Probabilistic Flood Damage Cost Estimation for Residential Buildings Using Bayesian Information Updating

Yitong Li[1] and Jie Gong[2]

[1]Department of Civil and Environmental Engineering, University of Alberta, Canada
[2]Department of Civil and Environmental Engineering, Rutgers University, USA
yitong2@ualberta.ca, jg931@soe.rutgers.edu

Abstract – Flood damage estimation for residential buildings is a critical component of post-disaster recovery planning, budgeting, and construction management. Conventional depth–damage curves, which remain the standard tool in flood risk assessment, rely on deterministic and monotonic assumptions that fail to capture the considerable variability observed in real-world flood losses. As a result, these methods often provide limited support for uncertainty-aware cost estimation and decision-making. This study proposes a probabilistic framework for residential flood damage cost estimation based on Bayesian inference and Markov Chain Monte Carlo simulation. Building damage is modeled as a percentage of replacement value and represented using beta distributions. The framework sequentially updates damage distributions as new post-flood observations become available and propagates uncertainty into cost estimates through Monte Carlo simulation. The proposed approach is demonstrated through case studies in Manville and Cranford, New Jersey, using residential flood damage data from Hurricane Ida. Results show that posterior predictive cost estimates consistently encompass observed damage costs within credible intervals, while traditional depth–damage curve-based estimates exhibit limited agreement with observations. The findings highlight the value of probabilistic, uncertainty-aware modelling through improved representation of cost uncertainty, thereby providing useful insights for residential construction cost estimation and post-disaster recovery management.

Keywords –

Flood damage cost estimation; Probabilistic modelling; Markov Chain Monte Carlo simulation; Uncertainty quantification

1 Introduction

Flooding is among the most impactful natural hazard worldwide, affecting the built environment and residential buildings that underpin community lifelines. As damage to residential buildings directly drives recovery needs and resource demands, accurate damage estimation is essential for construction-related decision-making, including defining repair scopes, allocating contingencies, and evaluating mitigation alternatives such as home elevation or buyout programs [1][2]. In practice, however, even small errors in translating physical flood characteristics into monetary losses can distort cost estimates. Such uncertainty complicates budget approval, delays reconstruction efforts, and weakens the economic justification for long-term residential flood mitigation investments [3].

The most commonly used approach for estimating flood damage to residential buildings is the depth–damage curve (DDC), which relates floodwater depth to an expected percentage of structural or contents loss. In the United States, standardized residential DDCs developed by the U.S. Army Corps of Engineers (USACE) and the Federal Emergency Management Agency (FEMA) are widely applied in housing damage assessments, insurance analyses, and construction cost estimation workflows [4][5]. Their simplicity makes them particularly attractive during early post-disaster stages, when rapid estimates are required and detailed building condition data are limited.

Despite their widespread adoption, existing studies have identified several critical limitations of DDC-based methods when applied to cost estimation. Specifically, standard DDCs are typically derived from aggregated historical data or expert judgment and represent an "average" response for broad residential building categories [6]. This approach doesn't consider the variability in residential construction practices, such as foundation type, framing systems, finish materials, and construction quality [3]. As a result, DDCs are typically used for high-level cost estimation, with limited ability to account for cost variability across individual homes.

Moreover, most conventional DDCs are built on the assumption that flood damage is a monotonic function of water depth, meaning that expected losses increase continuously as water levels rise [6]. However, recent study reveals that observed flood losses are often not monotonic and instead follow a bimodal distribution where losses are concentrated at the minimal and maximal extremes [6]. As a result, residential flood losses at a given depth exhibit variation, with similar inundation levels leading to different damage levels. From a construction management perspective, such systematic bias can influence residential recovery budget allocation and hinder effective sequencing and resourcing of repair activities, which highlights the importance of uncertainty-aware estimation approaches [7][3][8].

Although multivariate damage models that incorporate additional variables such as flood duration or building condition have been proposed [9][10][11], their extensive data requirements and limited transferability restrict their use in residential contexts. As a result, there is a need for probabilistic damage estimation frameworks tailored to residential buildings that can represent uncertainty while remaining practical for construction-oriented cost estimation. Probabilistic approaches, with particular emphasis on Bayesian inference–based approaches, are capable of addressing these limitations by producing distributions of cost estimates rather than single-point values. By combining prior knowledge of building information with post-flood damage observations, these methods enable uncertainty quantification and reliable cost estimation.

