



**HAL**  
open science

# Optimized non-parametric fragility curve estimation based on intensity measure data clustering and parametric model averaging

K. Trevelopoulos, C. Feau, I. Zentner

## ► To cite this version:

K. Trevelopoulos, C. Feau, I. Zentner. Optimized non-parametric fragility curve estimation based on intensity measure data clustering and parametric model averaging. 16Th European Conference On Earthquake Engineering, Jun 2018, Thessalonique, Greece. cea-02339334

**HAL Id: cea-02339334**

**<https://cea.hal.science/cea-02339334>**

Submitted on 15 Dec 2019

**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.

1 **Parametric models averaging for optimized non-parametric fragility curve estimation**  
2 **based on intensity measure data clustering**

3

4 Konstantinos Trevlopoulos<sup>a1</sup>, Cyril Feau<sup>a</sup>, Irmela Zentner<sup>b,c</sup>

5 <sup>a</sup>DEN - Service d'études mécaniques et thermiques (SEMT), CEA, Université Paris-Saclay,  
6 F-91191 Gif-sur-Yvette, France

7 <sup>b</sup>EDF R&D, EDF Lab Paris-Saclay, 7 Bvd Gaspard Monge, 91120 Palaiseau, France

8 <sup>c</sup>IMSIA, UMR CNRS-EDF-CEA-ENSTA ParisTech, Université Paris-Saclay, France

9 **ABSTRACT**

10 Seismic fragility curves give the probability of exceedance of the threshold of a damage state  
11 of a structure, or a non-structural component, conditioned on the intensity measure of the  
12 seismic motion. Typically, fragility curves are constructed parametrically assuming a  
13 lognormal shape. In some cases, which cannot be identified a priori, differences may be  
14 observed between non-parametric fragility curves, evaluated empirically based on a large  
15 number of seismic response analyses, and their estimations via the lognormal assumption.  
16 Here, we present an optimized Monte Carlo procedure for derivation of non-parametric fragility  
17 curves. This procedure uses clustering of the intensity measure data to construct the non-  
18 parametric curve and parametric models averaging for optimized assessment. In simplified  
19 case studies presented here as illustrative applications, the developed procedure leads to a  
20 fragility curve with reduced bias compared to the lognormal curve and to reduced confidence  
21 intervals compared to an un-optimized Monte Carlo-based approach. In the studied cases,  
22 this procedure proved to be efficient providing reasonable estimations even with as few as  
23 100 seismic response analyses.

---

<sup>1</sup>Corresponding author. E-mail address: konstantinos.trevlopoulos@cea.fr. Present address: French Alternative Energies and Atomic Energy Commission (CEA), DEN, 13108 Saint-Paul-lez-Durance CEDEX, France

24

25 **Keywords:** Seismic fragility curve; Non-parametric curve; Optimization; Data Clustering;  
26 Parametric Models Averaging

## 27 1 INTRODUCTION

28 A multitude of procedures is now available for probabilistic seismic assessment of structures  
29 [1]. Most notable is the framework by Yang et al. [2], which was the basis for the FEMA P-58  
30 guidelines [3]. Here, we focus on fragility curves giving the probability to exceed a damage  
31 state threshold conditioned on a measure of the intensity of the seismic motion, such as the  
32 fragility curves defined in [4]. Such fragility curves are used for probabilistic assessment of  
33 seismic risk [5] for structures and non-structural components in nuclear installations [6] and  
34 critical civil infrastructure, such as hospitals and ports of major urban areas in earthquake  
35 prone regions [7]. They can also be used to evaluate the impact of construction details on the  
36 structural performance of installations under seismic excitations [8–11] and in rapid response  
37 applications for risk management during a seismic crisis [12]. The use of fragility curves is not  
38 limited to earthquake-related problems, they are also used in the case of other types of loading  
39 such as wind [13].

40 The classical formulation of a fragility curve makes the hypothesis that the curve  
41 follows a lognormal shape. D’Ayala et al. [14] and FEMA [3] describe a series of procedures  
42 for analytical fragility curve estimation, which are commonly applied. Analytical fragility curve  
43 estimation is based on Engineering Demand Parameter (EDP) observations as a function of  
44 the Intensity Measure (IM). In order to obtain such observations, either cloud analysis,  
45 Incremental Dynamic Analysis (IDA) [15] or Multiple Stripes Analysis (MSA) [16] may be  
46 performed. Linear regression [17] is a common method for lognormal fragility curve estimation.  
47 The most well established methods for adjusting a lognormal fragility curve to observations  
48 from IDA or MSA were developed by Baker [4] and are based on the method of moments and  
49 Maximum Likelihood Estimation (MLE), respectively.

50           However, Mai et al. [18] observed differences between non-parametric fragility curves  
51 based on kernel density estimation and lognormal fragility curves according to different  
52 procedures and highlighted the effect of the derivation procedure on lognormal fragility curves.  
53 Noh et al. [19] also used kernel smoothing in order to construct non-parametric fragility curves  
54 showing that this can be an efficient solution when using sparse data. Lallement et al. [20]  
55 consider non-parametric fragility curves more truthful representations of observations and  
56 construct curves based on generalized additive models and Gaussian kernel smoothing.  
57 Furthermore, in [21], lognormal fragility curves for structural components did not represent  
58 effectively observations from simulations of the seismic response of a bridge.

59           The simplest construction of a non-parametric curve puts the EDP observations in bins  
60 according to the corresponding IM and estimates empirically the probability of exceeding the  
61 damage state threshold for every bin [22]. In practice, due to the prohibitive computational  
62 cost of most nonlinear mechanical models, the development of numerically efficient methods  
63 is required to evaluate such curves using a minimal number of computations.

64           Here, we propose a procedure based on Monte-Carlo (MC) simulations, which uses  
65 Parametric Models Averaging (PMA) in order to optimize the computation of non-parametric  
66 fragility curves, which are constructed based on k-means clustering [23] of the intensity  
67 measure data. Optimization is employed in order to obtain reduced confidence fragility curve  
68 intervals compared to the empirical estimations with an un-optimized MC approach. The key  
69 elements of the optimization are: (i) the EDP observations are computed with seismic  
70 response analyses using synthetic accelerograms, which are realizations of stochastic  
71 processes, (ii) the non-parametric fragility curve is expressed through the law of total  
72 probability as the weighted average of parametric fragility curves, each one of which is  
73 estimated based on the synthetic ground motions generated by a stochastic process. In the  
74 optimized approach, two alternative parametric models per process are proposed for the  
75 probability of exceedance of the damage state threshold. Finally, the range of applicability of  
76 each parametric model per process is analyzed.

77 To illustrate the proposed methodology, “simple” stochastic processes are defined  
78 generating synthetic accelerograms based on original seed acceleration records (Section 2).  
79 The generation results in a set of synthetic accelerograms reproducing the ground motion  
80 variability observed in the original set of ground motion records. The procedure for selection  
81 of the original seed records defining the processes is out of the scope of this work. Here, for  
82 simplicity, the initial set of ground motions are selected using magnitude and distance as  
83 criteria.

84 Here, the non-parametric fragility curves are estimated using as IM the Peak Ground  
85 Acceleration (PGA) or the spectral acceleration at the frequency of an oscillator. However, the  
86 developed procedure is independent of the selected IM. In the studied cases, the 95 %  
87 confidence interval (CI) of the estimated fragility curves is significantly reduced due to the  
88 optimization. Moreover, the bias of the fragility curves according to the optimization is tolerable  
89 or negligible with respect to the reference curve obtained with a very large number of  
90 observations, as long as the applicability recommendations are respected.

## 91 **2 SYNTHETIC GROUND MOTION GENERATION**

### 92 **2.1 Motivation**

93 Here, synthetic ground motions are employed in order to cover the range of IMs of  
94 interest and eventually obtain fragility curves based on IM clustering that are well discretized.  
95 Moreover, synthetic ground motions are used in order to exploit the statistical characteristics  
96 of the ground motions given by a process, such as the distribution of the IMs of the generated  
97 motions, in the context of the optimization of the computation of non-parametric fragility  
98 curves. A "simple" synthetic ground motion generator is developed, which reproduces the  
99 spectral variability of recorded accelerograms, because no hypothesis is introduced  
100 concerning their frequency content. Moreover, the original recorded accelerograms are  
101 selected from the European Strong Motion Database [24,25] using simple criteria, i.e.  $5.5 < M$   
102  $< 6.5$  and  $R < 20$  km. Selection of the original ground motions is out of the scope of this study.

103 It is worth noting that the main idea in the PMA methodology is that the synthetic ground  
 104 motion database consists of realizations of several stochastic processes. Therefore the  
 105 methodology herein could be used theoretically in conjunction with other procedures for  
 106 synthetic ground motion generation defining stochastic processes, such as the model in  
 107 Rezaeian and Der Kiureghian [26]. A study of the effect of the ground motion generator is out  
 108 of the scope of this article. As far as the most appropriate generator is concerned, that depends  
 109 on the problem at hand and the available data (e.g. response spectra, acceleration records,  
 110 see [1]).

## 111 2.2 Synthetic Ground Motion Generation Process

112 The generation process in this framework begins with retaining the FFT amplitude of  
 113 every real record in the original data set, replacing the phases with a vector of uniformly  
 114 distributed random values, computing the new ground motion via inverse FFT and imposing a  
 115 modulation function. The result is a series of unadjusted synthetic ground motions, which are  
 116 subsequently adjusted so that they are on average “spectrally equivalent” with the ground  
 117 motion records in the sense of acceleration response spectra. The  $i$ -th accelerogram ( $i = \{1,$   
 118  $\dots, N_r\}$ ) in a data set of  $N_r$  ground motion records may be expressed with Equation 1. The  
 119 amplitudes ( $A_{r,im}$ ) of the  $i$ -th real record ( $\alpha_{r,i}(t)$ ) are computed with the FFT algorithm and are  
 120 used in combination with random phase ( $\varphi_{s,ijm}$ ) in order to compute the  $j$ -th realization of a  
 121 stationary Gaussian process (Equation 2).

122

$$123 \quad \alpha_{r,i}(t) = \sum_{m=1}^n \left( A_{r,im} \sin(\omega_m + \varphi_{r,im}) \right)$$

$$124 \quad i = \{1, \dots, N_r\} \tag{1}$$

125

$$126 \quad \alpha_{s,ij}(t) = \sum_{m=1}^n \left( A_{r,im} \sin(\omega_m + \varphi_{s,ijm}) \right)$$

$$127 \quad i = \{1, \dots, N_r\}, \quad j = \{1, \dots, N_s\} \tag{2}$$

128

129 where  $\varphi_{s,jm}$  is the phase which is assumed to be a random variable with a uniform distribution  
130  $U(-\pi,\pi)$  according to Boore [27], and  $\omega_m$  is the m-th discrete angular frequency. The  $N_r$   
131 stationary Gaussian processes are converted to non-stationary processes using  $N_r$  modulation  
132 functions. Here the function by Housner and Jennings [28] (Equation 3) is used, however other  
133 modulation functions, e.g. [29], may be considered.

134

$$135 \quad q_i(t) = \begin{cases} \left(\frac{t}{T_{1,i}}\right)^3 & 0 \leq t \leq T_{1,i} \\ 1.0 & T_{1,i} < t \leq T_{2,i} \\ e^{-(t-T_{2,i})} & T_{2,i} < t \leq t_{d,i} \end{cases}$$

136  $i = \{1, \dots, N_r\}$  (3)

137

138 where  $T_{1,i}$  and  $T_{2,i}$  are the times defining the plateau of this modulation function and  $t_{d,i}$  is the  
139 total duration of the i-th acceleration record. Here,  $T_{1,i}$  and  $T_{2,i}$  are set equal to the times of  
140 observation of the 5 % and 95 % of the Arias intensity in the original acceleration record. The  
141 Arias intensity ( $I_{r,i}$ ) of the i-th acceleration record is given by Equation 4.

