Use of generalized lambda models for seismic fragility analysis

Xujia Zhu, Marco Broccardo, Bruno Sudret

▶ **To cite this version:**

Xujia Zhu, Marco Broccardo, Bruno Sudret. Use of generalized lambda models for seismic fragility analysis. 8th International Symposium on Reliability Engineering and Risk Management (ISRERM), Sep 2022, Hannover, Germany. hal-03788059

HAL Id: hal-03788059
https://hal.science/hal-03788059v1

Submitted on 26 Sep 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

ETH zürich

Use of generalized lambda models for seismic fragility analysis

X. Zhu, M. Broccardo and B. Sudret

Chair of Risk, Safety and Uncertainty Quantification
Stefano-Franscini-Platz 5
CH-8093 Zürich

Risk, Safety & Uncertainty Quantification

Data Sheet

Journal:	Proc. 8th International Symposium on Reliability Engineering and Risk Management (ISRERM), Hannover (Germany), September 4-7
Report Ref.:	RSUQ-2022-007
Arxiv Ref.:	
DOI:	-
Date submitted:	December 16, 2021
Date accepted:	September 5, 2022

Use of generalized lambda models for seismic fragility analysis

Xujia Zhu*[1], Marco Broccardo†[2], and Bruno Sudret‡[1]

[1] *Chair of Risk, Safety and Uncertainty Quantification, ETH Zürich, Switzerland*
[2] *Department of Civil, Environmental and Mechanical Engineering, University of Trento, Italy*

September 23, 2022

Abstract

Seismic structural performances are typically characterized by *fragility models*. A fragility model provides the probability of exceeding a certain level of damage in the structure given a set of intensity measures (IMs) of the seismic excitations. The damage level is defined as a function of a given Engineering Demand Parameter (EDP) (e.g., the maximal interstory drift for a multi-story building, the maximum lateral drift of piers for a bridge). In practice, performing fragility analysis is usually difficult due to limited seismic records and high cost of experiments or simulations. In this study, we model the seismic input by a stochastic artificial ground motion model. This stochastic ground motion model is a filtered white-noise parameterized by a set of engineering-meaningful parameters (i.e., a given set of parameters can generate an infinite number of seismic signals). As a result, the corresponding EDP is a random variable conditioned on the parameters of the ground motion model, and the input-output relationship can be viewed as a stochastic simulator. Given this representation, the fragility model can be defined as a function of the parameters of the ground motion model or any other IMs of interest (estimated from the input ground motion samples). To alleviate the computational burden of the fragility analysis, we propose using the generalized lambda surrogate model. The latter uses the flexible generalized lambda distribution to represent the distribution of the EDP for a given set of IMs. We illustrate the performance of the proposed method on a three-story shear frame. The results show that it outperforms a parametric linear model and a non-parametric kernel model.

1 Introduction

In Performance-Based Earthquake Engineering (PBEE) [1], the seismic risk is computed by convolving the output of Probabilistic Seismic Hazard Analysis (PSHA) with fragility, damage,

*zhu@ibk.baug.ethz.ch
†marco.broccardo@unitn.it
‡sudret@ethz.ch

and loss models. The results of the PSHA analysis are so-called hazard curves, which represent the rate of occurrence of a given intensity measure (IM). In its general form, the fragility model is a conditional probability distribution that relates a given IM (or vector of IMs) to structural performance. This is usually defined based on a quantity of interest (e.g., the maximum interstory drift for a multi-story building) called the *engineering demand parameter* (EDP). The development of a robust fragility model is key to accurate seismic risk assessments; therefore, it is not surprising that it has generated a significant amount of research by the structural engineering community. Among the broad landscape of fragility models, simulation-based fragility models have recently received more attention. In this setting, the structure is represented by a computational model, and the input is given by ground motion time series, which can be real or synthetic. In the first case, the ground motions are usually selected (and modified) based on the site and a return period of a given IM. In the second case, artificial ground motions are typically generated from a stochastic ground motion model fitted to a dataset of real ground motions. The latter case is of great interest because it allows a formal uncertainty quantification (UQ) framework to be defined to assess structural performance and, thus, to compute fragility models. Using this approach has significant advantages since it can unlock the use of the latest developments in forward UQ analysis to overcome the obstacle of traditional methods. These include (but are not limited to) the scarcity of actual ground motions for specific sites, the scaling of ground motions, and the high computational costs of time-history analysis.

Following this recent line of research, Abbiati et al. [2] developed a *replication*-based approach. In this framework, the ground motion parameters that generate the synthetic earthquake load are used as IMs. Multiple (e.g., 100) artificial ground motions are generated for the same ground motion parameters. The associated structural responses are realizations of the EDP conditioned on the given ground motion parameters. As a result, quantiles of the conditional EDP can be calculated from the realizations which are called *replications* in the sequel. This procedure is repeated for different ground motion parameters. The estimated quantiles (based on replications) are then represented by Kriging as deterministic functions of the ground motion parameters. These functions of quantiles characterize the distribution of the conditional EDP.

In this paper, we propose using the generalized lambda model developed by Zhu and Sudret [3] to emulate the distribution of EDP conditioned on IMs. This approach features no need for replications, i.e., the simulation is only performed once for each set of ground motion parameters. Moreover, we further extend the method to an adaptive approach which automatically selects the degree for the shape parameters of the model. The resulting surrogate model directly produces the conditional probability function which can be used for fragility analysis in a straightforward way.

