ARTICLE

Uncertainty Reduction Through Data Management in the Development, Validation, Calibration, and Operation of a Hurricane Vulnerability Model

Jean-Paul Pinelli¹ · Josemar Da Cruz¹ · Kurtis Gurley² · Andres Santiago Paleo-Torres² · Mohammad Baradaranshoraka² · Steven Cocke³ · Dongwook Shin³

Accepted: 22 October 2020 / Published online: 20 November 2020
© The Author(s) 2020

Abstract Catastrophe models estimate risk at the intersection of hazard, exposure, and vulnerability. Each of these areas requires diverse sources of data, which are very often incomplete, inconsistent, or missing altogether. The poor quality of the data is a source of epistemic uncertainty, which affects the vulnerability models as well as the output of the catastrophe models. This article identifies the different sources of epistemic uncertainty in the data, and elaborates on strategies to reduce this uncertainty, in particular through identification, augmentation, and integration of the different types of data. The challenges are illustrated through the Florida Public Hurricane Loss Model (FPHLM), which estimates insured losses on residential buildings caused by hurricane events in Florida. To define the input exposure, and for model development, calibration, and validation purposes, the FPHLM teams accessed three main sources of data: county tax appraiser databases, National Flood Insurance Protection (NFIP) portfolios, and wind insurance portfolios. The data from these different sources were reformatted and processed, and the insurance databases were separately cross-referenced at the county level with tax appraiser databases. The FPHLM hazard teams assigned estimates of natural hazard intensity measure to each insurance claim. These efforts produced an integrated and more complete set of building

descriptors for each policy in the NFIP and wind portfolios. The article describes the impact of these uncertainty reductions on the development and validation of the vulnerability models, and suggests avenues for data improvement. Lessons learned should be of interest to professionals involved in disaster risk assessment and management.

Keywords Catastrophe modeling · Data augmentation and integration · Florida Public Hurricane Loss Model (FPHLM) · Uncertainty reduction · Vulnerability model

1 Introduction

Catastrophe (cat) models for man-made infrastructure have four main components: a hazard component, which models the hazards, for example, hurricane or earthquake; an exposure model, which categorizes the exposure, for example, buildings, into generic classes; a vulnerability component, which models the effects of the hazard on the exposure and defines vulnerability functions for each building (or other type of exposure) class; and an actuarial component, which combines the vulnerability, the hazard, and the exposure, to quantify the risk in terms of physical damage, economic damage, or insured losses. Cat models address the primary needs of different user groups. One group includes the insurance industry and insurance regulators (Dong 2002; Shah et al. 2018). In this case, insurance portfolios are the input to the models, and the outputs are insured losses. Most of the cat models addressing the needs of the insurance industry are proprietary models from companies such as Risk Management Solutions (2019) and others. A notable exception is the Florida Public Hurricane Loss Model (FPHLM 2019). The other user group of cat

✉ Jean-Paul Pinelli
pinelli@fit.edu

¹ College of Engineering and Science, Florida Institute of Technology, Melbourne, FL 32901, USA

² Civil and Coastal Engineering, University of Florida, Gainesville, FL 32611, USA

³ Center for Ocean-Atmospheric Prediction Studies (COAPS), Florida State University, Tallahassee, FL 32306, USA

models includes economists (Michel-Kerjan et al. 2013), as well as disaster managers and city and emergency planners (Chian 2016; Biasi et al. 2017), where the focus is on emergency planning, post-disaster recovery, and increasingly on resilience studies (Muir-Wood and Stander 2016). In this case, the input can be databases of building or other infrastructure exposure from tax rolls or other sources, and the outputs are physical or monetary damage. In many cases, these models are in the public domain and can even be open source code. Some notable examples include MH-HAZUS (Schneider and Schauer 2006; FEMA 2015), and ShakeCast (Wald et al. 2008; Harvey et al. 2018).

These probabilistic natural hazard catastrophe (cat) models are the product of complex simulations, sometimes coupled with regression analyses, and involve many stochastic variables and sources of data. Figure 1 represents the interplay between the different types of data in a cat model, circled in red. For an insurance risk model, an insurance company will provide exposure data, which is the input to the cat model. The cat model filters the insurance exposure through its library of building classes, assigns vulnerability functions to each building class identified in the exposure, and applies hazard data, to produce the insured losses. In addition to the exposure data, an insurance company will also have claim data from past catastrophic events. Other noninsurance data (for example, topographic, geographic information system, and tax roll) can augment both exposure and claim data. The claim data, combined with hazard data, provide the basis for the validation of the output of the risk model (that is, the loss estimates). Current state-of-the-art vulnerability components of a cat model are probabilistic engineering models where the vulnerability functions result from Monte Carlo simulations on numerical models (Pita et al. 2014). In this case, the modelers need to validate the model vulnerability curves against empirical vulnerability curves, obtained through regression analyses of the enhanced claim data. Engineers can also derive vulnerability functions through direct or indirect regression analysis of the enhanced claim data versus hazard intensity (Pita et al. 2013), while semi-engineering models will generate vulnerability functions through a combination of engineering models and regression analyses. These validation and development efforts require multiple forms of information including: (1) actuarial exposure and claim data, such as values, deductibles, insured limits, loss value, and cause of loss; (2) building exposure and claim data, such as location, elevation, age, and building characteristics; and (3) hazard data, such as date and type of event, and more importantly, hazard intensity.

The same processes are at play independently of the type of hazard, or whether the cat model is used for insurance purposes, emergency management, or resilience

studies, or whether the exposure is residential, commercial, industrial, or a combination of these. It is clear that the uncertainty in the output of a cat model is highly dependent on the quality of the data (Khanduri and Morrow 2003), and that the uncertainty in the vulnerability model will depend on the quality of the data used for development or validation (Roueche et al. 2018). Consequently, cat models have high levels of uncertainty attached to their results (Kaczmarek et al. 2018). This uncertainty is a combination of aleatory and epistemic uncertainty, where the aleatory uncertainty is the result of natural variability in parameters like structural component strength capacities, and the epistemic uncertainty is due to lack of knowledge (Der Kiureghian and Ditlevsen 2009). Roueche et al. (2018) presented an exhaustive analysis of epistemic uncertainty for tornado damage fragility models, and some of their findings extend to other types of catastrophe models. A challenging aspect of cat modeling and operation is that the exposure and claim data (for example, building data in an insurance portfolio) are very often incomplete, inconsistent, absent, or erroneous. Also the hazard intensity data necessary to inform the claim data frequently are not available. Risk modelers must then combine different sources of data to enhance the quality of the exposure and claim data, not only to improve the development and validation processes, but also to improve the quality of the input data into the risk model and to reduce the resulting output uncertainty. These different datasets are typically very large, come from various sources with different formats, and might have different levels of quality.

This article is concerned with epistemic uncertainty in both the natural hazard catastrophe model output and in the vulnerability component of the model. One source of uncertainty is the lack of knowledge in the input data, while a second source is the lack of knowledge in the claim data used for development and validation of the model and in the hazard intensity measures associated with the claim data. Different strategies to reduce that uncertainty are presented and discussed. In particular, this article describes: the methodology for augmentation and integration of exposure and claim data with new and diverse information; and the impact of these enhanced data on the uncertainties attached to vulnerability model development, validation, and calibration and cat model output. The article is centered around the Florida Public Hurricane Loss Model as a case study, but the findings are applicable to other cat models, and other types of hazards as well.

Fig. 1 Schematic diagram of a catastrophe model and the interaction between data sources. Data sources for the catastrophe (cat) model are circled in red

2 Florida Public Hurricane Loss Model

The Florida Office of Insurance Regulation (FLOIR) sponsored the development of the Florida Public Hurricane Loss Model (FPHLM) as a tool for insurance regulation in the State of Florida (Hamid et al. 2011; FPHLM 2019). Its purpose is to estimate the potential insured losses in the State of Florida caused by hurricane events, including coastal flood (storm surge), inland flood, wind, and rain-water ingress.