142

$$143 \quad I_{r,i} = \frac{\pi}{2g} \int_0^{t_{d,i}} \alpha_{r,i}^2(t) dt, \quad i = \{1, \dots, N_r\} \quad (4)$$

144

145  $T_{1,i}$  and  $T_{2,i}$  are computed with Equations 5 and 6. As an example, Figure 1a shows the  
146 modulation function used for the synthetic ground motions based on real record No. 11.

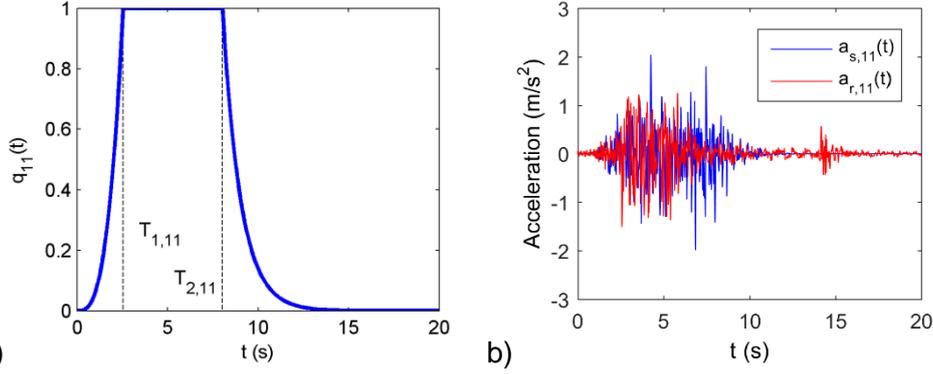
147

$$148 \quad \frac{\pi}{2g} \int_0^{T_{1,i}} \alpha_{r,i}^2(t) dt = 0.05 \cdot I_{r,i} \quad i = \{1, \dots, N_r\} \quad (5)$$

149

$$150 \quad \frac{\pi}{2g} \int_0^{T_{2,i}} \alpha_{r,i}^2(t) dt = 0.95 \cdot I_{r,i} \quad i = \{1, \dots, N_r\} \quad (6)$$

151



152

153 **Figure 1 a) Modulation function b) synthetic accelerogram and its original acceleration**  
 154 **record**

155

156 The  $j$ -th realization of an unadjusted synthetic accelerogram ( $\alpha_{s0,ij}(t)$ ) based on the  $i$ -  
 157 th acceleration record is given by Equation 7.

158

$$159 \quad \alpha_{s0,ij}(t) = q_i(t) \cdot \sum_{m=1}^n \left( A_{r,im} \cdot \sin(\omega_m + \varphi_{s,ijm}) \right)$$

$$160 \quad i = \{1, \dots, N_r\}, \quad j = \{1, \dots, N_s\}$$

161

(7)

162

163 Subsequently, the synthetic ground motions generated based on an acceleration  
 164 record are all scaled with the same scaling factor ( $c_i$ ), which minimizes the sum of the squares  
 165 of the differences between the acceleration response spectrum of the acceleration record for  
 166 5 % damping ( $S_{a,r,i}(f)$ ) and the median spectrum for 5 % damping of the scaled synthetic ground  
 167 motions ( $c \cdot \bar{S}_{a,s0,i}(f)$ ) over the frequencies between 0.2 and 25 Hz (Equation 8). The adjusted  
 168 synthetic ground motions ( $\alpha_{s,ij}(t)$ ) are given by Equation 9. As an example, Figure 1b shows  
 169 record No. 11 and one of its spectrally equivalent synthetic accelerograms. Figure 2 shows  
 170 the acceleration response spectrum of ground motion record No. 11, the spectra of all  
 171 synthetic ground motions generated based on this record and the median spectrum of the  
 172 synthetics ( $c_{11} \cdot \bar{S}_{a,s0,11}(f)$ ).

173

174 
$$c_i = \underset{(c)}{\operatorname{arg\,min}} \left( \sum_{f=0.2 \text{ Hz}}^{f=25 \text{ Hz}} (S_{a,r,i}(f) - c \cdot \bar{S}_{a,s0,i}(f))^2 \right)$$

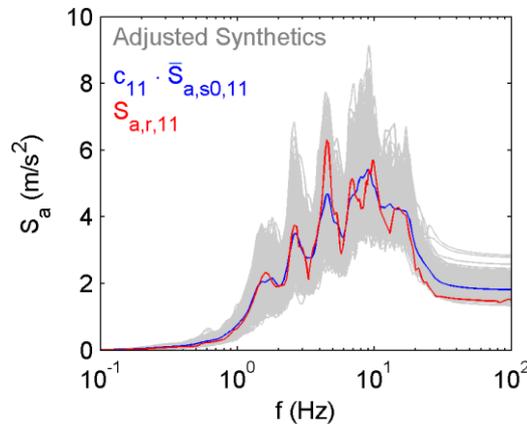
175 
$$i = \{1, \dots, N_r\}$$
 (8)

176

177 
$$\alpha_{s,ij}(t) = c_i \cdot q_i(t) \cdot \sum_{m=1}^n (A_{r,im} \cdot \sin(\omega_m + \varphi_{s,ijm}))$$

178 
$$i = \{1, \dots, N_r\}, \quad j = \{1, \dots, N_s\}$$
 (9)

179



180

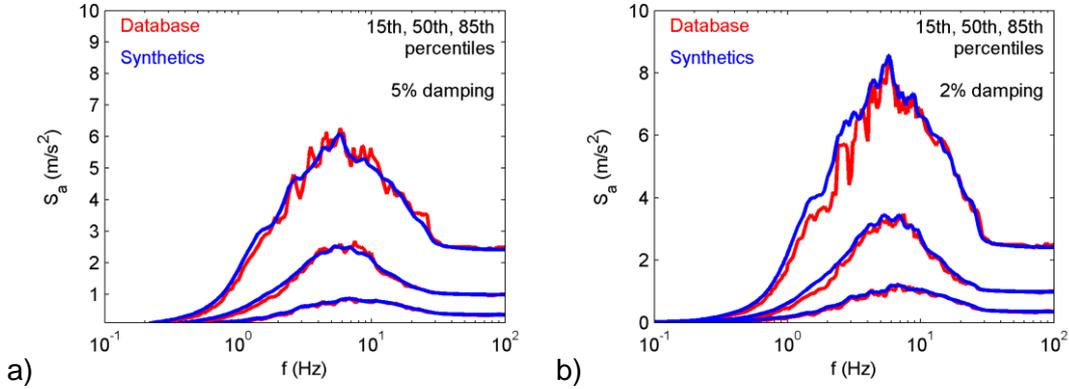
181 **Figure 2 Acceleration response spectra for 5 % damping of the adjusted synthetic**  
 182 **ground motions and their original ground motion**

183

184

185 Based on  $N_r = 96$  original acceleration records, a total of  $N_r \times N_s = 48000$  “spectrally  
 186 equivalent” synthetic accelerograms are generated ( $N_s = 500$  based on every acceleration  
 187 record) in order to be used in the analytical seismic fragility curve estimation. Figure 3a shows  
 188 the 15<sup>th</sup>, 50<sup>th</sup> and 85<sup>th</sup> percentiles of the acceleration response spectra for 5 % damping of the  
 189 ground motion records in the data set, and the corresponding percentiles of the spectra based  
 190 on the synthetic ground motions. The percentiles of the spectral values of the synthetic ground  
 191 motions match well that of the acceleration records and we consider that the ground motion  
 192 variability of the synthetics reproduces the variability in the original ground motion data set.  
 193 We observe in Figure 3b that the percentiles of the acceleration response spectra of the  
 194 synthetic ground motions for 2 % damping also match well the percentiles of the response

195 spectra of the acceleration records. Therefore, we consider that the adjustment technique is  
 196 quasi-independent of the damping value in the computation of the response spectra.  
 197



198 a) b)  
 199 **Figure 3 Percentiles of the acceleration response spectra for a) 5 % and b) 2 % damping**  
 200 **of the synthetic accelerograms and the ground motions in the original data set**  
 201

### 202 3 FRAGILITY CURVE CONSTRUCTION

#### 203 3.1 Structural Model

204 For the illustrative application of this framework and for verification of the PMA-based  
 205 methodology an inelastic single degree of freedom structure is employed. Its frequency is 5  
 206 Hz, it has a damping ratio of 5 % and yield displacement ( $u_y$ ) of  $3.3 \cdot 10^{-3}$  m. Its post-yield  
 207 stiffness, defining kinematic hardening, is equal to the 20 % of its elastic stiffness (Figure 4a).  
 208 The response of the structure is computed by solving Equation 10 with the central difference  
 209 method.

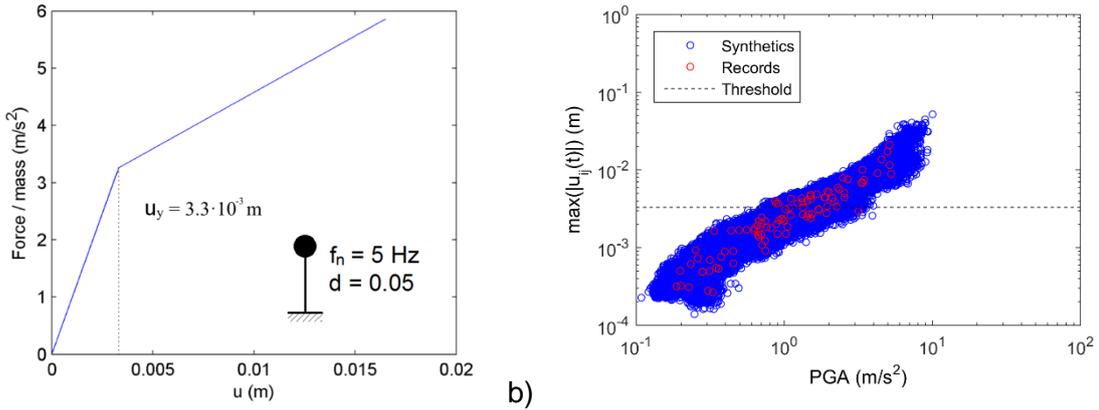
$$210$$

$$211 \quad m\ddot{u}_{ij}(t) + c\dot{u}_{ij}(t) + f_{ij}(t) = -m\alpha_{s,ij}(t) \quad (10)$$

212

213 where  $m$  is the mass of the oscillator,  $\ddot{u}_{ij}(t)$  and  $\dot{u}_{ij}(t)$  are the relative acceleration and velocity  
 214 of the mass, respectively, and  $f_{ij}(t)$  is the nonlinear resisting force.

215



216

a)

b)

217

218 **Figure 4 a) Backbone curve of the inelastic oscillator b) maximum oscillator**  
 219 **displacement ( $\max(|u_{ij}(t)|)$ ) observations as a function of the PGA**  
 220

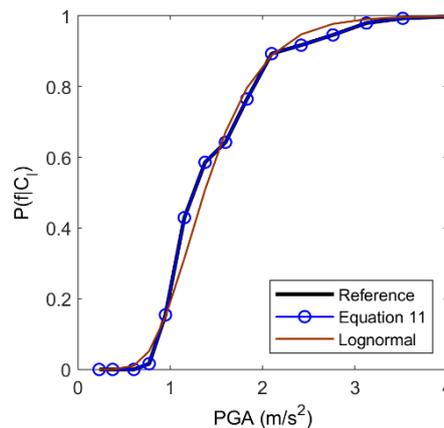
220

221 Figure 4b shows the maximum response of the inelastic oscillator under excitation with  
 222 the acceleration records and the generated synthetic ground motions. We observe that, in this  
 223 case, the responses under the synthetic ground motions are spread over an area between  
 224 and around the responses computed with the ground motion records. These data are used in  
 225 the different approaches here for deriving fragility curves.

