Available online at www.eccomasproceedia.org
Eccomas Proceedia COMPDYN (2017) 1880-1889

ECCOMAS
Proceedia

COMPDYN 2017
6th ECCOMAS Thematic Conference on
Computational Methods in Structural Dynamics and Earthquake Engineering
M. Papadrakakis, M. Fragiadakis (eds.)
Rhodes Island, Greece, 15–17 June, 2017

UNCERTAINTY IN SEISMIC INTENSITY MEASURES USED FOR FRAGILITY ANALYSIS

Alin Radu[1] **and Mircea Grigoriu**[2]

[1]University of Bristol
Bristol, BS8 1TR, UK
e-mail: alin.radu@bristol.ac.uk

[2] Cornell University
Ithaca, 14853, NY
e-mail: mdg12@cornell.edu

Keywords: Seismic fragility, seismic intensity measures, peak ground acceleration, spectral acceleration, correlation, independence, extreme value theory

Abstract. *Seismic fragility is the main tool used in the performance-based earthquake engineering, and represents the probability of a dynamic system to enter a damage state for a given intensity measure (IM). The most commonly-used seismic IMs are the peak ground acceleration (PGA), which is the absolute maximum of the ground acceleration, and the spectral acceleration (SA), which characterizes well the response of simple linear systems, but shows large uncertainties in the characterization of the seismic response of complex non-linear systems. The main assumption that underlies the basis of the selection of PGA or SA as IMs is that they correlate satisfactorily with the demand parameters (D) of realistic structures, i.e., non-linear, complex dynamic systems. Statistics and tools of random vibrations are employed to quantify the dependence relations between seismic IMs and structural demand parameters, for non-linear systems subjected to seismic ground-motion records.*

Peer-review under responsibility of the organizing committee of COMPDYN 2017.
doi: 10.7712/120117.5534.18210

1 INTRODUCTION

The main tools to characterize the seismic performance of buildings are seismic fragilities in performance-based earthquake engineering, and, similarly, vulnerability curves in the insurance industry. Seismic fragility functions are probabilities that a structure enters specified damage states, while seismic vulnerability functions are ratios of the repair cost to the total replacement cost of a structural system, both as functions of intensity measures (IMs). Traditionally, two intensity measures have been commonly used for constructing fragility functions, the peak ground acceleration (PGA) [1, 2], which is the absolute maximum of the seismic ground acceleration, and the spectral acceleration (SA) [3, 4], which is the absolute maximum of the response acceleration of single-degree-of-freedom (SDOF) linear systems. The limitations of these scalar intensity measures $IM = \{PGA,\ SA\}$ have been noticed before in [5, 6], which emphasize the need for vector-valued IMs.

This paper investigates critically the performance of these intensity measures as proxies for the seismic ground-motion processes and their usefulness in the accurate estimation of the structural response of realistic, complex structural systems. Demand parameters D conditional on IM need to be *efficient*, i.e. conditional random variables $D|IM$ have small variances, and *sufficient*, i.e. random variables $D|IM$ are completely defined for given intensity levels of the IMs. These properties are valid only under the critical assumption that IMs contain enough information about the seismic hazard in order to predict accurately the response of nonlinear, complex dynamic systems. This paper proposes to examine this assumption for the two intensity measures stated above, by using tools used in random vibration theory, and statistical tools such as correlation models, copula models and extreme value theory.

2 PROBLEM DEFINITION

Two simple dynamic systems are used for the purpose of our paper, a linear single-degree-of-freedom (SDOF) described by Eq.(1), and a non-linear SDOF system, known as the Bouc-Wen oscillator, described by Eq.(2).

$$\ddot{X}_{lin}(t) + 2\zeta_0\nu_0\dot{X}_{lin}(t) + \nu_0^2 X_{lin}(t) = A(t) \tag{1}$$

$$\ddot{X}(t) + 2\zeta_0\nu_0\dot{X}(t) + \nu_0^2(\lambda X(t) + (1-\rho)W(t)) = A(t) \tag{2}$$

$$\dot{W}(t) = -\gamma\dot{X}(t) + \alpha|W(t)|^{\eta-1}W(t) + \beta\dot{X}(t)|W(t)|^{\eta},$$

where $X_{lin}(t)$ and $X(t)$ denote the response displacements of the linear and the Bouc-Wen SDOF systems, respectively, subjected to a stochastic process $A(t)$, assumed to model the seismic ground motion. Coefficients ν_0 and ζ_0 are known as the fundamental frequency and damping ratio of the linear SDOF system, and $\alpha,\ \beta,\ \gamma,\ \rho,\ \eta$ define the properties of the non-linear system. For further numerical examples shown the following values have been used for the systems' parameters: $\nu_0 = 2\pi,\ \zeta_0 = 0.05,\ \alpha = 0.5,\ \beta = 5,\ \gamma = 3,\ \rho = 0.1,\ \eta = 1$.

