris management strategies.

It is also important to promote the acquisition and digitalization of data in underresourced and rural areas. The lack of good-quality data can jeopardize the ability of these communities to prepare and respond to natural disasters, for which platforms and policies that support data acquisition and free access to it in the context of natural hazards research should be promoted (Rathje et al. 2017).

Model Application

To showcase how the proposed models can be implemented, the lower-dimensionality model (i.e., model with 30 variables) at the intermediate resolution (250 m) was selected to predict the expected amount of debris in Galveston Island under present conditions when exposed to a scenario storm. To inform the multihazard storm features, ADCIRC+SWAN simulations of Storm FEMA 36 (SSPEED Center 2022) were used to estimate the surge depth, wave height and direction, flow velocity, momentum flux, and wind field characteristics. Storm FEMA36 is a probabilistic storm that approximately produces still-water elevation equivalent to a 500-year return period storm in the Houston-Galveston region (Ebersole et al. 2015).

The wind steadiness was computed using the hourly wind velocity estimates from the numerical simulation of the storm leveraging the procedure proposed by Berkovic (2018). The wind direction was obtained as the angle between the two velocity components of the wind field. Finally, the built-environment features were estimated using the totality of the building-stock data (i.e., not filtered for 2008 conditions) as described in the "Built Systems Parameters" section (Fereshtehnejad et al. 2021; Galveston Central Appraisal District 2020). Urban land cover and urban lag were collected using updated 2020 NLCD and calculated following the same workflow described in the "Built Systems Parameters" section (US Geological Service 2020). The road density was also computed as described in the "Built-Systems Parameters" section. Finally, the built-systems parameters associated with demographic data (i.e., number of households, housing units, and vacant and occupied houses) were collected from the most recent complete ACS 5-year estimate data (US Census Bureau 2019).

The natural-systems and human-systems parameters were computed as described in the "Natural-Systems Parameters" and "Human-Systems Parameters" sections using complete and updated data sets (GeoDa 2018; NOAA 2018; US Census Bureau 2019; US Geological Service 2020).

Fig. 5 shows the mean debris estimates in Galveston Island and the associated standard deviation of the prediction for the present conditions when subjected to FEMA36. Due to the influence of built-systems parameters in the model, highly developed and constructed areas in Galveston Island showed the highest debris accumulation. Comparatively, these areas also exhibited a larger value of standard deviation with respect to other regions; however, their magnitude was very low (less than 90 m^3) when considering that the predictive mean volumes are in the range of the thousands. Moreover, the large ranges of debris volume can be attributed to the more intense effects of the storm and the inclusion of more buildings in the model.

Beyond these spatial distributions, the total debris volume for the island was predicted to be 1.9×10^6 m^3—a 59% increase in debris relative to what was produced in Hurricane Ike. Such a predictive model—with the ability to provide uncertainty on the estimates—can be used to test the impact of alternative storm scenarios or development policies in the region.

Conclusions

This paper proposed a knowledge-informed data-driven methodology to explore the drivers of the debris process and assess debris accumulation and spreading in coastal regions exposed to hurricane and storm events. The methodology leverages domain knowledge to study the influence of human-built–natural systems and their interactions on debris behavior and makes use of these insights to develop a probabilistic data-driven model of debris volume.

