Bulletin of the Seismological Society of America, Vol. 106, No. 3, pp. 1223–1234, June 2016, doi: 10.1785/0120150369

On Multisite Probabilistic Seismic Hazard Analysis

by Massimiliano Giorgio and Iunio Iervolino

Abstract Seismic hazard assessment, in its classical format, models the stochastic process of occurrence of earthquakes causing the exceedance of ground-motion intensity measure (IM) thresholds at a specific site. This is because civil structures typically require probabilistic seismic hazard analysis (PSHA) for one location. On the other hand, there are cases in which it may be required to count the number of exceedances of a vector of IM thresholds at multiple sites over time. In these situations, in general, a form of stochastic dependence arises among the processes counting multiple earthquake exceedances of IM at the sites. The present study analyzes this dependence and shows how it is linked to the correlation among IMs at the sites in one earthquake. Indeed, it provides a formalization and derives closed-form solutions for multisite PSHA, showing that hazards at multiple sites are independent if and only if exceedances at the sites in one earthquake are mutually exclusive. The other key results of the work are: (1) probabilistically rigorous insights into the form of dependence among hazard at multiple sites are derived, and (2) it is shown that site-specific and multisite PSHA are unified, that is, how and why the latter is a special case of the former when only one location is considered. In addition, the applicative value of the formalization provided is illustrated by means of simple examples, including a loss assessment for a portfolio of structures and a hazard validation exercise.

Introduction

In those situations in which the analyst is interested in the exceedances of ground-motion intensity measure (IM) thresholds at more than one site, the key issue is to account for the existence of stochastic dependence among the site-specific processes, each counting the exceedances at a single site. This is the case, for example, of risk assessment for spatially distributed systems (e.g., Esposito *et al.*, 2015) that are possibly subjected to supply-chain or domino effects and for validation studies of hazard maps (e.g., Albarello and D'Amico, 2008). In the former case, to probabilistically quantify the loss, it is necessary to model the number of times the system's components that are deployed at different sites may fail at the same time; in the latter case, it is necessary to probabilistically characterize the exceedances at multiple sites consistent with the hazard map being validated in order to correctly compare with observations (e.g., Iervolino, 2013; Iervolino and Giorgio, 2015).

In fact, an approach similar to that of Weatherill *et al.* (2014) is required; such an approach has been studied by significant literature (e.g., McGuire, 1988; Eguchi, 1991) and is different with respect to probabilistic seismic hazard analysis (PSHA) (e.g., Cornell, 1968; McGuire, 2004). For example, according to classical PSHA, each process counting the occurrence over time of seismic events causing exceedance of an IM threshold at a single site is a homogeneous Poisson process (HPP). However, in general, the process describing the total number of exceedances collectively observed at the sites in a given time period is not a Poisson process.

On these premises and in the framework of the hypotheses consistent with those of site-specific PSHA, the study presented herein shows that it is possible to formally model the stochastic dependence between the processes regulating exceedances at different sites and then to formulate in closed-form a number of results, for example, the mean and the variance of the total number of exceedances in an arbitrary time interval or the joint distribution of the number of exceedances at two or more sites in any time period. It is also interesting to observe that, once the hazard at multiple sites is formalized, classical site-specific hazard results as the special case in which only one site is considered. In this sense, the study generalizes classical PSHA.

These results, as illustrated in the body of the article, are not obtained by formulating a new hazard theory, but by using the hypotheses of PSHA and established results in the field of stochastic point processes (e.g., Snyder, 1975), a class of random processes to which the HPP used to describe earthquake occurrence in PSHA belongs (Kingman, 1993; Resnick, 2002). In fact, the study analyzes the dependence among IMs at multiple sites in one earthquake and the well-known spatial correlation of IMs, which derive from the terms of ground-motion prediction equations (GMPEs). It is demonstrated that the correlation of IMs during one seismic event reflects in a nonobvious, yet quantifiable, way on the stochastic processes counting the exceedance of IM thresholds at the multiple sites over time (i.e., in multiple earthquakes).

