



A Bayesian network framework for probabilistic assessment of regional building losses from hurricanes

Yu Liang^a, Hao Zhang^{a,*}, Cao Wang^b

^a School of Civil Engineering, The University of Sydney, NSW 2006, Australia

^b School of Civil and Environmental Engineering, University of Technology Sydney, NSW 2007, Australia

ARTICLE INFO

Keywords:

Bayesian network
Probabilistic risk assessment
Hurricanes
Community resilience

ABSTRACT

This paper proposes a Bayesian network (BN) framework for the probabilistic assessment of regional building losses induced by hurricanes. To explicitly capture the spatial heterogeneity of damage, the study area is partitioned into geographic subdivisions, with the model formulated as a BN-based multi-output regression framework. This framework utilises key hurricane parameters, including translation speed, heading angle, location, and central pressure, as inputs to generate probabilistic loss statistics for each subdivision as outputs. This study introduces new schemes for the discretisation of the input and output variables of the BN. For hurricane variables, a supervised discretisation method is developed that integrates weighted principal component analysis with decision tree algorithms, while also considering the relative importance of each subdivision (e.g., building density) to improve accuracy in high-priority areas. For loss variables, a clustering-based discretisation method is applied to capture the characteristics of regional building losses. The proposed framework enables efficient assessment of the spatial distribution of hurricane-induced building losses in a community, accounting for uncertainties and spatial correlation in hurricane hazard predictions and structural performance. Its effectiveness is demonstrated through a numerical example.

1. Introduction

Hurricanes, also referred to as tropical cyclones or typhoons, are among the most destructive natural hazards, often causing widespread devastation with severe economic and social consequences. Recent events such as Hurricane Helene (2024) and Hurricane Milton (2024) resulted in economic losses amounting to billions of dollars [1]. Post-disaster emergency responses, including resource allocation, infrastructure restoration, and rescue prioritisation, rely on timely and comprehensive damage assessments [2]. Developing an efficient and reliable loss assessment framework is therefore essential to support emergency management and mitigate hurricane impacts.

Scenario-based probabilistic risk analysis is a widely used approach for assessing hurricane-induced damage and losses. It involves modelling wind speed fields, assessing structural damages, and estimating community losses. This framework explicitly incorporates uncertainties and correlations in wind field modelling and structural response, providing a more realistic estimation of potential impacts than deterministic methods [3–5]. However, hurricanes evolve dynamically across time and space, requiring repeated probabilistic analyses to deliver timely

updates during an event. This repeated analysis is computationally demanding, underscoring the need for more efficient methods of regional hurricane damage assessment.

Bayesian networks (BNs) provide a promising alternative, as they can explicitly account for uncertainties and correlations while reducing computational demands. A BN is a directed acyclic graphical model in which nodes represent random variables and arcs denote causal dependencies. A key advantage of BNs is their ability to infer unknown variables from partial information, making them well-suited for modelling complex input–output relationships. BNs can also integrate data from diverse sources, including quantitative measurements, semi-quantitative data, and expert judgment, while accommodating a wide range of variables such as hazards, climate conditions, losses, and decisions. This flexibility allows BNs to support both deterministic and probabilistic predictions [6,7]. As a result, they have been applied across various engineering domains such as engineering fault detection [8,9] and disaster-related risk assessment [10,11].

In hurricane risk assessment, BNs have been applied in several contexts. Wang [12] developed a BN model to investigate the factors influencing evacuation decisions during hurricanes, considering factors

* Corresponding author.

Email addresses: yu.liang@sydney.edu.au (Y. Liang), hao.zhang@sydney.edu.au (H. Zhang), cao.wang@uts.edu.au (C. Wang).

such as risk perception, media information, and social influence. Van Verseveld et al. [13] integrated hurricane-induced multi-hazard indicators (e.g., inundation depth, flow velocity, wave attack, scour depth) with building characteristics (e.g., type and property value) to estimate the probability of structural damage. However, this approach relies on post-hurricane observation data, which limits applicability in regions lacking reliable hurricane loss records. Furthermore, the model was developed at the building scale for relatively small study areas; hence, broader-scale applications may require alternative modelling strategies. Li et al. [14] proposed a BN for predicting power distribution system outages caused by hurricanes, incorporating variables related to hurricane dynamics and power grid topology. The assessment of power outages in this study was based on network connectivity, whereas actual power systems are more complex and may also depend on power flow conditions. Similar research efforts are reported in [15–17]. Overall, existing applications of BNs in hurricane risk assessment have focused mainly on post-disaster decision-making and building-level damage assessment. While these models are effective for localised or single-entity predictions, they often struggle with regional-scale assessments that necessitate high-dimensional, multi-output spatial evaluations. This limitation becomes particularly pronounced in large-scale regions where socio-economic significance is heterogeneously distributed, requiring prioritised consideration for specific regions.

A key challenge in BN modelling is the discretisation of continuous variables, which must balance model complexity against predictive accuracy. Three discretisation schemes are commonly used in BNs: manual, unsupervised, and supervised approaches. Manual discretisation relies on domain expertise and prior knowledge. This approach enhances interpretability and allows the integration of expert insights, but it introduces subjectivity and often lacks applicability across different contexts [6]. Unsupervised methods, such as equal-width and equal-frequency discretisation, are computationally efficient but often overlook important data patterns, reducing their predictive performance [6,18]. Supervised discretisation considers the relationship between inputs and outputs, retaining the information most relevant for prediction. Common supervised methods include minimum description length [19], decision trees [20], and Chi-squared-based methods [21]. A systematic review by Kaikkonen et al. [7], which analysed 72 studies using BNs in environmental and engineering risk assessments, found that manual and unsupervised discretisation methods remain dominant due to simplicity, interpretability, and low computational cost, while supervised discretisation methods were applied in fewer than 10% of the cases. Nevertheless, supervised discretisation has consistently shown superior performance. For example, Beuzen et al. [18] demonstrated that supervised discretisation methods achieved 5%–10% higher accuracy than manual and unsupervised methods.

This study proposes a BN-based multi-output spatial regression framework to evaluate regional hurricane-induced building losses. Unlike conventional BN models that typically predict an aggregated loss value, the proposed framework establishes conditional dependencies between hurricane parameters and the spatial distribution of losses across entire study regions. The methodological novelty lies in the introduction of a priority-area weighted learning scheme within a supervised discretisation process for input variables, alongside a clustering-based discretisation method for high-dimensional output variables. The effectiveness of this multi-output, priority-weighted framework is demonstrated through a numerical example.

2. Scenario-based hurricane risk assessment for regional buildings

Scenario-based hurricane risk assessment estimates the performance of civil infrastructures under a prescribed hurricane scenario characterised by parameters such as hurricane locations, central pressure, and translational speed. The probabilistic performance evaluation can be

expressed as:

$$P(Z \leq z | S) = \int_{\mathbf{v}} \int_{\mathbf{y}} F_{Z|DS}(z | \mathbf{v}) f_{DS|W}(\mathbf{v} | \mathbf{y}) f_{W|S}(\mathbf{y} | S) d\mathbf{v} d\mathbf{y}, \quad (1)$$

where $f_{W|S}(\mathbf{y}|S)$ is the probability density function (PDF) of wind speeds $\mathbf{W}$ at building locations given the hurricane scenario S . $f_{DS|W}(\mathbf{v}|\mathbf{y})$ is the PDF of physical damage states $\mathbf{DS}$ conditioned on $\mathbf{W}$. $F_{Z|DS}(z|\mathbf{v})$ is the cumulative distribution function (CDF) of the performance metric Z for building portfolios, conditioned on $\mathbf{DS}$. The probability $P(Z \leq z|S)$ represents the convolution of these three conditional distributions.

The wind speed distribution, $f_{W|S}(\mathbf{y}|S)$, is typically derived from a hurricane wind field model that produces gradient wind speeds. The gradient wind speeds are subsequently converted to surface wind speeds over a specified averaging duration using appropriate conversion factors. The conditional damage distribution $f_{DS|W}(\mathbf{v}|\mathbf{y})$ is commonly estimated using structural fragility functions, which describe the probability of exceeding a damage threshold for a given wind speed. These fragility functions are often modelled as lognormal CDFs, with inputs expressed in terms of 3-second gust wind speeds at 10 m elevation in open terrain [22,23]. Finally, $F_{Z|DS}(z|\mathbf{v})$ translates structural damage into regional performance indicators, such as economic losses or recovery time.

In practice, Eq. (1) cannot be solved in closed form due to the large number of buildings in a community, multiple sources of uncertainty, and the presence of spatial correlations. Spatial correlations arise both across sites, owing to similarities in wind speeds at nearby locations, and across structures, due to similarities in building performance resulting from common construction materials, structural types, and design codes [3]. Consequently, Monte Carlo simulation (MCS) is typically employed to approximate Eq. (1) [24]. In each simulation, a sample wind field is generated using a hurricane wind field model. A random field model is then applied to the hurricane wind field to account for the modelling uncertainties of the wind field model, which is a major contributor to hurricane risk assessment [23]. The post-hazard damage states of buildings are then evaluated using structural fragility functions, such as those provided in HAZUS [25] or other empirical/analytical studies [22,23]. In the structural damage assessment, spatial correlations, including both site-to-site and structure-to-structure correlations, should be incorporated. The system performance indicator Z is computed based on the relationship between the individual structure's damage and overall system performance. Repeating this procedure over a sufficiently large number of simulations yields the probability distribution and statistics of the system-level performance metric Z .

3. BN-based hurricane loss model for regional buildings

3.1. Fundamentals of Bayesian networks

A BN represents probabilistic dependencies among variables using a directed graphical structure. It consists of a set of nodes (random variables) and directed arcs (dependencies). The direct predecessors or successors of a node are referred to as parents and children, respectively. Nodes without parents are called root nodes and represent input variables, while nodes without children are called leaf nodes and represent output variables.