2.1 Overview

The FPHLM can analyze insurance portfolios of single-family homes, including manufactured homes, as well as commercial residential buildings (either apartment or condominium buildings) subject to hurricane hazard. The purpose of the model is to predict aggregated insured losses for residential properties in the form of annual expected losses and probable maximum losses. Insurance companies and state regulators use such loss estimates to define and evaluate rate filings. The model can also conduct scenario analyses to estimate losses for hypothetical events and historical storms.

The FPHLM differs from similar commercial proprietary models in that the science behind the model is publicly available through numerous peer-reviewed publications, many of which this article references. The Florida Commission on Hurricane Loss Projection Methodology (FCHLPM) has consistently certified the FPHLM, with the latest certification for FPHLM version 7.0 released in 2019. The public is able to download the

corresponding submission reports from the FCHLPM web site (FPHLM 2019). The FPHLM helps the FLOIR regulate the insurance industry in Florida, decide on the appropriateness of insurance premiums requests filed by the insurance industry, and verify the solvency of insurance companies through actuarial stress tests (Nicholson et al. 2018). Similarly, insurance and re-insurance companies can use the FPHLM to quantify their hurricane risk exposure in Florida.

2.2 Florida Public Hurricane Loss Model Vulnerability Models

The FPHLM has vulnerability models in the form of vulnerability matrices and curves for different building classes and for different hazards (wind or flood). The cells of a vulnerability matrix for a particular building class represent the probability of a given damage ratio occurring at a given hazard intensity measure (IM) of wind speed or flood height. The columns of the matrix are probability mass functions of damage ratio given an IM. The vulnerability curves are simply the mean values of the probability mass functions as a function of the hazard intensity. The FPHLM includes, for each building class, vulnerability functions for: building damage, which includes the damage to both the exterior and interior of the building; contents damage, where contents are anything inside the building but not attached to the building (for example, furniture, rugs, appliances, and so on); and time related expenses. The damage ratio is the cost of damage divided by the replacement value of the building or contents, where the replacement value is the cost to replace the property with a

new item of like kind and quality (Baradaranshoraka et al. 2019). In the case of time related expenses, the ratio is the time related expense divided by the maximum policy limit (Pinelli et al. 2011).

The vulnerability functions for the combined wind and rain hazard are the product of an engineering approach, which models the effect of the hazard on numerical models of each building class through Monte Carlo simulations (Pita et al. 2012; Johnson et al. 2018). Under the sponsorship of FLOIR, the FPHLM team has recently expanded the previously hurricane wind and rain-only scope of the FPHLM to include coastal and inland flood hazards. The team's strategy was to adapt the large body of tsunami-related building fragility curves, especially the work of Suppasri et al. (2013), to coastal flood, and to adapt the work of the U.S. Army Corp of Engineers (USACE 2006, 2015) for inland flood through a semi-engineering approach (Baradaranshoraka et al. 2017, 2019). Regression techniques using the flood claim data are the basis for the development of the flood contents vulnerability curves, as described later in this paper. Paramount to these modeling efforts is the validation and calibration of the vulnerability curves against insurance claim data.

3 Input Data for the Florida Public Hurricane Loss Model

There are about 6.33 million households in Florida. For historical and political reasons, the wind and flood perils are insured separately in the United States by the private sector for the case of wind, and by the federal government for the case of flood. About 97% of homeowners have wind insurance, which translates into about 6 million policies (risk) in the Florida Hurricane Catastrophe Fund exposure data. In 2018, Florida had 1.757 million NFIP flood policies in force for USD 440 billion exposure. Of the 6.33 million Florida households, about 28% had flood insurance. Flood insurance is only mandatory for homes in high risk locations with 1% or more chance of annual flooding. It is not required in low to moderate risk areas. Because of the different insurers, the wind and flood models of the FPHLM are independent separate models, each with their own input:

- The exposure portfolio file of the National Flood Insurance Program (NFIP) for the flood model.
- The exposure portfolio files of private insurers for the wind model.

Exposure files include all the policies insured by a given company. The datasets provided by these sources are briefly summarized below.

3.1 National Flood Insurance Program Exposure Database

The FLOIR provided the NFIP exposure portfolios to the FPHLM team. The NFIP exposure portfolios for Florida properties were made available for each year from 1992 to 2012, and include 22 million policies, which overlap from year to year. Each year's exposure file contains policy number, flood zone, address, original construction date, base flood elevation, and other data, but no structural building information.

3.2 Wind Insurance Exposure Portfolios

The wind exposure portfolios vary in the details provided, but generally include policy ID, zip code, county, construction type (frame or masonry), and year built. A few companies provide more details such as roof shape, number of stories, roof cover, and opening protection. The FPHLM engineering team has access to wind insurance exposure datasets from 23 different wind insurance companies, which provide the data to the FPHLM for processing. These data cover different years from 2011 to 2019, in three years intervals, with each interval including some 13.5 million policies, or roughly 4.5 million policies per year. The latitude and longitude values and/or addresses are provided starting with the 2012 exposure files in about 80% of the cases. The precise locations of the buildings in the exposure portfolios make it possible to relate the exposure information to the NFIP or tax appraiser databases, which is the grand challenge for integrating these datasets.

4 Data for Development and Validation of the Florida Public Hurricane Loss Model

The FPHLM team used two primary sources of claim data for development and validation purposes:

- The NFIP claims database for the case of the flood model.
- Twenty-three private wind insurance companies' claim files for the wind model. The claim files include only the policies that suffered a loss for a particular event. The next sections briefly summarize the datasets provided by these sources.

4.1 National Flood Insurance Program Claims Database

The FLOIR provided the NFIP claim portfolios to the FPHLM team. The claims database contains 153,751 claims between July 1975 and January 2014 for 126

different events. The hazard team analyzed the claim data locations and dates to associate a specific hazard to a given claim. From these datasets, they chose a dataset of 58,551 personal residential claims from the 12 storms with the most claims. Table 1 lists these storms, with the number of claims per storm, while Fig. 2 shows the tracks of the 12 storms, together with the distribution of the coastal flood claims. The coastal flood and wave team provided hazard information for all 12 storms, but the inland flood team was able to provide hazard outputs only for the post 1999 storms (Irene, Charlie, Francis, Jeanne, Dennis, Wilma, Ivan, and Katrina), which correspond to a subset of 43,552 personal residential claims. The NFIP claim files contain information such as the date of loss, policy number, physical address, cause of damage, total property value, financial damage to building and contents, and replacement cost. The property value in the claim files is more reliable than in the exposure files, since it is updated at the time of the claim. Fields are present in the files for structural information such as exterior wall type and foundation type, but do not contain values for 97% of the claims.

4.2 Wind Insurance Claims Portfolios

The wind insurance claim datasets represent 23 different wind insurance companies and nine hurricane events. There are 667,573 claim records among all the insurance portfolios and hurricane events. The data contained in the claim files vary by company and event, but generally include the policy ID, postal code, county, year built, construction type, and loss to structure, contents, appurtenant structures, and time related losses (additional living

expenses). The latitude and longitude values and/or addresses are in the claim data files for some of the 2004 and 2005 hurricanes only. The precise locations of the buildings make it possible to identify the location of these claim records, based on the policy ID.

5 Sources of Uncertainty in the Exposure and Claim Data

There are multiple sources of uncertainty in the exposure data used for input, and the claim data used for development and validation of the FPHLM. The subsections below identify these sources.

