



HAL
open science

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

Konstantinos Trevelopoulos, Cyril Feau, Irmela Zentner

► To cite this version:

Konstantinos Trevelopoulos, Cyril Feau, Irmela Zentner. Parametric models averaging for optimized non-parametric fragility curve estimation based on intensity measure data clustering. *Structural Safety*, 2019, 81, pp.101865. 10.1016/j.strusafe.2019.05.002 . hal-03484490

HAL Id: hal-03484490

<https://hal.science/hal-03484490>

Submitted on 20 Dec 2021

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.



Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License

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

3

4 Konstantinos Trevelopoulos^{a1}, Cyril Feau^a, Irmela Zentner^{b,c}

5 ^aDEN - Service d'études mécaniques et thermiques (SEMT), CEA, Université Paris-Saclay,
6 F-91191 Gif-sur-Yvette, France

7 ^bEDF R&D, EDF Lab Paris-Saclay, 7 Bvd Gaspard Monge, 91120 Palaiseau, France

8 ^cIMSIA, 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.

24

¹Present address: French Alternative Energies and Atomic Energy Commission (CEA), DEN, 13108 Saint-Paul-lez-Durance CEDEX, France

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 [26,27] 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 [24]. 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

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

124

$$125 \alpha_{s,ij}(t) = \sum_{m=1}^n (A_{r,im} \sin(\omega_m + \varphi_{s,ijm})) \quad i = \{1, \dots, N_r\} \quad j = \{1, \dots, N_s\} \quad (2)$$

126

127 where $\varphi_{s,ijm}$ is the phase which is assumed to be a random variable with a uniform distribution
 128 $U(-\pi, \pi)$ according to Boore [28], and ω_m is the m -th discrete angular frequency. The N_r

129 stationary Gaussian processes are converted to non-stationary processes using N_r modulation
 130 functions. Here the function by Housner and Jennings [29] (Equation 3) is used, however other
 131 modulation functions, e.g. [30], may be considered.

132

$$133 \quad q_i(t) = \begin{cases} \left(\frac{t}{T_{1,i}}\right)^3 & 0 \leq t \leq T_{1,i} \\ 1.0 & T_{1,i} < t \leq T_{2,i} \\ e^{-(t-T_{2,i})} & T_{2,i} < t \leq t_{d,i} \end{cases} \quad i = \{1, \dots, N_r\} \quad (3)$$

134

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

139

$$140 \quad I_{r,i} = \frac{\pi}{2g} \int_0^{t_{d,i}} \alpha_{r,i}^2(t) dt \quad i = \{1, \dots, N_r\} \quad (4)$$

141

142 $T_{1,i}$ and $T_{2,i}$ are computed with Equations 5 and 6. As an example, Figure 1a shows the
 143 modulation function used for the synthetic ground motions based on real record No. 11.

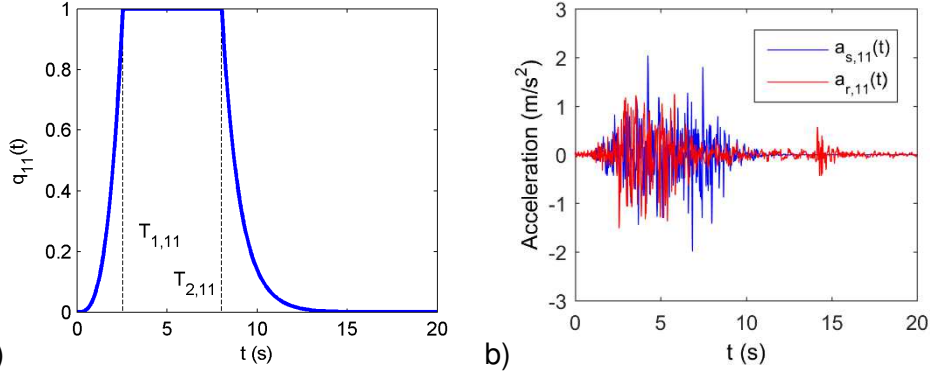
144

$$145 \quad \frac{\pi}{2g} \int_0^{T_{1,i}} \alpha_{r,i}^2(t) dt = 0.05 \cdot I_{r,i} \quad i = \{1, \dots, N_r\} \quad (5)$$

146

$$147 \quad \frac{\pi}{2g} \int_0^{T_{2,i}} \alpha_{r,i}^2(t) dt = 0.95 \cdot I_{r,i} \quad i = \{1, \dots, N_r\} \quad (6)$$

148



149

a)

b)

150 **Figure 1 a) Modulation function b) synthetic accelerogram and its original acceleration**
 151 **record**

152

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

155

$$156 \quad \alpha_{s0,ij}(t) = q_i(t) \cdot \sum_{m=1}^n \left(A_{r,im} \cdot \sin(\omega_m + \varphi_{s,ijm}) \right) \quad i = \{1, \dots, N_r\} \quad j = \{1, \dots, N_s\} \quad (7)$$

157

158 Subsequently, the synthetic ground motions generated based on an acceleration
 159 record are all scaled with the same scaling factor (c), which minimizes the sum of the squares
 160 of the differences between the acceleration response spectrum of the acceleration record for
 161 5 % damping ($S_{a,r,i}(f)$) and the median spectrum for 5 % damping of the scaled synthetic ground
 162 motions ($c \cdot \bar{S}_{a,s0,i}(f)$) over the frequencies between 0.2 and 25 Hz (Equation 8). The adjusted
 163 synthetic ground motions ($\alpha_{s,ij}(t)$) are given by Equation 9. As an example, Figure 1b shows
 164 record No. 11 and one of its spectrally equivalent synthetic accelerograms. Figure 2 shows
 165 the acceleration response spectrum of ground motion record No. 11, the spectra of all
 166 synthetic ground motions generated based on this record and the median spectrum of the
 167 synthetics ($c_{11} \cdot \bar{S}_{a,s0,11}(f)$).

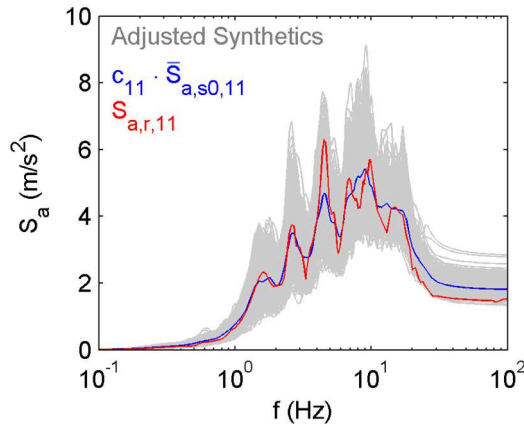
168

$$169 \quad c_i = \underset{(c)}{\arg \min} \left(\sum_{f=0.2 \text{ Hz}}^{f=25 \text{ Hz}} (S_{a,r,i}(f) - c \cdot \bar{S}_{a,s0,i}(f))^2 \right) \quad i = \{1, \dots, N_r\} \quad (8)$$

170

171 $\alpha_{s,ij}(t) = c_i \cdot q_i(t) \cdot \sum_{m=1}^n (A_{r,im} \cdot \sin(\omega_m + \varphi_{s,ijm})) \quad i = \{1, \dots, N_r\} \quad j = \{1, \dots, N_s\} \quad (9)$

172



173

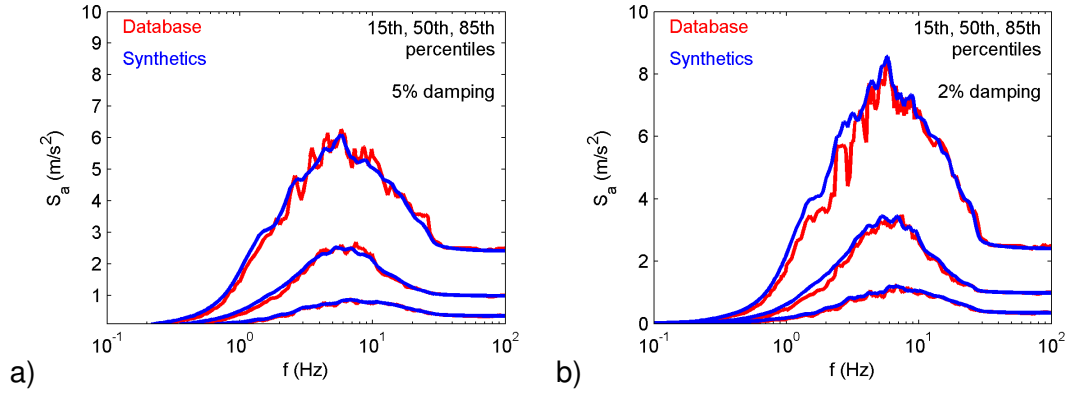
174 **Figure 2 Acceleration response spectra for 5 % damping of the adjusted synthetic**
 175 **ground motions and their original ground motion**

176

177

178 Based on $N_r = 96$ original acceleration records, a total of $N_r \times N_s = 48000$ “spectrally
 179 equivalent” synthetic accelerograms are generated ($N_s = 500$ based on every acceleration
 180 record) in order to be used in the analytical seismic fragility curve estimation. Figure 3a shows
 181 the 15th, 50th and 85th percentiles of the acceleration response spectra for 5 % damping of the
 182 ground motion records in the data set, and the corresponding percentiles of the spectra based
 183 on the synthetic ground motions. The percentiles of the spectral values of the synthetic ground
 184 motions match well that of the acceleration records and we consider that the ground motion
 185 variability of the synthetics reproduces the variability in the original ground motion data set.
 186 We observe in Figure 3b that the percentiles of the acceleration response spectra of the
 187 synthetic ground motions for 2 % damping also match well the percentiles of the response
 188 spectra of the acceleration records. Therefore, we consider that the adjustment technique is
 189 quasi-independent of the damping value in the computation of the response spectra.