Each node is associated with a conditional probability table (CPT), which specifies the probability distribution of the node given the states of its parents. For root nodes, the CPT is defined by a marginal distribution, which may be obtained from statistical data or expert judgment [26].

A key property of BNs is conditional independence: a node is conditionally independent of its non-descendants given its parents. This property allows a complex joint probability distribution to be decomposed into local conditional probabilities, significantly simplifying the

computation of the joint probability. According to the chain rule, the joint probability of variables X_1 to X_n can be expressed as:

$$P(X_1, X_2, \dots, X_n) = P(X_1)P(X_2 | X_1) \dots P(X_n | X_1, X_2, \dots, X_{n-1}). \quad (2)$$

Based on the property of conditional independence, the joint probability can be simplified to:

$$P(X_1, X_2, \dots, X_n) = \prod_{i=1}^n P(X_i | \text{Pa}(X_i)), \quad (3)$$

where $\text{Pa}(X_i)$ denotes the set of parent nodes of X_i , and $P(X_i | \text{Pa}(X_i))$ is the corresponding CPT of X_i .

Inference in BNs can be classified into two types depending on the position of evidence variables: predictive inference and diagnostic inference. Predictive inference involves reasoning forward from cause to effect, estimating the probability distribution of an outcome variable based on a known cause. In contrast, diagnostic inference reasons backward from observed effects to potential causes. For predictive inference, the distribution of a target variable X_i can be updated given evidence E , using the following expression:

$$P(X_i | E) = \frac{P(X_i, E)}{P(E)} = \frac{P(X_i, E)}{\sum_{X_i} P(X_i, E)}, \quad (4)$$

where $\sum_{X_i}$ denotes the summation over all possible states of the variable X_i .

For diagnostic inference, the posterior distribution of a potential cause variable X_j can be updated based on observed evidence E , using Bayes' Theorem:

$$P(X_j | E) = \frac{P(E | X_j) \cdot P(X_j)}{P(E)} = \frac{P(E | X_j) \cdot P(X_j)}{\sum_{X_j} P(E | X_j) \cdot P(X_j)}. \quad (5)$$

3.2. Construction of a BN-based hurricane loss model

The primary purpose of a BN-based hurricane loss model is to estimate the spatial distribution of loss based on fundamental hurricane parameters, without performing the scenario-based probabilistic risk assessment described in Section 2. The model consists of two main components: hurricane-related input data and damage-related output data. To capture the spatial distribution of damage, the study area is divided into multiple geographic subdivisions. Common subdivision units include ZIP codes, census tracts, and block groups [25], which generally decrease in size in that order. Alternatively, subdivisions may be custom-defined using algorithms that consider regional characteristics, such as SKATER [27].

Hurricane parameters include longitude, latitude, translation speed, heading angle, central pressure (P_c), maximum wind speed ($V_{\max}$), radius to maximum winds ($R_{\max}$), and the Holland parameter (B), among others. These variables can be grouped into three categories: geographical location, motion vector, and intensity. Selecting appropriate hurricane variables is essential to capture key characteristics while avoiding redundancy and controlling model complexity. In this study, five parameters are selected: the distance and azimuth of the hurricane relative to the study area, translation speed, heading angle, and central pressure. The first two describe the hurricane's geographical location. Polar coordinates are used instead of longitude and latitude because they more effectively represent the hurricane's position relative to the study area. Translation speed and heading angle together define the hurricane's motion vector. The final parameter, P_c , represents hurricane intensity and the potential extent of its impact. P_c is selected because it can be used to estimate other intensity-related parameters. For instance, the relationship between $V_{\max}$ and P_c is commonly expressed as $V_{\max} = m(P_{\text{ref}} - P_c)^n$ [28], where m and n are regression coefficients and P_{ref} denotes ambient atmospheric pressure. Previous studies have also established relationships between P_c and $R_{\max}$, as well as between P_c and B [25,29].

Thus, including P_c in the loss model effectively reflects the intensity characteristics of hurricanes.

Common performance metrics for regional buildings include the structural damage ratio, direct economic loss, economic loss ratio, and household displacement [24,30]. Among these, the loss ratio is frequently used to evaluate regional building performance because it provides a standardised measure that enables consistent comparisons across different regions and building types. The overall regional loss ratio, denoted as L , is defined as the ratio of total loss to the total replacement cost of all buildings and is given by Vitoontus [3]:

$$L = \frac{\sum_i L_i \cdot R_i}{\sum_i R_i}, \quad (6)$$

where L_i is the structural loss ratio of building i , and R_i is the full structural replacement cost of building i .

CPTs in BNs are typically constructed using two main approaches: knowledge-based expert assessment and data-driven statistical estimation. When sufficient data are available, conditional probabilities are estimated using methods such as maximum likelihood estimation (MLE) or Bayesian estimation with prior distributions. Conversely, when data are limited or incomplete, domain experts provide initial probability values based on their knowledge and experience. These expert-defined CPTs can then be refined using limited observational data through methods such as the Expectation-Maximisation (EM) algorithm or other probabilistic inference approaches [26]. In this study, the CPT of the hurricane loss model represents the probabilistic relationship between hurricane characteristics and building losses. It is constructed through three steps: (1) generating a set of hurricane scenarios defined by combinations of the five hurricane parameters introduced earlier; (2) calculating the regional loss ratio for each scenario using the probabilistic risk assessment framework described in Section 2; and (3) employing these hurricane scenarios and corresponding regional losses as training data for parameter learning via MLE, thereby constructing the CPT. For a variable X , the conditional probability $\hat{P}(X | \text{Pa}(X))$ estimated using MLE is given by:

$$\hat{P}(X = x | \text{Pa}(X) = \mathbf{p}) = \frac{\sum_i \mathbb{I}(X^{(i)} = x \wedge \text{Pa}(X)^{(i)} = \mathbf{p})}{\sum_i \mathbb{I}(\text{Pa}(X)^{(i)} = \mathbf{p})}, \quad (7)$$

where $\mathbf{p}$ is specific states or values of $\text{Pa}(X)$, x is a specific state or value of X , i denotes the i -th sample in the dataset, $\mathbb{I}(\cdot)$ is the indicator function that equals one if the condition is satisfied and zero otherwise, and $\wedge$ represents the logical AND operation.

This training procedure allows the BN to directly learn the statistical dependence between hurricane characteristics and regional losses, effectively transferring the accuracy of the probabilistic risk assessment framework into a computationally efficient surrogate model. Once trained, the BN can rapidly update loss predictions under new hurricane conditions while retaining the uncertainty and correlation structure captured in the training data.

4. Discretisation of continuous variables in hurricane loss models

4.1. Supervised discretisation for hurricane variables

The discretisation of hurricane variables directly affects the predictive accuracy of regional losses. Supervised discretisation methods are generally more effective than manual or unsupervised approaches because they preserve the predictive relationship between inputs and outputs [26,31]. Standard supervised discretisation is typically guided by a single output variable. In this study, multiple output variables are considered, specifically the losses associated with individual subdivisions. Discretising input variables separately for each output can result in inconsistent schemes across outputs. Conversely, combining all outputs into a single aggregated output during discretisation treats each subdivision equally, without considering differences in their relative importance, which may not reflect actual priorities. Some subdivisions,

Table 1
Types and quantities of single-family residential buildings [25].

Building type	Component type and proportion	Number
WSF1	Roof shape: Gable = 62%, Hip = 38%; Sheathing: 8d@6"/6" = 43%, 6d@6"/12" = 57%; Shutters: Yes = 30%, No = 70%.	72,379
WSF2	Same as WSF1	31,022
MSF1	Roof shape: Gable = 62%, Hip = 38%; Sheathing: 8d@6"/6" = 43%, 6d@6"/12" = 25%, 8d@6"/12" = 32%; Shutters: Yes = 30%, No = 70%.	240,325
MSF2	Same as MSF1	49,219

such as those with higher building density or greater economic value, should be given greater weight in the discretisation process.

PCA is a dimensionality reduction technique that transforms high-dimensional data into a lower-dimensional representation while retaining as much of the original variance (i.e., information) as possible. The standard procedure involves three steps. First, the original variables are standardised to have a zero mean and a variance of one to ensure comparability across different scales. Second, the loadings of the original variables are calculated, representing the contribution of each variable to the principal components. Finally, the number of principal components is determined using criteria such as the cumulative explained variance [32]. In this study, PCA is used not only for dimensionality reduction but also to incorporate the relative importance of the original variables. To achieve this, a weighting factor, $\sqrt{\omega}$, is applied to the standardised loss ratio. This adjustment scales the variance of each loss variable so that subdivisions with higher importance exhibit greater variance. More information from these variables is therefore retained during the PCA process. This procedure is expressed as:

$$L' = \sqrt{\omega} \cdot \frac{L - \mu}{\sigma}, \quad (8)$$

where L' is the weighted and standardised loss variable for a subdivision, and μ and σ denote the mean and standard deviation of L , respectively.