The intensity measures of interest are defined as:

$$PGA = \max_{t\geq 0}|A(t)| \tag{3}$$

$$SA = \nu_0^2 \max_{t\geq 0}|X_{lin}(t)|, \tag{4}$$

where PGA is the peak ground acceleration and SA is the pseudo spectral-acceleration calculated at frequency $\nu_0 = 2\pi$ (or in other words at period $T = 1s$), and damping ratio $\zeta = 5\%$. Note that the linear SDOF system in Eq.(1) is only used for the definition of $IM = SA$.

For simplicity's sake, we define the demand parameter for the non-linear system as its maximum displacement $D = \max_{t\geq 0}|X(t)|$. The fragility function can then be expressed as $P_f(im) = \mathbb{P}(D > d_{cr}|IM = im)$, which is the probability that the system enters a critical damage state, defined by the displacement threshold d_{cr}, for a given value of the intensity measure $IM = im$, where $IM \in \{PGA, SA\}$.

Intuition would suggest that PGA, SA and D are weakly dependent since they are obtained from stochastic processes $A(t)$, X_{lin} and $X(t)$, respectively, which have different properties as X_{lin} and $X(t)$ are solutions of the linear and non-linear differential equations with random input $A(t)$. If this intuition is correct, then fragilities as functions of $IM \in \{PGA, SA\}$ have large uncertainties and may have limited use in practice. The same conclusion can be drawn about vulnerability functions under the sensible assumption that damage ratios are strongly dependent on the demand parameters of the structural system. The goal of the paper is to look at the dependence between D and $IM \in \{PGA, SA\}$, which is quantified through several statistical tools. If the dependence between D and PGA, or D and SA is weak, then PGA and SA are unsatisfactory intensity measures and the fragility or vulnerability functions expressed as functions of PGA and SA provide limited information on the seismic structural performance.

A similar study was presented in [7] for a postulated model for the seismic ground motion $A(t)$ based on a seismological model. However, for the relevance of the current discussion, real ground motion records are used instead of samples of the stochastic process $A(t)$. The dataset used is the NGA West, a large dataset of ground-motion records from Western United States *NGA West*, described in [8] and available from the Pacific Earthquake Engineering Research Center (PEER) [9]. The dataset contains $n = 3,465$ records and is presented in Figure 1 as a scatter plot of their moment magnitudes versus their epicentral distances expressed in km.

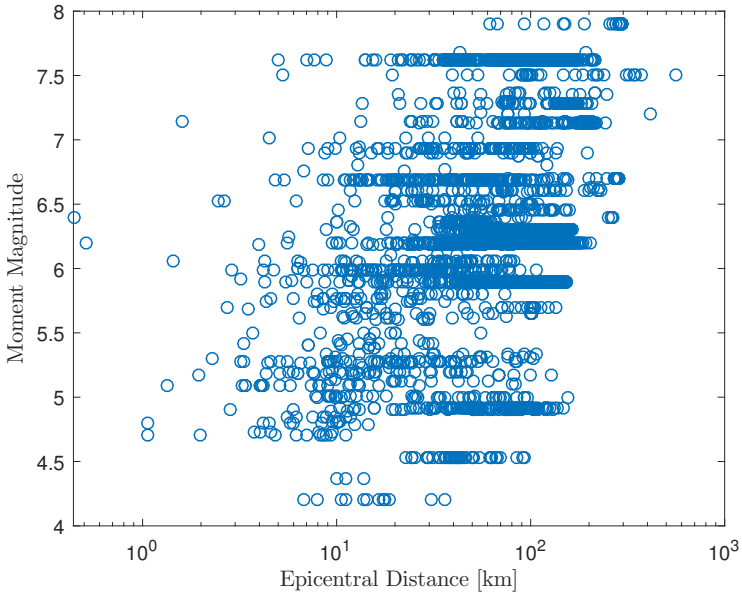

Figure 1: Ground motion records in the NGA West Dataset from PEER by moment magnitude and epicentral distance

Samples of the random variables PGA, SA, and D are calculated for each ground motion in the dataset, respectively, and are used for the investigations further described in the following sections.