To pursue the stated goals, the remainder of this article is structured such that the probabilistic basics of site-specific PSHA, and the issues arising when extending it to multiple sites, are reviewed first. Second, the sources of (spatial) dependence among exceedances at the multiple sites in one earthquake (e.g., correlation of IMs stemming from GMPEs) are analyzed. Third, the properties of the Poisson process are used to link dependence among IMs in one earthquake to dependence among processes counting exceedances in multiple earthquakes, and the conditions to obtain independent site-specific counting processes are identified. Hence, a strategy that allows modeling of the joint distribution of the number of exceedances at multiple sites and the distribution of the total number of exceedances at the sites, in any time interval, is presented. Finally, some illustrative applications are developed to show the practical value of the derived formulations for seismic risk analysis of distributed systems, as well as their implications for hazard validation studies.

Basics of Site-Specific Hazard and Probabilistic Issues in Its Extension to Multiple Sites

Site-specific PSHA, in its standard form, models the seismic hazard via a marked Poisson process, which is a process where a random variable (the mark) is associated with the occurrence of any event (e.g., Snyder, 1975) and carries information about the features of the event that occurred. In fact, it assumes that the number $N(t)$ of earthquakes occurring in the $(0, t]$ time interval at the site of interest (i.e., originating from a specific seismic source) follows an HPP with rate ν. Moreover, PSHA associates a mark to any arrival time of an earthquake (i.e., the multivariate random variable, $\mathbf{Z} = \{M, X, Y, \mathrm{IM}\}$), the components of which are the event magnitude M, a pair of coordinates identifying the source location $\{X, Y\}$, and the IM the earthquake produces at the site of interest. The coordinates of the source, together with the coordinates of the site for which the hazard is computed, determine the source-to-site distance; one of the covariates of GMPEs that are usually employed for hazard assessment, as described below.

The multivariate variables $\mathbf{Z}_i = \{M_i, X_i, Y_i, \mathrm{IM}_i\}$, $(i = 1, 2, \ldots)$ associated with different arrival times (i.e., to different earthquakes) are assumed to be identically distributed in PSHA, stochastically independent one of each other, and independent of $N(t)$ (Cornell, 1968). Therefore, it is also an HPP, the process that counts in $(0, t]$ the number of exceedances of a reference intensity value im* at the site, that is, $N_{\mathrm{IM}>\mathrm{im}^*}(t)$. It is well known that this Poisson process is characterized by a rate, which is a fraction of the rate of occurrence of earthquakes ν; that is, $\lambda_{\mathrm{IM}>\mathrm{im}^*} = \nu \cdot P(\mathrm{IM} > \mathrm{im}^*)$. In this relationship, $P(\mathrm{IM} > \mathrm{im}^*)$ is the probability of exceedance of the threshold in one earthquake.

Thus, classical PSHA consists of estimating the rate of exceedance $\lambda_{\mathrm{IM}>\mathrm{im}^*}$ (e.g., the mean number of exceedances per unit time) of an arbitrary IM value at a site. Given $\lambda_{\mathrm{IM}>\mathrm{im}^*}$, the probability that k exceedances occur in the time interval $(0, t]$ at the site $P[N_{\mathrm{IM}>\mathrm{im}^*}(t) = k]$ can be computed for any $t > 0$ and $k = 0, 1, 2, \ldots$ via the Poisson distribution in equation (1):

$$P[N_{\mathrm{IM}>\mathrm{im}^*}(t) = k] = \frac{(\lambda_{\mathrm{IM}>\mathrm{im}^*} \cdot t)^k}{k!} \cdot e^{-\lambda_{\mathrm{IM}>\mathrm{im}^*} \cdot t} \quad . \tag{1}$$

Computation of $\lambda_{\mathrm{IM}>\mathrm{im}^*}$ is often carried out as in equation (2), commonly referred to as the hazard integral, in which $P(\mathrm{IM} > \mathrm{im}^*)$ is obtained via the total probability theorem, which averages the conditional probability of exceedance of IM given the event's characteristics (i.e., M and $\{X, Y\}$) with respect to the joint distribution of M and $\{X, Y\}$. In the equation, $\Omega_{M,X,Y}$ is the domain of $\{M, X, Y\}$. (Usually, the source-to-site distance appears instead of the coordinates of the earthquake event, yet this representation is clearly equivalent.)

