Design of experiment based on a low fidelity model for seismic fragility estimation

Antoine Van Biesbroeck, Clément Gauchy, Cyril Feau, Josselin Garnier

Preprint of an article accepted for publication in
ESAIM: PROCEEDINGS AND SURVEYS, 2024 (to appear)

DESIGN OF EXPERIMENTS BASED ON A LOW FIDELITY MODEL FOR SEISMIC FRAGILITY CURVES ESTIMATION

Antoine Van Biesbroeck[1, 2], Clément Gauchy[3], Cyril Feau[2] and Josselin Garnier[1]

Abstract. Seismic fragility curves are key quantities of interest for Seismic Probabilistic Risk Assessment studies. They express the probability of failure of a mechanical structure of interest conditional to a scalar value derived from the ground motion signal coined Intensity Measure. In the literature, Bayesian approaches have emerged to enable their estimation within the difficult context of limited data availability. Yet, the log-normal modeling over which most of them are based requires the use of computationally expensive Markov chain Monte Carlo methods for providing Bayesian estimators. In this work, we propose an efficient modeling for the estimation of fragility curves in the Bayesian context, based on a low fidelity model of the structure's response to the ground motion signal and an objective prior. The analytical expression of our modeling allows fast generation of estimates. Also, the representative bias arisen by the modeling choice is handled with a judicious design of experiments methodology. Finally, our method is evaluated on a real case study, and the results highlight its efficiency and its ability to robustly overcome any bias when coupled with the design of experiments we propose.

Résumé. Les courbes de fragilité sismiques représentent un outil d'aide à la décision essentiel dans le cadre des études probabilistes de sûreté. Elles quantifient la probabilité de défaillance d'une structure mécanique d'intérêt conditionnellement à une valeur scalaire appelée Mesure d'Intensité et extraite du signal sismique. Dans la littérature, les approches bayésiennes s'illustrent par leurs capacités à estimer ces courbes dans un contexte où peu de données sont à disposition. Cependant, la modélisation log-normale sur laquelle la plupart se reposent nécessite l'emploi de méthodes Monte-Carlo par chaînes de Markov dont le coût computationel est élevé. Dans ce travail, nous proposons une modélisation efficace pour l'estimation de courbes de fragilité dans le canevas bayésien, basé sur un modèle basse fidélité de la réponse de la structure au signal sismique d'entrée et un prior objectif. L'expression analytique de notre modèle permet la génération rapide d'estimations *a posteriori*. Aussi, le biais de représentation issu du choix de modélisation est pris en considération, via la conception d'une méthodologie de planification d'expérience judicieuse. Enfin, notre méthode est évaluée sur un cas d'étude concret, et nos résultats mettent en avant son efficacité et sa capacité de s'affranchir avec robustesse de tout biais lorsqu'elle est couplée à la planification d'expérience que nous proposons.

[1] CMAP, CNRS, École polytechnique, Institut Polytechnique de Paris, 91120 Palaiseau, France ;
e-mail: antoine.van-biesbroeck@polytechnique.edu
[2] Université Paris-Saclay, CEA, Service d'Études Mécaniques et Thermiques, 91191 Gif-sur-Yvette, France
[3] Université Paris-Saclay, CEA, Service de Génie Logiciel pour la Simulation, 91191 Gif-sur-Yvette, France

Introduction

The probabilistic seismic risk assessment framework (SPRA) introduced in the 1980s for the nuclear industry is based on the estimation of seismic fragility curves, for the structures and components (SCs) of interest [1–5]. These curves are defined as the conditional probability that an engineering demand parameter (EDP) – such as the interstory drift ratio – exceeds a limit threshold, given a scalar value derived from the seismic ground motion and called intensity measure (IM). The IM can be for instance the peak ground acceleration (PGA) or a pseudo-spectral acceleration (PSA) evaluated for a given frequency and damping ratio [6–8]. As explained in [5], it is therefore assumed that the seismic hazard, on a given site, can be reduced to such a single indicator.

Practitioners have several data sources at their disposal to estimate fragility curves, namely: expert judgments supported by test data [1–3, 9], experimental data [3, 10, 11], results of damage collected on existing structures that have been subjected to earthquakes [12–14] as well as analytical results given by more or less refined numerical models using synthetic or real seismic excitations [15–20]. Over the years, many methods have been developed to estimate these curves [9, 12, 13]. Nowadays, even though machine learning techniques are becoming very popular [7, 19, 21–24], parametric fragility curves historically introduced in the SPRA framework are ubiquitous in practice and the log-normal model is the most widely used model due to its proven ability to handle limited data [12–20, 25–30].

Different strategies can be implemented to estimate the parameters that define the fragility curve in the log-normal model. Among these we distinguish the Bayesian framework [10, 19, 30–37]. This framework is interesting because it allows to solve the irregularity issues encountered when few data are available. This occurs with the widely used maximum likelihood estimation coupled with a bootstrap technique to estimate a confidence interval when the data are binary, that is, when they represent the failing or non-failing state of the structure [38, 39]. In practice, binary data are encountered when dealing with tests performed on shaking tables for instance.

In earthquake engineering, within the SPRA framework, Bayesian inference is often used to update log-normal fragility curves obtained beforehand by various approaches, by assuming independent distributions for the prior values of the parameters, such as log-normal distributions for instance [14, 19, 30, 33, 34]. Recently, based on the reference prior theory, the authors proposed the use of an objective prior [40, 41], in order to remove any subjectivity that could legitimately lead to inevitable open questions on the influence of the *a priori* on the quantities of interest [38]. In all these approaches, the use of Markov chain Monte Carlo (MCMC) methods is nevertheless necessary to sample the *a posteriori* distribution of the parameters, which can prove cumbersome to implement, particularly if we want a rapid first estimate of a fragility curve with limited data.

We circumvent this problem in this work by proposing an effective approach for the estimation of fragility curves, which avoids the use of the MCMC method. We rely on a low-fidelity linear model between the logarithm of the Engineering Demand Parameter and the one of the IM [9, 13, 25, 42]. Supported by the Bayesian framework, our model benefits from a fully analytical form; that former allows an efficient implementation and a solution for fast generation of estimates with limited data. The reliability of the Bayesian scheme w.r.t. its prior choice is answered as well with the derivation of an objective prior derived on the basis of the reference prior theory [41, 43]. Finally, since the low-fidelity model is a linear model, we also propose a sequential planning of experiments strategy to minimize the representation bias. The design we suggest relies on the maximization of the information brought by the observation of a new data item onto the posterior distribution. That one is measured through global sensitivity indices described in [44].

The remainder of this paper is organized as follows: the statement of our low-fidelity modeling strategy for the fragility curves estimation in a Bayesian framework is presented in section 1. After a brief review devoted to global sensitivity analysis, we describe in section 2 our design for a sequential planning of experiments, taking

the global sensitivity indices as a support. Section 3 is dedicated to the implementation of our methodology on a case study from the nuclear industry. Finally, Section 4 which precedes the conclusion offers a discussion on the performance of our method.

1. Low fidelity model for fragility curves

1.1. Linear regression model

The fragility curve which we seek to estimate is defined by

$$P_f(a) = \mathbb{P}(\text{EDP} > C \,|\text{IM} = a). \tag{1}$$

Up to a refinement of a multiplicative constant in the definition of the engineering demand parameter (EDP), C can be supposed to be equal to 1 in what follows. The EDP is supposed to be correlated with the intensity measure (IM) as follows

$$\log \text{EDP} = \rho \log \text{IM} + \epsilon, \tag{2}$$

where the random variable ϵ follows the distribution $\mathcal{N}(\mu, \sigma^2)$. Here ρ is supposed unknown, as well as μ and σ.