3 PERFORMANCE OF INTENSITY MEASURES

The dependence between the demand parameters D of the complex dynamic systems, in this case represented by the non-linear system in Eq.(2), and the two IMs, i.e. $IM = \{PGA,\ SA\}$, is studied by looking at three metrics: (1) the correlation, (2) the overall dependence, and (3) the tail dependence, i.e. simultaneous high values of D and IM.

3.1 Correlation

Correlation coefficients are attractively simple to calculate but they do not provide sufficient information on the dependence of random variables. For example, the normal standard random variable $X \sim N(0,1)$ and the non-Gaussian random variable $Y = X^2$ with mean 1 and variance 2 are not independent, since Y is known given X. Yet, they are uncorrelated since the correlation coefficient $\rho = Cov(X,Y)/\sigma_X\sigma_Y = 0$. However, correlation is still an acceptable measure of the linear relationship between two random variables, and correlation coefficients for the vectors (D, PGA) and (D, SA) are calculated, respectively.

3.1.1 Correlation of D and PGA

The left panel of Figure 2 shows a scatter plot of the n samples of the demand parameter for the linear system, i.e. $\max_{t\geq 0}|X_{lin}(t)|$ against the PGA, while the middle and right panels of the same figure show the demand parameter D defined previously for the Bouc-Wen oscillator plotted against PGA, for all n, and for the top $m = 100$ samples, respectively. Since the focus

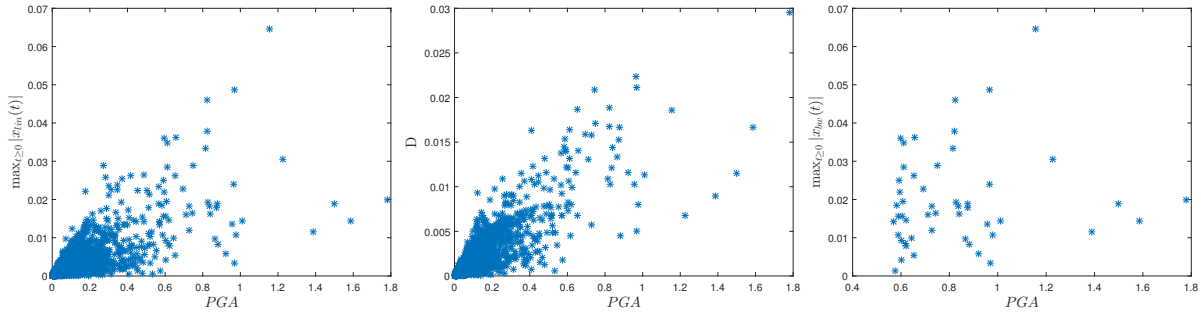

Figure 2: Scatter plot of $n = 3,465$ samples of $(PGA, \max_{t\geq 0}|X_{lin}(t)|)$.(left), $n = 3,465$ samples of (PGA, D) (middle), and top $m = 100$ samples of (PGA, D)(right).

of the paper is on the performance of the IMs with respect to complex, non-linear systems, the discussion will focus around the Bouc-Wen oscillator and the demand parameter D in relation with IM. In this case the correlation coefficient for (PGA, D) for the entire dataset is $\rho_{PGA,D} = 84.44\%$, and it decreases to 53.52% and 32.56% for the top 100 and 50 samples. Thus, the correlation between PGA and D decreases significantly as we look into the tail of (PGA, D).

3.1.2 Correlation of D and SA

Similar to the previous section, correlation between the demand parameters and the intensity measure SA is investigated. The left panel of Figure 3, which shows the scatter plot of the $n = 3,465$ samples of the maximum response of the linear system $\max_{t\geq 0}|X_{lin}(t)|$ against SA, is in accordance with findings of [6] that SA, unlike PGA, is a satisfactory IM for SDOF linear systems. The middle and right panels of Figure 3 show the scatter plots of (SA, D)

