Soil Dynamics and Earthquake Engineering 30 (2010) 1513–1527

Contents lists available at ScienceDirect

Soil Dynamics and Earthquake Engineering

journal homepage: www.elsevier.com/locate/soildyn

Disaggregation-based response weighting scheme for seismic risk assessment of structures

Ludovica Elefante *, Fatemeh Jalayer, Iunio Iervolino, Gaetano Manfredi

Dipartimento di Ingegneria Strutturale, Università degli Studi di Napoli, Federico II, Italy

ARTICLE INFO

Article history:
Received 24 September 2009
Received in revised form
8 July 2010
Accepted 11 July 2010

ABSTRACT

The results of seismic hazard disaggregation can be used to assign relative weights to a given ground motion record based on its corresponding magnitude, distance and deviation from the ground motion prediction model (epsilon) in order to make probability-based seismic assessments using non-linear dynamic analysis. In this paper, the implications of using the weighted ground motion records are investigated in terms of the mean annual frequency of exceedance of the critical component-based demand to capacity ratio in an existing reinforced concrete structure using both the peak ground acceleration and the first-mode spectral acceleration as intensity measures. It is demonstrated how site-specific seismic hazard disaggregation can be used in order to obtain the conditional probability distribution for a relevant ground motion characteristic given the chosen intensity measure. Distinguished by the amount of structural analysis required, two alternative non-linear dynamic analysis procedures, namely the cloud and the stripes method are implemented. The weighted cloud and the weighted stripes methods are then introduced as analysis procedures which modify the structural response to the selected ground motion records by employing the information provided from the seismic hazard analysis. It is demonstrated that the resulting annual frequencies based on weighted records are comparable to those obtained by using vector-valued intensity measures, while requiring less computational effort.

1. Introduction

The seismic input selection represents one of the main issues in assessing the seismic response of a structure through numerical dynamic analysis [1]. In some cases, the records are selected to have response spectra that approximate the uniform hazard spectrum or other "design" response spectrum (e.g., [2–4]). In general, it is reasonable to choose ground motion (GM) records whose magnitude, distance, site conditions and fault mechanisms are representative for the seismic hazard at the site of the structure under consideration. This choice may be guided by the disaggregation of the seismic hazard [5–7] for the site of interest. However, once the set of records is chosen, there are several techniques to evaluate the structural seismic response [8,9].

The choice of the GMs may be affected by the interface variable used to measure the intensity of GM, known as the intensity measure *IM*. According to the criteria proposed by Luco and Cornell [10] a preferred *IM* is both "sufficient" with respect to the GM characteristics and also "efficient". A sufficient *IM* renders the structural response conditionally statistically independent of other GM characteristics such as event magnitude, while an efficient *IM* predicts the structural response with (relatively) small record-to-record variability. Theoretically, careful record selection is not essential if the *IM* is demonstrated to be sufficient [11,12]. It has to be recalled that sufficiency of a specific *IM* depends on the structure, the structural response parameters and the GM characteristics. GM parameters such as site amplification and/or directivity may prove particularly troublesome because they may imply strong sensitivity of spectral shape to certain GM parameters [13]. A useful strategy, in cases where the adopted scalar intensity measure IM_1 does not prove to be sufficient, is to introduce an additional intensity measure, IM_2. That is, one can adopt a vector-valued $\boldsymbol{IM}=[IM_1,IM_2]$, consisting of two scalar *IM*'s, in order to render a more complete description of the GM characteristics [13].

In this paper, an approximate method based on linear regression is used in order to establish possible correlation between the structural response conditional on the primary IM_1 and the secondary IM_2. Moreover, a weighting scheme based on seismic hazard disaggregation is used, in the framework of the scalar IM_1, in order to adjust the structural response for possible correlations with a candidate secondary IM_2. This weighting scheme can be implemented in non-linear dynamic analysis

* Correspondence to: Department of Structural Engineering, University of Naples Federico II, Via Claudio 21, Naples 80125, Italy
E-mail address: ludovica.elefante@unina.it (L. Elefante).

0267-7261/$ - see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.soildyn.2010.07.003

procedure for both wide range of GM intensities and also limited range of GM intensities.

The efficiency of the weighting scheme is evaluated in terms of seismic risk which is represented herein by the mean annual frequency of exceeding the critical component demand to capacity ratio. The seismic risk curves are obtained by adoption of both scalar and vector-valued *IM*'s.

2. Probabilistic assessment based on non-linear dynamic analysis

Adopting the performance assessment methodology developed by the Pacific Earthquake Engineering Research (PEER) Center for buildings in the framework of Performance-Based Earthquake Engineering (PBEE) [14,15], a probabilistic performance-based criterion for seismic assessment of existing structures can be written as

$$\lambda_{EDP} \le P_0 \tag{1}$$

where λ_{EDP} refers to the (mean) annual frequency (MAF) of exceeding a specified damage level[1] expressed in terms of an engineering demand parameter (*EDP*) and P_0 the allowable probability threshold for the assessment.[2] In the framework of PBEE, an intermediate parameter known as the intensity measure *IM* is introduced in order to relate the characteristics of the GM record to structural performance. The annual rate of exceeding a specified limit state can be expanded, using the principles of probability theory, with respect to the adopted (scalar) *IM* in the following [15,9]:

$$\lambda_{EDP}(y) = \int P_{EDP|IM}(EDP > y|x)\left|d\lambda_{IM}(x)\right| \tag{2}$$

The first term in the integrand $P_{EDP|IM}(EDP > y|x)$ is the conditional probability of exceeding the structural response threshold y for a given value of $IM=x$. This term is also known as the structural *fragility*. The second term in the integrand is the absolute value of the derivative of the annual rate of exceeding $IM=x$; this second term is known as the *hazard* for the adopted *IM*. Ideally, the hazard function for the adopted *IM* is obtained from the results of site-specific probabilistic seismic hazard analysis (PSHA, see [17]).

The non-linear dynamic analysis procedures based on a limited suite of GM records can be used to estimate the fragility term in Eq. (2). Depending on the amount of structural analysis and also on the range of limit states for which the performance assessment is done, two alternative non-linear dynamic analysis procedures are considered in this work, the *cloud method* and the *stripes method*.

The cloud method [8] employs the linear least squares scheme to the specified *EDP* given *IM* based on non-linear structural response (*cloud response*) for a suite of GM records (un-scaled) in order to estimate the conditional mean and standard deviation of *EDP* given *IM*. The regression scheme is used to provide a power–law estimate of the median *EDP* for a given *IM*. Moreover, by assuming that the errors of the least-square estimate are independent and identically distributed (i.i.d.) and assuming a Gaussian distribution for the logarithm of *EDP* given *IM*, the $P_{EDP|IM}(EDP > y|x)$ term can be estimated using the Complementary Cumulative Distribution Function (CCDF)

$$P_{EDP|IM}(EDP > y|x) = G_{EDP|IM}(y|x) = 1-\Phi\left(\frac{\ln y-\eta(\ln y|x)}{\sigma(\ln y|x)}\right) \tag{3}$$

where $\eta(\ln y|x)$ and $\sigma(\ln y|x)$ are the conditional mean and standard deviation of the logarithm of *EDP* given a specific *IM* level, respectively.

The stripes method [9,8] provides the non-linear structural response parameters for the suite of records that are scaled to successively increasing *IM* levels: this is referred to as the *stripe response*. Subsequently, the statistical properties of the stripe response for various *IM* levels, calculated based on the response to the suite of records, can be employed to obtain the probability of exceeding a specified *EDP* level.

In the case where a vector-valued $\boldsymbol{IM}=[IM_1,IM_2]$ consisting of two scalar *IM*'s is adopted, the fragility term in Eq. (2) for the annual rate of exceeding $EDP=y$ can be expanded with respect to IM_2 and re-arranged as following [18]:

$$\lambda_{EDP}(y) = \iint P_{EDP|IM_2,IM_1}(EDP > y|x,z) f_{IM_2|IM_1}(z|x)\left|d\lambda_{IM_1}(x)\right| \tag{4}$$

The first term in the integrand is the conditional probability of exceeding $EDP=y$ given IM_1 and IM_2 and the second term is the conditional probability density function (PDF) for $IM_2=z$ given $IM_1=x$. Similar to the case regarding scalar *IM*, both cloud method and stripes method can be employed in order to perform probabilistic seismic risk assessments. In the cloud method, the two-variable ordinary least squares scheme can be used to estimate the statistical parameters for *EDP* conditional on both IM_1 and IM_2 [11]. Also in this case, by assuming that the errors of the multiple-linear least square estimate are independent and identically distributed (i.i.d.) and assuming a Gaussian distribution for the logarithm of *EDP* given IM_1 and IM_2, the $P_{EDP|IM_1,IM_2}(EDP > y|x,z)$ term can be estimated using the CCDF

$$P_{EDP|IM_1,IM_2}(EDP > y|x,z) = G_{EDP|IM_1,IM_2}(y|x,z) = 1-\Phi\left(\frac{\ln y-\eta(\ln y|x,z)}{\sigma(\ln y|x,z)}\right) \tag{5}$$

where $\eta(\ln y|x,z)$ and $\sigma(\ln y|x,z)$ are the conditional mean and standard deviation of the logarithm of *EDP*, respectively, given $IM_1=x$ and $IM_2=z$. Alternatively, using the stripes method, the simple linear least squares can be applied to the stripe response at various IM_1 level with IM_2 as the independent variable [18].

In this work, both scalar and vector *IM*'s are studied. As scalar *IM*'s, the peak ground acceleration (*PGA*) and the first-mode spectral acceleration ($S_a(T_1)$) are considered. As vector *IM*'s, the pairs consisting of *PGA* and magnitude *M* [11], $S_a(T_1)$ and the deviation from the GM prediction model epsilon (ε) [13] are considered. Epsilon is defined as the number of standard deviations by which an observed logarithmic spectral acceleration differs from the mean logarithmic spectral acceleration of a ground-motion prediction (attenuation) equation. The equation corresponding to this definition is:

$$\varepsilon = \frac{\ln S_a(T)-\hat{\mu}_{\ln S_a(T)}}{\hat{\sigma}_{\ln S_a(T)}} \tag{6}$$

where $\ln S_a(T)$ is the natural logarithm of the spectral acceleration at a specified period T and $\hat{\mu}_{\ln S_a(T)}$ and $\hat{\sigma}_{\ln S_a(T)}$ are, respectively, the mean and the standard deviation as predicted by a ground motion prediction relationship. In this work, ε is calculated based on the Sabetta and Pugliese [19] model for ground motion prediction. It should be noted that epsilon is defined with respect to the

[1] It is desirable to express the performance objectives in terms of life-cycle cost [16]. However, the focus of this work is on the ground motion record selection for the purpose of estimating non-linear structural response. Therefore, the performance objective is hereby stated in terms of an engineering demand parameter instead of economic indices.

[2] Note that in Eq. (1) it has been assumed that the numerical value for rate of exceedance is close to that of the probability of exceedance; this is true for small values of the exceedance probability representing very rare events, assuming a Poisson occurrence model.

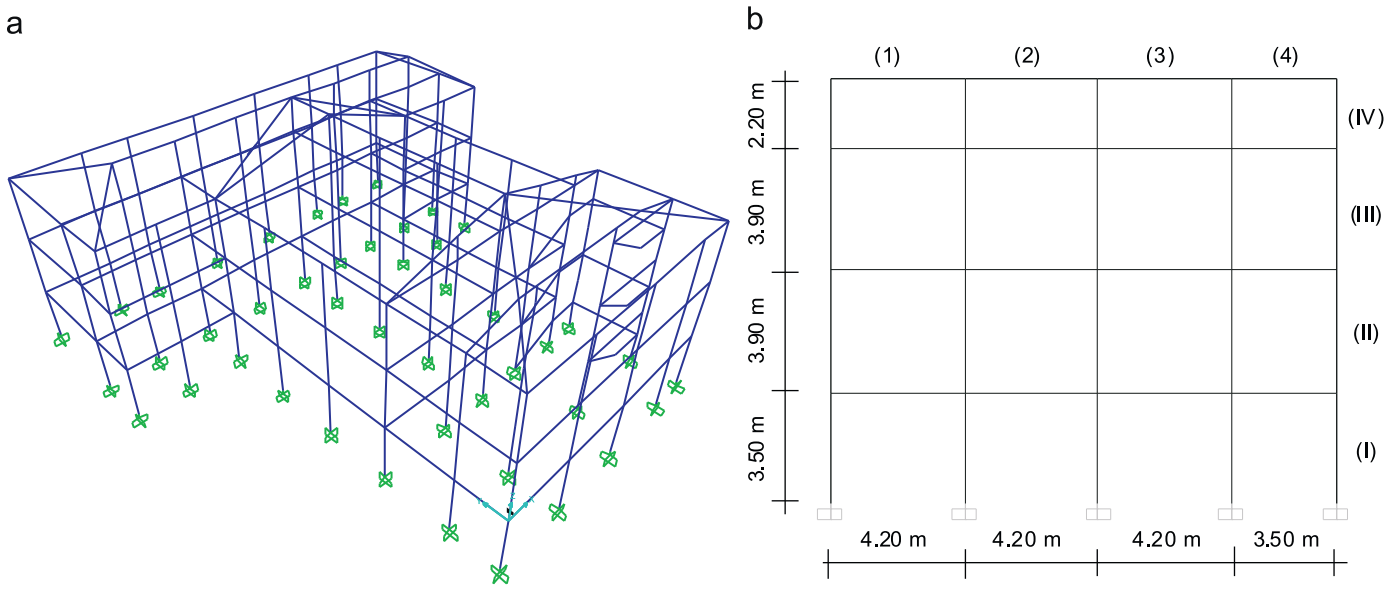

Fig. 1. (a) The tri-dimensional view of the scholastic building and (b) the central frame of the case-study building.

unscaled record and will not change in value when the record is scaled.