The remainder of the paper is structured as follows. In Section 2, we recap the principle of fragility analysis. In Section 3, we summarize the main ingredients of the generalized lambda model and present a novel method to build the surrogate model in an adaptive manner. We describe a computational example in Section 4. The detailed numerical results are given in

Section 5 to illustrate the performance of the proposed method. Finally, we conclude the main finding of the study and give outlook for future research in Section 6.

2 Fragility analysis

In PBEE, seismic loads are typically characterized by the IMs. An incomplete list of conventional IMs include peak ground acceleration, spectral acceleration, peak ground velocity, and Arias Intensity [4]. However, in general, an IM can represent any "optimal" feature of the seismic load. According to [4], optimal is defined as being practical, sufficient, effective, and efficient (see [4] for further details). Following this line, in [2], a set of IMs have been identified with a particular set of physically meaningful parameters of the stochastic ground motion model. In risk assessment, IMs are modeled by random variables to account for uncertain earthquake loads in the considered area. As a result, we represent the IMs by a random vector denoted by $\boldsymbol{X}$. The lower case $\boldsymbol{x}$ corresponds to a specific realization of IMs.

Because IMs are summary quantities, they cannot uniquely determine the detailed time series. In other words, for a given $\boldsymbol{x}$, one can find different earthquake signals that share the same values of IMs. Consequently, conditioned on $\boldsymbol{x}$, the EDP denoted by Y is still a random variable rather than a deterministic value.

Fragility analysis aims at computing the failure probability conditioned on the IMs. More precisely, the failure probability is defined as the probability that EDP exceeds a given threshold

$$p_f(\boldsymbol{x}) = \mathbb{P}\left(Y > \delta_0 \mid \boldsymbol{X} = \boldsymbol{x}\right), \tag{1}$$

where p_f is the fragility function, and δ_0 corresponds to the threshold that is selected based on the safety demand.

When aggregating the risk to evaluate the decision variables (e.g., monetary losses, casualties, downtime), the complete conditional probability is required [5]. As the fragility function in Eq. (1) can be directly evaluated by post-processing the conditional distribution, we focus on estimating the later for general purpose in this paper.

3 Generalized lambda models

To estimate the distribution of the EDP, Y, conditioned on the IMs, $\boldsymbol{X} = \boldsymbol{x}$, we propose using the generalized lambda model [3]. The generalized lambda model relies on the generalized lambda distribution to represent the conditional distribution. The distribution parameters as functions of $\boldsymbol{x}$ are represented by polynomial chaos expansions. In this section, we briefly review the principle of this model.

3.1 Introduction

A generalized lambda model assumes that the conditional distribution can be approximated by the generalized lambda distribution. This distribution family is a very flexible and can accurately represent most of the common parametric distributions, among others, uniform, normal, and extreme value distributions [6].

A generalized lambda distribution is given by a parameterization of its quantile function [7], that is,

$$Q(u;\boldsymbol{\lambda}) = \lambda_1 + \frac{1}{\lambda_2}\left(\frac{u^{\lambda_3}-1}{\lambda_3} - \frac{(1-u)^{\lambda_4}-1}{\lambda_4}\right), \tag{2}$$

where $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \lambda_3, \lambda_4)$ is the vector of the four distribution parameters: λ_1 is the location parameter, λ_2 is the scale parameter, λ_3 and λ_4 control the shape of the distribution. In order that $Q(u;\boldsymbol{\lambda})$ defined in Eq. (2) is a valid quantile function (increasing on $[0,1]$), it is required that $\lambda_2 > 0$. From the quantile function, we can derive the probability density function (PDF)

$$\begin{aligned} f_Y^{\mathrm{GLD}}(y;\boldsymbol{\lambda}) &= \frac{\lambda_2}{u^{\lambda_3-1}+(1-u)^{\lambda_4-1}}\mathbb{1}_{[0,1]}(u), \\ &\text{with } u = Q^{-1}(y;\boldsymbol{\lambda}), \end{aligned} \tag{3}$$

where $\mathbb{1}$ is the indicator function. As indicated by Eq. (3), the PDF does not have an explicit form, and one needs to solve a nonlinear equation to evaluate it numerically.

Under the assumption that the conditional distribution can be approximated by the generalized lambda distribution, the distribution parameters λ_1–λ_4 are functions of $\boldsymbol{x}$, denoted by $\boldsymbol{\lambda}(\boldsymbol{x})$. In the generalized lambda model, these functions are represented by polynomial chaos expansions [8]

$$\begin{aligned} \lambda_l(\boldsymbol{x}) &\approx \lambda_l^{\mathrm{PC}}(\boldsymbol{x};\boldsymbol{c}) = \sum_{\boldsymbol{\alpha}\in\mathcal{A}_l} c_{l,\boldsymbol{\alpha}}\psi_{\boldsymbol{\alpha}}(\boldsymbol{x}), \quad l = 1,3,4, \\ \lambda_2(\boldsymbol{x}) &\approx \lambda_2^{\mathrm{PC}}(\boldsymbol{x};\boldsymbol{c}) = \exp\left(\sum_{\boldsymbol{\alpha}\in\mathcal{A}_2} c_{2,\boldsymbol{\alpha}}\psi_{\boldsymbol{\alpha}}(\boldsymbol{x})\right). \end{aligned} \tag{4}$$