5.1 Uncertainty in Building Characteristics

For each insurance exposure portfolio, the different policies need to be mapped to the FPHLM building classes and corresponding vulnerability functions. Building vulnerability functions exist for various building classes, based on construction type, roof shape, roof cover, number of stories, opening protection, and for the year built in the case of wind hazard, and based on construction type, elevation, number of stories, and for year built (that is, strength) in the case of flood hazard. In many cases, in the insurance portfolios, detailed records of building characteristics are missing. Portfolio files have information on construction type, location, and year built, but roof shape, roof cover, number of stories, opening protection, and elevation are in general undefined. This is clearly a source of epistemic uncertainty, which makes the mapping of each portfolio policy to the right available vulnerability function challenging.

A study of the Florida building population (Michalski 2016) yielded the statistical distributions of number of stories, roof shapes, roof cover, and opening protection by age per region. The FPHLM team designed a mapping tool to read a policy and randomly assign unknown building characteristics, based on the building population statistics conditional on year built, where the year built serves as a proxy for the strength of the building. Once all the unknown parameters in the policy have been statistically assigned, a vulnerability function based on the corresponding combination of parameters can then be assigned. If the number of unknown parameters exceeds a certain threshold defined by the user, a weighted matrix is used instead. For each age group and region, the FPHLM defines a weighted vulnerability function for each construction type. The weighted functions are the sum of the corresponding vulnerability model functions weighted on the basis of the statistical distribution of their defining parameters (FPHLM 2019).

Table 1 Hurricane events with occurrence date currently included in the claims trial dataset from the National Flood Insurance Program (NFIP)

Storm name	Storm year	Number of claims
Andrew	1992	2525
Charley	2004	1317
Dennis	2005	3092
Elena	1985	5522
Frances	2004	3639
Georges	1998	3375
Irene	1999	12,215
Ivan	2004	7022
Jeanne	2004	2775
Katrina	2005	4872
Opal	1995	5422
Wilma	2005	6775
Total		58,551

Fig. 2 Tracks for the 12 storms with the most National Flood Insurance Program (NFIP) claims data in Florida, 1975–2014

These two approaches present problems. The random assignment of a parameter, when unknown, to each policy in a portfolio will reflect the distribution of the parameter for a sufficiently large portfolio, and in that case different runs of the model will converge to the same aggregated loss result. The individual results for each policy in the portfolio, however, might differ from run to run. The assignment of a weighted vulnerability overcomes these problems, but on the other hand overrides available information when only partial information is missing. Finally, the transformation of the damage into an insured loss, between deductible and limit, can introduce a nonlinearity in the actuarial model, such that the superposition principle might not apply: the sum of the individual policy losses, each with missing parameters randomly assigned, might not converge to the sum of the losses where all the policies are assigned weighted vulnerabilities.

The same issues affect the claim portfolios, and the corresponding development and validation efforts. In this case, the lack of data on building characteristics compromises the grouping of the claims corresponding to different building classes. The lack of granularity of the data results in regression models that might not be representative of the building classes.

5.2 Uncertainty in Building Location

Building location in the exposure and claim datasets might be missing. This is a source of epistemic uncertainty, and

results from several causes. Insurance data are proprietary and confidential, and some insurance companies are reluctant to release the exact location of the claims, and will provide only the postal code of the property. In this case, the modeler assumes that the property is located at the population centroid of the postal code zone. The uncertainty resulting from this problem can be evaluated by comparing results of regression models using the same data with exact address and with centroid location.

In other cases the address might be spelled incorrectly, the latitude or longitude coordinates might be incorrect, or the geocoding engine does not have the necessary precision. It is not uncommon for this to arise occasionally in portfolios of several hundred thousand properties. Finally, in some instances, the address provided by the insurance company is the address of the owner of the property instead of the address of the property itself. This uncertainty is more difficult to evaluate due to its arbitrariness. If the analysts detect an inconsistency in the address or location that they cannot resolve, they ask the insurance company to resolve the inconsistency.

5.3 Uncertainty in Property Value

The replacement property value (that is, building value and contents value) is a key parameter in cat model data. Vulnerability functions are nondimensional, and the insured monetary loss for a property is the damage ratio times the replacement value. This value has both aleatory

and epistemic uncertainty. Even if all the data were available, there is a subjective component in the value of a home, and a natural variability in its price. Variability is even greater in the case of contents. These variabilities are minimal compared to the epistemic uncertainty, however, due to lack of proper knowledge regarding the actual value of a property in an insurance portfolio. Traditionally, modelers use the insurance limits (also known as coverage) for building and contents as a proxy for their replacement values. In the case of private wind insurance, where insurers want to avoid under- or over-insurance, this assumption is reasonable for a building. Contents limit is usually 50% of the building limit, according to standard actuarial practice.

In the case of the NFIP, which is a subsidized government program not strictly subjected to actuarial accountability, the insured limit might not be a realistic representation of the true value of a building, since properties are many times underinsured. In some cases, NFIP provides an estimate of the property value alongside the coverage limit. The problem is that this property value is defined at the time of the inception of the insurance contract, and might not be updated over the years. As a result, in general, there are no reliable replacement value data provided in the exposure data, for either buildings or contents, and the coverage limits in the NFIP policies are not a true measure of the value of a building or its contents.

The impact of the uncertainty on the replacement value of building and contents is two-fold. It affects the final insured loss estimates (for example, probable maximum losses and annual expected losses). It also affects the quantification of the damage ratios from the claim data needed for the development, validation, and calibration processes.

5.4 Uncertainty in Claim Adjustment Value

The claim adjustment process is complex and involves large uncertainties. Different insurance companies have different adjustment guidelines, and within one company different individual adjusters can interpret the guidelines differently and can have different levels of experience. In the case of a large disaster involving large numbers of damaged structures, the adjusters do not have the time and resources to make detailed analyses. Due to these factors, there is a level of uncertainty, largely epistemic, in the values reported in the claim data. This uncertainty can be evaluated if a sufficiently large number of claims for similar types of buildings subjected to a similar hazard intensity is available. Reduction of this uncertainty involves training of the adjusters, cross-validation of the adjusted values between different adjusters, and assignment of sufficient resources to the adjustment process.

5.5 Uncertainty in Cause of Damage

The NFIP claims data have several fields that contain information that can be used to associate each claim to a flood hazard event. Among these fields are the catastrophe number, cause of loss, date of loss, and property location. The information provided by these fields is not always complete or accurate, and often does not contain sufficient detail concerning the flood hazard event. For example, the catastrophe number is often missing, or multiple hazard events may be assigned the same number if they occur around the same time. The cause of loss often appears to be incorrect based on other information. There are cases, for example, where the cause of loss was listed as “Tidal Water Overflow,” even though the property was too far from the coast to be affected by tidal water. The claims data do not provide important details of the hazard, such as the flood elevation or wave conditions.

For properties located in flood-prone areas, the damage might be a combination of flood water and wind, in which case the separation between the two causes of damage might be problematic, and the NFIP claims might be contaminated by wind damage, and vice versa the wind claims might include some damage due to flood. Barad-ranshoraka et al. (2017) provided guidance on how to separate wind and water damage. The uncertainty in the cause of damage affects the development and validation processes of the vulnerability models.

5.6 Uncertainty in Hazard Intensity Measurement

This is one of the most significant sources of uncertainty. In many cases, particularly for a wind and storm surge event, very limited data are available regarding the intensity of the hazard (for example, wind speed and direction, accumulated rainfall, surge and wave heights), and no measurement of the hazard intensity is available for every property in the claim data. To make up for the lack of data, the hazard intensity measurements from the few observation points, if available, are extrapolated to the claim locations. Field measurements and other sources of data (ground radar stations, aerial drone, satellites) can provide validation and calibration of numerical models of hazard intensity, which are then used to derive hazard intensity values at the claim locations.