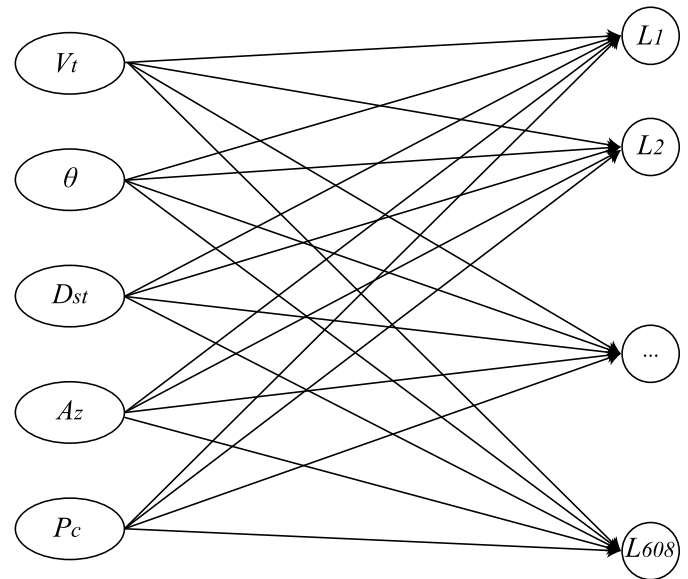


Fig. 2. Structure of the hurricane loss model for Miami-Dade County.

The principal components, denoted by $\mathbf{Y}$, are then computed from the adjusted loss ratios of all subdivisions, $\mathbf{L}'$, using PCA:

$$\mathbf{Y} = \mathbf{L}'\mathbf{Q}, \quad (9)$$

where $\mathbf{Q}$ is the loading matrix, which serves as a projection matrix that maps the original data into the PCA space. The number of principal components is selected based on a cumulative explained variance threshold, such as 95%.

Next, these principal components are used to guide the discretisation of hurricane variables using a decision tree algorithm. Decision trees are non-parametric supervised methods that identify optimal splits by maximising information gain, or equivalently, by minimising impurity [33]. During discretisation, the algorithm recursively partitions the hurricane variables to minimise the weighted variance of the resulting

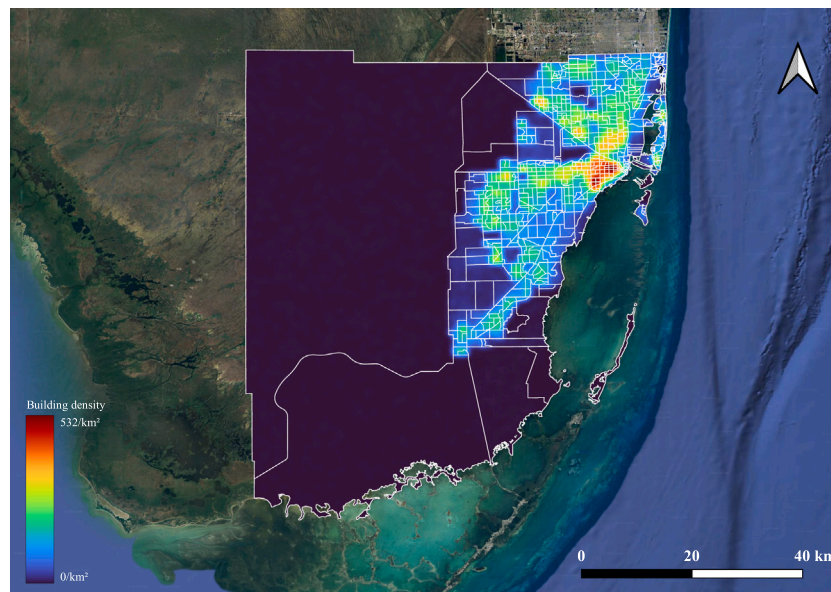


Fig. 1. Residential building distribution in Miami-Dade County with census tract boundaries (white lines). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

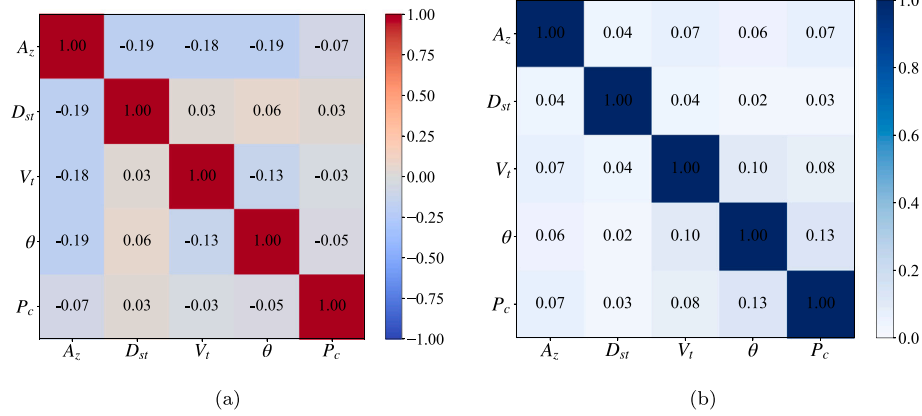


Fig. 3. Statistical Independence validation of hurricane variables: (a) Pearson correlation matrix; (b) normalised mutual information matrix.

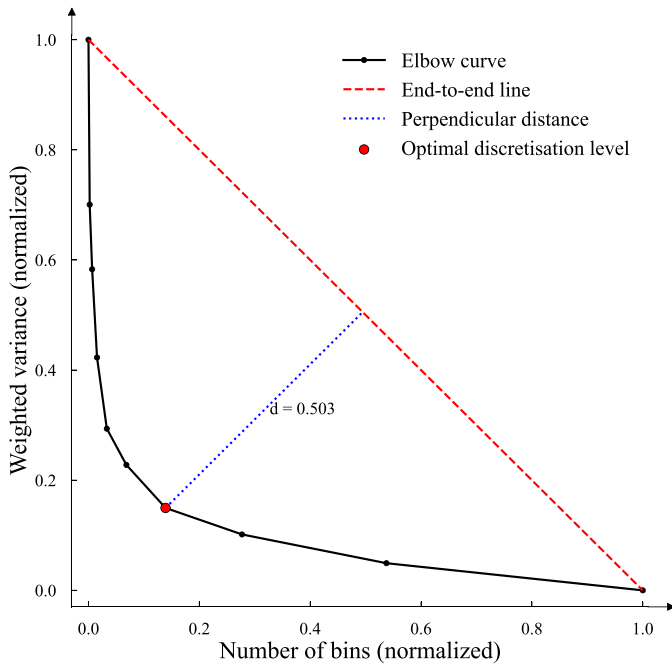


Fig. 4. Weighted variance versus number of discretisation bins for hurricane variables.

output subsets. The objective function is defined as:

$$\text{Minimise } \frac{1}{n} \sum_{m=1}^M \sum_{y \in S_m} (y - \bar{y})^2, \quad (10)$$

where S_m is the m -th subset, M is the total number of subsets, y is a sample of the principal components in S_m , $\bar{y}$ is the mean of the principal components in S_m , and n is the total number of samples. The number of discretisation intervals for each hurricane variable depends on both the depth of the decision tree and the variable's importance. Deeper trees generate more intervals, while variables with greater influence on the output variables are allocated more intervals. To determine the appropriate number of intervals, the elbow method can be applied. This graphical approach plots the relationship between the number of intervals and the weighted variance, identifying the point at which additional intervals result in only marginal reductions in variance.

4.2. Clustering-based discretisation for loss variables

Regional losses are influenced by factors such as structural types, terrain, and hurricane characteristics. Losses across different regions may be correlated, with adjacent areas typically exhibiting stronger spatial correlations [34]. Therefore, the discretisation of loss variables should capture these regional loss characteristics. Traditional methods, such as expert judgment or simple schemes like equal-width and equal-frequency discretisation, are inadequate due to the wide range of influencing factors and may lead to suboptimal outcomes.

To address these challenges, this study adopts a more flexible approach: K-means clustering. This method partitions the principal components into K clusters by iteratively updating cluster centroids to minimise the within-cluster variance, thereby capturing the underlying data structure. Conducting K-means clustering in the PCA space, rather than on individual loss variables, eliminates the need to cluster each variable separately, which reduces computational complexity and enables the global loss patterns represented by the principal components to be used to capture the overall characteristics of community damage.

The optimal number of clusters, K , is determined using the elbow method, as introduced in Section 4.1. The resulting cluster centroids, C , are then mapped back to the original loss variable space, denoted as C' , using the inverse PCA transformation:

$$C' = (CQ^T)\Omega^{-\frac{1}{2}}D + \mu, \quad (11)$$

where Ω , D , and μ represent the weight, standard deviation, and mean of the original loss variables, respectively. Discretisation thresholds are then determined as the midpoints between the centroids of adjacent clusters in C' .

5. An illustrative example

5.1. Introduction to the study area

This section uses Miami-Dade County, Florida, as an example to illustrate the development of a BN-based hurricane loss model and evaluate its effectiveness. Post-disaster assessments show that residential buildings are among the most frequently damaged structures [24]. Accordingly, this example focuses on single-family residential buildings in Miami-Dade County. Based on building statistics from the Hazus Inventory Technical Manual [25], single-family residential buildings are classified into four structural types: WSF1 (wood-framed, single-story), WSF2 (wood-framed, multi-story), MSF1 (masonry, single-story), and MSF2 (masonry, multi-story). Each type is further divided into subcategories according to variations in structural components such as roof shape, sheathing, and shutters. Specifically, WSF1 and WSF2 each have eight subcategories, while MSF1 and MSF2 each have twelve. The number of buildings in each subcategory is determined from the proportional

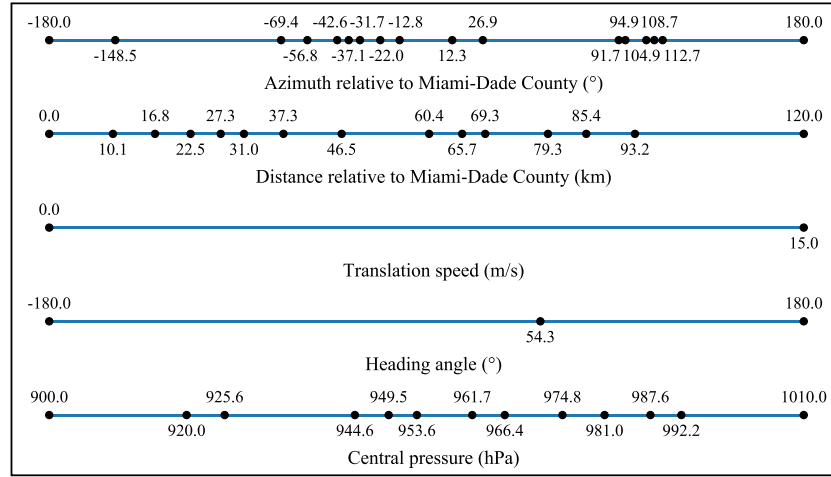


Fig. 5. Discretisation results of the five hurricane variables.