The exponential transform is used to guarantee the positiveness of λ_2. Here, $\psi_{\boldsymbol{\alpha}}$ is a multivariate polynomial chaos basis function defined by the multi-index $\boldsymbol{\alpha}$ whose j-th component α_j denotes the polynomial degree of $\psi_{\boldsymbol{\alpha}}$ in x_j, $c_{l,\boldsymbol{\alpha}}$ is the associated coefficient for λ_l^{PC}, and $\mathcal{A}_l$ is the truncation set that defines all the polynomial basis considered to represent λ_l. The most commonly used truncation scheme is the "full basis" of degree p which contains all the basis functions of total degree less than p

$$\mathcal{A}^{M,p} \overset{\mathrm{def}}{=} \left\{\boldsymbol{\alpha}\in\mathbb{N}^M : \sum_{j=1}^{M}\alpha_j \le p\right\}. \tag{5}$$

For independent random variables $\boldsymbol{X}$, the basis function is defined as a product of univariate polynomials:

$$\psi_{\boldsymbol{\alpha}}(\boldsymbol{x}) = \prod_{j=1}^{M}\phi_{\alpha_j}^{(j)}(x_j), \tag{6}$$

where $\left\{\phi_k^{(j)} : k \in \mathbb{N}\right\}$ is the orthogonal polynomial basis with respect to the marginal distribution f_{X_j}, i.e.,

$$\mathbb{E}\left[\phi_k^{(j)}(X_j)\,\phi_l^{(j)}(X_j)\right] = \begin{cases} 1 & \text{if } l = k \\ 0 & \text{otherwise} \end{cases}. \tag{7}$$

For normal, uniform, and exponential distributions, the associated univariate orthogonal polynomials are well-known as Hermite, Legendre and Laguerre polynomials [8].

When $\boldsymbol{X}$ has dependent components (as it is usually the case in earthquake engineering [2]), one common way to build the basis functions is to transform $\boldsymbol{X}$ into an auxiliary vector $\boldsymbol{\Xi} = \mathcal{T}(\boldsymbol{X})$ with independent components (e.g., a standard normal vector) using the Nataf or Rosenblatt transform [9]. Then, the polynomial basis is defined with respect to the transformed variables

$$\psi_{\boldsymbol{\alpha}}(\boldsymbol{x}) = \prod_{j=1}^{M} \phi_{\alpha_j}^{(j)}(\xi_j). \tag{8}$$

where $\left\{\phi_k^{(j)} : k \in \mathbb{N}\right\}$ is given by the marginal distribution of Ξ_j.

3.2 Build a generalized lambda model from data

To build a generalized lambda model, one needs to select a set of basis functions, i.e., a truncation set, for each component of $\boldsymbol{\lambda}^{\mathrm{PC}}$ and determine the associated coefficients, as shown in Eq. (4). In this section, we recall the method proposed in Zhu and Sudret [3] which features no need for replications. Furthermore, we will enrich this method by adaptively selecting appropriate basis functions for λ_3 and λ_4.

In the first step, we generate/collect data to build the surrogate. $\mathcal{X} = \left\{\boldsymbol{x}^{(1)}, \ldots, \boldsymbol{x}^{(N)}\right\}$ contains N realizations of IMs. For each point $\boldsymbol{x}^{(i)}$ of $\mathcal{X}$, the associated EDP $y^{(i)}$ is recorded in $\mathcal{Y} = \left\{y^{(1)}, \ldots, y^{(N)}\right\}$.

If the truncation sets $\{\mathcal{A}_l : l = 1, \ldots, 4\}$ are given, the maximum likelihood estimation can be used to calibrate the coefficients of the basis functions based on $\mathcal{X}$ and $\mathcal{Y}$:

$$\hat{\boldsymbol{c}} = \arg\max_{\boldsymbol{c}} \mathsf{L}\left(\boldsymbol{c}\right), \tag{9}$$

where L is the log-likelihood given by

$$\mathsf{L}\left(\boldsymbol{c}\right) = \sum_{i=1}^{N} \log\left(f^{\mathrm{GLD}}\left(y^{(i)}; \boldsymbol{\lambda}^{\mathrm{PC}}\left(\boldsymbol{x}^{(i)}; \boldsymbol{c}\right)\right)\right). \tag{10}$$

As suggested in Eq. (9), this estimator does not require replications.

To determine the truncation sets $\mathcal{A}_1$ and $\mathcal{A}_2$ for λ_1^{PC} and λ_2^{PC}, we plug the sparse solver hybrid-least angle regressions [10] into the modified feasible generalized least-squares framework proposed in Zhu and Sudret [3]. The latter consists in alternatively fitting the mean and the variance

function of EDP. The resulting basis functions for the mean form the truncation set $\mathcal{A}_1$ for λ_1^{PC}, and those for the variance function form $\mathcal{A}_2$ for λ_2^{PC}.

Zhu and Sudret [3] proposed selecting polynomials with relatively low degree, namely 1, for λ_3^{PC} and λ_4^{PC} if the shape of the distribution is not expected to change in a strongly nonlinear way. In this study, we propose Algorithm 1 based on the Bayesian information criterion (BIC) [11] to adaptively choose an appropriate degree for λ_3^{PC} and λ_4^{PC}. In this algorithm, we first build an generalized lambda model with λ_3^{PC} and λ_4^{PC} being constant (corresponding to the truncation set $\mathcal{A}^{M,0}$) and evaluate the associated BIC score:

$$\mathrm{BIC} \stackrel{\mathrm{def}}{=} -2\mathsf{L}(\hat{\boldsymbol{c}}) + \log(N)\|\hat{\boldsymbol{c}}\|_0, \tag{11}$$

where L is the log-likelihood function defined in Eq. (10), $\hat{\boldsymbol{c}}$ is the vector of the coefficients estimated by applying Eq. (9) to the selected truncation sets, and $\|\cdot\|_0$ is the ℓ^0-norm which gives the number of non-zero terms of a vector. Then, we increase the degree of λ_3^{PC} (λ_4^{PC}) by 1 and construct a generalized model accordingly, whose BIC score is calculated as BIC_3 (resp. BIC_4). The smaller value between BIC_3 and BIC_4 is defined by BIC_l with $l \in \{3, 4\}$ indicating the corresponding index. If BIC_l improves the current best BIC score, i.e., $\mathrm{BIC}_l < \mathrm{BIC}$, we update the optimal BIC score to BIC_l and keep the associated truncation set $\mathcal{A}_l$. We repeat this procedure until when increasing the degree for λ_3^{PC} and λ_4^{PC} does not further improve the BIC score.

