Contents lists available at ScienceDirect



Soil Dynamics and Earthquake Engineering

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

Implications of GMPE's structure for multi-site seismic hazard

Iunio Iervolino^{a, b, *}

^a Dipartimento di Strutture per l'Ingegneria e l'Architettura, Università degli studi di Napoli Federico II, via Claudio 21, 80125, Naples, Italy
^b IUSS – Scuola Universitaria Superiore di Pavia, piazza Vittoria 12, 27100, Pavia, Italy

ABSTRACT

Previous research has discussed the implications of the structure of a classical ground motion prediction equation (GMPE) on single-site probabilistic seismic hazard analysis and disaggregation. Classical refers to a GMPE where local site conditions (or any other factor) are accounted for via constant (with respect to magnitude and source-to-site distance) terms added to the mean and that do not affect the distribution of the residuals. Herein, the implications of such a structure of the GMPE are briefly discussed with respect to multi-site hazard assessment that, typically, requires a large number of simulations of random fields of ground motion intensity measures. It is shown that this type of GMPEs enables to run the simulations only once, independently of the soil conditions (or any other factor modeled in a similar way) eventually assigned to each site, which can represent a significant computational advantage in the case of spatially distributed assets.

1. Introduction

In [1], which this short note complements, ground motion prediction equations (GMPEs) of the type:

$$log(IM) = \mu_{m,r} + \theta + \varepsilon, \tag{1}$$

were discussed. This equation models the logarithm of a ground motion intensity measure (IM) at a specific site as a Gaussian random variable (RV) with mean, $\mu_{m,r} + \theta$, which depends on the earthquake magnitude (M) and source-to-site distance (R), as well as soil site conditions via the θ coefficient (usually, there is a reference condition for which $\theta = 0$; e.g., rock). The ε term of represents the residual, which is a zero mean Gaussian RV with variance σ^2 .

If the soil effect is constant with respect to $\{M = m, R = r\}$ and does not affect the variance of log(IM) (i.e., σ^2 is independent of θ), it was demonstrated that the results of single-site probabilistic seismic hazard analysis (PSHA) can be adjusted for different site classes shifting (in logarithmic scale) those for a reference soil condition by a factor equal to θ . It was also shown that disaggregation of probabilistic seismic hazard does not change with the soil category, a result that, in general, does not hold when PSHA is run by means of a logic tree, yet applies to each of its branches.

Although GMPEs with more complex functional forms, for which this reasoning does not apply, are becoming more common [2], those of the type in equation (1) are the still the majority [3], and they are still used in nationwide hazard studies [4] or ShakeMap [5]. Therefore, it is worthwhile to address also the advantages of a structure of type in

equation (1) with respect to multi-site hazard assessment, that is, when one is interested in probabilities of events involving IMs at a set of sites that can be hit by the same earthquake.

Typically, when seismic risk of spatially distributed assets is of concern, the simulation of IMs at the sites is the most computationally demanding part of the analysis [6]. On the other hand, the information on local conditions at a large scale is poor and may need to be adjusted eventually. Moreover, this kind of studies often requires *what-if* analyses, that is, the risk evaluation is preformed under different assumptions to assess sensitivity with respect to assigned conditions. In this framework, it is convenient to prove cases for which simulations do not need to be run again. This is the focus of the short study presented herein.

The remainder of the paper is structured such that modeling of random fields of ground motion intensity measures at multiple sites in one earthquake is recalled first. Then, the insensitivity of the covariance matrix to the site conditions is addressed. Two illustrative examples, referring to a large-scale multi-site hazard assessment in Italy, show the practical implications. A summary with final remarks closes the short note.

1.1. Random field of IMs conditional to the occurrence of one earthquake

The effect of one earthquake hitting multiple sites, *s* in number, may be defined as the vector collecting the (logarithms of) the IMs it produces, $\{log(IM_{1,i}), log(IM_{2,i}), ..., log(IM_{s,i})\}$, at the sites. With respect to equation (1), it is useful to introduce the subscripts j = 1, ..., s that identify each of the sites, and *i* that specifies the earthquake. In the

https://doi.org/10.1016/j.soildyn.2023.108022

Received 1 March 2023; Received in revised form 6 May 2023; Accepted 10 May 2023 Available online 23 May 2023 0267-7261/© 2023 Elsevier Ltd. All rights reserved.