PGA has been widely used as the ground motion *IM* in the past. The first-mode spectral acceleration $S_a(T_1)$ is verified to be a more suitable choice of an *IM*, as it reflects the elastic response of a single degree of freedom (SDOF) system with a period equal to the first-mode period of the structure. However, $S_a(T_1)$ is unable to reflect the effect of higher frequencies (higher modes in a structure with several degrees of freedom) or lower frequencies (severe non-linear behavior in the structural elements) or the near-source effects (a single low-frequency pulse dominating the ground motion record [20]). Ideally, one should use a vector consisting of $S_a(T_1)$ and an *IM* related to the spectral shape [11,21–24], for example the ratio $R(T)=S_a(T)/S_a(T_1)$ [11], where T is the second period whose spectral value is considered as important to the structural response. It has been demonstrated that *epsilon* introduced above may act as a proxy for the spectral shape [13].

Moreover, as vector *IM*'s, $S_a(T_1)$ is paired with the spectral acceleration at a period $T_2 \neq T_1$, $(S_a(T_2))$: as will be explained later, T_2 is the second period whose spectral acceleration is most efficient in predicting structural response.

3. Structural model

As the case-study, an existing school structure in the city of Avellino, Italy, is considered herein. Avellino is located in the Irpinia region, where the 1980 Irpinia Earthquake with moment magnitude (M_w) equal to 6.9 has taken place. The Irpinia region was classified as seismic Zone II (intermediate seismicity with design *PGA* on rock for the life safety limit state equal to 0.25 g) according to the Italian seismic guidelines [25]. The design seismic action considered in this study is that according to this code, although it is now superseded by DM 29, 2008 [26]. The structure consists of three stories and a semi-embedded story and its foundation lies on stiff soil (category *B* according to Eurocode 8 [27]). For the structure in question, the original design notes and graphics have been gathered. The building is constructed in the 1960s and it is designed for gravity loads only, as it is frequently encountered in the post-second world war construction. In Fig. 1a, the tri-dimensional view of the structure is illustrated; it can be observed that the building is irregular both in plane and elevation. The main central frame in the structure is extracted and used as the structural model (Fig. 1b).

The columns have rectangular section with the following dimensions: first storey: 40×55 cm^2, second storey: 40×45 cm^2, third storey: 40×40 cm^2 and fourth storey: 30×40 cm^2. The beams, also with rectangular section, have the following dimensions: 40×70 cm^2 at first and second floor and 30×50 cm^2 for the ultimate two floors. It can be inferred from the original design notes that the steel rebar is of the type Aq42, characterized by a nominal minimum yield strength $f_y=2300$ kg/cm^2 (225 MPa), and that the concrete has a minimum strength equal to 180 kg/cm^2 (18 MPa) [28]. However, the steel yield strength and the concrete compression strength in the structural model are taken equal to 3200 kg/cm^2 (314 MPa) and 165 kg/cm^2 (16 MPa), respectively. These are mean values extracted from a statistical survey of material properties in existing buildings constructed in the 1960s [29,30].

The finite element model of the frame is constructed, using the Open System for Earthquake Engineering Simulation (OpenSees) software, assuming that the non-linear behavior in the structure is concentrated in plastic hinges located at the element ends [31]. The plastic hinges can form both in beams and columns. Hinge characteristics such as, yield and ultimate chord rotation and the plastic hinge length are calculated from the semi-empirical formulas provided in the Italian code [32] for existing buildings. The concrete behavior is modeled based on the Mander–Priestly [33] constitutive relation for un-confined concrete. The reinforcing steel is assumed to have elastic–plastic behavior. The plastic hinges take into account the interaction of flexural and axial action and the flexural strength degradation (shear failure is not considered in hinge modeling). As it regards the post-peak behavior, it is assumed that the section resistance drops to a value of the moment equal to one tenth of the ultimate moment (M^*) corresponding to a rotation value equal to two times the ultimate value (θ^*), resulting in a tri-linear curve. The tri-linear moment-rotation backbone curve is demonstrated in Fig. 2.[3] The P-Delta effects are included in the structural analysis. The

[3] These post-peak values for moment and rotation are chosen rather arbitrarily in order to avoid numerical in-convergence problems.

structural damping is modeled based on the Rayleigh model and is assumed to be equal to 5% for the first two modes. The small-amplitude period for the first two vibration modes are equal to 0.73 and 0.26 s, respectively.

The performance objective is expressed in terms of a (scalar) system *EDP* which reflects how far away the structure is from the threshold of the limit state defined as $EDP=1$ [34]

$$\lambda_{LS} = \lambda_{EDP>1} \leq P_0 \tag{7}$$

The *EDP* can be defined as the ratio of system demand D to system capacity C_{LS}, (e.g., ratio of θ_{max} to θ_{CLS}) or it can be defined as a function of component demand and capacities, which is equal to the one at the onset of failure. This latter formulation is adopted herein based on the system reliability concept of *cut-sets* [35]. In this work, the scalar global *EDP*, denoted by Y, is a critical demand to capacity ratio defined as the demand to capacity ratio of the strongest component of the weakest structural mechanism or cut-set [34]

$$Y = \max_{l=1}^{N_{mech}} \min_{j=1}^{N_l} \frac{D_{jl}}{C_{jl}} \tag{8}$$

where N_{mech} is the number of considered cut-sets or potential failure mechanisms and N_l the number of components taking part in the lth cut-set. A cut-set is defined as any set of components whose joint failure leads to system failure. Three types of potential failure mechanisms (cuts-sets) are considered herein:

(a) Ultimate chord rotation capacity in the external columns of a storey;
(b) Ultimate chord rotation capacity in the central columns of a storey and
(c) Yield chord rotation in the columns of a storey (a.k.a., soft story mechanism).

The components of each mechanism are the column plastic hinges involved in it. Shear failure has not been considered.

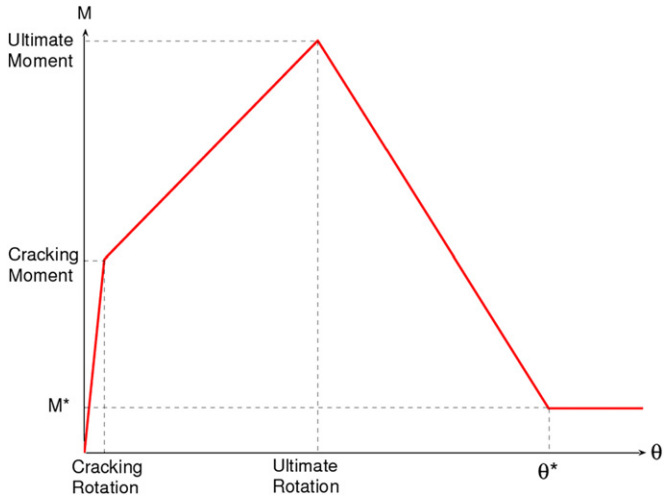

Fig. 2. Schematic diagram of the typical tri-linear behavior characterizing the rigid-plastic hinge.

4. The suites of ground motion records and their properties

Two different suites, respectively, of 21 (*Sel_A*) and 20 (*Sel_B*) GM records, all based on Mediterranean events, have been selected for this study. They are all main-shock recordings recorded on stiff soil ($400\ \text{m/s} < V_{s30} < 700\ \text{m/s}$), which is consistent with the soil-type for the site. The first suite of records is taken from European Strong-Motion Database or ESD (http://www.isesd.cv.ic.ac.uk/ESD/Database/Database.htm) (17 recordings) and Pacific Earthquake Engineering Research Next Generation Attenuation or PEER NGA Database (http://peer.berkeley.edu/nga/flatfile.html) (4 recordings). The earthquake events have moment magnitude (M_w) between 5.3 and 7.2, and closest distances ranging between 7 and 30 km. This suite of records has been chosen in order to cover a wide range of moment magnitude values. The second suite of records, is selected from the ESD. It consists of earthquake events with M_w between 5.9 and 7.2, and closest distances ranging between 0 and 71 km. This second suite is chosen so that it covers a wide range of ε values. Tables 1 and 2 illustrate the GM recordings, their M_w, the fault

Table 1
Selection *A* of ground motion records.

Record	M_w	FM	V_{s30} (m/s)	*ED* (km)	*FD* (km)	*PGA* (g)	$S_a(T_1)$ (g)	ε
Basso Tirreno	6.0	Oblique	?	18	16	0.15	0.17	−0.121
Valnerina	5.8	Normal	?	23	21	0.04	0.03	−0.529
Camp. Lucano	6.9	Normal	529	16	13	0.16	0.31	−0.519
Preveza	5.4	Thrust	?	28	7	0.14	0.10	−0.244
Umbria	5.6	Normal	546	19	19	0.21	0.02	0.230
Lazio Abruzzo	5.9	Normal	?	36	28	0.07	0.05	−0.219
Etolia	5.3	Thrust	405	20	12	0.04	0.01	−0.518
Montenegro	5.4	Thrust	399	18	?	0.07	0.09	−0.227
Kyllini	5.9	Strike slip	490	14	11	0.15	0.15	−0.231
Duzce 1	7.2	Oblique	662	26	13	0.13	0.18	−0.722
Umbria Marche	5.7	Normal	400	32	28	0.04	0.05	−0.334
Potenza	5.8	Strike slip	494	28	29	0.10	0.08	−0.003
Ano Liosia	6.0	Normal	411	20	9	0.16	0.06	−0.308
Adana	6.3	Strike slip	?	39	30	0.03	0.05	−0.749
South Iceland	6.5	Strike slip	?	15	10	0.21	0.13	−0.344
Tithorea	5.9	Normal	665	25	?	0.03	0.02	−0.639
Patras	5.6	Strike slip	665	30	?	0.05	0.02	−0.184
Friuli Italy-01	6.5	Reverse	425	20	21	0.35	0.35	0.168
Friuli, Italy-02	5.9	Reverse	412	18	18	0.21	0.08	0.110
Fruili, Italy-03	5.5	Reverse	412	20	21	0.11	0.21	0.034
Irpinia, Italy-01	6.9	Normal	600	15	18	0.13	0.30	−0.466
(average)	6.0		501	23	18	0.12	0.12	−0.277

Table 2
Selection *B* of ground motion records.