6 Uncertainty Reduction Through Augmentation and Integration of the Insurance Databases

One way to reduce the epistemic uncertainties is to augment the quality of the exposure and claim datasets. Missing information in the insurance databases might be

found in other sources of data. In particular, the state tax rolls can provide substantial information on building characteristics, location, and value. Data from a variety of sources can provide information on cause of damage and hazard intensity.

6.1 Tax Appraiser Databases

The FPHLM team collected tax appraiser (TA) datasets for 64 out of 67 Florida counties comprising approximately 99% of the state's population. The TA databases vary in size from 457,492 properties for large urban counties like Broward, to 4584 properties for small rural counties like Madison. An ideal TA dataset would include all building properties and components that are vulnerable to hurricane damage: interior and exterior wall composition, roof shape, roof cover, floor cover, year built, year of retrofit (if any), and number of stories. In addition, building location, elevation, footprint area, and the number of apartment units per building are desired.

Finally, the TA databases include several estimates of the value of each property for tax purposes. However, these values include the value of the land, except in some counties where the building-only value is also provided, whereas, the vulnerability model is concerned only with the value of the building.

The TA databases consist of data tables, typically spread across several files, linked through unique parcel identification numbers, which hold the building attribute information, and the geographic information system (GIS) shapefiles, which typically contain the polygons defining the geographic boundaries of each parcel.

6.2 Sources of Uncertainty in the Tax Appraiser Databases

There is no uniform format or content standard across the different county TA databases, and the amount and quality of data collected differs from county to county. This is far from an ideal situation, and is the consequence of the different level of resources available to the TA, from county to county, and the administrative decentralization typical in Florida. The lack of a uniform nomenclature and format across the different counties adds to the uncertainty, since it leads to possible errors of interpretation of the data. In particular, most of the building characteristics in the TA databases (as well as in many insurance portfolios) relate to fire classification rather than wind resistance.

In an extreme case like the Miami-Dade TA database, which contains virtually no building information useful for the purposes of FPHLM calibration and validation, apart from location, year built, and value, there is no augmentation of the data possible. The modelers in such cases rely

on exposure statistics from neighboring counties with a similar exposure mix to randomly assign the missing parameters or define weighted matrices.

6.3 Reformatting and Standardization of Building Attributes

The first step in processing the databases was to reformat into a consistent standard nomenclature the building attribute data contained within the exposure, claims, and TA databases. Table 2 provides the standard nomenclature for some of the main building attributes in these databases related to the hurricane vulnerability of the buildings. The FPHLM has four different building models: (1) personal residential for single-family homes (per); (2) commercial residential for one to three story low-rise apartment or condominium buildings (com); (3) manufactured homes (manuf); and (4) four stories of higher mid-rise to high-rise commercial residential apartment or condominium buildings (MHR). The exterior walls of the personal and low-rise commercial residential buildings can be made of timber (Tbr), unreinforced, reinforced, or generic masonry (MsryU, MsryR, Msry), or some other material (Other). The roof cover can be some variant of shingles (shng): shngR (rated), shngU (unrated)), shngH (high strength), shng (generic), tile, metal, or other material. The roof shape can be some variant of gable (gbl): gblB (braced), gblU (unbraced), or gbl (generic)), hip, or other roof shape. For each attribute, within each database, a link table relates the unique values for a given attribute within the database to a value matching the common nomenclature provided in Table 2. Table 3 illustrates this process, where unique values of exterior wall type from the Martin County TA database match the standard nomenclature. Generating the link tables was intuitive for most counties and attributes, but documentation on the exact definition of attribute values were generally not available from the counties, which introduces uncertainty in the link tables. For example, in Table 3, it is unclear what type of structural

Table 2 Common nomenclature for primary building attributes

Building model	Exterior wall	Roof cover	Roof shape
per	Tbr	shngR	gblB
com	MsryR	shngU	gblU
manuf	MsryU	shngH	gbl
MHR	Msry	shng	hip
	Other	tile	Other
		metal	
		Other	

Table 3 Example link table to reformat attributes into common nomenclature

Tax appraiser-assigned exterior wall	Standard nomenclature
Wood frame	Tbr
Concrete block	Msry
Concrete block 12	Msry
6 Concrete block	Msry
8 Concrete block	Msry
Concrete	Msry
Reinforced concrete	MsryR
Fireproof steel	Other
Fire resistant	Other
Metal	Other
Stone local cut	Other

system “Fire Resistant” actually refers to, and so without any other supporting documentation, it is coded as “Other.”

6.4 Hazard Information

To overcome the lack of hazard intensity data within the claims data, additional hazard-related information and land surface characteristics were added to the claims data, for each claim record, from a variety of sources. This information can help identify the cause or likelihood of a hazard event or attribute a particular hazard event to a loss claim. In the case of flood hazard, these data include:

Distance to Coast This is the distance of each claim location to the nearest coast. This information is useful to determine if coastal flood was likely to occur. The distance was computed using the 2011 Multi-Resolution Land Characteristics Consortium (MRLC) National Land Cover Database (NLCD) to identify water and land locations (Homer et al. 2015). The NLCD has approximately 30 m resolution, allowing for a reasonably accurate calculation.

Distance to Nearest Body of Water This information was computed in the same manner as the distance to coast, but the distance is to nearest body of water, regardless of water body type. This can help determine if the flooding could be due to river, stream, or lake overflow.

Precipitation Maximum These data were derived from the high resolution (4 km) PRISM (Parameter-Elevation Regression on Independent Slopes Model)¹ daily rainfall data (Daly et al. 2008). The largest precipitation amount that occurred with ± 1 day and within 10 km of the claim property location was recorded. This information helps

determine if the cause of loss could be due to accumulation of rainfall.

Storm Identification Many of the claims are due to tropical storms or hurricanes. We used the revised Atlantic hurricane database (HURDAT2) of the National Hurricane Center (2018) to determine if a storm was in the vicinity of the claim property during the claimed date of loss. The date and time of the storm track positions in HURDAT2 were compared to the claim location and time of loss using a spatial and temporal metric to determine possible storm influence. Additional information concerning the storm intensity level (depression, tropical storm, or hurricane) was included.

6.5 Cause of Loss Revision

The above information permitted a revision of the cause of loss where the original claim loss cause was in obvious error or ambiguous. The algorithm for the revision is as follows: If the original cause of loss was tidal water overflow, but the distance to the coast was greater than 20 km, the precipitation maximum for the location was checked. If it is greater than 1 inch of rainfall, the cause was revised to be “accumulation of rainfall”; otherwise the cause was designated as questionable. If the original cause of loss was accumulation of rainfall, but the precipitation maximum was less than 0.2 inch, then the cause of loss is marked as questionable. For hurricane events, if the original cause of loss is unknown for a property that is located less than 20 km from the coast and the precipitation maximum is greater than 1 inch, then the cause of loss is marked as undetermined, but is either “tidal water overflow” or “accumulation of rainfall,” and further investigation is warranted. Otherwise, the cause of loss is marked as tidal water overflow. For the cases flagged as “questionable” or “undetermined,” the modelers assign the final cause of loss based on the results of the numerical flood model (see Sect. 6.7).

6.6 Federal Emergency Management Agency Estimates of Flood Elevations

In the NFIP claim datasets, the claim data are assigned flood heights above ground level. These flood heights are based on the FPHLM hazard team outputs or from data obtained from the Federal Emergency Management Agency (FEMA 2016). In each case, the hazard intensity (flood inundation) is obtained by subtracting the ground elevation from the flood elevation. The ground and flood elevation are measured with respect to the North American Vertical Datum of 1988 (NAVD88).²

¹ <https://prism.oregonstate.edu/>.