226 **3.2 Empirical Non-Parametric Fragility Curves Based On MC Simulations and IM**  
 227 **Clustering**

228 The class of non-parametric fragility curves constructed here is based on MC  
 229 simulations and clustering of the Intensity Measure observations. In the illustrative example,  
 230 the maximum oscillator displacement is used as the EDP and the PGA is selected as IM for  
 231 simplicity while acknowledging that other IMs may be more efficient [30]. The total Intensity  
 232 Measure (IM) observations of all recorded and synthetic ground motions are classified to a  
 233 number of clusters with k-means clustering [23]. K-means clustering is an iterative optimization  
 234 procedure, which groups the IM observations in a selected number  $N_c$  of clusters. This  
 235 procedure also returns an IM value as the centroid of each cluster. The centroid of each cluster  
 236 is equal to the mean of the IM observations grouped in that cluster and the optimization  
 237 procedure consists in minimizing the sum of squares of its differences from the observations

238 in its cluster, i.e. the variance. Here, the IM observations are grouped into  $N_c = 20$  clusters  
 239 using the function “kmeans” in MATLAB [31], while the effect of IM discretization is out of the  
 240 scope of this work. Subsequently, the point probabilities are classically computed at the IM  
 241 cluster centroids ( $C_l, l = \{1, \dots, N_c\}$ ) as the ratio of the number of exceedances of the damage  
 242 state threshold, which are observed in the analyses corresponding to the IMs in a cluster, to  
 243 the number of total observations in the cluster. In this case, the damage state threshold is  
 244 equal to the yield displacement ( $3.3 \cdot 10^{-3}$  m) without loss of generality. Figure 5 shows the non-  
 245 parametric fragility curve computed in this case with 48096 seismic response analyses using  
 246 all available recorded and synthetic accelerograms. Whenever the entirety of original and  
 247 synthetic ground motions is used, the empirical Monte-Carlo-based non-parametric fragility  
 248 curve will be called “reference”. The derivation of the other curves in Figure 5 follows.  
 249



250

251 **Figure 5 Lognormal, reference and fragility curve according to Equation 11**

252

### 253 3.3 New Formulation Of The Non-Parametric Fragility Curves

254 The proposed PMA methodology in this paper for optimized estimation of non-  
 255 parametric fragility curves is based on Equation 11. This equation expresses the discrete

256 fragility curve  $P(f|C_l)$ , which is defined at  $N_c$  cluster centroids ( $C_l$ ), by means of the law of total  
257 probability.

258

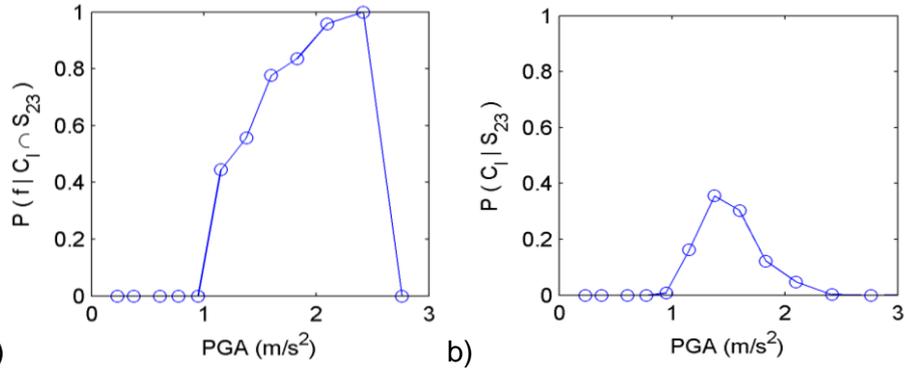
$$259 \quad P(f|C_l) = \frac{\sum_{i=1}^{N_r} (P(f|C_l \cap S_i) \cdot P(C_l|S_i) \cdot P(S_i))}{\sum_{i=1}^{N_r} (P(C_l|S_i) \cdot P(S_i))} \quad (11)$$

260

261 The conditional probability  $P(f|C_l \cap S_i)$  corresponds to the probability of exceeding the  
262 damage state threshold at cluster centroid  $C_l$  under excitation with ground motions originating  
263 from random process  $S_i$ . This is practically the fragility curve estimated with the ground  
264 motions originating from process  $S_i$ . The conditional probability  $P(C_l|S_i)$  is the probability of  
265 sorting the IM observations, which correspond to the ground motions belonging to process  $S_i$ ,  
266 in the  $l$ -th cluster. As an example, Figures 6a and 6b show  $P(f|C_l \cap S_{23})$  and  $P(C_l|S_{23})$ ,  
267 respectively, which result from an empirical computation. Finally, the probability  $P(S_i)$  equals  
268 the fraction of the number of ground motions used, which belong to random process  $S_i$ , to the  
269 total number of ground motions used to estimate the fragility curve. If we generate an equal  
270 number of synthetic ground motions for every one of  $N_r$  acceleration records, and all available  
271 ground motions are used in the computation, then  $P(S_i) = 1/N_r$ . This is the case in the validation  
272 of the Equation 11 which is presented in Figure 5. We use 96x500 synthetic ground motions  
273 generated by the random processes  $S_i$  in addition to the 96 ground motions in the original data  
274 set. Figure 5 shows that, as expected, the fragility curve defined by Equation 11 coincides with  
275 the empirical fragility curve used as reference.

276

277



278

279 **Figure 6 a) Empirical probability of exceeding the damage state threshold ( $\max(|u_{ij}(t)|)$   
 280  $= 3.3 \cdot 10^{-3}$  m) based on the synthetics generated based on the acceleration record  $S_{23}$  at  
 281 the cluster centroids ( $C_l$ ) b) probability of observing a PGA value in the synthetics  
 282 based on acceleration record  $S_{23}$   
 283**

### 284 3.4 Lognormal Curve Adjusted To The Non-Parametric Curve

285 In order to observe potential differences between lognormal fragility curves and the non-  
 286 parametric curves estimated with the different approaches herein, a Maximum Likelihood  
 287 Estimation of the lognormal cumulative distribution function is employed. The MLE of the  
 288 lognormal curve uses the point probabilities constituting the empirical fragility curve based on  
 289 the selected IM and corresponding EDP observations. The MLE is performed with Equations  
 290 12-15 and the estimated lognormal curve is given by Equation 16.

291

$$292 P(n_l, r_l, C_l) = \frac{n_l!}{r_l!(n_l - r_l)!} \cdot P(f|C_l)^{r_l} \cdot (1 - P(f|C_l))^{n_l - r_l} \quad (12)$$

293

$$294 L = \prod_{l=1}^{N_c} P(n_l, r_l, C_l) \quad (13)$$

295

$$296 \ln(L) = \sum_{l=1}^{N_c} \left[ \ln \left( \frac{n_l!}{r_l!(n_l - r_l)!} \right) + r_l \cdot \ln \Phi \left( \frac{\ln(C_l) - \ln(A)}{\beta} \right) + (n_l - r_l) \cdot \ln \left( 1 - \Phi \left( \frac{\ln(C_l) - \ln(A)}{\beta} \right) \right) \right] \quad (14)$$

297

$$298 \{\bar{A}, \bar{\beta}\} = \arg \max_{(A, \beta)} (\ln(L)) \quad (15)$$

299

$$300 \quad P(f|IM) = \Phi \left( \frac{\ln IM - \ln \bar{A}}{\bar{\beta}} \right) \quad (16)$$

301

302 Where  $n_l$  is the number of EDP observations corresponding to the IM observations in the l-th  
 303 cluster,  $r_l$  is the number of EDP observations, which correspond to the IM observations in the  
 304 l-th cluster, that exceed the damage state threshold,  $C_l$  the IM centroid of the l-th cluster,  $P(f|C_l)$   
 305 the empirical fraction of EDP observations exceeding the damage state threshold in the l-th  
 306 cluster,  $P(n_l, r_l, C_l)$  the binomial distribution, L the likelihood function,  $\Phi$  the standard normal  
 307 cumulative distribution function, A and  $\beta$  the median and the dispersion of the lognormal  
 308 distribution, respectively,  $\bar{A}$  and  $\bar{\beta}$  their estimations,  $P(f|IM)$  the probability of exceeding the  
 309 damage state threshold given the IM. The difference with the curve fitting by Baker [4] is that  
 310 the fractions of damage state threshold exceedances at the cluster IM centroids are used  
 311 instead of the fractions at the IMs of EDP stripes. Figure 5 includes a lognormal curve  
 312 computed with this approach using the point probabilities, which constitute the reference  
 313 fragility curve.

#### 314 **4 OPTIMIZATION WITH PARAMETRIC MODELS AVERAGING**

315 In order to illustrate the optimization of the non-parametric clustering fragility curve  
 316 estimation, we are employing five approaches: (i) MC un-optimized, (ii) lognormal un-  
 317 optimized, (iii) lognormal optimized, (iv) PMA – Model 1 and PMA – Model 2, and (v) reference.  
 318 The reference curve has already been described and used in the validation of Equation 11.  
 319 PMA – Model 1 and PMA – Model 2 are the two forms of the optimized approach which are  
 320 described in Sections 4.1-4.2.

321 In the *MC un-optimized approach*, the number of seismic response analyses is firstly  
 322 selected. Subsequently, an equal number of IM observations are selected from every cluster,  
 323 equal to the number of total analyses divided by the number of clusters (rounded down to the  
 324 closest integer). If there are less IM observations in some clusters than required, we select  
 325 those available and we select the rest by selecting an even number of observations from the

326 rest clusters and so on. After determining the number of IM observations per cluster that will  
327 be selected, the actual selection of the IM observations is made. This selection is based on  
328 the results of k-means clustering of the IM observations based on all synthetic and recorded  
329 accelerograms. K-means returns for every IM observation the index of the cluster to which the  
330 observation is sorted. Based on the returned indices, lists of the IM observations per cluster  
331 are made and the required observations per cluster are randomly selected from the  
332 corresponding lists. Subsequently, the seismic ground motions, which correspond to the  
333 selected IM observations, are used as excitations in dynamic time-history analyses of the  
334 oscillator in order to compute EDP observations. In the MC un-optimized approach, as in the  
335 reference, the probability of exceeding the damage state threshold is estimated empirically at  
336 the  $l$ -th cluster centroid as the observed fraction of EDP observations exceeding the damage  
337 state threshold to the total number of EDP observations corresponding to the IM observations  
338 in the cluster. The lognormal curve derived using the data used in the MC un-optimized  
339 approach will be called *lognormal un-optimized*.

340 The optimized PMA approach is based on Equation 11 and follows the procedure of the  
341 MC un-optimized approach with three modifications. First, the conditional probability  $P(C_i/S_i)$   
342 is not estimated with the selected IM observations, but with a very large number of IM  
343 observations in order to obtain a very precise estimation. Here, each  $P(C_i/S_i)$  distribution is  
344 empirically estimated with all available 501 IM observations; 500 observations corresponding  
345 to the synthetics and 1 to the original acceleration record. Practically, this means that the  
346 estimation of  $P(C_i/S_i)$  in the optimized approach and in the computation for the reference  
347 fragility curve are identical. It is worth noting that the estimation of  $P(C_i/S_i)$  does not require  
348 any seismic response analyses, but it requires only IM observations based on synthetic ground  
349 motions, which has a small computational cost. Second, IM observations (and the  
350 corresponding seismic ground motions used to compute EDP observations through seismic  
351 response analyses) are selected only if they are sorted in a cluster  $k_i$  where  $P(C_i/S_i)$  reaches  
352 its maximum. This is one of the key elements of the optimization process. In order to do so,  
353 the IM observations sorted in clusters other than the cluster, where  $P(C_i/S_i)$  of their process of

354 origin is maximized, are expunged from the lists of IM observations per cluster, from which IM  
355 observations are randomly drawn. The third and most important modification concerns the  
356 conditional probability of exceeding the damage state threshold in the case of each process  
357 ( $P(f|C_l \cap S_i)$ ). Instead of the empirical estimation of  $P(f|C_l \cap S_i)$ , the optimized approach employs  
358 two alternative parametric models. The first model (parametric model 1) assumes that  
359  $P(f|C_l \cap S_i)$  remains constant as a function of the IM, and that it is equal to  $P(f|C_{k_l} \cap S_i)$ . The  
360 second model (parametric model 2) uses a lognormal curve for  $P(f|C_l \cap S_i)$ . In the following, the  
361 parametric models 1 and 2 are analyzed.