1.2. Likelihood

We have at our disposal a data-set composed by the tuples $(a_i, y_i)_{i=1}^k$, $a_i \in \mathcal{A} \subset (0, \infty)$ denoting the measured IM from the i-th seismic ground motion signal and $y_i \in \mathcal{Y} \subset (0, \infty)$ denoting the structure's EDP. Conditionally to the parameter $\theta = (\rho, \mu, \sigma)$, we suppose the observations to be independent and identically distributed. As a_i follows a distribution assumed to admit a density $a \mapsto p(a)$ w.r.t. the Lebesgue measure and to be independent of θ, the distribution of the y_i is known conditionally to (a_i, θ):

$$\log y_i | a_i, \theta \sim \mathcal{N}(\rho \log a_i + \mu, \sigma^2). \tag{3}$$

The likelihood for the parameter θ is therefore:

$$\ell_k^0(\hat{\mathbf{y}}, \hat{\mathbf{a}}|\theta) = \prod_{i=1}^{k} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(\hat{y}_i - \rho\hat{a}_i - \mu)^2}{2\sigma^2}\right) p(\hat{a}_i), \tag{4}$$

where $\hat{y}_i$ (resp. $\hat{a}_i$) denotes $\log y_i$ (resp. $\log a_i$) and $\hat{\mathbf{y}}$ (resp. $\hat{\mathbf{a}}$) denotes the vector $(\log y_i)_{i=1}^k$ (resp. $(\log a_i)_{i=1}^k$).

This likelihood introduces a challenge due to the lack of clear separation among the three parameters that constitute θ. Within the Bayesian framework, which we develop later, this challenge could result in a posterior distribution that is hardly tractable. We address this issue by introducing the following quantities:

$$\rho_k = \frac{\text{Cov}_k(\hat{\mathbf{y}}, \hat{\mathbf{a}})}{\text{Var}_k\, \hat{\mathbf{a}}}, \quad \mathbf{z} = \tilde{P}_{\mathbf{a}}^T(\hat{\mathbf{y}} - \rho_k \hat{\mathbf{a}}), \tag{5}$$

where Cov_k, Var_k respectively denote the empirical covariance and variance, and the matrix $\tilde{P}_{\mathbf{a}}$ is defined in appendix A. Their conditional distributions are given by:

$$\rho_k | \mathbf{a}, \theta \sim \mathcal{N}\left(\rho, \frac{\sigma^2}{k\,\text{Var}_k\, \hat{\mathbf{a}}}\right), \tag{6}$$

$$\mathbf{z} | \mathbf{a}, \theta \sim \mathcal{N}(\mu \tilde{P}_{\mathbf{a}}^T \mathbf{1}, \sigma^2 D); \quad D = \text{diag}\left(1, \ldots, 1, \frac{\hat{\mathbf{a}}^T\hat{\mathbf{a}}}{k\,\text{Var}_k\, \hat{\mathbf{a}}}\right), \tag{7}$$

where $\mathbf{a}$ is the vector $(a_i)_{i=1}^k$, $\mathbf{1}$ denotes the vector of $\mathbb{R}^k$ composed only of ones, and $\mathrm{diag}(\lambda_1, \dots, \lambda_{k-1})$ refers to the diagonal matrix of $\mathbb{R}^{(k-1)\times(k-1)}$ whose diagonal coefficients are the $(\lambda_i)_{i=1}^{k-1}$.

Let us denote by $(w_i)_{i=1}^k$ the orthogonal columns of the matrix $P_{\mathbf{a}}$ defined in appendix A. Noticing $\rho_k = w_k^T \hat{\mathbf{y}}$ and $\mathbf{z}$ is spanned by the $(w_i)_{i=1}^{k-1}$, we deduce that $\mathbf{z}$ and ρ_k are independent conditionally to $(\mathbf{a}, \theta)$, and that the knowledge of $(\mathbf{z}, \rho_k, \mathbf{a})$ is equivalent to the one of $(\mathbf{y}, \mathbf{a})$ or $(\hat{\mathbf{y}}, \hat{\mathbf{a}})$. Thus, the likelihood issued from the observation of $(\mathbf{z}, \rho_k, \mathbf{a})$ is

$$\ell_k(\mathbf{z}, \rho_k, \mathbf{a}|\theta) = p(\mathbf{a}) \frac{\|\hat{\mathbf{a}} - \overline{\hat{\mathbf{a}}}\|}{\sqrt{2\pi \hat{\mathbf{a}}^T \hat{\mathbf{a}} \sigma}} \exp\left(- \frac{(z_{k-1} - \mu\sqrt{k})^2}{2\sigma^2 \frac{\hat{\mathbf{a}}^T \hat{\mathbf{a}}}{\|\hat{\mathbf{a}} - \overline{\hat{\mathbf{a}}}\|^2}} \right) \exp\left(- \frac{(\rho_k - \rho)^2}{2\sigma^2 / (k \operatorname{Var}_k \hat{\mathbf{a}})} \right) \prod_{i=1}^{k-2} \frac{1}{\sqrt{2\pi}\sigma} \exp\left(- \frac{z_i^2}{2\sigma^2} \right), \quad (8)$$

where $\overline{\hat{\mathbf{a}}} = \frac{1}{k} \sum_{i=1}^k \log a_i$.

1.3. Prior and posterior

Within a Bayesian context, the parameter of interest θ is itself a random variable taking values in a space $\Theta \subset \mathbb{R}^2 \times (0, \infty)$ and following a distribution called the prior. We take as a support the reference prior theory [41, 43] to justify the choice of the Jeffreys's prior for θ, derived from the likelihood expressed in equation (8). Conditionally to $\mathbf{a}$, that former is a Gaussian density, making the associated Fisher information matrix being:

$$\mathcal{I}(\theta) = \int_{\mathcal{A}^k} \begin{pmatrix} k \frac{\|\hat{\mathbf{a}} - \overline{\hat{\mathbf{a}}}\|^2}{\sigma^2 \hat{\mathbf{a}}^T \hat{\mathbf{a}}} & 0 & 0 \\ 0 & k \frac{2}{\sigma^2} & 0 \\ 0 & 0 & \frac{k \operatorname{Var}_k \hat{\mathbf{a}}}{\sigma^2} \end{pmatrix} \prod_{i=1}^k p(a_i) \, \mathrm{d}a_i. \quad (9)$$

The Jeffreys' prior being defined as the one whose density J w.r.t. the Lebesgue measure is proportional to $\sqrt{|\mathcal{I}(\theta)|}$, we obtain

$$J(\theta) \propto \frac{1}{\sigma^3}. \quad (10)$$

Finally, the posterior distribution of θ is given by its density, which is proportional to the product of the likelihood (from equation (8)) with the prior:

$$p(\theta|\mathbf{z}, \mathbf{a}, \rho_k) \propto \frac{1}{\sigma^{k+3}} \exp\left(- \frac{\sum_{i=1}^{k-2} z_i^2}{2\sigma^2} \right) \exp\left(- \frac{k\|\hat{\mathbf{a}} - \overline{\hat{\mathbf{a}}}\|^2}{\hat{\mathbf{a}}^T \hat{\mathbf{a}}} \frac{(z_{k-1} k^{-1/2} - \mu)^2}{2\sigma^2} \right) \exp\left(- \frac{(\rho_k - \rho)^2}{2\sigma^2 / k \operatorname{Var}_k \hat{\mathbf{a}}} \right). \quad (11)$$

We recognize the above as a product of square inverse gamma distributions. More precisely, σ^{-2} follows a gamma distribution, and μ and ρ follow independent Gaussian distributions conditionally to σ.

This posterior allows to elucidate the distribution of what expresses the fragility curve:

$$P_f(a)|\theta \sim \mathbb{P}(\hat{y} > 0 | \hat{a}, \theta) = \Phi\left(\frac{\rho \log a + \mu}{\sigma} \right), \quad (12)$$

with Φ being the c.d.f. of a standard Gaussian distribution.