^{*} Dipartimento di Strutture per l'Ingegneria e l'Architettura, Università degli studi di Napoli Federico II, via Claudio 21, 80125, Naples, Italy. *E-mail address: iunio.iervolino@unina.it.*

framework of multi-site hazard, a GMPE models the logarithm of IM at the generic site j due to earthquake i, as:

$$log(IM_{j,i}) = E \left[log(IM_{j,i}) \middle| m_i, r_{j,i}, \theta_j \right] + \eta_i + \varepsilon_{j,i},$$
⁽²⁾

where $E[log(IM_{j,i})|m_i, r_{j,i}, \theta_j]$ is the mean of $log(IM_{j,i})$ conditional to magnitude, source-to-site distance and the site condition θ_j (or any other factor modeled in the same manner). The term η_i denotes the inter-event residual, which is constant for all sites in the *i*-th earthquake, while $\varepsilon_{j,i}$ is the intra-event residual of the logarithms of IM at site *j* in earthquake *i*. Inter-event residuals are assumed to be stochastically independent from intra-event residuals, and both are assumed to be normally distributed at each site, with zero mean and with variance σ_{inter}^2 and σ_{intra}^2 , respectively. Thus, $log(IM_{j,i})$, at site *j* and conditional to $\{M = m_i, R = r_{j,i}\}$ of earthquake *i*, is a Gaussian RV, with mean $E[log(IM_{j,i})|m_i, r_{j,i}, \theta_j] = \mu_{m_i, r_{ij}} + \theta_j$ and variance $\sigma^2 = \sigma_{inter}^2 + \sigma_{intra}^2$.

It is generally assumed that { $log(IM_{1,i}), log(IM_{2,i}), ..., log(IM_{s,i})$ } form a Gaussian random field (GRF), that is, the logarithms of the IMs have a multivariate normal distribution conditional to the features of the *i*-th earthquake. The mean vector of the GRF is given by $E[log(IM_{i,j})|m_i,r_{j,i},\theta_j]$, i = 1,...,s, and the covariance matrix, Σ , is of the type as in equation (3):

$$\boldsymbol{\Sigma} = \sigma_{inter}^{2} \cdot \begin{bmatrix} 1 & 1 & \cdots & 1 \\ 1 & 1 & \cdots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \cdots & 1 \end{bmatrix} + \sigma_{intra}^{2} \cdot \begin{bmatrix} 1 & \rho_{1,2} & \cdots & \rho_{1,s} \\ \rho_{2,1} & 1 & \cdots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{s,1} & \rho_{s,2} & \cdots & 1 \end{bmatrix},$$
(3)

where, $\rho_{j,h}$, is the correlation coefficient between intra-event residuals at two generic sites $\{j,h\}$, j = 1, 2, ..., s, h = 1, 2, ...s [13]. In equation (3) the subscript *i* has been dropped because, per equations (1) and (2), the variance of the residuals does not depend on magnitude and location of the earthquake and, consequently, also Σ does not explicitly depend on *i*. Most importantly, because equation (1) implies that the RV representing the logarithms of IM for a site with conditions represented by θ , is obtained adding such a coefficient to the RV representing the logarithms of IM for a reference condition for which $\theta = 0$, the covariance matrix of the GRF is also independent of the conditions of the sites, that is:

$$Cov[log(im_{j,i}), log(im_{h,i})] = Cov\{E[log(IM_{j,i})|m_i, r_{j,i}, \theta_j] + \eta_i \\ + \varepsilon_{j,i}, E[log(IM_{h,i})|m_i, r_{h,i}, \theta_h] + \eta_i + \varepsilon_{h,i}\}$$
(4)
$$= \sigma_{inter}^2 + \rho_{j,h} \cdot \sigma_{intra}^2,$$

where Cov stands for covariance. In other words, the covariance of the logarithms of the IMs at different sites in the same earthquake is simply the covariance of the residuals, regardless of site conditions.¹

The described modeling of random fields of IMs has important practical consequences for multi-site hazard assessment.

