

A Framework for Characterizing Uncertainty Factors in Postdisaster Structural Performance Assessment Data

Tori Tomiczek, M.ASCE

Associate Professor, Dept. of Naval Architecture and Ocean Engineering, United States Naval Academy, 590 Holloway Rd., Mail Stop 11D, Annapolis, MD 21403 (corresponding author). ORCID: <https://orcid.org/0000-0003-4116-7547>. Email: vjohnson@usna.edu

Jennifer Helgeson

Research Economist, Engineering Laboratory, National Institute of Standards and Technology, 100 Bureau Dr., Mail Stop 8611, Gaithersburg, MD 20899-8611. ORCID: <https://orcid.org/0000-0002-3692-7874>. Email: jennifer.helgeson@nist.gov

Elaina Sutley, M.ASCE

Associate Professor, Dept. of Civil, Environmental, and Architectural Engineering, Kansas Univ., 2150 Learned Hall, Lawrence, KS 66045. Email: enjstutley@ku.edu

Donghwan Gu

Professional Research Experience Program (PREP) Postdoctoral Researcher, Community Resilience Group, National Institute of Standards and Technology, 100 Bureau Dr., Mail Stop 8611, Gaithersburg, MD 20899-8611. ORCID: <https://orcid.org/0000-0002-8189-2455>. Email: donghwan.gu@nist.gov

Sara Hamideh

Assistant Professor, School of Marine and Atmospheric Sciences, Stony Brook Univ., Stony Brook, NY 11794. ORCID: <https://orcid.org/0000-0001-5298-9525>. Email: sara.hamideh@stonybrook.edu

P. Shane Crawford, M.ASCE

Guest Research Associate, National Institute of Standards and Technology, 100 Bureau Dr., Mail Stop 8611, Gaithersburg, MD 20899-8611. ORCID: <https://orcid.org/0000-0003-3848-1447>. Email: patrick.crawford@nist.gov

Forum papers are thought-provoking opinion pieces or essays founded in fact, sometimes containing speculation, on a civil engineering topic of general interest and relevance to the readership of the journal. The views expressed in this Forum article do not necessarily reflect the views of ASCE or the Editorial Board of the journal.

[https://doi.org/10.1061/\(ASCE\)NH.1527-6996.0000604](https://doi.org/10.1061/(ASCE)NH.1527-6996.0000604)

This work is made available under the terms of the Creative Commons Attribution 4.0 International license, <https://creativecommons.org/licenses/by/4.0/>.

Introduction

Postdisaster structural performance assessments provide quantifiable evidence of infrastructure system performance under hazard conditions, help to prioritize emergency response and recovery activities, and identify broader design issues. Such assessments can form a basis for empirical fragility functions, refine existing process-based models based on validation exercises, and establish

baseline conditions for recovery tracking and modeling. Therefore, understanding the uncertainty associated with an assessment and its effect on damage, loss, or recovery models is critical for the verification and validation processes of these models. Uncertainties in a model's characterization of a hazard and the resulting structural response should be quantified when possible. Many models propagate these uncertainties to evaluate their effects on model outputs or perform sensitivity tests to changes in model parameters. However, modal, temporal, and human factors that may cause variability in the assessment of a given system's condition or state at a specific point in time are rarely characterized to assess how sources of uncertainty can compound in postdisaster performance assessments. This study identified and analyzed sources of uncertainty in postdisaster performance assessments including *how* and *when* assessments are performed and *who* is involved with postdisaster assessments. The discussion of *how* covers the mode of assessment, including field-based assessments, remote observation, and/or combined observation and measurement. *When* considers the temporal nature of an assessment, including the timing of performance assessments following a disaster and the effects of cascading or successive disasters. Last, *who* captures human factors, including surveyor bias, training, and familiarity with the affected area. A framework for quantifying uncertainty factors in field-based performance assessments is proposed. The framework follows the methodology presented in FEMA P-695 (FEMA 2009), which characterizes total system collapse uncertainty for seismic hazards based on four uncertainty (β) factors. Repositories of publicly available damage data hosted on platforms such as DesignSafe-CI [a cyberinfrastructure component of the Natural Hazards Engineering Research Infrastructure (NHERI) collaboration funded by the National Science Foundation] (Rathje et al. 2017) may be leveraged in future work to develop quantitative values for proposed uncertainty parameters in field study measurements and modeling applications.

Background

While many previous studies have investigated infrastructure and social systems subjected to natural hazard events through postevent reconnaissance efforts, few report uncertainty as a standard element in assessed damages or performance. In recent years, several studies have explicitly addressed uncertainty in various aspects of performance assessment; several recent examples are presented in Table 1. However, a robust, standardized framework for systematically and consistently quantifying these inherent uncertainties is necessary for comparison across hazard types and individual events. Such a framework is critical for verification and validation processes when collected data and assessments are used to inform vulnerability and resilience models.

A Theoretical Framework for Characterizing Uncertainty in Postdisaster Performance Assessments

Uncertainty in postdisaster performance assessments can stem from various factors, all of which are affected by the nature of the event

Table 1. Selected previous damage studies that considered uncertainty in damage state assessment