Its distribution is known *a posteriori*, given that $\frac{\rho \log a + \mu}{\sigma}$, conditionally to $(a, \mathbf{a}, \mathbf{y})$ for any $a \in \mathcal{A}$, is distributed as the sum of a variable with Gaussian distribution and the square root of a variable with Gamma distribution (both variables being independent):

$$\frac{\rho \log a + \mu}{\sigma} | a, \mathbf{a}, \mathbf{y} \sim \mathcal{N}\left(0, \frac{\log^2 a}{k \operatorname{Var}_k \hat{\mathbf{a}}} + \frac{\hat{\mathbf{a}}^T \hat{\mathbf{a}}}{k\|\hat{\mathbf{a}} - \overline{\hat{\mathbf{a}}}\|^2} \right) + \left(\rho_k \log a + \frac{z_{k-1}}{\sqrt{k}} \right) \Gamma^{1/2}(\tilde{c}, \tilde{d}), \quad (13)$$

$$\tilde{c} = k/2, \qquad \tilde{d} = \frac{1}{2} \sum_{i=1}^{k-2} z_i^2. \quad (14)$$

2. Sensitivity index for design of experiments

2.1. A review of the global sensitivity analysis

Global sensitivity analysis (GSA) is a cornerstone of uncertainty quantification studies of computer simulators. It aims at quantifying how the uncertainties within the observed output of a model are influenced by the uncertainties of one or several of its inputs [45]. More formally, in classical GSA settings, a system outputs an observed variable Y, supposed to be a function of input variables $Y = \eta(X_1, \dots, X_p)$, where the input X_i's are assumed to follow a known distribution and to be mutually independent. Since the first indices introduced by Sobol' [46], GSA's tools measure statistically how Y is impacted by one or some of the X_i [47]. Global sensitivity indices [44] are quantities whose class regroups a large range of these tools. According to their definition, letting d be a dissimilarity measure between probability distributions, the impact of input X_i onto the output Y can be derived as

$$S_i = \mathbb{E}_{X_i}\left[d(\mathbb{P}_Y || \mathbb{P}_{Y|X_i})\right], \tag{15}$$

where $\mathbb{P}_Y$ is the distribution of Y, $\mathbb{P}_{Y|X_i}$ is the conditional distribution of Y given X_i, and d is a dissimilarity measure. The choice of d can depend on the expected properties. A classical example is to set $d(P||Q) = \|\mathbb{E}_{X\sim P}[X] - \mathbb{E}_{X\sim Q}[X]\|^2$ which gives the un-normalized Sobol' index [46].

2.2. Sequential planning of experiments via global sensitivity index maximization

Following the idea of Da Veiga [44], a judicious data acquisition strategy would be to minimize the sensitivity that the posterior would get from the observations. This way, the IM a_{k+1} that has to be chosen for the next simulation which would output an EDP y_{k+1}, after having observed $(\mathbf{y}, \mathbf{a}) = (y_i, a_i)_{i=1}^k$ is the one such that the following index is maximized:

$$\mathbb{E}_{y_{k+1}|a_{k+1},\mathbf{y},\mathbf{a}}[d(\mathbb{P}_{\theta|\mathbf{y},\mathbf{a}}||\mathbb{P}_{\theta|y_{k+1},a_{k+1},\mathbf{y},\mathbf{a}})], \tag{16}$$

where $\mathbb{E}_{y_{k+1}|a_{k+1},\mathbf{y},\mathbf{a}}$ is the expectation with respect to the distribution of y_{k+1} given $a_{k+1}, \mathbf{y}, \mathbf{a}$. Within the GSA's scope, this makes the next experiment being chosen as the one such that the resulting observation of the structure's response provides the most impact onto the parameter of interest θ. Sequentially, the observations are chosen to maximize the evolution of the posterior distribution. We invite one to notice that this viewpoint joins the reference prior theory one. Indeed, the reference prior is built to be the one such that the posterior distribution is expected to evolve the most from the prior [43].

If the relation between the logarithm of the EDP and that of the IM is "very close" to a linear relation, equation (16) is sufficient to improve the learning of the fragility curve. In other words, a strategy based on this equation makes it possible to sufficiently explore the space of the IMs, in order to maximize their empirical variance and thus reduce the *a posteriori* variance of the estimation of the fragility curve, all things being equal (cf. equation (13)). Note that there is no mathematical proof of this in this paper, but it has been tested numerically.

In practice, since the linear model is expected to be biased, the way to reduce the bias is to localize the learning, even if it is not optimal with respect to the *a posteriori* variance of the estimation of the fragility curve. In our work, the locality of interest corresponds to the values of IMs for which the fragility curve evolves "significantly" from 0 to 1. For this reason, we propose a refinement of the data acquisition strategy of equation (16) which includes the researched information:

$$\mathbb{E}_{s_{k+1}|a_{k+1},\mathbf{y},\mathbf{a}}[d(\mathbb{P}_{\theta|\mathbf{y},\mathbf{a}}||\mathbb{P}_{\theta|s_{k+1},a_{k+1},\mathbf{y},\mathbf{a}})], \tag{17}$$

with $s_{k+1} = \mathbb{1}_{\hat{y}_{k+1}>0}$.

As a dissimilarity measure, we suggest the following, defined as a Sobol' index of the fragility curve:

$$\begin{aligned} d(\mathbb{P}_{\theta|\mathbf{y},\mathbf{a}}||\mathbb{P}_{\theta|s_{k+1},a_{k+1},\mathbf{y},\mathbf{a}}) &= \|\mathbb{E}[P_f|\mathbf{y},\mathbf{a}] - \mathbb{E}[P_f|s_{k+1},a_{k+1},\mathbf{y},\mathbf{a}]\|_{\mathrm{L}^2}^2 \\ &= \int_{\mathcal{A}} |\mathbb{E}[P_f(a)|\mathbf{y},\mathbf{a}] - \mathbb{E}[P_f(a)|s_{k+1},a_{k+1},\mathbf{y},\mathbf{a}]|^2 \, \mathrm{d}a, \end{aligned} \tag{18}$$

where for any $a \in \mathcal{A}$, $P_f(a) = \Phi\left((\rho \log a + \mu)/\sigma\right)$ inherits from the distribution of θ. Conditionally to $(\mathbf{y}, \mathbf{a})$, its distribution has been elucidated in section 1.3. Also,

$$p(\theta|s_{k+1}, a_{k+1}, \mathbf{y}, \mathbf{a}) = \frac{p(s_{k+1}|a_{k+1}, \mathbf{y}, \mathbf{a}, \theta)}{\mathbb{E}[p(s_{k+1}|a_{k+1}, \mathbf{y}, \mathbf{a}, \theta)|\mathbf{y}, \mathbf{a}]} p(\theta|\mathbf{y}, \mathbf{a}) \tag{19}$$

with $p(s_{k+1}|a_{k+1}, \mathbf{y}, \mathbf{a}, \theta) = P_f(a_{k+1})^{s_{k+1}}(1 - P_f(a_{k+1}))^{1-s_{k+1}}$. Thus, samples of θ conditionally to $(\mathbf{y}, \mathbf{a})$ allow the approximation of both expectations in equation (18) by Monte-Carlo averages. The integrals in a are estimated by Simpson's rule. In the following example, a regular subdivision of $\mathcal{A} = [0, A_{\max}]$ is suggested (see section 3.2).

The calculation of this index necessitates several initial observations. Actually, the derivations of both the likelihood and the posterior, as outlined in sections 1.2 and 1.3, require that $k > 2$ and $a_1 \neq a_2$ (refer to appendix A). In our experiments, we randomly select $k_0 = 3$ initial seismic signals with distinct IMs from their original distribution. The planning of experiment strategy is then sequentially implemented to select subsequent IM values by maximizing the numerical approximation of the index expressed in equation (17). The optimization in one dimension is carried out using the BFGS algorithm.

3. Numerical application

3.1. **Case study presentation**

This case study concerns the seismic behavior of a piping system forming part of the secondary line of a French pressurized water reactor. Figure 1 presents a perspective of the mock-up positioned on the Azalee shaking table at the EMSI laboratory of CEA/Saclay. Simultaneously, figure 1-right depicts the finite element model (FEM), employing beam elements and implemented through the proprietary FE code CAST3M [48]. The validation of the FEM was carried out thanks to an experimental campaign described in [49].

The mock-up comprises a carbon steel TU42C pipe with an outer diameter of 114.3 mm, a thickness of 8.56 mm, and a 0.47 elbow characteristic parameter. This pipe, filled with water without pressure, includes three elbows, with a valve-mimicking mass of 120 kg, constituting over 30% of the mock-up's total mass. One end of the mock-up is clamped, while the other is guided to restrict displacements in the X and Y directions. Additionally, a rod is positioned atop the specimen to limit mass displacements in the Z direction (refer to Figure 1-right). During testing, excitation was applied exclusively in the X direction.

