Surrogate modeling based methodology for developing limit state free tsunami building fragility models and performing sensitivity analysis

Fragility Modeling

Scope

The present document provides details on the surrogate model-based fragility assessment methodology.

Methodology

Logistic regression is used in this work as a surrogate model to develop parametrized limit state free fragility functions due to its good prediction accuracy [38,40] and an easy to communicate closed form expression.

The proposed fragility modeling approach is schematically shown in Figure 1 highlighting the main steps involved in the procedure. The first step involves selecting the key structural parameters, such as material strength, factor for opening ratio, and building's drag coefficient, and tsunami intensity measures such as tsunami inundation depth and speed. In addition to structural and hazard characteristics, the proposed approach also parameterizes the fragility model on an engineering demand parameter (EDP) which can be used to define the capacity limit states at various damage level thresholds. Inclusion of the EDP as an explanatory variable makes the fragility model *limit state free* and enables the fragility function to predict the probability of exceeding any given value of EDP demand (edp_d). Using the limit state free fragility function, the probability of exceeding an EDP value (edp_d) can be obtained as:

$$P(EDP \geq edp_d | x_b, x_{IM}) = \frac{1}{1+\exp\left(-g(x_b, x_{IM}, edp_d)\right)} \quad (1)$$

In the above equation, $g(x_b, x_{IM}, edp_d)$ is a polynomial representing the log of odds in favor of $EDP \geq edp_d$ for given values of structural characteristics (x_b) and tsunami intensity measures x_{IM}. Collectively, the parameters X_b, X_{IM}, and EDP_d can be represented as a vector $X_p = \{X_b, X_{IM}, EDP_d\}$.

In order to derive the logistic regression model described in Eq. 1, step 2 performs an experimental design using Latin hypercube sampling (LHS) [41] to span the space of parameters in X_{IM} and X_b uniformly within appropriate lower and upper bounds and generate a large set of parameters consisting of n combinations. Using the same process, a smaller set of parameters, with about $n/5$ samples, is generated to test the predictive performance of the logistic regression model. In step 3, for each of the n parameter combinations, a finite element model of the building is generated and its tsunami performance is assessed using the EDP. During this analysis, tsunami loads applied on the building are incrementally increased until failure. The results of the finite element analysis, particularly the EDP and the corresponding tsunami IM values, are recorded at each step of the analysis. Next, in the fourth step, n_{EDP} equally spaced values of the EDP are sampled between the lower and the upper bounds of the EDP, which act as capacities (EDP_c). This range for EDP can be selected based upon the seismic capacity limit states of buildings at different damage states (e.g. slight, moderate, extensive, or complete) or left as a variable that spans a sufficient range based on expert judgement or information collected based of experiments or in the field following disasters. In each of the n simulations, the tsunami IM values (IM_c) corresponding to each of the n_{EDP} EDP values are obtained using the results recorded during the finite element analyses. These IM_c values are compared against the tsunami IM sampled in step 2 during LHS – im_d. For the cases where the $im_d \geq \boldsymbol{IM_c}$, an indicator binary variable marked as one indicating that the demand exceeded the capacity, otherwise it is marked as zero indicating non-exceedance. Thus, in step 5, a binary vector ($\boldsymbol{y}$) with $n \times n_{EDP}$ entries is generated. Correspondingly, the set of n input parameters is also replicated n_{EDP} times. Additionally, for each of the $n \times n_{EDP}$ cases the drift capacity limits are also recorded in a separate vector. The parameter combinations, intensity measures, and the drift capacity limit values are stored together in a matrix ($\boldsymbol{X}$). The same procedure is repeated for the data set consisting of the test samples.

The binary vector ($\boldsymbol{y}$), described above, is used to develop the logistic regression model presented in Eq. 1 using a Bayesian approach. For this purpose, all the input variables stored in the matrix $\boldsymbol{X}$ are scaled such that they have zero mean and a standard deviation of 0.5 [42,43]. Thereby, a normalized matrix of inputs ($\boldsymbol{X_n}$) is obtained. Next, in step 6, the order of the polynomial $g(\boldsymbol{X_n})$ is chosen and priors are defined in step 7 for each of the coefficients in the polynomial, the coefficients are represented in a vector form as b. Herein, t location-scale distribution is chosen for each of the coefficients in b with location parameter set as zero, scale set at 25, and the shape parameter is set as 7.0. These particular parameters are chosen for the prior distributions to ensure that the coefficients, b, do not become unstable for perfectly separable data. The priors for all the coefficients are assumed to be statistically independent. This particular distribution is chosen to define priors since it has been shown to be effective for problems where perfect separation of data points is possible [42,43], which was observed in the preliminary analyses.