This study proposes a Bayesian framework that integrates residential building attributes with Markov Chain Monte Carlo (MCMC) simulation for building cost estimation following flood impacts. Rather than assuming a monotonic depth–damage relationship, building damage is modelled independently at each inundation depth. For each depth and residential building type (e.g., two-story buildings with basements), the proportion of building damage, bounded between zero and one, is modelled using a beta distribution. MCMC is then employed to update these distributions as new damage observations become available. The resulting damage distributions are subsequently converted into cost distributions by incorporating building replacement values. Through this updating process, cost estimates are continuously refined while uncertainty is explicitly quantified, thereby supporting more robust budgeting, contingency planning, and construction management decision-making in flood-prone residential areas.

2 Literature Review

One of the most widely used approaches for estimating flood damage to residential buildings is the DDC, which relates inundation depth directly to expected economic loss. Although there is no universally accepted framework for assessing flood damage in urban build environments, most existing damage models rely on DDCs because of their conceptual simplicity and ease of implementation [12]. In practice, these curves are frequently used to support cost estimation, post-disaster budgeting, and recovery planning, particularly during early-stage damage assessments when detailed building inspections are not yet available.

DDCs are typically constructed under the assumption that flood damage increases monotonically with water depth. In practice, however, this assumption is frequently violated. Recent analyses of more than two million National Flood Insurance Program (NFIP) claims show that observed losses are often non-monotonic and instead exhibit a bimodal distribution, with damage clustered near both minimal and maximal loss levels [6]. From a construction management perspective, this behaviour complicates repair cost estimation, as similar flood depths can lead to vastly different repair needs, ranging from minor finishes replacement to full system and structural rehabilitation [3]. Studies further indicate that conventional DDCs tend to overestimate damage from shallow flooding while underestimating losses associated with deep inundation. In many cases, their predictive performance is sufficiently limited that a simple mean of observed damages outperforms the curve-based estimates [6]. Another limitation of traditional DDCs is their reliance on a “one-to-one” relationship between depth and damage, where each inundation depth is associated with a single representative damage value, typically expressed as the median or modal loss. Empirical evidence suggests that flood damage at a given depth is highly variable and poorly described by central tendency measures. For example, at one foot of inundation, the middle 50% of observed claims may span from approximately 4% to 30% of building value [6].

The inability to explicitly model uncertainty is widely recognized as a major bottleneck in flood risk assessment, as deterministic approaches typically produce single-point estimates that obscure the inherent variability and stochastic nature of flood damage processes. To address this limitation, numerous studies have developed multivariable flood damage models that extend beyond depth–damage formulations by incorporating additional factors, such as flood duration, the timing and quality of early warnings, building quality, and residents’ socioeconomic characteristics [9][10][11]. A range of statistical techniques has been applied, including random forest models [13], tree-structured models [14], and empirically derived equations [15]. Despite being useful, these approaches require large, high-quality datasets, and limitations in data availability and consistency can introduce bias [16]. Moreover, these multivariable models are often highly site-specific, limiting their transferability across regions.

A growing body of research has shifted toward probabilistic flood vulnerability modeling to improve the reliability and transparency of flood damage cost estimates [8][3]. Probabilistic frameworks, particularly those grounded in Bayesian inference, allow uncertainty in depth–damage relationships to be explicitly represented and propagated into cost estimates. By combining prior knowledge with limited empirical data, Bayesian methods yield robust, observation-informed cost estimates while explicitly quantifying uncertainty [17][18]. When empirical damage and cost data are scarce, prior information, such as established damage functions or historical repair cost data, can guide early-stage cost estimation and budget planning. As additional building-specific damage observations become available, estimates can be progressively refined to reflect observed cost estimates. Rather than producing single-point damage or cost estimates, Bayesian models generate probability distributions of expected losses, with credible intervals that reflect data availability. For residential building categories with small sample sizes, wider credible intervals reflect increased uncertainty, whereas narrower intervals correspond to greater confidence under improved data availability.

3 Methodology

This study proposes a Bayesian framework that integrates residential building attributes with MCMC simulation for building cost estimation following flood impacts. The detailed steps are illustrated in Figure 1.

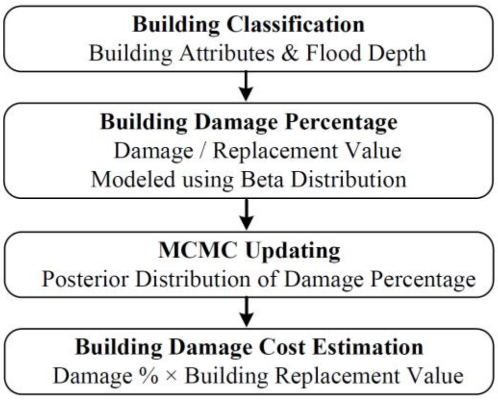

Figure 1. Methodology framework.