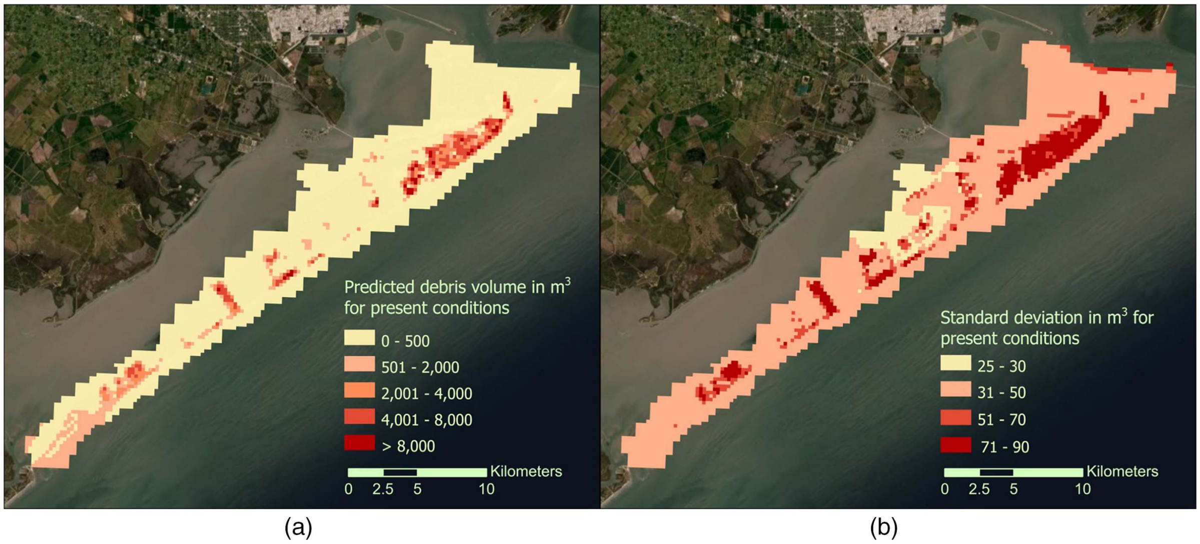

Fig. 5. (a) Mean debris volume prediction; and (b) associated uncertainty (standard deviation) for present conditions in Galveston Island, Texas. (Base maps from Esri, DigitalGlobe, GeoEye, i-cubed, USDA FSA, USGS, AEX, Getmapping, Aerogrid, IGN, IGP, swisstopo, and the GIS User Community.)

A comprehensive database of relevant human-built–natural systems features was developed for Galveston Island, Texas, under Hurricane Ike conditions to inform the predictive debris volume model.

Moreover, a data-processing framework was developed to inform debris volume using a debris removal database from Hurricane Ike. Gaussian process regression was leveraged to develop the debris predictive model due to its ability to capture spatial correlations and provide uncertainty in debris estimates, setting the basis for informed decision making and policy development on resilient debris management. Probabilistic estimates of debris quantities are key to develop risk estimates, support informed decision-making and mitigation planning strategies, as well as to evaluate different scenarios and potential consequences of debris accumulation on coastal communities. The effect of spatial resolution on the debris estimates was also investigated by proposing three different models at the resolutions of 500 m (low resolution), 250 m (intermediate resolution), and 125 m (high resolution).

Results showed that the model performance was better at the low and intermediate resolutions with a normalized RMSE of 0.49 and 0.59, respectively. Comparisons of the performance of the models in estimating the total debris volume in Galveston Island were also presented, with errors of less than 2% for the low- and intermediate-resolution models, and a percent error of approximately 10% for the high-resolution model. These models improve upon existing models, which can add up to errors of 50% in debris estimation and include a limited set of predictors, failing to capture the complex interactions on multihazard storm with human-built–natural systems on coastal regions.

Due to the better performance of the intermediate- and low-resolution models, they were leveraged to analyze the relative importance of the different predictors in the model. The built-systems parameters showed a high influence on the intermediate-resolution model, especially the ones related to damage-dependent predictors and urban features. In the case of the low-resolution model, parameters that captured the general built and natural characteristics of the region showed a higher influence, along with the multihazard storm parameters. The human-systems parameters showed a lower relative importance in the models when compared with the storm, built-systems, and natural-systems variables. The variables with a higher relative importance were then used to propose two lower dimensionality models at the 500- and 250-m scales using feature selection strategies. Results showed that reducing the number of variables does not hinder the performance of the predictive model significantly, with changes in the RMSE in the range of 8%.

To showcase how the proposed models can be implemented, an example of the expected debris accumulations and distributions in Galveston Island under present conditions and a 500-year return period storm were presented. The intermediate-resolution (250 m) model was selected to lay the foundation for future debris cascading impacts studies, such as the loss of connectivity in the aftermath of a storm due to road closures attributed to debris accumulation. Even though models with different spatial resolutions can show similar predictive performance, the selection of the model depends on multiple factors including the availability of data in the region, the computational resources, the time span given for the database collection and model development, as well as the end goal of the model implementation. In this regard, there is a need to improve data acquisition, data digitalization, and automation prior to the onset of natural disasters, especially in communities with low resources.