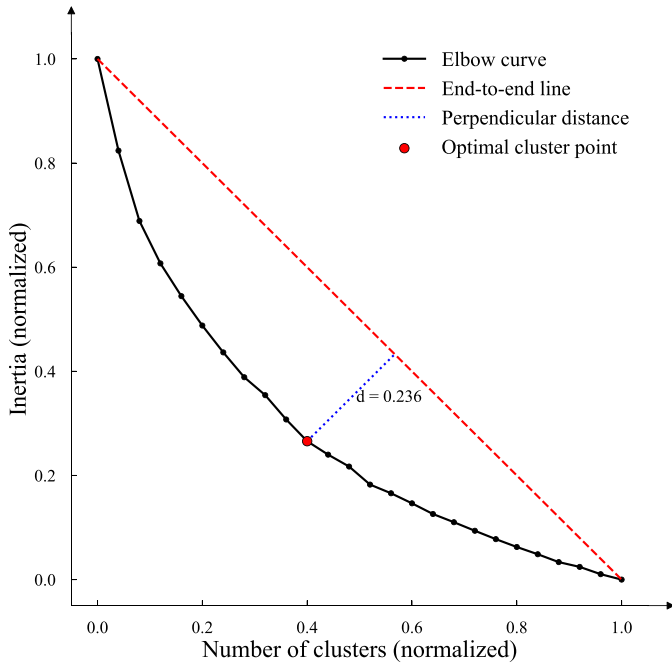


Fig. 6. Reduction of inertia with an increasing number of clusters.

combinations of these component types relative to the total number of buildings in the category, as summarised in Table 1. The spatial distribution of residential buildings, represented by building density, is shown in Fig. 1, where building density is calculated using the number of buildings within each census tract and the corresponding tract area [25].

Fragility functions for each building type are obtained from HAZUS-MH [25]. In this framework, damage states (DSs) are classified into five categories: no damage, minor damage, moderate damage, severe damage, and destruction. Fragility functions are typically modelled using the CDF of a lognormal distribution. The probability of a specific damage state is given by:

$$P(DS = v | U = u) = \Phi\left(\frac{\ln(u) - \lambda_v}{\xi_v}\right) - \Phi\left(\frac{\ln(u) - \lambda_{v+1}}{\xi_{v+1}}\right), \quad (12)$$

where $\Phi(\cdot)$ is the standard normal CDF; λ_v and ξ_v are the logarithmic mean and standard deviation of the fragility function for $DS = v$; and λ_{v+1} and ξ_{v+1} are the corresponding parameters for $DS = v + 1$. Loss

ratios are calculated using the empirical relationship between damage state and loss ratio [35]: no damage = 0%, minor damage = 20%, moderate damage = 40%, severe damage = 80%, and destruction = 100%. Because absolute structural replacement costs can vary over time, while the relative cost ratios among building types tend to remain stable [25], this study uses replacement cost ratios and assumes they remain constant across time and census tracts. According to construction cost estimates from [36], the average cost is approximately \$30 per square foot for wood-framed structures and \$47 per square foot for masonry structures. Based on these values, the structural replacement cost ratio of MSF1 to WSF1 is set to 1.56. In addition, the average height of one-story buildings is about 9 feet, while two-story buildings average about 17 feet [25]. Accordingly, the structural replacement cost ratios for WSF2/WSF1 and MSF2/MSF1 are both set to 1.88.

Miami-Dade County contains 706 census tracts, which are used as the spatial units for loss estimation in this study. Census tracts are selected because they generally have relatively uniform population sizes and building counts, represent an intermediate spatial scale (larger than block groups but smaller than ZIP code areas) and have clearly defined boundaries. These characteristics make census tracts widely used in regional risk assessments [37,38].

5.2. Configuration for scenario-based hurricane loss assessment

The gradient wind field, V_g , for a given hurricane scenario is estimated using the model proposed by Georgiou [39]:

$$V_g = \frac{1}{2} (V_t \sin \alpha - fr) + \sqrt{\frac{1}{4} (V_t \sin \alpha - fr)^2 + B \frac{\Delta P}{\rho} \left(\frac{R_{\max}}{r}\right)^B \exp\left(-\left(\frac{R_{\max}}{r}\right)^B\right)}, \quad (13)$$

where V_t = the translation speed of the cyclone; r = the radial distance from the cyclone centre; α = the clockwise angle between the cyclone's translation direction and the radial position of the sites; f = the Coriolis parameter; ρ = the air density, which is equal to 1.15 kg/m³. The radius of maximum wind speed, $R_{\max}$, is estimated following Vickery and Wadhwa [29]:

$$R_{\max} = \exp(2.377 - 4.825 \times 10^{-5} \Delta P^2 + 0.0483\psi), \quad (14)$$

where ψ is the latitude of the hurricane centre. The Holland pressure profile exponent parameter, B , is calculated by Vickery and Wadhwa [29]:

$$B = 1.38 + 0.00184 \Delta P - 0.00309 R_{\max}. \quad (15)$$

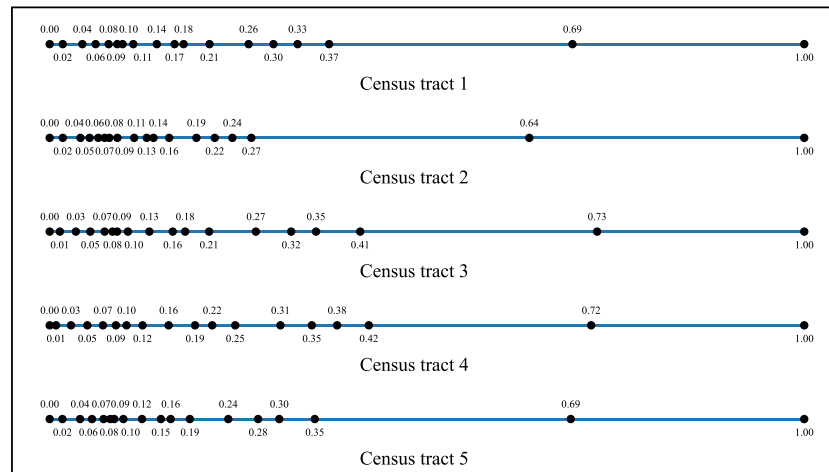
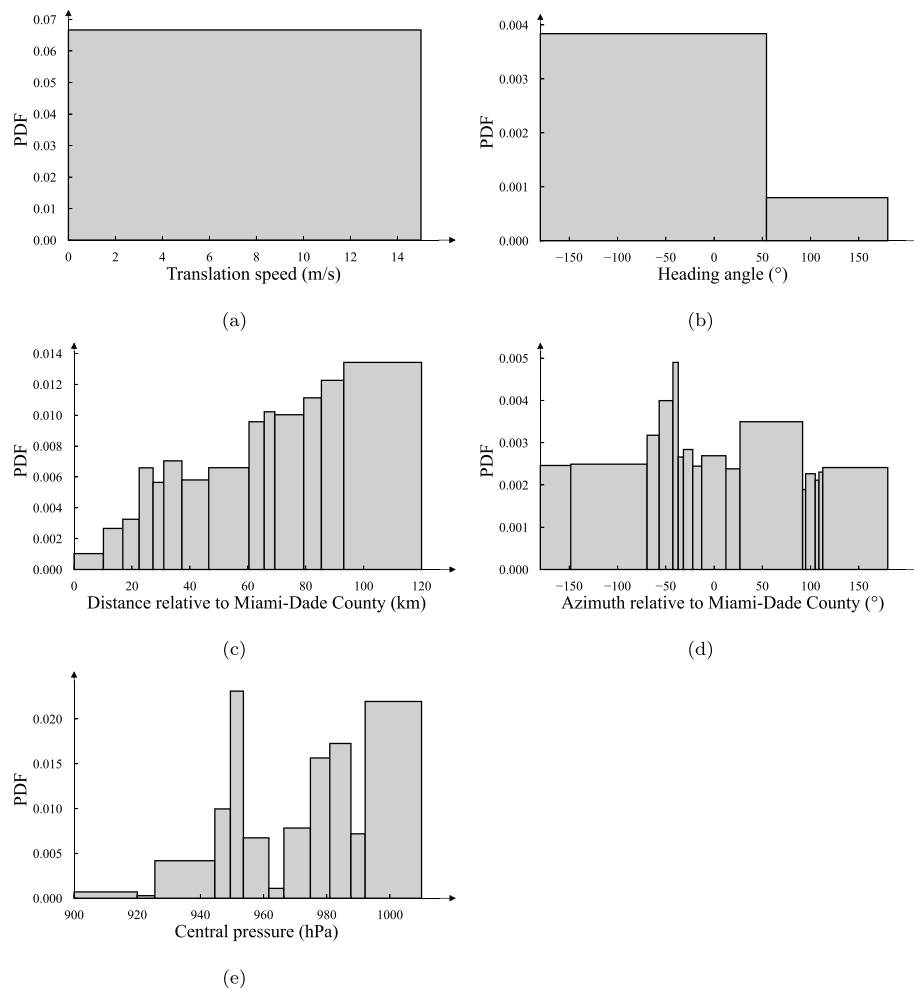


Fig. 7. Discretisation of loss ratios for census tracts 1–5.