Study	Event (year)	Assessed system (scale)	Mode of collection considered	Uncertainty considerations
Roueché et al. (2018)	Joplin tornado (2011)	Buildings	Data set of 1,241 residential structures	Investigated epistemic uncertainties in fragility functions based on postdisaster assessments of structural damage. Uncertainty in hazard intensity measurement (wind speed) was identified as the greatest contributor to uncertainty in fragilities; potential misclassification and sampling error were small contributors to uncertainty with a sufficient number of samples.
Marvi (2020)	General flooding	Residential, commercial, industrial buildings: structure and contents	Review paper describing empirical models and analytical models for flood damage analysis	Reviewed previous works on flood damage analysis for building structures and contents. For both empirical and analytical models, the study identified a lack of reliable data for model construction and validation as a methodological challenge. The study noted that previous studies identified the inability to communicate uncertainty in models.
Nofal and van de Lindt (2020a)	General flooding	Community resilience	Review paper describing flood risk components (hazard, exposure, and vulnerability) and deterministic and stochastic approaches for modeling damage, loss, and recovery	Described the need for probabilistic approaches for propagating uncertainties and identified the need for a standardized assessment framework to measure the impact of approaches to community resilience.
Helgeson et al. (2021)	Hurricane Florence (2018)	Households, businesses	Field-based performance assessments; in-person surveys of residents and businesses	Addressed uncertainty due to repeated flood events in a community and proposed a methodology for disaggregating the impacts of first and second flood events.
Khajwal and Noshadravan (2021)	Hurricane Harvey (2017)	Buildings	Crowdsourced surveys based on postdisaster imagery	Implemented a decision rule to infer damage state based on participant responses and used an information theoretic model to obtain probabilistic descriptions of inferred damage states while quantifying reliability of respondents and ambiguity of images.
Sutley et al. (2021)	Hurricane Matthew (2016)	Households, businesses	In-person surveys of residents and businesses	Addressed uncertainty due to mode of performance assessment, comparing household-reported damage to engineering assessments; addressed temporal uncertainties associated with time of data collection.

as well as the typology and scale of the assessed system. A framework for characterizing these uncertainties is shown in Fig. 1. The framework comprises four steps to identify, characterize, quantify, and propagate uncertainty in postevent performance assessment and modeling. This study focused on the first two steps of the framework: (1) identifying factors affecting sources of uncertainty; and (2) characterizing sources of uncertainty in postevent performance assessments.

In Step 1, factors affecting sources of uncertainty include the nature of the event and the type and scale of the assessed system. The nature of the event (red box) refers to hazard intensity and damage mechanisms that may impact a system or component of a system. It also includes factors that may alter the capabilities of researchers to establish baseline conditions: for example, for

characterizing multihazard events, cascading or successive hazard events, or longitudinal recovery. The type and scale of the assessed system (cyan box) affect the sample size, sample methodology, and system archetypes chosen for performance assessment.

In Step 2, nature, type, and scale affect epistemic uncertainties in performance assessments due to modal (blue box, the method by which damage data are collected), temporal (orange box, assessment timelines in comparison to damage or recovery timelines), and human (yellow box, potential bias in damage observations) factors, all of which comprise aleatory and epistemic uncertainties. While aleatory uncertainty cannot be reduced, aleatory and epistemic uncertainties arising from modal, temporal, and human factors can be characterized. By evaluating the uncertainties in these three factors, research teams can (1) take steps to mitigate epistemic

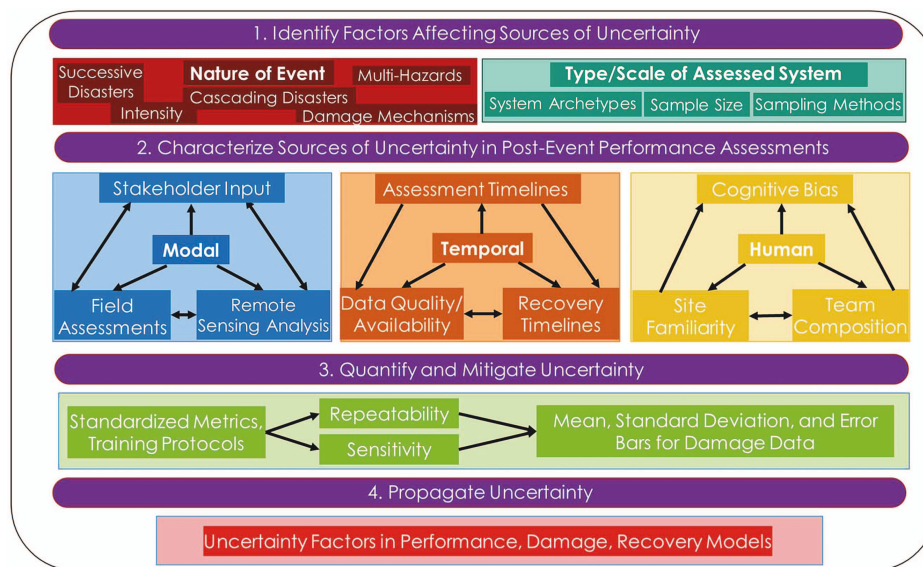


Fig. 1. (Color) Framework for uncertainty characterization in postdisaster performance assessments.

uncertainty; and (2) quantify uncertainty factors that can be used in damage or recovery models to better communicate risks to decision makers and stakeholders. The following sections characterize factors affecting aleatory and epistemic uncertainties and offer mechanisms to reduce epistemic uncertainties.

Nature of the Event

The nature of the event comprises the type and intensity of hazard (s), warning time, and impacted area or footprint. This study focused on natural hazards, including earthquakes, tsunamis, floods, droughts, wildfires, and wind. Many of these hazards can occur (1) simultaneously in multihazard events, such as hurricanes (surge + wave + wind + rain + debris) and tornadoes (wind + rain + debris), (2) successively, such as a second flood event occurring before a community is able to recover from an earlier event (e.g., Helgeson et al. 2021; Crawford et al. 2022), or (3) with cascading effects, such as seismic events that cause fire or fire events that cause flooding.