For example, one may seek the probability of observing, in a given time interval, a specific number of exceedances of a vector of ground motion intensity measure thresholds, $\{log(im_1^*), log(im_2^*), ..., log(im_s^*)\}$, each one corresponding to one of the sites. Multi-site hazard typically requires simulating several times, say *n*, GRF realizations, that is, to simulate the vector collecting the IMs at the sites conditional to the occurrence of one earthquake: $\{log(IM_{1,i}), log(IM_{2,i}), ..., log(IM_{s,i})\}, i = 1, 2, ...n.$ If the simulations are carried out considering the same reference site condition for all sites (e.g., rock) a consequence of the covariance structure discussed is that if the analyst wants to assign site conditions different from the reference one for an arbitrary subset of sites, then the simulations do not require to be run again. The realizations on different soil conditions may be

readily obtained from those for the reference case just adding to them the desired soil coefficients, $\{\theta_1, \theta_2, ..., \theta_s\}$, from the GMPE. In other words, the *i*-th realization on arbitrary soil conditions is simply $\{log(im_{1,i}) + \theta_1, log(im_{2,i}) + \theta_2, ..., log(im_{s,i}) + \theta_s\}$ $\forall i$. In fact, it is even simpler, yet has the same effect, to keep the simulations from the reference conditions and compare them with the thresholds adjusted subtracting the soil coefficients as: $\{log(im_1^*) - \theta_1, log(im_2^*) - \theta_2, ..., log(im_s^*) - \theta_s\}$.

2. Illustrative applications

2.1. The number of sites experiencing at least one exceedance in a time interval

For this application, the sixty-eight seismic station sites from the study of [7] that aimed at validating the official Italian hazard map [8], are considered. Sites and local site conditions, in terms of Eurocode 8 or EC8 [9] classification, are given in Fig. 1a.

The considered IM is the peak ground acceleration (PGA). The scope is to compute the probability mass function (PMF) of the number of sites, among those considered, experiencing in thirty years at least one exceedance of 0.1g, that is the threshold vector is $log(im_i^*) = log(0.1), j =$ 1, 2, ..., 68. (Thresholds are equal for the sites for simplicity only.) To compute the sought distribution, it is required to simulate the random fields of PGA at the sites. To this aim, the source model for the Italian hazard map is considered; it is made of thirty-six areal source zones shown in Fig. 1a. As it regards the seismic features of each zone (i.e., earthquake rates and magnitude distributions), the branch named 921 of the study by Ref. [8], is considered. The seismicity rates for this branch are per magnitude bins (above a minimum magnitude for each zone) and are given in Ref. [10]. The GMPE of branch 921 is that of [11], which has only one residual term. Herein, although not strictly necessary, the more recent GMPE by Ref. [12] is used, as it is consistent with equation (2). Moreover, the spatial correlation model of [13], for intra-event residuals, is also considered. Then, the seismic history spanning thirty years at the sixty-eight sites is simulated one-hundred-thousand times, using the REASSESS software [14], via the following steps.

- (i) A realization of the number of earthquakes, say *n*, occurring at the sources zones in thirty years is obtained sampling a Poisson distribution with mean 30·ν_{tot}, where ν_{tot} is the sum of the annual rates (ν_z) of occurrence of earthquakes above the minimum magnitude, of the thirty-six zones: ν_{tot} = Σ³⁶_{z=1} ν_z.
- (ii) For each earthquake, i = 1, 2,...,n, simulated in the previous step, the seismic source zone where it occurs is sampled according to the probability that when a seismic event occurs it is from zone z: P[z is the zone of the earthquake |earthquake i occurs] = \nu_z/\nu_{tot}, \forall z.
- (iii) The earthquake's magnitude and location for earthquake *i* are simulated. The magnitude value is sampled from the distribution for the source where the earthquake occurs, while the location coordinates are simulated assuming that these are uniformly distributed over the source zone.
- (iv) Magnitude and location of the *i*-th earthquake are used to compute the mean of the logarithms of PGA on rock (A-type site conditions according to EC8, that is *rock* site conditions) at each of the sites.
- (v) One value of the interevent residual of the GMPE for earthquake *i* is sampled from a Gaussian distribution with zero mean and variance σ_{inter}^{2} .
- (vi) Sixty-eight values of the intra-event residuals, one for each site, of the GMPE for earthquake *i* are sampled from a multivariate Gaussian distribution with null mean vector and covariance matrix given by the second term at the right-hand side of equation (3).
- (vii) The inter-event residual realization (the same for all sites) and the intra-event residual realization (one for each site) from the steps

¹ It also follows that the covariance of IMs recorded at different sites in different earthquakes is zero, which also applies to different IMs recorded in different earthquakes at the same site. It is because classical assumptions of PSHA [15].