² <https://www.ngs.noaa.gov/datums/vertical/north-american-vertical-datum-1988.shtml>.

There are relatively few observations of flood elevation and wave conditions for historical flood events, particularly at the resolution of claims data. For Hurricane Ivan in 2004, FEMA has published some resources on flood elevation (FEMA 2016). These include high water marks, flood surge height contours, and an inundation map. The flood contours are based on the high water marks along with engineering judgment. The inundation maps indicate all locations that experienced flooding, regardless of surge height. These data are highly correlated to observed flood damage. We derived estimated surge heights at NFIP claims data locations for Ivan by using an interpolation algorithm between the closest points to the FEMA surge contour elevations and the target claim location. The method used an inverse distance squared weighting method when the target location was between two contours, and a nearest neighbor extrapolation otherwise. The domain of the FEMA data, which covers four counties (Escambia, Okaloosa, Santa Rosa, and Walton), was divided into zones, so that interpolation was not done across contours that were in separate regions divided by dry land masses.

6.7 Historical Reconstruction

Except in cases like Ivan, where we used actual coastal flood measurements combined with wave model data, observations of flood elevations and wave conditions are not generally available or are very limited. To obtain estimates of these variables, we used the FPHLM hazard models and other data to reconstruct selected historical flood events that are listed in Table 1. For storm surge, the Coastal Estuarine Storm Tide (CEST) model (Zhang et al. 2008) was used. The CEST was forced by estimated observed winds from the H*Wind hurricane wind analyses (Powell et al. 1998) when available. Otherwise, modeled winds derived from the FPHLM wind model were used (Powell et al. 2005). Wave conditions were determined by a wave model from the FPHLM that is based on the Steady-State Spectral Wave Model (STWAVE) described by Massey et al. (2011). For inland flooding, a modeling framework based on the EPA Storm Water Management Model (SWMM) was used (Simon and Tryby 2018). For historical reconstruction, the inland flood model was driven by observed Next-Generation Radar (NEXRAD)³ rainfall data. The modeled hazard data were interpolated to the claim locations, taking into account high resolution 5 m digital elevation map data from the Florida Digital Elevation Model mosaic (FLDEM 2013). Based on the modeled results, each claim in the NFIP portfolio for the select historical events can be assigned to one of four

hydrological states: (1) inland flood with no waves; (2) coastal flood with minor waves; (3) coastal flood with moderate waves; and (4) coastal flood with severe waves.

6.8 Building and Contents Value Estimation

Modelers and actuaries assume in general that the coverage limit in private wind peril insurance policies is an acceptable measure of the true value of the insured building. The situation is different in the case of the NFIP exposure data. Several possibilities exist to remedy the situation when the property value in the NFIP file is either missing or unreliable. The NFIP (2013) provides adjustment formulas for underinsurance factors, which apply to the insurance limits of both building and contents. Alternatively, a triangulation between the tax roll databases, the NFIP claim files, and the wind insurance claim files can provide a solution. It is possible to compare the coverage limits of the wind policies (which is a better measurement of the true value of building or contents) to the values reported in the TA data for a large number of properties, and derive statistically a factor to adjust all the TA property values. These adjusted TA values would then act as a proxy for the building replacement values in the NFIP exposure portfolios.

In the case of NFIP claim data, the files contain updated building values, collected at the time of the claim. This is not the case for the contents value, however. To make up for the underestimation of contents value, the FPHLM team multiplies the coverage contents value by a factor equal to the ratio between the building property value and the building coverage. The assumption is that both building and contents coverage are underestimated in the same proportion. The result is an adjusted contents coverage value.

6.9 Geocoding and Integration of the Databases

For a given building, part of an insurance portfolio, or with a claim due to flood-induced and/or wind-induced damage during a hurricane, the following data sources could be joined: (1) the standardized building attributes information from the county TA database as well as the TA property values; (2) the wind insurance exposure portfolios; (3) the wind claims portfolios; (4) the NFIP exposure portfolio; (5) the NFIP claims portfolio; and (6) the hazard data from either model output, or field observations (for example, FEMA observations (FEMA 2016) for the case of Hurricane Ivan). In the case of the claim data, this combined information allows for the classification of the claims by building class and one of the four hydrological states (Sect. 6.7) in the case of flood hazard, where the relationship between wave height and inundation depth defines the four different flood condition states. This facilitates the

³ <https://www.ncdc.noaa.gov/data-access/radar-data/nexrad-products>.

Fig. 3 Process for linking National Flood Insurance Program (NFIP), wind, tax appraiser, and hazard data in a hurricane vulnerability model

development of semiempirical vulnerability curves, elaborated in the next section, as well as the comparison between the empirical vulnerability curves and the semi-engineering or engineering model curves used for the purpose of validation and calibration.

Figure 3 illustrates the specific links that join the various databases. The links can be broadly classified into table joins and spatial joins. In the table join, a common field between the two tables provides a link to map the different databases together. For instance, in the first level of the figure, table joins based on policy number map the NFIP claims into the NFIP exposure to form the NFIP portfolio, and also the unique parcel identification numbers join the tax appraiser data tables, the TA values, and the GIS shapefiles (containing the individual parcel polygons) to form the tax appraiser portfolio. Table joins also match standardized addresses (Goldberg et al. 2014) among all three datasets. Notice that the main strategy for joining the datasets is by matching standardized addresses. This is because this procedure yields high level of accuracy since it requires an exact match of address fields.

When combining data from heterogeneous sources, it is very common to have extraneous information in the address fields, which leads to a high percentage of non-matching cases. All records that fail the first attempt to match using standardized addresses attempt a rematch using a geospatial approach. For that, a within-spatial-join

Fig. 4 Buffer radius of 50 m to improve geocoded-based matching

combines the TA database with the NFIP or wind databases. A Python coding procedure reads the county shapefiles and finds the nearest elements in the NFIP or wind exposure or claim portfolios that is within a parcel polygon, employing the Global Positioning System (GPS) coordinates obtained from geocoding the physical address, projected into the U.S. National Atlas Equal Area system for enhanced metrics. A search buffer for valid parcels in the TA databases mitigates imprecisions of the geocoding engine, as Fig. 4 shows. The geocoding engines can resolve to incorrect parcels. For example, in Fig. 4, the blue marker represents an insurance claim, which after geocoding the address falls inside the green area, which is a park. To mitigate this issue, a search beam of 50 meters is

generated around the point, which intersects with polygons with parcel information. The closest valid parcel, represented by the orange marker, is assigned to the claim/point.

For each insurance claim or policy that matches with a TA record, the associated building attributes from the TA database are appended to the NFIP or wind databases, providing a combined or augmented dataset linking claims or exposure policies, and building attributes.

On the other hand, the spatial join has a second variant with which to generate the hybrid database linking the hazard intensities with the augmented NFIP or wind claim portfolios. In that case, the geocoding of the physical address in the NFIP or wind claim database provides their corresponding GPS coordinates. Then the coastal flood heights and/or inland flood heights, and/or wind speed intensity at the geocoded location for each claim (generated by the hazard model or from field data), and the claim record are joined for the same GPS coordinates (match-spatial-join). This match-spatial-join can be done on either the raw insurance claim portfolios or on the augmented insurance portfolios to produce either a raw or an augmented hybrid claim portfolio.

7 Development, Validation, and Calibration of Flood Vulnerability Curves

The processing of the data and the matching of the TA, hazard, and insurance data result in augmented NFIP and wind claims subsets that contain loss, value, hazard intensity, and structural characteristics. As case studies, this article describes the impact of the augmented NFIP claims data on the validation and calibration of a coastal flood building vulnerability curve, and on the development of a coastal flood contents vulnerability curve. Since the vulnerability curves are damage ratios as a function of hazard intensity, the modelers transformed the claim values into damage ratios by dividing the building and contents claim values by their respective replacement values, described in Sect. 2.2.