$$\begin{aligned}\lambda_{\mathrm{IM}>\mathrm{im}^*} &= \nu \cdot P(\mathrm{IM} > \mathrm{im}^*) \\ &= \nu \cdot \iiint_{M,X,Y \in \Omega_{M,X,Y}} P(\mathrm{IM} > \mathrm{im}^* | m, x, y) \\ &\quad \cdot f_{M,X,Y}(m, x, y) \cdot dm \cdot dx \cdot dy \end{aligned} \tag{2}$$

Issues in Hazard Assessment When Multiple Sites are Concerned

If one is interested in modeling the number of exceedances at multiple sites (s in number) located in an area affected by the same seismic source, then the situation is more complicated. Indeed, the process, say $N_{\mathrm{IM}_j>\mathrm{im}_j^*}(t)$, that counts the number of exceedances of im_j^* at the site j is an HPP with rate $\lambda_{\mathrm{IM}_j>\mathrm{im}_j^*}$, $\forall\ j = 1, 2, \ldots, s$. This can be obtained via equation (2) for each of the sites; however, two items are worth noting.

1. The processes $N_{\mathrm{IM}_1>\mathrm{im}_1^*}(t), N_{\mathrm{IM}_2>\mathrm{im}_2^*}(t), \ldots, N_{\mathrm{IM}_3>\mathrm{im}_3^*}(t)$ are, in general, stochastically dependent.
2. The total rate $\lambda_{\#Ex}$ (i.e., the mean number of exceedances collectively observed per unit time at the ensemble of the sites, $\#Ex$) can be calculated as the sum of the rates referring to the individual sites: $\lambda_{\#Ex} = \sum_{j=1}^{s} \lambda_{\mathrm{IM}_j>\mathrm{im}_j^*}$. Nevertheless, because of item (1), the process $N_{\#Ex}(t) = \sum_{j=1}^{s} N_{\mathrm{IM}_j>\mathrm{im}_j^*}(t)$, that counts the total number of exceedances over time, is not a Poisson process.

Therefore, an equation like (1) cannot be used to compute the probability of observing a certain number of exceedances collectively at the sites within a particular time interval. This is why, for risk assessment of distributed systems, it is not straightforward to separate multisite hazard from the behavior of the components and, thereby, why Monte Carlo simulations of ground motions and component failure are typically employed together; whereas for individual structures, site-specific hazard is evaluated separately from fragility and then integrated to obtain the risk (e.g., Cornell and Krawinkler, 2000).

Similarly, it is not easy to formulate the joint distribution of the number of exceedances at multiple sites—for example, to compute the probability of observing, within a particular time interval, a certain number of exceedances at some sites and not at the others. In fact, as mentioned, the formalization of both these issues can be considered as unaddressed in the seismic field. On the other hand, a formal probabilistic framework to count exceedances at multiple sites (i.e., multisite PSHA), and that degenerates in classical site-specific PSHA when only one site is considered, can be developed and is the ultimate aim of this study.

To derive such a framework, it is helpful to distinguish between (1) the dependence among IMs at different sites affected by one earthquake (i.e., spatial correlation of IMs) and (2) how the dependence in one earthquake probabilistically reflects on the joint distribution of the number of exceedances at the sites in multiple earthquakes (i.e., the dependence among the processes counting exceedances over time at each of the sites, that is, the $N_{\mathrm{IM}_j > \mathrm{im}_j^*}(t)$, $j = 1, 2, \ldots, s$). Point (1) is discussed in the next section, followed by a section addressing some properties of the Poisson process (point 2), which are relevant for the following developments.

Dependence of IMs at Multiple Sites in One Event: Mean and Residuals of GMPEs

The nature and form of stochastic dependence existing among the processes counting exceedances over time of ground-motion thresholds at multiple sites is related to the probabilistic characterization of the effects of a common earthquake at the different sites. The latter is an issue referred to often in recent literature as the spatial correlation of IMs (e.g., Jayaram and Baker, 2009; Esposito and Iervolino, 2011, 2012).