In a performance assessment, it can be challenging to isolate damage sources. For example, during tornadoes, damage to a building structure may be initially caused by strong uplift and torsional wind loads or from windborne debris, among other factors, whereas total damage may be caused by a combination of multiple hazards (Minor et al. 1977). Depending on whether debris missiles remain on site, their existence and impact can be difficult to discern. Multi-hazard events such as hurricanes can similarly create challenges in distinguishing between wind and water damage. Properly identifying the source and magnitude of structural damage due to either wind and wind-driven rainfall or storm surge and wave action can be critical in allocating separate insurance policies, which may cover only one hazard. While wind loading and storm surge and wave loading cause different damage signatures for affected structures, structural expertise and experience as well as knowledge of construction and structural responses to either loading type are necessary to accurately identify the occurrence and magnitude of either source of damage during a hurricane (Peraza et al. 2014). Accurately assessing these differences may require specific reconnaissance methods (i.e., field inspections rather than resident surveys or assessment of remote aerial imagery) and significant investment of time and resources. Some types of damage may be more visible than other types in remote sensing data such as

aerial imagery. For example, building collapse due to an earthquake or roof damage due to a wind storm is often evident from above a structure, whereas intermediate damage or flood damage to a building's walls or interior is often not easily identifiable from aerial images. If on-the-ground surveyors cannot see the roof or cannot enter building interiors, they may not be able to accurately assess damage to these systems.

The warning time and footprint of a hazard event affect preparation, sheltering, and evacuation decisions and can ultimately affect the magnitude of infrastructure and social system disruptions. The magnitude of a disruption of infrastructure and social systems may alter a survey team's arrival depending on available resources, transportation, and safety concerns. In turn, a survey team's arrival and the proportion of residents who evacuated versus sheltered in place during an event may affect a research team's ability to seek occupant input on predisaster conditions and peak hazard levels (e.g., peak inundation height during a flood or peak wind speeds during a tornado). Last, hazard intensity impacts accuracy in assessment. For example, systems experiencing complete damage/loss can be more obvious to a surveyor, whereas minor or intermediate damage levels may be more subjective. Characterizations of minor or intermediate damage levels may require site familiarity or information being confirmed by an occupant, particularly for interior damage that may occur due to flooding or rainwater intrusion (e.g., Tomiczek et al. 2020; Sutley et al. 2021). Furthermore, the evolution of damage with increasing or sustained hazard conditions is difficult to identify following an event. For example, floods leave watermarks on structures at the height of sustained inundation levels. It can be challenging for surveyors to determine whether watermarks indicate peak inundation height or represent a lower level of inundation that was sustained for a longer duration. All of these aspects create uncertainty surrounding the nature of an event.

Type and Scale of the Assessed System

The type and scale of the assessed system can affect the epistemic uncertainty and level of detail of postevent reconnaissance. Both the research question and system of interest drive the effective sample size and sampling methodology for detecting and mitigating uncertainties. In addition, other parameters such as statistical significance and expected prevalence of an investigated outcome (e.g., damage) influence decisions on sample size and sampling

methods. Using an appropriate sample size and sampling methods is critical to avoid Type I (incorrectly rejecting the null hypothesis) or Type II (incorrectly accepting the null hypothesis) errors in interpretation of collected data (Jones et al. 2003; Groves et al. 2011; Martínez-Mesa et al. 2014). For example, if residential buildings are the system of interest, then a sample size sufficient to represent the entire geographical area and all construction types of interest should be designed. If residential buildings with a specific foundation (e.g., crawlspace, pile-elevated) are the system of interest, then a smaller (relative to all residential buildings) sample can be identified. If specific building components constitute the system of interest, a targeted sample should be designed, and sample design methodology may require in-person investigation rather than remote data collection. Including a control group (i.e., of undamaged systems) is another crucial part of sample design, because it helps identify uncertainties related to a hazard event and common construction and maintenance practices of the local area. For example, to assess resilience and recovery in Lumberton, North Carolina, following a flood precipitated by a hurricane, van de Lindt et al. (2018) used a nonproportional stratified cluster sampling technique to select housing units with high and low probabilities of being in the inundation zone at a three to one ratio by census block. Given that the goal was to evaluate community-level resilience and recovery, it was critical for the sample to include housing units inside and outside the flooded area.

Analysis of the archetypal performance of different infrastructure components under hazard conditions based on field observations can be applied by researchers in inter- or transdisciplinary settings. Such analysis allows researchers to study resilience and recovery issues from a broad range of theoretical and methodological standpoints to identify, for example, differences in damage and recovery of buildings with elevated or nonelevated foundation types following a flood event. The use of systems archetypes is critical when considering community-level resilience. However, scaling archetypes across size and complexity is complicated. Moreover, the scope and detail of reconnaissance efforts will be affected by the archetype and system scale considered, and it is difficult to note general trends in the formation and types of unintended consequences when dealing with similar hazards across different system scales. The effects of risk factors for unintended consequences may only be known in hindsight and may require a high level of sample frequency across historical data to be recognized as anything more than noise.

Modal Uncertainties

Modal uncertainties refer to epistemic and aleatory uncertainties; epistemic uncertainties are affected by the methodology or combination of methodologies by which damage data are collected. Damage observations may be collected in person or virtually via field-based assessments, remotely-sensed data and imagery, stakeholder input, or a combination of these data collection modes. Depending on the nature of an event and the type and scale of an assessed system, each mode has advantages and limitations, and all modes face challenges of access, data availability, and data quality. For example, while field-based performance assessments readily identify damage to structural components, observations may be limited to structural exteriors and may therefore limit the ability of researchers to assess interior damage. Stakeholders (e.g., residents, owners, or operators) can provide key details about interior damage or changes in damage state between the time of an event and the time of a survey team's assessment. However, stakeholder-reported damages can be affected by a respondent's trauma, bias, and experience with contractors (e.g., a contractor

may overimply damage to a respondent to gain more work). Field reconnaissance efforts also must be based on an awareness that resources in affected areas may be needed for disaster response (e.g., hotel rooms for temporary shelter and/or for housing emergency responders; generators for power; limited food, water, and gas supplies for disaster victims and emergency responders from local restaurants, grocery stores, and disaster relief efforts) and face ethical considerations regarding the timing, duration, and logistics of field-based performance assessments or surveys (Gaillard and Peek 2019). While satellite imagery or other remotely sensed data can allow surveyors to quickly assess a hazard's footprint and building-level impact without requiring travel to an affected community and associated fieldwork resources, such assessments are often insufficient to capture intermediate damage states (Brown et al. 2012) and are affected by the quality and extent of available imagery.