Record	M_w	FM	V_{s30} (m/s)	*ED* (km)	*FD* (km)	*PGA* (g)	$S_a(T_1)$ (g)	ε
Friuli	6.5	Thrust	?	42	34	0.06	0.22	−0.015
Friuli	6.5	Thrust	?	87	71	0.05	0.11	0.003
Camp. Lucano	6.9	Normal	472	48	33	0.11	0.25	−0.204
Camp. Lucano	6.9	Normal	529	16	13	0.16	0.31	−0.493
Kalamata	5.9	Normal	486	10	0	0.22	0.48	−0.231
Kalamata	5.9	Normal	399	11	0	0.24	0.48	−0.233
Umbria Marche	6.0	Normal	546	11	1	0.52	0.56	−0.216
Umbria Marche	6.0	Normal	450	38	27	0.09	0.17	0.062
South Iceland	6.5	Strike slip	?	7	6	0.63	0.54	−0.288
Duzce 1	7.2	Oblique	662	26	13	0.13	0.18	−0.893
Friuli	6.5	Thrust	?	42	34	0.09	0.25	0.031
Friuli	6.5	Thrust	?	87	71	0.07	0.12	−0.002
Camp. Lucano	6.9	Normal	472	48	33	0.14	0.26	−0.187
Camp.Lucano	6.9	Normal	529	16	13	0.18	0.31	−0.484
Kalamata	5.9	Normal	486	10	0	0.30	0.63	−0.120
Kalamata	5.9	Normal	399	11	0	0.27	0.51	-0.208
Umbria Marche	6.0	Normal	546	11	1	0.46	0.64	-0.156
Umbria Marche	6.0	Normal	450	38	27	0.10	0.18	0.065
South Iceland	6.5	Strike slip	?	7	6	0.51	0.74	−0.154
Duzce 1	7.2	Oblique	662	26	13	0.16	0.14	−1.004
(average)	6.4		506	30	20	0.22	0.35	−0.236

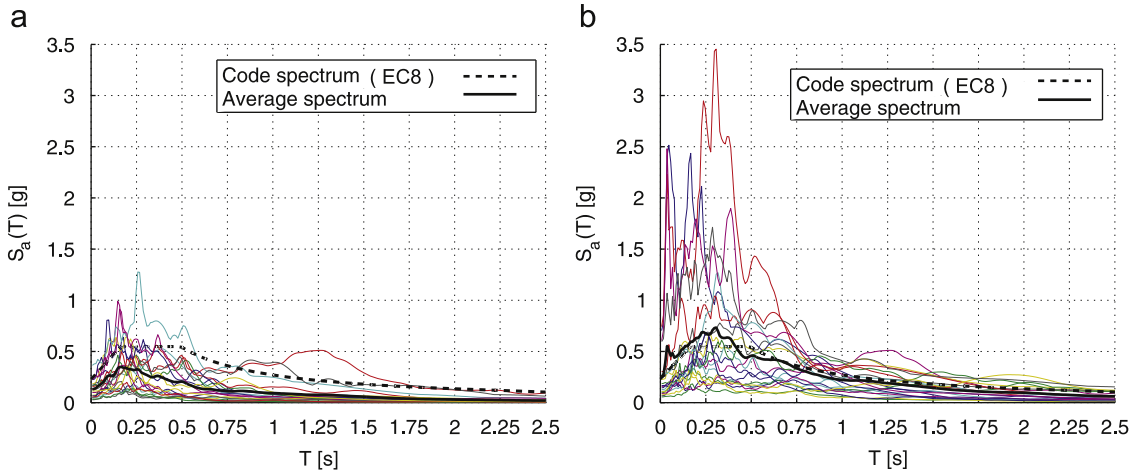

Fig. 3. (a) Acceleration spectra *Sel_A* and (b) acceleration spectra *Sel_B*.

mechanism (FM), the velocity of propagation of the shear waves (V_{s30}, symbol "?" indicates that the value is not available in the database), epicentral distance (*ED*), fault distance (*FD*), peak ground acceleration (*PGA*), spectral acceleration at the first mode ($S_a(T_1)$) and ε values for each record of *Sel_A* and *Sel_B*, respectively. It has to be noted that epsilon values of *Sel_B* are all included in a range of values between −1.1 and 0.07. The acceleration spectra for the original (un-scaled) records for the two chosen selections, *Sel_A* and *Sel_B*, are plotted in Fig. 3a and b, respectively.

5. The disaggregation of seismic hazard

In order to adopt a vector-valued ***IM*** for representing the GM intensity in the seismic assessment outlines in Eq. (4), it is necessary to obtain the conditional probability distribution for the second *IM* given the occurrence of the original *IM*.

This section employs a site-specific seismic hazard analysis performed based on the Italian seismic zonation (ZS9, areal seismic zones, [36], Fig. 4) inside a Bayesian framework for inference in order to obtain the conditional probability distribution for magnitude m, distance r and the deviation from the attenuation law ε given the original *IM* adopted. As mentioned earlier, the GM prediction relation adopted in this work is the Sabetta and Pugliese relation [20].

It should be noted that the website of INGV (Istituto Nazionale di Geofisica e Vulcanologia, Progetto DPC-INGV-S1, http://esse1.mi.ingv.it) provides the results of site-specific seismic hazard analysis based on the Italian seismic zonation, but only in terms of *PGA*.

The seismic hazard can be disaggregated with respect to, for example, magnitude using the Bayes theorem

$$f(m|IM) = \frac{f(IM,m)}{\sum f(IM|m)f(m)} = \frac{f(IM,m)}{f(IM)} \tag{9}$$

Suppose that the site of interest is surrounded by areal seismic zones. Using the total probability theorem, the term $f(IM,m)$ can be further expanded with respect to the seismogenic areas surrounding the site, the distance r of the points within each area to the site of interest and epsilon values

$$f(m|IM) = \frac{\sum_i \sum_R \sum_\varepsilon \alpha_i I(IM|m,r,\varepsilon) f(m) f(r) f(\varepsilon)}{f(IM)} \tag{10}$$

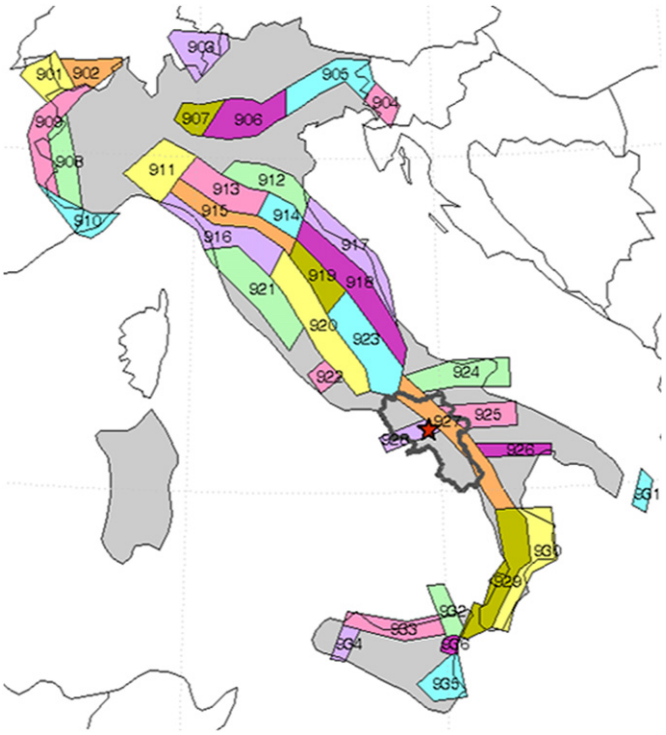

Fig. 4. Seismogenic zonation ZS9; the different zones are identified by number; the Campania region is highlighted by a gray polyline and the site of interest is indicated by a pentagram.

where α_i is the relative seismicity of seismic area A_i with respect to other seismic zones considered ($\alpha_i=\lambda_i/\Sigma\lambda$, where λ_i is the seismicity of zone i). The term $I(IM|m, r, \varepsilon)$ is an indicator function reflecting the fact that given the epsilon, distance and magnitude, the *IM* value is going to be known deterministically from the GM prediction relationship. That is, $I(IM|m, r, \varepsilon)$ can assume only two values, namely, zero and unity. It should be noted that, given that an earthquake takes place in seismic zone A_i, it is assumed that it is equally likely to have its epicenter located anywhere inside the area. Therefore, the probability $f(r)$ of having an earthquake with its epicenter inside the areal increment dA, whose center is individuated by the distance r from the site, can be calculated as dA/A where A is the total area of the seismic zone.

The probability density function (PDF) for the magnitude can be calculated from the Gutenberg–Richter truncated distribution

$$f(m)=\frac{\beta e^{-\beta m}}{e^{-\beta m_l}-e^{-\beta m_u}} \tag{11}$$

where β is the Richter b-value times ln(10) and m_l and m_u are a lower and an upper bound magnitude, respectively. Table 3 illustrates the herein used relative seismicity λ, the Richter b-value and a lower and an upper magnitude bound (m_l and m_u, respectively) for those ZS9 seismic zones surrounding the Campania region.

It should be noted that the term $f(IM)$, in the denominator of Eq. (10) acts as a scaling constant on the nominator. Therefore, it can be calculated by calculating the nominator for an interval covering all possible m values, for a given value of *IM*, and summing them up. This is because the resulting probability distribution needs to sum to one for all possible m values.

In order to disaggregate the seismic hazard with respect to ε of the GM prediction relationship, the same as above, the Bayes theorem and the total probability theorem can be used in order

Table 3
Parameters of those ZS9 seismic zones surrounding the Campania region.

Zone	λ	b-value	m_l	m_u
920	0.0600	1.9600	4.7600	6.1400
922	0.3700	2.0000	4.7600	5.4500
923	0.1400	1.0500	4.7600	7.0600
924	0.1300	1.0400	4.7600	6.8300
925	0.1700	0.6700	4.7600	6.8300
926	0.1000	1.2800	4.7600	6.1400
927	0.4300	0.7400	4.7600	7.0600
928	0.2100	1.0400	4.7600	5.9100
929	0.1700	0.8200	4.7600	7.2900
930	0.1700	0.9800	4.7600	6.6000

to calculate

$$f(\varepsilon|IM)=\frac{\sum_i\sum_M\sum_R\alpha_i I(IM|m,r,\varepsilon)f(m)f(r)f(\varepsilon)}{f(IM)} \tag{12}$$

where $f(IM,\varepsilon)$ is expanded, using the total probability theorem, with respect to the seismogenic areal zones, magnitude m and source-to-site distance r. In a similar manner, as described above, the constant term $f(IM)$ in the denominator can be calculating by summing up the nominator in Eq. (12) for all possible values of epsilon. It should be noted that the expansion of the nominator in Eq. (12) is done assuming that the probability distribution for ε is independent of other GM parameters. Moreover, it is assumed that the probability distribution for magnitude is the same for all the points within a given seismic area.

The conditional probability distributions of magnitude given *PGA* and of epsilon of the prediction law given $S_a(T_1)$, have been obtained from Eqs. (10) and (12) through the disaggregation of the seismic hazard for the site of the case-study structure.

The distributions $f(m|PGA)$ obtained for different levels of *PGA* are illustrated in Fig. 5a. The histogram of magnitude for record selection *A* is also shown at the bottom of the figure. In the same manner, the distributions $f(\varepsilon|S_a(T_1))$ obtained for different levels of $S_a(T_1)$ are illustrated in Fig. 5b. The histogram of epsilon for record selection *B* is plotted at the bottom of the figure.

6. Sufficiency and the weighted method

Linear regression is a useful statistical tool for investigating the sufficiency criteria for a candidate IM_1 [10,21,11]. The *EDP*, *Y*, can be predicted as a function of IM_1 (e.g., *PGA* or $S_a(T_1)$), by performing linear regression (usually in the logarithmic scale). The efficiency of an *IM* can be measured by the standard error of the regression analysis. In order to establish the sufficiency of IM_1, the effectiveness of GM characteristic variables as additional regression variables (i.e., in addition to IM_1) can be investigated. In other words, GM characteristic variables cause very little improvement in the regression prediction as regression variables in addition to a "sufficient" *IM*. This improvement may be judged by the reduction in the dispersion of the regression residuals and/or the statistical significance of the regression coefficients corresponding to the GM characteristic variables [11].

6.1. Residual–residual plots

In this study, a simplified statistical approach based on regression is implemented for measuring the effectiveness of GM characteristics as additional regression variables. This method uses a graphical statistical tool known as the residual–residual plot. The main advantage of the residual–residual plots is that

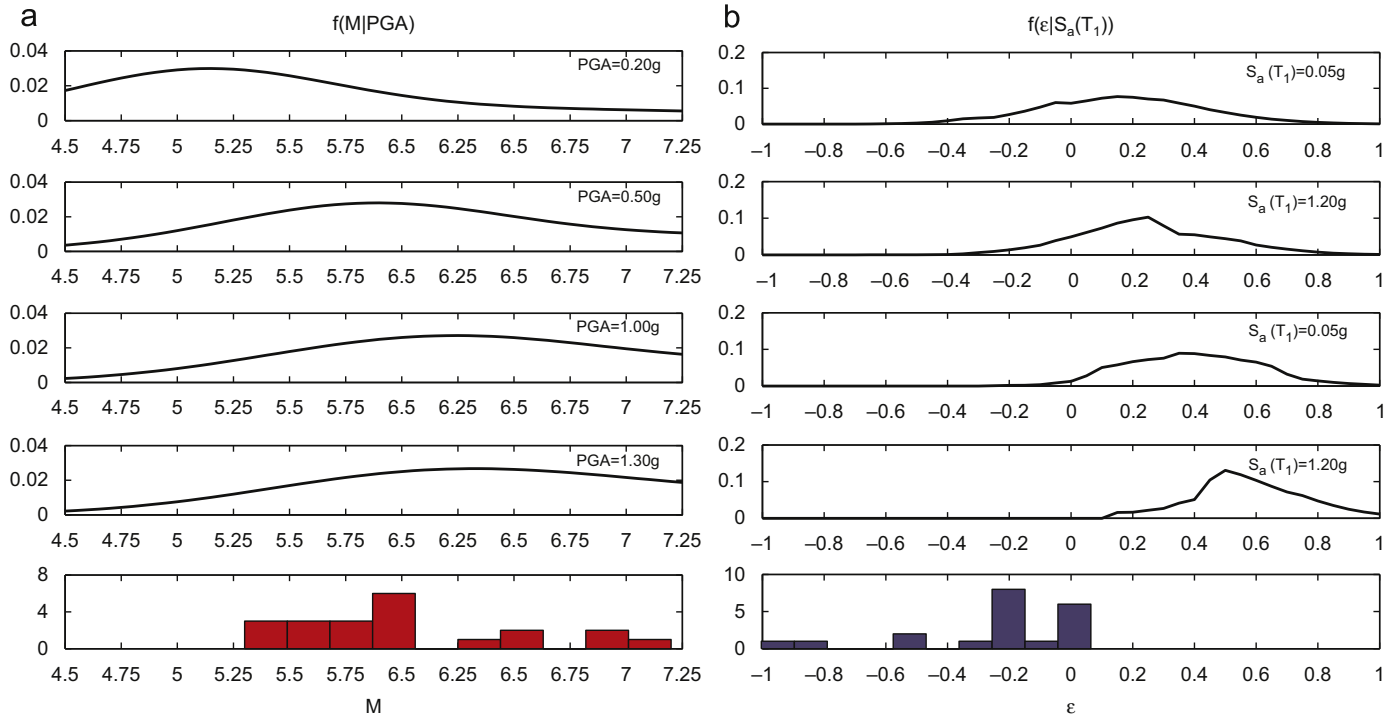

Fig. 5. (a) Conditional probability distribution of M given PGA and histogram of M values for *Sel_A* and (b) conditional probability distribution of ε given $S_a(T_1)$ and histogram of ε values for *Sel_B*.

they offer visual means for judging the improvement caused by an additional regression variable.