Fig. 8. Marginal distributions of hurricane variables: (a) V_t , (b) θ , (c) D_{st} , (d) A_z , and (e) P_c .

It should be noted that the statistical relationships proposed by Vickery and Wadhera [29] incorporate random residual terms, which represent additional modelling uncertainty. In principle, including these residuals would yield a more comprehensive characterisation of the total uncertainty. However, previous studies have demonstrated that the uncertainty associated with these residuals is relatively marginal compared

to that arising from hurricane wind field modelling, which is widely recognised as the dominant source of uncertainty in hurricane risk assessment [23,40]. Consequently, these residual terms are not explicitly considered in the present study to simplify the computational process.

Converting V_g to the 3-second gust wind speed at 10 m elevation in open terrain, which serves as the input for fragility functions, involves

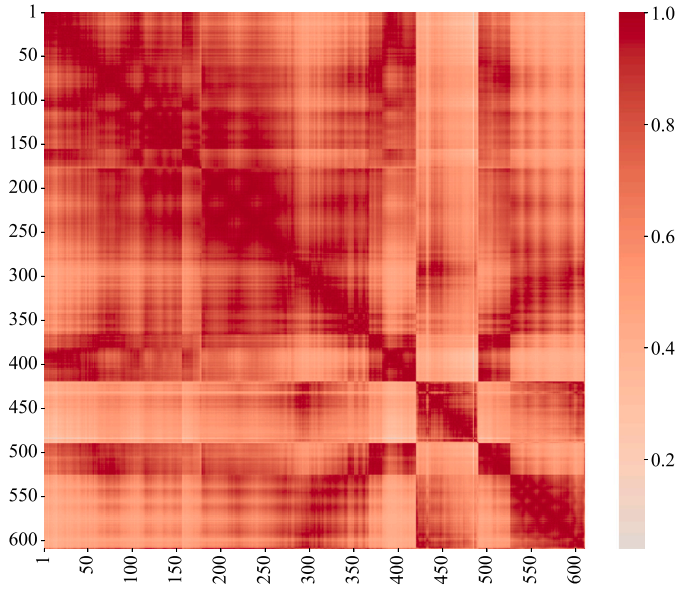


Fig. 9. Correlation heatmap of loss ratios across the 608 census tracts.

three steps: gradient-to-surface conversion, water-to-open-terrain conversion, and sustained-to-gust conversion. These conversions follow the procedures outlined by Georgiou [39] and Vickery and Skerlj [41].

The uncertainty of the wind field is estimated using the method proposed in Zeng et al. [42], which evaluates the uncertainty of the Georgiou model based on data from 34 recorded hurricane wind fields across six historical North Atlantic hurricanes. Their analysis indicated that the uncertainty follows a lognormal distribution, with the mean of the logarithmic error given by $-0.0515(r/R_{\max}) - 0.08321$ and a constant standard deviation of 0.12. Zeng et al. [42] also calculated that the spatial correlation of wind speeds between locations i and j in the North Atlantic, denoted as ρ_{ij} , can be estimated as:

$$\rho_{ij} = \exp\left(-\frac{h_{ij}^2}{1.322R_{\max}^2}\right), \quad (16)$$

where h_{ij} represents the distance between locations i and j .

Adjacent buildings often exhibit correlated structural resistance due to similarities in construction materials, design practices, and building codes. Because census tracts are defined with consideration of homogeneity in living conditions [43], structural capacity correlation is assumed to exist among buildings within the same census tract. The correlation coefficient is set to 0.7 for buildings of the same type and 0.4 for buildings of different types [44].

Using the parameters defined in Eq. (13) through Eq. (16) and the procedure described in Section 2, the mean, standard deviation, and probability distribution of the loss ratio for each census tract are estimated. Given the large number of buildings, random sampling is employed to improve computational efficiency. Lin and Wang [30] evaluated different sample sizes for a 6 km² study area containing 4246 buildings and found that the sampling estimates converged to the exact solution when the sample size reached 100. In the present study, the average census tract area is 7.3 km², with an average of 557 buildings. Although the average area is slightly larger than that considered by Lin and Wang [30], the average number of buildings is substantially lower. Accordingly, a sample size of 100 is used for tracts with more than 100 buildings, while all buildings are included when the total count is below 100.

5.3. BN-based hurricane loss model for Miami-dade county

The BN-based hurricane loss model incorporates five key hazard variables that are both readily available in databases and effective in characterising hurricane location, movement and intensity. These variables include the distance (D_{st}) and azimuth (A_z) relative to the study area's centroid, translation speed (V_t), heading angle (θ), and central pressure (P_c). Regarding the damage variables, the model defines loss ratios at the census-tract level for Miami-Dade County. To balance spatial resolution with computational efficiency, this study focuses on 608 census tracts (represented by variables L_1 to L_{608}) that contain more than 10 residential buildings.

The topology of the BN is established based on physical interpretability. Specifically, the five hurricane variables serve as root nodes, each directed toward the census tract loss nodes, representing the collective impact of hurricane characteristics on regional vulnerability. The resulting BN structure is illustrated in Fig. 2. A primary assumption in this structure is the mutual independence of the five hurricane variables.

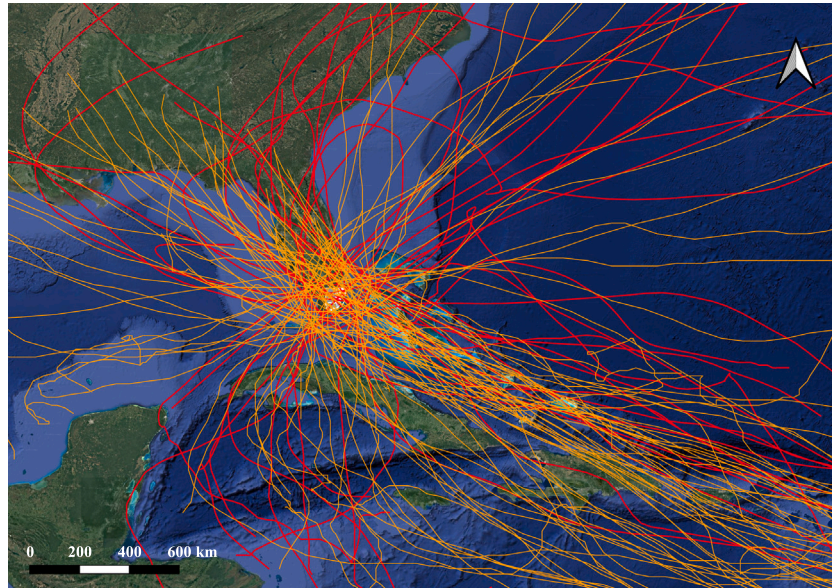


Fig. 10. Spatial distribution of 100 test hurricanes used for model validation.

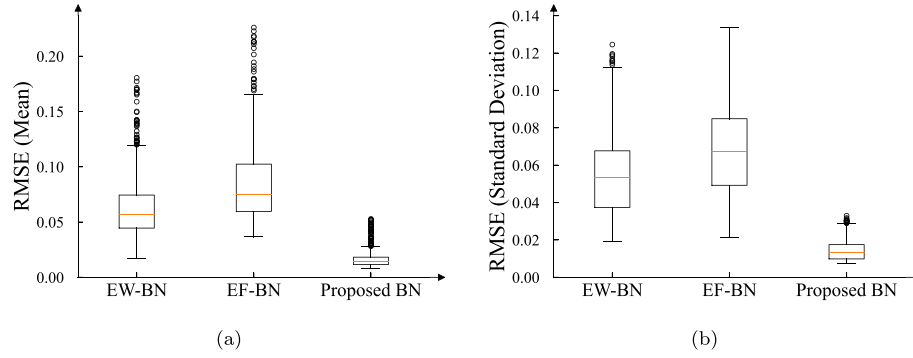


Fig. 11. Comparison of prediction accuracy of BN models with different discretisation methods: (a) RMSE of mean loss ratio (b) RMSE of standard deviation of loss ratio.

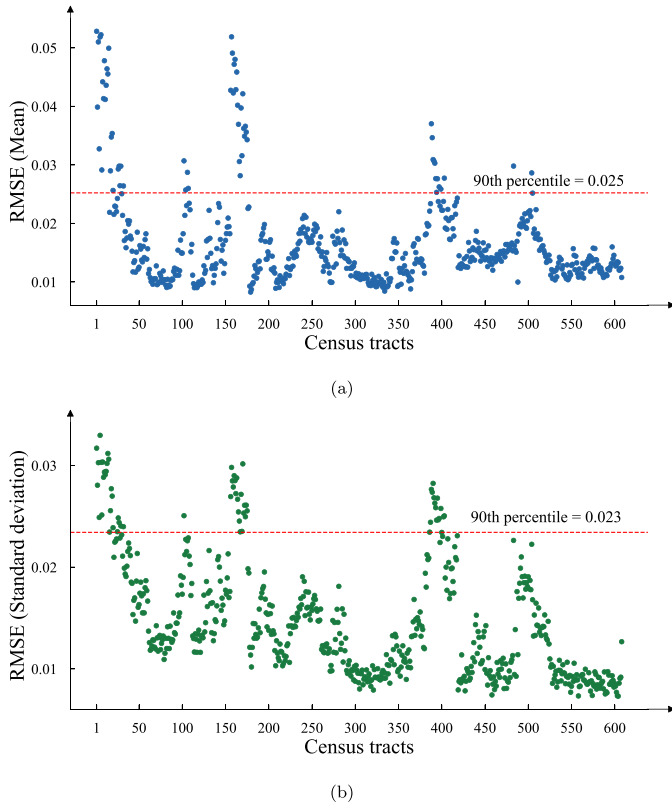


Fig. 12. RMSE of predicted loss ratios across 608 census tracts using 100 test hurricanes: (a) mean, (b) standard deviation.