To provide insights into this correlation, it is worthwhile to recall that, for PSHA purposes, ground motion is typically modeled via a GMPE, which factually is the probabilistic distribution of the chosen IM conditional on earthquake magnitude, source-to-site distance, and other parameters such as local geological conditions. In other words, GMPEs provide the $P(\mathrm{IM} > \mathrm{im}^* | m, x, y)$ term of equation (2). In their classical format, GMPEs are obtained by regression and model the logs of IM, at a site j due to earthquake i, as in equation (3):

$$\log \mathrm{IM}_{j,i} = E(\log \mathrm{IM} | m_i, r_{j,i}, \theta) + \eta_i + \varepsilon_{j,i}, \qquad (3)$$

in which $E(\log \mathrm{IM} | m_i, r_{j,i}, \theta)$ is the mean of $\log \mathrm{IM}_{j,i}$ conditional on parameters such as magnitude M source-to-site distance R, and others (θ); η_i denotes the interevent residual, which is a constant term for all sites in a given earthquake and quantifies how much the mean of $\log \mathrm{IM}_{j,i}$ in the ith earthquake differs from $E(\log \mathrm{IM} | m_i, r_{j,i}, \theta)$; and $\varepsilon_{j,i}$ models the effect of intraevent variability of ground motion at site j in earthquake i. Interevent and intraevent residuals are usually assumed to be stochastically independent random variables (RVs), normally distributed with zero mean and variance $\sigma^2_{\mathrm{inter}}$ and $\sigma^2_{\mathrm{intra}}$, respectively. Then, for site-specific hazard, IM is modeled as a lognormal random variable in which the logarithm has variance $\sigma_T^2 = \sigma^2_{\mathrm{inter}} + \sigma^2_{\mathrm{intra}}$.

The spatial correlation of ground motion formally depends on the GMPE. In fact, when dealing with multiple sites, given magnitude and location of the earthquake (i.e., given the occurrence of one event with those features), it is generally assumed that the logs of IMs at the sites form a Gaussian random field (GRF), in which the components of the mean vector are given by the $E(\log \mathrm{IM} | m_i, r_{j,i}, \theta)$ terms (one for each of the sites j, $j = 1, 2, \ldots, s$). In addition, the covariance matrix Σ of the GRF is taken as in equation (4), in which $\rho_{j,h}$ is the correlation coefficient between intraevent residuals at two sites $\{j, h\}$ in the region (e.g., Park *et al.*, 2007; Malhotra, 2008). Assigning the mean vector and the covariance matrix completely defines the GRF.

$$\Sigma = \sigma^2_{\mathrm{inter}} \cdot \begin{bmatrix} 1 & 1 & \cdots & 1 \\ 1 & 1 & \cdots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \cdots & 1 \end{bmatrix} + \sigma^2_{\mathrm{intra}} \cdot \begin{bmatrix} 1 & \rho_{1,2} & \cdots & \rho_{1,s} \\ \rho_{2,1} & 1 & \cdots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{s,1} & \rho_{s,2} & \cdots & 1 \end{bmatrix} \qquad (4)$$

Regarding correlation, it is important to note the following:

1. Correlation among IMs is generated by the fact that all components of the mean vector share the same event features, such as magnitude and location. (In particular, the magnitude determines positive correlation because, if the event is strong for one site, it is strong for all the sites; whereas, given M, the distance can determine positive or negative correlation between IMs at the sites, depending on the relative positions of the sites with respect to the source.)
2. Correlation among IMs is also generated because the first term on the right side of equation (4) produces interevent residuals that are perfectly and positively correlated at all sites in one event, and this is another source of stochastic dependence among IMs in one earthquake (given i, this contribution is the same for each j).
3. The last source of correlation among IMs in a single earthquake is the second term on the right side of equation (4); the (symmetrical) matrix produces nonperfectly (positively) correlated intraevent residuals.

In other words, the primary source of spatial dependence among IMs at different sites stems from the fact that all sites share the same rupture's features, including the interevent residual. Another source of stochastic dependence among ground motions, generated at different sites by the same event, is then represented by intraevent residuals (usually, correlation among these residuals is assumed to decrease as the separation distance between two sites increases). In summary, the correlation of interevent and intraevent residuals is responsible only for a part of the correlation of IMs at multiple sites, and both can only determine positive correlation.