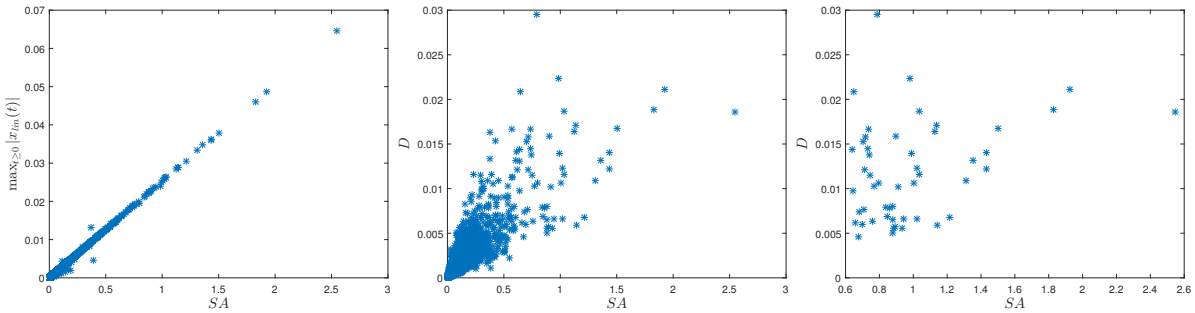

Figure 3: Scatter plot of $n = 3,465$ samples of $(SA, \max_{t\geq 0}|X_{lin}(t)|)$.(left), $n = 3,465$ samples of (SA, D) (middle), and top $m = 100$ samples of (SA, D)(right).

for the entire number n and the top $m = 100$ samples of the dataset, respectively. However, similar observation as before are noticed for the case of the non-linear system. The correlation coefficient for (SA, D) for the entire dataset of $n = 3,465$ samples is $\rho_{SA,D} = 83.75\%$, and it decreases to 53.69% and further down to 33.00% for the top 100 and 50 samples.

3.2 Overall dependence

The overall dependence of the demand parameter D and the intensity measures IM is tested by using the definition of independence, i.e., $F(im, d) = F_{IM}(im)F_D(d)$, where $F(im, d) = \mathbb{P}(IM \leq im, D \leq d)$ is the joint distribution function of (IM, D) and $F_{IM}(im) = \mathbb{P}(IM \leq im)$ and $F_D(d) = \mathbb{P}(D \leq d)$ are the cumulative distribution functions of random variables IM and D, respectively.

The joint distribution of $F(im, d)$ may be calculated by either (1) using copulas, which have been extensively used for this purpose in applications [10], or (2) empirically using direct observations. Details on the representation of distribution of random vectors by copulas and their uniqueness are stated by Sklar theorem [11]. Empirical estimates' accuracy depends on the sample size of the vector (IM, D), and they may not provide accurate information about the components of (IM, D) beyond the range of the data.

Both approaches were used to calculate the joint distribution of $F(im, d)$ yielding similar results. The copula estimator $\hat{F}_C(im, d)$ of $F(im, d)$ is calculated as

$$\hat{F}_C(im, d) = \left(F_{IM}(im)^{-\theta} + F_D(d)^{-\theta} - 1\right)^{-\frac{1}{\theta}}, \tag{5}$$

where $\theta \in [-1, \infty)\backslash\{0\}$ is the parameter in the Clayton copula defined by $C(u, v) = (\max\{u^{-\theta} + v^{-\theta} - 1, 0\})^{-1/\theta}$, with $0 \leq u, v \leq 1$, and estimated from data. The empirical estimator of $F(im, d)$ is calculated by

$$\hat{F}(im, d) = \frac{1}{n}\sum_{k=1}^{n} \mathbb{1}\left(\{IM_k \leq im\} \cap \{D_k \leq d\}\right), \tag{6}$$

where IM_k and D_k are the samples of IM and D calculated for the sample ground motions in the NGA dataset.

3.2.1 Overall dependence of D and PGA

Figure 4 shows the Clayton copula $C(u, v)$ for (PGA, D), fitted to data (left panel) and the corresponding joint distribution $\hat{F}_C(im, d)$ (middle panel). The right panel of Figure 4 shows

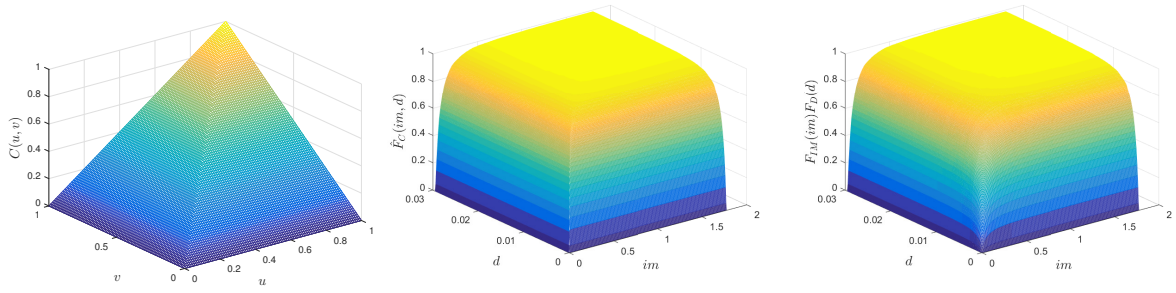

Figure 4: Clayton fitted-copula $C(u,v)$ (left); joint fitted Clayton distribution $\hat{F}_C(im,d)$ (middle); empirical estimate distribution $\hat{F}_C(im,d)$ (right), for $IM = PGA$.