To justify this, historical hurricane data within a 100 km radius of the study area are extracted from the HURDAT2 database for events occurring after 1900 [45]. The independence is validated by calculating the correlation matrix and the normalised mutual information (NMI), the latter of which is computed using arithmetic-mean normalisation [46]. As shown in Fig. 3, the weak coupling between these variables confirms that the independence assumption is statistically sound.

A training dataset is constructed to support the data-driven discretisation of both hurricane and loss variables. The generation of this dataset follows the scenario-based probabilistic loss assessment framework detailed in Section 2. Initially, hurricane scenarios are randomly sampled from the parameter space of the five hazard variables, defined based on historical hurricane records near Miami-Dade: $V_i \in [0, 15]$ m/s, $\theta \in [0, 360^\circ]$, $D_{st} \in [0, 120]$ km, $A_z \in [0, 360^\circ]$, and $P_c \in [900, 1010]$ hPa. To determine an optimal dataset size, the relationship between sample

size and the stability of the discretisation results is evaluated. We tested sample sizes ranging from 1000 to 12,000, with an increment of 1000. For this sensitivity analysis, the decision tree depth is fixed at 4, and P_c is selected as the representative variable to examine the discretisation stability. This configuration provides a moderate discretisation level with five intervals in this case, ensuring a consistent comparison across different sample sizes. For each size, 10 independent experiments are conducted to calculate the normalised root mean square error (RMSE) of the resulting split points. While previous literature suggests that a normalised RMSE of 5–10% indicates acceptable variation [47], a stricter threshold of 3% is adopted here, given the significant impact of discretisation on BN performance. The results indicate that the stability threshold is satisfied at a sample size of 10,000, which is subsequently adopted for the training set. For each sampled scenario, 1000 MCS are performed to derive the probability distribution of the loss ratio for each census tract.

The first step in discretising hurricane variables is to perform a weighted PCA on the loss variables to reduce dimensionality while preserving the relative importance of each tract in the principal components. In this study, building density is adopted as the importance factor (see Fig. 1). The rationale for selecting building density is that it directly reflects the spatial concentration of exposed assets; higher-density areas typically correlate with greater potential losses and, consequently, demand higher priority in risk assessment and emergency response planning. It should be noted that, depending on the specific application context and data availability, alternative indicators, such as economic valuation, population density, or composite indices, could also be employed to represent regional priorities. The threshold for the cumulative explained variance ratio is set to 95%. The first five principal components, which collectively explain 96.8% of the variance, are selected for subsequent analysis. In the second step, these five principal components are used to guide the discretisation of the hurricane variables through a decision tree, implemented using the DecisionTreeRegressor class in the scikit-learn library. By specifying the tree depth, the algorithm identifies optimal split points for each hurricane variable, generates the corresponding discretisation thresholds, and determines the total number of bins. Because the decision tree uses binary splitting, the total number of bins approximately follows a 2^n relationship, where n is the tree depth. To select an appropriate discretisation level, Fig. 4 plots the weighted variance, defined in Eq. (10), against the number of bins, with each point corresponding to a specific tree depth. Both the weighted variance and the number of bins are normalised to the range [0,1] to facilitate identification of the elbow point. Following the method proposed by Satopaa et al. [48], a straight line is drawn between the first and last points, and the perpendicular distance from each point on the curve to this line is computed. The maximum distance is taken as the optimal discretisation level. At this level, the discretisation numbers and thresholds for each hurricane variable are shown in Fig. 5. The number of bins for A_z , D_{st} ,

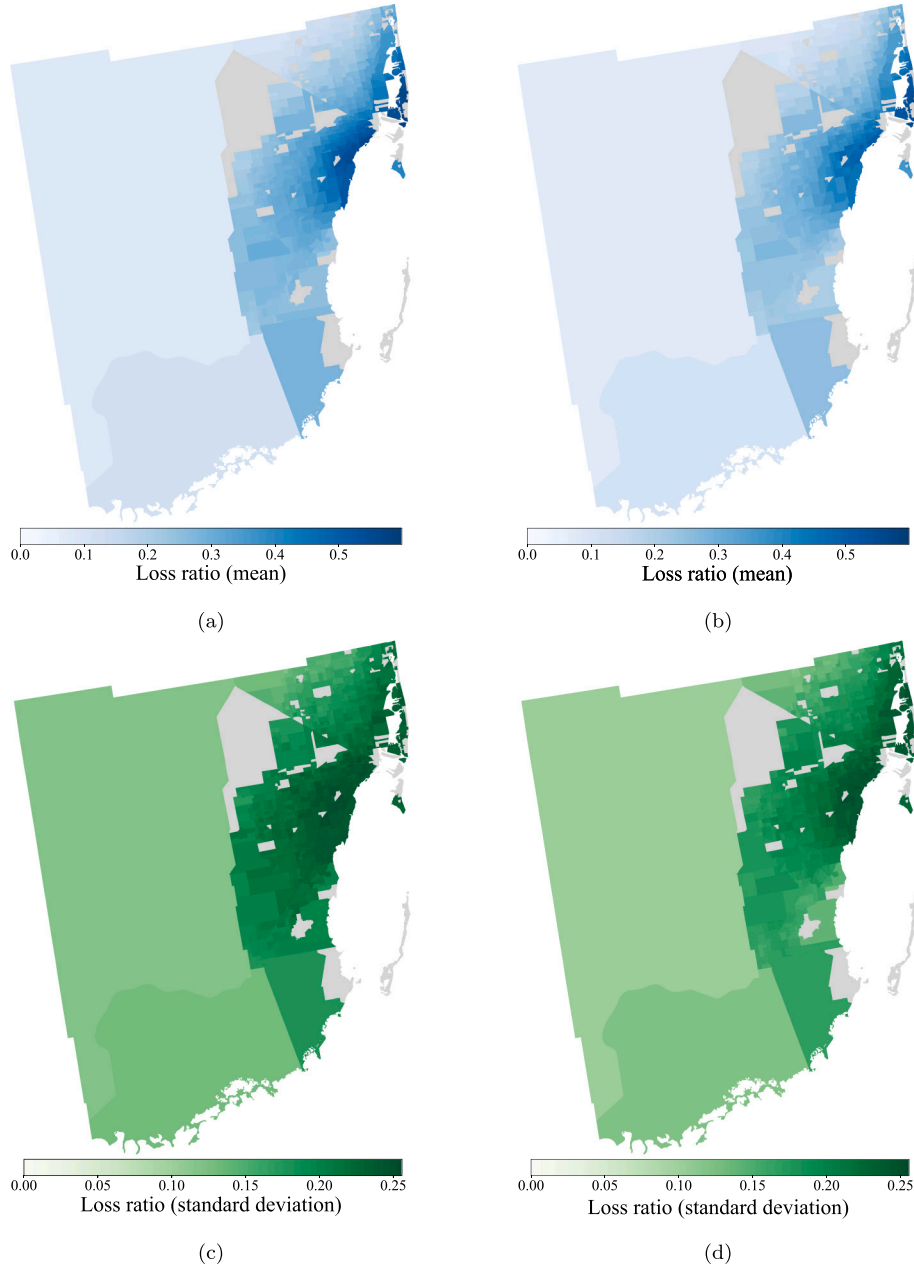


Fig. 13. Comparison of reference and BN-based model estimated loss ratios for Hurricane Andrew (1992): (a) mean (reference), (b) mean (BN-based), (c) standard deviation (reference), and (d) standard deviation (BN-based).

and P_c is substantially higher than for V_i and θ , suggesting that the former three variables have a greater influence on losses.

The discretisation of loss variables is performed by applying K-means clustering to the five principal components, implemented using the KMeans class in the scikit-learn library. Fig. 6 shows the number of clusters ranging from 5 to 30 and the corresponding reduction in inertia, defined as the sum of squared distances to the cluster centres. Following the same procedure used to determine the optimal discretisation level for hurricane variables, the optimal number of clusters for loss variables in this example is 15. The cluster centroids in the PCA space are then mapped back to the original variable space using Eq. (11). Discretisation points are determined as the midpoints between adjacent cluster centroids, with 0 and 1 added as the lower and upper bounds of the discretisation intervals, respectively. The discretisation results

vary across tracts, and Fig. 7 presents the outcomes for L_1 to L_5 as an illustration.

The construction of CPTs requires defining the marginal distributions of the five hurricane variables (A_z , D_{st} , V_i , θ , and P_c) and the conditional distributions linking these hazard variables to the census tract loss nodes. The marginal distributions are derived from historical hurricane records in the vicinity of the study area, as illustrated in Fig. 8. In this figure, the area of each bin corresponds to the marginal probability of the respective discretisation interval. The conditional distributions are established using the scenario-based probabilistic loss assessment framework. While the initial training set contains 10,000 scenarios, parameter learning for CPTs requires comprehensive coverage of the state space (i.e., all possible combinations of the discretised hurricane variables). To ensure that each potential scenario combination is represented at

least once, an additional 5000 targeted samples were generated and integrated with the initial set, resulting in a total of 15,000 training cases for CPT estimation. Each case provides a 608×1000 matrix, representing the probability distribution of loss ratios across all 608 census tracts for a specific scenario. Notably, this dataset inherently captures the spatial correlations in regional damage arising from geographical proximity. Fig. 9 displays a heatmap of the correlation coefficients among the 608 loss variables, where darker colours indicate stronger positive correlations. The results confirm that correlations are significantly higher between geographically adjacent census tracts. By performing parameter learning on this comprehensive dataset, these spatial dependencies are naturally preserved within the CPTs of the BN model.