Residual–residual plots are constructed by: (a) performing regression of the dependent variable EDP versus the (first) independent variable IM_1 (e.g., PGA or $S_a(T_1)$), (b) performing regression of the second independent variable IM_2 (e.g., M or ε) on the first variable IM_1 and (c) plotting the residuals of the two regressions mentioned above against each other. Roughly speaking, the two regressions on IM_1 "eliminate" possible dependence of both EDP and IM_2 on IM_1. This facilitates investigating the potential dependence of EDP on IM_2, by observing a (statistically) significant trend, in the linear regression between the two sets of residuals explained above. The significance of the trend is measured by both the variability in the residuals of such regression and/or testing the following hypothesis: "*The slope of the regression line is zero*" (i.e., test of hypothesis). The significance of the slope is usually measured by a quantity known as the *p-value*, assuming that the slope of the regression line is a random variable described by *Student's T-distribution* [37]. The hypothesis is rejected (i.e., the slope is non-zero) if the *p-value* is smaller than a certain (small) value, e.g., 0.01.

An effective IM_2 is going to "explain" part of the variability in the data EDP that is not captured by IM_1. In the context of sufficiency, this means that if the standard error of the residual–residual regression is (significantly) smaller than the standard error of the regression of EDP versus IM_1, this will also confirm (in addition to the test of hypothesis) that IM_1 is not sufficient with respect to IM_2. It also indicates that IM_1 and IM_2 together can provide a more "*efficient*" prediction of the EDP, recalling that the efficiency criterion is based on the variability of Y for a given value of the IM_1.

However, it should be noted that in this (simplified) approach the sufficiency of IM_1 is questioned for one IM_2 at the time. This would ignore possible interactions between the GM characteristics themselves. A more thorough approach consists of performing a multi-variable regression of EDP on IM_1 and all of the GM characteristics in question, and test the (joint) hypothesis of whether all the regression coefficients corresponding to the GM characteristics are simultaneously zero. Nevertheless, it is believed that the residual–residual plot approach is still effective in un-covering potential dependencies of EDP on GM characteristics.

6.2. The cloud method

As described above, in order to estimate the statistical properties of the *cloud response*, the linear least squares is applied on EDP versus IM for a suite of GM records (un-scaled) in order to estimate the conditional mean and standard deviation of EDP given IM [38]. This is equivalent to fitting a power–law curve of the form, aIM^b, to the median EDP. The standard deviation of the regression is assumed to be constant with respect to IM over the range of IM's in the cloud.[4] If the adopted IM is sufficient in relation to other GM characteristics, the cloud method is efficient and easy to apply. In cases where the adopted IM_1 is not sufficient with respect to an other GM characteristic IM_2 a vector-valued $\mathbf{IM}=[IM_1,IM_2]$ consisting of two scalar IM's could be adopted. The two-variable linear least squares scheme can be used to estimate the statistical parameters for the EDP conditional on both IM_1 and IM_2 [11,18].

If IM_1 is not sufficient with respect to the GM characteristics a weighted regression scheme may help in reducing the dependence of the residuals (of the "original" regression on IM_1) on GM characteristics. Shome and Cornell [21] implemented the weighted regression scheme in order to take into account the effect of the shape factor in predicting the response of special moment resisting frames. Baker and Cornell [13] have implemented the weighted regression scheme in order to incorporate the information in the GM ε in addition to $S_a(T_1)$. The weighted regression scheme weights each error term

[4] This assumption may be unconservative, in fact for example, for the spectral acceleration it has been proved in [11] that the standard deviation tends to increase for the larger values of spectral acceleration. This stresses the importance of performing linear regression locally, i.e., in a region of IM values of interest.

(residual) proportional to its corresponding variance [39,37]. It can be argued that the variance of each error term and hence the corresponding weight is positively related to the following ratio:

$$w_i \propto \frac{f_{IM_2|IM_1}(z_i|x)_{disaggregation}}{f_{IM_2|IM_1}(z_i|x)_{data}} \tag{13}$$

where $f_{IM_2|IM_1}(z|x)$ is the fraction of the GMs with IM_2 equal to z for a given IM_1 equal to x. In this work, it is assumed that it is equally likely to observe IM_2 given IM_1 for each record in the set; therefore, $f_{IM_2|IM_1}(z_i|x)_{data}$ is going to be equal to $1/N_T$, where N_T is the total number of records. $f_{IM_2|IM_1}(z_i|x)_{disaggregation}$ is the probability that IM_2 is equal to z_i for a given IM_1 equal to x, estimated by disaggregation of seismic hazard.

6.3. The multiple-stripe analysis

In the multiple-stripe analysis, the suite of GM records are scaled to successively increasing values of the *IM* parameter. The set of corresponding *EDP*s calculated for each *IM* value can be referred to as the *stripe response* for the *IM* level [8]. The statistical properties of the stripe responses for various *IM* levels, can be evaluated based on the response to the suite of records, in order to estimate the fragility using Eq. (3) [8,9].[5]

If the adopted IM_1 is not sufficient, a vector-valued $\boldsymbol{IM}=[IM_1,IM_2]$ consisting of two scalar *IM*'s could be adopted. In this case, the linear regression analysis can be employed in order to investigate the dependence of the stripe response for each IM_1 level on the secondary IM_2 parameter [18]. In contrast to the cloud method, the constant coefficients a and b and the standard deviation of the regression residuals, estimated from the observed prediction errors, are re-estimated at every IM_1 stripe and depend on the IM_1 value at each stripe.

If the "primary" IM_1 is not sufficient with respect to the GM characteristics, instead of adopting a vector-valued $\boldsymbol{IM}$, a procedure similar to the one used in the weighted cloud method can be employed, in order to account for additional information available on the correlation between the adopted primary IM_1 and a candidate secondary IM_2 by employing the information extracted from seismic hazard disaggregation. For a given suite of GM records, the stripe response at each IM_1 level can be weighted in relation to the conditional probability distribution $f(IM_2|IM_1)$. This approximate method, which has been proposed by Shome and Cornell [21] and Jalayer [11], incorporates the available information about IM_2 given IM_1 in the seismic assessments based on the scalar IM_1.

The (log of the) median of the stripe response for each *IM* level can be approximated by the expected value (mean) of the (log of the) relevant *EDP*, *Y*, for a given *IM* level x [37,11].[6] Expanding the estimated median (of the log) with respect to IM_2, the mean (of the log) of the stripe response can also be calculated as

$$\ln \hat{\eta}_{Y|IM_1}(x) \cong E[\ln Y(x)] = \sum_{j=1}^{N_{IM_2}} E[\ln Y(x,z_j)|IM_1,IM_2] f_{IM_1|IM_2}(z_j|x) \tag{14}$$

where $\hat{\eta}(.)$ denotes the estimated median for the stripe response. For a selection consisting of N_T GM records the above expression can be calculated by dividing the chosen records into N_{IM_2} bins, in which each bin is represented by value z_j. The $E[\ln Y(x, z_j)]$ is the estimated conditional mean (i.e., sample average) for the natural logarithm of the *EDP* in each bin, represented by value z_j, for a given IM_1 value x, and $f_{IM_2|IM_1}(z_j|x)$ is the probability that IM_2 is equal to z (in bin j) given IM_1 equal to x (e.g. it can be obtained from disaggregation).

In a similar manner, the conditional variance of the natural logarithm of response for a given $IM_1=x$, can be expanded with respect to a candidate IM_2, as

$$\hat{\beta}^2_{\ln Y|IM_1}(x) \cong \sum_{j=1}^{N_{IM_2}} E[(\ln Y(x,z_j))^2|IM_1,IM_2] f_{IM_1|IM_2}(z_j|x) - (E[\ln Y|IM])^2 \tag{15}$$

Here, $E[(\ln Y(x, z_j))^2|IM_1,IM_2]$ denotes the expected squares of the natural logarithm of the structural response *EDP* for the GMs in each bin represented by the value z_j (for a given $IM_1=x$).

As it was stated before, IM_1 is sufficient if it renders the *EDP* conditionally independent of the GM characteristics, for all intensity levels. For a given *IM* level $IM_1=x$, the statistical equivalent to this statement is to establish that the conditional probability distribution for the *EDP* given IM_1 is independent of other GM characteristics, namely

$$f_{Y|IM_1}(x) = f_{Y|IM_1,IM_2}(x,z_j) \tag{16}$$

for any z_j value. The sufficiency criterion can be "approximated" in terms of the (conditional) statistical moments of the response (e.g., conditional mean and variance). For example, a "first-order measure" of the sufficiency criterion can be obtained by establishing that the first (conditional) moment of the response for a given IM_1 level, x, is independent of other GM characteristics

$$\ln \hat{\eta}_{Y|IM_1}(x) \cong E[\ln Y(x)|IM_1] = E[\ln Y(x,z_j)|IM_1,IM_2] \tag{17}$$

If an *IM* is sufficient, it can be demonstrated that the two sides of Eq. (14) will always be equal and does not depend on the probability distribution $f_{IM_2|IM_1}$.[7]

A "second-order measure" of sufficiency criterion can be expressed in terms of the (conditional) second moment of the stripe response

$$E[(\ln Y(x))^2|IM_1] = E[(\ln Y(x,z_j))^2|IM_1,IM_2] \tag{18}$$

for all z_j value. Similarly, it can be demonstrated that the two sides of Eq. (15) will always be equal, independent of $f_{IM_2|IM_1}(z_j|x)$, if the second-order sufficiency is established.

When multiple-stripe analysis is performed for high *IM* levels, it happens quite often that the structural analysis cases do not yield meaningful values. This could either imply that the structure has lost its load bearing capacity or may signal numerical in-convergence. Since both cases are characterized by very large *EDP* values, they are both referred to, for simplicity, as the "collapse cases". In this study, the logistic regression is used in order take into account the collapse cases, that happen in the structural analysis for increasing level of IM_1. In order to explicitly take into account the collapse cases, the stripe response is divided into two parts, namely, the non-collapse and the collapse parts. The logistic regression [40] is applied to the collapse data in order to predict the probability of collapse as a function of the second IM_2 (rather than estimating the probability of collapse as the fraction of records in an IM_1 stripe that cause collapse). Using the indicator variable C to designate the occurrence

[5] The multiple-stripe analysis (MSA) [8] and the incremental dynamic analysis (IDA) [9] would lead to (almost) the same results in terms of the prediction of *EDP* given *IM*. However, unlike MSA, which could (potentially) be done using different sets of records at different intensities, the IDA results pertain to a fixed set of ground motion records. Moreover, unlike MSA, which is performed for a given candidate *IM*, the IDA can be used for screening alternative candidate *IM*'s without redoing the structural analyses.

[6] For a lognormal distribution, the log of the median is exactly equal to the mean of the logarithm.