3.3 Fragility analysis with generalized lambda models

After building a generalized lambda model, we can post-process it to calculate the fragility function defined in Eq. (1). For given values of IMs $\boldsymbol{x}$, one can evaluate Eq. (4) to calculate the distribution parameters $\boldsymbol{\lambda} = \boldsymbol{\lambda}^{\mathrm{PC}}(\boldsymbol{x}; \hat{\boldsymbol{c}})$ for the associated EDP. As the generalized lambda distribution is parameterized by its quantile function Eq. (2), one needs to inverse this nonlinear function to evaluate the cumulative probability function (CDF) that is necessary to calculate the exceeding probability.

More precisely, for a given threshold δ_0, we first solve

$$\delta_0 = Q(u; \boldsymbol{\lambda}). \tag{12}$$

Because δ_0 can be outside the range of the distribution, a solution between $[0, 1]$ may not exist. In this case, the CDF will take the value 0 or 1 depending on the range of the distribution and the value of δ_0. Therefore, the CDF evaluated at δ_0 is given by

$$\mathbb{P}(Y \le \delta_0 \mid \boldsymbol{X} = \boldsymbol{x}) = \begin{cases} 0 & Q(0; \boldsymbol{\lambda}) \ge \delta_0 \\ Q^{-1}(\delta_0; \boldsymbol{\lambda}) & Q(0; \boldsymbol{\lambda}) < \delta_0 < Q(1; \boldsymbol{\lambda}) \\ 1 & Q(1; \boldsymbol{\lambda}) \le \delta_0 \end{cases} \tag{13}$$

Finally, the fragility function is computed by

$$p_f(\boldsymbol{x}) = \mathbb{P}(Y \ge \delta_0 \mid \boldsymbol{X} = \boldsymbol{x}) = 1 - \mathbb{P}(Y \le \delta_0 \mid \boldsymbol{X} = \boldsymbol{x}). \tag{14}$$

Algorithm 1 Adaptive selection of truncation sets for λ_3^{PC} and λ_4^{PC}

Input: $\mathcal{A}_1$, $\mathcal{A}_2$, $\mathcal{X}$, $\mathcal{Y}$
$\mathcal{A}_3 \leftarrow \mathcal{A}^{M,0}$, $\mathcal{A}_4 \leftarrow \mathcal{A}^{M,0}$
Build a generalized lambda model following Eq. (9)
Calculate BIC associated with the model
$p_3 \leftarrow 0$, $p_4 \leftarrow 0$, Stop $\leftarrow$ **false**
while not Stop **do**
  $\mathcal{A}_3 \leftarrow \mathcal{A}^{M,p_3+1}$, $\mathcal{A}_4 \leftarrow \mathcal{A}^{M,p_4}$
  Build a generalized lambda model following Eq. (9)
  Calculate BIC_3 associated with the model
  $\mathcal{A}_3 \leftarrow \mathcal{A}^{M,p_3}$, $\mathcal{A}_4 \leftarrow \mathcal{A}^{M,p_4+1}$
  Build a generalized lambda model following Eq. (9)
  Calculate BIC_4 associated with the model
  $l \leftarrow \arg\min \{\mathrm{BIC}_l : l \in \{3,4\}\}$
  if $\mathrm{BIC}_l < \mathrm{BIC}$ **then**
    $p_l \leftarrow p_l + 1$, $\mathrm{BIC} \leftarrow \mathrm{BIC}_l$
  else
    Stop $\leftarrow$ **true**
  end if
end while
Output: $\mathcal{A}_3 \leftarrow \mathcal{A}^{p_3}$, $\mathcal{A}_4 \leftarrow \mathcal{A}^{p_4}$

4 Computational example

4.1 Synthetic ground motion

In this paper, we use a simplified version of the site-based ground motion model introduced in [12] for broad-band motions. The model is the frequency-domain version of the original time-domain model implemented in [13]. Specifically, the model is defined via spectral representation using an Evolutionary Power Spectral Density (EPSD), which preserves the key advantage of separating the temporal and spectral components of the process.

The spectral content of the process is represented by a normalized stationary Kanai-Tajimi Power Spectral Density (KT-PSD), which is completely characterized by the main frequency parameter, ω_g, and the bandwidth, ζ_g. Therefore, in this preliminary study, we neglect the non-stationary spectral characteristics of the ground motion. Moreover, we estimate the main frequency at the strong ground motion phase and fix the bandwidth parameter to 0.9 (representative of broad-band motions). The stationary process is modulated in time by a gamma modulating function ([13], [12]), which is completely defined by the expected Arias intensity I_a, the time at which 45% level of the expected Arias intensity is reached, t_{mid}, and the effective duration of the

motion, D_{5-95}. The complete EPSD is derived by combining the normalized KT-PDS and the modulating function; moreover, to ensure zero residual velocity and displacement, a high-pass filter is applied according to the evolutionary theory of Priestley as described in details in [12]. To summarize, the free parameters of the model are $\boldsymbol{x} = [I_a, t_{\text{mid}}, D_{5-95}, \omega_g]$.

