

Bayesian Multimodal Models for Risk Analyses of Low-Probability High-Consequence Events

O. Arda Vanli

Abstract This paper reviews a set of Bayesian model updating methodologies for quantification of uncertainty in multi-modal models for estimating failure probabilities in rare hazard events. Specifically, a two-stage Bayesian regression model is proposed to fuse an analytical capacity model with experimentally observed capacity data to predict failure probability of residential building roof systems under severe wind loading. The ultimate goals are to construct fragility models accounting for uncertainties due to model inadequacy (epistemic uncertainty) and lack of experimental data (aleatory uncertainty) in estimating failure (exceedance) probabilities and number of damaged buildings in building portfolios. The proposed approach is illustrated on a case study involving a sample residential building portfolio under scenario hurricanes to compare the exceedance probability and aggregate expected loss to determine the most cost-effective wind mitigation options.

1 Introduction

In wind engineering and insurance underwriting it is of interest to predict the failure probability of built structures under hurricane force winds. Fragility functions, or conditional probability of failure under a given wind loading, are used to model the failure occurrence. Failure is assumed to occur when the wind pressure exceeds the component capacity for a given mode of failure. Fragility curve for a residential structure depend on many factors, including shape of roof (gable or hip), frame material (wood or masonry), number of stories, roof to wall connection type (toe-nail or hurricane clip) and terrain roughness. For example, HAZUS [28], a commonly used hazard analysis software package for loss estimate analysis, contains hundreds of fragility curves for single family residential structures, depending on the combination

O. Arda Vanli

Department of Industrial and Manufacturing Engineering, Florida A& M University – Florida State University College of Engineering, Tallahassee, FL 32310, USA, e-mail: oavanli@eng.fsu.edu

of these factors. One of the main challenges in predicting failures of built structures against extreme hurricane events from available loss data, that come in the form of either insurance claims or field surveys, is that extreme events are rare therefore the loss data is highly sparse, resulting in large uncertainties in performance predictions. The lack of available data of component performance in high wind speed domains is particularly problematic. This region is referred to as the low-probability high-consequence event and is thought to have significant influence in a cost-benefit assessment [13, 6, 15]. In order to reduce uncertainty in the predictions it is of interest to combine physics-based failure prediction functions for structural components with data from field surveys or insurance claims. This chapter presents a review of Bayesian model updating approaches to deal with uncertainty in predicting failure in rare but high consequence events. Specifically, we study methods to develop fragility functions of building components that can in turn be used quantify uncertainty of failure probabilities against extreme wind loading and make decisions for aggregate wind loss of building inventories. Many of the ideas presented in this chapter have been previously discussed in the author's earlier research, including [27, 31] and [32].

2 Models to Predict Hurricane Risk

The probability of failure P_f of a structural component for a certain failure mode is defined as:

$$P_f = \int F_v p(v) dv \quad (1)$$

in which $p(v)$ is the probability density function of the hazard, in this case hurricane wind speed v , and F_v is the fragility function, the probability that the structure fails at this wind speed. Fragility analysis using this equation uncouples the hazard (probability distribution of wind load) from the structure reliability (fragility function), thus, the analyses for determining hazard likelihood and fragility of various components can be conducted separately. It is often the case that accurate information on hazard is not available, hence it is desirable to make safety decisions against a range of hazard intensities through the use of fragility analysis [14]. The fragility is the conditional probability that capacity of the component is less than a wind load D_v for a given wind speed v

$$F_v = P(C \leq D_v) = \int_{\eta \in \Omega} f(\eta) d\eta$$

in which $f(\eta)$ is the joint probability density function of the parameters η that define the capacity and wind load models, and Ω is the set of values of η such that $C \leq D_v$. The integral is usually evaluated using numerical integration or first-order reliability methods (Madsen et al. [16]). In wind engineering, fragility functions are often

assumed to follow certain probability function, where the lognormal distribution is one of the most commonly used distribution (Rosowsky and Ellingwood [23], Li and Ellingwood [14]):

$$F_v = \Phi [\ln (D_v/m)/\zeta] \quad (2)$$

in which $\Phi[\bullet]$ is the standard normal probability integral, m is the median capacity, and ζ is the logarithmic standard deviation of capacity. Li and Ellingwood [14] found the fragility curves for hurricane winds but did not provide bounds on the curve. Gardoni et al., [7] Straub and der Kiureghian [26] considered quantifying uncertainty in fragility by provided confidence bounds, however, the application was limited to earthquake loss analysis.

In a recent survey of hurricane vulnerability analysis methods Pita et al.,[19] mentioned that uncertainty in the loss estimates depends on the wind speed domain under study. Uncertainty is typically larger in the lower and the higher wind speed ranges. The former uncertainty is because insurance claim data typically employed to fit fragility functions do not include any damage lower than the deductibles, while the latter one is because there is scarce past loss data due to the rare occurrence of strong hurricanes. To circumvent issues with scarce performance data, researchers have investigated physics-based analytical approaches to model building vulnerability. For example Ellingwood et al., [5] and Zhang et al., [29] developed capacity models of building components from which fragility curves can be found to assess the response of a light-frame wooden construction exposed to extreme winds and earthquakes. It is assumed that the severity of a catastrophe is based on annual probability of exceeding the design hazard or its return period.

Most analytical approaches to estimate the load carrying capacity of building components based on Newtonian-mechanics principles do not account for uncertainty that arises due to modeling assumptions (Mottershead and Friswell, [18]). Several assumptions made at the modeling stage can contribute significantly to model uncertainty that results in biased capacity predictions: variation of material properties during manufacture, inexact modeling of material constitutive behavior, inaccurate modeling of the boundary conditions, and insufficient refinement in spatial discretization of distributed parameters. To correct the bias between computer predicted deterministic model and experimental observations, a model-updating method is employed. The major purpose of model updating is to modify the model parameters to obtain a better agreement between numerical results and the test data. According to Straub and der Kiureghian [26], the uncertainties in the model gives rise to statistical dependence among observations and can have significant effect on fragility. Gardoni et al., [7] have used Bayesian framework for constructing univariate and multivariate predictive capacity model based on experimental observations and develop fragility curves with uncertainty bounds. Beck and Au [3] have proposed a Bayesian framework for finding probability distribution of the model parameters for structural analysis and Beck and Katafygiotis [4] have used an adaptive Markov chain Monte Carlo simulation to find updated posterior probability using a sampling based approach rather than closed-form expressions.

In engineering statistics improving the predictive accuracy of computer models with physical or experimental data has been studied extensively and referred to as model updating and model calibration (Kennedy and O'Hagan [12], Reese et al., [22], Qian and Wu [20]). Vanli and Jung [27] have used the probabilistic model updating method to improve the accuracy of damage size and location prediction of a structural health monitoring system with help of a finite element analysis method. Model uncertainty is typically categorized into the forms epistemic and aleatory uncertainty. Epistemic uncertainty derives from a lack of knowledge about the appropriate value to use for a quantity that is assumed to have a fixed value in the context of a particular analysis. Aleatory uncertainty arises because the system under study can behave in many different ways (e.g. many different accidents are possible at a power station). Thus, aleatory uncertainty is a property of the system under study and epistemic uncertainty is a property of the analysis (Helton et al., [11]). Often, some type of random sampling through a set of event trees is used to sample from aleatory uncertainty and random sampling from the input distributions is used to sample from epistemic uncertainty. In addition, to provide a better coverage of low probability/high consequence events and enhance the effectiveness of the computational effort, importance sampling is recommended as a more effective sampling procedure than random sampling.

3 Bayesian Model Updating for Hurricane Risk

The parameters of both functions $p(v)$ and F_v are determined from data and therefore are subject to estimation error. In this study we represent the fragility function in terms of random parameters using a Bayesian approach and evaluate the integral in Equation (1) with respect to the posterior distribution of the parameters. A two-stage Bayesian model updating approach (Reese et al., [22]) is presented to fuse capacity predictions from an analytical model and capacity data from experiments. The Bayesian posterior distribution of the capacity is used to develop posterior distribution and confidence bounds of fragility functions, total failure probability and aggregate hurricane loss.