190



191

192

193

194

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

195

3 FRAGILITY CURVE CONSTRUCTION

196

3.1 Structural Model

197

198

199

200

201

202

203

204

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

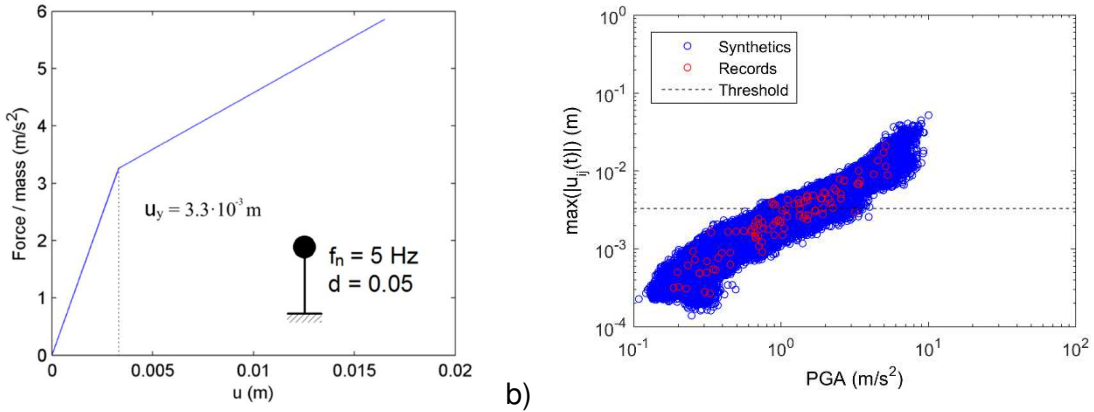
205

206

207

208

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



209

a)

b)

210

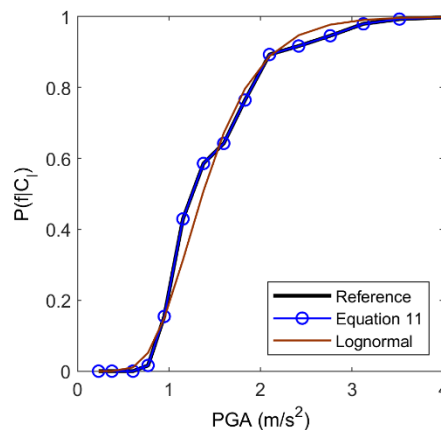
211 **Figure 4 a) Backbone curve of the inelastic oscillator b) maximum oscillator**
 212 **displacement ($\max(|u_{ij}(t)|)$) observations as a function of the PGA**
 213

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

219 **3.2 Empirical Non-Parametric Fragility Curves Based On MC Simulations and IM**
 220 **Clustering**

221 The class of non-parametric fragility curves constructed here is based on MC
 222 simulations and clustering of the Intensity Measure observations. In the illustrative example,
 223 the maximum oscillator displacement is used as the EDP and the PGA is selected as IM for
 224 simplicity while acknowledging that other IMs may be more efficient [31]. The total Intensity
 225 Measure (IM) observations of all recorded and synthetic ground motions are classified to a
 226 number of clusters with k-means clustering [23]. K-means clustering is an iterative optimization
 227 procedure, which groups the IM observations in a selected number N_c of clusters. This
 228 procedure also returns an IM value as the centroid of each cluster. The centroid of each cluster
 229 is equal to the mean of the IM observations grouped in that cluster and the optimization
 230 procedure consists in minimizing the sum of squares of its differences from the observations

231 in its cluster, i.e. the variance. Here, the IM observations are grouped into $N_c = 20$ clusters
 232 using the function “kmeans” in MATLAB [32], while the effect of IM discretization is out of the
 233 scope of this work. Subsequently, the point probabilities are classically computed at the IM
 234 cluster centroids ($C_l, l = \{1, \dots, N_c\}$) as the ratio of the number of exceedances of the damage
 235 state threshold, which are observed in the analyses corresponding to the IMs in a cluster, to
 236 the number of total observations in the cluster. In this case, the damage state threshold is
 237 equal to the yield displacement ($3.3 \cdot 10^{-3}$ m) without loss of generality. Figure 5 shows the non-
 238 parametric fragility curve computed in this case with 48096 seismic response analyses using
 239 all available recorded and synthetic accelerograms. Whenever the entirety of original and
 240 synthetic ground motions is used, the empirical Monte-Carlo-based non-parametric fragility
 241 curve will be called “reference”. The derivation of the other curves in Figure 5 follows.
 242



243

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

245

246 3.3 New Formulation Of The Non-Parametric Fragility Curves

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

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

251

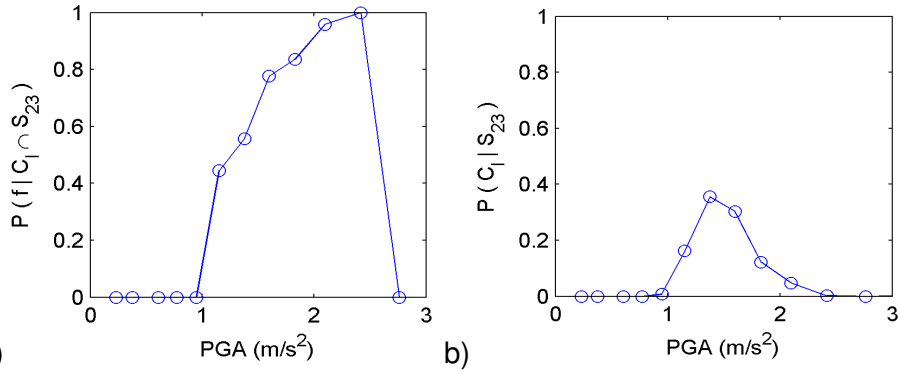
$$252 \quad P(f|C_l) = \frac{\sum_{i=1}^{N_r} (P(f|C_l \cap S_i) \cdot P(C_l|S_i) \cdot P(S_i))}{\sum_{i=1}^{N_r} (P(C_l|S_i) \cdot P(S_i))} \quad (11)$$

253

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

269

270



271

272 **Figure 6 a) Empirical probability of exceeding the damage state threshold ($\max(|u_{ij}(t)|)$
 273 $= 3.3 \cdot 10^{-3}$ m) based on the synthetics generated based on the acceleration record S_{23} at
 274 the cluster centroids (C_l) b) probability of observing a PGA value in the synthetics
 275 based on acceleration record S_{23}
 276**

277 3.4 Lognormal Curve Adjusted To The Non-Parametric Curve

278 In order to observe potential differences between lognormal fragility curves and the non-
 279 parametric curves estimated with the different approaches herein, a Maximum Likelihood
 280 Estimation of the lognormal cumulative distribution function is employed. The MLE of the
 281 lognormal curve uses the point probabilities constituting the empirical fragility curve based on
 282 the selected IM and corresponding EDP observations. The MLE is performed with Equations
 283 12-15 and the estimated lognormal curve is given by Equation 16.