Next, the model is fitted to a catalog of recorded far-field ground motions retrieved from the PEER NGA-West2 database. Specifically, the catalog includes 71 ground motions recorded at a range of distances (10-90 km) and site conditions from reverse earthquakes with magnitude between 6 and 7.6. The two horizontal components of each record are rotated into the major and intermediate principal directions. Only the major component is used in this study. The procedure used to estimate ω_g is described in details in [12]. However, in [12], ω_g is a time-varying function, while in this study ω_g corresponds to the main frequency of the ground motions at t_{mid} (which is considered the strong phase of the ground motion). The procedure to estimates the parameters $I_a, t_{\text{mid}}, D_{5-95}$ follows [13].

From the 71 estimates of the free parameters, we define and fit a joint-probability model to account for the parameter variability and their dependence structure. In essence, we transform the vector of model parameters, $\boldsymbol{x}$, in a random vector, $\boldsymbol{X}$. The probabilistic model is based on log-normal marginal distributions and a Gaussian copula (i.e., a joint log-normal distribution). The parameters defining the log-normal marginal distributions together with the Gaussian copula are reported in Table 1.

Provided with this joint probability model, the simulation of the ground motions is given by a two-step procedure. First, the model parameters are sampled from the joint log-normal distribution; then, time series are generated by filtering white noise Gaussian vectors with the EPSD and the high-pass filter. It is important to recognize that this is a stochastic simulator setting, i.e., for a given set of model parameters $\boldsymbol{X} = \boldsymbol{x}$, multiple time series can be generated. It follows that when these time series are used as inputs into a computational model, the model response of interest (i.e., the EDP) is a random variable even when $\boldsymbol{X} = \boldsymbol{x}$.

Table 1: Ground motion parameters

Name	Distribution
I_a	$\mathcal{LN}\left(-4.61, 1.45^2\right)$
t_{mid}	$\mathcal{LN}\left(2.55, 0.90^2\right)$
D_{5-95}	$\mathcal{LN}\left(2.67, 0.53^2\right)$
ω_g	$\mathcal{LN}\left(1.42, 0.59^2\right)$
Correlation matrix	$R = \begin{pmatrix} 1 & 0.015 & -0.23 & -0.13 \\ 0.015 & 1 & 0.68 & -0.36 \\ -0.23 & 0.68 & 1 & -0.11 \\ -0.13 & -0.36 & -0.11 & 1 \end{pmatrix}$

4.2 Three-story frame

In this Section, we introduce the properties of a three-story shear frame idealized as a three degree of freedom system. We are interested in the dynamic response of the system subjected to the ground motions generated according to Section 4.1 The interstory behavior is inelastic, with a force-interstory-drift relationship based on a Bouc-Wen hysteretic model [14]. Specifically, the i-th interstory restoring force is written as

$$q_i(v_i(t), \dot{v}_i(t)) = k_i \left[\alpha v_i(t) + (1-\alpha) z(t)\right], \tag{15}$$

where $v_i(t)$ denotes the interstory-drift, α is a parameter that controls the degree of inelasticity (i.e., $\alpha = 1$ corresponds to the linear case), k_i is the initial elastic interstory stiffness, and $z(t)$ is the hysteretic response governed by the following law

$$\dot{z}(t) = -\gamma \left|\dot{v}(t)\right| \left|z(t)\right|^{n-1} - \eta \left|z(t)\right|^{n} \dot{v}_i(t) + A\dot{v}(t), \tag{16}$$

where γ, n, A and η are the model parameters. The values of structural properties, including the local masses m_i and damping c_i, and model parameters are reported in Table 2. The story yield displacement, δ_y, is set to 0.01[m] and the post-hardening stiffness is set at 10% of the elastic stiffness k_i for all the three stories. The EDP of interest is the maximum interstory-drift, i.e.,

$$Y = \max\left[\max[v_1(t), v_2(t), v_3(t)]\right]. \tag{17}$$

Table 2: Structural properties and Bouc-Wen parameters

	m_i [kg]	c_i [Ns/m]	k_i [N/m]	α	n	$\gamma[1/\mathrm{m}^n]$	$\eta[1/\mathrm{m}^n]$	A
Storey 1	1E6	1.73E6	3.0E8	0.1	5	$1/(2\delta_y)^n$	$1/(2\delta_y)^n$	1
Storey 2	1E6	1.73E6	2.4E8	0.1	5	$1/(2\delta_y)^n$	$1/(2\delta_y)^n$	1
Storey 3	1E6	1.73E6	1.5E8	0.1	5	$1/(2\delta_y)^n$	$1/(2\delta_y)^n$	1