Fig. 1. (a) Considered sites with conditions according to EC8, and the seismic source zone model used in Ref. [8]. (b) PMF of the number of sites experiencing in thirty years at least one exceedance of 0.1g.

(v) and (vi) are added to $E[log(IM_{j,i})|m_i,r_{j,i}]$, to compute the PGA at each site in the *i*-th earthquake event (zero PGA is assigned to sites distant more than 200 km from the earthquake location, because of the applicability limits of the GMPE.). Then, the number of sites where exceedance of 0.1g is observed, in the *i*-th earthquake, is counted.

- (viii) Repeating steps from (ii) to (vii) for each of the n earthquakes simulated in the thirty years under simulation, enables determining the number of different sites, among the sixty-eight considered, experiencing at least one exceedance of 0.1g, say they are k.
- (ix) Repeating one-hundred-thousand times (an arbitrary number of times used in this application) steps (i-v) enables to collect the values of k to build the distribution of the number of sites experiencing at least one exceedance of 0.1g in thirty years; i.e., P[k sites experience exceedance of 0.1g in 30 years], k = 0, 1, ..., 68.

The resulting distribution, which is – by design – the PMF when the site condition is A-type for all sites, is given in Fig. 1b. It has mean and variance equal to 8 and 14 respectively.

Recognizing that, in fact, the site conditions are different for some sites (Fig. 1a) it is sufficient to update the thresholds. The updated thresholds are obtained as $log(im_j^*) = log(0.1) - \theta_j \forall j$, where θ_j is the appropriate coefficient for the site according to the GMPE [12]. (It adopts the same EC8 soil classification, that is, it has five coefficients for conditions from A-type to E-type.) At this point, counting exceedances with respect to the updated thresholds, which does not require to simulate again the PGA at the sites, immediately allows to retrieve the PMF of the sites experiencing at least one exceedance of 0.1g in thirty years considering the actual site conditions. The resulting distribution is also given in Fig. 1b; its mean and variance are equal to 15 and 23, respectively.

2.2. The total number of exceedances at multiple sites in a time interval

As a second application, the distribution of the total number of exceedances of 0.1g at the sixty-eight sites in thirty years is computed. This sought PMF is that of a RV defined between zero and plus infinity. To compute it, the same steps from (i) to (ix) of the simulation above

could be retained. In fact, the IMs simulated in step (vii) serve to count how many exceedances are observed, collectively among the sites, in the *i*-th earthquake (i.e., between zero and sixty-eight). Step (viii) serves to sum-up the exceedances observed in the *n* earthquakes constituting the thirty-years span in question, and (ix) means repeating the same procedure for several thirty-years' realizations.

In fact, the simulations were performed in a way that features two phases and can be considered smarter: (1) the distribution of the total number of exceedances conditional to the occurrence of one *generic* earthquake (i.e., of unspecified magnitude and location) is computed; (2) such a distribution is combined with the Poisson distribution to get the PMF of the total number of exceedances in thirty years. The steps of the phase (1) are summarized here.

- (i) For each source zone, the distribution of the total number of exceedances at the sixty-eight sites, conditional to the occurrence of one generic earthquake, is built. If the RV is indicated as U, then the built distribution is that providing $P[u \text{ exceedances of } 0.1g \text{ at the sites } | \text{occurrence of one} earthquake on source } z], <math>u = 0, 1, ..., 68$. To compute it requires sampling earthquake magnitude and location for zone z.
- (ii) The magnitude and location of the earthquake from the previous step are used to compute the mean of the logarithms of PGA on rock (A-type soil conditions) at each of the sites. Moreover, one inter-event and sixty-eight intra-event residual values are sampled and added to the mean of the GMPE, as in steps from (iv) to (vii) of the previous application, to compute the PGAs at the sites and to count the exceedances in the earthquake.
- (iii) Repeating steps (i) and (ii) (for one-hundred-thousand earthquakes herein) allows to build the distribution of the total number of exceedances at the sites given the occurrence of a generic earthquake on the source in question.