284

$$285 \quad P(n_l, r_l, C_l) = \frac{n_l!}{r_l!(n_l - r_l)!} \cdot P(f|C_l)^{r_l} \cdot (1 - P(f|C_l))^{n_l - r_l} \quad (12)$$

286

$$287 \quad L = \prod_{l=1}^{N_c} P(n_l, r_l, C_l) \quad (13)$$

288

$$289 \quad \ln(L) = \sum_{l=1}^{N_c} \left[\ln \left(\frac{n_l!}{r_l!(n_l - r_l)!} \right) + r_l \cdot \ln \Phi \left(\frac{\ln(C_l) - \ln(A)}{\beta} \right) + (n_l - r_l) \cdot \ln \left(1 - \Phi \left(\frac{\ln(C_l) - \ln(A)}{\beta} \right) \right) \right] \quad (14)$$

290

$$291 \quad \{\bar{A}, \bar{\beta}\} = \arg \max_{(A, \beta)} (\ln(L)) \quad (15)$$

292

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

294

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

307 4 OPTIMIZATION WITH PARAMETRIC MODELS AVERAGING

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

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

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

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

347 origin is maximized, are expunged from the lists of IM observations per cluster, from which IM
 348 observations are randomly drawn. The third and most important modification concerns the
 349 conditional probability of exceeding the damage state threshold in the case of each process
 350 ($P(f|C_l \cap S_i)$). Instead of the empirical estimation of $P(f|C_l \cap S_i)$, the optimized approach employs
 351 two alternative parametric models. The first model (parametric model 1) assumes that
 352 $P(f|C_l \cap S_i)$ remains constant as a function of the IM, and that it is equal to $P(f|C_{k_l} \cap S_i)$. The
 353 second model (parametric model 2) uses a lognormal curve for $P(f|C_l \cap S_i)$. In the following, the
 354 parametric models 1 and 2 are analyzed.

355 4.1 Parametric Model 1

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

363

$$364 \quad k_l = \underset{(l)}{\arg \max} P(C_l|S_i) \quad i = \{1, \dots, N_r = 96\} \quad l = \{1, \dots, N_c = 20\} \quad (17)$$

365

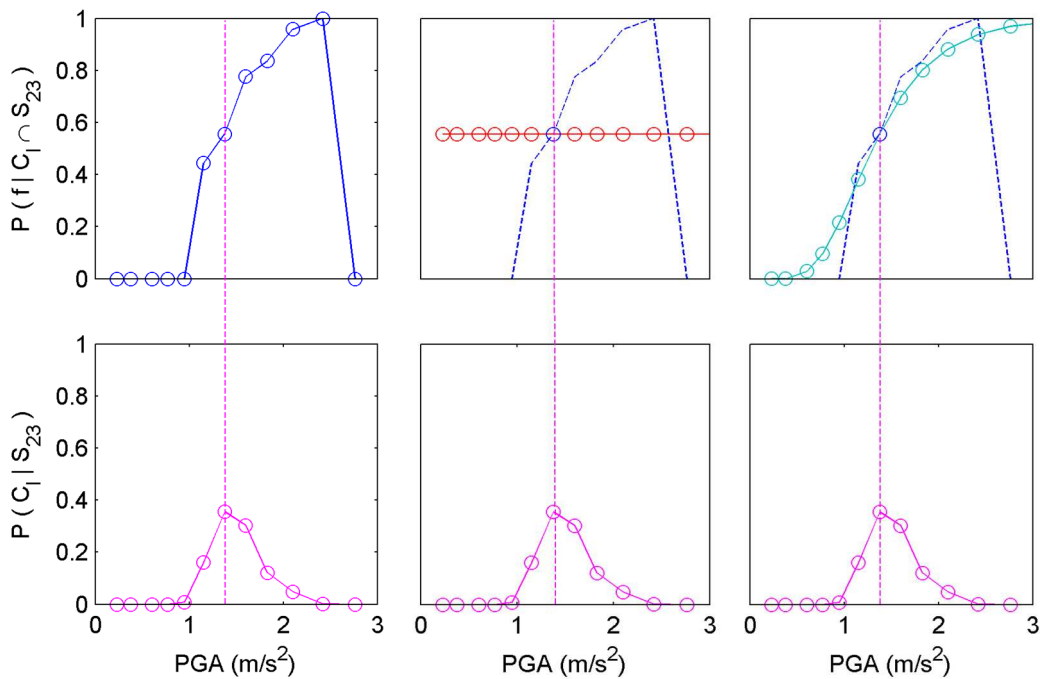
$$366 \quad P_{fi} = P(f|C_l \cap S_i) = P(f|C_{k_l} \cap S_i) \quad i = \{1, \dots, N_r = 96\} \quad l = \{1, \dots, N_c = 20\} \quad (18)$$

367

368 As an example, Figure 7 (top left) shows the empirical conditional probability
 369 $P(f|C_{k_{23}} \cap S_{23})$ estimated with the observations corresponding to the ground motions based
 370 on the 23rd accelerogram record. Moreover, Figure 7 (top middle) shows the corresponding
 371 model used in the optimized approach, which assumes a constant probability (red curve),
 372 which is estimated at the cluster IM centroid for which $P(C_l|S_{23})$ is maximized (Figure 7

373 bottom). When employing parametric model 1, an error is introduced with respect to $P(f|C_i \cap S_i)$.
 374 In specific, $P(f|C_i \cap S_i)$ is under- and overestimated at IM cluster centroids where $C_i > C_k$ and C_i
 375 $< C_k$, respectively. The extent to which $P(f|C_i \cap S_i)$ is under- or overestimated varies, and
 376 generally increases with the distance between C_i and C_k . However, the introduced error is
 377 mitigated by the fact that Equation 11 computes the product $P(f|C_i \cap S_i) \cdot P(C_i|S_i)$. The further C_i
 378 is found from C_k , the smallest the introduced error, because $P(C_i|S_i)$ diminishes with the
 379 distance from C_k (e.g. Figure 7 bottom). Moreover, the fact that $P(f|C_i \cap S_i)$ is simultaneously
 380 under- and overestimated (e.g. Figure 7 middle) at $C_i > C_k$ and $C_i < C_k$, respectively, also
 381 mitigates the global error in the estimation of the fragility curve, as the underestimation on one
 382 side balances to some extent the overestimation on the other.

383



384

385 **Figure 7 Top left: Empirical fragility curve based on the ground motions originating**
 386 **from acceleration record 23. Top middle: parametric model 1 (constant probability of**
 387 **damage state threshold exceedance). Top right: parametric model 2 (lognormal model)**
 388 **Bottom: conditional probability $P(C_i|S_{23})$.**
 389

390 4.2 Parametric Model 2

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

405

$$406 \quad A_i = \exp\left(\ln(C_{k_i}) - \beta_i \cdot \Phi^{-1}\left(P(f|C_{k_i} \cap S_i)\right)\right) \quad \beta_i = \beta \quad (19)$$

407

$$408 \quad P(f|C_i \cap S_i) = \Phi\left(\frac{\ln(C_i) - \ln(A_i)}{\beta_i}\right) \quad (20)$$

409

410 As an example, Figure 7 (top right) shows $P(f|C_i \cap S_{23})$ as estimated with the lognormal
 411 parametric model (cyan curve). In this case, the model approximates well the empirical
 412 estimation of the probability of exceeding the damage state threshold. This figure illustrates
 413 that the largest differences between the probabilities given by the model and the empirical
 414 estimation are observed where $P(C_i|S_i)$ is close to zero. However, the empirical probabilities
 415 $P(C_i|S_i)$ are equal to zero at the IM centroids of clusters without any IM observations.

416 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
417 zero, which means that no error is introduced at these clusters due to the use of a parametric
418 model for $P(f|C_i \cap S_i)$. As shown in the following, parametric model 2 is particularly necessary
419 when the dispersion of the lognormal optimized fragility curve is small (approximately for $\beta <$
420 0.3). In such cases, we consider justified to impose a common dispersion on all the parametric
421 models corresponding to the processes S_i .