To reduce modal uncertainty, a field study team can combine modes of performance assessment. For example, preevent remote sensing can provide a baseline for quality control of postevent assessments and can identify preexisting issues, such as structural damage due to deferred maintenance. Postevent virtual and field-based remote sensing [e.g., data from ground, aerial/unmanned aerial vehicles (UAV) or satellite sensors including, but not limited to, photography, multispectral, or light detection and ranging (LiDAR)] and virtual and field-based repair/recovery data collection can also enhance the overall characterization of disaster impacts. However, combined methods must have sufficient overlap (i.e., field-based and remote sensing assessments for the same structures) to characterize performance robustly, and care must be taken to maintain quality standards for combined modes. More rapid measurement methods may be extrapolated to a broader sample than the subset of combined modes provided that uncertainty is characterized through comparison of the rapid method with more detailed assessment methods.

Temporal Uncertainties

Temporal uncertainties refer to the timing after a hazard event when a performance assessment is made. While aleatory uncertainties (e.g., natural variations in damage or recovery progression over time) will exist for any infrastructure and social system affected by a hazard event, epistemic uncertainties related to a field team's knowledge of the immediate impacts and long-term response of infrastructure and social systems can be identified and reduced. Although performance assessments are generally performed in the immediate aftermath of a disaster, the timing can vary and is often a function of the nature of the event (e.g., when travel is possible based on safety concerns in the immediate aftermath, resource allocation during response efforts, or survey team coordination). For example, Sutley et al. (2021) assessed damage three days after a tornado; van de Lindt et al. (2018) assessed damage six weeks after a flood; Highfield et al. (2014) assessed damage approximately three months after a hurricane; and Marshall et al. (2019) assessed damage seven weeks after a hurricane. The extent to which assessment delay time creates uncertainty in a performance assessment is difficult to quantify and can change based on the severity and footprint of the hazard. For example, damage sustained from an enhanced Fujita scale (EF) 1 tornado that is localized to the neighborhood level may be cleaned up and repaired within days, whereas damage sustained from a Category 5 hurricane that affects multiple states may remain untouched for weeks or months in some areas. In addition, remnants of damage that remain unrepaired long after initial damage occurs may be worse than initial damage. For example, in a study of tornado impacts, Hamideh et al. (2021) observed

cascading water-related damage during the winter resulting from delayed repairs.

Temporal uncertainty can also be associated with recall bias when residents are asked to provide an assessment of damage after a period of time. Sutley et al. (2021) reassessed damage on a five-point scale more than a year after a flood by asking building occupants questions about the damage they observed during the event. Identical categorizations had been utilized one month after the flood in performance assessments by engineers (van de Lindt et al. 2018, 2020). A comparison of damage states assigned in van de Lindt et al. (2018) and Sutley et al. (2021) showed that it was common for damage states to be higher in the occupant assessment than in the engineering assessment. This difference likely stemmed from temporal, modal, and human factors, in which the latter includes recall bias (i.e., the tendency to remember previous events or experiences inaccurately or omit details) and is described in the following section.

Temporal uncertainty can be associated with cascading or subsequent hazard events that create cumulative damage, especially when damage from a previous event is unrepaired. As reported in Helgeson et al. (2021), damage from a 2016 flood event was still evident in 2018 after a second flood occurred. The evidence of damage from two different events separated by two years was only realized because the field study team had been conducting a longitudinal study and was familiar with the site and sample (van de Lindt et al. 2018, 2020; Sutley et al. 2021; Crawford et al. 2022). The field study team in Helgeson et al. (2021) would have been able to assess total, visible damage without prior field experience and associated knowledge of conditions in the area from previous events. However, they were only able to discern the root cause of damage for a recent flood because of the team's familiarity with the site and their knowledge of a past event, as well as the level of deferred maintenance or disrepair.

Last, temporal uncertainty can be related to variability in repair pace across a community when some areas are already in repair and some have not started repairs when performance assessments take place. This type of temporal uncertainty can be reduced by conducting assessments as early as possible, by combining different modes of assessment, including resident surveys and interviews with expert observation, or by completing follow-up performance assessments.

Human Uncertainties

Human-induced uncertainty through bias and other sources is to be expected and is both epistemic and aleatory in nature. Therefore, it is important that key sources of epistemic uncertainty be recognized, documented, and reduced to the extent possible. Key sources of human-related uncertainty can stem from the design of a data collection instrument, responses from survey participants or stakeholders, or assessments by a researcher.

Cognitive biases can influence a survey instrument and cause participants to reply in an imprecise or inaccurate manner. Data collection instruments, which often aid researchers as guides for observational record keeping or facilitate data collection through interaction with stakeholders (e.g., residents, business operators, emergency personnel) are vulnerable to the application of heuristics and biases. Typical sources of such bias include, but are not limited to, social desirability bias, response bias arising from a lack of randomness in the sample, or a failure to meet sample quotas or weighting factors applied to data. Data collection instruments are developed by humans with their own views and opinions; therefore, there are often subtle biases in the manner in which questions are worded or results are recorded or interpreted (Bloomfield et al. 2016).

Response bias is a general term for a wide range of tendencies for participants and researchers to populate data collection instruments with inaccurate information. This type of bias may be particularly relevant when disaster victims are asked to substantiate or describe physical damage. For stakeholders providing information to researchers, question order bias (i.e., items that appear earlier in a survey can affect responses to later questions or even impact whether later questions are answered at all), social desirability bias (i.e., the tendency to underreport undesirable outcomes and to overreport more desirable attributes or outcomes), and respondent bias/trauma, including post-event time of questioning, all contribute to response bias (Krumpal 2013). There is widespread agreement that disaster-related psychological and psychosocial impacts and recovery are prevalent. However, relative findings differ depending upon hazard type and level of severity, timing of data collection, and theoretical perspectives employed (Tierney 2000).