First, buildings are classified according to the following attributes: occupancy type, which describes the building use (e.g., single-family residence or two- to four-unit residential building); number of stories (e.g., one-story or two-story structures); basement type (e.g., no basement, finished basement, or crawlspace); and flood depth, defined as the difference between the flood elevation and the building's first-floor elevation.

Next, beta distributions are used to model the distribution of building damage percentage. This percentage is used as the damage metric because depth–damage curves, which remain the standard tool in flood damage assessment, express damage as the percentage of the estimated building damage value to the total building replacement value. Modelling damage in percentage terms provides a normalized representation of loss that is independent of absolute building value and allows direct compatibility with existing depth–damage–based cost estimation workflows. The resulting damage ratios are subsequently converted into monetary losses by incorporating building replacement values.

The beta distribution is selected to model building damage percentage because this metric is defined as a proportion bounded between zero and one. Unlike unbounded distributions, the beta distribution ensures physically meaningful damage ratios and can flexibly represent a wide range of empirical damage patterns, including skewed and highly variable loss behaviour observed in flood events. In addition, the beta distribution is well suited for Bayesian inference and MCMC updating, enabling progressive refinement of damage estimates as new observations become available.

For a given inundation depth and residential building class, observed damage is expressed as a percentage,

$$y_i = \frac{estimated\ building\ damage\ value_i}{building\ replacement\ value_i} \tag{1}$$

where $y_i \in (0,1)$, with 0 indicating no damage and 1 indicating total loss. These observations are modelled using a beta likelihood, which inherently enforces the physical bounds of building damage percentages. The parameters of the beta distribution, $\theta = (\alpha, \beta)$, are updated using Bayes' rule as new building damage percentage observations D_{new} become available:

$$p(\alpha, \beta | D_{old}, D_{new}) \propto p(D_{new}|\alpha, \beta) \times p(\alpha|D_{old}) \times p(\beta|D_{old}) \tag{2}$$

with the likelihood given by:

$$p(D_{new}|\alpha, \beta) = \prod_{i=1}^{n} Beta\,(y_i|\alpha, \beta). \tag{3}$$

The parameters α and β are assigned as independent gamma prior distributions,

$$p(\alpha|D_{old}) = Gamma(\alpha|a_1, b_1), \tag{4}$$
$$p(\beta|D_{old}) = Gamma(\beta|a_2, b_2), \tag{5}$$

which can also be specified as informative or weakly informative depending on data availability. When substantial prior knowledge of residential building damage exists, informative priors can be selected to reflect established damage behaviour. In data-scarce contexts, weakly informative priors (e.g.,

$\mathrm{Gamma}(0.001, 0.001)$) can be used minimize prior influence.

MCMC sampling is used to approximate the posterior distribution of building damage percentage. Each MCMC draw represents a plausible realization of damage percentage for the given depth and building class. These sampled damage percentages are subsequently converted into building damage cost samples by multiplying them by the corresponding building replacement value. This transformation propagates uncertainty from damage percentage directly into cost estimation, resulting in a probabilistic representation of residential flood damage cost estimation.

4 Results

To demonstrate the applicability of the proposed framework, a case study was conducted using residential flood damage data from Hurricane Ida. Two townships in New Jersey—Manville and Cranford—were selected due to their repeated exposure to riverine flooding and their history of residential flood damage. Both communities experienced inundation during Hurricane Ida, resulting in a large number of reported residential losses. The analysis focuses on a representative residential building class: single-family dwellings with two stories and basements, evaluated at an inundation depth of 1 ft above the first-floor elevation. Building damage percentage data for this building class were obtained from the OpenFEMA National Flood Insurance Program (NFIP) dataset [19], which provides publicly available, claim-level flood damage information derived from post-event assessments.