Steps (i)-(iii) are repeated for each source zone. The obtained distributions are combined, via the total probability theorem, to get the distribution of the total number of exceedances at the sites given the occurrence of one generic earthquake randomly across all zones:

 $P[u \text{ exceedances of } 0.1g \text{ at the sites } | \text{occurrence of one generic earthquake}] = \sum_{z=1}^{30} (\nu_z / v_{tot}) \cdot P[u \text{ exceedances of } 0.1g \text{ at the sites } | \text{occurrence of one earthquake on source } z].$



Fig. 2. (a) PMF of the total number of exceedances of 0.1g at the sites conditional to the occurrence of one generic earthquake; (b) PMF of the total number of exceedances of 0.1g at the sites in thirty years.

The obtained distribution is given in Fig. 2a. The distribution refers to A-type (i.e., rock) site conditions. To adjust it for the actual soil conditions of the sites is sufficient, in performing step (ii), to adjust the threshold subtracting the soil coefficient from the GMPE from 0.1g, in log scale. In fact, the PMF obtained adjusting the thresholds for soil conditions is also given in Fig. 2a. At this point, the phase (2) can be performed as described in the following.

- (iv) A realization of the number of earthquakes, say *n*, occurring in the country in thirty years is obtained sampling from a Poisson distribution with rate $30 \cdot \nu_{tot}$.
- (v) For each earthquake i = 1, 2, ..., n the PMF from the phase (1) is sampled; i.e., for each earthquake, a realization of the total number of exceedances of 0.1g at sites is obtained. The PMF to sample is the one providing the total number of exceedances in one generic earthquake (i.e., that of Fig. 2a).
- (vi) Repeating steps (iv)-(v) for the *n* earthquakes allows to count, in the thirty years under simulation, the total number of exceedances of 0.1g among the sixty-eight considered, say they are *k*.
- (vii) Repeating one-hundred-thousand times steps (iv)-(vi) allows to collect the *k*-values to build the distribution of the total number of exceedances of 0.1g in thirty years at the considered sites;
 i.e., *P*[*k* exceedances of 0.1g are observed collectively at sites in 30 years], *k* = 0, 1, ..., +∞.

The resulting distribution is given for two cases in Fig. 2b. The curve referring to A-type conditions has been computed sampling the corresponding PMF of Fig. 2a. The one referring to soil has been computed sampling the distribution of Fig. 2a adjusted for the actual soil site conditions for each of the sites. It is worthwhile to note that the mean and the variance of the total number of exceedances at the sites in thirty years are 9 and 18, respectively, for all-rock conditions, while they are 18 and 43, respectively, when the actual soil site conditions are considered.

3. Summary

This short note dealt with the implications of the structure of ground motion modeling on multi-site probabilistic seismic hazard analysis. In particular, the practical advantages of ground motion prediction equations, where the site effect is represented by a coefficient added to the mean, invariant with respect to magnitude and distance, and that does not affect the standard deviation of the residuals, were discussed.

In the context of typical modeling of random fields of ground motion intensity measures in one earthquake, the covariance matrix of IMs at multiple sites is insensitive to the site conditions. Because multi-site hazard requires large simulations of the mentioned random fields, this is a significant advantage for large scale studies involving many sites, as these simulations are computationally demanding and the information about soil conditions may be poor and may need to be adjusted eventually, or what-if analyses may be required. The practical implications were illustrated via two nationwide examples, showing that the sought results can be obtained simulating the IMs on rock and adjusting the thresholds for the actual sites' conditions eventually.

The discussed results, which may be useful in the context of risk assessment of spatially distributed assets, apply to any other factor, beyond site conditions, which is modeled similarly in the GMPE.

4. Data and resources

The location of the recording stations in the applications have been kindly provided by Dario Albarello (Università degli Studi di Siena, Italy) and Vera D'Amico (*Istituto Nazionale di Geofisica e Vulcanologia*, Italy). The soil conditions of the recording stations were retrieved from http://itaca.mi.ingv.it/(last accessed in December 2016).

Author statement

Iunio Iervolino: Conceptualization, Methodology; Investigation; Writing- Original draft, Writing- Review & editing.