3.1 Bayesian Two-stage Regression Model

Suppose $C(\mathbf{x})$ is the capacity defined as a function of $\mathbf{x} = (x_1, x_2, \dots, x_d)$, a $d \times 1$ feature (regressor) vector that contains the variables denoting the number and types of clips/nails and wood of the roof to wall connection. A polynomial regression is used to represent the relation between the connection capacity and the connection features:

$$C(\mathbf{x}) = \beta' \mathbf{x} + \epsilon \quad (3)$$

where $\boldsymbol{\beta} = (\beta_1, \beta_2, \dots, \beta_d)$, a $d \times 1$ vector of unknown parameters (regression coefficients) that quantifies the effects of these variables on capacity. Polynomial regression model is a class of multiple linear regression models, that are linear in the regression parameters, but contain polynomial functions of the independent variables, including, x , x^2 , and x_1x_2 (Montgomery et al., [17]). A normally distributed random error term $\epsilon \sim N(0, \sigma^2)$ is assumed to capture all un-modeled sources of variation.

A first stage of the model describes the analytical capacities (described in Section 3.2) as a function of the features using Bayesian regression. A second stage of the model updates the first stage model with the experimental observations of capacity. The same model form as in Equation (3) is assumed in both first and second stages. Bayesian inference relies on specifying prior probability distributions on the unknown parameters and developing posterior distributions (conditional on observed data). In the first stage a *non-informative prior* is assumed on the analytical capacity model parameters as $p(\boldsymbol{\beta}, \sigma^2) \propto 1/\sigma^2$. The random error is assumed to be normally distributed, therefore, the likelihood function is derived based on this assumption and the regression model structure. Based on the regression model and the random error distribution, the likelihood function for an individual capacity for a regressor vector $\mathbf{x}$ is

$$p(C(\mathbf{x})|\boldsymbol{\beta}, \sigma^2) = N(\boldsymbol{\beta}'\mathbf{x}, \sigma^2). \quad (4)$$

The *first stage* model posterior distributions of $\boldsymbol{\beta}$ and σ^2 are obtained by updating the noninformative prior with the observed analytical capacity data (Gelman et al., [9], pp 355-356) using the Bayes' theorem. Let $\mathbf{C}_a = (c_{a1}, \dots, c_{an_1})$ be the vector of analytical capacity data for n_1 different connection configurations (e.g., wood type, connection type), specified by a $n_1 \times d$ matrix $\mathbf{X}_1$ of regressor where each row represents the features of a specific connection. The posteriors are therefore obtained as

$$\begin{aligned} p(\boldsymbol{\beta}|\sigma^2, \mathbf{C}_a) &= MVN(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1) \\ p(\sigma^2|\mathbf{C}_a) &= Inv - \chi^2(\nu_1, s_1^2) \end{aligned}$$

where $MVN(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1)$ denotes a d -dimensional multivariate normal probability distribution with mean vector $\boldsymbol{\mu}_1$ and variance-covariance matrix $\boldsymbol{\Sigma}_1$, and $Inv - \chi^2(\nu_1, s_1^2)$ denotes a scaled-inverse chi-square distribution with degree of freedom ν_1 and scale parameter s_1^2 . The parameters of the posterior distributions are

$$\begin{aligned} \boldsymbol{\mu}_1 &= (\mathbf{X}_1'\mathbf{X}_1)^{-1} \mathbf{X}_1'\mathbf{C}_a \\ \boldsymbol{\Sigma}_1 &= \sigma^2 (\mathbf{X}_1'\mathbf{X}_1)^{-1} \\ \nu_1 &= n_1 - d \\ s_1^2 &= (n_1 - d)^{-1} (\mathbf{C}_a - \mathbf{X}_1\boldsymbol{\mu}_1)' (\mathbf{C}_a - \mathbf{X}_1\boldsymbol{\mu}_1) \end{aligned}$$

The second stage model posteriors are obtained by taking the posteriors of the first stage as the prior of the parameters and updating the priors with the experimental data. The second stage posteriors are therefore the informative prior distributions on the parameters β and σ^2 as follows:

$$\begin{aligned} p(\beta|\sigma^2) &= MVN(\beta_0, \sigma^2 \Lambda_0) \\ p(\sigma^2) &= Inv - \chi^2(\nu_0, s_0^2). \end{aligned}$$

in which, β_0 , Λ_0 , ν_0 and s_0^2 are the parameters of the prior distributions to be specified by the user according to the posterior of the first stage. We tune the parameters of the priors of the second stage so that they are equal to those of the posterior of the first stage as

$$\begin{aligned} \beta_0 &= \mu_1 \\ \Lambda_0 &= (X_1' X_1)^{-1} \\ \nu_0 &= \nu_1 \\ s_0^2 &= s_1^2. \end{aligned}$$

These set of priors are referred to as *conjugate* priors for a normal regression model (Gelman et al., [9]). When the prior is conjugate for the likelihood model, the posterior distribution follows the same parametric form as the prior distribution. Conjugate family has the convenience that the posterior can be represented as a closed-form expression. Let $C_e = (c_{e1}, \dots, c_{en_2})$ be the vector of capacity measurements obtained for n_2 different connection configurations, under a set of regressors specified by a $n_2 \times d$ matrix X_2 . Note that in general the number of data n_1 and n_2 and the connector configurations given in the regressor matrices X_1 and X_2 of the first and second stages can be different. By using these priors and assuming a normal likelihood function for the observed experimental capacity data C_e , it can be shown that the posteriors of β and σ^2 for the *second stage* are:

$$p(\beta|\sigma^2, C_e, C_a) = MVN(\mu_2, \Sigma_2) \quad (5)$$

$$p(\sigma^2|C_e, C_a) = Inv - \chi^2(\nu_2, s_2^2) \quad (6)$$

for which the parameters are defined as

$$\mu_2 = (\Lambda_0^{-1} + X_2' X_2)^{-1} (\Lambda_0 \beta_0 + X_2' C_e) \quad (7)$$

$$\Sigma_2 = \sigma^2 (X_2' X_2)^{-1} \quad (8)$$

$$\nu_2 = n_2 + \nu_1 \quad (9)$$

$$s_2^2 = \frac{1}{\nu_2} [\nu_1 s_1^2 + (C_e - X_2 \mu_2)' (C_e - X_2 \mu_2) + (\mu_1 - \mu_2)' \Lambda_0 (\mu_1 - \mu_2)] \quad (10)$$

Equations (7) to (10) show how the noninformative prior on β and σ^2 is updated with n_1 analytical capacity data C_a (first stage) and n_2 experimental capacity data C_e (second stage). Note that the dependence of the posteriors on the analytical capacity C_a is through the parameters s_1^2 and μ_1 .

Note that the normality assumption of the error term in the Bayesian regression models is made to satisfy the conjugate prior condition and to obtain the posterior distributions in closed-form. However, this assumption does not have to be satisfied in applications and for such cases a typical approach is to apply a variable transformation on the response variable in the form $c = f_T(\tilde{c})$ so that the regression on the transformed variable has normal errors ([20], pp 161). Here $\tilde{c}$ is the untransformed capacity data, c is the transformed capacity to be used in the regression analysis and $f_T(\cdot)$ is a suitable transformation function. The Bayesian model is fitted to the transformed capacity data and the prediction on the original (untransformed) capacity is obtained by applying the inverse transformation. In the case study presented in Section 4.2 the experimental data required a logarithmic transformation, $f_T(\tilde{c}) = \log(\tilde{c})$, be used.