[7] The above conclusion is based on the fact that the sum of the fractions $f_{IM_2|IM_1}$ is equal to unity.

of collapse (C equals 1 if the record causes collapse and 0 otherwise), the following functional form is fitted [18]:

$$P(C|IM_1=x,IM_2=z)=\frac{e^{a(x)+b(x)z}}{1+e^{a(x)+b(x)z}} \tag{19}$$

where a and b are the coefficients to be estimated for the stripe response at $IM_1=x$. Using the total probability theorem the first term in the integrand of Eq. (4) can be expanded in this way

$$P_{DM|IM_1,IM_2}(DM>y|x,z)=P_{DM|IM_1,IM_2}(DM>y|x,z,NC)\cdot P(NC|IM_1=x,IM_2=z)+1\cdot P(C|IM_1=x,IM_2=z) \tag{20}$$

where

$$P(NC|IM_1=x,IM_2=z)=1-P(C|IM_1=x,IM_2=z) \tag{21}$$

is the probability of not having collapses given $IM_1=x$ and $IM_2=z$. The term $P_{DM|IM_1,IM_2}(DM>y|x,z,NC)$ may be estimated by a lognormal probability distribution whose mean and standard deviation are calculated using Eqs. (14) and (15). Alternatively, in the case of a vector-valued ***IM***, it can be calculated using linear least squares in order to estimate the mean and the standard deviation for the probability distribution $P_{DM|IM_1,IM_2}(DM>y|x,z,NC)$.

7. Numerical results

Distinguished by number of analyses carried out, two alternative procedures are considered in this work: the cloud method and the multiple-stripe method.

7.1. The cloud method

For record selection A (*Sel_A*) the primary IM (IM_1) is *PGA* paired in vector form with M. Instead, for record selection B (*Sel_B*) the primary IM (IM_1) is the $S_a(T_1)$ paired in vector form, with the deviation from the GM prediction relationship ε, that can be regarded as a "shape factor" of the single spectrum in a range of periods in the vicinity of T_1.

In Figs. 6a and 7a the results obtained using the cloud method in order to predict Y adopting as IM_1, *PGA* based on selection A, and $S_a(T_1)$ based on selection B, are, respectively, illustrated. As stated before, a graphical statistical tool known as the residual–residual plot is used in order to establish the sufficiency of IM_1 and the effectiveness of GM characteristic variables adopted as IM_2. The main advantage of the residual–residual plots is that they offer visual means for judging the improvement caused by IM_2. In Figs. 6b and 7b the residual–residual plot related to the introduction of M (for *Sel_A*) and ε (for *Sel_B*) as IM_2, together

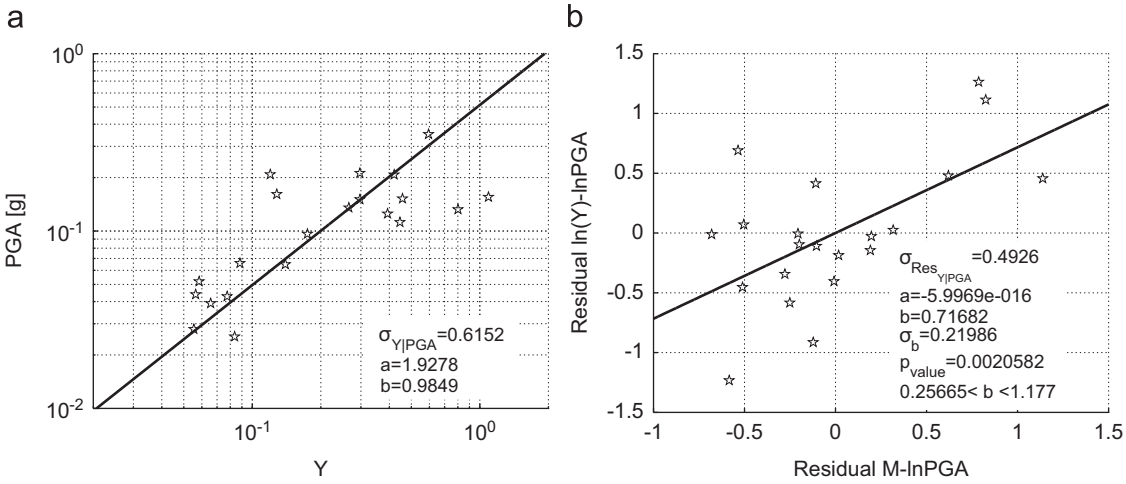

Fig. 6. (a) Simple regression *Y–PGA*, *Sel_A* and (b) residual–residual plot for M as IM_2.

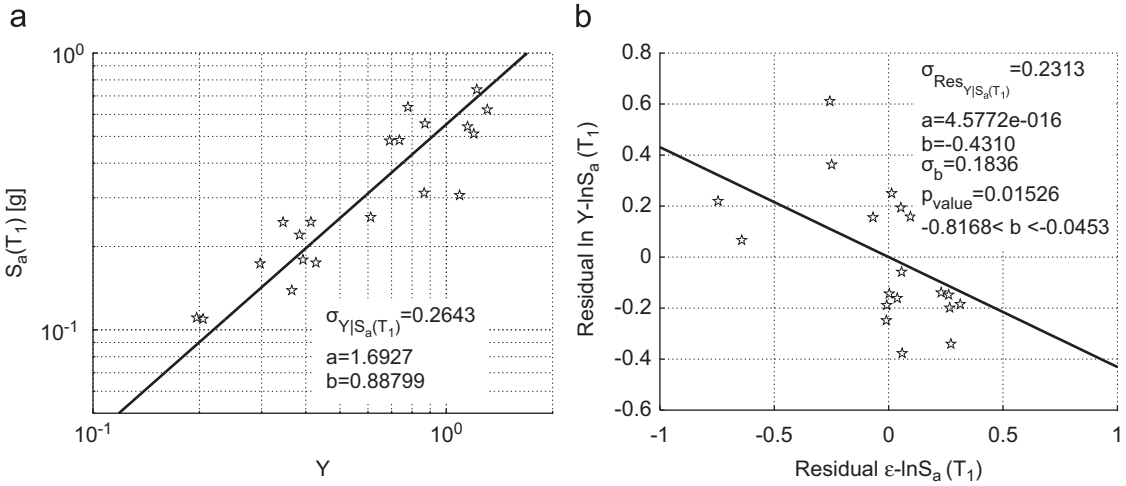

Fig. 7. (a) Simple regression Y–$S_a(T_1)$, *Sel_B* and (b) residual–residual plot for ε as IM_2.

with the *p*-values calculated for the hypotheses test, are, respectively, shown. In Fig. 6b a significant positive trend in the plot can be observed, confirmed by both the *p*-value and the sigma value; this indicates that, based on *Sel_A*, *PGA* is not sufficient with respect to *M*. In Fig. 7b a (statistically significant) negative trend can be observed. This means that, based on *Sel_B*, $S_a(T_1)$ is not sufficient with respect to ε.

Figs. 8a and 9a show the results obtained by the cloud method adopting the vector intensity measure [*PGA*, *M*] for *Sel_A* and [$S_a(T_1)$, ε] for *Sel_B*, respectively. In both cases the multiple regression is used to predict the structural *EDP*, *Y*, as a function of the two chosen *IM*'s.

It would be interesting to study how the seismic risk, represented herein by the MAF of exceeding *Y*, is affected by the weighed regression scheme. For both record selections, the results of weighted regression (Figs. 8b and 9b, respectively) can be used in order to predict the conditional mean and standard deviation of *Y* as a function of IM_1; these parameters are then incorporated in Eq. (3) in order to calculate the structural fragility. It should be noted in Figs. 8b and 9b that (IM_{1i}, Y_i) data pairs are plotted by circles with diameters proportional to the corresponding weight which are proportional to $f(IM_{2i}|IM_{1i})$. The MAF of exceeding *Y* can be calculated from Eq. (2) by integrating the structural fragility and the IM_1 hazard curve. Figs. 14b and 15b illustrate the *PGA* hazard curve and the $S_a(T_1)$ hazard curve extracted from the site of INGV for the coordinates of the site (lat. 40.915; lon. 14.78), respectively.

The MAF of exceeding *Y* has been calculated from Eq. (4) by adopting a vector-valued ***IM***, [*PGA*, *M*] for *Sel_A* and [$S_a(T_1)$, ε] for *Sel_B*. It serves as a benchmark for judging if the weighted regression is helpful in adjusting for the dependence on IM_2 (*M* for *Sel_A*, ε for *Sel_B*). Fig. 10a and b illustrates the results obtained by following the above-mentioned alternative methods for the two considered record selections, respectively.

In Fig. 10a and b the thick lines represent the MAF of exceeding the *Y* adopting as ***IM*** the pair [*PGA*, *M*] and [$S_a(T_1)$, ε], for *Sel_A* and for *Sel_B*, respectively. The thin lines represent the MAF of exceeding *Y* using a scalar IM_1 (*PGA* for *Sel_A* and $S_a(T_1)$ for *Sel_B*). The dashed lines represent the MAF of exceeding *Y* using a scalar IM_1 (*PGA* for *Sel_A* and $S_a(T_1)$ for *Sel_B*) but adjusting for the dependence on IM_2 (*M* for *Sel_A*, ε for *Sel_B*) by weighted regression.

It can be observed that the weighted regression manages to take into account some of the information provided by IM_2 and its corresponding MAF of exceeding *Y* ends up somewhere between those corresponding to the original cloud method and the multiple-regression, respectively. Fig. 10a shows that the weighting scheme is partially effective in taking into account the *M*

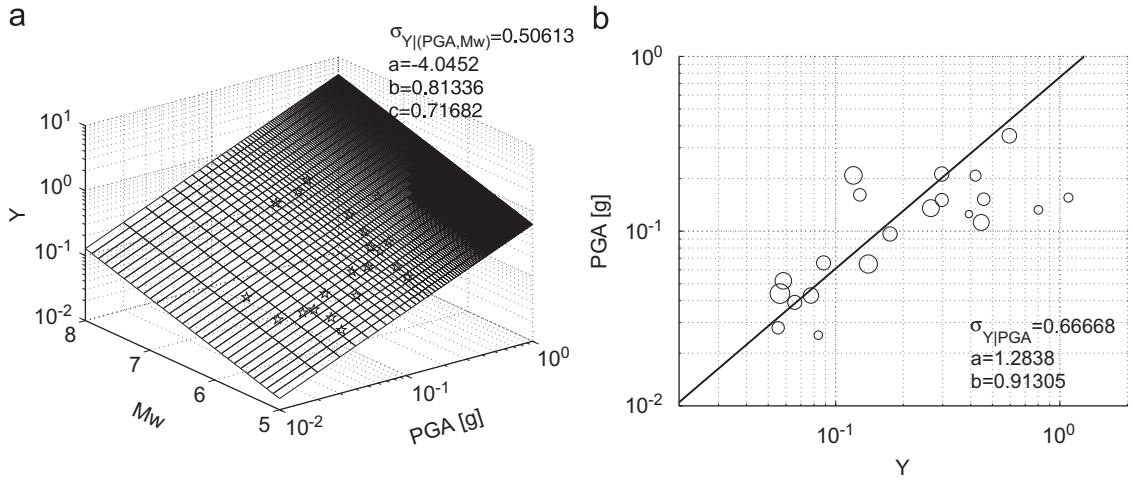

Fig. 8. (a) Multiple regression *PGA–M–Y*, *Sel_A* and (b) weighted regression *PGA–M–Y*, *Sel_A*.

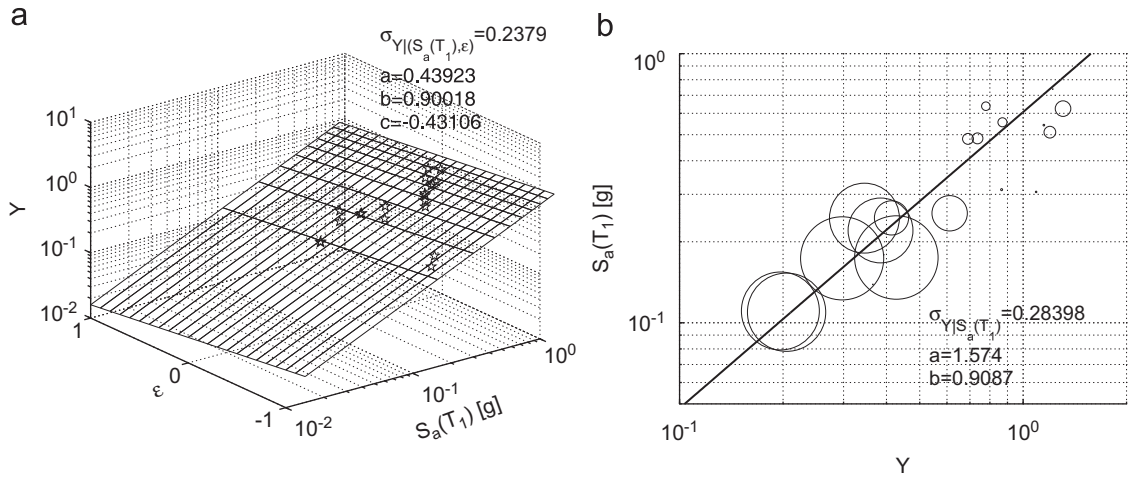

Fig. 9. (a) Multiple regression $S_a(T_1)$–*Y*–ε, *Sel_B* and (b) weighted regression $S_a(T_1)$–*Y*–ε, *Sel_B*.