Declaration of competing interest

The author, Iunio Iervolino, declares no competing interests.

Data availability

No data was used for the research described in the article.

Acknowledgements

This note benefitted of discussions with Massimiliano Giorgio (Università degli Studi di Napoli Federico II, Italy). Pasquale Cito (Università degli Studi di Napoli Federico II, Italy) contributed to the illustrative examples.

References

- Iervolino I. Soil-invariant seismic hazard and disaggregation. Bull Seismol Soc Am 2016;106(4):1900–7. https://doi.org/10.1785/0120160072.
- [2] Gregor N, Abrahamson NA, Atkinson GM, Boore DM, Bozorgnia Y, Campbell KW, et al. Comparison of NGA-west2 GMPEs. Earthq Spectra 2014;30(3):1179–97. https://doi.org/10.1193/070113EQS186M/ASSET/IMAGES/LARGE/10.1193_ 070113EQS186M-FIG12.JPEG.
- [3] Douglas J. Fifty years of ground-motion models. In: 2ECEES 2nd European conference on earthquake engineering and seismology. Instanbul, Turkey; 2014. https://doi.org/10.1785/0120050137.
- [4] Meletti C, Marzocchi W, D'Amico V, Lanzano G, Luzi L, Martinelli F, et al. The new Italian seismic hazard model (MPS19). Ann Geophys 2021;64(1):1–29. https://doi. org/10.4401/ag-8579.
- [5] Michelini A, Faenza L, Lanzano G, Lauciani V, Jozinović D, Puglia R, et al. The new shakemap in Italy: progress and advances in the last 10 yr. Seismol Res Lett 2020; 91(1):317–33. https://doi.org/10.1785/0220190130.
- [6] Park J, Bazzurro P, Baker JW. Modeling spatial correlation of ground motion Intensity Measures for regional seismic hazard and portfolio loss estimation. In: Kanda, Takada Furuta, editors. Applications of statistics and probability in civil engineering. Taylor & Francis Group; 2007.

I. Iervolino

- [7] Albarello D, D'amico V. Testing probabilistic seismic hazard estimates by comparison with observations: an example in Italy. Geophys J Int 2008;175(3): 1088–94. https://doi.org/10.1111/J.1365-246X.2008.03928.X/3/175-3-1088-FIG002.JPEG.
- [8] Stucchi M, Meletti C, Montaldo V, Crowley H, Calvi GM, Boschi E. Seismic hazard assessment (2003-2009) for the Italian building code. Bull Seismol Soc Am 2011; 101(4):1885–911. https://doi.org/10.1785/0120100130.
- [9] EN 1998-1 Eurocode 8. Design of structures for earthquake resistance Part 1: general rules, seismic actions and rules for buildings. European Committee for Standardization. 2004. Brussels, Belgium.
- [10] Cito P, Iervolino I. Rarity, proximity, and design actions: mapping strong earthquakes in Italy. Ann Geophys 2020;63(6):SE671. https://doi.org/10.4401/ag-8516.
- [11] Ambraseys NN, Simpson KA, Bommer JJ. Prediction of horizontal response spectra in Europe. Earthq Eng Struct Dynam 1996;25(4):371–400. https://doi.org/ 10.1002/(SICI)1096-9845(199604)25:4<371::AID-EQE550>3.0.CO;2-A.
- [12] Bindi D, Pacor F, Luzi L, Puglia R, Massa M, Ameri G, et al. Ground motion prediction equations derived from the Italian strong motion database. Bull Earthq Eng 2011;9(6):1899–920. https://doi.org/10.1007/s10518-011-9313-z.
- [13] Esposito S, Iervolino I. PGA and PGV spatial correlation models based on European multievent datasets. Bull Seismol Soc Am 2011;101(5):2532–41.
- [14] Chioccarelli E, Cito P, Iervolino I, Giorgio M. REASSESS V2.0: software for singleand multi-site probabilistic seismic hazard analysis. Bull Earthq Eng 2019;17(4): 1769–93. https://doi.org/10.1007/s10518-018-00531-x.
- [15] Giorgio M, Iervolino I. On multisite probabilistic seismic hazard analysis. Bull Seismol Soc Am 2016;106(3):1223–34. https://doi.org/10.1785/0120150369.