Figures 2 and 3 illustrate three posterior predictive distributions (mean, lower 95%, upper 95%) of building damage percentage for Manville and Cranford, respectively, obtained through Bayesian updating using MCMC simulation. Weakly informative gamma priors were assigned to the beta shape parameters α and β to represent minimal prior knowledge of building damage behaviour. In the figures, the histogram shows the empirical distribution of observed damage percentages. The three curves represent beta distributions corresponding to the posterior mean and the lower and upper 95% credible bounds of the shape parameters, derived from the updated MCMC samples. While only three representative distributions are shown for clarity, they reflect two sources of uncertainty implied by the posterior: variability within each beta distribution, which represents the range of possible damage percentages for a given set of shape parameters, and variability across beta distributions arising from uncertainty in the shape parameters themselves. Together, these distributions capture the range of plausible damage behaviours, including cases with density concentrated near both low and high damage extremes, demonstrating the flexibility of the beta model to represent outcomes from minimal damage to near-total loss.

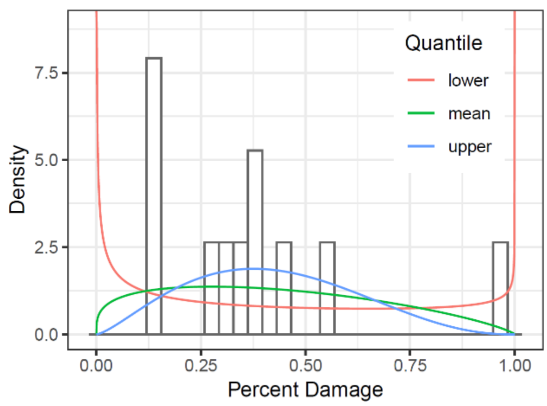

Figure 2. Posterior predictive distributions (mean, lower 95%, upper 95%) of residential building damage percentage for Manville, NJ.

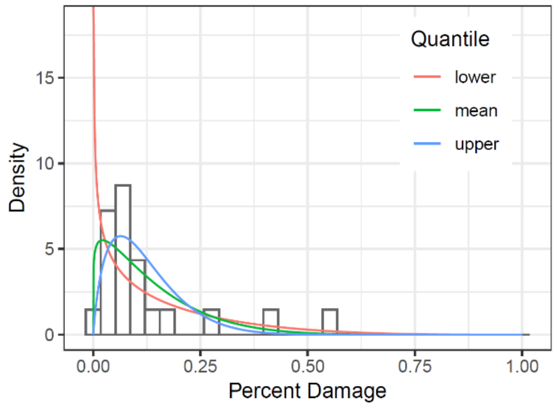

Figure 3. Posterior predictive distributions (mean, lower 95%, upper 95%) of residential building damage percentage for Cranford, NJ.

For practical cost estimation, these damage percentages can be directly translated into monetary losses by multiplying them by the building replacement value. As a result, each beta distribution corresponds to a distribution of potential damage costs rather than a single estimate. This enables probabilistic cost estimation, allowing practitioners to quantify not only expected losses (e.g., using the mean) but also the associated uncertainty (e.g., using credible intervals), which is critical for risk-informed decision-making, budgeting, and resilience planning.

Figure 4 compares estimated residential flood damage costs against observed damage costs for Manville, including results from both the proposed probabilistic approach and traditional depth–damage curves. For the probabilistic approach, the vertical lines represent the 95% credible intervals of damage cost derived from posterior predictive distributions of building damage percentage, with blue circles indicating the corresponding mean estimates. These cost estimates are obtained by transforming the predicted damage percentages into monetary values using building replacement costs. The diagonal line represents perfect agreement between estimated and observed costs. Intersections between the credible intervals and the diagonal line indicate that the observed damage cost is captured within the uncertainty bounds of the proposed method. The results show that a substantial portion of the observed costs fall within the estimated credible intervals, demonstrating the ability of the probabilistic framework to represent uncertainty and capture observed outcomes.

In contrast, the red triangles represent deterministic cost estimates obtained from traditional depth–damage curves. These point estimates are frequently farther from the diagonal line and, unlike the probabilistic intervals, do not account for uncertainty. Comparing the mean probabilistic estimates (blue circles) with the deterministic estimates (red triangles), the probabilistic approach generally yields estimates closer to the diagonal, indicating improved agreement with observed costs. While a few cases show the deterministic estimates performing comparably or slightly better, this is likely due to limited data informing the posterior in those instances. Overall, the results highlight the advantage of the proposed method in providing uncertainty-aware estimates while maintaining improved accuracy relative to traditional approaches.