7.1 Validation of Building Vulnerability Curves for Coastal Flood

This section reports on coastal flood building vulnerability validation as a case study for hurricane Ivan, because this is the only hurricane for which there is substantial field-observed flood hazard intensity data. The hazard data for coastal flood (surge) is based on FEMA observations in the aftermath of Hurricane Ivan for the counties of Escambia, Santa Rosa, Okaloosa, and Walton (FEMA 2016).

In the NFIP hybrid claim databases, whether raw or augmented, the claims contained in any inundation depth

interval result in a large variability of reported damage ratios ranging between no damage to maximum damage. A two-way statistical analysis produces the empirically expected damage ratio values (EDVs) within each discrete 0.5 m inundation depth interval. The EDV is the mean or expected damage ratio given that the inundation depth is within a given interval. In the raw NFIP claim data, the year built and information on whether or not the structure is elevated is available, but no information exists on exterior wall type or number of stories. The authors computed the EDVs from a total of 1540 claims for pre-1994 nonelevated buildings, which represent houses built according to less stringent building codes, and conform the majority of the data. For example, pre-1994 masonry buildings tend to be unreinforced masonry. The EDVs were calculated for the different 0.25 m hazard intervals, up to 2.25 m depth for the case of coastal flood with minor waves. The augmented data contain exterior wall as well as number of story information in addition to the year built, so the authors could identify 248 buildings as pre-1994 masonry, one-story, slab-on-grade buildings, and computed the EDVs for this subset of claims.

Figure 5 shows the EDVs from both the raw data (solid triangles) and the augmented data (solid squares), together with the FPHLM vulnerability curve for the case of one-story, single family on grade (SFG), residential, unreinforced-masonry structure with a zero foot first floor elevation relative to ground level affected by coastal flood with minor waves. The figure shows a better agreement between the FPHLM vulnerability model and the augmented data points: the average percent difference relative to the model is 7.2% for the augmented data, and 16.3% for the raw data.

No reliable claim data exists beyond 2.25 m, due to a number of causes. They include: (1) a lack of sufficient data points for the higher hazard intensity intervals; (2) a need to refine the hazard intensity assignment strategy (for example, the higher inundation cases are all coastal, where perhaps localized geographical effects need further consideration); or (3) data pollution of various types, such as inaccurate claims information or structure elevation not properly reported. These potential causes are under investigation.

7.2 Development and Calibration of Contents Vulnerability Curves for Coastal Flood

This section reports on the coastal flood contents vulnerability curve development as a case study for all 12 storms of Table 1 combined. From this set, only the coastal flood claims corresponding to nonelevated residential structures were selected. With respect to the contents, the empirical EDVs, as a function of inundation depth, exhibit non-

Fig. 5 Validation of the vulnerability curve of a one-story, single family on grade (SFG) unreinforced, masonry-residential structure for coastal flood with minor waves. Hurricane Ivan 2004

monotonic trends, preventing the direct development of contents vulnerability curves based purely on the claim data. To address this problem, the FPHLM team converted the building vulnerability curves into contents vulnerability curves using a transfer function derived from the NFIP claim data. To derive this transfer function, the building damage ratio and contents damage ratio are calculated for each claim, for both the raw and the augmented data. In addition, for the case of the contents damage ratio, the authors tested two possibilities to define the contents replacement value (that is, the denominator of the contents damage ratio). The contents replacement value can be either the policy contents coverage limit, or the adjusted contents coverage limit described in Sect. 6.8.

Using a two-way statistical analysis, similar to the one used in the previous section, the authors computed the contents EDVs, given that the building damage ratio is within a certain interval, for the four cases of raw and augmented claim data, with either contents coverage value or adjusted contents coverage value. The raw data include all coastal flood claims for pre-1994 nonelevated buildings, while the augmented data include only the coastal flood claims for pre-1994 one-story masonry buildings. Figure 6a shows a plot of these contents EDVs as a function of the building damage ratio for all four cases with the corresponding curve fits. These curve fits are the potential transfer function between the building and contents

vulnerability curves. They are similar to a vulnerability curve, where the contents damage ratio is a function of the building damage ratio, instead of being a function of the hazard intensity.

With the FPHLM building vulnerability curve and the transfer function, it is possible to derive a contents vulnerability curve through a mapping procedure. For a given hazard intensity, the building vulnerability curve yields the building damage ratio. Then, using that building damage ratio as the input of the transfer function, the contents damage ratio is produced, which corresponds to the given hazard intensity. Repeating this process for different hazard intensities will result in the contents vulnerability curve. Figure 6b is an example of the resulting contents vulnerability curves for the hydrological state of surge with minor waves, for a one-story on grade unreinforced masonry structure.

This methodology resolves the issue of the uncertainty attached to the assignment of a specific hazard intensity to each claim. The claims are simply grouped, according to their structural characteristics and flood condition (coastal versus inland flood). The method also resolves to a certain extent the lack of claim data for higher hazard intensities; it accomplishes this result by linking the claim data to the entire range of the building vulnerability model, which extends to the whole range of hazard intensity. Finally, the procedure also ensures compatibility between the building

Fig. 6 Contents versus building damage ratios (a); and contents vulnerability curves derived from the Florida Public Hurricane Loss Model (FPHLM) building vulnerability curves and claim data, for one-story, single family on grade (SFG), unreinforced masonry-residential structures subjected to coastal flood with minor waves (b)

vulnerability model, the claim data, and the resulting contents vulnerability model.

The authors computed the maximum relative differences between two contents, vulnerability curves 1 and 2, as the maximum ratio $(EDV2-EDV1)/EDV1$. This ratio is equal to 7.1% for the case of the augmented NFIP curve with contents coverage (black curve) versus the raw NFIP curve with contents coverage (blue curve). This ratio gives a measurement of the epistemic uncertainty attached to the lack of building information in the claim data. It is 6.5% for the case of the augmented NFIP curve with adjusted

contents coverage (red curve) versus the augmented NFIP curve with contents coverage (black curve). This ratio gives a measurement of the epistemic uncertainty attached to the contents coverage value. It is 7.5% for the case of the augmented NFIP curve with adjusted contents coverage (red curve) versus the raw NFIP curve with contents coverage (blue curve). This ratio gives a measurement of the epistemic uncertainty attached to the combined contents coverage value and the lack of building information in the claim data. These metrics have to be considered with caution, since they are also dependent on the number of

claim data points in each subset, and their distribution over the different inundation depth intervals. The metrics illustrate nonetheless substantial differences in the final vulnerability curves, which will have an impact on the predicted losses.

8 Discussion

The case studies above illustrate the potential epistemic uncertainties associated with the use of claims data for risk model development, calibration, and validation, and should temper the temptation to treat claims data as an “exact solution.” These case studies can be replicated for other hazard conditions, and other building classes, and the lessons learned are not restricted to hurricane models in Florida. The examples illustrate both the benefits and the problems linked to data augmentation. The main issues in the data are the lack of detailed building information in both the insurance and the tax roll datasets. The case studies show how even modest improvements in the claim data can have an effect on model vulnerability curves, which in turn are used to compute insured losses to guide what the insurance companies will charge as a premium.