3.2 Posterior Predictive Distribution

We make inferences about the capacity of a component based on the posterior probability (predictive) distribution of the capacity $C(\mathbf{x})$ from the two-stage updating of the model (3) and using the posteriors we developed for the parameters. The posterior predictive distribution of the capacity from the two-stage updated model is (Gelman et al., [9], pp 358-359) obtained from the likelihood given by specified by Equation (4) and the posteriors of parameters β and σ^2 given by the Equations (5) and (6), respectively, by computing the following integral

$$p(C(\mathbf{x})|C_e, C_d) = \int p(C(\mathbf{x})|\beta, \sigma^2) p(\beta|\sigma^2, C_e, C_a) p(\sigma^2|C_e, C_a) d\beta d\sigma^2. \quad (11)$$

This equation obtains the predictive distribution, by integrating the likelihood function $p(C(\mathbf{x})|\beta, \sigma^2)$, a conditional distribution based on specific values of parameters β and σ^2 , over the posteriors $p(\beta|\sigma^2, C_e, C_a)$ and $p(\sigma^2|C_e, C_a)$. Through this integration the Bayesian predictive distribution accounts for uncertainty in the model parameters. The fragility at a given wind load defined by Equation (2) is determined by integrating the predictive distribution (11) of the capacity to find the probability that capacity is less than the wind load, $F_v = P(C(\mathbf{x}) \leq D_v)$, as follows

$$F_v = \int_0^{D_v} p(C(\mathbf{x})|C_e, C_a) dC. \quad (12)$$

in which the wind load D_v for the bound of the integral is determined according to ANSI/ASCE-7 [4] for given wind speed v and roof configuration (as discussed in the case study). Based on (4), the fragility conditional on β, σ^2 is

$$F_v|\boldsymbol{\beta}, \sigma^2 = P(C(\mathbf{x}) \leq D_v|\boldsymbol{\beta}, \sigma^2) = \Phi\left(\frac{D_v - \boldsymbol{\beta}'\mathbf{x}}{\sigma}\right) \quad (13)$$

where, as before, $\Phi(\cdot)$ is the cumulative distribution function of the standard normal. In structural reliability literature [16], sometimes this is expressed as $F_v|\boldsymbol{\beta}, \sigma^2 = 1 - \Phi(\gamma_v)$ where $\gamma_v = (\boldsymbol{\beta}'\mathbf{x} - D_v) / \sigma$ is the reliability index. To obtain the unconditional fragility, this expression is averaged over the posteriors, similarly as in Equation (11):

$$F_v = \int \Phi\left(\frac{D_v - \boldsymbol{\beta}'\mathbf{x}}{\sigma}\right) p(\boldsymbol{\beta}|\sigma^2, C_e, C_a) p(\sigma^2|C_e, C_a) d\boldsymbol{\beta} d\sigma^2.$$

Fragility curve is obtained point-by-point by finding F_v for a range of wind speeds using the computational approach presented in Section 3.3. The case study will illustrate the application of this integration.

3.3 Computational Methods to Estimate Rare Event Probabilities

In this section we discuss how to compute the probability of failure by computing the convolution integral (1) from the fragility curve F_v and the probability density function $p(v)$. Specifically, we highlight here the computational challenges in finding probabilities of rare events and propose an efficient computational approach to overcome some of them.

In many applications, in particular to compute the probability of failure, the predictive distribution of capacity or the fragility function is typically not needed to be obtained in closed-form. A common approach is to use Monte Carlo simulation [12] to estimate this distribution from randomly generated realizations, and will be used in this chapter. Samples of C can be drawn from the predictive distribution (11) using the following steps (i) Draw σ^2 from distribution (6); (ii) Conditional on σ^2 , draw $\boldsymbol{\beta}$ from distribution (5); (iii) Conditional on σ^2 and $\boldsymbol{\beta}$, draw C from distribution (4); (iv) Conditional on σ^2 and $\boldsymbol{\beta}$, find F_v using (13).

Figure 1 shows, as an illustration, a histogram of wind speed values from a given distribution $p(v)$ overlayed with the fragility curve F_v for a specific roof to wall connection (2 hurricane clips and SYP wood) evaluated with this approach. We discuss the definition in more detail in the case study below how the fragility curve and the wind speed probability distribution are defined. As it can be seen, the two distributions overlap only at the right tail of the wind speed distribution (with very few common occurrences) that makes integration with standard Monte Carlo methods very challenging resulting in large sampling errors. To reduce the computational effort involved with standard Monte Carlo, we use importance sampling (Rubinstein and Kroese [24], pp. 135) to estimate P_f .

In importance sampling a proposal distribution $g(v)$ is used in place of the density function of wind speed $p(v)$ in the integral to find the expectation of the weighted fragility as follows:

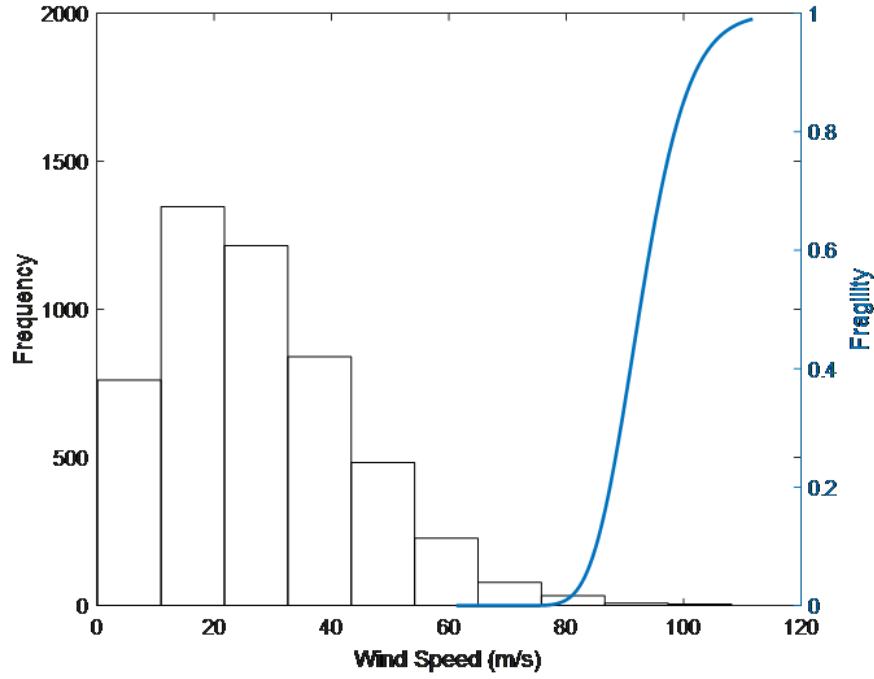


Fig. 1 Wind speed distribution and the fragility curve for a roof-to-wall connection with SYP wood and 2 hurricane clips

$$P_f = \int_0^{\infty} F_v \frac{p(v)}{g(v)} g(v) dv = \int_0^{\infty} F_v^* g(v) dv \quad (14)$$

The new integral can be viewed as the convolution of the weighted fragility $F_v^* = [p(v)/g(v)] F_v$ with the proposal distribution $g(v)$. As proposal distribution, in this study we use a Weibull distribution “translated” to the right so that samples in the right-tail (low probability) region of wind speed can be generated more frequently and failure probability can be more adequately estimated. It turns out that if the amount of translation is chosen carefully, the Monte Carlo integration will result in high precision with reasonable computational burden. We use Algorithm 1 below to obtain the probability of failure based on importance sampling.

Algorithm 1: Obtain posterior distribution and estimate of probability of failure

1. Given wind speed distribution $p(v)$ and proposal distribution $g(v)$. Set Monte Carlo simulation counter $i = 1$.

2. Simulate wind speed v_i from $g(v)$
3. Compute wind load D_{v_i}
4. Simulate σ^2 from the posterior $Inv - \chi^2(v_2, s_2^2)$
5. Simulate β from the posterior $MVN(\mu_2, \Sigma_2)$ based on the simulated σ^2
6. Find failure probability F_{v_i} from Equation (11).
7. Go to step 4. Repeat m times. The simulated fragility values are $(F_{v_i}^1, \dots, F_{v_i}^m)$
8. Compute weighted fragility values $F_{v_i}^{j*} = F_{v_i}^j p(v_i) / g(v_i)$ for $j = 1, 2, \dots, m$, a random draw from posterior of fragility at wind load D_{v_i}
9. Find the $\alpha/2$ -th and $(1 - \alpha/2)$ -th quantiles of the posterior $(F_{v_i}^{1*}, \dots, F_{v_i}^{m*})$; this is a 100 $(1 - \alpha)$ % confidence interval on the true fragility at wind load D_{v_i} .
10. Find the failure probability in i -th simulation as

$$P_{f_i} = \frac{1}{m} \sum_{j=1}^m F_{v_i}^{j*}$$

11. If $i < M$ then set $i = i + 1$ and go to Step 2 to continue with drawing a new wind speed v_i . If $i = M$ then go to Step 12.
12. The sampled values $(P_{f_1}, \dots, P_{f_M})$ is a random draw from the posterior of the failure probability. The importance sampling estimate of the probability of failure is

$$\hat{P}_f = \frac{1}{M} \sum_{i=1}^M P_{f_i}$$

13. Find the $\alpha/2$ -th and $(1 - \alpha/2)$ -th quantiles of the posterior. This is a 100 $(1 - \alpha)$ % confidence interval on the true failure probability.