the empirical estimate $\hat{F}(im,d)$ of $F(im,d)$. Figure 5 shows the difference between the joint distribution $\hat{F}(im,d)$ and the distribution $F_{IM}(im)F_D(d)$ in the left panel and the contour lines of the two estimators of $F(im,d)$. At the scale of the figure, the two estimates of $F(im,d)$ look very similar, suggesting that $IM = PGA$ and D are almost independent.

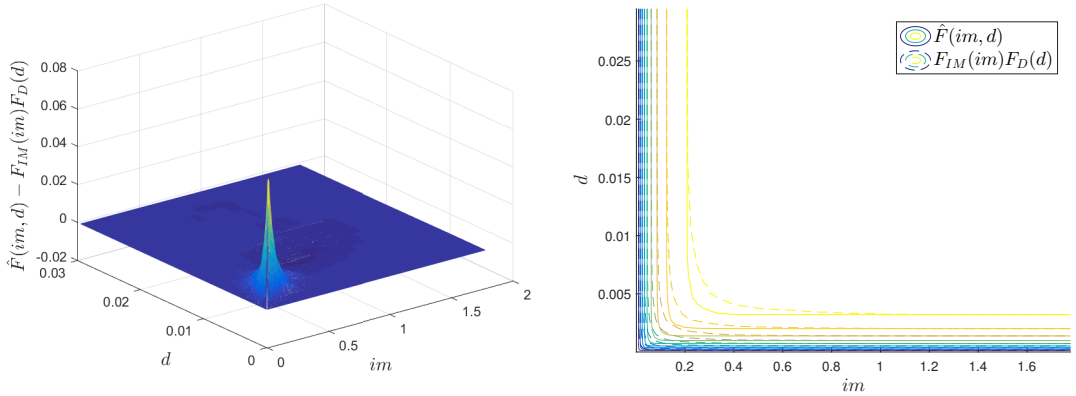

Figure 5: Difference between distribution functions $\hat{F}(im,d) - F_{IM}(im)F_D(d)$ (left) and counter plots of $\hat{F}(im,d)$ and $F_{IM}(im)F_D(d)$ (right), for $IM = PGA$

3.2.2 Overall dependence of D and SA

Plots similar to the ones for $IM = PGA$ in Figures 4 and 5 are shown for $IM = SA$ in Figures 6 and 7, respectively. Similar observations as for PGA and D, based on the graphic examinations of the plots in Figure 7, can be made regarding SA and D, which also seem to be nearly independent.

3.3 Tail dependence

Empirical or copula estimates of joint distributions of random vectors may fail to calculate joint extremes of their components and misrepresent the dependence of their components, in particular at large values. Therefore, elements of the multivariate extreme-value theory are employed to study the relationship between large values of demand parameters D and intensity measures IM. Their dependence is particularly important at these extreme values since large demand parameters are likely to cause extensive damage or even failure, which is reflected in the tails of the loss distributions.

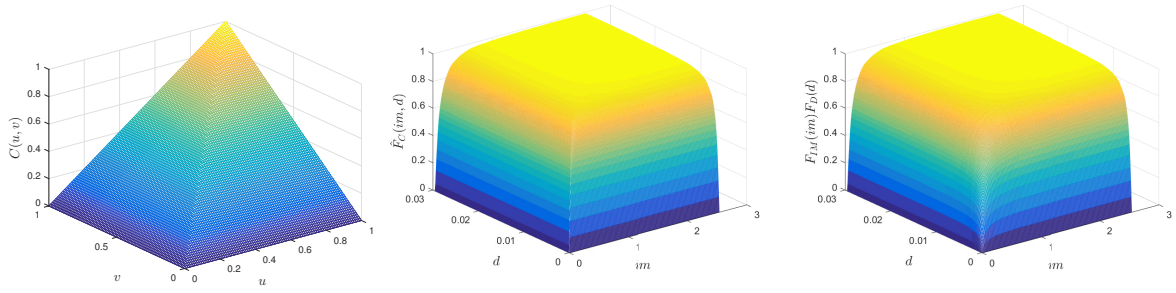

Figure 6: Clayton fitted-copula $C(u,v)$ (left); joint fitted Clayton distribution $\hat{F}_C(im,d)$ (middle); empirical estimate distribution $\hat{F}_C(im,d)$ (right), for $IM = SA$.