This study also showcased how convergent studies can help to develop strategies and methodologies that can give a holistic and comprehensive understanding of natural disasters, focusing on the risk drivers and the potential requirements and modeling techniques needed for resilient and inclusive decision making. Future studies should aim to include data from different storms and from different regions in order to improve the predictive capabilities of the model and its reliability. Moreover, given the flexibility of Gaussian process regression, opportunities also exist to develop customized kernel functions that can model different trends in the data by combining standard kernel functions or adding constraints to them. Ongoing work is investigating the coupled effects of a changing climate, human development patterns, and policies on storm-induced debris, and the cascading effects and consequences of hurricane-induced debris on distributed infrastructure systems and vulnerable populations.

Data Availability Statement

Some data used during the study were provided by a third party (wind field intensity parameters and the debris removal database). Direct requests for these materials may be made to the provider as indicated in the Acknowledgments. All the other data (human-built–natural system parameter database), models, and codes that support the findings of this study are available from the corresponding author upon reasonable request.

Acknowledgments

The authors gratefully acknowledge the support of this research by the National Science Foundation under Awards OISE-1545837 and CMMI-2002522. In addition, the second author is supported in part by the NIST Center of Excellence for Risk-Based Community Resilience Planning under Cooperative Agreement 70NANB15H044 between the National Institute of Standards and Technology (NIST) and Colorado State University. Any opinions, findings, and conclusions, or recommendations expressed in this paper are those of the authors and do not necessarily reflect the views of the sponsors. The authors give a special acknowledgement to Dr. Genevera Allen, who gave important guidance on the development of the models and their interpretation. In addition, Miku Fukatsu and Henry Ulrich, undergraduate students doing internships in Padgett's Research group, are also gratefully acknowledged for their support in the preprocessing of the data used in this study. Rice University's Fondren Library GIS center is acknowledged for their support in identifying the data used in the testbed analysis, along with the resources of the NHERI DesignSafe Cyberinfrastructure CMMI-2022469. The authors also acknowledge Dr. Kayode Atoba for providing the data associated to the HAZUS debris model. Finally, the authors gratefully acknowledge RMS for providing the HWind data used to inform the wind field parameters and Tetra Tech for providing the debris removal database. Maps throughout this paper were created using ArcGIS-software by Esri. ArcGIS and ArcMap are the intellectual property of Esri.

Supplemental Materials

Tables S1–S4 are available online in the ASCE Library (www.ascelibrary.org).

References

Beatley, T. 2012. *Planning for coastal resilience: Best practices for calamitous times*. Washington, DC: Island Press.

Berkovic, S. 2018. "Wind regimes and their relation to synoptic variables using self-organizing maps." *Adv. Sci. Res.* 15 (Jan): 1–9. https://doi.org/10.5194/asr-15-1-2018.

CGH (Computational Hydraulics Group at the University of Texas at Austin). 2017. "The computational hydraulics group." Accessed April 3, 2022. https://chg.oden.utexas.edu/.

Charvet, I., J. Macabuag, and T. Rossetto. 2017. "Estimating tsunami-induced building damage through fragility functions: Critical review and research needs." *Front. Built Environ.* 3 (Aug): 36. https://doi.org/10.3389/fbuil.2017.00036/full.

Cutter, S. L., B. J. Boruff, and W. Lynn Shirley. 2012. "Social vulnerability to environmental hazards." In *Hazards vulnerability and environmental justice*, 143–160. London: Routledge.

DeltaIV. 2017. "Why is the mean function in Gaussian process uninteresting?" Accessed July 1, 2022. https://stats.stackexchange.com/q/222304.

Dukes, B. 2019. "Undergraduate Research Experience (REU), NHERI 2019: A data processing and visualization framework for hurricane debris modeling."Accessed September 1, 2022. https://search.datacite.org/works/10.17603/ds2-jt8d-sx39.

Ebersole, B. A., T. C. Massey, J. A. Melby, N. C. Nadal-Caraballo, D. L. Hendon, T. W. Richardson, and R. W. Whalin. 2015. *Interim report—Ike dike concept for reducing hurricane storm surge in the Houston-Galveston region*. Jackson, MS: Jackson State Univ.

Escobedo, F. J., C. J. Luley, J. Bond, C. Staudhammer, and C. Bartel. 2009. "Hurricane debris and damage assessment for Florida urban forests." *Arboriculture Urban For.* 35 (2): 100–106. https://doi.org/10.48044/jauf.2009.018.

FEMA. 2009. *Hurricane Ike in Texas and Louisiana: Mitigation Assessment Team report, building performance observations, recommendations, and technical guidance*. FEMA P-757. Washington, DC: FEMA.