4 Case Study: Miami-Dade County Residential Building Inventory Wind Loss Estimation

In this section we present the proposed Bayesian multimodal analysis method to estimate failure probability of roof systems under hurricane wind loading. Catastrophic failures of one and two story, light-frame residential buildings are the most frequently observed types of loss in a hurricane. The failure of roof-to-wall connections is a dominant cause of the breach of the building envelope (roof sheathing). Breach of the building envelope constitutes a significant component of hurricane loss because possible subsequent water and wind damage to the interior and the contents of the building can be very high. Roof-to-wall connections play a key role in load transfer from roof to walls during heavy wind and is the most commonly failing component under hurricane winds under tension load due to uplift forces acting on the roof [25, 23].

A gable roof, site-built (one or two story) residential building is considered as the model building, as shown in Figure 2, with dimensions $B = 17.1$ meters, $L = 13.4$ meters, $h1 = 3.05$ meters and $h2 = 2.8$ meters. The house considered in this example is adopted from Florida Public Hurricane Loss Projection Model [13]. The roof-to-wall connections are at every connection between the rafter and the top plate. With this configuration, there are 31 connections on each side of the house and end trusses have 8 connections, therefore, the roof structure has a total of 70 connections. The roof-to-wall connection types considered in this chapter are toe nails and clips, representing houses built in 1990s and after 2006, respectively, and will be referred to as “unmitigated” and “mitigated” houses. Mishra et al., [31] estimated the probability of failure for the first connection. Mishra et al., [32] presented a system-level reliability analysis of the roof system to estimate probability of failure for a given damage severity (e.g., failure of 10% of the connections). Roof system components representing mitigation actions (hurricane clips) and no mitigation (toe-nailing the rafters) are considered to evaluate the benefit of mitigations.

4.1 Bayesian two-stage model to predict failure probability of roof systems under hurricane winds

This section illustrates the application of the two-stage Bayesian updating approach for finding the predictive distribution of the capacity of a roof-to-wall connection and finding the probability of failure against wind loading. The analytical capacity model for the connections is described in detail in [31] based on the failure mode

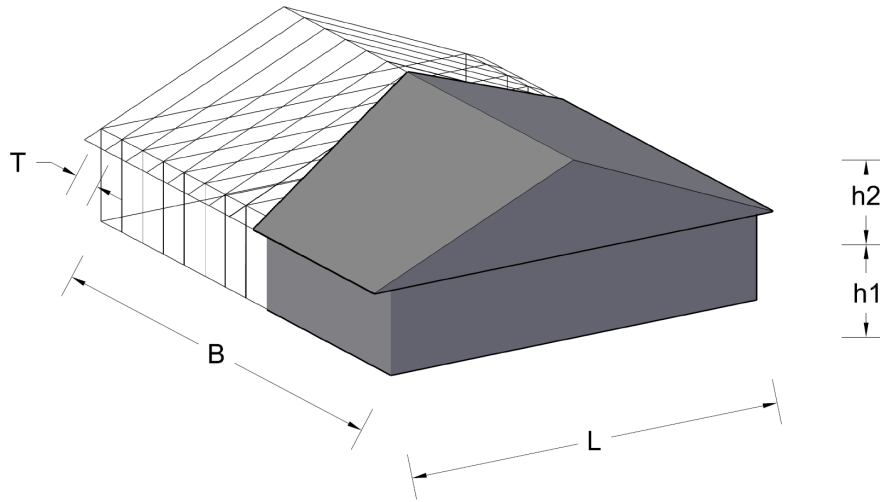


Fig. 2 The schematic of the house considered in the analyses

of nail pull out for toenails and the failure modes of nail pull out, clip tearing and wood rupture following for hurricane clips. The nail pullout force is determined by the individual contribution from each nail in the clip. According to this formula, the analytical capacity C_a [kN] is the smallest of resistances against three failure modes: the combined resistance offered by the nails, the resistance against wood rupture and the resistance against clip deformation. The experimental capacity C_e [kN] data for toe nails (reported in [21]) and for clips (reported in [1]) are used for Bayesian updating of analytical capacity models. We will compare the proposed Bayesian model updating approach to the traditional lognormal fragility method [14].

We considered the capacity data for Spruce Pine Fir (SPF), Southern Yellow Pine (SYP) and Douglas Fir (DF) wood types and 1, 2 and 4 hurricane clip configurations and the toe-nail connection. For the hurricane clips capacity, we fit a second order polynomial as the regression model to capacity data (after applying some suitable transformation)

$$C(\mathbf{x}) = \delta_0 + \delta_1 x_1 + \delta_{11} x_1^2 + \delta_2 x_2 + \delta_3 x_3 + \epsilon \quad (15)$$

where $C(\mathbf{x})$ is the connection capacity, $\boldsymbol{\beta} = (\delta_0, \delta_1, \delta_2, \delta_3, \delta_{11})$ is the vector of regression coefficients, containing 5 parameters to be estimated, and $\mathbf{x} = (1, x_1, x_2, x_3, x_1^2)$ is the vector of regressors. The regressor x_1 denotes the number of clips (takes values of 1, 2 and 4) and x_2, x_3 the dummy regressors that specify the 3 wood types: for SPF $(x_2, x_3) = (0, 0)$, for SYP $(x_2, x_3) = (1, 0)$ and for DF $(x_2, x_3) = (0, 1)$. A quadratic term x_1^2 for the number of connections is included to accommodate for possible curvature in capacity due to varying number of connections. The quadratic term and the first order term for the dummy variables, (e.g. x_2 and x_3^2), cannot be included together in the regression model due to multicollinearity between these terms. For the toenails, no information on a feature is available, therefore, $C(\mathbf{x}) = \delta_0$, with only 1 parameter (intercept) to be estimated.

We implemented the analytical capacity models of the clips and toe nails for the failure models. Figure 3 shows the analytical and experimental capacity data, for each of the 1, 2, and 4 clips and SPF, SYP and DF combinations. For the 1 and 2 clip cases, 5 capacity data are available for all 3 wood types, for the 4 clip case, 5 capacity data are available for SPF and SYP, thus, the regressor matrices $\mathbf{X}_1$ and $\mathbf{X}_2$ are 40×5 matrices and data vectors $\mathbf{C}_e$ and $\mathbf{C}_a$ are both length 40 (i.e., $n_1 = n_2 = 40$). For the toe nail case, 13 capacity data is available for an 8 mm diameter toe nail and SYP wood, therefore, $\mathbf{X}_1$ and $\mathbf{X}_2$ are 13×1 vectors of 1's and $\mathbf{C}_e$ and $\mathbf{C}_a$ are 13×1 vectors (i.e., $n_1 = n_2 = 13$).

To determine the wind load D_v [kN] acting on the connection, the wind load acting on the roof is evaluated using the ASCE-7 [2] methodology. The net force is the force acting on a single interior truss. Thus, truss of the roof is assumed to be simply supported, with two supports on both ends of the truss (see Figure 2). Hence to find the wind load acting on each support, this force is divided equally into the support locations. Figure 4(a) shows the wind load as a function of wind speed.