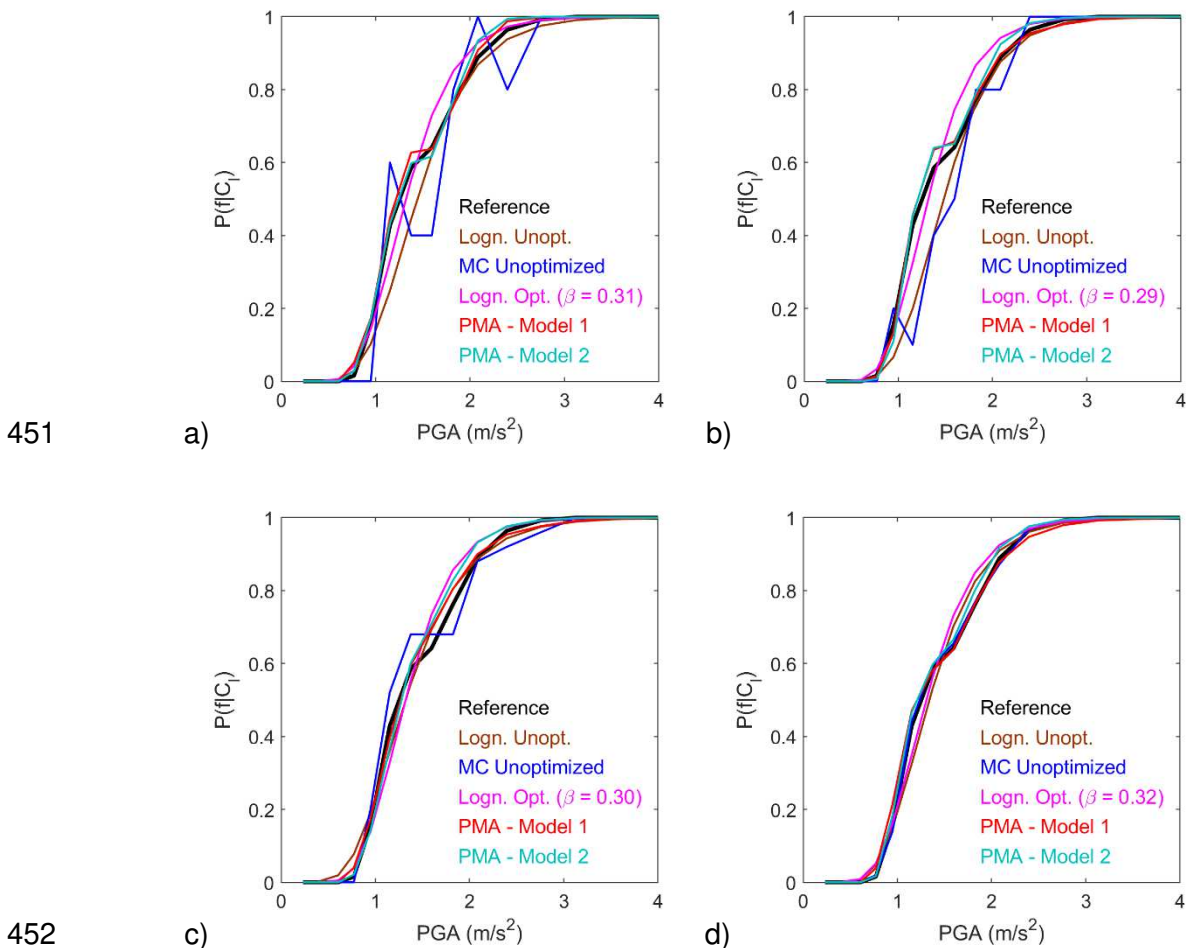
422 **5 APPLICATION OF THE METHODOLOGY**

423 In order to offer insight to the wider field of application of the developed methodology,
424 which is essentially a MC procedure, we use it to compute the fragility curves in three cases:
425 (i) inelastic oscillator without structural uncertainties and ground motions selected at random
426 from the data set of all recorded and synthetic ground motions, (ii) inelastic oscillator without
427 structural uncertainties and ground motions resulting from scaling of a single recorded
428 accelerogram, (iii) inelastic oscillator with structural uncertainties. Based on the results of
429 these three cases, we make our recommendations for practice. In the third case, the fragility
430 curves are derived using as IM the PGA and the spectral acceleration at the frequency of the
431 oscillator ($S_a(5 \text{ Hz})$). To evaluate the effectiveness of the optimized procedures, we are
432 comparing the estimated fragility curves with the reference curve and the 95 % CI according
433 to the different approaches. The CI are computed based on bootstrap resampling [33] with a
434 different set of 500 samples for each case.

435 **5.1 Structural Model Without Uncertainties And Data Set Of Acceleration Records**

436 The developed PMA-based optimization is firstly applied it in the case of the inelastic
437 oscillator employed previously in the description of the methodology (Figure 4a). The ground
438 motion data set used consists of the 96 recorded accelerograms and the 48,000
439 corresponding synthetic ground motions generated with the described procedure in section 2.
440 The fragility curves are computed as a function of the PGA, for a damage state threshold

441 defined by a maximum oscillator displacement of $3.3 \cdot 10^{-3}$ m, and according to the different
 442 approaches are shown in Figure 8 in the case of 100, 200, 500, and 10,000 analyses,
 443 respectively. The curves *MC un-optimized* and *lognormal un-optimized* are computed based
 444 on the same set of seismic response analyses, which is different from the set of analyses used
 445 for the optimized non-parametric curves. Every set of seismic response analyses is performed
 446 using a different and randomly selected set of ground motions according to the optimized or
 447 un-optimized approaches. Additionally, the reference non-parametric fragility curve, which is
 448 estimated based on 48096 analyses with all recorded and synthetic accelerograms, is included
 449 in the figures in order to observe any potential statistical error or bias in the evaluated curves.
 450

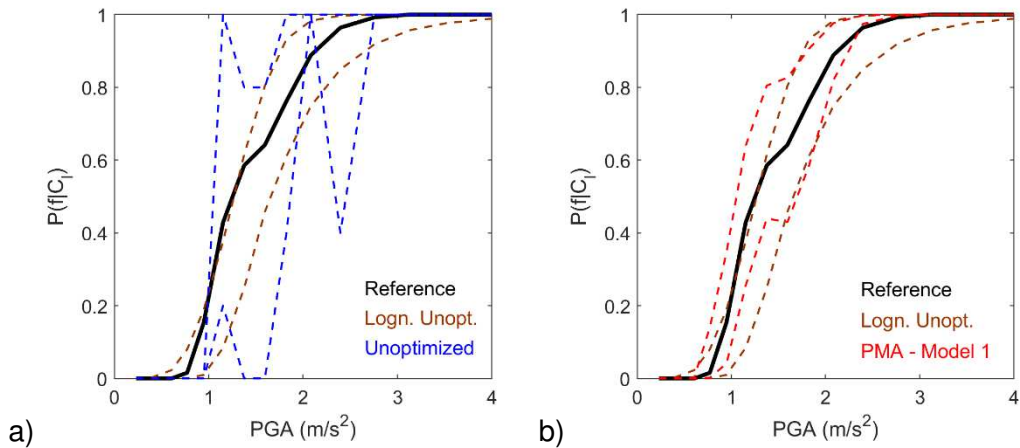


453 **Figure 8 Fragility curves for maximum oscillator displacement ($\max(|u_{ij}(t)|)$) threshold**
 454 **of $3.3 \cdot 10^{-3}$ m evaluated without structural uncertainties and with the enriched ground**
 455 **motion data set based all considered records and based on a) 100 b) 200 c) 500 and d)**
 456 **10,000 analyses, and the reference non-parametric fragility curve**
 457

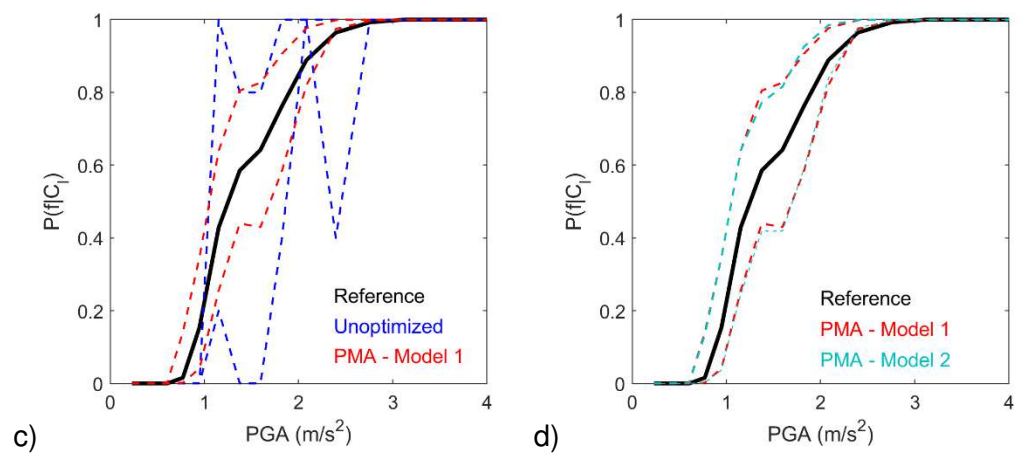
458 In the case of 100, 200 and 500 seismic response analyses (Figure 8a-c), the
459 differences between the reference and the rest fragility curves is primarily due to error of
460 estimation. However, in the case of 10,000 analyses, the difference is rather due to a bias in
461 the computation, given that the fragility curves are evaluated with a very large number of
462 analyses. As far as the MC un-optimized and PMA curves are concerned, they practically
463 converge with the reference curve as the number of analyses increases, which means that no
464 bias is introduced due to the assumptions in this case. In Figure 8d, we observe differences
465 between the reference and the lognormal un-optimized curve based on 10,000 analyses.
466 Given the number of analyses, we consider that the lognormal curve is biased. More important
467 differences between lognormal and non-parametric curves may be observed, when –among
468 other reasons– the studied structures are more complex than a single-degree-of-freedom
469 oscillator, as in [15]. As a measure of the estimation error, the 95 % CI of the fragility curve
470 based on 100 analyses and according to the different approaches are shown in Figure 9. As
471 expected, the MC un-optimized approach gives a poor estimation (Figure 9a) due to the small
472 amount of data and the lognormal un-optimized is more effective. The CI of the curves
473 according to the lognormal un-optimized and the PMA – Model 1 approaches appear to be
474 equivalent (Figure 9b). However, the median lognormal un-optimized curve may converge
475 towards a biased estimation (e.g. Figure 8). Therefore, its confidence interval is not
476 necessarily representative of the goodness of the estimation. This is a weakness of the
477 parametric models and it is beforehand unknown whether there is bias in the fragility curve in
478 complex cases. We observe that the PMA – Model 1 approach results in CI which are
479 significantly smaller than the CI according to the MC un-optimized approach (Figure 9c).
480 According to the curves in Figure 9d, we conclude that the two forms of the PMA optimization
481 are equally effective in this case.