The numerous simulations carried out for this case study were obtained with artificial seismic signals generated with the stochastic generator proposed by Rezaeian and Der Kiureghian [50]. This generator implemented in [7] was calibrated from 97 real accelerograms selected in the European Strong Motion Database for a magnitude M such that $5.5 \leq M \leq 6.5$, and a source-to-site distance $R < 20$ km [51]. Note that enrichment is not a necessity in the Bayesian framework – especially if a sufficient number of real signals is available – but it allows comparative performance studies, such as those presented in this work.

As in practice the piping system is located in a building, the artificial signals were filtered using a fictitious 2% damped linear single-mode building at 5 Hz, which corresponds to the first eigenfrequency of the 1% damped piping system. The chosen failure criterion is based on the assessment of excessive out-of-plane rotation of the elbow near the clamped end of the mock-up, following the recommendation in [52]. The chosen IM is the PSA which is calculated here at 5 Hz for a damping ratio of 1%.

In order to evaluate the effectiveness of the proposed method, we considered the nonlinear seismic behavior of the piping system. Regarding the nonlinear constitutive law of the material, a bilinear law exhibiting kinematic hardening was used to reproduce the overall nonlinear behaviour of the mock-up with satisfactory agreement compared to the results of the seismic tests [49].

In this work, the critical rotation threshold is set at $C = 4.1°$, representing the 90%-level quantile derived from a sample of 10^4 nonlinear numerical simulations.

Finally, the fragility curve that we will call "reference" in the following was obtained by Monte-Carlo averages on clusters of the IM using the K-means algorithm, following the suggestion of Trevlopoulos et al. [29], from the

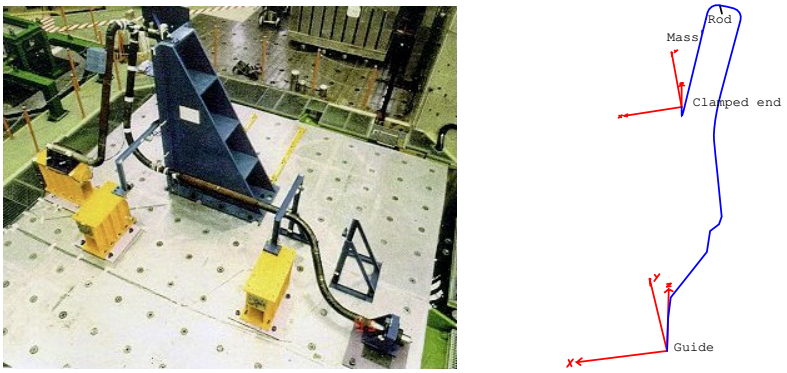

FIGURE 1. (left) Overview of the piping system on the Azalee shaking table and (right) Mock-up Finite Element Model.

10^4 data that we dispose. The estimation procedure, not described here for the sake of brevity, is also presented in the reference [38] where the same system is studied. In this method, the average goes along its confidence intervals; in the computation of the metrics that are suggested in the next section, the average is considered as the reference.

3.2. **Benchmarking metrics**

In order to evaluate the effects that our planning of experiments method has over the fragility curve estimates, we consider four quantitative metrics described hereafter. Those are later implemented on the *a posteriori* estimates conditional to two types of dataset: one derived from our planning of experiments methodology, and one without. Considering a sample $(\mathbf{a}, \mathbf{y})$, we denote by $a \mapsto P_f^{|\mathbf{a},\mathbf{y}}(a)$ the random process defined as the fragility curve conditionally to the sample. $P_f^{|\mathbf{a},\mathbf{y}}(a) = \Phi((\rho \log a + \mu)/\sigma)$ inherits from the *a posteriori* distribution of θ. For each value a the r-quantile of the random variable $P_f^{|\mathbf{a},\mathbf{y}}(a)$ is denoted by $q_r^{|\mathbf{a},\mathbf{y}}(a)$, and its median is denoted by $m^{|\mathbf{a},\mathbf{y}}(a)$. Also, we take into account the reference fragility curve $a \mapsto P_f^{\text{ref}}(a)$, implemented in the same fashion as in [38], and as evoked in section 3.1; and we consider a bounded set $\mathcal{A} = [0, A_{\max}]$ for the IM, the truncation is set to the maximal IM within the database of 10^4 seismic signals we have at disposal for this work, as disclosed in section 3.1. We define:

- The square bias to the median: $\mathcal{B}^{|\mathbf{a},\mathbf{y}} = \|m^{|\mathbf{a},\mathbf{y}} - P_f^{\text{ref}}\|_{\mathrm{L}^2}^2$; where $\|P\|_{\mathrm{L}^2}^2 = \dfrac{1}{A_{\max}} \displaystyle\int_0^{A_{\max}} P(a)^2 \, \mathrm{d}a.$
- The quadratic error: $\mathcal{E}^{|\mathbf{a},\mathbf{y}} = \mathbb{E}\left[\|P_f^{|\mathbf{a},\mathbf{y}} - P_f^{\text{ref}}\|_{\mathrm{L}^2}^2 | \mathbf{a}, \mathbf{y}\right]$.
- The $1-r$-square credibility width: $\mathcal{W}^{|\mathbf{a},\mathbf{y}} = \|q_{1-r/2}^{|\mathbf{a},\mathbf{y}} - q_{r/2}^{|\mathbf{a},\mathbf{y}}\|_{\mathrm{L}^2}^2$.
- The $1-r$-coverage probability: $\mathcal{P}^{|\mathbf{a},\mathbf{y}} = \dfrac{1}{A_{\max}} \displaystyle\int_0^{A_{\max}} \mathbb{1}_{P_f^{\text{ref}}(a) \in \left[q_{1-r/2}^{|\mathbf{a},\mathbf{y}}(a), q_{r/2}^{|\mathbf{a},\mathbf{y}}(a)\right]} \, \mathrm{d}a.$

For the forthcoming implementation of these metrics, numerous *a posteriori* samples of the process $P_f^{|\mathbf{a},\mathbf{y}}$ are generated from their known distribution (see equation (11)) and serve the computation of the medians, quantiles and means through Monte-Carlo derivation. The integrals are approximated numerically from Simpsons' interpolation on sub-intervals of regular size $0 = A_0 < \cdots < A_p = A_{\max}$. In our computations, we use $A_{\max} = 55\,\mathrm{m/s^2}$, and $p = 200$.

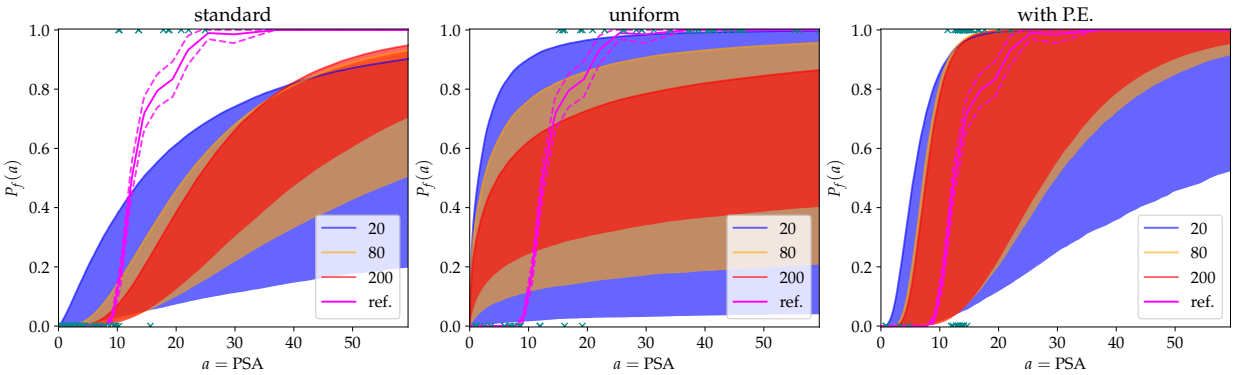

FIGURE 2. Examples of fragility curve estimations for different number of observations, with the PSA considered as IM. Are plotted the 95% credibility interval for 3 datasets of respective sizes 20 (blue), 80 (orange), and 200 (red); the reference curve P_f^{ref} is drawn in magenta and is accompanied with the confidence interval of the procedure in dashed lines. The observations are chosen (a) w.r.t. the standard distribution of the IM; or (b) w.r.t. a uniform distribution on $[0, A_{\max}]$; or (c) using our planning of experiments method. The green crosses represent 50 pairs $(a_i, \mathbb{1}_{y_i>C})$ drawn for each method.