The probability distribution $p(v)$ of wind speed in Miami-Dade county that incorporates the uncertainty in the wind loading. In this study, we assume a Weibull

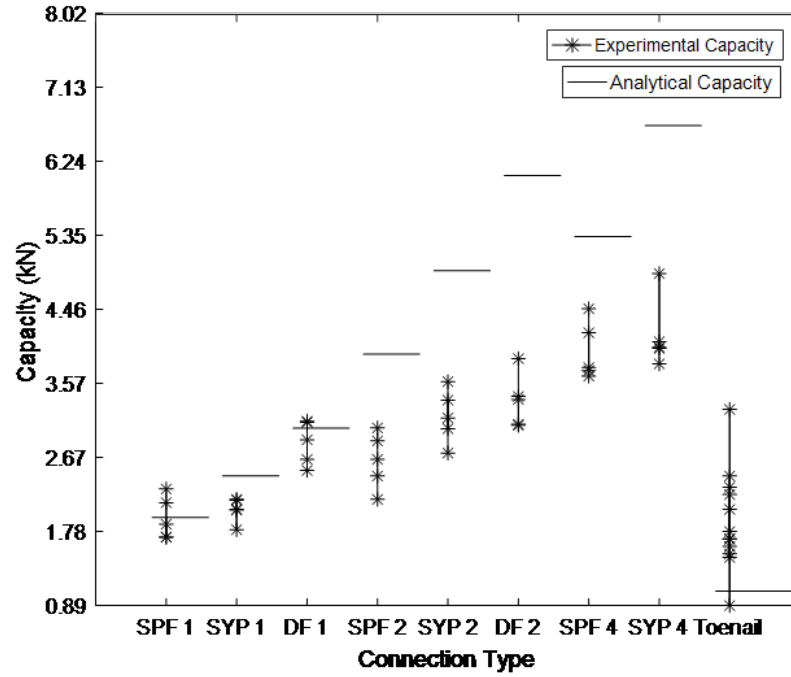


Fig. 3 Experimental and analytical capacities for different wood and hurricane clip combinations.

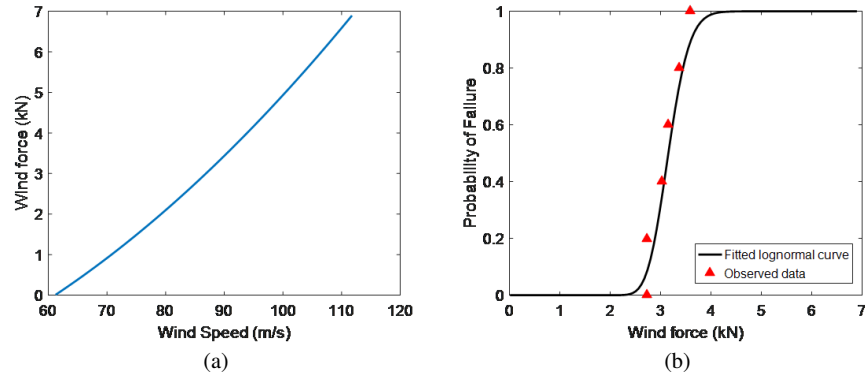


Fig. 4 (a) Calculated wind loads on the roof for a range of wind speeds (b) Lognormal fragility curve fitted to capacity data for two hurricane clips made out of SYP wood.

distribution with scale parameter 68.33 and shape parameter 1.738 as discussed in [14] as the hurricane wind speed distribution. Figure 1, shown previously, compares

the Weibull wind speed PDF and the Fragility function with SYP wood and 2 hurricane clips.

The Bayesian predictive distribution of the capacity for the SYP wood 2 hurricane clips connection and the toe nail connection are show in Figure 5. Figure 4(b) shows the fragility curve based on the lognormal approach of Equation (2) fitted to data for the connection with SYP wood and 2 clips. The parameters of the lognormal curve estimated for SYP wood and 2 clips are median $m = 3.1532$ kN and logarithmic standard deviation $\zeta = 0.1046$ kN. Figure 5 compares the Bayesian predictive distribution of capacity (solid curve), lognormal capacity distribution (dashed curve), deterministic capacity (vertical line), and the measured capacity data (triangles). The fragility is calculated from these predictive distributions. Lognormal method uses the median capacity and logarithmic standard deviation of the data, therefore distribution mostly covers the observed data. By contrast, Bayesian updated model is a compromise between the experimental data (green markers) and the analytical model (red line), which is evidenced from the fact that the peak of the probability distribution of the updated model (blue curve) is moved from the lognormal distribution (black curve) towards the analytical capacity (red line).

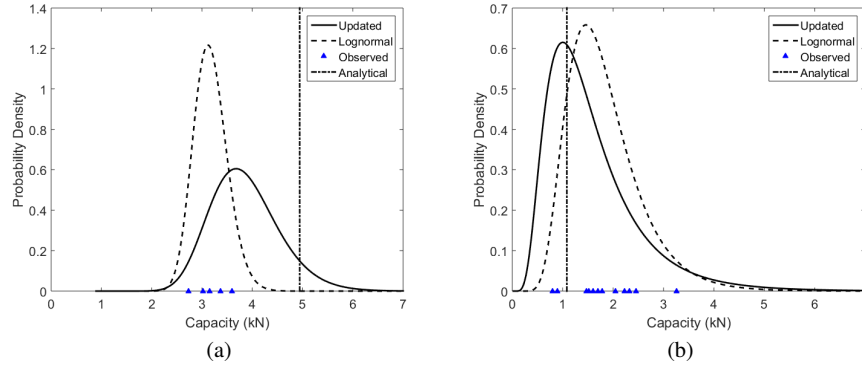


Fig. 5 Predictive distribution function of capacity of (a) SYP wood and 2 clips (b) toenails.

The results of Algorithm 1 to generate the posterior of probability of failure is illustrated in Figure 6 with data of SYP-2 clips connection and $D_v = 3$ kN. The predictive density curves of capacity with different realizations of β and σ^2 from their posteriors are shown in Figure 6(a) and the vertical line indicates 3 kN load. The fragility $F_v = P(C_u < D_v)$ at this wind load is the area to the left of the line under the density curves. Figure 6(b) is the histogram of the fragility values found from 10,000 simulated capacity distributions. The 2.5th and 97.5th percentiles of the fragility values, shown as vertical lines, specify a 95% confidence interval on the true fragility. The procedure in Algorithm 1 is repeated for a range of wind loads to find the fragility and confidence the bounds. Figure 7 shows the fragility curves along with the confidence intervals obtained with the updated model for 2

hurricane clips and toe nails cases. For clips, the analytical capacity is higher than the measured values thus the fragility of the updated model is lower than that from the lognormal approach (which considers only the measured data). For toenails, the analytical model underestimates the capacity therefore the fragility of the updated model is higher than that from the lognormal approach. In addition, with much lower capacities than hurricane clips, the fragility of toenails are much higher at lower forces than those of hurricane clips.

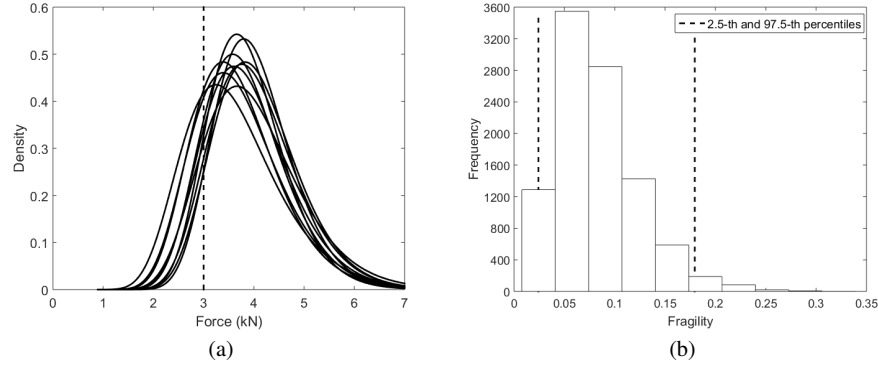


Fig. 6 (a) Predictive distribution function of capacity of SYP wood and 2 clips with random parameters. The vertical line indicates 3kN force (b) Fragility values found with distribution functions at 3 kN wind force and Monte Carlo estimate of 2.5th and 97.5th percentiles.