#### 362 4.1 Parametric Model 1

363 The first model for  $P(f|C_l \cap S_i)$  is defined by a single parameter for every process. When  
364 using this model, the optimized approach will be called *PMA – Model 1*. This one parameter  
365 is taken equal to the empirical estimation of the probability of exceeding the damage state  
366 threshold at the  $k_l$ -th IM cluster centroid, where  $P(C_l|S_i)$  obtains its maximum value (Equation  
367 17). The one-parameter models are defined by Equation 18 and model the probability of  
368 exceeding the damage state threshold ( $P_{fi}$ ) per process as constant throughout all cluster IM  
369 centroids.

370

$$371 \quad k_l = \underset{(l)}{\arg \max} P(C_l|S_i)$$

$$372 \quad i = \{1, \dots, N_r = 96\}, \quad l = \{1, \dots, N_c = 20\} \quad (17)$$

373

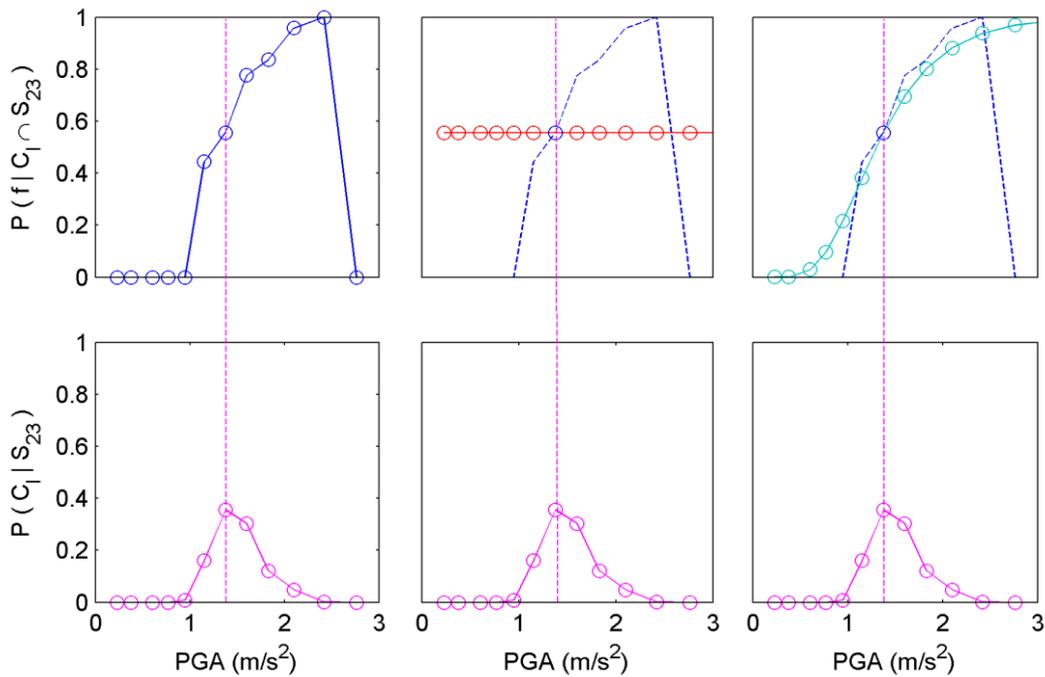
$$374 \quad P_{fi} = P(f|C_l \cap S_i) = P(f|C_{k_l} \cap S_i)$$

$$375 \quad i = \{1, \dots, N_r = 96\}, \quad l = \{1, \dots, N_c = 20\} \quad (18)$$

376

377 As an example, Figure 7 (top left) shows the empirical conditional probability  
378  $P(f|C_{k_{23}} \cap S_{23})$  estimated with the observations corresponding to the ground motions based  
379 on the 23<sup>rd</sup> accelerogram record. Moreover, Figure 7 (top middle) shows the corresponding

380 model used in the optimized approach, which assumes a constant probability (red curve),  
 381 which is estimated at the cluster IM centroid for which  $P(C_l|S_{23})$  is maximized (Figure 7  
 382 bottom). When employing parametric model 1, an error is introduced with respect to  $P(f|C_l \cap S_i)$ .  
 383 In specific,  $P(f|C_l \cap S_i)$  is under- and overestimated at IM cluster centroids where  $C_l > C_k$  and  $C_l$   
 384  $< C_k$ , respectively. The extent to which  $P(f|C_l \cap S_i)$  is under- or overestimated varies, and  
 385 generally increases with the distance between  $C_l$  and  $C_k$ . However, the introduced error is  
 386 mitigated by the fact that Equation 11 computes the product  $P(f|C_l \cap S_i) \cdot P(C_l|S_i)$ . The farther  $C_l$   
 387 is found from  $C_k$ , the smallest the introduced error, because  $P(C_l|S_i)$  diminishes with the  
 388 distance from  $C_k$  (e.g. Figure 7 bottom). Moreover, the fact that  $P(f|C_l \cap S_i)$  is simultaneously  
 389 under- and overestimated (e.g. Figure 7 middle) at  $C_l > C_k$  and  $C_l < C_k$ , respectively, also  
 390 mitigates the global error in the estimation of the fragility curve, as the underestimation on one  
 391 side balances to some extent the overestimation on the other.  
 392



393  
 394 **Figure 7 Top left: Empirical fragility curve based on the ground motions originating**  
 395 **from acceleration record 23. Top middle: parametric model 1 (constant probability of**  
 396 **damage state threshold exceedance). Top right: parametric model 2 (lognormal model)**  
 397 **Bottom: conditional probability  $P(C_l|S_{23})$ .**  
 398

399 **4.2 Parametric Model 2**

400 The lognormal curve is used as the second alternative parametric model for  $P(f|C_i \cap S_i)$   
 401 for every process. This form of the optimized approach will be called *PMA – Model 2*. In order  
 402 to define this model for every process, two parameters are required: the dispersion and the  
 403 median of the lognormal curve. These two parameters could be computed, if two or more point  
 404 probabilities were available, to which the lognormal curve might be fitted. Since the optimized  
 405 approach selects only IMs (and corresponding accelerograms) in cluster  $k_i$ , where  $P(C_i|S_i)$  is  
 406 maximized, and computes the corresponding EDPs and  $P(f|C_{k_i} \cap S_i)$ , the only available point  
 407 probability is  $(C_{k_i}, P(f|C_{k_i} \cap S_i))$ . Therefore, we assume that the dispersion of the lognormal  
 408 curve for every process ( $\beta_i$ ) is equal to the dispersion of the lognormal fragility curve ( $\bar{\beta}$ , which  
 409 will be referred to as  $\beta$  for simplicity), which is derived with Equations 12-16 using the data  
 410 selected according to the optimized approach. This curve will be called *lognormal optimized*  
 411 (there is no actual optimization here, this is simply part of the naming scheme). This allows us  
 412 to compute the median of the curve for every process ( $A_i$ ) with Equation 19. Based on  $A_i$  and  
 413  $\beta_i$ , the parametric model for every process is subsequently defined with Equation 20.

414

$$415 \quad A_i = \exp\left(\ln(C_{k_i}) - \beta_i \cdot \Phi^{-1}\left(P(f|C_{k_i} \cap S_i)\right)\right)$$

$$416 \quad \beta_i = \beta \tag{19}$$

417

$$418 \quad P(f|C_i \cap S_i) = \Phi\left(\frac{\ln(C_i) - \ln(A_i)}{\beta_i}\right) \tag{20}$$

419

420 As an example, Figure 7 (top right) shows  $P(f|C_i \cap S_{23})$  as estimated with the lognormal  
 421 parametric model (cyan curve). In this case, the model approximates well the empirical  
 422 estimation of the probability of exceeding the damage state threshold. This figure illustrates  
 423 that the largest differences between the probabilities given by the model and the empirical  
 424 estimation are observed where  $P(C_i|S_i)$  is close to zero. However, the empirical probabilities

425  $P(C_i|S_i)$  are equal to zero at the IM centroids of clusters without any IM observations.  
426 Therefore, at such cluster IM centroids, the product  $P(f|C_i \cap S_i) \cdot P(C_i|S_i)$  in Equation 11 is always  
427 zero, which means that no error is introduced at these clusters due to the use of a parametric  
428 model for  $P(f|C_i \cap S_i)$ . As shown in the following, parametric model 2 is particularly necessary  
429 when the dispersion of the lognormal optimized fragility curve is small (approximately for  $\beta <$   
430  $0.3$ ). In such cases, we consider justified to impose a common dispersion on all the parametric  
431 models corresponding to the processes  $S_i$ .

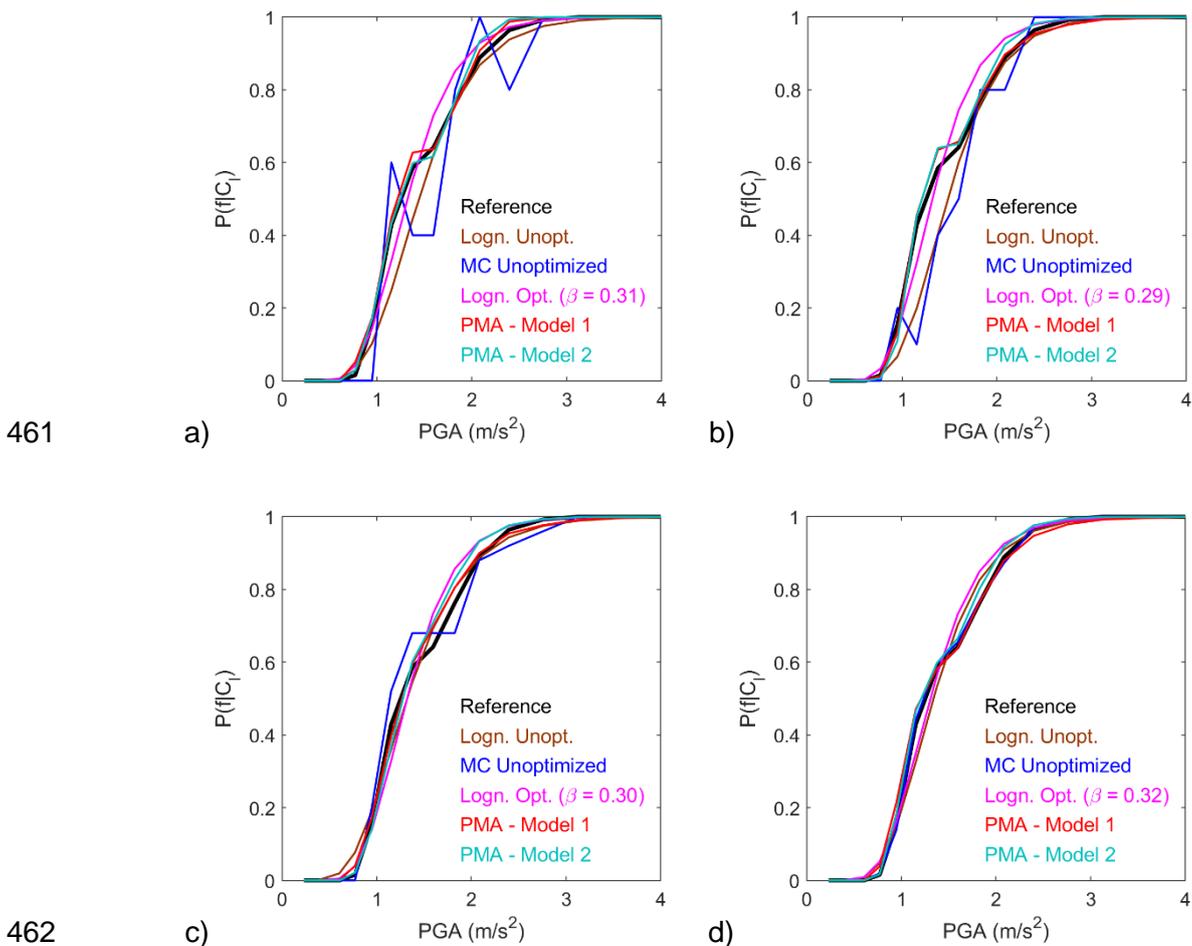
## 432 **5 APPLICATION OF THE METHODOLOGY**

433 In order to offer insight to the wider field of application of the developed methodology,  
434 which is essentially a MC procedure, we use it to compute the fragility curves in three cases:  
435 (i) inelastic oscillator without structural uncertainties and ground motions selected at random  
436 form the data set of all recorded and synthetic ground motions, (ii) inelastic oscillator without  
437 structural uncertainties and ground motions resulting from scaling of a single recorded  
438 accelerogram, (iii) inelastic oscillator with structural uncertainties. Based on the results of  
439 these three cases, we make our recommendations for practice. In the third case, the fragility  
440 curves are derived using as IM the PGA and the spectral acceleration at the frequency of the  
441 oscillator ( $S_a(5 \text{ Hz})$ ). To evaluate the effectiveness of the optimized procedures, we are  
442 comparing the estimated fragility curves with the reference curve and the 95 % CI according  
443 to the different approaches. The CI are computed based on bootstrap resampling [32] with a  
444 different set of 500 curves for each case.