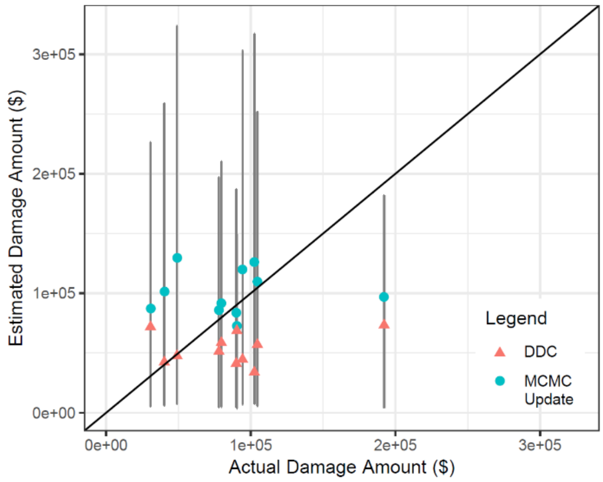

Figure 4. Observed versus estimated residential flood damage costs for Manville, NJ: probabilistic (95% credible intervals) and deterministic comparisons.

Figure 5 compares estimated and observed residential flood damage costs for Cranford, NJ, including both the proposed probabilistic approach and traditional depth–damage curve estimates. Similar to the Manville results, the posterior predictive cost estimates are represented by 95% credible intervals, with mean estimates shown as blue circles. A large proportion of these intervals intersect the diagonal line, indicating that the observed costs are well captured within the uncertainty bounds of the proposed method. In contrast, the deterministic depth–damage curve estimates (shown as red triangles) are more frequently distant from the diagonal.

With a larger dataset available for Cranford, the posterior distributions of the model parameters are better constrained, resulting in more accurate and consistent cost estimates. This is reflected in the mean probabilistic estimates, which are generally closer to the diagonal line than the deterministic estimates. The results demonstrate that, with sufficient data, the proposed probabilistic framework not only captures uncertainty but also provides improved accuracy compared to traditional depth–damage approaches.

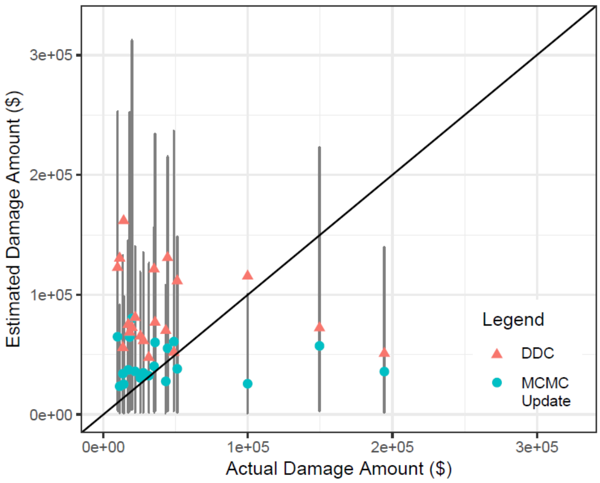

Figure 5. Observed versus estimated residential flood damage costs for Cranford, NJ: probabilistic (95% credible intervals) and deterministic comparisons.

Overall, the Manville and Cranford case studies demonstrate the practical applicability of the proposed Bayesian framework for residential flood damage cost estimation. Across both communities, the probabilistic approach consistently captures observed damage costs within estimated uncertainty ranges, while traditional depth–damage curve–based estimates show more limited agreement with observed outcomes. Additionally, with increased data availability in Cranford, the probabilistic approach provides not only reliable uncertainty quantification but also improved accuracy in mean cost estimates compared to the deterministic method. These results highlight the value of uncertainty-aware modelling for residential flood damage cost estimation.

5 Conclusion

This study presents a probabilistic framework for estimating residential flood damage costs that explicitly accounts for uncertainty in building damage behaviour. By modelling building damage percentage using a Bayesian inference through MCMC simulation, the proposed method moves beyond conventional deterministic depth–damage curves that rely on single-point estimates and monotonic assumptions. Modelling damage independently at each inundation depth allows the framework to represent multiple plausible damage states for the same flood level, better reflecting observed residential flood loss patterns.

The case studies conducted in Manville and Cranford, New Jersey, using post-event data from Hurricane Ida, demonstrate the practical applicability and advantages of the proposed approach. Across both communities, posterior predictive cost estimates consistently encompassed observed residential damage costs within their credible intervals, whereas traditional depth–damage curve–based estimates showed limited agreement with observations. These results highlight the limitations of deterministic damage functions for construction cost assessment and illustrate the value of uncertainty-aware modelling for capturing the variability inherent in residential flood damage.