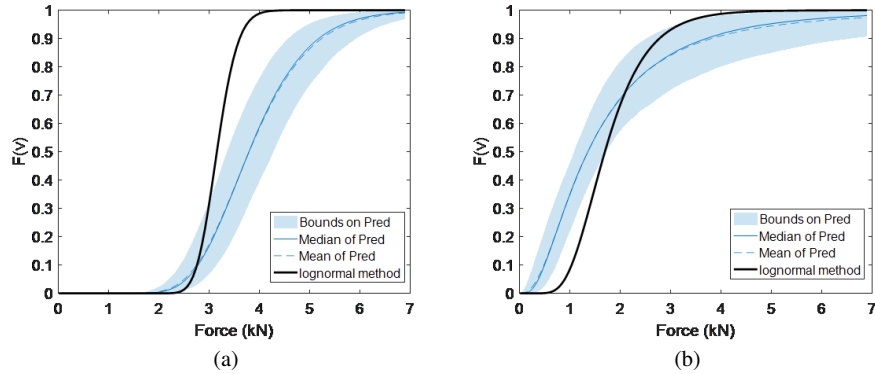


Fig. 7 (a) Fragility curves and their uncertainty bounds from the Bayesian approach. Fragility curve with lognormal approach is shown with black line. (a) For SYP wood and 2 clips. (b) For toenails.

We obtained the estimates of the probability of failure using Algorithm 1. 10,000 Monte Carlo simulations are performed and Equation (14) was evaluated using

the fragility curve of the updated model. Samples are drawn from the posterior of the probability of failure P_f of connections with different wood types. Figure 8 shows the median and upper 95-th percentile of probability of failure from the Bayesian approach and the mean failure probability from the lognormal approach, in increasing order of the median P_f . It can be seen that with the Bayesian approach (median) is more conservative than lognormal approach in reliability prediction for all cases, however, the upper prediction bound encompass the lognormal approach results. Overall, reliability decreases with fewer clips and the toe-nail has the lowest reliability. The wood types are ordered as DF, SYP and SPF in decreasing order of reliability.

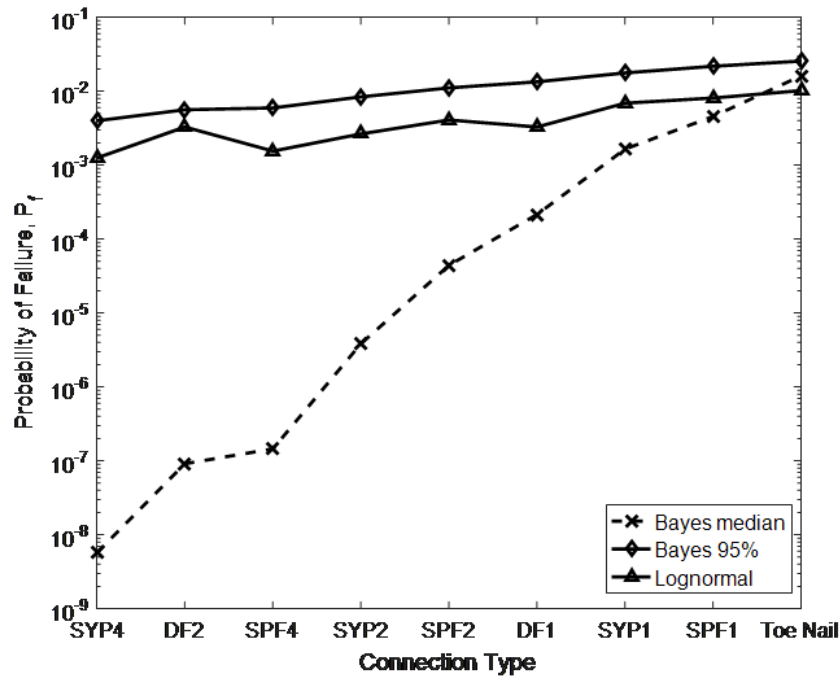


Fig. 8 Probability of failure (median and 95-th percentile) of a single roof-to-wall connection for different connection types using proposed Bayesian and existing lognormal approaches.

4.2 Bayesian Wind Loss Estimation for an Inventory of Buildings

To accurately account for the building stock subject to the wind loading, the building inventory is developed using digitized building footprints in Open Street Map [30].

Figure 9 shows the building stock in a Miami-Dade County, FL region extracted from the OSB Buildings repository, where an inventory of $n = 100$ single-family wood frame, gable roof type residential buildings are shown. It is assumed that all residential buildings in this region can be represented using the roof model shown in Figure 2 and discussed in Section 4 for all houses.

In this section we will model the aggregate loss of this building inventory using the Bayesian posterior distribution of failure probabilities computed using Algorithm 1. We will compute the exceedance probability of aggregate loss in the building inventory and the associated uncertainty bounds from the failure probability distribution. The exceedance probability will in turn be used to compare mitigation options, that is, for a given loss value if the exceedance probability of a connection type much smaller than an alternative then it is preferred. As the mitigation options it toe-nail connection type will be compared to the SYP wood and 2-clips connection. The aggregate loss will be computed by assuming that all houses in the region use a specific connection type in each roof-to-wall connection.

Exceedance probability (EP) curve is a graphical representation of the probability that a certain level of loss will be exceeded over a future time period [8]. For illustration purposes, for the sake of illustration, we assume each house in the inventory is worth \$100,000 and the damage ratio is 20% due to failure of a single roof to wall connection, that is, the loss per house due to a connection failure is $L_0 = 0.20 \times 100,000 = \$20,000$.

The probability of failure P_f is computed using Algorithm 1 for the roof-to-wall connections and the wind speed distribution of this region. Note that it is assumed the probability of the hurricane affecting all buildings throughout the region is equal and the connections are identical, therefore P_f is the same for all houses in the inventory. Letting w denote the number of failed clips and assuming it follows a binomial distribution $p(w)$ with failure probability P_f , we have

$$p(w) = \binom{K}{w} P_f^w (1 - P_f)^{K-w}, \text{ for } w = 0, 1, \dots, K \quad (16)$$

where $K = N \times n$ in which N is the number of connections per house and n is the number of houses in the inventory. The aggregate loss in the inventory is $L = L_0 w$ therefore the expected loss is $E[L] = L_0 E[w] = L_0 K P_f$. The probability that the aggregate loss exceeds a specific value l , or the *exceedance probability*, is

$$EP_l = P(L > l) = P(w > l/L_0) = \sum_{w=\lceil l/L_0 \rceil}^M p(w) \quad (17)$$

where $\lfloor x \rfloor$ is the floor function, giving as output the greatest integer less than or equal to x . For example, the exceedance probability for 1 million dollars in this inventory can be found from the probability mass function as:

$$EP_{\$1 \times 10^6} = P(w > \$1,000,000/\$20,000) = P(w > 50) = \sum_{w=50}^K p(w)$$

Note that more than a single connection may fail during a hurricane per single building. For any loss value l , the posterior distribution of the exceedance probability EP_l , is generated using Equations (16) and (17) and the posterior distribution of P_f . From the posterior distribution of EP_l , the median, and 95-th percentile are used as an estimate of EP and upper confidence bound, respectively for each l . Figure 10 shows the median and the 95-th percentile of the exceedance probability for loss values in the range of 0 to \$200,000 from the Bayesian and lognormal approaches. To create these figures, EP estimates and upper bounds for the building inventory are computed using the Bayesian approach by assuming that $N = 2$ hurricane clips and SYP wood or toenails are used for each connection of each house. Lognormal approach uses a constant total probability of failure to find EP. By contrast, the Bayesian approach considers the distribution of P_f therefore is able to provide an estimate of the median loss and an upper confidence bound on loss.

With the Bayesian approach the median aggregate loss in the building inventory that will be exceeded in a 100-year return hurricane (i.e., an annual probability of exceedance 10^{-2}) with 2-clips configuration is almost zero with a 95% upper confidence bound of \$61K. For toenails, the Bayesian approach estimates the median aggregate loss as \$93K with a 95% upper confidence bound of \$128K. The lognormal approach gives the aggregate loss with 2-clips as \$29K and with toenails as \$69K. In estimating the exceedance probability for a given loss, the Bayesian approach is less conservative (provides smaller estimates of probability) than the lognormal approach for the clips, while it is more conservative (provides larger estimates of probability) for the toenails. This is partly because the lognormal approach looks at only the measured capacity while the Bayesian approach considers both the measured and the analytical capacity.