482

483



484



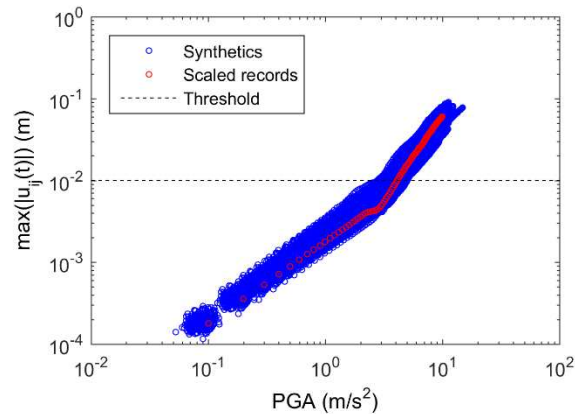
485 **Figure 9 95 % confidence intervals of fragility curves for maximum oscillator**
 486 **displacement ($\max(|u_{ij}(t)|)$) threshold of $3.3 \cdot 10^{-3}$ m evaluated without structural**
 487 **uncertainties and with the enriched ground motion data set based all considered**
 488 **records and based on 100 analyses, and the reference non-parametric fragility curve**
 489

490 **5.2 Structural Model Without Uncertainties And Data Set Of A Multiply Scaled**
 491 **Acceleration Record**

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

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

502

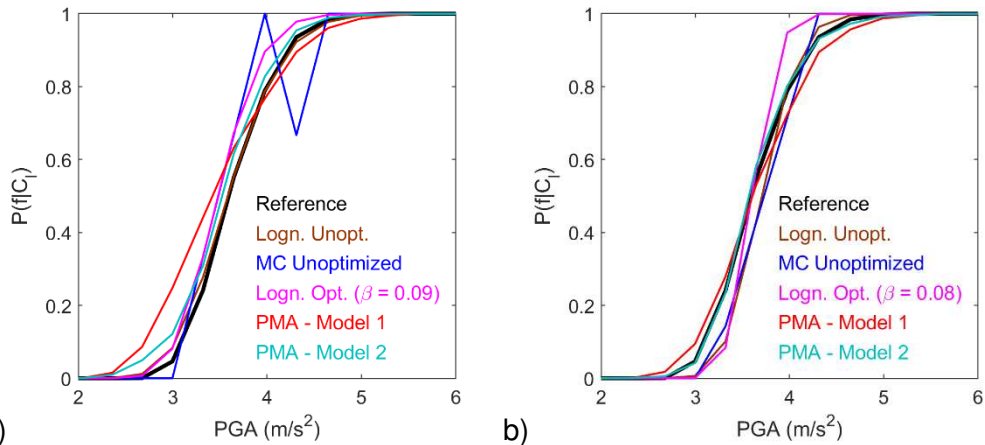


503

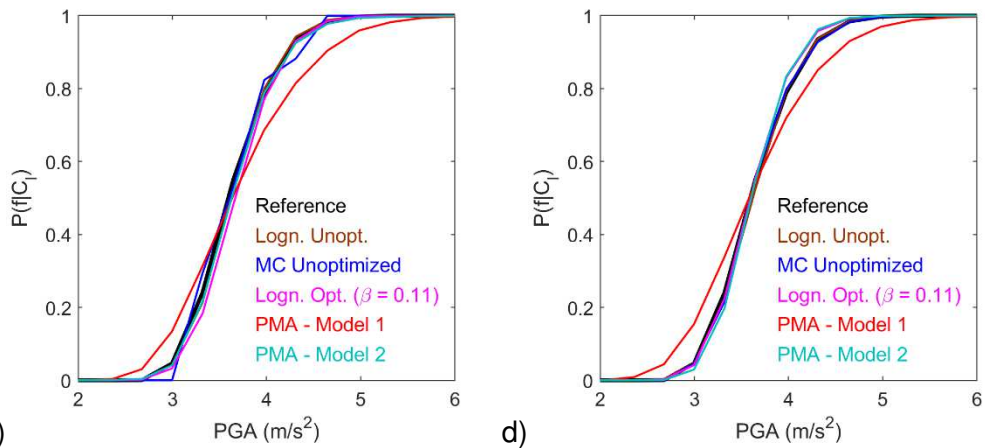
504 **Figure 10 Maximum oscillator displacement ($\max(|u_{ij}(t)|)$) as a function of PGA**
505 **computed without structural uncertainties and with the enriched ground motion data**
506 **set based on scaling of a single randomly selected recorded accelerogram**

507

508 The fragility curves for a damage state threshold defined by a maximum oscillator
509 displacement of $1.0 \cdot 10^{-2}$ m according to the different considered approaches using 100, 200,
510 500, and 10,000 seismic response analyses are shown in Figure 11. In this case, all fragility
511 curves converge to the reference with the exception of the PMA – Model 1 curve, which is
512 based on the optimization assuming models of constant $P(f|C_i \cap S_i)$ per process. It is concluded
513 that PMA model 1 produces a biased curve contrary to PMA model 2 (Figure 11d).
514 Nevertheless, PMA model 2 may also result in bias in other cases (not shown here), when the
515 dispersion of the unoptimized lognormal curve is very small ($\beta < 0.1$). Indeed, in such cases,
516 the reference fragility curve tends towards a step function, which cannot be approximated by
517 the PMA-based procedures presented here unless a finer IM discretization is considered. It
518 should also be taken into account that the observed difference between the reference curve
519 and the curves according to the different approaches in the case of 100 and 200 seismic
520 response analyses is principally an estimation error due to the limited number of seismic
521 response analyses used.



523



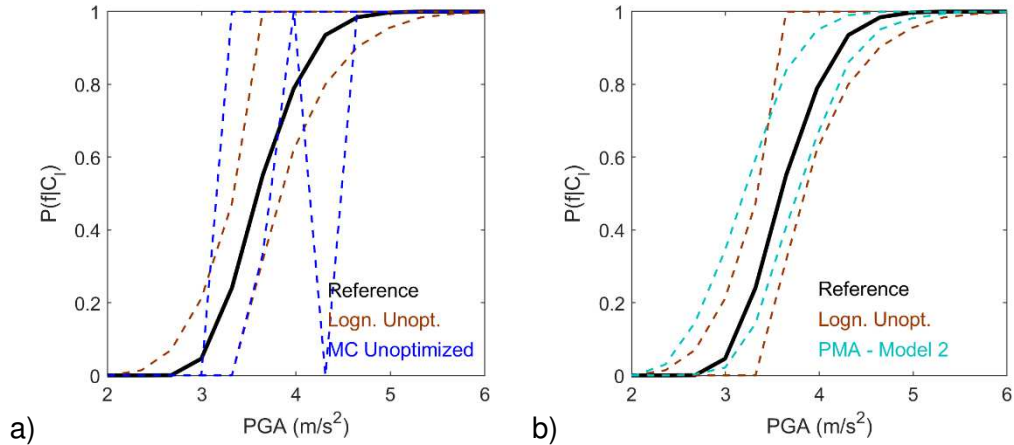
524

525 **Figure 11 Fragility curves for maximum oscillator displacement ($\max(|u_{ij}(t)|)$) threshold**
 526 **of $1.0 \cdot 10^{-2}$ m evaluated without structural uncertainties and with the enriched ground**
 527 **motion data set based on scaling of acceleration record 27 and based on a) 100 b) 200**
 528 **c) 500 and d) 10,000 analyses, and the reference non-parametric fragility curve**
 529