Using Bayes theorem, in step 8, the joint posterior probability density function of the coefficients $\left(p(b|\boldsymbol{D})\right)$ for observed data $\boldsymbol{D}$**, i.e.** $\boldsymbol{X_n}$ and $\boldsymbol{y}$, can be written as:

$$p(b|\boldsymbol{D}) = \frac{p(\boldsymbol{D}|b)p(b)}{p(\boldsymbol{D})} \propto p(\boldsymbol{D}|b)p(b) \tag{2}$$

In the above equation, $p(\boldsymbol{D}|b)$ is the likelihood of the data conditional on the polynomial coefficients b, $p(b)$ is the prior probability density function of the polynomial coefficients**,** and $p(\boldsymbol{D})$ is the probability of the observed data $\boldsymbol{D}$. Since $p(\boldsymbol{D})$, does not vary for a given data set, the posterior distribution of b is proportional to the product of the prior and the likelihood, as shown in Eq. 2. Since the prior distributions of the coefficients b are assumed to be statistically independent, $p(b)$ can be written as the product of prior distributions of the individual coefficients ($f(b_c)$):

$$p(b) = \prod_c^{n_c} f(b_c)\,; \forall\, b_c \in b \tag{3a}$$

$$f(b_c) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sigma\sqrt{\nu\pi}\,\Gamma\left(\frac{\nu}{2}\right)}\left[1 + \frac{1}{\nu}\left(\frac{b_c-\mu}{\sigma}\right)^2\right]^{-\left(\frac{\nu+1}{2}\right)}\,; \mu = 0, \sigma = 25; \nu = 7 \tag{3b}$$

where, n_c is the number of coefficients in the polynomial $g(\cdot)$, i.e. the cardinality of b. Assuming that the outcomes $\boldsymbol{y}$ are binomially distributed, the likelihood of the data for given coefficients b can be written as:

$$p(\boldsymbol{D}|b) = \prod_{i}^{n_{tot}} \theta_i^{y_i}(1-\theta_i)^{1-y_i} \tag{4a}$$

$$\theta_i = \frac{1}{1+e^{-X_i\boldsymbol{b}}} \tag{4b}$$

where n_{tot} is the total number of parameter combinations in $\boldsymbol{X_n}$, which is equal to $n \times n_{EDP}$; and X_i is the ith row in $\boldsymbol{X_n}$.

The posterior distribution of b obtained using Eqs. 2 to 4 can be maximized to obtain a maximum a posteriori probability (MAP) estimate, which provides a point estimate of the coefficients. Alternatively, in step 9, Markov Chain Monte Carlo (MCMC) simulations can be performed to sample the posterior probability density function and obtain large number of instances of the coefficients b. Overall, this approach requires a significantly smaller number of structural response simulations in comparison to the typically used Monte Carlo Simulations based fragility assessment approach, making the developed approach computationally more efficient. Furthermore, the approach described in this section can also be used to obtain traditional limit state specific (e.g. slight, moderate, extensive and complete) fragility functions by using a single value of the capacity threshold in step 4 instead of n_{EDP} equally spaced values.

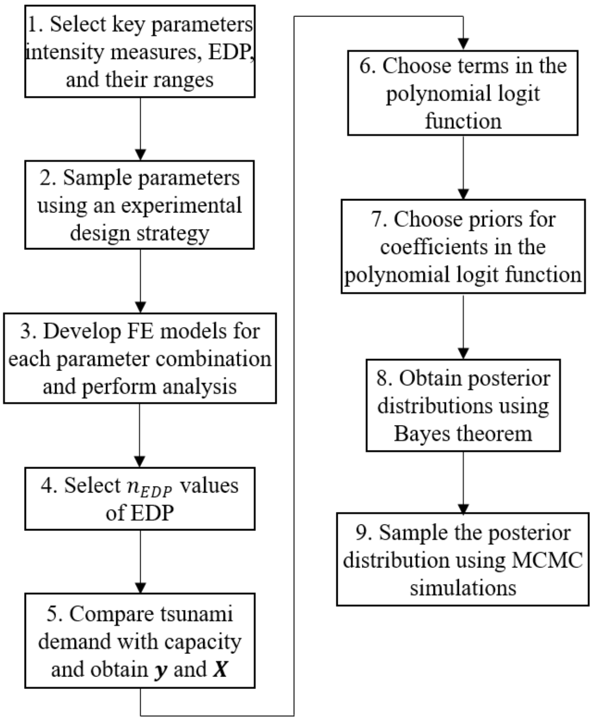
1. Select key parameters intensity measures, EDP, and their ranges
2. Sample parameters using an experimental design strategy
3. Develop FE models for each parameter combination and perform analysis
4. Select n_{EDP} values of EDP
5. Compare tsunami demand with capacity and obtain $\boldsymbol{y}$ and $\boldsymbol{X}$
6. Choose terms in the polynomial logit function
7. Choose priors for coefficients in the polynomial logit function
8. Obtain posterior distributions using Bayes theorem
9. Sample the posterior distribution using MCMC simulations

Figure 1. Schematic representation of the fragility modeling approach.