FEMA. 2012. *HAZUS–MH 2.1 Hurricane model technical manual*. Washington, DC: FEMA.

FEMA. 2019. *FEMA public assistance funded projects detail—Open government initiative*. Washington, DC: FEMA.

Fereshtehnejad, E., I. Gidaris, N. Rosenheim, T. Tomiczek, J. E. Padgett, D. T. Cox, S. Van Zandt, and W. Gillis Peacock. 2021. "Probabilistic risk assessment of coupled natural-physical-social systems: Cascading impact of hurricane-induced damages to civil infrastructure in Galveston, Texas." *Nat. Hazards Rev.* 22 (3): 04021013. https://doi.org/10.1061/(ASCE)NH.1527-6996.0000459.

Galveston Central Appraisal District. 2020. "2020 Galveston CAD Roll." Data Set. Accessed September 1, 2020. http://www.galvestoncad.org/index.php/Appraisal_Exports.

Gelfand, A. E., and E. M. Schliep. 2016. "Spatial statistics and Gaussian processes: A beautiful marriage." *Spatial Stat.* 18 (Nov): 86–104. https://doi.org/10.1016/j.spasta.2016.03.006.

GeoDa. 2018. "Applications of spatial weights." Accessed February 1, 2022. https://geodacenter.github.io/workbook/4d_weights_applications/lab4d.html.

Gonzalez Duenas, C., C. Bernier, and J. Padgett. 2019. "Probabilistic assessment of bridges subjected to waterborne debris." In *Proc., Coastal Structures Conf.*, 356–365. Karlsruhe, Germany: Bundesanstalt für Wasserbau. https://doi.org/10.18451/978-3-939230-64-9_036.

González-Dueñas, C., C. Bernier, and J. E. Padgett. 2022. "A data-driven approach to hurricane debris modeling." In *Proc., IMAC-XXXVIII Conf.* Bethel, CT: Society for Experimental Mechanics.

Green, D., D. Yu, I. Pattison, R. Wilby, L. Bosher, R. Patel, P. Thompson, K. Trowell, J. Draycon, and M. Halse. 2017. "City-scale accessibility of emergency responders operating during flood events." *Nat. Hazards Earth Syst. Sci.* 17 (1): 1–16. https://doi.org/10.5194/nhess-17-1-2017.

Guo, S. 2021. "Implement a Gaussian process from scratch." Accessed July 1, 2022. https://towardsdatascience.com/implement-a-gaussian-process-from-scratch-2a074a470bce.

H-GAC (Houston-Galveston Area Council). 2011. *Regional storm debris management assessment*. Houston: H-GAC.

H-GAC (Houston-Galveston Area Council). 2018. "LiDAR building footprints 2018." Accessed November 1, 2020. http://www.h-gac.com/gis-applications-and-data/datasets.aspx.

H-GAC (Houston-Galveston Area Council). 2022. "Houston-Galveston Area Council (H-GAC)." Accessed May 1, 2019. https://www.h-gac.com/Home.

Hoffer, J. G., A. B. Ofner, F. M. Rohrhofer, M. Lovrić, R. Kern, S. Lindstaedt, and B. C. Geiger. 2022. "Theory-inspired machine learning—Towards a synergy between knowledge and data." *Weld. World* 66 (7): 1291–1304. https://doi.org/10.1007/s40194-022-01270-z.

Jia, B., X. C. Jianzhong Zhou, Z. He, and H. Qin. 2019. "Deriving operating rules of hydropower reservoirs using Gaussian process regression." *IEEE Access* 7 (Oct): 158170–158182. https://doi.org/10.1109/ACCESS.2019.2948760.

Kameshwar, S., H. Park, D. T. Cox, and A. R. Barbosa. 2021. "Effect of disaster debris, floodwater pooling duration, and bridge damage on immediate post-tsunami connectivity." *Int. J. Disaster Risk Reduct.* 56 (Feb): 102119. https://doi.org/10.1016/j.ijdrr.2021.102119.

Luther, L. G., and Science Resources and Industry Division. 2006. *Disaster debris removal after Hurricane Katrina: Status and associated issues*. Washington, DC: Congressional Research Service, Library of Congress.