The irregular quality of the tax roll data used to augment the insurance data limits the benefits of data augmentation, since the size of the augmented datasets tends to be smaller than the size of the raw datasets. The improvement in data quality might be lost to reduction in data size. The solution lies in an effort at the county level, supported by the state, to improve the quality of the tax roll data collection. Coordination between building departments and tax appraiser offices at the county level, and the departments of insurance regulation, revenue, and emergency management at the state level, coupled with training of the data collectors and increased resources should ensure a uniform level of quality of the tax roll data throughout the state. Similar efforts from the public and private insurers should also result in increased quality of the insurance portfolio data for both exposure and claim data, which would render data augmentation less critical.

Modern technology, such as combining machine learning with advanced instrumentation (Lidar, drones, 3D cameras, and so on), can provide an alternative path to produce reliable complete datasets of both exposure and post-disaster building inventories (Catbas and Kijewski-Correa 2013; Kijewski-Correa et al. 2014; Koc et al. 2019). Another avenue of improvement lies in the gathering of field hazard intensity data thanks to more extensive networks of instrumentation. These advances necessitate continuous investments from agencies like the U.S. Geological Survey or the National Oceanic and Atmospheric

Administration in the United States and their counterparts in other countries (Kohler et al. 2017).

Extreme hazard events have very large return periods and are underrepresented in the historical claim data. In addition, for any given event, the density of claims gravitates toward the lower intensity, since the footprint of the hazard itself shrinks with intensity. For example, the number of NFIP claims within the State of Florida decreases as the hazard intensity increases. The result is that development, validation, and calibration of vulnerability models based on claim data are more accurate in lower hazard intensities, as illustrated in Fig. 5. The method described in Fig. 6 offers a partial solution to that problem, since it decouples the model development from the hazard intensity data.

9 Conclusion and Recommendations

Catastrophe models are complex probabilistic nonlinear systems, whose results are very sensitive to the quality of the input data. Not only is it critical to ensure a certain level of input data quality, but the different components of the model (hazard, vulnerability, and actuarial) need to be validated and calibrated to the greatest extent possible. In particular, the analysis of insurance claims data provides a way for developing, validating, and calibrating various aspects of the vulnerability component of the cat model, in order to improve the credibility of the model outputs. Typically, insurance claims data need to go through extensive processing and interpretation. This article describes some of the challenges risk modelers face when utilizing insurance claims data.

The article discusses the different data sources involved in this development and validation process: insurance exposure and claim data; tax roll data; geographic information system data; elevation or topographic data; and hazard data from either observations or simulations. The article walks through the steps needed to integrate these different sources of data into a meaningful ensemble. The result is augmented insurance data, which result in higher quality input data for the cat model, and facilitate the development and validation process of the vulnerability model. In particular, the article shows how the augmented claim data can validate and calibrate the building vulnerability curves. A 3-step methodology then transforms these building vulnerability curves into new contents vulnerability curves. In that process, the contents damage to the building damage relationship derived using insurance claims data acts as a transfer function between building and contents vulnerabilities. This method produces contents vulnerability curves compatible with both the claim data and the building vulnerability models, and it does not rely

on estimates of the hazard intensities that produced the claims. In addition, the matching of tax appraiser databases with the wind and flood insurance portfolios has the potential to increase the accuracy of the portfolio analyses, since more data will be available during the analyses.

Yet the data augmentation itself suffers from the lack of quality assurance of the tax roll data. Substantial efforts are needed on the part of state agencies to remedy this situation, as well as on the part of insurance companies to reduce the need for data augmentation. This is a work in progress. Additional work includes the evaluation of the uncertainty attached to other sources of data like digital elevation data, or surface roughness data, the validation of the hazard model, the validation of the combined wind and flood model, and extending the case studies presented herein to a larger number of events.

Acknowledgements The Florida Office of Insurance Regulation (FLOIR) provided financial support for this work. The opinions, findings, and conclusions presented in this article are those of the authors alone, and do not necessarily represent the views of the FLOIR. Special thanks to Yuepeng Li from Florida International University, and Andrew Kennedy from University of Notre Dame, who provided hazard information, as well as to Daysiry Rodriguez and Amssatou Diagne who processed the databases.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Baradaranshoraka, M., J.-P. Pinelli, K. Gurley, M. Zhao, X. Peng, and A. Paleo-Torres. 2019. Characterization of coastal flood damage states for residential buildings. *ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems, Part A: Civil Engineering* 5(1): Article 04019001.
- Baradaranshoraka, M., J.-P. Pinelli, K. Gurley, X. Peng, and M. Zhao. 2017. Hurricane wind versus storm surge damage in the context of a risk prediction model. *Journal of Structural Engineering* 143(9): Article 04017103.
- Biasi, G., M.S. Mohammed, and D.H. Sanders. 2017. Earthquake damage estimations: ShakeCast case study on Nevada bridges. *Earthquake Spectra* 33(1): 45–62.
- Catbas, F.N., and T. Kijewski-Correa. 2013. Structural identification of constructed systems: Collective effort toward an integrated approach that reduces barriers to adoption. *Journal of Structural Engineering* 139(10): 1648–1652.
- Chian, S.C. 2016. A complementary engineering-based building damage estimation for earthquakes in catastrophe modeling. *International Journal of Disaster Risk Science* 7(1): 88–107.
- Daly, C., M. Halbleib, J.I. Smith, W.P. Gibson, M.K. Doggett, G.H. Taylor, J. Curtis, and P.P. Pasteri. 2008. Physiographically sensitive mapping of climatological temperature and precipitation across the conterminous United States. *International Journal of Climatology* 28(15): 2031–2064.
- Der Kiureghian, A.D., and O. Ditlevsen. 2009. Aleatory or epistemic? Does it matter? *Structural Safety* 31(2): 105–112.
- Dong, W. 2002. Engineering models for catastrophe risk and their application to insurance. *Earthquake Engineering and Engineering Vibration* 1(1): 145–151.
- FEMA (Federal Emergency Management Agency). 2015. *HAZUS-MH 2.1 hurricane model technical manual*. Washington, DC: Mitigation Division, Federal Emergency Management Agency.
- FEMA (Federal Emergency Management Agency). 2016. *Geographic information systems data*. <https://www.fema.gov/geographic-information-systems-data>. Accessed 11 Jul 2019.
- FLDEM (Florida Digital Elevation Model). 2013. *Florida Digital Elevation Model (DEM) mosaic—5-meter cell size—Elevation units centimeters*. Gainesville, FL: University of Florida GoPlan Center. https://www.fgdl.org/metadata/fgdl_html/flidar_mosaic_cm.htm. Accessed 24 Oct 2020.
- FPHLM (Florida Public Hurricane Loss Model). 2019. *Florida public hurricane loss model 7.0*. Miami, FL: Laboratory for Insurance, Financial, and Economic Research, Florida International University, International Hurricane Research Center (IHRC). <https://www.sbafla.com/methodology/ModelerSubmissions/CurrentYearModelSubmissions.aspx>. Accessed 24 Oct 2020.
- Goldberg, D.W., J.N. Swift, and J.P. Wilson. 2014. *Address standardization*. Technical report 12. Los Angeles, CA: GIS Research Laboratory, University of Southern California.
- Hamid, S., J.-P. Pinelli, S.-C. Cheng, and K. Gurley. 2011. Catastrophe model based assessment of hurricane risk and estimates of potential insured losses for the State of Florida. *Natural Hazard Review* 12(4): 171–183.
- Harvey Jr., P.S., S.K. Heinrich, and K.K. Muraleetharan. 2018. A framework for post-earthquake response planning in emerging seismic regions: An Oklahoma case study. *Earthquake Spectra* 34(2): 503–525.
- Homer, C.G., J.A. Dewitz, L. Yang, S. Jin, P. Danielson, G. Xian, J. Coulston, N.D. Herold, et al. 2015. Completion of the 2011 national land cover database for the conterminous United States—Representing a decade of land cover change information. *Photogrammetric Engineering and Remote Sensing* 81(5): 345–354.
- Johnson, T., J.-P. Pinelli, T. Baheru, A.G. Chowdhury, J. Weekes, and K. Gurley. 2018. Simulation of rain penetration in buildings and associated damage within a hurricane vulnerability model. *Natural Hazard Review* 19(2): Article 04018004.
- Kaczmarek, J., S. Jewson, and E. Bellone. 2018. Quantifying the sources of simulation uncertainty in natural catastrophe models. *Stochastic Environmental Research and Risk Assessment* 32(3): 591–605.
- Khanduri, A.C., and G.C. Morrow. 2003. Vulnerability of buildings to windstorms and insurance loss estimation. *Journal of Wind Engineering and Industrial Aerodynamics* 91(4): 455–467.
- Kijewski-Correa, T., N. Smith, A. Taflanidis, A. Kennedy, C. Liu, M. Krusche, and C. Vardeman II. 2014. CyberEye: Development of integrated cyber-infrastructure to support rapid hurricane risk assessment. *Journal of Wind Engineering and Industrial Aerodynamics* 133: 211–224.
- Koc, E., B. Cetiner, L. Soibelman, and E. Taciroglu. 2019. System-based vulnerability and resilience assessment in mega-scale transportation systems: Towards data and model-driven