The computational efficiency of the proposed BN framework is significantly higher than that of conventional MCS. For a given hurricane scenario in the study area, scenario-based hurricane risk assessment using MCS typically requires computation times on the order of minutes (approximately 2–3 min on a standard desktop computer with an Intel i7 processor and 32 GB RAM). In contrast, once trained, the BN model performs inference through probabilistic propagation and can provide predictions almost instantaneously. This advantage is particularly important in hurricane risk assessment, where events may last for several days or even weeks and require frequent updates (e.g., every 15 min) [29]. Under such conditions, the proposed BN framework can substantially reduce computational demand and significantly improve efficiency for decision-making. Furthermore, the computational cost of MCS increases significantly with the size and complexity of the study area, whereas the inference time of the BN model remains relatively stable after training. This makes the proposed approach especially suitable for large-scale regional applications.

5.4. Model validation

To evaluate the effectiveness of the proposed hurricane loss model, this section validates its predictive performance using the landfall moments of 100 hurricane events. The selection criteria require hurricanes to reach at least Category 2 intensity on the Saffir–Simpson Hurricane Wind Scale and to pass within 100 km of Miami-Dade County. These thresholds are chosen because post-hurricane reports indicate that noticeable community damage typically begins with Category 2 storms [49]. The 100 km distance ensures the inclusion of hurricanes with substantial local impact, representing approximately twice the typical

$R_{\max}$ of 40–50 km [50]. Among the selected events, 27 are historical hurricanes recorded since 1900, while the remaining 73 are obtained from a synthetic hurricane database [51]. Fig. 10 illustrates the tracks of these 100 hurricanes, where red and yellow lines denote historical and synthetic events, respectively.

A key contribution of this study is the development of a supervised discretisation method that integrates weighted PCA with decision tree algorithms. To demonstrate its superiority, we compare the proposed method with two conventional approaches: equal-width (EW-BN) and equal-frequency (EF-BN) discretisation. Fig. 11(a) and (b) present the RMSE distributions for the mean and standard deviation of loss ratios across all census tracts, calculated based on the 100 test hurricanes. The results clearly indicate that the proposed supervised method yields significantly lower errors compared to traditional methods, validating its ability to capture the complex relationships between hurricane parameters and regional losses. Specifically, Fig. 12(a) and (b) present the RMSE distributions for the mean and standard deviation of loss ratios across the 608 census tracts based on the proposed model. The results demonstrate robust predictive performance, with the 90th percentile RMSE values for the mean and standard deviation being 0.025 and 0.023, respectively.

To evaluate the model's capability in capturing spatial variations, Hurricane Andrew (1992) is selected as a representative example to illustrate the spatial distribution of the mean and standard deviation of loss ratios across Miami-Dade County. Fig. 13(a) and (c) show the reference values obtained from a scenario-based risk assessment framework, while Fig. 13(b) and (d) present the corresponding estimates from the BN model. In these maps, darker colours indicate higher loss ratios, and grey areas denote regions with fewer than 10 residential buildings, which are excluded from the analysis. For this specific event, the RMSE for the estimated mean and standard deviation across all census tracts are 0.030 and 0.021, respectively. Beyond point estimates, the model's ability to predict the probability distribution is further verified in Fig. 14. Six census tracts representing different loss levels (at the 0.05, 0.20, 0.40, 0.60, 0.80, and 0.95 quantiles of the sorted loss ratios) are selected. As shown in the comparison between the BN-based predictions and the MCS reference results, the probability distributions align closely, further confirming the model's validity.

Finally, the spatial heterogeneity of the prediction errors is investigated. As observed in Figs. 11 and 12, some census tracts exhibit noticeably higher RMSE values. This variation is likely attributable to

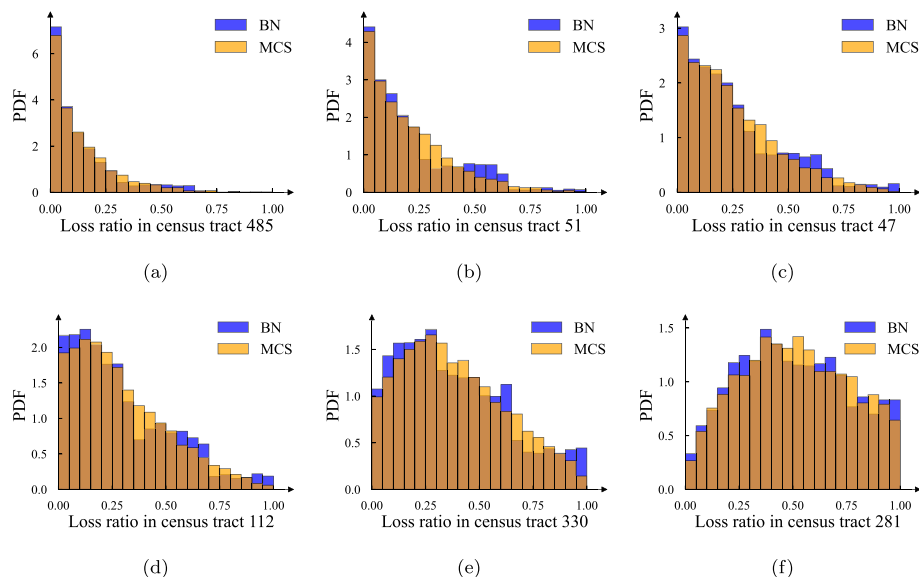


Fig. 14. Comparison of BN-based estimated loss ratio distributions and reference results for Hurricane Andrew (1992) across selected census tracts at different quantile levels: (a) 5th, (b) 20th, (c) 40th, (d) 60th, (e) 80th, and (f) 95th percentiles.

Table 2

RMSE comparison of loss ratios between high- and low-density census tracts.

Region type	RMSE (mean)	RMSE (standard deviation)
High-density regions (top 5%)	0.0145	0.0117
Low-density regions (bottom 5%)	0.0200	0.0170

the weighting factor (building density) incorporated into the discretisation process. To test this hypothesis, a comparative analysis is performed between census tracts in the top 5% and bottom 5% of building density, as summarised in Table 2. The results indicate that in high-density areas, the average RMSE for both the mean and standard deviation is approximately 30% lower than in low-density areas. These findings demonstrate that incorporating importance factors into the discretisation process effectively prioritises and reduces prediction errors in high-priority regions.

The above validation demonstrates that the proposed BN-based hurricane loss model can effectively capture the spatial distribution of losses across the study area, with improved predictive accuracy in high-priority regions. In future work, the proposed framework will be extended to multiple major cities along the US Atlantic and Gulf coasts. This will enable a more comprehensive evaluation of its applicability across regions with diverse building stocks and terrain characteristics. Furthermore, the computational efficiency of the BN model will be leveraged to support large-scale and long-term hurricane risk assessments under climate change scenarios.

6. Conclusion

This paper presents a BN framework for estimating regional building losses caused by hurricanes. The framework efficiently and accurately quantifies the spatial distribution of losses across large geographic areas under specified hurricane scenarios. Because the discretisation of continuous variables significantly affects model accuracy, a new discretisation scheme is introduced. The main findings are summarised as follows:

1. The model incorporates two types of variables: hurricane variables (including translation speed, heading angle, distance and azimuth of the hurricane centre relative to the study area, and central pressure) and loss variables (loss ratios for each geographic subdivision). The relationship between hurricanes and losses is established through a scenario-based probabilistic risk assessment framework.
2. Hurricane variables are discretised using a supervised approach that combines weighted PCA with decision tree algorithms, allowing the relative importance of each subdivision to be incorporated into the discretisation process. Loss variables are discretised using a clustering-based method to capture the characteristics of regional losses.
3. The framework is demonstrated through a numerical example, showing high predictive accuracy. The 90th percentile RMSE values for the mean and standard deviation of loss estimates across the study area are 0.025 and 0.023, respectively. In high-importance areas, prediction errors are approximately 30% lower than in low-importance areas due to the integration of importance factors into the discretisation process.

CRedit authorship contribution statement

Yu Liang: Writing – review & editing, Writing – original draft, Validation, Methodology, Formal analysis, Data curation, Conceptualization. **Hao Zhang:** Writing – review & editing, Validation, Supervision, Methodology, Conceptualization. **Cao Wang:** Writing – review & editing, Validation, Supervision, Conceptualization.

Declaration of generative AI and AI-assisted technologies in the writing process

During the preparation of this work the author(s) used ChatGPT to translate, paraphrase and correct grammar. After using this tool/service, the author(s) reviewed and edited the content as needed and take(s) full responsibility for the content of the publication.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This research has been supported by the China Scholarship Council Postgraduate Research Scholarship and the China Scholarship Council Supplementary Scholarship from the Faculty of Engineering, the University of Sydney, Australia. Cao Wang was supported by the Australian Government through the Australian Research Council's Discovery Early Career Researcher Award (DE240100207). These supports are gratefully acknowledged.