Both respondents and researchers can contribute to human-related uncertainties in postdisaster reconnaissance. For example, cognitive biases can influence a researcher toward more severe or mild performance assessments. A standard method to quantitatively define damage is desirable and allows for comparison to different geographic areas, building stocks, and hazard characteristics (Rathfon et al. 2013). Standardizing damage classification can be particularly challenging for multihazard events such as combined wind and storm surge during hurricanes. While several component-based damage schemes for hurricane-induced damage have been proposed, ranging from a binary scale (survive-fail) to an ordinal scale with a number of intermediate damage states (i.e., Friedland 2009; Kennedy et al. 2011; Tomiczek et al. 2017, 2020), no assessment methodology has been universally adopted for this type of hazard.

Even given standardized metrics to assess damage, human factors in the field can contribute to assessment biases and uncertainties. These factors include, but are not limited to, experience, educational background, team composition, site familiarity, and field conditions during performance assessments. For example, in a study of assessor bias and the effect of site familiarity on performance assessments, photographic damage data from 44 structures in Galveston, Texas, and the Bolivar peninsula in Texas following Hurricane Ike (2008) were assessed by survey participants using a five-point damage scale. Individually-assessed structures varied by as many as two damage state categories among surveyors, although agreement and overall damage tended to increase when prestorm imagery was provided (Brodersen et al. 2017). When assessing component-based damage, bias may arise based on the severity of damage in the area that a researcher assesses first (e.g., looking at heavily or lightly damaged areas may set the baseline superficially high or low). Furthermore, drawing correlations between impacts and physical damage may be challenging to researchers, causing bias; such challenges may arise when assessing postevent loss of customers to a business and physical damage to a business structure.

Reducing Variation in Performance Assessment

By recognizing modal, temporal, and human factors leading to uncertainty in performance assessment, steps can be taken to reduce variation associated with these factors. Standardizing damage measurements and conducting longitudinal studies will likely lead to reductions in assessment uncertainties collectively and enhance the ability to quantify sources of uncertainty.

Using and assigning common archetypes provides a mechanism for comparing damages across buildings or infrastructure systems

by grouping units with similar attributes. Eisenack et al. (2019) highlighted four major insights on the use of archetypes applicable in the study of hazard and resilience impacts on infrastructure and social systems that can be mapped to community assessment. These insights included the following: (1) identify domains of validity for specific archetypes, (2) ensure that archetypes can be aggregated to assess a community impact, (3) clearly establish levels of abstraction (i.e., differences between essential characteristics of idealized archetypes and characteristics of actual systems), and (4) determine relationships between attribute configurations, theoretical models, and ranges of validity over hazard conditions. The use of archetypes can be either inductive or deductive in nature. For example, the US Geological Survey developed detailed structure type categories based on construction type (e.g., wood frame, steel, reinforced concrete, reinforced masonry) and details such as low- or high-rise construction in order to develop a country-specific inventory for assessing earthquake vulnerability (Jaiswal and Wald 2008). Additionally, Nofal and van de Lindt (2020b) developed 15 building archetypes for investigating building vulnerability to flood hazards. Appropriate use of archetypes should avoid overgeneralization by identifying reappearing but nonuniversal patterns that hold for well-defined subsets of cases. There is a need for quality criteria and practices in the selection and use of archetypes—which may range from defining built infrastructure to organizational structures and hazard types—to guide the design of theoretically rigorous and practical archetype analyses. Challenges in the use of archetypes include, but are not limited to, demonstration of the validity of the analysis, delineation of archetype boundaries, and selection of appropriate attributes to define them. For example, varying hazard intensity–system response relationships may exist among different archetypes. Therefore, grouping archetypes or inaccurately delineating archetype boundaries can increase the uncertainty in empirical fragility modeling. When possible, archetypes should be determined or planned for in designing a postevent assessment sample.

A standardized performance assessment module with associated field protocols for disaggregating archetypes and determining the performance of affected systems can be a breakthrough for identifying uncertainties over multiple field-based performance assessments. Proper equipment (e.g., safety gear, horizontal levels, batteries for cameras or specialized equipment), preparation (e.g., safety briefings, preparations for working in conditions with no power, water, or restrooms), and documentation can contribute to the success of data collection, interpretation, and dissemination (Peraza et al. 2014; Roueche et al. 2019, 2020, 2021). Designing and adopting valid and consistent measurements will provide damage data that are directly comparable to published and publicly available data based on the same measurements. For example, the USDA Food Security Survey Module has been incorporated into national level surveys (Pérez-Escamilla et al. 2004) as well as independent studies, because researchers can use a core module for required measurements and optional modules that can accommodate varying needs in field studies (Bickel et al. 2000). Using the USDA compatible survey modules has enabled researchers to test the validity and reliability of their own data by comparing them with national statistics and shared data from other researchers. In addition, subsidiary damage data offer a comparable layer for performance assessment outcomes from field studies. For example, pre- and postevent property tax records can provide an indirect measurement of damage by decreased appraised values (Zhang and Peacock 2009; Hamideh et al. 2018). Google street view and satellite images can also be employed to check preevent physical conditions related to nondisaster damage by deferred maintenance and housing abandonment (van de Lindt et al. 2018). Last, follow-up

assessments and longitudinal studies can enable comparison of assessments made at different points in time to evaluate the optimal timing for measuring initial damage and can provide a means of quantifying modal, temporal, and human-induced uncertainty.

Discussion and Future Work

This paper introduced a framework for identifying and characterizing uncertainties in postdisaster performance assessments. The nature of an event, the type of system being assessed, the scale of assessment, and mode, timing, and human factors all contribute to uncertainty in performance assessment. The study did not attempt to quantify sources of uncertainty but rather articulated the scope of potential sources of uncertainty and the interdependencies among them. The framework is an attempt to isolate each source. However, many sources of uncertainty inherently overlap and may be correlated. For example, cognitive biases are affected by the mode and timing of performance assessment. More data are required for robust quantification of the various sources of uncertainty across hazards, events, and geographies and their correlations. It is critical to measure how each source independently influences performance assessment and how sources combine to jointly influence performance assessment. A method for addressing correlations is also needed to identify factors that may increase other sources of epistemic uncertainty.