Marchesini, G., H. Beraud, and B. Barroca. 2021. "Quantification of disaster waste: Review of the available methods." *Int. J. Disaster Risk Reduct.* 53 (Feb): 101996. https://doi.org/10.1016/j.ijdrr.2020.101996.

Matthews, A. G., M. Van Der Wilk, T. Nickson, K. Fujii, A. Boukouvalas, P. León-Villagrá, Z. Ghahramani, and J. Hensman. 2017. "GPflow: A Gaussian process library using tensorflow." *J. Mach. Learn. Res.* 18 (40): 1–6.

Mauti, G., T. Takabatake, J. Stolle, I. Nistor, N. Goseberg, and A. Mohammadian. 2020. "Experimental investigation of loading due to debris dams on structures." *J. Hydraul. Eng.* 146 (5): 04020029. https://doi.org/10.1061/(ASCE)HY.1943-7900.0001731.

Molina, R., C. Gonzalez, and J. Padgett. 2019. *Undergraduate Research Experience (REU), NHERI 2019: A data processing framework for the advancement of hurricane debris modeling*. Seattle: DesignSafe-CI. https://doi.org/10.17603/ds2-tzhh-7w14 v1.

NOAA (National Oceanic and Atmospheric Administration). 2007. "Galveston, Texas, coastal digital elevation model." Accessed February 1, 2022. https://www.ncei.noaa.gov/metadata/geoportal/rest/metadata/item/gov.noaa.ngdc.mgg.dem:403/html#.

NOAA (National Oceanic and Atmospheric Administration). 2018. "2018 TPWD Lidar: Coastal Texas." Accessed February 1, 2022. https://coast.noaa.gov/dataviewer/#/lidar/search/.

NOAA (National Oceanic and Atmospheric Administration). 2022. "National shoreline vector data." Accessed February 1, 2022. https://www.ngs.noaa.gov/NSDE/.

Paananen, T., J. Piironen, M. Riis Andersen, and A. Vehtari. 2020. "Variable selection for Gaussian processes via sensitivity analysis of the posterior predictive distribution." In *Proc., AISTATS 2019—22nd Int. Conf. on Artificial Intelligence and Statistics*, 1743–1752. Maastricht, The Netherlands: ML Research Press.

Park, H., and D. T. Cox. 2019. "Effects of advection on predicting construction debris for vulnerability assessment under multi-hazard earthquake and tsunami." *Coastal Eng.* 153 (Sep): 103541. https://doi.org/10.1016/j.coastaleng.2019.103541.

Park, H., D. T. Cox, P. J. Lynett, D. M. Wiebe, and S. Shin. 2013. "Tsunami inundation modeling in constructed environments: A physical and numerical comparison of free-surface elevation, velocity, and momentum flux." *Coastal Eng.* 79 (Sep): 9–21. https://doi.org/10.1016/j.coastaleng.2013.04.002.

Peek, L., J. Tobin, R. M. Adams, H. Wu, and M. Clay Mathews. 2020. "A framework for convergence research in the hazards and disaster field: The natural hazards engineering research infrastructure CONVERGE facility." *Front. Built Environ.* 6 (Jul): 110.

Powell, M. D., S. H. Houston, L. R. Amat, and N. Morisseau-Leroy. 1998. "The HRD real-time hurricane wind analysis system." *J. Wind Eng. Ind. Aerodyn.* 77 (Sep): 53–64. https://doi.org/10.1016/S0167-6105(98)00131-7.

Rasmussen, C. E., and C. K. Williams. 2006. *Gaussian processes for machine learning*. Cambridge, MA: MIT Press.

Rathje, E. M., et al. 2017. "DesignSafe: New cyberinfrastructure for natural hazards engineering." *Nat. Hazards Rev.* 18 (3): 1–7. https://doi.org/10.1061/(ASCE)NH.1527-6996.0000246.

Reible, D. D., C. N. Haas, J. H. Pardue, and W. J. Walsh. 2006. "Toxic and contaminant concerns generated by Hurricane Katrina." *J. Environ. Eng.* 132 (6): 565–566. https://doi.org/10.1061/(ASCE)0733-9372(2006)132:6(565).

Santi, P. M., K. Hewitt, D. F. VanDine, and E. Barillas Cruz. 2011. "Debris-flow impact, vulnerability, and response." *Nat. Hazard.* 56 (1): 371–402. https://doi.org/10.1007/s11069-010-9576-8.