5 Numerical results

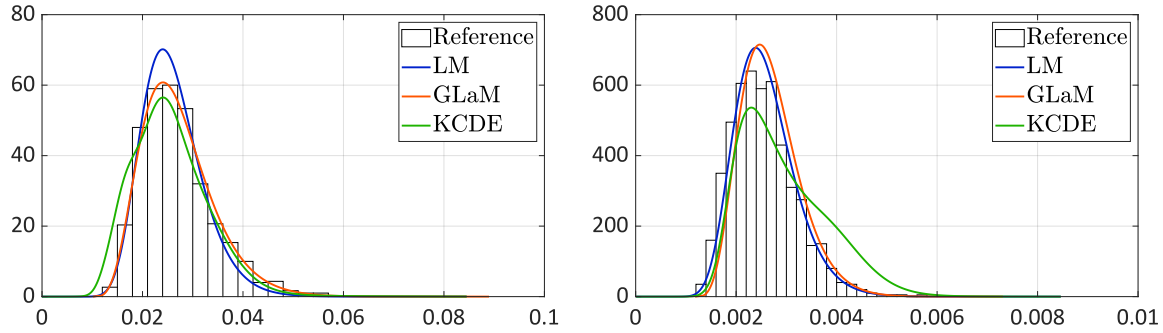
In this section, we compare the generalized lambda model (GLaM) in presented Section 3 with two other statistical models for fragility analysis of the case study in Section 4. The first model is a simple linear model (LM), where we assumes that the model output is a linear function of the IMs (in the log-space) with a homoskedastic Gaussian noise, i.e.,

$$\log\left(Y_{\boldsymbol{x}}\right) = \sum_{j=1} \beta_j \log(x_j) + e. \tag{18}$$

Here, β_j's are the coefficients, and $e \sim \mathcal{N}(0, \sigma^2)$. $\boldsymbol{\beta}$ and σ are estimated by ordinary least-squares [15]. The second model is a state-of-the-art kernel conditional density estimator (KCDE) from the package `np` [16] implemented in R.

We select the four ground motion parameters in Table 1 as IMs. This choice allows for not only using replications to validate the surrogate models but also capturing the dependency between the source and the EDP without going through intermediate variables [2].

To get data, we generate 10^5 realizations of the random ground motion parameters with Latin hypercube sampling (LHS) [17]. The 3-DOF nonlinear Bouc-Wen model is evaluated once for each realization. To build the surrogate model, we randomly *subsample* this big data set to obtain training dataset of a desired size N.



(a) $\boldsymbol{x} = (0.03, 16.9, 9.6, 1.6)$ (b) $\boldsymbol{x} = (0.001, 8.9, 12.6, 6.7)$

Figure 1: Comparisons of emulated PDFs, $N = 1{,}000$.

Fig. 1 illustrates the PDFs of the maximum interstory drift for two sets of IMs. The surrogate models in comparison are built on a training set of size $N = 1{,}000$. The results demonstrate that GLaM outperforms the other two models. In the left plot Fig. 1a, GLaM more accurately represents the tail of the distribution. In the right plot Fig. 1b, GLaM has a similar performance to the linear model: both models approximate well the underlying distribution, whereas the nonparametric model KCDE overestimates the spread of the distribution.

To quantitatively compare the accuracy of the surrogate models for representing the conditional distribution, we define an error metric by

$$\varepsilon = \frac{\mathbb{E}_{\boldsymbol{X}}\left[d^2_{\mathrm{WS}}\left(Y_{\boldsymbol{X}}, \tilde{Y}_{\boldsymbol{X}}\right)\right]}{\mathrm{Var}\left[Y\right]} \tag{19}$$

where $Y_{\boldsymbol{x}}$ is the EDP for given IMs $\boldsymbol{x}$, $\tilde{Y}_{\boldsymbol{x}}$ corresponds to that of the emulator, and d_{WS} is the *Wasserstein distance of order two* [18] defined by

$$\begin{aligned} d^2_{\mathrm{WS}}\left(Y_1, Y_2\right) &\overset{\mathrm{def}}{=} \|Q_1 - Q_2\|_2^2 \\ &= \int_0^1 \left(Q_1(u) - Q_2(u)\right)^2 \mathrm{d}u\,, \end{aligned} \tag{20}$$

where Q_1 and Q_2 are the quantile functions of two random variables Y_1 and Y_2, respectively.

In this study, we use a test set of 50 to calculate the expectation in Eq. (19). For each point of the test set, we repeatedly run the simulator 1,000 times as a reference to evaluate Eq. (20). To compare the convergence behaviors of the surrogate models, we vary the size of the training

set $N \in \{250; 500; 1{,}000; 2{,}000; 4{,}000\}$. Each scenario is run 20 times with independent training samples to account for statistical uncertainty.

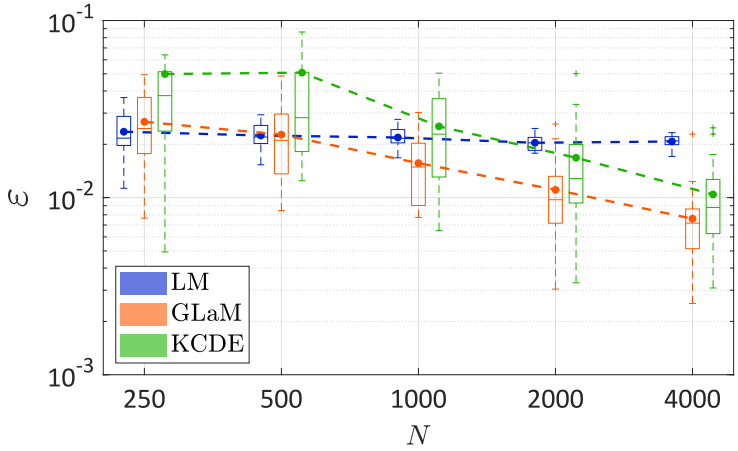


Figure 2: Comparison of the convergence among the surrogate models in terms of the normalized Wasserstein distance. The dashed lines denote the average value over 20 repetitions of the full analysis.