Uncertainty quantification for performance assessments requires a significant and coordinated effort from researchers from multiple hazards and fields. A foundation for quantifying these uncertainties has been utilized by FEMA P-695 (FEMA 2009), which quantified a series of β factors capturing uncertainty for seismic hazard analysis. Four independent sources of uncertainty were identified (seismic records, design requirements, test data, and modeling) to estimate total structural collapse uncertainty. Record-to-record uncertainty was found using a lognormal distribution, while uncertainty in design requirements, test data, and modeling were determined using expert opinions based on the completeness and robustness of the uncertainty factor and confidence in the basis of these factors (FEMA 2009). A similar approach could be employed to inform hazard, damage, and recovery models such that fragility functions and subsequent use of such functions provide more accurate depictions of system performance. For example, comparisons for the same structure of performance assessments based on field surveys, remote sensing, and/or resident surveys may identify trends in modal uncertainties, allowing for similar uncertainty factors to be determined based on the completeness, robustness, or confidence of a given performance assessment data set. Longitudinal studies of the same sample (e.g., Sutley et al. 2021; Helgeson et al. 2021) can provide insight into parameters characterizing temporal effects on damage assessment, while assessments by multiple researchers of the same sample (or sample subset) can be analyzed to quantify inherent variability due to human uncertainties. Identification and quantification of these sources of uncertainty may also lead to future research on strategies to reduce uncertainty.

Deliberate uncertainty quantification can be time consuming and challenging during already time- and resource-intensive reconnaissance efforts following disasters. Therefore, it may not always be feasible to document uncertainty parameters during field reconnaissance, given that field teams prioritize collection of perishable performance data for a sufficient sample. A subset of a sample may be analyzed in the field or evaluated based on imagery and other data collected during reconnaissance efforts using photography or emerging street-view image capture technology

(e.g., Roueche et al. 2021). However, it is important for teams to recognize trade-offs and to achieve a balance between uncertainty quantification and robust data collection. Efforts to characterize modal uncertainties from photographic and field-based assessments may allow teams to perform uncertainty characterization after returning from field efforts focused primarily on data collection.

Recommendations

To fully characterize the modal, temporal, and human uncertainties affecting postdisaster performance assessments, a comprehensive data-synthesis effort spanning hazards, geographies, and systems is required.

- Field teams must seek to identify potential modal, temporal, and human sources of uncertainty in performance assessments while taking steps to minimize the effects of these uncertainties on assessed samples by, for example, using combined modes of performance assessment.
- When possible, assessment uncertainty should be quantified and communicated in published data archives, with multiple researchers participating in quality assurance processes for individual records to identify error rates and reduce uncertainties (e.g., Roueche et al. 2021). Objective damage state classifications standardized across scale and granularity (i.e., number of damage states) may facilitate replicability as well as cross-event comparison, although targeted efforts utilizing researcher-developed performance assessment schemes can be useful in answering specific questions.
- Efforts to quantify uncertainty in field-based performance assessments can allow for uncertainty propagation into data-driven modeling so that researchers can provide more informed guidance to practitioners and community stakeholders. Such efforts are only possible with the support of collective, collaborative community practices toward the replicability and reproducibility of research processes.
- Data sharing and reuse are essential to allow quantification of performance assessment variation for different events and geographies. Ongoing initiatives to promote data sharing and reuse include those taken by Structural Extreme Events Reconnaissance (StEER), DesignSafe, and other research groups or repositories for damage data following suggested replication standards (data quality assurance and quality control, e.g., Roueche et al. 2019, 2020) as well as the Natural Hazards Center Weather Ready Research Award Program (e.g., First et al. 2022) and NHERI CONVERGE training modules (e.g., Adams et al. 2021; Evans et al. 2021). These efforts can enable repeated performance assessments for the quantification of human uncertainty as well as investigation into modal and temporal uncertainties, based on data availability.
- Last, intentionally designed longitudinal studies or follow-up visits to the same sites at different points in time after a disaster will glean new data and further advance the identification and quantification of uncertainty in field assessments.

Data Availability Statement

Some or all data, models, or code generated or used during the study are available from the corresponding author by request, including framework, performance assessment data from the longitudinal Lumberton study, and performance assessment data collected following Hurricane Ike.

Acknowledgments

This research was conducted as part of the NIST Center of Excellence for Risk-Based Community Resilience Planning under Cooperative Agreement 70NANB20H008 between the NIST and Colorado State University. The views expressed in this article are those of the authors and do not necessarily represent the opinions or views of NIST or the US Department of Commerce. The authors would like to thank Professor John van de Lindt, Professor Andre Barbosa, and Dr. Omar Nofal, who contributed valuable discussions informing the content of this article.