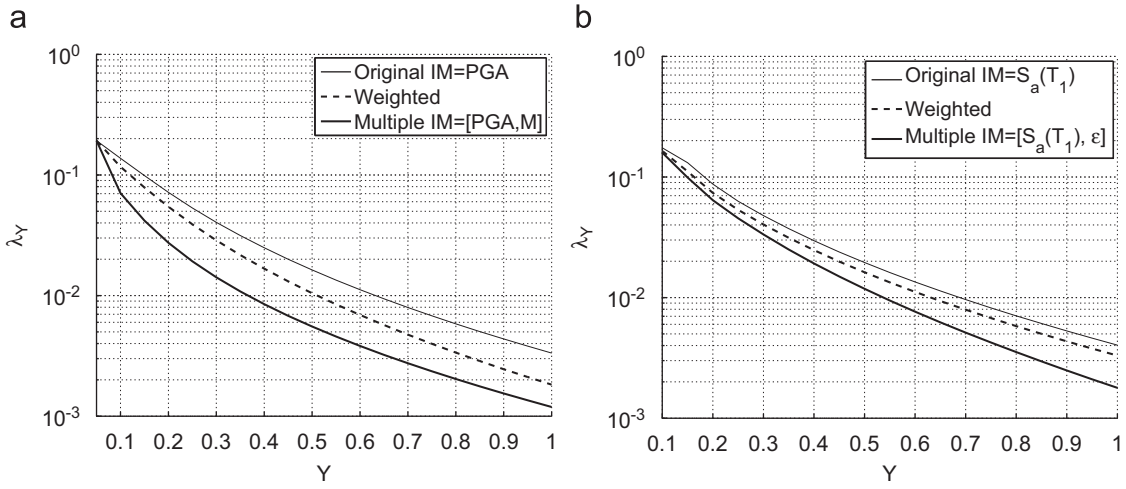

Fig. 10. (a) The MAF of exceeding *Y*, cloud analysis, *Sel_A* and (b) the MAF of exceeding *Y*, cloud analysis, *Sel_B*.

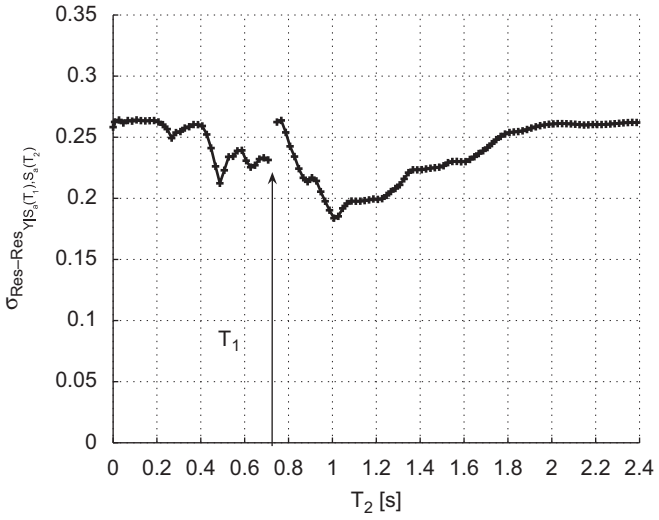

Fig. 11. The standard error of the residual–residual plot for $S_a(T_2)$ as IM_2 for predicting the structural response for different values of the period T_2.

dependence in the prediction of seismic risk for *Sel_A* using *PGA* as *IM*. In the same way, Fig. 10b demonstrates that the seismic risk curve calculated using the weighted regression manages to adjust partially for the dependence on ε, using $S_a(T_1)$ as IM_1 for *Sel_B*. Obviously, the resulting improvement depends both on the distribution of the epsilon values for *Sel_B* and the GM prediction relationship used in the hazard/disaggregation analysis (Fig. 5b).

For *Sel_B*, the choice of ε as IM_2 has been inspired by the fact that it acts as a proxy for spectral shape in the vicinity of the fundamental period of the structure. Therefore, it is interesting to investigate the dependence of *Y* given $S_a(T_1)$ on spectral acceleration at another period $T_2 \neq T_1$, $S_a(T_2)$, in order to consider in an explicit way the shape of the elastic acceleration spectrum [22–24]. The procedure followed for establishing the dependence of *Y* given $S_a(T_1)$, on $S_a(T_2)$ is similar to the one followed in order to investigate the sufficiency of $S_a(T_1)$ with respect to ε. $S_a(T_2)$ is regarded as a second regression variable IM_2, and the residual–residual plots are used to study its efficiency in reducing the dispersion.

It would be interesting to identify the period whose corresponding spectral acceleration is most efficient in predicting the *EDP* conditional on $S_a(T_1)$. The efficiency of the spectral acceleration at a period, $T_2 \neq T_1$, as the second regression variable, has been studied from the point of view of the reduction in the standard error of the regression [11]. This leads to finding an optimal period T_2 at which the dispersion is minimum and the pair $[S_a(T_1), S_a(T_2)]$ has maximum efficiency as an ***IM***. The plot in Fig. 11, illustrates the standard error of the residual–residual regressions performed for a range of periods. It can be observed that choosing the spectral acceleration at $T_2=1$ s leads to the maximum reduction in the residual–residual dispersion. The period corresponding to the first mode of the structure is also marked on the figure. The efficiency of $S_a(T_2=1\text{ s})$ in reducing the standard error of the residual–residual plot signals non-linear behavior in the structural elements resulting in the elongation of the vibration period.

In order to characterize the joint probability distribution for the logarithm of the spectral acceleration at two periods T_1 and T_2, the mean and standard deviation values for both can be extracted from the GM prediction equation [41,42]. In order to completely specify the first and second moments for this pair of spectral values, one needs to evaluate the correlation between $\ln S_a$ values at the two periods. An empirically determined relationship for such correlation is given by Inoue and Cornell [43]

$$\rho_{\ln S_a(T_1),\ln S_a(T_2)} = 1-0.33\left|\ln(T_1/T_2)\right| \quad 0.1\,\text{s} \le T_1,T_2 \le 4\,\text{s} \tag{22}$$

Previous research has established that $\ln S_a(T_1)$ and $\ln S_a(T_2)$ are each marginally normally distributed [44]. Under the mild assumption that they are also jointly normally distributed, one can obtain the conditional mean of $\ln S_a(T_2)$, given $\ln S_a(T_1)$ [41]

$$\mu_{\ln S_a(T_1)|\ln S_a(T_2)=x} = \mu_{\ln S_a(T_2)} + \rho_{\ln S_a(T_1),\ln S_a(T_2)}\sigma_{\ln S_a(T_2)}\left(\frac{x-\mu_{\ln S_a(T_1)}}{\sigma_{\ln S_a(T_1)}}\right) \tag{23}$$

The conditional standard deviation of $\ln S_a(T_2)$ is given as

$$\sigma_{\ln S_a(T_2)|\ln S_a(T_1)=x} = \sigma_{\ln S_a(T_2)}\sqrt{1-\rho^2_{\ln S_a(T_1),\ln S_a(T_2)}} \tag{24}$$

These statistics have been used in order to obtain the MAF of exceeding the critical component demand to capacity ratio *Y*. The conditional mean and standard deviation derived from Eqs. (23) and (24) can be used in order to construct the conditional probability distribution $f(S_a(T_2)|S_a(T_1))$.

Fig. 12a shows the results obtained by the cloud method and the introduction of $S_a(T_2)$ with $T_2=1$ s as the IM_2, using the multiple regression. Fig. 12b illustrates the results of weighted regression analysis exploiting the additional information provided by the conditional probability distribution $f(S_a(T_2)|S_a(T_1))$. The

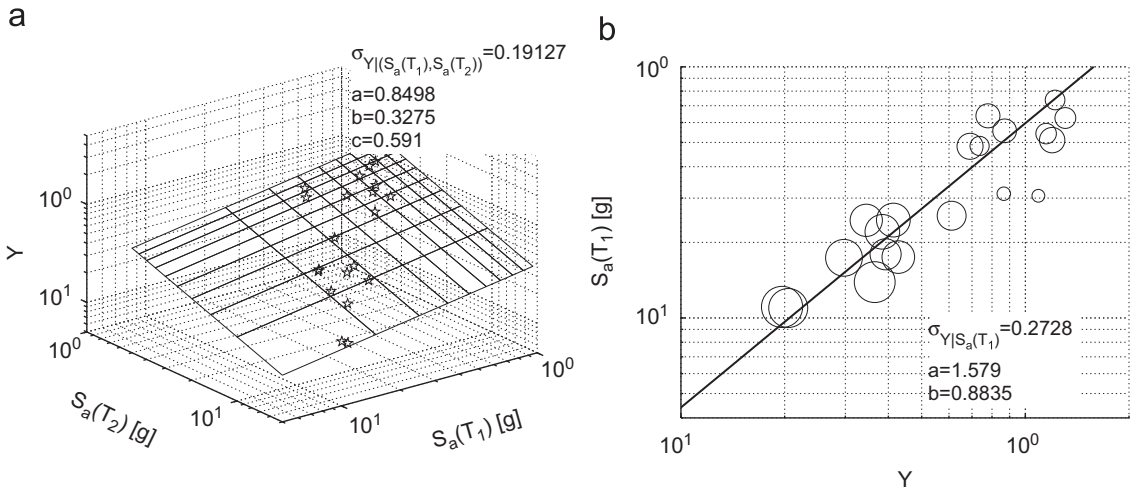

Fig. 12. (a) Multiple regression $S_a(T_1)$–$S_a(T_2)$–Y, *Sel_B* and (b) weighted regression $S_a(T_1)$–$S_a(T_2)$–Y, *Sel_B*.

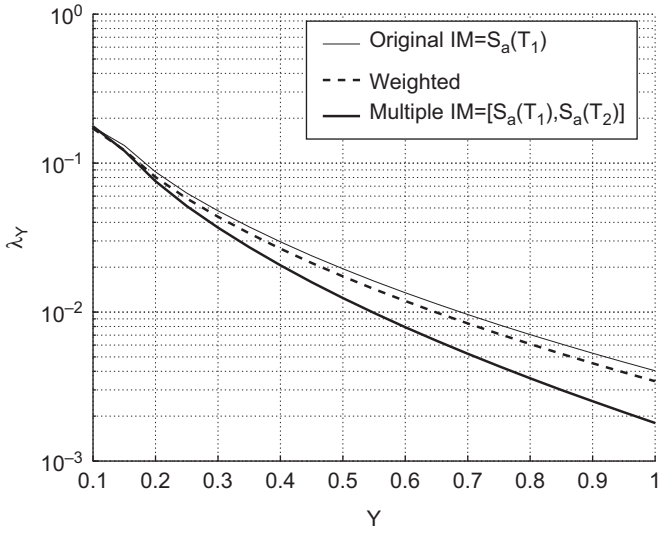

Fig. 13. The MAF of exceeding Y, cloud analysis, *Sel_B*.

($S_{a,i}(T_1)$, Y_i) data pairs are plotted by circles with diameters proportional to the corresponding weight which are proportional to $f(S_a(T_{2i})|S_a(T_{1i}))$. Fig. 13 illustrates the seismic risk curves calculated by following the alternative approaches just discussed.

It can be observed that there is only a small gain in information resulting from incorporating the correlation between the two spectral values inside the weighted regression scheme. The same as the case of $IM_2=\varepsilon$, the weights applied inside the weighted regression are sensitive to both the interval covered by the $S_a(T_2)$ values of the selection of records and the characterization of $f(S_a(T_2)|S_a(T_1))$.

7.2. The multiple-stripe analysis

The results of multiple-stripe analysis are used in order to investigate the efficiency of the weighting scheme based on the information provided by seismic hazard, in the prediction of the MAF of exceeding Y. Figs. 14a and 15a illustrate the results of multiple-stripe analysis, using the same IM's used in the cloud method, for the case-study structure subjected to record selection A and record selection B, respectively.

The critical demand to capacity ratio Y (the stripe response) is calculated for increasing levels of *PGA* (*Sel_A*) and $S_a(T_1)$ (*Sel_B*) using the multiple-stripe method. The lines connecting the (counted) 16th, 50th and 84th percentiles of the stripe response at each IM level are also shown. The numbers near black diamonds in Figs. 14a and 15a indicate the number of "collapse cases" encountered for each level of *PGA* and $S_a(T_1)$, respectively. In Figs. 14b and 15b the peak ground acceleration hazard curve and the spectral acceleration hazard curve (extracted from the site of INGV for the coordinates of the site (lat. 40.915; lon. 14.78)) are shown, respectively.