- methodologies. In *Computing in civil engineering 2019: Smart cities, sustainability, and resilience*. Selected Papers from the ASCE International Conference on Computing in Civil Engineering 2019, 17–19 June 2019, Atlanta, GA, USA, ed. Y.K. Cho, F. Leite, A. Behzadan, and C. Wang, 444–450. Reston, VA: American Society of Civil Engineers.
- Kohler, M.D., E.S. Cochran, D. Given, S. Guiwits, D. Neuhauser, I. Henson, R. Hartog, P. Bodin, et al. 2017. Earthquake early warning ShakeAlert system: West coast wide production prototype. *Seismological Research Letters* 89(1): 99–107.
- Massey, T.C., M.E. Anderson, J. McKee Smith, J., Gomez, and R. Jones. 2011. *STWAVE: Steady-State Spectral Wave Model user's manual for STWAVE, version 6.0*. Washington, DC: U.S. Army Corps of Engineers.
- Michalski, J. 2016. *Building exposure study in the State of Florida and application to the Florida Public Hurricane Loss Model*. Master thesis. Melbourne, FL, USA: Florida Tech.
- Michel-Kerjan, E., S. Hochrainer-Stigler, H. Kunreuther, J. Linne-rooth-Bayer, R. Mechler, R. Muir-Wood, N. Ranger, P. Vaziri, and M. Young. 2013. Catastrophe risk models for evaluating disaster risk reduction investments in developing countries. *Risk Analysis* 33(6): 984–999.
- Muir-Wood, R., and D. Stander. 2016. Resilience analytics and the public sector management of disasters. In *Extended abstract collection*. Vol. 2 of the Proceedings of the 6th International Disaster and Risk Conference: Integrative Risk Management—Towards Resilient Cities, IDRC Davos 2016, 28–31 August 2016, Davos, Switzerland, ed. M. Stal, D. Sigrist, S. Wahlen, J. Portmann, J. Glover, N. Bernabe, D. Moy de Vitry, and W.J. Ammann, 436–439. Davos: International Development Research Centre.
- National Hurricane Center. 2018. *Best track data (HURDAT2)*. <https://www.nhc.noaa.gov/data/#hurdat>. Accessed 17 Nov 2019.
- NFIP (National Flood Insurance Program). 2013. *Technical documentation of NFIP actuarial assumptions and methods*. Supporting rates effective 1 October 2013. Washington, DC: FEMA.
- Nicholson, J.E., K. Clark, and G. Daraskevich. 2018. The Florida insurance market: An analysis of vulnerabilities to future hurricane losses. *Journal of Insurance Regulation* 37(3): 57–89.
- Pinelli, J.-P., G. Pita, K. Gurley, B. Torkian, S. Hamid, and C. Subramanian. 2011. Damage characterization: Application to Florida public hurricane loss model. *Natural Hazard Review* 12(4): 190–195.
- Pita, G., J.-P. Pinelli, S. Cocke, K. Gurley, J. Mitrani-Reiser, J. Weekes, and S. Hamid. 2012. Assessment of hurricane-induced internal damage to low-rise buildings in the Florida Public Hurricane Loss Model. *Journal of Wind Engineering & Industrial Aerodynamics* 104–106: 76–87.
- Pita, G., J.-P. Pinelli, K. Gurley, and S. Hamid. 2013. Hurricane vulnerability modeling: Evolution and future trends. *Journal of Wind Engineering & Industrial Aerodynamics* 114: 96–105.
- Pita, G., J.-P. Pinelli, K. Gurley, and J. Mitrani-Reiser. 2014. State of the art of hurricane vulnerability estimation methods: A review. *Natural Hazard Review* 16(2): Article 04014022.
- Powell, M.D., S.H. Houston, L.R. Amat, and N. Morisseau-Leroy. 1998. The HRD real-time hurricane wind analysis system. *Journal of Wind Engineering and Industrial Aerodynamics* 77–78: 53–64.
- Powell, M.D., G. Soukup, S. Cocke, S. Gulati, N. Morisseau-Leroy, S. Hamid, N. Dorst, and L. Axe. 2005. State of Florida hurricane loss projection model: Atmospheric science component. *Journal of Wind Engineering and Industrial Aerodynamics* 93(8): 651–674.
- Risk Management Solutions. 2019. *North Atlantic hurricane models, version 18.1 (Build 1945)*. Submitted in compliance with the 2017 Standards of the Florida Commission on Hurricane loss projection methodology, 21 May 2019.
- Roueche, D.B., D.O. Prevatt, and F.T. Lombardo. 2018. Epistemic uncertainties in fragility functions derived from post-disaster damage assessments. *ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems, Part A: Civil Engineering* 4(2): Article 04018015.
- Schneider, P.J., and B.A. Schauer. 2006. HAZUS—Its development and its future. *Natural Hazards Review* 7(2): 40–44.
- Shah, H.C., W. Dong, P. Stojanovski, and A. Chen. 2018. Evolution of seismic risk management for insurance over the past 30 years. *Earthquake Engineering and Engineering Vibration* 17(1): 11–18.
- Simon, M., and M. Tryby. 2018. *Open source SWMM: Community-based software development for storm water management modeling*. Washington, DC: U.S. Environmental Protection Agency.
- Suppasri, A., E. Mas, I. Charvet, R. Gunasekera, K. Imai, Y. Fukutani, Y. Abe, and F. Imamura. 2013. Building damage characteristics based on surveyed data and fragility curves of the 2011 Great East Japan tsunami. *Natural Hazards* 66(2): 319–341.
- USACE (United States Army Corps of Engineers). 2006. *Depth-damage relationships for structures, contents, and vehicles and content-to-structure value ratios (CSV) in support of the Donaldsonville to the Gulf, Louisiana, feasibility study*. New Orleans District, Louisiana: US Army Corps of Engineers.
- USACE (United States Army Corps of Engineers). 2015. *US North Atlantic coast comprehensive study: Resilient adaptation to increasing risk*. Physical damage function summary report. US Army Corps of Engineers. <https://www.nad.usace.army.mil>. Accessed 24 Oct 2020.
- Wald, D., K. Lin, and K. Porter. 2008. ShakeCast: Automating and improving the use of ShakeMap for post-earthquake decision-making and response. *Earthquake Spectra* 24(2): 533–553.
- Zhang, K., C. Xiao, and J. Shen. 2008. Comparison of the CEST and SLOSH models for storm surge flooding. *Journal of Coastal Research* 24(2): 489–499.