### 445 **5.1 Structural Model Without Uncertainties And Data Set Of Acceleration Records**

446 The developed PMA-based optimization is firstly applied it in the case of the inelastic  
447 oscillator employed previously in the description of the methodology (Figure 4a). The ground  
448 motion data set used consists of the 96 recorded accelerograms and the 48,000  
449 corresponding synthetic ground motions generated with the described procedure in section 2.

450 The fragility curves are computed as a function of the PGA, for a damage state threshold  
 451 defined by a maximum oscillator displacement of  $3.3 \cdot 10^{-3}$  m, and according to the different  
 452 approaches are shown in Figure 8 in the case of 100, 200, 500, and 10,000 analyses,  
 453 respectively. The curves *MC un-optimized* and *lognormal un-optimized* are computed based  
 454 on the same set of seismic response analyses, which is different from the set of analyses used  
 455 for the optimized non-parametric curves. Every set of seismic response analyses is performed  
 456 using a different and randomly selected set of ground motions according to the optimized or  
 457 un-optimized approaches. Additionally, the reference non-parametric fragility curve, which is  
 458 estimated based on 48096 analyses with all recorded and synthetic accelerograms, is included  
 459 in the figures in order to observe any potential statistical error or bias in the evaluated curves.  
 460



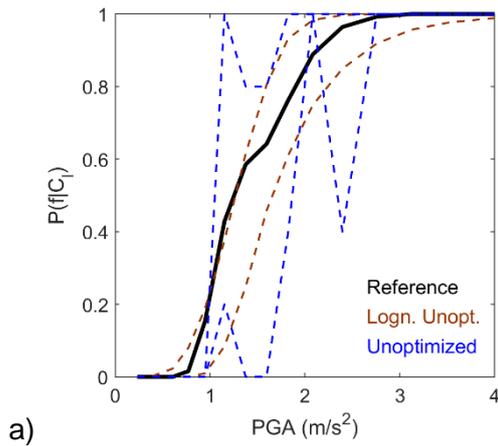
463 **Figure 8 Fragility curves for maximum oscillator displacement ( $\max(|u_{ij}(t)|)$ ) threshold**  
 464 **of  $3.3 \cdot 10^{-3}$  m evaluated without structural uncertainties and with the enriched ground**

465 **motion data set based all considered records and based on a) 100 b) 200 c) 500 and d)**  
466 **10,000 analyses, and the reference non-parametric fragility curve**  
467

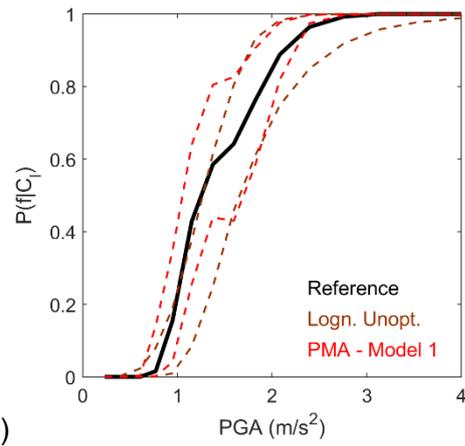
468 In the case of 100, 200 and 500 seismic response analyses (Figure 8a-c), the  
469 differences between the reference and the rest fragility curves is primarily due to error of  
470 estimation. However, in the case of 10,000 analyses, the difference is rather due to a bias in  
471 the computation, given that the fragility curves are evaluated with a very large number of  
472 analyses. As far as the MC un-optimized and PMA curves are concerned, they practically  
473 converge with the reference curve as the number of analyses increases, which means that no  
474 bias is introduced due to the assumptions in this case. In Figure 8d, we observe differences  
475 between the reference and the lognormal un-optimized curve based on 10,000 analyses.  
476 Given the number of analyses, we consider that the lognormal curve is biased. More important  
477 differences between lognormal and non-parametric curves may be observed, when –among  
478 other reasons– the studied structures are more complex than a single-degree-of-freedom  
479 oscillator, as in [18]. As a measure of the estimation error, the 95 % CI of the fragility curve  
480 based on 100 analyses and according to the different approaches are shown in Figure 9. As  
481 expected, the MC un-optimized approach gives a poor estimation (Figure 9a) due to the small  
482 amount of data and the lognormal un-optimized is more effective. The CI of the curves  
483 according to the lognormal un-optimized and the PMA – Model 1 approaches appear to be  
484 equivalent (Figure 9b). However, the median lognormal un-optimized curve may converge  
485 towards a biased estimation (e.g. Figure 8). Therefore, its confidence interval is not  
486 necessarily representative of the goodness of the estimation. This is a weakness of the  
487 parametric models and it is beforehand unknown whether there is bias in the fragility curve in  
488 complex cases. We observe that the PMA – Model 1 approach results in CI which are  
489 significantly smaller than the CI according to the MC un-optimized approach (Figure 9c).  
490 According to the curves in Figure 9d, we conclude that the two forms of the PMA optimization  
491 are equally effective in this case.

492

493

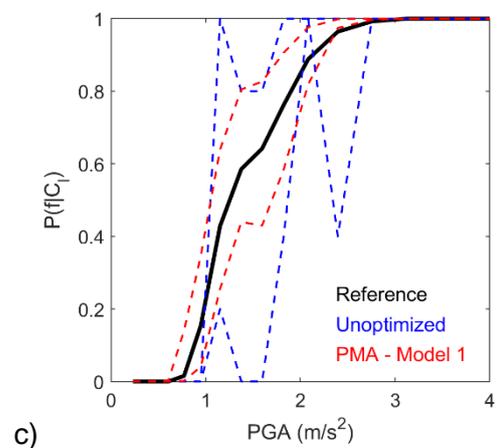


a)

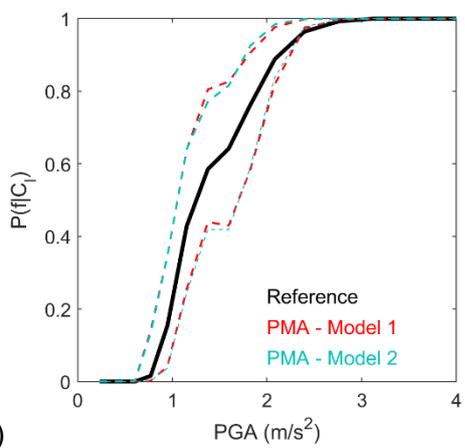


b)

494



c)



d)

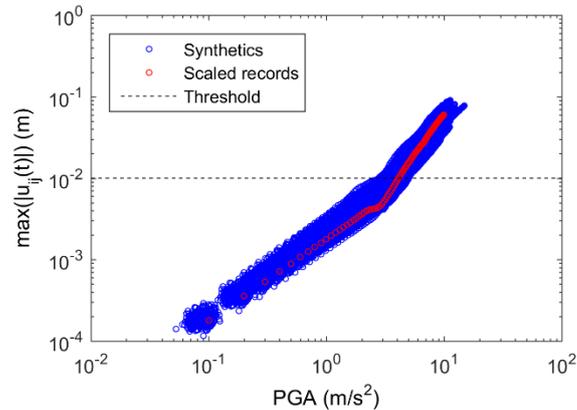
495 **Figure 9 95 % confidence intervals of fragility curves for maximum oscillator**  
 496 **displacement ( $\max(|u_{ij}(t)|)$ ) threshold of  $3.3 \cdot 10^{-3}$  m evaluated without structural**  
 497 **uncertainties and with the enriched ground motion data set based all considered**  
 498 **records and based on 100 analyses, and the reference non-parametric fragility curve**  
 499

500 **5.2 Structural Model Without Uncertainties And Data Set Of A Multiply Scaled**  
 501 **Acceleration Record**

502 Here we study a case with limited ground motion variability in order to demonstrate that  
 503 the applicability of the developed procedures for non-parametric fragility curve estimation  
 504 depends on the dispersion of the lognormal optimized curve, which is fitted to the data in an  
 505 intermediate step of the computation. In this case, the synthetic ground motions are generated  
 506 based on artificial accelerograms, which result from scaling multiple times (100 in this case) a  
 507 randomly selected acceleration record from the 96 original real records. Based on each  
 508 artificial accelerogram, 500 synthetic ground motions are generated with the procedure in

509 Section 2.2. Again, the oscillator in Figure 4a is employed. The maximum oscillator  
510 displacements as a function of the PGA based on all synthetic and artificial records, which are  
511 used for the reference non-parametric curve for this case, are shown in Figure 10.

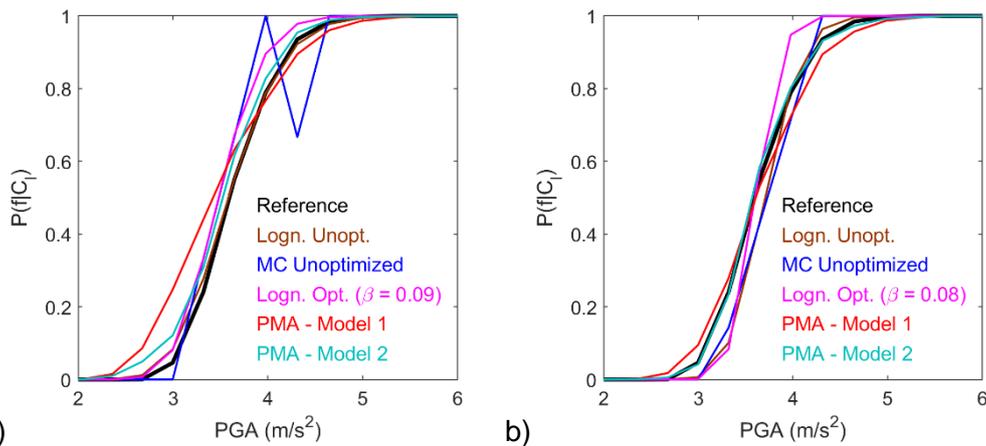
512



513

514 **Figure 10 Maximum oscillator displacement ( $\max(|u_{ij}(t)|)$ ) as a function of PGA**  
515 **computed without structural uncertainties and with the enriched ground motion data**  
516 **set based on scaling of a single randomly selected recorded accelerogram**  
517

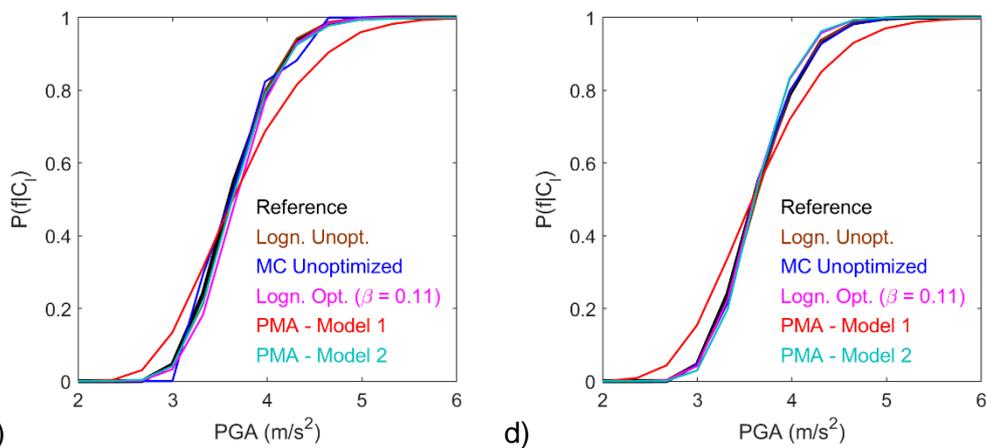
518 The fragility curves for a damage state threshold defined by a maximum oscillator  
519 displacement of  $1.0 \cdot 10^{-2}$  m according to the different considered approaches using 100, 200,  
520 500, and 10,000 seismic response analyses are shown in Figure 11. In this case, all fragility  
521 curves converge to the reference with the exception of the PMA – Model 1 curve, which is  
522 based on the optimization assuming models of constant  $P(f|C \cap S_i)$  per process. It is concluded  
523 that PMA model 1 produces a biased curve contrary to PMA model 2 (Figure 11d).  
524 Nevertheless, PMA model 2 may also result in bias in other cases (not shown here), when the  
525 dispersion of the unoptimized lognormal curve is very small ( $\beta < 0.1$ ). Indeed, in such cases,  
526 the reference fragility curve tends towards a step function, which cannot be approximated by  
527 the PMA-based procedures presented here unless a finer IM discretization is considered. It  
528 should also be taken into account that the observed difference between the reference curve  
529 and the curves according to the different approaches in the case of 100 and 200 seismic  
530 response analyses is principally an estimation error due to the limited number of seismic  
531 response analyses used.