References

- Adams, R. M., C. M. Evans, and L. Peek. 2021. "Broader ethical considerations for hazards and disaster researchers." *NHERI CONVERGE*. Accessed May 17, 2022. <https://converge-training.colorado.edu/courses/broader-ethical-considerations-for-hazards-and-disaster-researchers/>.
- Bickel, G., M. Nord, C. Price, W. Hamilton, and J. Cook. 2000. *Guide to measuring household food security*. Alexandria, VA: USDA, Food and Nutrition Service.
- Bloomfield, R., M. W. Nelson, and E. Soltes. 2016. "Gathering data for archival, field, survey, and experimental accounting research." *J. Accounting Res.* 54 (2): 341–395. <https://doi.org/10.1111/1475-679X.12104>.
- Brodersen, S., S. Hamideh, W. G. Peacock, and D. Gu. 2017. "Evaluating the accuracy of post-disaster damage assessments using pre-impact images of residential structures." Accessed December 14, 2021. <https://hazards.colorado.edu/workshop/2017/abstract/poster-session#evaluating-the-accuracy-of-post-disaster-damage-assessments-using-pre-impact-images-of-residential-structures>.
- Brown, D., K. Suito, M. Liu, R. Spence, E. So, and M. Ramage. 2012. "The use of remotely sensed data and ground survey tools to assess damage and monitor early recovery following the 12.5.2008 Wenchuan earthquake in China." *Bull. Earthquake Eng.* 10 (3): 741–764. <https://doi.org/10.1007/s10518-011-9318-7>.
- Eisenack, K., S. Villamayor-Tomas, G. Epstein, C. Kimmich, N. Magliocca, D. Manuel-Navarrete, C. Oberlack, M. Roggero, and D. Sietz. 2019. "Design and quality criteria for archetype analysis." *Ecol. Soc.* 24 (3): 6. <https://doi.org/10.5751/ES-10855-240306>.
- Evans, C. M., R. M. Adams, and L. Peek. 2021. "Collecting and sharing perishable data." *NHERI CONVERGE*. Accessed May 17, 2022. <https://converge-training.colorado.edu/courses/collecting-and-sharing-perishable-data/>.
- FEMA. 2009. *Quantification of building seismic performance factors*. FEMA P-695. Washington, DC: Applied Technology Council, FEMA.
- First, J. M., K. Ellis, and S. Strader. 2022. "Examining public response and climate conditions during overlapping tornado and flash flood warnings." *Weather Ready Report WR1. Natural Hazards Center*. Accessed May 17, 2022. <https://hazards.colorado.edu/weather-ready-research/examining-public-response-and-climate-conditions-during-overlapping-tornado-and-flashflood-warnings>.
- Friedland, C. 2009. "Residential building damage from hurricane storm surge: Proposed methodologies to describe, assess, and model building damage." Ph.D. thesis, Dept. of Civil and Environmental Engineering, Louisiana State Univ.
- Gaillard, J. C., and L. Peek. 2019. "Disaster-zone research needs a code of conduct." *Nature* 575 (7783): 440–442. <https://doi.org/10.1038/d41586-019-03534-z>.
- Groves, R. M., F. J. Fowler Jr., M. P. Couper, J. M. Lepkowski, E. Singer, and R. Tourangeau. 2011. *Survey methodology*. New York: Wiley.
- Hamideh, S., M. Freeman, S. Sritharan, and J. Wolseth. 2021. *Damage, dislocation, and displacement after low-attention disasters: Experiences of renter and immigrant households*. Natural Hazards Center Quick Response Grant Rep. Series No. 304. Boulder, CO: Natural Hazards Center, Univ. of Colorado Boulder.
- Hamideh, S., W. G. Peacock, and S. Van Zandt. 2018. "Housing recovery after disasters: Primary versus seasonal and vacation housing markets in coastal communities." *Nat. Hazards Rev.* 19 (2): 04018003. [https://doi.org/10.1061/\(ASCE\)NH.1527-6996.0000287](https://doi.org/10.1061/(ASCE)NH.1527-6996.0000287).