From a construction management perspective, the proposed framework provides decision-support outputs that are directly applicable to budgeting, contingency planning, and post-disaster recovery management. Rather than producing a single expected cost, the method generates distributions cost estimations, enabling stakeholders to assess financial risk and allocate resources more effectively under uncertainty. The ability to update damage estimates as new data become available further enhances the framework's suitability for post-disaster contexts, where information is often incomplete and evolves over time. While this study focuses on a specific residential building class and flood depth, the framework is flexible and can be extended to additional building types, inundation levels, and geographic regions as data become available. Future research will incorporate additional building attributes to enable more detailed cost estimation. Overall, this research demonstrates that probabilistic, Bayesian-based damage modelling offers a robust and practical alternative to traditional depth–damage curves for residential flood damage cost estimation, thereby informing reliable planning and budget allocation for residential building construction and recovery.

References

[1] Y. Li and W. Ji. Robustness of stakeholder response networks for infrastructure system protection. *Journal of Management in Engineering*, 37(6):04021063, 2021.

[2] Y. Li, F. Zhang, and W. Ji. Integrated data-driven and equity-centered framework for highway restoration following flood inundation. *Journal of Management in Engineering*, 39(3):04023012, 2023.

[3] H. McGrath, A. Abo El Ezz, and M. Nastev. Probabilistic depth–damage curves for assessment of flood-induced building losses. *Natural Hazards*, 97(1):1–14, 2019.

[4] Federal Emergency Management Agency. Hazus Flood Model Technical Manual. On-line: https://www.fema.gov/sites/default/files/documents/fema_rsl_hazus-7-fltm_06272025_0.pdf, Accessed: 24/04/2026.

[5] Federal Emergency Management Agency. Hazus. On-line: https://www.fema.gov/flood-maps/products-tools/hazus, Accessed: 24/04/2026.

[6] O. E. Wing, et al. New insights into US flood vulnerability revealed from flood insurance big data. *Nature Communications*, 11(1):1444, 2020.

[7] W. Mobley, K. O. Atoba, and W. E. Highfield. Uncertainty in flood mitigation practices: Assessing the economic benefits of property acquisition and elevation in flood-prone communities. *Sustainability*, 12(5):2098, 2020.

[8] K. Angeles and T. Kijewski-Correa. Bayesian data integration framework for the development of component-level fragilities derived from multiple post-disaster datasets. *Structural Safety*, 99:102260, 2022.

[9] B. Merz, H. Kreibich, and U. Lall. Multi-variate flood damage assessment: A tree-based data-mining approach. *Natural Hazards and Earth System Sciences*, 13(1):53–64, 2013.

[10] K. Schröter, et al. How useful are complex flood damage models? *Water Resources Research*, 50(4):3378–3395, 2014.

[11] A. H. Thieken, et al. Estimating changes in flood risks and benefits of non-structural adaptation strategies—a case study from Tyrol, Austria. *Mitigation and Adaptation Strategies for Global Change*, 21(3):343–376, 2016.

[12] R. Hasanzadeh Nafari and T. Ngo. Predictive applications of Australian flood loss models after a temporal and spatial transfer. *Geomatics, Natural Hazards and Risk*, 9(1):416–430, 2018.

[13] M. B. Malgwi, M. Schlögl, and M. Keiler. Expert-based versus data-driven flood damage models: A comparative evaluation for data-scarce regions. *International Journal of Disaster Risk Reduction*, 57:102148, 2021.

[14] T. Sieg, et al. Tree-based flood damage modeling of companies: Damage processes and model performance. *Water Resources Research*, 53(7):6050–6068, 2017.
[15] A. Paleo-Torres, et al. Vulnerability of Florida residential structures to hurricane induced coastal flood. *Engineering Structures*, 220:111004, 2020.
[16] T. Endendijk, et al. Flood vulnerability models and household flood damage mitigation measures: An econometric analysis of survey data. *Water Resources Research*, 59(8):e2022WR034192, 2023.
[17] Y. Li and W. Ji. Bayesian-based forecasting of infrastructure restoration progress following extreme events. *International Journal of Disaster Risk Reduction*, 103519, 2022.
[18] Y. Li, C. Yang, and W. Ji. Enhanced prediction of highway flood inundation through Bayesian generalized linear geostatistical models. *Advanced Engineering Informatics*, 61:102451, 2024.
[19] Federal Emergency Management Agency. OpenFEMA Data Sets. On-line: https://www.fema.gov/about/openfema/data-sets, Accessed: 24/04/2026.