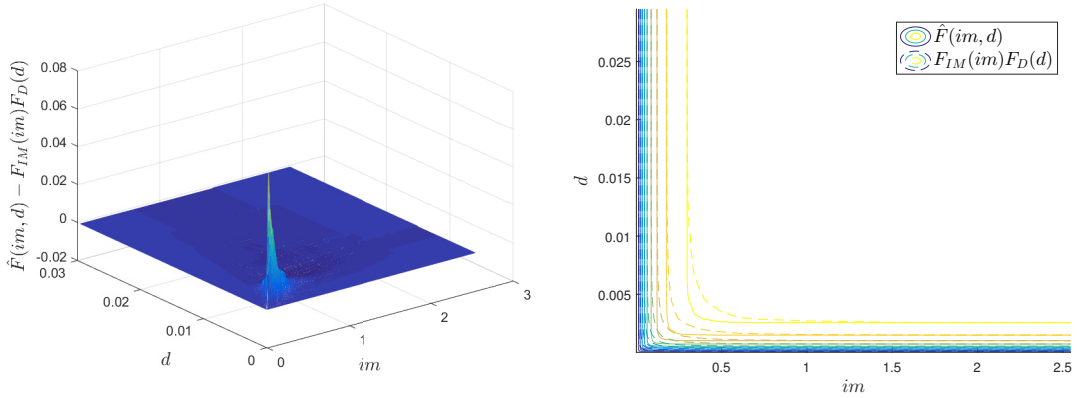

Figure 7: Difference between distribution functions $\hat{F}(im,d) - F_{IM}(im)F_D(d)$ (left) and counter plots of $\hat{F}(im,d)$ and $F_{IM}(im)F_D(d)$ (right), for $IM = SA$

The method of multivariate extreme-value theory used to quantify the dependence between simultaneous large values of IM and D is the ranks method [12], which does not require prior knowledge on the distributions of IM and D. The output of the method is a spectral measure $s(\theta)$ with support $\theta \in [0, \pi/2]$. If most of the mass of $s(\theta)$ is concentrated around the extremes 0 and $\pi/2$, then extremes of IM and D are nearly independent, and if the mass is concentrated around $\pi/4$, then the two variables are strongly dependent. The implementation of the method follows two steps. First, the samples of IM_k and D_k of IM and D are mapped into descending-order ranks

$$r_{IM}^{(k)} = \sum_{l=1}^{n} \mathbb{1}(IM_l \geq IM_k) \tag{7}$$

$$r_{D}^{(k)} = \sum_{l=1}^{n} \mathbb{1}(D_l \geq D_k), \tag{8}$$

where $k = 1, ..., n$ and n is the number of samples available. The second step is to represent vector $\left(n_0/r_{IM}^{(k)}, n_0/r_D^{(k)}\right)$ in polar coordinates, i.e.,

$$\left(n_0/r_{IM}^{(k)}, n_0/r_D^{(k)}\right) = \left(\rho^{(k)}\cos(\theta^k), \rho^{(k)}\sin(\theta^k)\right), \tag{9}$$

where $0 < n_0 << n$ is an integer. The spectral measure $s(\theta)$ is represented by the histogram of $\{\theta^{(k)}\}$ whose support is $[0, \pi/2]$. The selection of n_0 is important and it represents the number

of the top samples selected to construct the measure $s(\theta)$. If n_0 is too small, the uncertainty in $s(\theta)$ can be significant, but if n_0 is too large, the tail properties can be misrepresented.

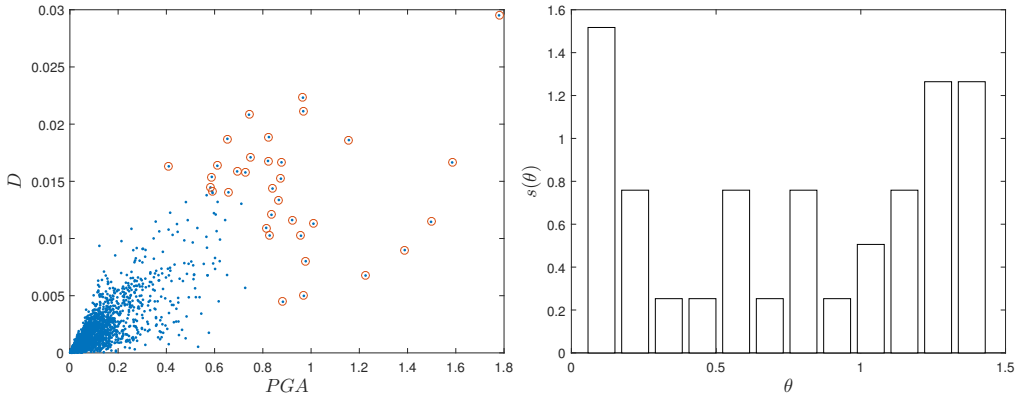

Figure 8: Selected samples of (PGA, D) (left) and the angular measure $s(\theta)$ (right)