3.3. **Numerical results**

Figure 2 shows examples of fragility curve estimations based on different dataset sizes. The results presented in Figure 2-(c) come from our planning-of-experiments (PE) methodology while the results presented in Figures 2-(a) and 2-(b) come from independent random samples, with IMs that have been drawn w.r.t. their original standard distribution or w.r.t. a uniform distribution. These qualitative results clearly illustrate the contribution of our methodology.

When samples are randomly drawn from the original IM distribution or from a uniform distribution, the results show a rapid convergence of the estimates – in the sense that the associated credibility intervals decrease rapidly – towards biased estimations of the fragility curves. Conversely, when the samples come from the PE methodology, the bias decreases but the convergence is slower.

These observations are confirmed on a larger scale by the results presented in Figures 3 and 4. These ones are issued from computations of the metrics $\mathcal{B}^{|\mathbf{a},\mathbf{y}}$, $\mathcal{E}^{|\mathbf{a},\mathbf{y}}$, $\mathcal{W}^{|\mathbf{a},\mathbf{y}}$ and $\mathcal{P}^{|\mathbf{a},\mathbf{y}}$ described in section 3.2, for various observation sets $(\mathbf{a}, \mathbf{y})$. They compare the performances of our PE methodology with the two methods that are based on independently drawn observations: the one that involves IM samples drawn from their standard distribution, and the one that involves IM samples drawn from a uniform distribution over the range $[0, A_{\max}]$.

Figure 3 shows empirical comparisons of the bias and of the quadratic error between the method involving a design of experiments and without. These two results clearly illustrate that the PE approach outperforms the standard and the uniform approaches. Although the quadratic error is strongly related to the variance of the estimates it is significantly offset by the fact that the bias is smaller with the PE approach. Indeed, figure 4-left illustrates that the 95%-square credibility width is smaller with the standard and uniform approaches than with the PE-based approach. This good result nevertheless masks a lack of robustness of the standard or uniform approaches since the estimate turns out to be strongly biased, as shown in figure 4-right. This figure shows indeed the coverage probability for the both methods, as a function of the dataset size. It measures the average inclusion of the reference fragility curves to the *a posteriori* credibility intervals.

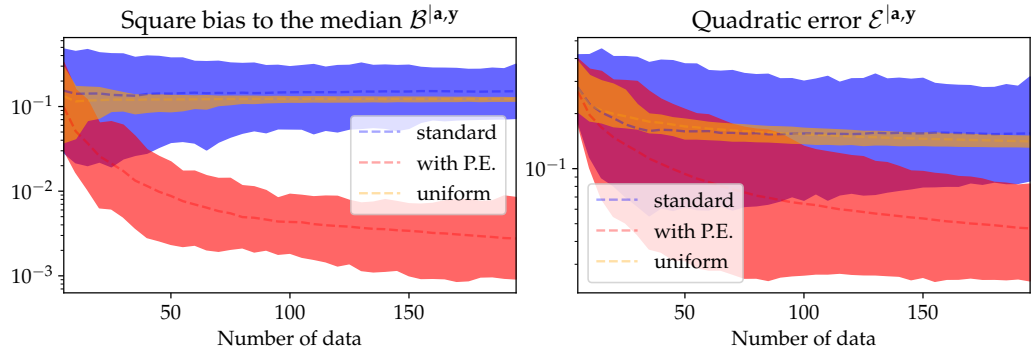

FIGURE 3. Confidence intervals and means w.r.t. the $(\mathbf{a}, \mathbf{y})$ for (left) the square bias to the median $\mathcal{B}^{|\mathbf{a},\mathbf{y}}$ and (right) the quadratic error $\mathcal{E}^{|\mathbf{a},\mathbf{y}}$; as a function of the number of observations. For each value of $k = 5, 10, \ldots, 200$, a number of $L = 200$ dataset have been drawn following the standard distribution of the IM firstly (for the blue curves), following a uniform distribution on $[0, A_{\max}]$ secondly (for the orange curves), and following the planning of experiments method thirdly (for the red curves).

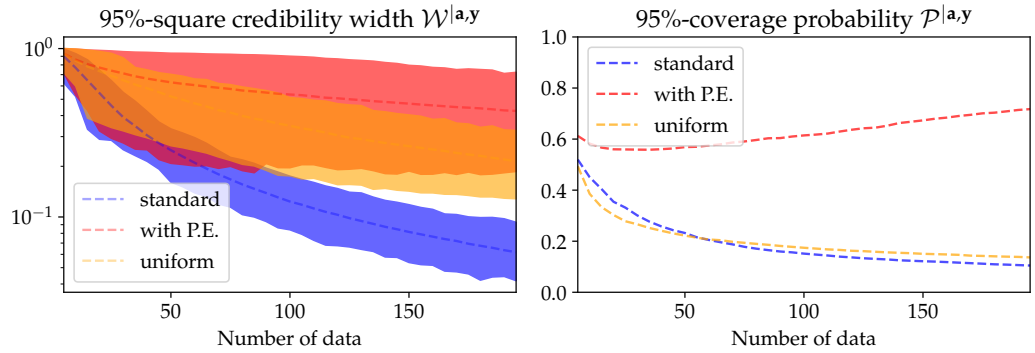

FIGURE 4. (left) 95%-confidence intervals and means w.r.t. $(\mathbf{a}, \mathbf{y})$ for the 95%-square credibility width $\mathcal{W}^{|\mathbf{a},\mathbf{y}}$, as a function of the number of observations. (right) mean w.r.t. $(\mathbf{a}, \mathbf{y})$ for the 95%-coverage probability $\mathcal{P}^{|\mathbf{a},\mathbf{y}}$, as a function of the number of observations. For each value of $k = 5, 10, \ldots, 200$, $L = 200$ datasets have been drawn following the standard distribution of the IM firstly (for the blue curves), following a uniform distribution on $[0, A_{\max}]$ secondly (for the orange curves), and following the planning of experiments method thirdly (for the red curves).

4. DISCUSSION

The results presented in the previous section clearly illustrate the superiority of the PE-based approach over the standard and uniform approaches. Similar results, not presented here for the sake of brevity, and obtained with the PGA as IM, as well as with other types of structures, also confirm these results.

The strength of the approach proposed in this work is its completely analytical nature, which avoids the use of MCMC methods for the *a posteriori* estimation of fragility curves. To do this, however, it is necessary to assume that the logarithm of the EDP evolves linearly as a function of the logarithm of the IM.

So, to solve this problem, we are faced with a contradiction. In order to satisfy the linearity assumption, it is necessary, on the one hand, that the learning zone is local, that is to say restricted to the vicinity of the IMs for which the fragility curve evolves significantly from 0 to 1. On the other hand, the equation (13) shows that an empirical variance of the IMs that is too small is penalizing from the point of view of the variance of the

estimation of the fragility curves. The variance of the latter is in fact inversely proportional to that of the IMs considered for learning.

As the numerical results show, the proposed learning method localizes the learning in the area of interest and significantly reduces the model bias. As a result, this is accompanied by a slight reduction in the size of the credibility interval with the number of training data.

This therefore suggests that the proposed method is effective for samples with limited size (for instance, given the results presented in section 3.3, an appropriate limit could be a sample size smaller than 100 for the case study treated in this paper). Beyond that, given the cost associated with each training data, it seems preferable to move towards less constrained and more sophisticated methods [7, 24, 38].

Conclusion

Assessing the seismic fragility of structures and components when few data are available is a challenging task and the Bayesian framework is known to be effective for these types of problems.

In this work we proposed an efficient Bayesian methodology whose strength lies in its fully analytical nature, which avoids the use of MCMC methods for the *a posteriori* estimation of fragility curves. The effectiveness of the method comes from the assumption of linearity between the logarithm of the EDP and that of the IM of interest. As this hypothesis implies a model bias in most practical cases, we proposed a strategy in order to minimize this bias, by concentrating the learning in the vicinity of the IMs for which the fragility curve evolves significantly from 0 to 1.