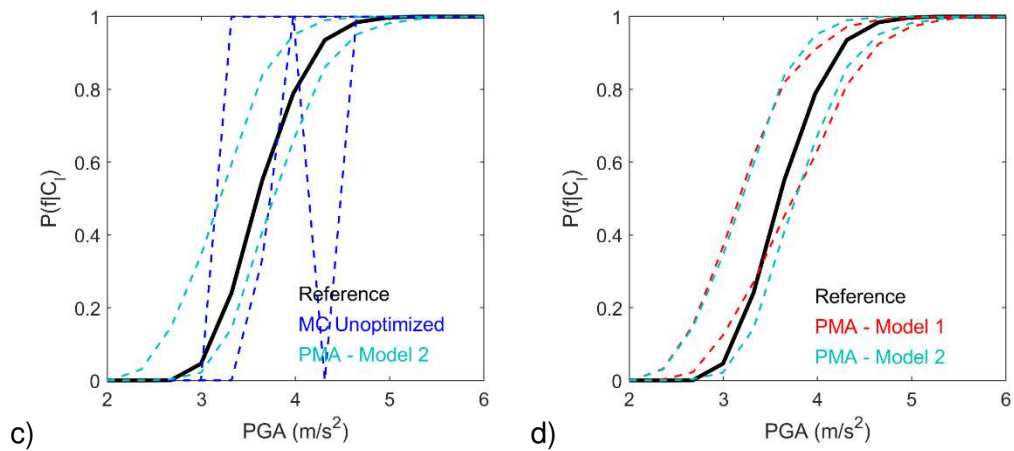
530 Figure 12 includes the 95 % CI of the fragility curves based on 100 seismic response
 531 analyses. Also in this case, the lognormal un-optimized is more effective than the MC un-
 532 optimized. The CI of the lognormal un-optimized and PMA – Model 2 curves indicate that both
 533 approaches are effective in this case, with PMA – Model 2 being slightly better. Once more,
 534 the estimation error according to the PMA – Model 2 approach is significantly less than the
 535 error in the case of the MC un-optimized computation with 100 analyses. Contrary to the CI of
 536 PMA – Model 1 curve, the CI of the PMA – model 2 curve envelopes the reference and
 537 indicates that this is the preferable approach in this case.

538

539



540



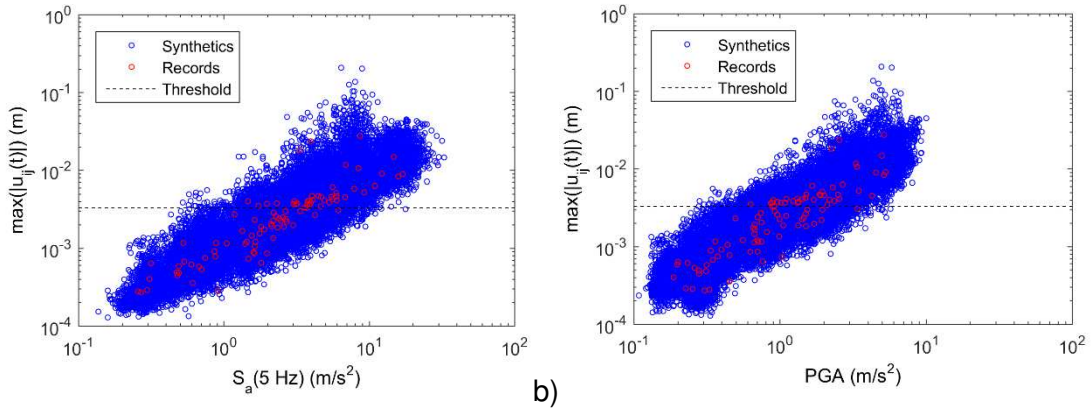
541 **Figure 12 95 % confidence intervals of fragility curves for maximum oscillator**
542 **displacement ($\max(|u_{ij}(t)|)$) threshold of $1.0 \cdot 10^{-2}$ m evaluated without structural**
543 **uncertainties and with the enriched ground motion data set based on scaling of**
544 **acceleration record 27 and based on 100 analyses, and the reference non-parametric**
545 **fragility curve**
546

547 5.3 Structural Model With Uncertainties And Data Set Of Acceleration Records

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

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

558



559

a)

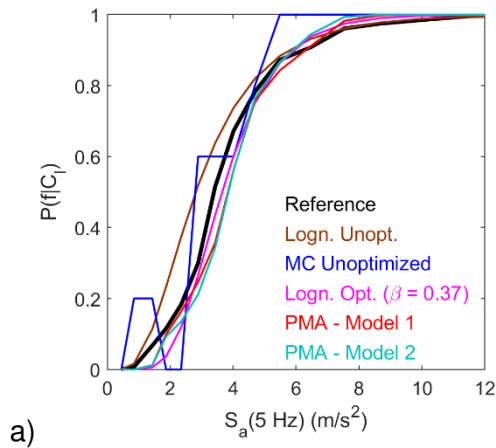
b)

560 **Figure 13 Maximum oscillator displacement ($\max(|u_{ij}(t)|)$) as a function of a) the spectral**
 561 **acceleration at 5 Hz and b) the PGA in the case of the oscillator with frequency and**
 562 **yield displacement uncertainty**
 563

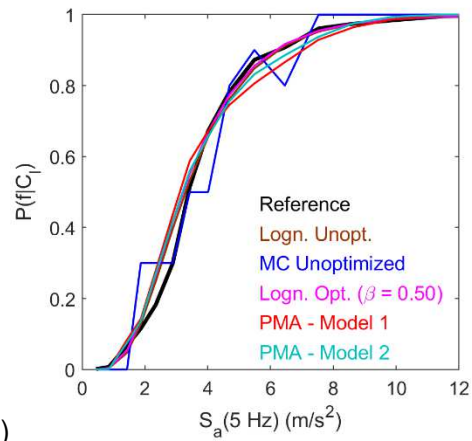
564 The fragility curves for a damage state threshold of $3.3 \cdot 10^{-3}$ m maximum oscillator
 565 displacement as a function of $S_a(5\text{ Hz})$ and PGA are shown in Figure 14 and 15. As expected,
 566 (i) the introduction of uncertainties leads to increase of the dispersion of the lognormal fragility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

573

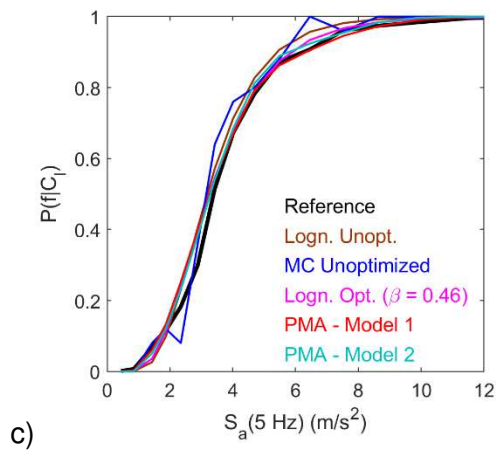
574



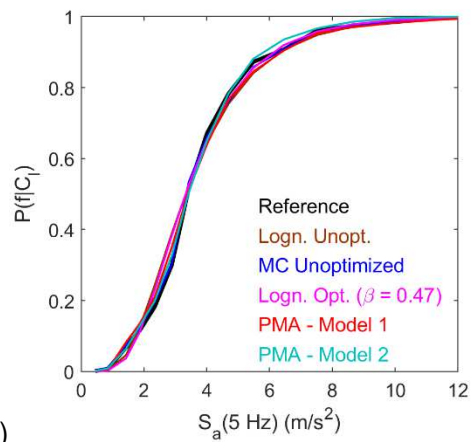
b)



575

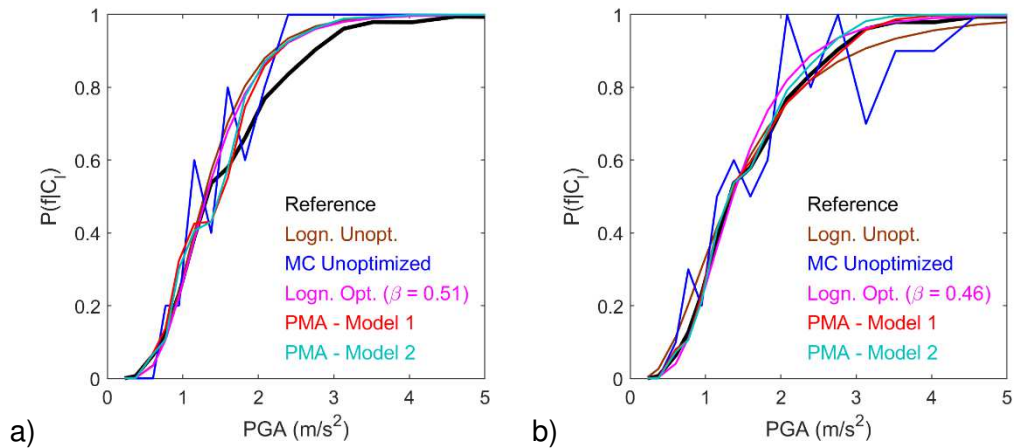


d)