5 Conclusions

This chapter presented Bayesian model updating approaches for quantification of uncertainty in predicting failure in rare but high consequence events. Specifically, a two-stage Bayesian regression model is discussed to fuse an analytical capacity model with experimentally observed capacity data to quantify uncertainties in performance of residential building roof systems under high velocity wind conditions. It is shown how the model updating approach can be used for decision making about various mitigation options of buildings, such as strengthening roof to wall connections in a regional aggregate loss analysis. In order to conduct regional aggregate loss analysis with a realistic information about house inventories we discussed how OSM Buildings repository can be integrated into the analysis.

A case study for quantifying Miami-Dade County hurricane wind risk is presented to show how analytical capacity models of connections can be updated with

experimental data available from the literature and how the reliability of roof systems under various mitigated and unmitigated configurations can be assessed using the posterior distribution of probability of failure. It is shown that proposed Bayesian capacity model has lower bias compared to the analytical prediction model, which in turn results in the better estimation of the exceedance probabilities. An aggregate loss analysis with associated confidence bounds for a given wind model and building portfolio is presented to translate the posterior distribution of the capacity model into monetary terms, which has significant implications for policy and insurance underwriting applications.

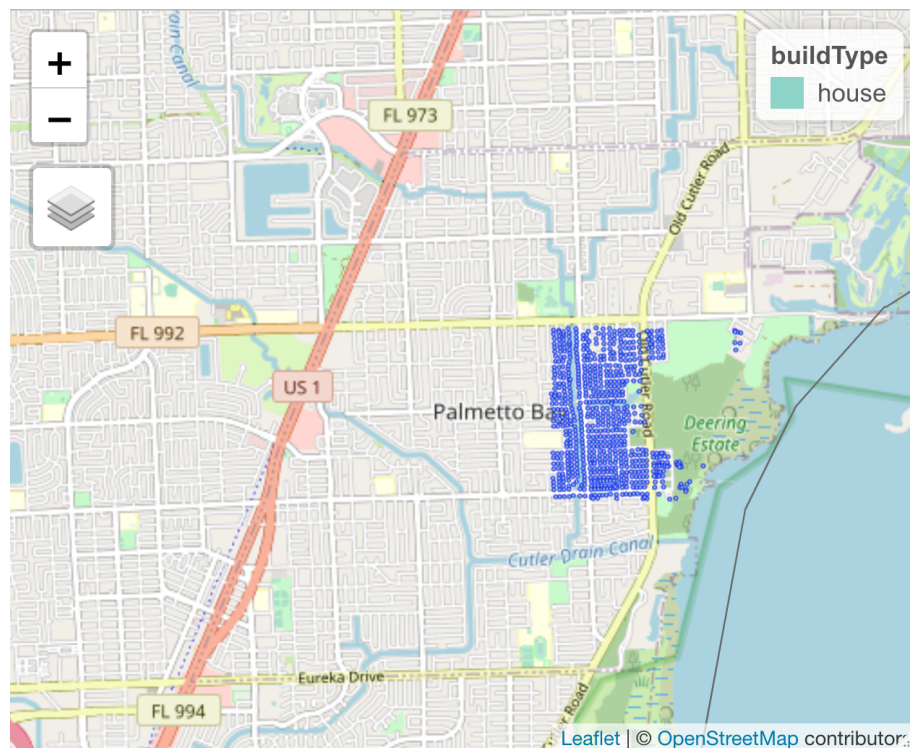
The methodology incorporated uncertainty in the capacity models only and the wind speed Weibull probability models are known without any error. As further future work, the uncertainty in the wind speed models can also be considered in the failure probability and aggregate loss calculations within the proposed Bayesian framework.

Acknowledgements This publication was supported by National Science Foundation (NSF) grant CMMI-2101091 and the National Sea Grant College Program of the U.S. Department of Commerce's National Oceanic and Atmospheric Administration (NOAA), Grant No. NA 14OAR4170108, Subcontract No UFDSP00010509 made to Florida State University. The views expressed are those of the authors and do not necessarily reflect the view of these organizations.

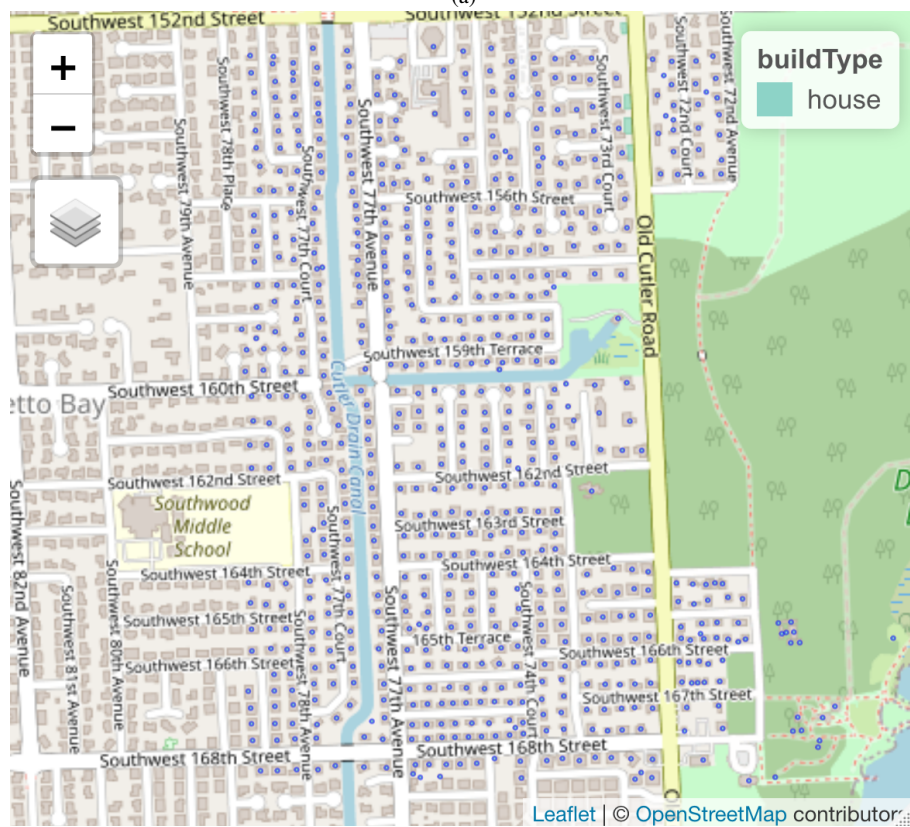
References

1. Ahmed, S. S., Canino, I., Chowdhury, A. G., Mirmiran, A., & Suksawang, N. (2011). Study of capability of multiple mechanical fasteners in roof-to-wall connections of timber residential buildings. *Practice Periodical on Structure Design and Construction*, 16(1), 2-9.
2. American Society of Civil Engineers (ASCE). (1998). Minimum design loads for buildings and other structures, ANSI/ASCE 7-98-1998.
3. Beck, J. L., & Au, S.-K. (2002). Bayesian Updating of Structural Models and Reliability using Markov Chain Monte Carlo Simulation. *Journal of Engineering Mechanics*, 128(4), 380–391.
4. Beck, J. L., & Katafygiotis, L. S. (1998). Updating Models and Their Uncertainties. II: Model Identifiability. *Journal of Engineering Mechanics*, 124(4), 463–467.
5. Ellingwood, B. R., Rosowsky, D. V., Li, Y., & Kim, J. H. (2004). Fragility Assessment of Light-Frame Wood Construction Subjected to Wind and Earthquake Hazards. *Journal of Structural Engineering*, 130(12), 1921–1930.
6. Ellingwood, B. R., & Wen, Y. K. (2005). Risk-benefit-based design decisions for low probability/high consequence earthquake events in Mid-America, *Progress in Structural Engineering and Materials*, 7(2), 56-70
7. Gardoni, P., der Kiureghian, A., & Mosalam, K. M. (2002). Probabilistic Capacity Models and Fragility Estimates for Reinforced Concrete Columns based on Experimental Observations. *Journal of Engineering Mechanics*, 128(10), 1024–1038.
8. Grossi, P., & Kunreuther, H. (2005). Catastrophe modeling: a new approach to managing risk (Vol. 25). Springer Science & Business Media.
9. Gelman, A., Carlin, J. B., Stern, H. S., & Rubin, D. B. (2014). *Bayesian data analysis* (Vol. 2). London: Chapman & Hall/CRC.
10. Gurley, K., Pinelli, J.-P., Subramnian, C., Cope, A., Zhang, L., Murphree, J., et al. (2005). Florida public hurricane loss projection model engineering team. Final report, International Hurricane Research Center, Florida International University, Miami, USA.