Song, J., R. De Risi, and K. Goda. 2017. "Influence of flow velocity on tsunami loss estimation." *Geosciences* 7 (4): 114. https://doi.org/10.3390/geosciences7040114.

SSPEED Center. 2022. "Severe storm prediction, education, & evacuation from disasters center." Accessed July 1, 2022. https://www.sspeed.rice.edu/publications.

Stearns, M., and J. E. Padgett. 2012. "Impact of 2008 Hurricane Ike on bridge infrastructure in the Houston/Galveston region." *J. Perform. Constr. Facil.* 26 (4): 441–452. https://doi.org/10.1061/(ASCE)CF.1943-5509.0000213.

Stolle, J., N. Goseberg, I. Nistor, and E. Petriu. 2018. "Probabilistic investigation and risk assessment of debris transport in extreme hydrodynamic conditions." *J. Waterw. Port Coastal Ocean Eng.* 144 (1): 04017039. https://doi.org/10.1061/(ASCE)WW.1943-5460.0000428.

Stolle, J., I. Nistor, N. Goseberg, and E. Petriu. 2020. "Multiple debris impact loads in extreme hydrodynamic conditions." *J. Waterw. Port Coastal Ocean Eng.* 146 (2): 04019038. https://doi.org/10.1061/(ASCE)WW.1943-5460.0000546.

Tetra Tech. 2022. "About." Accessed May 1, 2019. https://www.tetratech.com/en/about.

Thompson, B. K., F. J. Escobedo, C. L. Staudhammer, C. J. Matyas, and Y. Qiu. 2011. "Modeling hurricane-caused urban forest debris in Houston, Texas." *Landscape Urban Plann.* 101 (3): 286–297. https://doi.org/10.1016/j.landurbplan.2011.02.034.

Tomiczek, T., A. Kennedy, and S. Rogers. 2014. "Collapse limit state fragilities of wood-framed residences from storm surge and waves during Hurricane Ike." *J. Waterw. Port Coastal Ocean Eng.* 140 (1): 43–55. https://doi.org/10.1061/(ASCE)WW.1943-5460.0000212.

TxDOT (Texas DOT). 2022. "TxDOT roadway inventory." Accessed February 1, 2022. https://gis-txdot.opendata.arcgis.com/datasets/TXDOT::txdot-roadway-inventory/explore?location=31.057868%2C-100.081515%2C6.50.

Umpierre, D., and G. Margoles. 2005. "Broward county's web-based hurricane debris estimation tool (HurDET)." In *Proc., 2005 ESRI Int. User Conf.* Redlands, CA: ESRI.

USACE. 2017. "Disaster impact models." Accessed June 1, 2022. http://www.usace.army.mil/Missions/Emergency-Operations/Disaster-Impact-Models/.

US Census Bureau. 2009. "American community survey 5-year estimates: Population (Table DP05)." Accessed February 1, 2022. https://data.census.gov/cedsci/table?q=DP05&g=0100000US&y=2009&tid=ACSDP5Y2009.DP05&hidePreview=true.

US Census Bureau. 2019. "American community survey 5-year estimates: Population (Table DP05)." Accessed February 1, 2022. https://data.census.gov/cedsci/table?q=DP05&g=0100000US&y=2019&tid=ACSDP5Y2019.DP05&hidePreview=true.

USEPA. 2008. *Planning for natural disaster debris*. Washington, DC: USEPA.

US Geological Service. 2008. "National land cover dataset 2008." Accessed February 1, 2022. https://www.mrlc.gov/data.

US Geological Service. 2020. "National land cover dataset 2020." Accessed February 1, 2022. https://www.mrlc.gov/data.

West, A. J., C. W. Lin, T. C. Lin, R. G. Hilton, S. H. Liu, C. T. Chang, K. C. Lin, A. Galy, R. B. Sparkes, and N. Hovius. 2011. "Mobilization and transport of coarse woody debris to the oceans triggered by an extreme tropical storm." *Limnol. Oceanogr.* 56 (1): 77–85. https://doi.org/10.4319/lo.2011.56.1.0077.

Zhu, C., R. H. Byrd, P. Lu, and J. Nocedal. 1997. "Algorithm 778: L-BFGS-B: Fortran subroutines for large-scale bound-constrained optimization." *ACM Trans. Math. Software* 23 (4): 550–560. https://doi.org/10.1145/279232.279236.