Fig. 2 summarizes the convergence of the surrogate models. For a few data $N = 250$, LM provides slightly more accurate results than GLaM. This is because the linear model is very simple and restrictive, which can result in a large bias but a small variance. Increasing N almost does not improve the performance of the linear model. This implies that the error is mainly dominated by the bias. For large data set, namely $N \geq 2{,}000$, LM performs the worst among the three models. In contrast, GLaM and KCDE are flexible models, and thus they have a smaller bias but a bigger variance. With increasing N, both models show a clear decay of the errors. For $N = 500$, GLaM has a similar performance to LM. For $N \geq 1{,}000$, GLaM outperforms the other two models.

Now, we study the accuracy of the surrogate models for estimating failure probabilities in the context of fragility analysis. In this study, we fix the threshold to $\delta_0 = 0.02$[m], i.e., the structure fails when the maximum interstory drift exceeds 0.02m.

Following Eq. (1), the failure probability is a function of the IMs. As a result, we use the normalized mean-squared error as a performance indicator

$$\varepsilon_p = \frac{\mathbb{E}_X\left[\left(p_f(\boldsymbol{X}) - \tilde{p}_f(\boldsymbol{X})\right)^2\right]}{\mathrm{Var}\left[p_f(\boldsymbol{X})\right]}, \tag{21}$$

where $p_f(\boldsymbol{x})$ is the reference value calculated from 1,000 replications at $\boldsymbol{x}$, $\tilde{p}_f(\boldsymbol{x})$ is the failure probability predicted by the surrogate model. The expectation and variance are estimated from the test set (of size 50).

Fig. 3 compares the accuracy of the surrogate models in terms of estimating the fragility function with the threshold value $\delta_0 = 0.02$m. Similar to Fig. 2, the performance of LM does not improve by using more data. Despite the decay of the error, KCDE estimates poorly the failure probability

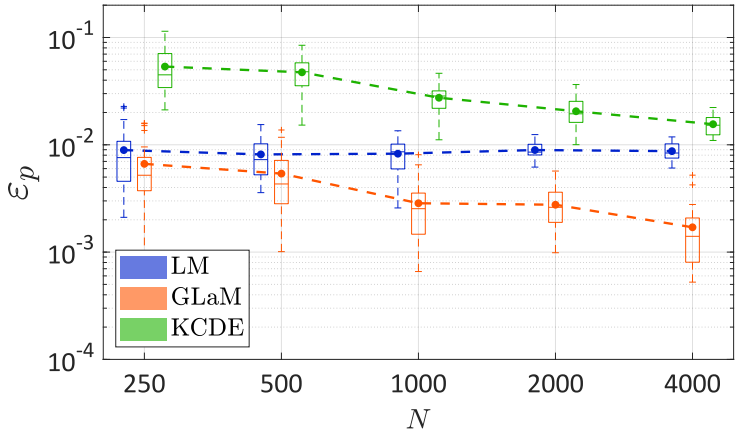


Figure 3: Comparisons of the surrogate models for the estimation of the fragility function with the threshold value $\delta_0 = 0.02$m. The dashed lines denote the average value over 20 repetitions of the full analysis.

even for a large number of data. GLaM clearly outperforms the other models for all sizes of the data. Moreover, for $N = 4{,}000$, the average error of GLaM (1.7×10^{-3}) is almost one order of magnitude less than that of LM (8.7×10^{-3}) and KCDE (1.6×10^{-2}).

Finally, we plot the fragility function in the $I_a - \omega_g$ plan of a GLaM built upon 1,000 model evaluations in Fig. 4. The other variables are fixed at their mean value. As a comparison, we run the simulator for a validation set of nine points obtained by the Cartesian product of $I_a \in \{0.02, 0.06, 0.1\}$ and $\omega_g \in \{2, 6, 10\}$. The reference failure probability associated with each validation point is computed by 1,000 replications (i.e., totally 9,000 model runs for validation). As seen in Fig. 4, the diamonds representing the reference points lie fairly well on the estimated fragility surface. The maximum absolute error of GLaM among the 9 validation points is 0.08 at $I_a = 0.1$ and $\omega_g = 6$.

6 Conclusions

Fragility analysis plays a central role in performance-based earthquake engineering. This consists in computing the probability that the engineering demand parameter exceeds a certain level given the values of intensity measures. In this paper, we propose using the generalized lambda model to emulate the dependence of EDP on IMs. In such a surrogate, the conditional distribution is represented by the flexible generalized lambda distribution. The four distribution parameters are modelled by polynomial chaos expansions as functions of IMs. As this model produces the conditional distribution, the fragility function can be directly computed by post-processing the surrogate.

Compared to the original approach for GLaM presented in [3], we propose in this contribution an adaptive algorithm to select appropriate truncation degrees for the shape parameters. This novel approach is compared with a simple linear model and a state-of-the-art non-parametric