The statistics of the stripe response (i.e., mean and standard deviation of the logarithm) are calculated for each IM_1 level and incorporated in Eq. (3) in order to calculate the structural fragility. For the vector-valued intensity measures, [*PGA*, M] for *Sel_A* and [$S_a(T_1)$, ε] for *Sel_B*, a simple regression analysis is performed for the stripe response, at each IM_1 level on logarithm of the stripe response versus moment magnitude, $IM_2=M$, and epsilon, $IM_2=\varepsilon$, respectively. The regression results are then used in order to calculate the conditional mean and standard deviation of the logarithm of Y given *PGA* and M (*Sel_A*) and given $S_a(T_1)$ and ε (*Sel_B*). These statistics are incorporated in Eq. (5) in order to calculate the structural fragility. Alternatively, the weighted multiple-stripe analysis is used in order to weigh the stripe response based on the results of seismic hazard disaggregation, $f(M|PGA)$ for *Sel_A* and $f(\varepsilon|S_a(T_2))$ for *Sel_B*. The weighted statistics (mean and standard deviation of logarithm) for the stripe response are calculated from Eqs. (14) and (15). The structural fragility can be calculated from Eq. (3) for the scalar IM based on the weighted statistics. It should be mentioned that the presence of the collapse cases in the stripe response is accounted for by employing a logistic regression scheme as discussed previously. Through numerical integration of the structural fragility with the IM hazard curve, the seismic risk curves for the critical component demand to capacity are calculated (Fig. 16a and b).

Given the increased sophistication and computational effort associated with the multiple-stripe method compared to the cloud method, the position of the hazard curve calculated using the weighted method indicates that the weighting scheme for the stripes analysis is more effective in taking into account the IM_2 dependence (M for *Sel_A* and ε for *Sel_B*) in the prediction of the MAF of exceeding Y with respect to the curve obtained through the cloud method. In other words, in the cloud method the efficiency of the weighted regression scheme depends strongly on the suite of GM records selected, since they are not scaled. Instead, when applying the stripes method, the GM records are

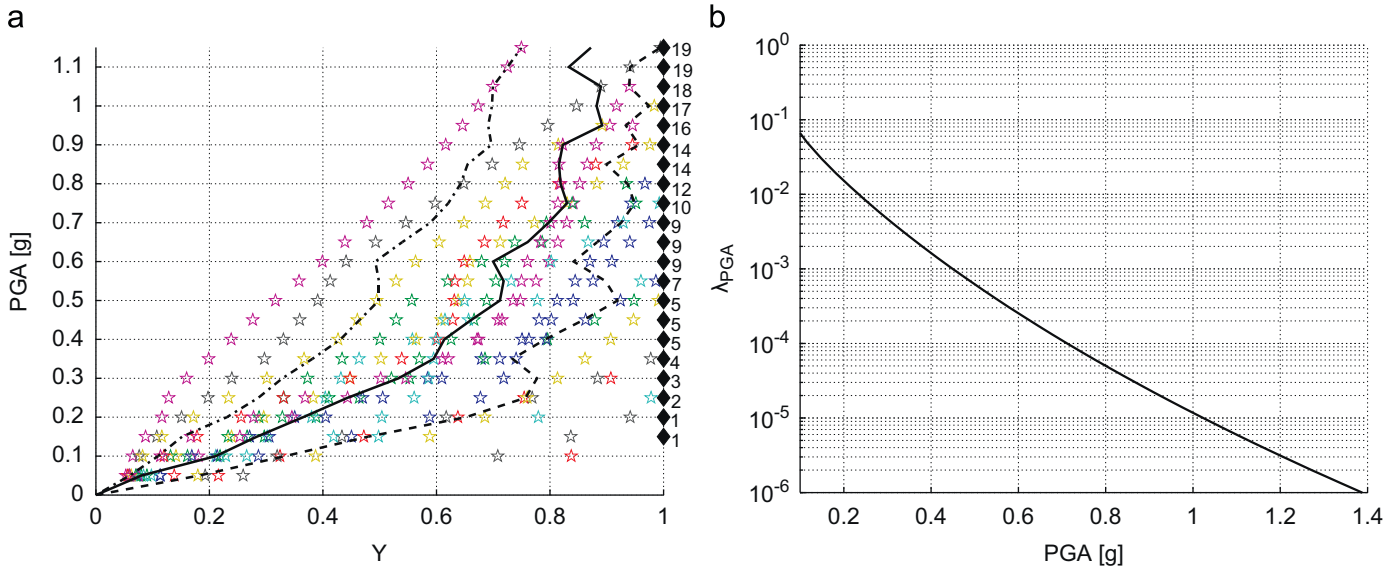

Fig. 14. (a) Results of multiple-stripe analysis for *Sel_A* and (b) peak ground acceleration hazard curve.

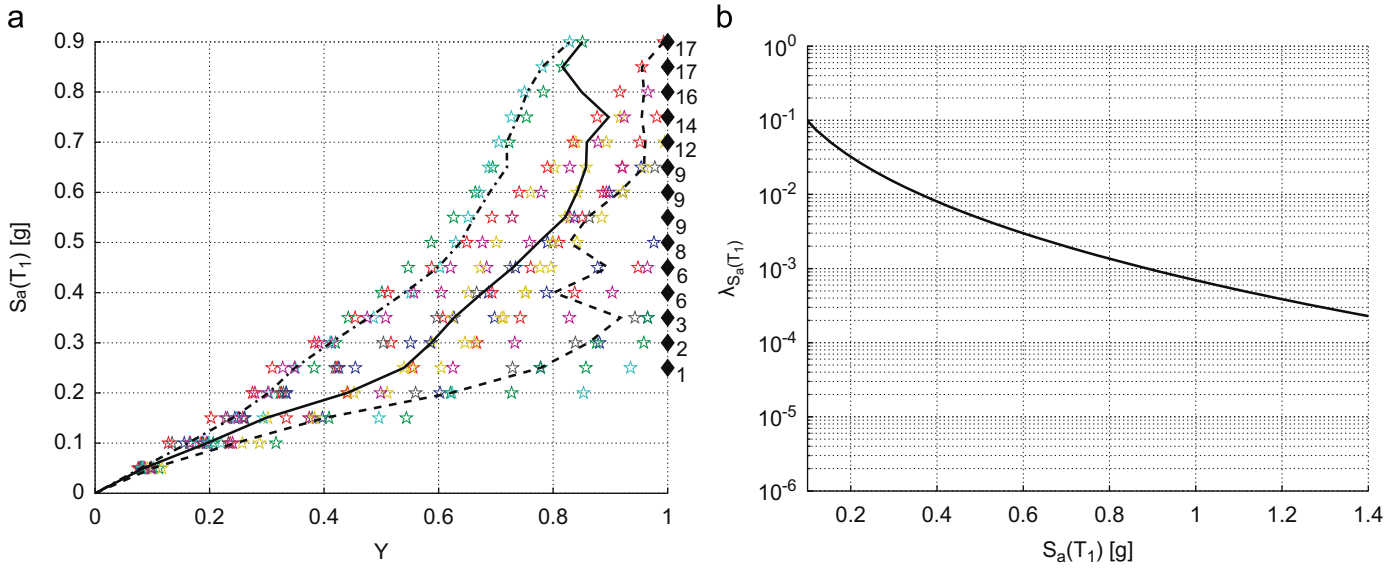

Fig. 15. (a) Results of multiple-stripe analysis for *Sel_B* and (b) spectral acceleration hazard curve for $T_1=0.75$ s.

scaled to successively increasing levels of *IM* and can cover an ample range of the GM intensities.

Next, the pair of *IM*'s $S_a(T_1)$ and $S_a(T_2)$ are also studied in the framework of the multiple-stripe analysis, for *Sel_B*. The conditional probability distributions $f(S_a(T_2)|S_a(T_1))$ of spectral acceleration at $T=T_2$ given spectral acceleration at the fundamental period of the structure are illustrated in Fig. 17; moreover, the histogram of the values of $S_a(T_2)$ of the record selection *B* is shown at the bottom of the figure.

Fig. 18 illustrates the MAF of exceeding *Y*, which is calculated through numerical integration of the structural fragility and the spectral acceleration hazard curve in Fig. 15b.

It is observed that the weighted multiple-stripe analysis does not succeed in including the additional information provided by $IM_2=S_a(T_2)$. It can be noted from Fig. 17, that the conditional probability distributions $f(S_a(T_2)|S_a(T_1))$ do not defer drastically from the histogram of the $S_a(T_2)$ values for selection *B* (i.e., the weights are going to be close to unity). Moreover, the conditional probability distributions seem to remain invariant for large values of $S_a(T_1)$. Hence, the weighted stripe method leads to little improvement compared to the original stripes.

8. Conclusions

In the framework of the performance assessment methodology developed by the PEER Center, the mean annual frequency (MAF) of exceeding a specified engineering demand parameter (*EDP*) is calculated by adopting the ground motion intensity measure (*IM*), as an intermediate parameter to relate the ground motion record characteristics to the structural performance. Non-linear structural dynamic analysis method can be incorporated in this framework in order to evaluate the structural performance given the *IM*. The cloud method (suitable for a limited range of *IM* levels) and the multiple-stripe method (suitable for a wide range of *IM* levels) [8,9] are two such methods considered herein.

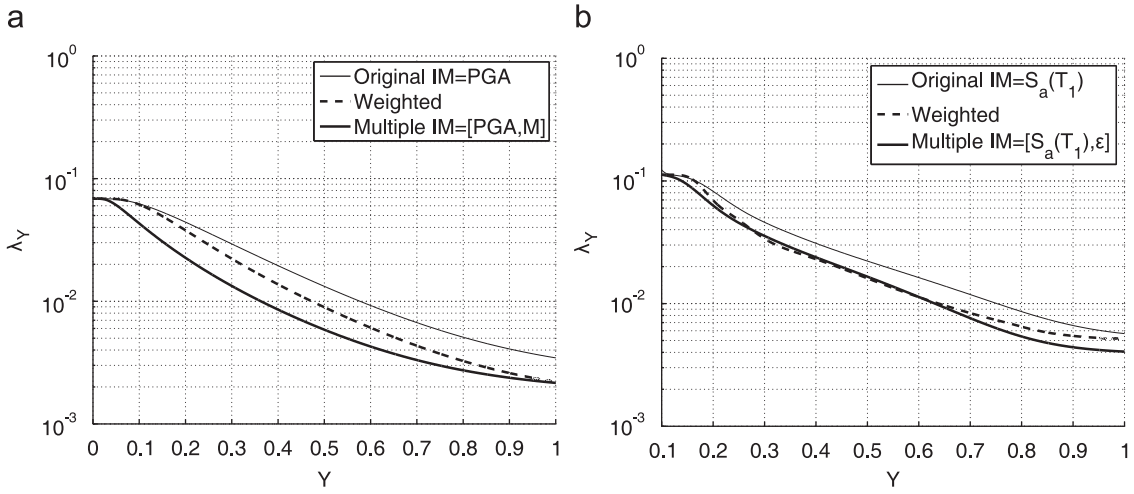

Fig. 16. (a) The MAF of exceeding *Y*, multiple-stripe analysis, *Sel_A* and (b) the MAF of exceeding *Y*, multiple-stripe analysis, *Sel_B*.

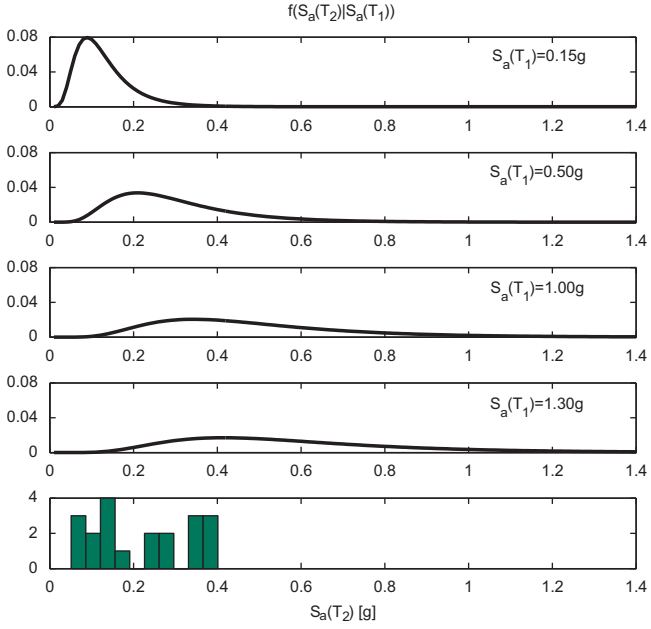

Fig. 17. Conditional probability distribution of $S_a(T_2)$ given some values of $S_a(T_1)$ and histogram of $S_a(T_1)$ values for *Sel_B*.