533

a)

b)



534

c)

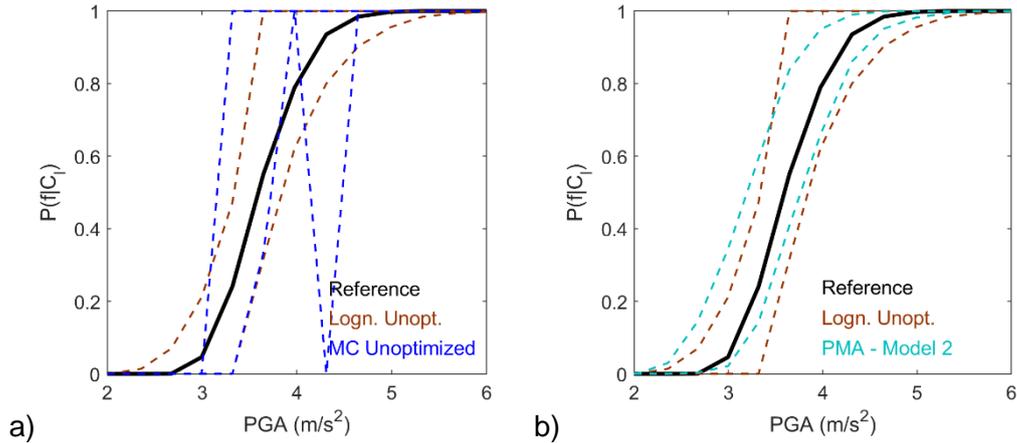
d)

535 **Figure 11 Fragility curves for maximum oscillator displacement ( $\max(|u_{ij}(t)|)$ ) threshold**  
 536 **of  $1.0 \cdot 10^{-2}$  m evaluated without structural uncertainties and with the enriched ground**  
 537 **motion data set based on scaling of acceleration record 27 and based on a) 100 b) 200**  
 538 **c) 500 and d) 10,000 analyses, and the reference non-parametric fragility curve**  
 539

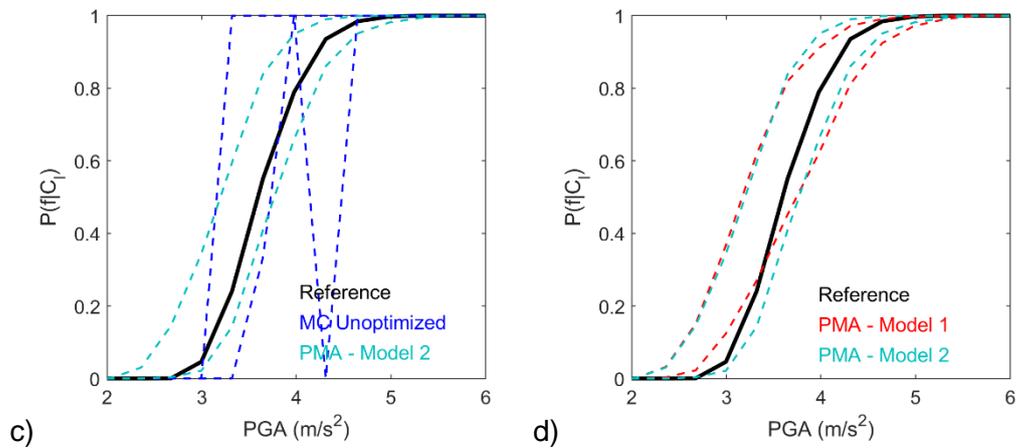
540 Figure 12 includes the 95 % CI of the fragility curves based on 100 seismic response  
 541 analyses. Also in this case, the lognormal un-optimized is more effective than the MC un-  
 542 optimized. The CI of the lognormal un-optimized and PMA – Model 2 curves indicate that both  
 543 approaches are effective in this case, with PMA – Model 2 being slightly better. Once more,  
 544 the estimation error according to the PMA – Model 2 approach is significantly less than the  
 545 error in the case of the MC un-optimized computation with 100 analyses. Contrary to the CI of  
 546 PMA – Model 1 curve, the CI of the PMA – model 2 curve envelopes the reference and  
 547 indicates that this is the preferable approach in this case.

548

549



550



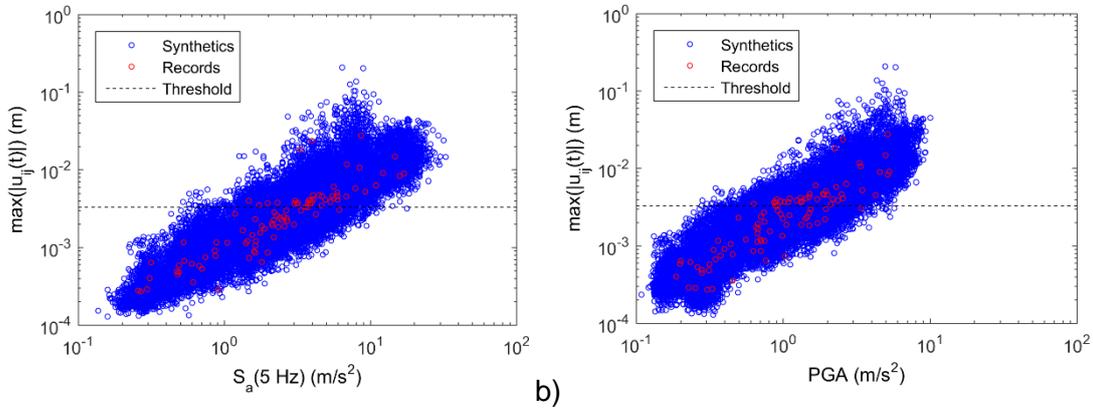
551 **Figure 12 95 % confidence intervals of fragility curves for maximum oscillator**  
 552 **displacement ( $\max(|u_{ij}(t)|)$ ) threshold of  $1.0 \cdot 10^{-2}$  m evaluated without structural**  
 553 **uncertainties and with the enriched ground motion data set based on scaling of**  
 554 **acceleration record 27 and based on 100 analyses, and the reference non-parametric**  
 555 **fragility curve**  
 556

557 **5.3 Structural Model With Uncertainties And Data Set Of Acceleration Records**

558 The developed optimization procedure is also applied in the case of uncertain structural  
 559 parameters. In specific, the oscillator in Figure 4a is employed and uncertainty is introduced  
 560 by considering the elastic frequency and the yield displacement of the oscillator as random  
 561 parameters with a coefficient of variation of 0.2. To do so, in every simulation, i.e. seismic  
 562 response analysis, the elastic frequency (5.0 Hz) and the yield displacement ( $3.3 \cdot 10^{-3}$  m) are  
 563 multiplied with random independent values sampled from two identical normal distributions  
 564 with mean and standard deviation equal to 1.0 and 0.2, respectively. Such pairs of random  
 565 values are sampled with Latin Hypercube Sampling for all 96 records and 48000 synthetic

566 ground motions in the data set. Figure 13 shows the damage state threshold and the computed  
 567 EDPs ( $\max(|u_{ij}(t)|)$ ) as a function of  $S_a(5 \text{ Hz})$  and PGA as IM, respectively.

568



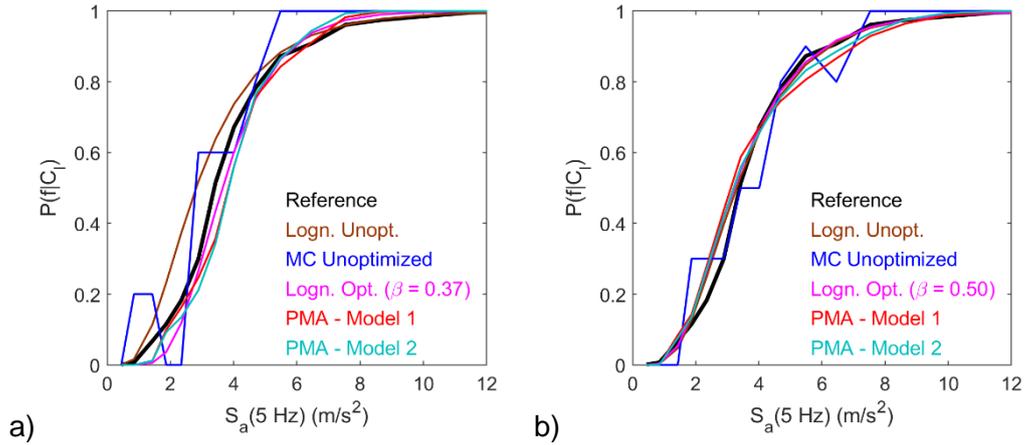
569

570 **Figure 13 Maximum oscillator displacement ( $\max(|u_{ij}(t)|)$ ) as a function of a) the spectral**  
 571 **acceleration at 5 Hz and b) the PGA in the case of the oscillator with frequency and**  
 572 **yield displacement uncertainty**  
 573

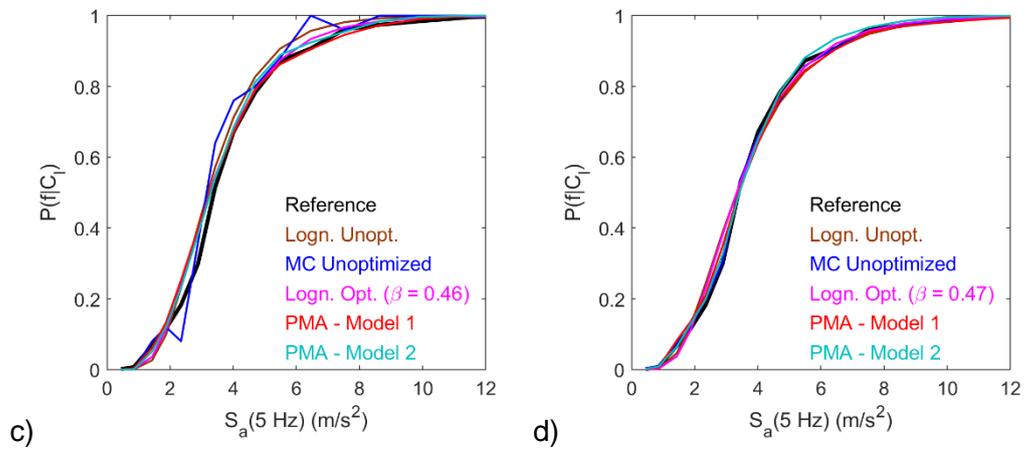
574 The fragility curves for a damage state threshold of  $3.3 \cdot 10^{-3}$  m maximum oscillator  
 575 displacement as a function of  $S_a(5 \text{ Hz})$  and PGA are shown in Figure 14 and 15. As expected,  
 576 (i) the introduction of uncertainties leads to increase of the dispersion of the lognormal fragility  
 577 curves and (ii) the dispersion of the lognormal fragility curves is slightly larger when PGA is  
 578 considered as IM. The optimized fragility curves and un-optimized non-parametric fragility  
 579 curves converge with the reference fragility curves in both cases (Figures 14d, 15d) and  
 580 present small differences from the lognormal curves. It is worth noting that the lognormal  
 581 optimized curve has a dispersion of 0.48 and 0.52 when using as IM the  $S_a(5 \text{ Hz})$  and the  
 582 PGA, respectively.