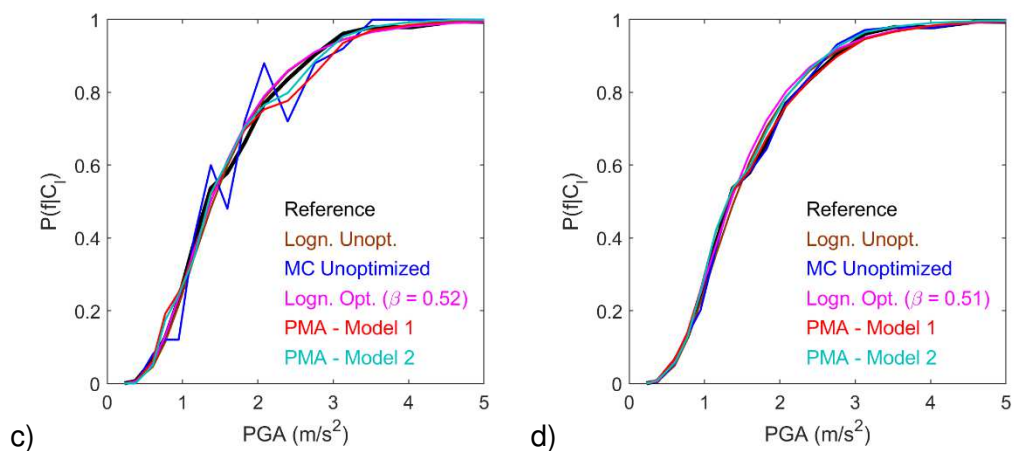


576 **Figure 14 Spectral acceleration ($S_a(5\text{ Hz})$)-based fragility curves for maximum oscillator**
577 **displacement ($\max(|u_{ij}(t)|)$) threshold of $3.3 \cdot 10^{-3}$ m evaluated with structural**
578 **uncertainties and with the enriched ground motion data set based all considered**
579 **records and based on a) 100 b) 200 c) 500 and d) 10,000 analyses, and the reference**
580 **non-parametric fragility curve**
581

582



583



584 **Figure 15 PGA-based fragility curves for maximum oscillator displacement ($\max(|u_{ij}(t)|)$)**
585 **threshold of $3.3 \cdot 10^{-3}$ m evaluated with structural uncertainties and with the enriched**
586 **ground motion data set based all considered records and based on a) 100 b) 200 c) 500**
587 **and d) 10,000 analyses, and the reference non-parametric fragility curve**
588

589 6 CONCLUSION

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

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

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

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

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

652 the resources for 10,000 or more analyses are available, the MC un-optimized approach can
653 be used. In the area between 100 and 10,000 analyses, which is of practical interest, we
654 suggest either a parametric curve, or a non-parametric optimized fragility curve computation
655 with one of the two proposed alternatives. In this area, the dispersion of the optimized
656 lognormal curve fitted to the selected data dictates the optimal approach. In the case of a large
657 dispersion ($0.3 \leq \beta$), the optimization with the constant probability of damage threshold
658 exceedance per process is sufficient, while in the case of a limited dispersion ($0.1 \leq \beta < 0.3$),
659 the optimization with the lognormal model per process is recommended. When PMA – Model
660 1 and 2 use a large number of seismic response analyses and give drastically different results
661 (as in the case with an original data set consisting of ground motions resulting from scaling a
662 single acceleration record), PMA – Model 2 should be preferred, unless the dispersion of the
663 associated optimized lognormal curve is very small ($\beta < 0.1$). In such cases, the presented
664 PMA approaches are not efficient and a lognormal model for the fragility curve is
665 recommended.

666

667 **Table 1 Recommended type of fragility curve based on the number of seismic analyses**
668 **(N) and the dispersion of the lognormal (un-optimized) curve fitted to the empirical non-**
669 **parametric curve (β)**

670

	$\beta < 0.1$	$0.1 \leq \beta < 0.3$	$0.3 \leq \beta$
$N < 100$	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

671

672 **Our procedure has also been applied in the case of a realistic finite element model of a low-**
673 **rise reinforced concrete bare frame (modelling details may be found in [34]), not presented**
674 **here for the sake of brevity. The results lead to the same conclusions.** Should one attempt to
675 apply the procedure herein in the case of complex structures, they will face a series of
676 challenges, which are, however, not specific to our methodology. A major concern would be
677 the selection of an efficient IM. IMs are considered efficient [35], when the seismic response

678 as a function of the IM has a low dispersion. However, scalar IMs are not efficient in every
679 case. For example, the spectral acceleration at the first eigenfrequency is a common scalar
680 IM, which is efficient in the case of structures, whose response is mostly affected by their first
681 mode. However, it is not efficient in the case of tall buildings [36]. In the case of structures with
682 multiple degrees of freedom the use of more adapted IMs, or even a vector of different IMs
683 [37] may be a solution. That said, further investigations should be made to see if the
684 procedures herein can be modified to use a vector IM. Although, the procedure herein is—in
685 principle— independent of the selected IM and the damage state, it should be adapted to more
686 severe damage states such as collapse. Indeed, the simulation of severe damage states may
687 be computationally demanding and may require to take into account P-delta effects [38], to
688 simulate brittle failure modes [39] and consider alternative IMs [40]. In addition, a validation of
689 our methodology with a very large number of seismic response analyses in the case of
690 complex structures has a prohibitive computational cost. To test the usefulness of our
691 procedure in the case of complex structures, fragility curves given by our procedure based on
692 a reasonable number of seismic response analyses (e.g. a few hundred) could be compared
693 with curves given by other procedures, which reduce the computational cost. Such procedures
694 may rely, amongst others, on metamodeling strategies based on neural networks [41], or
695 support vector machines [42]. **Finally, further studies of the developed procedure using
696 realistic structural models and fragility curves conditioned on failure, instead of curves
697 conditioned on an engineering demand parameter threshold, should provide additional
698 insights.**

699 **ACKNOWLEDGMENTS**

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

703 **REFERENCES**

- 704 [1] Fragiadakis M, Vamvatsikos D, Karlaftis MG, Lagaros ND, Papadrakakis M. Seismic
705 assessment of structures and lifelines. *J Sound Vib* 2015;334:29–56.
706 doi:10.1016/j.jsv.2013.12.031.
- 707 [2] Yang TY, Moehle J, Stojadinovic B, Der Kiureghian A. Seismic Performance Evaluation
708 of Facilities: Methodology and Implementation. *J Struct Eng* 2009;135:1146–54.
709 doi:10.1061/(ASCE)0733-9445(2009)135:10(1146).
- 710 [3] FEMA. Seismic Performance Assessment of Buildings – Volume 1 – Methodology.
711 WASHINGTON, DC: 2012.
- 712 [4] Baker JW. Efficient Analytical Fragility Function Fitting Using Dynamic Structural
713 Analysis. *Earthq Spectra* 2015;31:579–99. doi:10.1193/021113EQS025M.
- 714 [5] Silva V, Crowley H, Bazzurro P. Exploring Risk-Targeted Hazard Maps for Europe.
715 *Earthq Spectra* 2016;32:1165–86. doi:10.1193/112514EQS198M.
- 716 [6] Berge-Thierry C, Svay A, Laurendeau A, Chartier T, Perron V, Guyonnet-Benaize C,
717 et al. Toward an integrated seismic risk assessment for nuclear safety improving current
718 French methodologies through the SINAPS@ research project. *Nucl Eng Des* 2017;323:185–
719 201. doi:10.1016/j.nucengdes.2016.07.004.
- 720 [7] Tsionis G, Mignan D, Pinto A, Giardini D, European Commission, Joint Research
721 Centre. Harmonized approach to stress tests for critical infrastructures against natural
722 hazards. Luxembourg: Publications Office; 2016.
- 723 [8] Zhang J, Huo Y. Evaluating effectiveness and optimum design of isolation devices for
724 highway bridges using the fragility function method. *Eng Struct* 2009;31:1648–60.
725 doi:10.1016/j.engstruct.2009.02.017.
- 726 [9] Saha SK, Matsagar V, Chakraborty S. Uncertainty quantification and seismic fragility
727 of base-isolated liquid storage tanks using response surface models. *Probabilistic Eng Mech*
728 2016;43:20–35. doi:10.1016/j.probengmech.2015.10.008.