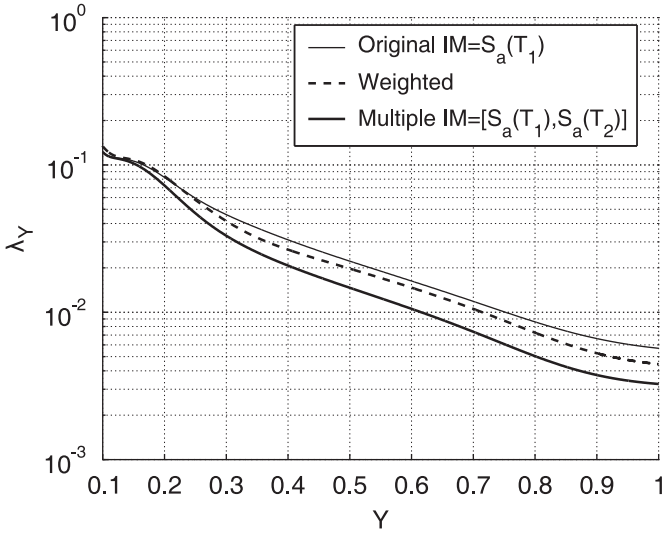

Fig. 18. The MAF of exceeding *Y*, multiple-stripe analysis and *Sel_B*.

This *IM*-based performance assessment methodology implicitly assumes that the structural response depends only on the chosen *IM* and not on any other properties of the ground motion (GM). This criterion is called "*sufficiency*" and states that the structural response *EDP* for a given *IM* is statistically independent of (other) GM characteristics. If the sufficiency criterion is not met, the conditional probability distribution for *EDP* given *IM* will be dependent on (other) GM characteristics. Hence, a biased estimate of the structural response would be obtained if the selected GM records do not match the records that the real structure will be subjected to. In order to increase the sufficiency of an *IM*, additional parameters can be introduced so that the resulting vector-valued ***IM*** would describe more completely the properties of the GM. In particular, a vector intensity measure $\boldsymbol{IM}=[IM_1,IM_2]$ consisting of two parameters can be adopted, where IM_2 strives to capture the GM characteristics not already described by IM_1.

A simple statistical/graphical tool known as the residual–residual plot is employed in this work in order to reveal the possible dependence of the *EDP* conditional on the adopted IM_1, on a candidate IM_2. In cases where the sufficiency for IM_1 is not established, a weighting scheme based on the results of the seismic hazard analysis can be adopted in order to implement the additional information provided by a candidate IM_2. In the cloud method, a weighted regression scheme was applied to the cloud response, weighting each square residual term in relation to the conditional probability distribution $f(IM_2|IM_1)$, obtained through seismic hazard disaggregation. In multiple-stripe analysis, an analogous weighting procedure was employed. In this approach, the stripe response is discretized into a set of IM_2 bins; each bin is then weighted by the probability $f(IM_2|IM_1)$ obtained from seismic hazard and/or seismic disaggregation analysis. In particular the *IM* pairs [*PGA*, *M*], $[S_a(T_1), \varepsilon]$ and $[S_a(T_1), S_a(T_2)]$, $T_2 \neq T_1$ were considered in this work. The conditional probability distributions $f(M|PGA)$ and $f(\varepsilon|S_a(T_1))$ have been calculated through the disaggregation of the seismic hazard for the site of the case-study structure using the Bayesian updating.

The implication of using the weighting scheme has been studied in terms of seismic risk represented herein by the MAF of exceeding the critical component demand to capacity ratio *Y*. The seismic risk curves obtained by adopting the scalar IM_1 and the vector-valued $\boldsymbol{IM}=[IM_1,IM_2]$ are used to benchmark the efficiency

of the weighting scheme. It is observed that the weighting scheme manages to take into account some of the information provided by IM_2 and its corresponding MAF of exceeding Y ends up somewhere between those obtained adopting the scalar IM_1 and the vector-valued $\boldsymbol{IM}=[IM_1,IM_2]$. In this case-study, the weighting scheme proves to be more efficient for multiple-stripe analysis compared to the cloud analysis. This can be attributed to the fact that, the multiple-stripe analysis spans over a wider range of IM levels and therefore, in comparison to the cloud method, may be less sensitive to the selection of GM records. In general, it is observed that the weighted scheme enhances the assessment of the MAF of exceeding EDP for the structure considered.

References

[1] Shome N, Cornell CA, Bazzurro P, Carballo JE. Earthquakes, records, and nonlinear responses. Earthquake Spectra 1998;14(3):411–567.

[2] Iervolino I, Maddaloni G, Cosenza E. Eurocode 8 compliant real record sets for seismic analysis of structures. Journal of Earthquake Engineering 2008;12(1):54–90.

[3] Iervolino I, Maddaloni G, Cosenza E. A note on selection of time-histories for seismic analysis of bridges in Eurocode 8. Journal of Earthquake Engineering 2009;13(8):1125–52.

[4] Iervolino I, Galasso C, Cosenza E. REXEL: computer aided record selection for code-based seismic structural analysis. Bulletin of Earthquake Engineering 2010;8:339–62.

[5] McGuire R. Probabilistic seismic hazard analysis and design earthquakes: closing the loop. Bulletin of the Seismological Society of America 1995;85(5):1275–84.

[6] Bazzurro P, Cornell CA. Disaggregation of seismic hazard. Bulletin of the Seismological Society of America 1999;89(2):501–20.

[7] Convertito V, Iervolino I, Herrero A. Importance of mapping design earthquakes: insights for the Southern Appennines, Italy. Bulletin of the Seismological Society of America 2009;99(5):2979–91.

[8] Jalayer F, Cornell CA. Alternative nonlinear demand estimation methods for probability-based seismic assessments. Earthquake Engineering and Structural Dynamics 2009;38(8):951–72.

[9] Vamvatsikos D, Cornell CA. Incremental dynamic analysis. Earthquake Engineering and Structural Dynamics 2001;31(3):491–514.

[10] Luco N, Cornell CA. Structure-specific scalar intensity measures for near-source and ordinary earthquake ground motions. Earthquake Spectra 2003;23(2):357–92.

[11] Jalayer F. Direct probabilistic seismic analysis: implementing non-linear dynamic assessments. PhD Thesis, Department of Civil and Environmental Engineering, Stanford, CA, 2003.

[12] Iervolino I, Cornell CA. Record selection for nonlinear seismic analysis of structures. Earthquake Spectra 2005;21(3):685–713.

[13] Baker JW, Cornell CA. A vector-valued ground motion intensity measure consisting of spectral acceleration and epsilon. Earthquake Engineering and Structural Dynamics 2005;34:1193–217.

[14] Jalayer F., Cornell CA. A technical framework for probability-based demand and capacity factor design (DCFD) seismic formats, PEER report 2003/08, 2003.

[15] Cornell CA, Krawinkler H. Progress and challenges in seismic performance assessment. PEER Center News 2000;3:2.

[16] Wen YK. Reliability and performance-based design. Structural Safety 2001;23:407–28.

[17] McGuire RK. Seismic Hazard and Risk Analysis. 1st ed. Earthquake Engineering Research Institute; 2004.

[18] Baker JW. Probabilistic structural response assessment using vector-valued intensity measures. Earthquake Engineering and Structural Dynamics 2007;36(13):1861–83.

[19] Sabetta F, Pugliese A. Estimation of response spectra and simulation of non-stationary earthquake ground motions. Bulletin of the Seismological Society of America 1996;86(2):337–52.

[20] Chioccarelli E., Iervolino I. Near-source seismic demand and pulse-like records: a discussion for L'Aquila earthquake. Earthquake Engineering & Structural Dynamics, Early View doi: 10.1002/eqe.987, 2010.

[21] Shome N., Cornell C.A. Probabilistic seismic demand analysis of non-linear structures. Report no. RMS-35, Stanford University, Stanford, CA, (1999).

[22] Vamvatsikos D, Cornell CA. Developing efficient scalar and vector intensity measures for IDA capacity estimation by incorporating elastic spectral shape information. Earthquake Engineering and Structural Dynamics 2005;34(13):1573–600.

[23] Baker JW, Cornell CA. Vector-valued intensity measures incorporating spectral shape for prediction of structural response. Journal of Earthquake Engineering 2008;12(4):534–54.

[24] Mehanny SSF. A broad-range power–law form scalar-based seismic intensity measure. Engineering Structures 2009;31(7):1354–68.

[25] Ordinanza del Presidente del Consiglio dei Ministri (OPCM) no. 3519. "Criteri per l'individuazione delle zone sismiche e la formazione e l'aggiornamento degli elenchi delle medesime zone", Gazzetta Ufficiale della Repubblica Italiana, vol. 108, 2006, (in Italian).

[26] CS..LL.PP, DM 14 gennaio, "Norme Tecniche per le Costruzioni". Gazzetta Ufficiale della Repubblica Italiana, vol. 29, 2008, (in Italian).

[27] CEN, European Committee for Standardisation TC250/SC8/, Eurocode 8: Design Provisions for Earthquake Resistance of Structures, Part 1.1: General rules, seismic actions and rules for buildings, PrEN1998-1, 2003.

[28] Regio Decreto Legge (R.D.L.) 2229. "Norme per l'esecuzione delle opere in conglomerato cementizio semplice o armato", 1939, (in Italian).

[29] Verderame G.M., Stella A., Cosenza E. Le proprietà meccaniche degli acciai impiegati nelle strutture in cemento armato realizzate negli anni '60. *X Convegno Nazionale "L'Ingegneria Sismica in Italia"*, Potenza e Matera 9–13 Settembre 2001b (in Italian), 2001.

[30] Verderame GM., Manfredi G., Frunzio G. Le proprietà meccaniche dei calcestruzzi impiegati nelle strutture in cemento armato realizzate negli anni '60. *X Congresso Nazionale "L'ingegneria Sismica in Italia"*, Potenza-Matera 9–13 settembre 2001a (in Italian), 2001.

[31] Scott MH, Fenves GL. Plastic hinge integration methods for force-based beam-column elements. Journal of Structural Engineering, ASCE 2006;132(2):244–52.

[32] Ordinanza del Presidente del Consiglio dei Ministri (OPCM) no. 3431, "Ulteriori modifiche ed integrazioni all'ordinanza del Presidente del Consiglio dei Ministri no. 3274 del 20 marzo". Gazzetta Ufficiale della Repubblica Italiana no. 107 del 10-5-2005 (Suppl. Ordinario no.85) (in Italian), 2003.

[33] Mander JB, Priestley MJN, Park R. Theoretical stress–strain model for confined concrete. Journal of Structural Engineering 1988;114(8):1804–26.

[34] Jalayer F, Franchin P, Pinto PE. A scalar decision variable for seismic reliability analysis of RC frames. Special issue of Earthquake Engineering and Structural Dynamics on Structural Reliability 2007;36(13):2050–79.

[35] Ditlevsen O, Masden H. Structural Reliability Methods. John Wiley & Sons Inc.; 1996.

[36] Meletti C, Galadini F, Valensise G, Stucchi M, Basili R, Barba S, Vannucci G, Boschi E. A seismic source zone model for the seismic hazard assessment of the Italian territory. Tectonophysics 2008;450:85–108.

[37] Rice JA. Mathematical statistics and data analysis. Belmont, CA: Duxbury/Wadsworth; 1995.

[38] Cornell CA, Jalayer F, Hamburger RO, Foutch DA. Probabilistic basis for 2000 SAC Federal emergency management agency steel moment frame guidelines. Journal of Structural Engineering 2002;128(4):526–33.

[39] Weisberg S. Applied linear regression, Wiley Series in Probability and Mathematics, 2nd ed., 1985.

[40] Neter J, Kutner MH, Nachtsheim CJ, Wasserman W. Applied linear statistical models. Boston: McGraw-Hill; 1996.

[41] Baker JW, Cornell CA. Spectral shape, epsilon and record selection. Earthquake Engineering and Structural Dynamics 2006;35:1077–95.

[42] Cordova PP., Deierlein GG., Mehanny SSF., Cornell CA. Development of a two-parameter seismic intensity measure and probabilistic assessment procedure. In: Proceedings of the 2nd US–Japan workshop on performance-based seismic design methodology for reinforced concrete building structures, PEER report 2000/10, Pacific Earthquake Engineering Research Center, University of California, Berkeley, 2000.

[43] Inoue T., Cornell CA. Seismic hazard analysis of multi-degree-of-freedom structures. Reliability of Marine Structures, RMS-8, Stanford, CA, 1990, 70pp.

[44] Abrahamson NA. State of the practice of seismic hazard evaluation. GeoEng 2000, Melbourne, Australia, 19–24 November 2000, pp. 659–85.