583

584

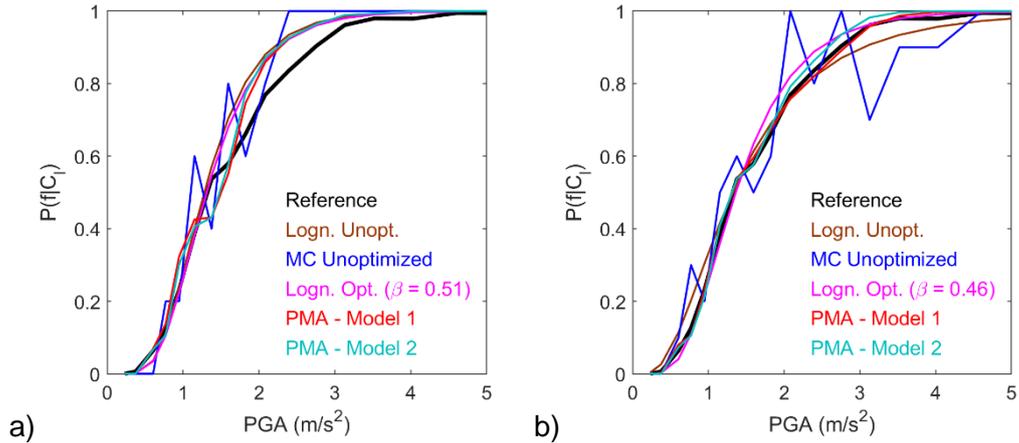


585

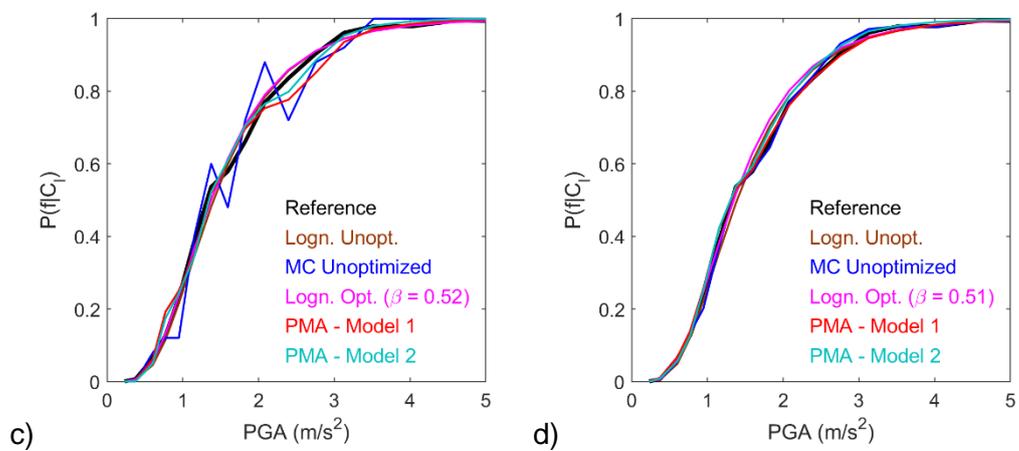


586 **Figure 14 Spectral acceleration ( $S_a(5 \text{ Hz})$ )-based fragility curves for maximum oscillator**  
587 **displacement ( $\max(|u_{ij}(t)|)$ ) threshold of  $3.3 \cdot 10^{-3} \text{ m}$  evaluated with structural**  
588 **uncertainties and with the enriched ground motion data set based all considered**  
589 **records and based on a) 100 b) 200 c) 500 and d) 10,000 analyses, and the reference**  
590 **non-parametric fragility curve**  
591

592



593



594 **Figure 15 PGA-based fragility curves for maximum oscillator displacement ( $\max(|u_{ij}(t)|)$ )**  
595 **threshold of  $3.3 \cdot 10^{-3}$  m evaluated with structural uncertainties and with the enriched**  
596 **ground motion data set based all considered records and based on a) 100 b) 200 c) 500**  
597 **and d) 10,000 analyses, and the reference non-parametric fragility curve**  
598

## 599 6 CONCLUSION

600 Here, we present a procedure for optimized derivation of non-parametric fragility curves  
601 using synthetic accelerograms. The fragility curves given by the presented procedure are  
602 intended for use for a specific structure rather than for a class of structures. A simple synthetic  
603 accelerogram generator is used, which reproduces the ground motion variability observed in  
604 a data set of ground motion records. However, the presented procedure is more general since  
605 it can use synthetic ground motions from other generators as long as they define random  
606 processes similar to those defined here. Also, note that the presented procedure is

607 independent of the selected IM. The optimization relies on the fact that the generated synthetic  
608 signals are realizations of a series of stochastic processes, each of which is based –in this  
609 work– on an acceleration record in the original data set. Using the EDP observations based  
610 on the synthetic ground motions, a parametric fragility curve is estimated for each process.  
611 Two alternative parametric models per process are proposed: a lognormal model and a model  
612 of constant probability of exceeding the damage state threshold. Based on the estimated  
613 models for all processes considered, a non-parametric fragility curve is estimated based on  
614 PMA, which computes the weighted average of the parametric models according to the law of  
615 total probability.

616 For the illustrative cases herein, synthetic ground motions are generated with a “simple”  
617 generator, which uses an original set of acceleration records. The generator produces  
618 synthetic ground motions with acceleration response spectra, whose 15<sup>th</sup>, 50<sup>th</sup> and 85<sup>th</sup>  
619 percentiles match well the corresponding percentiles of the spectral values of the ground  
620 motions in the original data set. All recorded and synthetic accelerograms are used as  
621 excitations of an inelastic single degree of freedom oscillator in order to obtain EDP  
622 observations as a function of the IM and estimate a reference fragility curve. The entirety of  
623 the IM observations of the recorded and synthetic ground motions is classified to clusters with  
624 k-means clustering. The number of clusters is selected based on engineering judgment, since  
625 the effect of the number of clusters is not studied here, and may nevertheless be a factor  
626 limiting the applicability of this methodology in some cases. Subsequently, the probability of  
627 exceeding the damage state threshold is estimated empirically at the cluster IM centroids  
628 using the EDP observations corresponding to the IM observations in each cluster. The result  
629 is the MC-based empirical non-parametric fragility curve, which is used as reference, as it is  
630 considered the best estimation possible based on the IM clustering approach and the available  
631 data. In the MC un-optimized approach, the same procedure is followed, but instead of using  
632 all data, an as constant as possible number of IM observations per cluster is selected so that  
633 the total number of analyses is in accordance with the available computational time.

634 The PMA optimized approach builds upon the MC un-optimized estimation by introducing  
635 an additional IM observation selection criterion. The IM observations in every cluster eligible  
636 for selection are those found in clusters with maximum probability of observation given the  
637 process, which generated the corresponding synthetic ground motions. Based on the EDP  
638 observations, which correspond to the selected IM observations, the probabilities of exceeding  
639 the damage state threshold at the IM cluster centroids are empirically estimated. These  
640 probabilities are used to define the parametric fragility curve, i.e. the parametric mode, which  
641 is related to each random process. Subsequently, the parametric models are averaged with  
642 the probabilities of occurrence of each random process in the clusters which are estimated  
643 with a very large number of synthetic ground motions, with practically no computational cost,  
644 since it requires no seismic response analyses. As in [18] or [21], we observe that non-  
645 parametric curves based on the proposed procedures may present differences from lognormal  
646 curves based on the same data. Here, the smallest differences between lognormal un-  
647 optimized and non-parametric fragility curves are observed when the dispersion of the  
648 lognormal curves are either very small (e.g.  $< 0.1$ ) or considerable (e.g.  $> 0.5$ ). As far as the  
649 uncertainty of the estimated non-parametric curve is concerned, we employ non-parametric  
650 bootstrap resampling to estimate the 95 % CI of the fragility curves. Moreover, the 95 % CI of  
651 the PMA curve is reduced with respect to the CI of the MC un-optimized curve for the same  
652 number of seismic response analyses in all cases in the study. In conclusion, the developed  
653 methodology is an efficient and useful procedure for fragility curve estimation and has wider  
654 applicability than a parametric model (e.g. the lognormal), which may lead to biased  
655 estimations.

656 Our recommendations are summarized in Table 1. The criteria that guide us are two: the  
657 dispersion of the lognormal optimized curve fitted to the selected data and the discretization  
658 of the IM observations, i.e. the number of clusters. When applying the proposed procedure,  
659 estimating a fragility curve while using a very coarse IM discretization can be considered  
660 equivalent to the estimation of a fragility curve with a very small dispersion. In the area of 100  
661 or less analyses, use of a typical un-optimized lognormal fragility curve is recommended. If

662 the resources for 10,000 or more analyses are available, the MC un-optimized approach can  
663 be used. In the area between 100 and 10,000 analyses, which is of practical interest, we  
664 suggest either a parametric curve, or a non-parametric optimized fragility curve computation  
665 with one of the two proposed alternatives. In this area, the dispersion of the optimized  
666 lognormal curve fitted to the selected data dictates the optimal approach. In the case of a large  
667 dispersion ( $0.3 \leq \beta$ ), the optimization with the constant probability of damage threshold  
668 exceedance per process is sufficient, while in the case of a limited dispersion ( $0.1 \leq \beta < 0.3$ ),  
669 the optimization with the lognormal model per process is recommended. When PMA – Model  
670 1 and 2 use a large number of seismic response analyses and give drastically different results  
671 (as in the case with an original data set consisting of ground motions resulting from scaling a  
672 single acceleration record), PMA – Model 2 should be preferred, unless the dispersion of the  
673 associated optimized lognormal curve is very small ( $\beta < 0.1$ ). In such cases, the presented  
674 PMA approaches are not efficient and a lognormal model for the fragility curve is  
675 recommended.

676

677 **Table 1 Recommended type of fragility curve based on the number of seismic analyses**  
678 **(N) and the dispersion of the lognormal (un-optimized) curve fitted to the empirical non-**  
679 **parametric curve ( $\beta$ )**

680

	$\beta < 0.1$	$0.1 \leq \beta < 0.3$	$0.3 \leq \beta$
$N < 100$	Un-optimized Lognormal	Un-optimized Lognormal	Un-optimized Lognormal
$100 \leq N < 10,000$	Un-optimized Lognormal	PMA – Model 2	PMA – Model 1
$10,000 \leq N$	MC Un-optimized	MC Un-optimized	MC Un-optimized

681

682 Our procedure has also been applied in the case of a realistic finite element model of a low-  
683 rise reinforced concrete bare frame (modelling details may be found in [33]), not presented  
684 here for the sake of brevity. The results lead to the same conclusions. Should one attempt to  
685 apply the procedure herein in the case of complex structures, they will face a series of  
686 challenges, which are, however, not specific to our methodology. A major concern would be  
687 the selection of an efficient IM. IMs are considered efficient [34], when the seismic response

688 as a function of the IM has a low dispersion. However, scalar IMs are not efficient in every  
689 case. For example, the spectral acceleration at the first eigenfrequency is a common scalar  
690 IM, which is efficient in the case of structures, whose response is mostly affected by their first  
691 mode. However, it is not efficient in the case of tall buildings [35]. In the case of structures with  
692 multiple degrees of freedom the use of more adapted IMs, or even a vector of different IMs  
693 [36] may be a solution. That said, further investigations should be made to see if the  
694 procedures herein can be modified to use a vector IM. Although, the procedure herein is –in  
695 principle– independent of the selected IM and the damage state, it should be adapted to more  
696 severe damage states such as collapse. Indeed, the simulation of severe damage states may  
697 be computationally demanding and may require to take into account P-delta effects [37], to  
698 simulate brittle failure modes [38] and consider alternative IMs [39]. In addition, a validation of  
699 our methodology with a very large number of seismic response analyses in the case of  
700 complex structures has a prohibitive computational cost. To test the usefulness of our  
701 procedure in the case of complex structures, fragility curves given by our procedure based on  
702 a reasonable number of seismic response analyses (e.g. a few hundred) could be compared  
703 with curves given by other procedures, which reduce the computational cost. Such procedures  
704 may rely, amongst others, on metamodeling strategies based on neural networks [40], or  
705 support vector machines [41]. Finally, further studies of the developed procedure using  
706 realistic structural models and fragility curves conditioned on failure, instead of curves  
707 conditioned on an engineering demand parameter threshold, should provide additional  
708 insights.