- Helgeson, J., S. Hamideh, and E. J. Sutley, eds. 2021. *Community resilience-focused technical investigation of the 2016 Lumberton, NC Flood: Community impact and recovery following successive flood events*. NIST Special Publication 1230-3. Gaithersburg, MD: NIST.
- Highfield, W. E., W. G. Peacock, and S. Van Zandt. 2014. "Mitigation planning: Why hazard exposure, structural vulnerability, and social vulnerability matter." *J. Plann. Educ. Res.* 34 (3): 287–300. <https://doi.org/10.1177/0739456X14531828>.
- Jaiswal, K. S., and D. J. Wald. 2008. *Creating a global building inventory for earthquake loss assessment and risk management*. US Geological Survey Open-File Rep. No. 2008-1160. Denver: USGS.
- Jones, S. R., S. Carley, and M. Harrison. 2003. "An introduction to power and sample size estimation." *Emergency Med. J.* 20 (5): 453–458. <https://doi.org/10.1136/emj.20.5.453>.
- Kennedy, A., S. Rogers, A. Sallenger, U. Gravois, B. Zachry, M. Dosa, and F. Zarama. 2011. "Building destruction from waves and surge on the Bolivar Peninsula during Hurricane Ike." *J. Waterway, Port, Coastal, Ocean Eng.* 137 (3): 132–141. [https://doi.org/10.1061/\(ASCE\)WW.1943-5460.0000061](https://doi.org/10.1061/(ASCE)WW.1943-5460.0000061).
- Khajwal, A. B., and A. Noshadran. 2021. "An uncertainty-aware framework for reliable disaster damage assessment via crowdsourcing." *Int. J. Disaster Risk Reduct.* 55 (Mar): 102110. <https://doi.org/10.1016/j.ijdrr.2021.102110>.
- Krumpal, I. 2013. "Determinants of social desirability bias in sensitive surveys: A literature review." *Qual. Quantity* 47 (4): 2025–2047. <https://doi.org/10.1007/s11135-011-9640-9>.
- Marshall, J., et al. 2019. "StEER–Hurricane Dorian: Field assessment structural team (FAST-1) Early Access Reconnaissance Report (EARR)." *DesignSafe-CI*. https://aub.ie/Steer_dorian_earr.
- Martínez-Mesa, J., D. A. González-Chica, J. L. Bastos, R. R. Bonamigo, and R. P. Duquia. 2014. "Sample size: How many participants do I need in my research?" *Anais Brasileiros Dermatologia* 89 (4): 609–615. <https://doi.org/10.1590/abd1806-4841.20143705>.
- Marvi, M. T. 2020. "A review of flood damage analysis for a building structure and contents." *Nat. Hazards* 102 (3): 967–995. <https://doi.org/10.1007/s11069-020-03941-w>.
- Minor, J. E., J. R. McDonald, and K. C. Mehta. 1977. *The tornado: An engineering-oriented perspective*. NOAA Technical Memorandum ERL NSSL-82. Norman, OK: National Weather Service.
- Nofal, O. M., and J. W. van de Lindt. 2020a. "Understanding flood risk in the context of community resilience modeling for the built environment: Research needs and trends." *Sustainable Resilient Infrastruct.* 7 (3): 171–187. <https://doi.org/10.1080/23789689.2020.1722546>.
- Nofal, O. M., and J. W. van de Lindt. 2020b. "Minimal building flood fragility and loss function portfolio for resilience analysis at the community level." *Water* 12 (8): 2277. <https://doi.org/10.3390/w12082277>.
- Peraza, D. B., W. L. Coulbourne, and M. Griffith. 2014. *Engineering investigations of hurricane damage: Wind versus water*. Reston, VA: ASCE.
- Pérez-Escamilla, R., A. M. Segall-Correia, L. Kurdian Maranha, M. de Fátima Archanjo Sampaio, L. Marín-León, and G. Panigassi. 2004. "An adapted version of the US Department of Agriculture Food Insecurity Module is a valid tool for assessing household food insecurity in Campinas, Brazil." *J. Nutr.* 134 (8): 1923–1928. <https://doi.org/10.1093/jn/134.8.1923>.
- Rathfon, D., R. Davidson, J. Bevington, A. Vicini, and A. Hill. 2013. "Quantitative assessment of post-disaster housing recovery: A case study of Punta Gorda, Florida, after Hurricane Charley." *Disasters* 37 (2): 333–355. <https://doi.org/10.1111/j.1467-7717.2012.01305.x>.
- Rathje, E., et al. 2017. "DesignSafe: New cyberinfrastructure for natural hazards engineering." *Nat. Hazards Rev.* 18 (3): 06017001. [https://doi.org/10.1061/\(ASCE\)NH.1527-6996.0000246](https://doi.org/10.1061/(ASCE)NH.1527-6996.0000246).
- Roueche, D., et al. 2021. "Field assessment structural teams: FAST-1, FAST-2, FAST-3." In *StEER–Hurricane Laura*. Salem, OR: NHERI. <https://doi.org/10.17603/ds2-dha4-g845>.
- Roueche, D. B., T. Kijewski-Correia, K. Mosalam, D. O. Prevatt, and I. Robertson. 2019. *Virtual assessment structural team (VAST) handbook: Data enrichment and quality control (DE/QC) for US windstorms, Version 2.0*. StEER. Notre Dame, IN : Structural Extreme Events Reconnaissance.
- Roueche, D. B., T. Kijewski-Correia, K. Mosalam, D. O. Prevatt, and I. Robertson. 2020. *Virtual assessment structural team (VAST) handbook: Data enrichment and quality control (DE/QC) for StEER earthquake rapid evaluation form, Version 1.0*. StEER. Notre Dame, IN : Structural Extreme Events Reconnaissance.
- Roueche, D. B., D. O. Prevatt, and F. T. Lombardo. 2018. "Epistemic uncertainties in fragility functions derived from post-disaster damage assessments." *ASCE-ASME J. Risk Uncertainty Eng. Syst. Part A: Civ. Eng.* 4 (2): 04018015. <https://doi.org/10.1061/AJRUA6.0000964>.
- Crawford, P. S., J. Mitrani-Reiser, E. J. Sutley, T. Q. Do, T. Tomiczek, O. M. Nofal, J. M. Weigand, M. Watson, J. W. van de Lindt, and A. J. Graettinger. 2022. "Measurement approach to develop flood-based damage fragilities for residential buildings following repeat inundation events." *ASCE-ASME J. Risk Uncertainty Eng. Syst. Part A: Civ. Eng.* 8 (2): 04022019. <https://doi.org/10.1061/AJRUA6.0001219>.
- Sutley, E. J., M. K. Dillard, and J. W. van de Lindt, eds. 2021. *Community resilience focused technical investigation of the 2016 Lumberton, NC Flood: Community recovery one year later*. NIST Special Publication 1230-2. Gaithersburg, MD: NIST.
- Tierney, K. 2000. "Controversy and consensus in disaster mental health research." *Prehospital Disaster Med.* 15 (4): 55–61. <https://doi.org/10.1017/S1049023X00025292>.
- Tomiczek, T., A. Kennedy, Y. Zhang, M. Owensby, M. E. Hope, N. Lin, and A. Flory. 2017. "Hurricane damage classification methodology and fragility functions derived from Hurricane Sandy's effects in coastal New Jersey." *J. Waterway, Port, Coastal, Ocean Eng.* 143 (5): 04017027. [https://doi.org/10.1061/\(ASCE\)WW.1943-5460.0000409](https://doi.org/10.1061/(ASCE)WW.1943-5460.0000409).
- Tomiczek, T., K. O'Donnell, K. Furman, B. Webbmartin, and S. Scyphers. 2020. "Rapid damage assessments of shorelines and structures in the Florida Keys after Hurricane Irma." *Nat. Haz. Rev.* 21 (1): 05019006. [https://doi.org/10.1061/\(ASCE\)NH.1527-6996.0000349](https://doi.org/10.1061/(ASCE)NH.1527-6996.0000349).
- van de Lindt, J. W., et al. 2018. *Community resilience-focused technical investigation of the 2016 Lumberton, North Carolina flood: Multi-disciplinary approach*. NIST Special Publication 1230. Gaithersburg, MD: NIST.
- van de Lindt, J. W., et al. 2020. "Community resilience-focused technical investigation of the 2016 Lumberton, North Carolina, Flood: An interdisciplinary approach." *Nat. Hazards Rev.* 21 (3): 04020029. [https://doi.org/10.1061/\(ASCE\)NH.1527-6996.0000387](https://doi.org/10.1061/(ASCE)NH.1527-6996.0000387).
- Zhang, Y., and W. G. Peacock. 2009. "Planning for housing recovery? Lessons learned from Hurricane Andrew." *J. Am. Plann. Assoc.* 76 (1): 5–24. <https://doi.org/10.1080/01944360903294556>.