The numerical results clearly illustrate the superiority of the proposed approach over an approach without a learning strategy. They emphasize the robustness of a design of experiments which is based on a sensitivity analysis of the posterior distribution. Such construction is not limited to the modeling we derive in this work in particular, and could still be adapted to another to increase its learning abilities. They also suggest that the proposed method is effective for a limited sample size (about 100 in our settings). Beyond that, given the cost associated with each training data, it seems preferable to move towards less constrained and more sophisticated methods, in order to more effectively minimize both the biases and the variance of the estimates.

For practitioners, this method therefore constitutes a rapid and robust tool for first estimates of fragility curves in a context where the datasets are of limited size.

A. Construction of $\mathbf{z}$

Conditionally to $(\mathbf{a}, \theta)$, we derive the distribution of $\hat{\mathbf{y}} - \rho_k \hat{\mathbf{a}}$:

$$\hat{\mathbf{y}} - \rho_k \hat{\mathbf{a}} | \mathbf{a}, \theta \sim \mathcal{N}(\mu \mathbf{1}, \sigma^2 U_{\mathbf{a}}), \quad \text{with} \quad U_{\mathbf{a}} = I - \frac{\hat{\mathbf{a}}(\frac{1}{2}\hat{\mathbf{a}} - \overline{\hat{\mathbf{a}}})^T}{k \operatorname{Var}_k \hat{\mathbf{a}}} - \frac{(\frac{1}{2}\hat{\mathbf{a}} - \overline{\hat{\mathbf{a}}})\hat{\mathbf{a}}^T}{k \operatorname{Var}_k \hat{\mathbf{a}}}, \quad \overline{\hat{\mathbf{a}}} = \frac{1}{k}\sum_{i=1}^{k} \log a_i, \tag{20}$$

where $\mathbf{1}$ denotes the vector of $\mathbb{R}^k$ which contains only ones. Below is suggested a diagonalization of the matrix $U_{\mathbf{a}}$: we define $P_{\mathbf{a}}$ such that $P_{\mathbf{a}}^T P_{\mathbf{a}} = I$ and $P_{\mathbf{a}}^T U_{\mathbf{a}} P_{\mathbf{a}}$ is a diagonal matrix. To define $\mathbf{z}$, we denote by $\tilde{P}_{\mathbf{a}}$ the matrix in $\mathbb{R}^{k\times(k-1)}$ composed by the $k-1$ first columns of $P_{\mathbf{a}}$. In what follows, we assume that $k > 2$ and that the coordinates of $\mathbf{a}$ are not all identical.

The matrix $U_{\mathbf{a}}$ takes the form of $I - uv^T - vu^T$ for some vectors u and v of $\mathbb{R}^k$, which are linearly independent. It is clear that the diagonalization of $U_{\mathbf{a}}$ is linked with the one of $V = uv^T + vu^T$.

First of all notice that $v^\perp$ and $u^\perp$ are two different hyperplanes because of the linear independence of u and v. That makes $u^\perp \cap v^\perp$ a subspace of dimension $k-2$. Therefore, as one would notice that $u^\perp \cap v^\perp \subset \ker V$ and $\operatorname{im} V \subset \operatorname{Span}(u, v)$, the converse inclusions stand.

This way, while 0 is the first eigenvalue of V with rank $k-2$, an other eigenvalue r must admit eigenvectors in $\mathrm{Span}(u,v)$, which should make the system

$$\left\{\begin{array}{ll} r\gamma & = \gamma v^T u + \delta v^T v \\ r\delta & = \gamma u^T u + \delta u^T v \end{array}\right. \tag{21}$$

admitting an infinity of solutions w.r.t. (γ,δ). Equivalently, its determinant must be null which lead to the two solutions

$$r = v^T u \pm \|v\|\|u\|. \tag{22}$$

As u and v are linearly independent, the equation above defines two different eigenvalues r_+ and r_-, both of rank 1. Let r be one of those, a resolution of the equation system (21) gives that the eigenspace associated with r is $\mathrm{Span}(v^T v u + (r - v^T u)v)$.

Coming back to $U_{\mathbf{a}}$, the arguments above show that the eigenvalues of $U_{\mathbf{a}}$ are 1, $1-r_-$ and $1-r_+$ with respective ranks $k-2$, 1 and 1, and with:

$$r_+ = 1\ , \quad r_- = \frac{-k\bar{\hat{\mathbf{a}}}^2}{\|\hat{\mathbf{a}} - \bar{\hat{\mathbf{a}}}\|^2}, \tag{23}$$

because

$$\|\hat{\mathbf{a}}\|^2\|\frac{1}{2}\hat{\mathbf{a}} - \bar{\hat{\mathbf{a}}}\|^2 = \sum_{i=1}^k \hat{a}_i^2 \left(\frac{1}{4}\sum_{i=1}^k \hat{a}_i^2 + \left(\sum_{i=1}^k \hat{a}_i^2\right)\frac{1}{k}\left(\sum_{i=1}^k \hat{a}_i\right)^2 - \left(\sum_{i=1}^k \hat{a}_i^2\right)\frac{1}{k}\left(\sum_{i=1}^k \hat{a}_i\right)^2\right) = \frac{1}{4}\left(\sum_{i=1}^k \hat{a}_i^2\right)^2. \tag{24}$$

Now, let us choose $w_1,\dots,w_{k-2}$ an orthonormal basis of $\hat{\mathbf{a}}^\perp \cap (\frac{1}{2}\hat{\mathbf{a}} - \bar{\hat{\mathbf{a}}})^\perp$. Let us define

$$w_{k-1} = \frac{1}{\sqrt{k}}\mathbf{1}, \quad w_k = \frac{\hat{\mathbf{a}} - \bar{\hat{\mathbf{a}}}}{\|\hat{\mathbf{a}} - \bar{\hat{\mathbf{a}}}\|}. \tag{25}$$

We remind $\mathbf{1}$ is the vector whose coordinates are ones. Therefore, denoting $P_{\mathbf{a}}$ the matrix whose columns are the w_i, $i=1,\dots,k$; it comes $P_{\mathbf{a}}^T P_{\mathbf{a}} = I$ and

$$P_{\mathbf{a}}^T U_{\mathbf{a}} P_{\mathbf{a}} = \mathrm{diag}\left(1,\dots,1,\frac{\hat{\mathbf{a}}^T\hat{\mathbf{a}}}{\|\hat{\mathbf{a}} - \bar{\hat{\mathbf{a}}}\|^2},0\right). \tag{26}$$

The complete definition of $P_{\mathbf{a}}$ depends on the chosen orthonormal basis $w_1,\dots,w_{k-2}$ of $\hat{\mathbf{a}}^\perp \cap (\frac{1}{2}\hat{\mathbf{a}} - \bar{\hat{\mathbf{a}}})^\perp$. In practice, we proceed as follows to construct it, starting from $w_{k-1} = (w_{k-1}^{(j)})_{j=1}^k$ and $w_k = (w_k^{(j)})_{j=1}^k$ as defined in equation (25). We denote by e_i the canonical vectors of $\mathbb{R}^k$ (the jth coordinate of e_i is equal to 0 iff $j \neq i$). As the coordinates of $\mathbf{a}$ are not all the sames, there exist j,p such that $w_k^{(j)} \neq w_k^{(p)}$ (in practice, we select the minimal j and the minimal p such that this property is verified). Thus, we can show that the vectors $w_k, w_{k-1}, (e_i)_{i\neq j,p}$ form a basis of $\mathbb{R}^k$ by computing their determinant:

$$\det(w_k, w_{k-1}, (e_i)_{i\neq j,p}) = w_k^{(j)} w_{k-1}^{(p)}(-1)^{j+p+1} - w_{k-1}^{(j)} w_k^{(p)}(-1)^{j+p+1} \neq 0, \tag{27}$$

as $w_{k-1}^{(j)} = w_{k-1}^{(p)}$. Eventually, the family of the w_i, $i=1,\dots k$ is the result of the Gram-Schmidt process applied to $u_k,\dots,u_1 = w_k, w_{k-1}, (e_i)_{i\neq j,p}$:

$$w_{k-i} = \frac{\tilde{w}_{k-i}}{\|\tilde{w}_{k-i}\|}, \qquad \tilde{w}_{k-i} = u_{k-i} - \sum_{j<i} w_{k-j}^T u_{k-i} \cdot w_{k-j}, \tag{28}$$

Note that this process leaves the expressions of w_k and w_{k-1} unchanged. The family $w_1, \dots, w_{k-2}$ thus forms a basis of $\text{Span}(w_k, w_{k-1})^\perp = \hat{\mathbf{a}}^\perp \cap (\frac{1}{2}\hat{\mathbf{a}} - \bar{\hat{\mathbf{a}}})^\perp$.