3.3.1 Tail dependence of D and PGA

Following the ranks-method procedure described above, the spectral measure $s(\theta)$ is calculated and shown in Figure 8 (right) for the n_0 extreme-value samples of (D, PGA) selected and marked by red circles in Figure 8 (left). It is noticed that the mass of $s(\theta)$ with support on $[0, \pi/2]$ is distributed away from the center point of the interval, with high concentration around the extreme values 0 and $\pi/2$, respectively. This suggests that there is a weak dependence also between simultaneous high values of D and PGA. This suggests that the conditional random variables $D|PGA$ and D have similar distributions. In other words, the fragility defined earlier becomes $\mathbb{P}(D > d_{cr}|PGA) \simeq \mathbb{P}(D > d_{cr})$, which suggests that the fragility for the non-linear system considered is nearly independent of PGA.

3.3.2 Tail dependence of D and SA

Similar conclusions as the ones drawn for PGA can be inferred also in the case of SA. The left panel of Figure 9 shows the samples with large values of (SA, D) selected (red circles) to calculate the spectral measure $s(\theta)$, plotted in the right panel of the same figure. As for the previous case, most the mass of $s(\theta)$ is distributed away from the mid point $\pi/4$ with higher concentrations towards the boundaries of the support $[0, \pi/2]$. Thus, it can be concluded that the dependence between simultaneous large values of D and SA is also weak and that the fragility is almost independent of SA.

On the other hand, if $s(\theta)$ is calculated for the demand parameter defined as the maximum displacement of the linear oscillator $\max_{t\geq 0}|X_{lin}(t)|$ and SA, then the opposite trend as for vector (SA, D) is noticed. Figure 10 shows that the mass of $s(\theta)$ is almost exclusively concentrated around $\pi/4$, value which suggests a strong dependence between the components of the vector $(SA, \max_{t\geq 0}|X_{lin}(t)|)$. The same trend has already been noticed already in the scatter plot in Figure 3 (left panel), which shows a clear linear functional dependence between SA and $\max_{t\geq 0}|X_{lin}(t)|$, as expected in the case of linear SDOF systems.

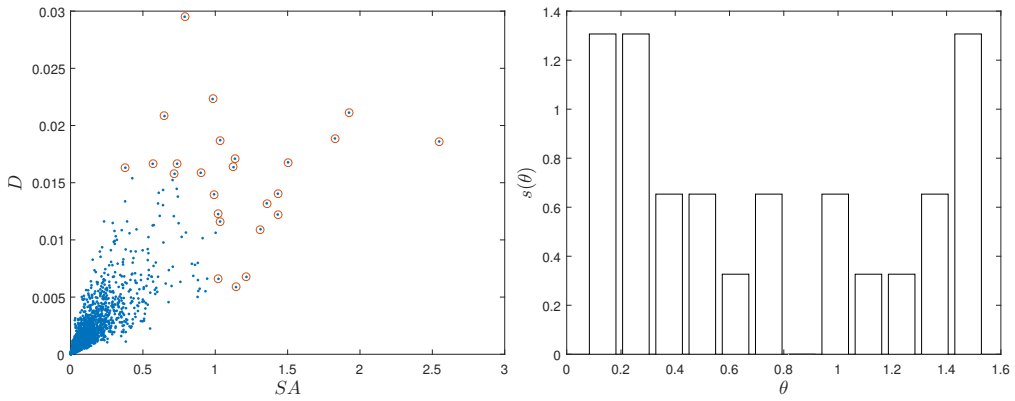

Figure 9: Selected samples of (SA, D) (left) and the angular measure $s(\theta)$ (right)