References

- [1] Smith AB. An active year of U.S. billion-dollar weather and climate disasters; 2024. <https://www.climate.gov/news-features/blogs/beyond-data/2024-active-year-us-billion-dollar-weather-and-climate-disasters> [Accessed: 15 March 2025].
- [2] Mushtaha AW, Alaloul WS, Baarimah AO, Musarat MA, Alzubi KM, Khan AM. A decision-making framework for prioritizing reconstruction projects in post-disaster recovery. *Results Eng* 2025;25:103693.
- [3] Vitoontus S. Risk assessment of building inventories exposed to large scale natural hazards [Ph.D. thesis]. Georgia Institute of Technology, United States; 2012.
- [4] Ma L, Christou V, Bocchini P. Framework for probabilistic simulation of power transmission network performance under hurricanes. *Reliab Eng Syst Saf* 2022;217:108072.
- [5] Amini K, Padgett JE. Probabilistic risk assessment of hurricane-induced debris impacts on coastal transportation infrastructure. *Reliab Eng Syst Saf* 2023;240:109579.
- [6] Zheng Y, Xie Y, Long X. A comprehensive review of Bayesian statistics in natural hazards engineering. *Nat Hazard* 2021;108:63–91.
- [7] Kaikkonen L, Parviainen T, Rahikainen M, Uusitalo L, Lehtikainen A. Bayesian networks in environmental risk assessment: a review. *Integr Environ Assess Manage* 2021;17:62–78.
- [8] Guo Y, Wang H, Guo Y, Zhong M, Li Q, Gao C. System operational reliability evaluation based on dynamic Bayesian network and XGBoost. *Reliab Eng Syst Saf* 2022;225:108622.
- [9] Liu C, Sun J, Wang F, Ning S, Xu G. Bayesian network method for fault diagnosis of civil aircraft environment control system. *Proc Inst Mech Eng I J Syst Control Eng* 2020;234:662–74.
- [10] Aydin M, Akyuz E, Boustras G. A holistic safety assessment for cargo holds and decks fire & explosion risks under fuzzy Bayesian network approach. *Saf Sci* 2024;176:106555.
- [11] Zhou T, Zhang L, Hu J, Modarres M, Droguett EL. A critical review and benchmark study of dependency modeling for seismic probabilistic risk assessment in the nuclear power industry. *Reliab Eng Syst Saf* 2024;245:110009.
- [12] Wang HS. A unified dynamic model of electrodermal activity [Ph.D. thesis]. Northeastern University, United States; 2023.
- [13] Van Verveveld HCW, Van Dongeren AR, Plant NG, Jäger WS, Den Heijer C. Modelling multi-hazard hurricane damages on an urbanized coast with a Bayesian network approach. *Coast Eng* 2015;103:1–14.
- [14] Li B, Chen Y, Huang S, Guan H, Xiong Y, Mei S. A Bayesian network model for predicting outages of distribution system caused by hurricanes. In: *IEEE power & energy society general meeting*. Quebec, Canada; 2020. p. 1–5.
- [15] Cai B, Zhang Y, Wang H, Liu Y, Ji R, Gao C, Kong X, Liu J. Resilience evaluation methodology of engineering systems with dynamic-Bayesian-network-based degradation and maintenance. *Reliab Eng Syst Saf* 2021;209:107464.
- [16] Lu Q, Zhang W. Integrating dynamic Bayesian network and physics-based modeling for risk analysis of a time-dependent power distribution system during hurricanes. *Reliab Eng Syst Saf* 2022;220:108290.
- [17] Wang C, Liu Y, Zhang X, Li X, Paramygin V, Sheng P, Zhao X, Xu S. Scalable and rapid building damage detection after hurricane using causal Bayesian networks and InSAR imagery. *Int J Disaster Risk Reduct* 2024;104:104371.
- [18] Beuzen T, Marshall L, Splinter KD. A comparison of methods for discretizing continuous variables in Bayesian networks. *Environ Model Softw* 2018;108:61–6.

- [19] Fayyad UM. Multi-interval discretization of continuous-valued attributes for classification learning. In: International joint conference on artificial intelligence. Savoie, France; 1993. p. 1022–9.
- [20] Costa VG, Pedreira CE. Recent advances in decision trees: an updated survey. *Artif Intell Rev* 2023;56:4765–800.
- [21] Kerber R. ChiMerge: discretization of numeric attributes. In: Proceedings of the tenth national conference on artificial intelligence. California, the United States; 1992. p. 123–8.
- [22] Ellingwood BR, Rosowsky DV, Li Y, Kim JH. Fragility assessment of light-frame wood construction subjected to wind and earthquake hazards. *J Struct Eng* 2004;130:1921–30.
- [23] Li Y, Ellingwood BR. Hurricane damage to residential construction in the US: importance of uncertainty modeling in risk assessment. *Eng Struct* 2006;28:1009–18.
- [24] Ellingwood BR, Wang N, Harris JR, McAllister TP. Performance-based engineering to achieve community resilience. In: Sustainable and resilient infrastructure. Routledge, Oxfordshire, England; 2018. p. 94–111.
- [25] FEMA. Multi-hazard loss estimation methodology: hurricane model. Technical Report MH 2.1, Department of Homeland Security, Washington, D.C.; 2014.
- [26] Fenton N, Neil M. Risk assessment and decision analysis with Bayesian networks. 2nd ed. CRC Press, United States; 2018.
- [27] Teixeira LV, Assuncao RM, Loschi RH. A generative spatial clustering model for random data through spanning trees. In: IEEE international conference on data mining. New Jersey, the United States; 2015. p. 997–1002.
- [28] Harper B. Tropical cyclone parameter estimation in the Australian region: wind-pressure relationships and related issues for engineering planning and design. Technical Report J0106-PR003E, Woodside Australian Energy, Australia; 2002.
- [29] Vickery PJ, Wadhera D. Statistical models of Holland pressure profile parameter and radius to maximum winds of hurricanes from flight-level pressure and H*Wind data. *J Appl Meteorol Climatol* 2008;47:2497–517.
- [30] Lin P, Wang N. Building portfolio fragility functions to support scalable community resilience assessment. *Sustain Resilient Infrastruct* 2016;1:108–22.
- [31] Chen X, Ma X, Jia L, Zhang Z, Chen F, Wang R. Causative analysis of freight railway accident in specific scenes using a data-driven Bayesian network. *Reliab Eng Syst Saf* 2024;243:109781.
- [32] Greenacre M, Groenen PJF, Hastie T, D'Enza AI, Markos A, Tuzhilina E. Principal component analysis. *Nat Rev Methods Primers* 2022;2:1–21.
- [33] Jun S. Evolutionary algorithm for improving decision tree with global discretization in manufacturing. *Sensors* 2021;21:2849.
- [34] Liang Y. Probabilistic cyclone damage assessment on large spatially distributed civil infrastructure systems [Ph.D. thesis]. The University of Sydney, Australia; 2025.
- [35] Wang C, Teh LH, Feng K. A probabilistic framework for resilience quantification of residential building portfolios exposed to tropical cyclone winds. *ASCE-ASME J Risk Uncertain Eng Syst A Civ Eng* 2023;9:4023040.
- [36] Homeguide. How much does it cost to frame a house; 2025. <https://homeguide.com/costs/cost-to-frame-a-house> [Accessed: 28 August 2025].
- [37] Salim MZ, Qiang Y, Dixon B, Collins J. A disparate disaster: spatial patterns of building damage caused by hurricane Ian and associated socio-economic factors. *Remote Sens* 2024;16:3792.
- [38] Rouhana F, Zhu J, Bagtzoglou AC, Burton CG. Analyzing structural inequalities in natural hazard-induced power outages: a spatial-statistical approach. *Int J Disaster Risk Reduct* 2025;117:105184.
- [39] Georgiou P. Design wind speeds in tropical cyclone-prone regions [Ph.D. thesis]. The University of Western Ontario, Canada; 1986.
- [40] Vickery PJ, Masters FJ, Powell MD, Wadhera D. Hurricane hazard modeling: the past, present, and future. *J Wind Eng Ind Aerodyn* 2009;97:392–405.
- [41] Vickery PJ, Skerlj PF. Hurricane gust factors revisited. *J Struct Eng* 2005;131:825–32.
- [42] Zeng D, Zhang H, Li Q, Ellingwood BR. Tropical cyclone damage assessment of distributed infrastructure systems under spatially correlated wind speeds. *Struct Saf* 2021;91:102080.
- [43] Krieger N, Chen JT, Waterman PD, Soobader M-J, Subramanian SV, Carson R. Geocoding and monitoring of US socioeconomic inequalities in mortality and cancer incidence: does the choice of area-based measure and geographic level matter. *Am J Epidemiol* 2002;156:471–82.
- [44] Zeng D, Zhang H, Dai H, Beer M. Scalable risk assessment of large infrastructure systems with spatially correlated components. *Struct Saf* 2023;101:102311.
- [45] NOAA. Atlantic hurricane dataset - HURDAT2; 2023. <http://www.aoml.noaa.gov/hrd/hurdat/hurdat2.html> [Accessed: 22 February 2023].
- [46] Jerdee M, Kirkley A, Newman M. Normalized mutual information is a biased measure for classification and community detection. *Nat Commun* 2025;16:11268.
- [47] Law AM. Simulation modeling and analysis. McGraw-Hill Education, Dubuque, the United States; 2013.
- [48] Satopaa V, Albrecht J, Irwin D, Raghavan B. Finding a kneedle in a haystack: detecting knee points in system behavior. In: International conference on distributed computing systems workshops. Washington, D.C. the United States; 2011. p. 166–71.
- [49] Wikipedia. List of costliest tropical cyclones; 2025. https://en.wikipedia.org/w/index.php?title=List_of_costliest_tropical_cyclones&oldid=1300647065 [Accessed: 17 July 2025].
- [50] Hsu SA, Yan Z. A note on the radius of maximum wind for hurricanes. *J Coastal Res* 1998;14:667–8.
- [51] Bloemendaal N, Haigh ID, De Moel H, Muis S, Haarsma RJ, Aerts JCJH. Generation of a global synthetic tropical cyclone hazard dataset using STORM. *Sci Data* 2020;7:40.