We invite the reader to notice that in any way, the construction of those $k-2$ first columns of $P_{\mathbf{a}}$ has no influence on the resulting posterior distribution of interest (given by equation (11)). Indeed, that latter only involves the expressions of z_{k-1} and of $\sum_{i=1}^{k-2} z_i^2$. Concerning the first one, it is equal to $w_{k-1}^T(\hat{\mathbf{y}} - \rho_k \hat{\mathbf{a}})$, and concerning the second one, it is equal to:

$$\begin{aligned}\sum_{i=1}^{k-2} z_i^2 = \sum_{i=1}^{k-2} |w_i^T(\hat{\mathbf{y}} - \rho_k\hat{\mathbf{a}})|^2 &= \sum_{i=1}^{k} |w_i^T(\hat{\mathbf{y}} - \rho_k\hat{\mathbf{a}})|^2 - |w_{k-1}^T(\hat{\mathbf{y}} - \rho_k\hat{\mathbf{a}})|^2 - |w_k^T(\hat{\mathbf{y}} - \rho_k\hat{\mathbf{a}})|^2 \\ &= \|\hat{\mathbf{y}} - \rho_k\hat{\mathbf{a}}\|^2 - |w_{k-1}^T(\hat{\mathbf{y}} - \rho_k\hat{\mathbf{a}})|^2 - |w_k^T(\hat{\mathbf{y}} - \rho_k\hat{\mathbf{a}})|^2.\end{aligned}$$

Acknowledgements. This research was supported by the CEA (French Alternative Energies and Atomic Energy Commission) and the SEISM Institute (www.institut-seism.fr/en/).

References

[1] Robert P. Kennedy, C. Allin Cornell, Robert D. Campbell, Stan J. Kaplan, and Harold F. Perla. Probabilistic seismic safety study of an existing nuclear power plant. *Nuclear Engineering and Design*, 59(2):315–338, 1980. doi:10.1016/0029-5493(80)90203-4.

[2] Robert P. Kennedy and Mayasandra K. Ravindra. Seismic fragilities for nuclear power plant risk studies. *Nuclear Engineering and Design*, 79(1):47–68, 1984. doi:10.1016/0029-5493(84)90188-2.

[3] Young-Ji Park, Charles H. Hofmayer, and Nilesh C. Chokshi. Survey of seismic fragilities used in PRA studies of nuclear power plants. *Reliability Engineering & System Safety*, 62(3):185–195, 1998. doi:10.1016/S0951-8320(98)00019-2.

[4] Robert P. Kennedy. Risk based seismic design criteria. *Nuclear Engineering and Design*, 192(2):117–135, 1999. doi:10.1016/S0029-5493(99)00102-8.

[5] C. Allin Cornell. Hazard, ground motions and probabilistic assessments for PBSD. In *Proceedings of the International Workshop on Performance-Based Seismic Design - Concepts and Implementation*, pages 39–52, University of California, Berkeley, 2004. PEER Center.

[6] Matteo Ciano, Massimiliano Gioffrè, and Mircea Grigoriu. The role of intensity measures on the accuracy of seismic fragilities. *Probabilistic Engineering Mechanics*, 60:103041, 2020. doi:10.1016/j.probengmech.2020.103041.

[7] Rémi Sainct, Cyril Feau, Jean-Marc Martinez, and Josselin Garnier. Efficient methodology for seismic fragility curves estimation by active learning on support vector machines. *Structural Safety*, 86:101972, 2020. doi:10.1016/j.strusafe.2020.101972.

[8] Matteo Ciano, Massimiliano Gioffrè, and Mircea Grigoriu. A novel approach to improve accuracy in seismic fragility analysis: The modified intensity measure method. *Probabilistic Engineering Mechanics*, 69:103301, 2022. doi:10.1016/j.probengmech.2022.103301.

[9] Irmela Zentner, Max Gündel, and Nicolas Bonfils. Fragility analysis methods: Review of existing approaches and application. *Nuclear Engineering and Design*, 323:245–258, 2017. doi:10.1016/j.nucengdes.2016.12.021.

[10] Paolo Gardoni, Armen Der Kiureghian, and Khalid M. Mosalam. Probabilistic capacity models and fragility estimates for reinforced concrete columns based on experimental observations. *Journal of Engineering Mechanics*, 128 (10):1024–1038, 2002. doi:10.1061/(ASCE)0733-9399(2002)128:10(1024).

[11] Do-Eun Choe, Paolo Gardoni, and David Rosowsky. Closed-form fragility estimates, parameter sensitivity, and bayesian updating for rc columns. *Journal of Engineering Mechanics*, 133(7):833–843, 2007. doi:10.1061/(ASCE)0733-9399(2007)133:7(833).

[12] Masanobu Shinozuka, Maria Qing Feng, Jongheon Lee, and Toshihiko Naganuma. Statistical analysis of fragility curves. *Journal of Engineering Mechanics*, 126(12):1224–1231, 2000. doi:10.1061/(ASCE)0733-9399(2000)126:12(1224).

[13] David Lallemant, Anne Kiremidjian, and Henry Burton. Statistical procedures for developing earthquake damage fragility curves. *Earthquake Engineering & Structural Dynamics*, 44(9):1373–1389, 2015. doi:10.1002/eqe.2522.

[14] Daniel Straub and Armen Der Kiureghian. Improved seismic fragility modeling from empirical data. *Structural Safety*, 30(4):320–336, 2008. doi:10.1016/j.strusafe.2007.05.004.

[15] Irmela Zentner. Numerical computation of fragility curves for NPP equipment. *Nuclear Engineering and Design*, 240(6):1614–1621, 2010. doi:10.1016/j.nucengdes.2010.02.030.

[16] Fan Wang and Cyril Feau. Influence of Input Motion's Control Point Location in Nonlinear SSI Analysis of Equipment Seismic Fragilities: Case Study on the Kashiwazaki-Kariwa NPP. *Pure and Applied Geophysics*, 2020. doi:10.1016/j.engstruct.2018.02.024.

[17] Tushar K. Mandal, Siddhartha Ghosh, and Nikil N. Pujari. Seismic fragility analysis of a typical indian PHWR containment: Comparison of fragility models. *Structural Safety*, 58:11–19, 2016. doi:10.1016/j.strusafe.2015.08.003.

[18] Zhiyi Wang, Nicola Pedroni, Irmela Zentner, and Enrico Zio. Seismic fragility analysis with artificial neural networks: Application to nuclear power plant equipment. *Engineering Structures*, 162:213–225, 2018. doi:10.1016/j.engstruct.2018.02.024.

[19] Zhiyi Wang, Irmela Zentner, and Enrico Zio. A Bayesian framework for estimating fragility curves based on seismic damage data and numerical simulations by adaptive neural networks. *Nuclear Engineering and Design*, 338:232–246, 2018. doi:10.1016/j.nucengdes.2018.08.016.

[20] Chnfeng Zhao, Na Yu, and Yilung Mo. Seismic fragility analysis of AP1000 SB considering fluid-structure interaction effects. *Structures*, 23:103–110, 2020. doi:10.1016/j.istruc.2019.11.003.

[21] Joonam Park and Peeranan Towashiraporn. Rapid seismic damage assessment of railway bridges using the response-surface statistical model. *Structural Safety*, 47:1–12, 2014. doi:10.1016/j.strusafe.2013.10.001.