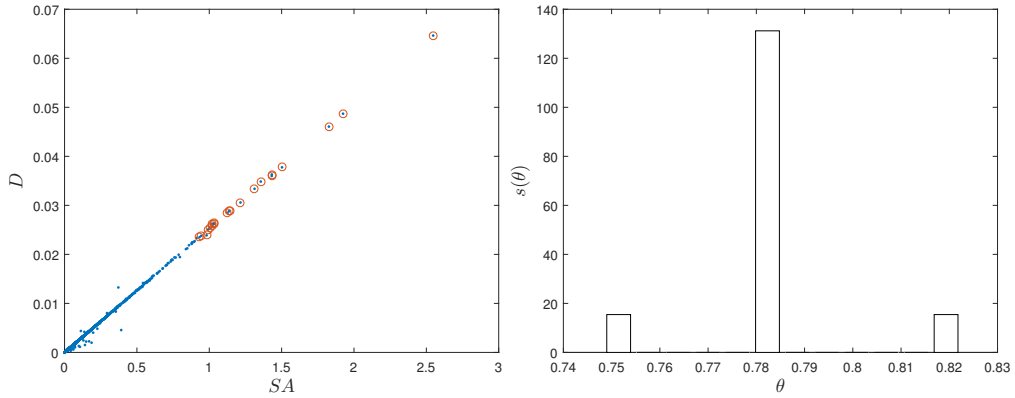

Figure 10: Selected samples of $(SA, \max_{t\geq 0} |X_{lin}(t)|)$ (left) and the angular measure $s(\theta)$ (right)

4 CONCLUSIONS

Fragility and vulnerability are defined often as functions of extensively-used intensity measures, such as the peak ground acceleration PGA and the spectral acceleration SA. These intensity measures are meaningful only under the assumption that they capture sufficient information on the seismic hazard such that the demand parameters D of complex, non-linear structures correlate satisfactorily with them. The paper examines whether the dependence between D and PGA, and D and SA, respectively, is adequate for fragility/vulnerability analyses. It has been shown that both the overall dependence between (IM, D), for $IM \in \{PGA, SA\}$, is weak as well as the dependence at simultaneous large values of IM and D. Thus, PGA and SA provide insufficient information on the demand parameters of non-linear, complex structural systems, and fragility and vulnerability functions for these systems are almost independent of the intensity measures IM.

ACKNOWLEDGEMENTS

The work reported in this paper has been supported by the Marie Skłodowska-Curie Actions of the European Union's Horizon 2020 Program under the grant agreement 704679 - PARTNER, and by the National Science Foundation under grants CMMI-095714 and CMMI-1639669. This support is gratefully acknowledged.

REFERENCES

[1] Hazus-MH MR5, Multi-hazard loss estimation methodology: earthquake model *Department of Homeland Security, Federal Emergency Management Agency, Washington, DC*.

[2] M. Shinozuka, M.Q. Feng, H. Kim, T. Uzawa and T. Ueda, Statistical analysis of fragility curves *Technical report MCEER: 106-E-7.3.5 and 106-E-7.6*, 2001.

[3] Y.S. Choun and A.S. Elnashai, A simplified framework for probabilistic earthquake loss estimation *Probabilistic Engineering Mechanics*. **25**, 355–364, 2010.

[4] A. Liel, C.B. Haselton, G.G. Deierlein and J.W. Baker, Incorporating modeling uncertainties in the assessment of seismic collapse risk of buildings *Structural Safety*. **31**, 197–211, 2009.

[5] J.W. Baker, Probabilistic structural response assessment using vector-valued intensity measures *Earthquake Engineering and Structural Dynamics*. **36**, 1861–1883, 2007.

[6] C. Kafali and M. Grigoriu, Seismic fragility analysis: Application to simple linear and nonlinear systems *Earthquake Engineering and Structural Dynamics*. **36**, 1885-1900, 2007.

[7] M. Grigoriu, Do seismic intensity measures (IMs) measure up? *Probabilistic Engineering Mechanics*. **46**, 80–93, 2016.

[8] B. Chiou, R. Darragh, N. Gregor, and W. Silva, NGA Project Strong-Motion Database *Earthquake Spectra*. **24(1)**, 23–44, 2008.

[9] Next Generation Attenuation Relationships for Western US (NGA West) *http://peer.berkeley.edu/ngawest/*. Last accessed on 23/02/2017.

[10] C. Genest and A.C. Favre, Everything you wanted to know about copula modeling but you were afraid to ask. *Journal of Hydrological Engineering*. **12(4)**, 347–368, 2007.

[11] L. Ruschendorf, Dependence, Risk Bounds, Optimal Allocations and Portfolios (Chap.1). *Mathematical Risk Analysis*. Ed. Springer, 2013.

[12] S. Resnick, Heavy-tail Phenomena: Probabilistic and Statistical Modeling. *Ed. Springer*. New York, 2007.