- 729 [10] Patil A, Jung S, Kwon O-S. Structural performance of a parked wind turbine tower
730 subjected to strong ground motions. *Eng Struct* 2016;120:92–102.
731 doi:10.1016/j.engstruct.2016.04.020.
- 732 [11] Gidaris I, Taflanidis AA, Mavroeidis GP. Kriging metamodeling in seismic risk
733 assessment based on stochastic ground motion models: Seismic Risk Assessment Through
734 Kriging Metamodeling. *Earthq Eng Struct Dyn* 2015;44:2377–99. doi:10.1002/eqe.2586.
- 735 [12] Parolai S, Haas M, Pittore M, Fleming K. Bridging the Gap Between Seismology and
736 Engineering: Towards Real-Time Damage Assessment. In: Pitilakis K, editor. *Recent Adv.*
737 *Earthq. Eng. Eur.*, vol. 46, Cham: Springer International Publishing; 2018, p. 253–61.
738 doi:10.1007/978-3-319-75741-4_10.
- 739 [13] Quilligan A, O'Connor A, Pakrashi V. Fragility analysis of steel and concrete wind
740 turbine towers. *Eng Struct* 2012;36:270–82. doi:10.1016/j.engstruct.2011.12.013.
- 741 [14] D'Ayala D, Meslem A, Vamvatsikos D, Porter K, Rossetto T, Silva V. Guidelines for
742 Analytical Vulnerability Assessment - Low/Mid-Rise. GEM; 2015.
- 743 [15] Vamvatsikos D, Cornell CA. Incremental dynamic analysis. *Earthq Eng Struct Dyn*
744 2002;31:491–514. doi:10.1002/eqe.141.
- 745 [16] Jalayer F, Cornell CA. Alternative non-linear demand estimation methods for
746 probability-based seismic assessments. *Earthq Eng Struct Dyn* 2009;38:951–72.
747 doi:10.1002/eqe.876.
- 748 [17] Zentner I. A general framework for the estimation of analytical fragility functions based
749 on multivariate probability distributions. *Struct Saf* 2017;64:54–61.
750 doi:10.1016/j.strusafe.2016.09.003.
- 751 [18] Mai C, Konakli K, Sudret B. Seismic fragility curves for structures using non-parametric
752 representations. *Front Struct Civ Eng* 2017;11:169–86. doi:10.1007/s11709-017-0385-y.
- 753 [19] Noh HY, Lallemand D, Kiremidjian AS. Development of empirical and analytical fragility
754 functions using kernel smoothing methods: DEVELOPMENT OF FRAGILITY FUNCTIONS
755 USING KERNEL SMOOTHING METHODS. *Earthq Eng Struct Dyn* 2015;44:1163–80.
756 doi:10.1002/eqe.2505.

- 757 [20] Lallemand D, Kiremidjian A, Burton H. Statistical procedures for developing earthquake
758 damage fragility curves: STATISTICAL PROCEDURES FOR DAMAGE FRAGILITY
759 CURVES. *Earthq Eng Struct Dyn* 2015;44:1373–89. doi:10.1002/eqe.2522.
- 760 [21] Karamlou A, Bocchini P. Computation of bridge seismic fragility by large-scale
761 simulation for probabilistic resilience analysis: BRIDGE SEISMIC FRAGILITY BY LARGE-
762 SCALE SIMULATION FOR RESILIENCE. *Earthq Eng Struct Dyn* 2015;44:1959–78.
763 doi:10.1002/eqe.2567.
- 764 [22] Porter K, Kennedy R, Bachman R. Creating Fragility Functions for Performance-Based
765 Earthquake Engineering. *Earthq Spectra* 2007;23:471–89. doi:10.1193/1.2720892.
- 766 [23] Jain AK, Murty MN, Flynn PJ. Data clustering: a review. *ACM Comput Surv*
767 1999;31:264–323. doi:10.1145/331499.331504.
- 768 [24] Rezaeian S, Der Kiureghian A. Simulation of synthetic ground motions for specified
769 earthquake and site characteristics. *Earthq Eng Struct Dyn* 2010:n/a-n/a.
770 doi:10.1002/eqe.997.
- 771 [25] Zentner I, Poirion F. Enrichment of seismic ground motion databases using Karhunen-
772 Loève expansion. *Earthq Eng Struct Dyn* 2012;41:1945–57. doi:10.1002/eqe.2166.
- 773 [26] Ambraseys N, Smit P, Sigbjornsson R, Suhadolc P, Margaris B. Internet-Site for
774 European Strong-Motion Data. European Commission, Research-Directorate General,
775 Environment and Climate Programme; 2002.
- 776 [27] Ambraseys N, Smit P, Douglas J, Margaris B, Sigbjornsson R, Olafsson S, et al.
777 Internet site for European strong-motion data. *Boll Geofis Teor E Appl* 2004;45:113–29.
- 778 [28] Boore DM. Simulation of Ground Motion Using the Stochastic Method. *Pure Appl*
779 *Geophys* 2003;160:635–76. doi:10.1007/PL00012553.
- 780 [29] Housner G, Jennings P. Generation of Artificial Earthquakes. *J Eng Mech Div*
781 1964;90:113–52.
- 782 [30] Rodolfo Saragoni G, Hart GC. Simulation of artificial earthquakes. *Earthq Eng Struct*
783 *Dyn* 1973;2:249–67. doi:10.1002/eqe.4290020305.

- 784 [31] Kostinakis K, Fontara I-K, Athanatopoulou AM. Scalar Structure-Specific Ground
785 Motion Intensity Measures for Assessing the Seismic Performance of Structures: A Review. *J*
786 *Earthq Eng* 2018;22:630–65. doi:10.1080/13632469.2016.1264323.
- 787 [32] MATLAB. MathWorks; 2015.
- 788 [33] Iervolino I. Assessing uncertainty in estimation of seismic response for PBEE. *Earthq*
789 *Eng Struct Dyn* 2017;46:1711–23. doi:10.1002/eqe.2883.
- 790 [34] Trevelopoulos K, Guéguen P. Period elongation-based framework for operative
791 assessment of the variation of seismic vulnerability of reinforced concrete buildings during
792 aftershock sequences. *Soil Dynamics and Earthquake Engineering* 2016;84:224–37.
793 doi:10.1016/j.soildyn.2016.02.009.
- 794 [35] Luco N, Cornell CA. Structure-Specific Scalar Intensity Measures for Near-Source and
795 Ordinary Earthquake Ground Motions. *Earthq Spectra* 2007;23:357–92.
796 doi:10.1193/1.2723158.
- 797 [36] Jayaram N, Lin T, Baker JW. A Computationally Efficient Ground-Motion Selection
798 Algorithm for Matching a Target Response Spectrum Mean and Variance. *Earthq Spectra*
799 2011;27:797–815. doi:10.1193/1.3608002.
- 800 [37] Kohrangi M, Bazzurro P, Vamvatsikos D. Vector and Scalar IMs in Structural Response
801 Estimation, Part II: Building Demand Assessment. *Earthq Spectra* 2016;32:1525–43.
802 doi:10.1193/053115EQS081M.
- 803 [38] Eads L, Miranda E, Krawinkler H, Lignos DG. An efficient method for estimating the
804 collapse risk of structures in seismic regions: AN EFFICIENT METHOD FOR ESTIMATING
805 THE COLLAPSE RISK OF STRUCTURES. *Earthq Eng Struct Dyn* 2013;42:25–41.
806 doi:10.1002/eqe.2191.
- 807 [39] Kyriakides NC, Pantazopoulou SJ. Collapse Fragility Curves for RC Buildings
808 Exhibiting Brittle Failure Modes. *J Struct Eng* 2018;144:04017207.
809 doi:10.1061/(ASCE)ST.1943-541X.0001920.

810 [40] Eads L, Miranda E, Lignos DG. Average spectral acceleration as an intensity measure
811 for collapse risk assessment: Average Spectral Acceleration as an IM for Collapse Risk
812 Assessment. *Earthq Eng Struct Dyn* 2015;44:2057–73. doi:10.1002/eqe.2575.

813 [41] Wang Z, Pedroni N, Zentner I, Zio E. Seismic fragility analysis with artificial neural
814 networks: Application to nuclear power plant equipment. *Eng Struct* 2018;162:213–25.
815 doi:10.1016/j.engstruct.2018.02.024.

816 [42] Sainct R, Feau C, Martinez J, Garnier J. Efficient Seismic fragility curve estimation by
817 Active Learning on Support Vector Machines (submitted). *Struct Saf* n.d.

818