[22] Junwon Seo and Daniel G. Linzell. Use of response surface metamodels to generate system level fragilities for existing curved steel bridges. *Engineering Structures*, 52:642–653, 2013. doi:10.1016/j.engstruct.2013.03.023.

[23] Ioannis Gidaris, Alexandros A. Taflanidis, and George P. Mavroeidis. Kriging metamodeling in seismic risk assessment based on stochastic ground motion models. *Earthquake Engineering & Structural Dynamics*, 44(14):2377–2399, 2015. doi:10.1002/eqe.2586.

[24] Clement Gauchy, Cyril Feau, and Josselin Garnier. Uncertainty quantification and global sensitivity analysis of seismic fragility curves using kriging. *International Journal for Uncertainty Quantification*, 14(4):39–63, 2024. doi:10.1615/Int.J.UncertaintyQuantification.2023046480.

[25] Mohammad Amin Hariri-Ardebili and Victor E. Saouma. Probabilistic seismic demand model and optimal intensity measure for concrete dams. *Structural Safety*, 59:67–85, 2016. doi:10.1016/j.strusafe.2015.12.001.

[26] Bruce R. Ellingwood. Earthquake risk assessment of building structures. *Reliability Engineering & System Safety*, 74(3):251–262, 2001. doi:10.1016/S0951-8320(01)00105-3.

[27] Sang-Hoon Kim and Masanobu Shinozuka. Development of fragility curves of bridges retrofitted by column jacketing. *Probabilistic Engineering Mechanics*, 19(1):105–112, 2004. Fourth International Conference on Computational Stochastic Mechanics. doi:10.1016/j.probengmech.2003.11.009.

[28] Chu Mai, Katerina Konakli, and Bruno Sudret. Seismic fragility curves for structures using non-parametric representations. *Frontiers of Structural and Civil Engineering*, 11(2):169–186, 2017. doi:10.1007/s11709-017-0385-y.

[29] Konstantinos Trevlopoulos, Cyril Feau, and Irmela Zentner. Parametric models averaging for optimized non-parametric fragility curve estimation based on intensity measure data clustering. *Structural Safety*, 81:101865, 2019. doi:10.1016/j.strusafe.2019.05.002.

[30] Yoshifumi Katayama, Yasuki Ohtori, Toshiaki Sakai, and Hitoshi Muta. Bayesian-estimation-based method for generating fragility curves for high-fidelity seismic probability risk assessment. *Journal of Nuclear Science and Technology*, 58(11):1220–1234, 2021. doi:10.1080/00223131.2021.1931517.

[31] P. Steve Koutsourelakis. Assessing structural vulnerability against earthquakes using multi-dimensional fragility surfaces: A Bayesian framework. *Probabilistic Engineering Mechanics*, 25(1):49–60, 2010. doi:10.1016/j.probengmech.2009.05.005.

[32] Guillaume Damblin, Merlin Keller, Alberto Pasanisi, Pierre Barbillon, and Éric Parent. Approche décisionnelle bayésienne pour estimer une courbe de fragilité. *Journal de la Societe Française de Statistique*, 155(3):78–103, 2014.

[33] Sashi Kanth Tadinada and Abhinav Gupta. Structural fragility of t-joint connections in large-scale piping systems using equivalent elastic time-history simulations. *Structural Safety*, 65:49–59, 2017. doi:10.1016/j.strusafe.2016.12.003.

[34] Shinyoung Kwag and Abhinav Gupta. Computationally efficient fragility assessment using equivalent elastic limit state and Bayesian updating. *Computers & Structures*, 197:1–11, 2018. doi:10.1016/j.compstruc.2017.11.011.

[35] Jong-Su Jeon, Sujith Mangalathu, Junho Song, and Reginald Desroches. Parameterized seismic fragility curves for curved multi-frame concrete box-girder bridges using bayesian parameter estimation. *Journal of Earthquake Engineering*, 23(6):954–979, 2019. doi:10.1080/13632469.2017.1342291.

[36] Armin Tabandeh, Pouyan Asem, and Paolo Gardoni. Physics-based probabilistic models: Integrating differential equations and observational data. *Structural Safety*, 87:101981, 2020. doi:10.1016/j.strusafe.2020.101981.

[37] Sangwoo Lee, Shinyoung Kwag, and Bu-seog Ju. On efficient seismic fragility assessment using sequential bayesian inference and truncation scheme: A case study of shear wall structure. *Computers & Structures*, 289:107150, 2023. doi:10.1016/j.compstruc.2023.107150.

[38] Antoine Van Biesbroeck, Clément Gauchy, Cyril Feau, and Josselin Garnier. Reference prior for Bayesian estimation of seismic fragility curves. *Probabilistic Engineering Mechanics*, 76:103622, 2024. doi:10.1016/j.probengmech.2024.103622.

[39] Antoine Van Biesbroeck, Clément Gauchy, Cyril Feau, and Josselin Garnier. Influence of the choice of the seismic intensity measure on fragility curves estimation in a bayesian framework based on reference prior. In *Proceedings of the 5th Thematic Conference on Uncertainty Quantification in Computational Sciences and Engineering*. ECCOMAS, 2023. doi:10.7712/120223.10327.19899.

[40] Robert E. Kass and Larry Wasserman. The selection of prior distributions by formal rules. *Journal of the American Statistical Association*, 91(435):1343–1370, 1996. doi:10.1080/01621459.1996.10477003.

[41] James O. Berger, José M. Bernardo, and Dongchu Sun. The formal definition of reference priors. *The Annals of statistics*, 37(2):905–938, 2009. doi:10.1214/07-AOS587.

[42] Swarup Ghosh and Subrata Chakraborty. Seismic fragility analysis of structures based on bayesian linear regression demand models. *Probabilistic Engineering Mechanics*, 61:103081, 2020. doi:10.1016/j.probengmech.2020.103081.

[43] Antoine Van Biesbroeck. Generalized mutual information and their reference priors under Csizar f-divergence. 2023. arXiv.2310.10530. doi:10.48550/arXiv.2310.10530.

[44] Sébastien Da Veiga. Global sensitivity analysis with dependence measures. *Journal of Statistical Computation and Simulation*, 85(7):1283–1305, 2015. doi:10.1080/00949655.2014.945932.

[45] Bertrand Iooss and Paul Lemaître. *A Review on Global Sensitivity Analysis Methods*, pages 101–122. Springer US, Boston, MA, 2015. ISBN 978-1-4899-7547-8. doi:10.1007/978-1-4899-7547-8_5.

[46] Ilya M. Sobol'. Sensitivity estimates for non linear mathematical models. *Mathematical Modeling and Computer Experiments*, 1:407–414, 1993.

[47] Sébastien Da Veiga, Fabrice Gamboa, Bertrand Iooss, and Clémentine Prieur. *Basics and Trends in Sensitivity Analysis*. Computational Science & Engineering. Society for Industrial and Applied Mathematics, Philadelphia, PA, 1 edition, 2021. ISBN 978-1-61197-668-7. doi:10.1137/1.9781611976694.

[48] CEA. CAST3M, 2019. URL http://www-cast3m.cea.fr/.

[49] Françoise Touboul, Pierre Sollogoub, and Nadine Blay. Seismic behaviour of piping systems with and without defects: experimental and numerical evaluations. *Nuclear Engineering and Design*, 192(2):243–260, 1999. doi:10.1016/S0029-5493(99)00111-9.

[50] Sanaz Rezaeian and Armen Der Kiureghian. Simulation of synthetic ground motions for specified earthquake and site characteristics. *Earthquake Engineering & Structural Dynamics*, 39(10):1155–1180, 2010. doi:10.1002/eqe.997.

[51] Nicholas N. Ambraseys, Patrick Smit, Raniero Berardi, Dario Rinaldis, Fabrice Cotton, and Catherine Berge. Dissemination of european strongmotion data, 2000. CD-ROM collection. European Commission, Directorate-General XII, Environmental and Climate Programme, ENV4-CT97-0397, Brussels, Belgium.

[52] Françoise Touboul, Nadine Blay, Pierre Sollogoub, and Stéphane Chapuliot. Enhanced seismic criteria for piping. *Nuclear Engineering and Design*, 236(1):1–9, 2006. doi:10.1016/j.nucengdes.2005.07.002.