11. Helton, J. C., Johnson, J. D., Sallaberry, C. J., & Storlie, C. B. (2006). Survey of sampling-based methods for uncertainty and sensitivity analysis. *Reliability Engineering & System Safety*, 91(10), 1175-1209.
12. Kennedy, M.C. and A. O'Hagan (2001), Bayesian calibration of computer models. *Journal of the Royal Statistical Society Series B-Statistical Methodology*, 63, 425-450.
13. Kleindorfer, P. R., & Kunreuther, H. (1999). The complementary roles of mitigation and insurance in managing catastrophic risks. *Risk Analysis*, 19(4), 727-738
14. Li, Y., & Ellingwood, B. R. (2006). Hurricane damage to residential construction in the US: Importance of uncertainty modeling in risk assessment. *Engineering structures*, 28(7), 1009-1018
15. Luxhøj, J. T., & Coit, D. W. (2006). Modeling low probability/high consequence events: an aviation safety risk model, In *Reliability and Maintainability Symposium*, 2006. RAMS'06. Annual (pp. 215-221). IEEE.
16. Madsen, H. O., Krenk, S., & Lind, N. C. (2006). *Methods of structural safety*. Courier Corporation.
17. Montgomery D., Peck E., Vining G. (2006). *Introduction to Linear Regression Analysis*. Fourth Edition. John Wiley & Sons.
18. Mottershead, J. E., & Friswell, M. I. (1993). Model Updating In Structural Dynamics: A Survey. *Journal of Sound and Vibration*, 167(2), 347-375.
19. Pita, G., Pinelli, J. P., Gurley, K., & Mitrani-Reiser, J. (2014). State of the Art of Hurricane Vulnerability Estimation Methods: A Review. *Natural Hazards Review*, 16(2), 04014022.
20. Qian, P. Z. G., and Wu, C. F. J. (2008). Bayesian hierarchical modeling for integrating low-accuracy and high-accuracy experiments, *Technometrics* 50, 192-204.
21. Reed, T. D., Rosowsky, D., V., & Schiff, S., D. (1997). Uplift capacity of light-frame rafter to top plate connections. *Journal of Architectural Engineering*, 3(4), 156-163.
22. Reese, C. S., Wilson, A.G., Hamada, M., Martz, H.F., & Ryan, K.J. (2004). Integrated analysis of computer and physical experiments. *Technometrics*, 46, 153-164.
23. Rosowsky, D. V., & Ellingwood, B. R. (2002). Performance-based engineering of wood frame housing: Fragility analysis methodology. *Journal of Structural Engineering*, 128(1), 32-38.
24. Rubinstein, R. Y., & Kroese, D. P. (2011). *Simulation and the Monte Carlo method* (Vol. 707). John Wiley & Sons.
25. Shanmugam, B., Nielson, B. G., & Prevatt, D. O. (2009). Statistical and analytical models for roof components in existing light-framed wood structures. *Engineering Structures*, 31(11), 2607-2616.
26. Straub, D., & der Kiureghian, A. (2008). Improved seismic fragility modeling from empirical data. *Structural Safety*, 30(4), 320-336.
27. Vanli, O. A., & Jung, S. (2014). Statistical updating of finite element model with Lamb wave sensing data for damage detection problems. *Mechanical Systems and Signal Processing*, 42(1-2), 137-151.
28. Vickery, P. J., Skerlj, P. F., Lin, J., Twisdale Jr, L. A., Young, M. A., & Lavelle, F. M. (2006). HAZUS-MH hurricane model methodology. II: Damage and loss estimation. *Natural Hazards Review*, 7(2), 94-103.
29. Zhang, S., Nishijima, K., & Maruyama, T. (2014). Reliability-based modeling of typhoon induced wind vulnerability for residential buildings in Japan. *Journal of Wind Engineering and Industrial Aerodynamics*, 124, 68-81.
30. OSM Buildings (2023). OpenStreetMap Buildings. Retrieved from <https://osmbuildings.org/data/> [Accessed 04-01-2023]
31. Mishra, S., Vanli, O. A., Alduse, B. P., & Jung, S. (2017). Hurricane loss estimation in wood-frame buildings using Bayesian model updating: Assessing uncertainty in fragility and reliability analyses. *Engineering Structures*, 135, 81-94.
32. Mishra, S., Vanli, O. A., Kakareko, G., & Jung, S. (2019). Preventive maintenance of wood-framed buildings for hurricane preparedness. *Structural safety*, 76, 28-39.

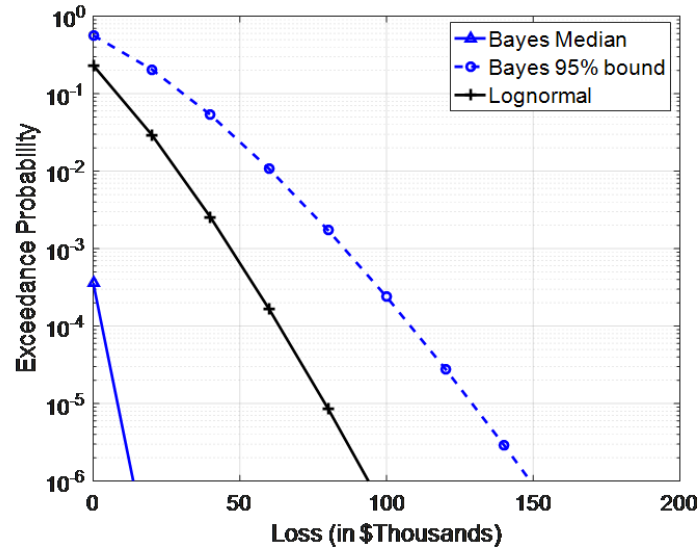


(a)

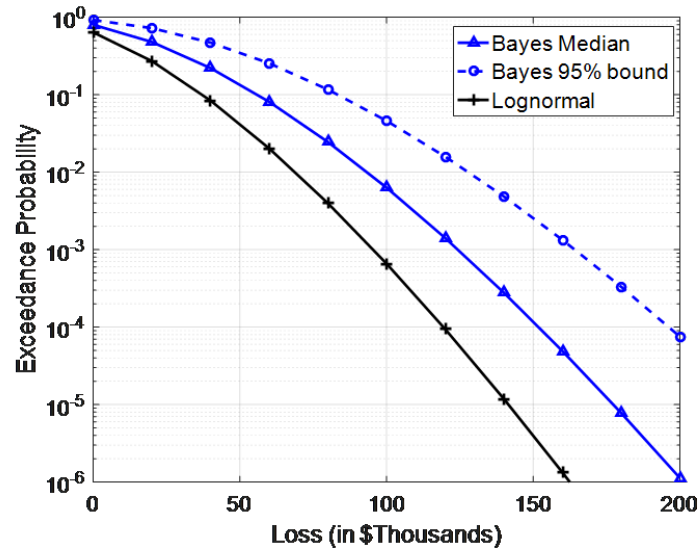


(b)

Fig. 9 (a) Single family, gable roof building inventory of a region in Miami-Dade County (b) The study region containing 100 buildings.



(a)



(b)

Fig. 10 (a) Exceedance probability curves for total loss using the Bayesian (median and 95th percentile) and lognormal approaches. (a) SYP wood and 2 clips. (b) Toe nails.