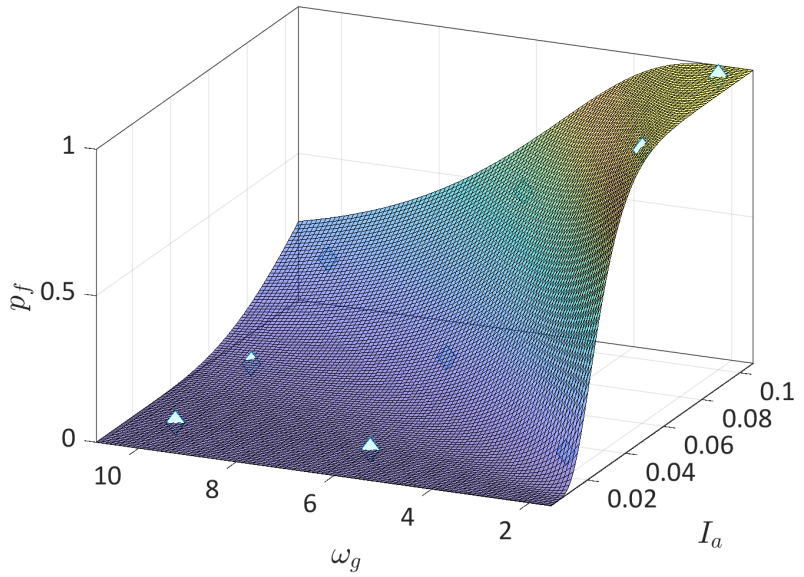


Figure 4: Fragility function in the $I_a - \omega_g$ plan of a GLaM built on 1,000 samples. The diamond points correspond to the reference value computed from 1,000 replications.

model on a numerical example using synthetic ground motions calibrated from records. In the case study of a three-story frame, the four parameters used to generate artificial ground motions are considered as IMs, and the maximum interstory displacement is selected as EDP. Based on this choice, we use replications to obtain a reference distribution of the EDP conditioned on given values of the IMs.

The numerical results show that the proposed approach can accurately represent the EDP-IMs dependence. In terms of approximating the entire conditional distribution, the generalized lambda model generally outperforms its two counterparts. Compared with the simple linear model, it demonstrates a similar performance for relatively small sample sizes, but it shows a much steeper decay of the errors. Compared with the non-parametric model, it always yields smaller errors. Moreover, the generalized lambda model always gives the most accurate estimation of the fragility function for all sizes of the training data.

For future research, we will examine the accuracy of the generalized lambda model in estimating fragility curves with other classical IMs, e.g., peak ground acceleration and spectral acceleration [19]. Besides, we plan to incorporate the model in the PBEE framework to evaluate the decision variables [5], as the conditional distribution is directly given by the surrogate.

Acknowledgment

This paper is a part of the project “Surrogate Modeling for Stochastic Simulators (SAMOS)” funded by the Swiss National Science Foundation (Grant #200021_175524), whose support is

gratefully acknowledged.

References

[1] C. A. Cornell and H. Krawinkler. Progress and challenges in seismic performance assessment. *PEER center news*, 3(2):1–3, 2000.

[2] G. Abbiati, M. Broccardo, S. Marelli, and F. Paolacci. Seismic fragility analysis based on artificial ground motions and surrogate modeling of validated structural simulators. *Earthq. Eng. Struct. Dyn.*, 9:2314–2333, 2021.

[3] X. Zhu and B. Sudret. Emulation of stochastic simulators using generalized lambda models. *SIAM/ASA J. Unc. Quant.*, 9:1345–1380, 2021.

[4] Kevin Mackie and Božidar Stojadinović. *Seismic demands for performance-based design of bridges*. Pacific Earthquake Engineering Research Center Berkeley, 2003.

[5] J. W. Baker and C. A. Cornell. Uncertainty propagation in probabilistic seismic loss estimation. *Structural Safety*, 30:236–252, 2008.

[6] X. Zhu and B. Sudret. Replication-based emulation of the response distribution of stochastic simulators using generalized lambda distributions. *Int. J. Uncertainty Quantification*, 10:249–275, 2020.

[7] M. Freimer, G. Kollia, G.S. Mudholkar, and C.T. Lin. A study of the generalized Tukey lambda family. *Comm. Stat. Theor. Meth.*, 17:3547–3567, 1988.

[8] D. Xiu and G.E. Karniadakis. The Wiener-Askey polynomial chaos for stochastic differential equations. *SIAM J. Sci. Comput.*, 24(2):619–644, 2002.

[9] E. Torre, S. Marelli, P. Embrechts, and B. Sudret. Data-driven polynomial chaos expansion for machine learning regression. *J. Comput. Phys.*, 388:601–623, 2019.

[10] G. Blatman and B. Sudret. Adaptive sparse polynomial chaos expansion based on Least Angle Regression. *J. Comput. Phys.*, 230:2345–2367, 2011.

[11] G. Schwarz. Estimating the dimension of a model. *Ann. Statist.*, 6(2):461–464, 1978.

[12] Marco Broccardo and Mayssa Dabaghi. A spectral-based stochastic ground motion model with a non-parametric time-modulating function. In *12th International Conference on Structural Safety and Reliability; Vienna*, volume 2017, pages 1–10, 2017.

[13] A. Rezaeian and A. Der Kiureghian. Simulation of synthetic ground motions for specified earthquake and site characteristics. *Earthq. Eng. Struct. Dyn.*, 39:1155–1180, 2010.

[14] Yi-Kwei Wen. Method for random vibration of hysteretic systems. *J. Eng. Mech. Div. (ASCE)*, 102(2):249–263, 1976.

[15] J.M. Wooldridge. *Introductory Econometrics: A Modern Approach*. Cengage Learning, 5th

edition, 2013.

[16] T. Hayfield and J.S. Racine. Nonparametric Econometrics: The np Package. *J. Stat. Softw.*, 27:1015–1026, 2008.

[17] M. D. McKay, R. J. Beckman, and W. J. Conover. A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics*, 21(2):239–245, 1979.

[18] C. Villani. *Optimal transport, old and new*. Cambridge Series in Statistical and Probabilistic Mathematics. Springer, Cambridge, 2000.

[19] C. Mai, K. Konakli, and B. Sudret. Seismic fragility curves for structures using non-parametric representations. *Front. Struct. Civ. Eng.*, 11:169–186, 2017.