## 709 **ACKNOWLEDGMENTS**

710 Funding: This work was supported by the research project *SINAPS@* (ANR-11-RSNR-0022),  
711 a project of the SEISM Institute (<https://institut-seism.fr/>) funded by The French National  
712 Research Agency in the context of its program *Investments for the Future*.

713 **REFERENCES**

- 714 [1] Fragiadakis M, Vamvatsikos D, Karlaftis MG, Lagaros ND, Papadrakakis M. Seismic  
715 assessment of structures and lifelines. *J Sound Vib* 2015;334:29–56.  
716 doi:10.1016/j.jsv.2013.12.031.
- 717 [2] Yang TY, Moehle J, Stojadinovic B, Der Kiureghian A. Seismic Performance Evaluation  
718 of Facilities: Methodology and Implementation. *J Struct Eng* 2009;135:1146–54.  
719 doi:10.1061/(ASCE)0733-9445(2009)135:10(1146).
- 720 [3] FEMA. Seismic Performance Assessment of Buildings – Volume 1 – Methodology.  
721 WASHINGTON, DC: 2012.
- 722 [4] Baker JW. Efficient Analytical Fragility Function Fitting Using Dynamic Structural  
723 Analysis. *Earthq Spectra* 2015;31:579–99. doi:10.1193/021113EQS025M.
- 724 [5] Silva V, Crowley H, Bazzurro P. Exploring Risk-Targeted Hazard Maps for Europe.  
725 *Earthq Spectra* 2016;32:1165–86. doi:10.1193/112514EQS198M.
- 726 [6] Berge-Thierry C, Svay A, Laurendeau A, Chartier T, Perron V, Guyonnet-Benaize C,  
727 et al. Toward an integrated seismic risk assessment for nuclear safety improving current  
728 French methodologies through the SINAPS@ research project. *Nucl Eng Des* 2017;323:185–  
729 201. doi:10.1016/j.nucengdes.2016.07.004.
- 730 [7] Tsionis G, Mignan D, Pinto A, Giardini D, European Commission, Joint Research  
731 Centre. Harmonized approach to stress tests for critical infrastructures against natural  
732 hazards. Luxembourg: Publications Office; 2016.
- 733 [8] Zhang J, Huo Y. Evaluating effectiveness and optimum design of isolation devices for  
734 highway bridges using the fragility function method. *Eng Struct* 2009;31:1648–60.  
735 doi:10.1016/j.engstruct.2009.02.017.
- 736 [9] Saha SK, Matsagar V, Chakraborty S. Uncertainty quantification and seismic fragility  
737 of base-isolated liquid storage tanks using response surface models. *Probabilistic Eng Mech*  
738 2016;43:20–35. doi:10.1016/j.probengmech.2015.10.008.

- 739 [10] Patil A, Jung S, Kwon O-S. Structural performance of a parked wind turbine tower  
740 subjected to strong ground motions. *Eng Struct* 2016;120:92–102.  
741 doi:10.1016/j.engstruct.2016.04.020.
- 742 [11] Gidaris I, Taflanidis AA, Mavroeidis GP. Kriging metamodeling in seismic risk  
743 assessment based on stochastic ground motion models: Seismic Risk Assessment Through  
744 Kriging Metamodeling. *Earthq Eng Struct Dyn* 2015;44:2377–99. doi:10.1002/eqe.2586.
- 745 [12] Parolai S, Haas M, Pittore M, Fleming K. Bridging the Gap Between Seismology and  
746 Engineering: Towards Real-Time Damage Assessment. In: Pitilakis K, editor. *Recent Adv.*  
747 *Earthq. Eng. Eur.*, vol. 46, Cham: Springer International Publishing; 2018, p. 253–61.  
748 doi:10.1007/978-3-319-75741-4\_10.
- 749 [13] Quilligan A, O'Connor A, Pakrashi V. Fragility analysis of steel and concrete wind  
750 turbine towers. *Eng Struct* 2012;36:270–82. doi:10.1016/j.engstruct.2011.12.013.
- 751 [14] D'Ayala D, Meslem A, Vamvatsikos D, Porter K, Rossetto T, Silva V. Guidelines for  
752 Analytical Vulnerability Assessment - Low/Mid-Rise. GEM; 2015.
- 753 [15] Vamvatsikos D, Cornell CA. Incremental dynamic analysis. *Earthq Eng Struct Dyn*  
754 2002;31:491–514. doi:10.1002/eqe.141.
- 755 [16] Jalayer F, Cornell CA. Alternative non-linear demand estimation methods for  
756 probability-based seismic assessments. *Earthq Eng Struct Dyn* 2009;38:951–72.  
757 doi:10.1002/eqe.876.
- 758 [17] Zentner I. A general framework for the estimation of analytical fragility functions based  
759 on multivariate probability distributions. *Struct Saf* 2017;64:54–61.  
760 doi:10.1016/j.strusafe.2016.09.003.
- 761 [18] Mai C, Konakli K, Sudret B. Seismic fragility curves for structures using non-parametric  
762 representations. *Front Struct Civ Eng* 2017;11:169–86. doi:10.1007/s11709-017-0385-y.
- 763 [19] Noh HY, Lallemand D, Kiremidjian AS. Development of empirical and analytical fragility  
764 functions using kernel smoothing methods: DEVELOPMENT OF FRAGILITY FUNCTIONS  
765 USING KERNEL SMOOTHING METHODS. *Earthq Eng Struct Dyn* 2015;44:1163–80.  
766 doi:10.1002/eqe.2505.

- 767 [20] Lallemand D, Kiremidjian A, Burton H. Statistical procedures for developing earthquake  
768 damage fragility curves: STATISTICAL PROCEDURES FOR DAMAGE FRAGILITY  
769 CURVES. *Earthq Eng Struct Dyn* 2015;44:1373–89. doi:10.1002/eqe.2522.
- 770 [21] Karamlou A, Bocchini P. Computation of bridge seismic fragility by large-scale  
771 simulation for probabilistic resilience analysis: BRIDGE SEISMIC FRAGILITY BY LARGE-  
772 SCALE SIMULATION FOR RESILIENCE. *Earthq Eng Struct Dyn* 2015;44:1959–78.  
773 doi:10.1002/eqe.2567.
- 774 [22] Porter K, Kennedy R, Bachman R. Creating Fragility Functions for Performance-Based  
775 Earthquake Engineering. *Earthq Spectra* 2007;23:471–89. doi:10.1193/1.2720892.
- 776 [23] Jain AK, Murty MN, Flynn PJ. Data clustering: a review. *ACM Comput Surv*  
777 1999;31:264–323. doi:10.1145/331499.331504.
- 778 [24] Ambraseys N, Smit P, Sigbjornsson R, Suhadolc P, Margaris B. Internet-Site for  
779 European Strong-Motion Data. European Commission, Research-Directorate General,  
780 Environment and Climate Programme; 2002.
- 781 [25] Ambraseys N, Smit P, Douglas J, Margaris B, Sigbjornsson R, Olafsson S, et al.  
782 Internet site for European strong-motion data. *Boll Geofis Teor E Appl* 2004;45:113–29.
- 783 [26] Rezaeian S, Der Kiureghian A. Simulation of synthetic ground motions for specified  
784 earthquake and site characteristics. *Earthq Eng Struct Dyn* 2010;39:1155–1180.  
785 doi:10.1002/eqe.997.
- 786 [27] Boore DM. Simulation of Ground Motion Using the Stochastic Method. *Pure Appl*  
787 *Geophys* 2003;160:635–76. doi:10.1007/PL00012553.
- 788 [28] Housner G, Jennings P. Generation of Artificial Earthquakes. *J Eng Mech Div*  
789 1964;90:113–52.
- 790 [29] Rodolfo Saragoni G, Hart GC. Simulation of artificial earthquakes. *Earthq Eng Struct*  
791 *Dyn* 1973;2:249–67. doi:10.1002/eqe.4290020305.
- 792 [30] Kostinakis K, Fontara I-K, Athanatopoulou AM. Scalar Structure-Specific Ground  
793 Motion Intensity Measures for Assessing the Seismic Performance of Structures: A Review. *J*  
794 *Earthq Eng* 2018;22:630–65. doi:10.1080/13632469.2016.1264323.

- 795 [31] MATLAB. MathWorks; 2015.
- 796 [32] Iervolino I. Assessing uncertainty in estimation of seismic response for PBEE. *Earthq*  
797 *Eng Struct Dyn* 2017;46:1711–23. doi:10.1002/eqe.2883.
- 798 [33] Trevelopoulos K, Guéguen P. Period elongation-based framework for operative  
799 assessment of the variation of seismic vulnerability of reinforced concrete buildings during  
800 aftershock sequences. *Soil Dynamics and Earthquake Engineering* 2016;84:224–37.  
801 doi:10.1016/j.soildyn.2016.02.009.
- 802 [34] Luco N, Cornell CA. Structure-Specific Scalar Intensity Measures for Near-Source and  
803 Ordinary Earthquake Ground Motions. *Earthq Spectra* 2007;23:357–92.  
804 doi:10.1193/1.2723158.
- 805 [35] Jayaram N, Lin T, Baker JW. A Computationally Efficient Ground-Motion Selection  
806 Algorithm for Matching a Target Response Spectrum Mean and Variance. *Earthq Spectra*  
807 2011;27:797–815. doi:10.1193/1.3608002.
- 808 [36] Kohrangi M, Bazzurro P, Vamvatsikos D. Vector and Scalar IMs in Structural Response  
809 Estimation, Part II: Building Demand Assessment. *Earthq Spectra* 2016;32:1525–43.  
810 doi:10.1193/053115EQS081M.
- 811 [37] Eads L, Miranda E, Krawinkler H, Lignos DG. An efficient method for estimating the  
812 collapse risk of structures in seismic regions: AN EFFICIENT METHOD FOR ESTIMATING  
813 THE COLLAPSE RISK OF STRUCTURES. *Earthq Eng Struct Dyn* 2013;42:25–41.  
814 doi:10.1002/eqe.2191.
- 815 [38] Kyriakides NC, Pantazopoulou SJ. Collapse Fragility Curves for RC Buildings  
816 Exhibiting Brittle Failure Modes. *J Struct Eng* 2018;144:04017207.  
817 doi:10.1061/(ASCE)ST.1943-541X.0001920.
- 818 [39] Eads L, Miranda E, Lignos DG. Average spectral acceleration as an intensity measure  
819 for collapse risk assessment: Average Spectral Acceleration as an IM for Collapse Risk  
820 Assessment. *Earthq Eng Struct Dyn* 2015;44:2057–73. doi:10.1002/eqe.2575.

821 [40] Wang Z, Pedroni N, Zentner I, Zio E. Seismic fragility analysis with artificial neural  
822 networks: Application to nuclear power plant equipment. Eng Struct 2018;162:213–25.  
823 doi:10.1016/j.engstruct.2018.02.024.

824 [41] Saint R, Feau C, Martinez J, Garnier J. Efficient Seismic fragility curve estimation by  
825 Active Learning on Support Vector Machines (submitted). Struct Saf  
826 <https://arxiv.org/abs/1810.01240>

827