Before discussing how these issues translate in the stochastic dependence observed among processes counting exceedances caused at different sites by multiple earthquakes (i.e., over time), it is worthwhile to mention that the same reasoning can be applied to more recent forms of GMPEs, which further decompose the residuals, including the so-called single-site sigma (e.g., Rodriguez-Marek *et al.*, 2011). Therefore, the results presented in the following are not affected by such a representation of the GMPE.

Dependence among Processes Counting Exceedances at Different Sites in Multiple Earthquakes: The Properties of Poisson Processes

Let us consider a marked Poisson process; in particular, let $N(t)$ denote a homogeneous Poisson process, and let Z_i, denote the RV associated with its ith arrival time. Thus, we postulate that:

1. the random variables $\{Z_1, Z_2, ...\}$, associated with different arrival times, are stochastically independent (of one another) and identically distributed (i.i.d.); and
2. the random variables $\{Z_1, Z_2, ...\}$ and $N(t)$ are stochastically independent.

Consider now r mutually exclusive subsets $\{A_1, A_2, ..., A_r\}$ of the domain of Z_i; that is, $P[(Z_i \in A_k) \cap (Z_i \in A_m)] = \emptyset$, $\forall\ i, \forall\ k \neq m$. In this case, the consequences to points (1) and (2) are:

3. the process $N_{A_k}(t)$, which counts the arrival times for which $Z_i \in A_k$ for any k, is still an HPP, with mean function $E[N_{A_k}(t)] = E[N(t)] \cdot P(Z \in A_k)$; and
4. the processes $\{N_{A_1}(t), N_{A_2}(t), \ldots, N_{A_r}(t)\}$ are stochastically independent (Kingman, 1993; Resnick, 2002).

The operation that produces the processes in point (3) is called filtering or thinning, and the obtained processes are called filtered or thinned processes.

It is worth remarking that, subsequent to points (3–4), the mutual exclusivity of subsets $\{A_1, A_2, \ldots, A_r\}$ is a necessary condition for stochastic independence of the filtered processes. Thus, considering mutual exclusivity implies that the events $(Z_i \in A_1), (Z_i \in A_2), \ldots, (Z_i \in A_r)$ cannot occur simultaneously, the result is that stochastic independence among the filtered processes is not obtained in the case in which the counted events, given i, are stochastically independent. This can appear somewhat counterintuitive, yet the reader should focus on it, because it will be useful later on.

Formalizing Multisite Probabilistic Seismic Hazard

This section directly follows from the previous one, translating it into the seismic hazard context. In agreement with classical PSHA, let $N(t)$ denote the HPP that counts the earthquakes occurring on a specified seismic source. Let ν be the rate of such a process. At this point, note that points (1) and (2) of the previous section are exactly the hypotheses of site-specific hazard discussed when summarizing the basics of PSHA. In particular, the magnitude, the earthquake location, and the IM at the site for which the hazard is evaluated are independent and identically distributed in multiple earthquakes; moreover, they are independent of the process that counts earthquakes $N(t)$ (see Cornell, 1968). The objective is now to derive insights into the process counting exceedances, when extending these hypotheses to multiple sites and taking advantage of the results (3) and (4) of the previous section.

Consider s sites of interest to which are associated s thresholds $\{\mathrm{im}_1^*, \mathrm{im}_2^*, \ldots, \mathrm{im}_s^*\}$ in terms of the IM of interest (i.e., an arbitrary threshold for each site). Also assume that each mark associated with the arrival times of the counting process $N(t)$ has the following components: the magnitude M of the earthquake, the earthquake location $\{X, Y\}$, and the vector $\mathbf{IM} = \{\mathrm{IM}_1, \mathrm{IM}_2, ..., \mathrm{IM}_s\}$ that specifies the effect one arrival (i.e., one earthquake) produces at the sites. Then $\mathbf{Z}_i$, associated with the ith arrival time, is now the multivariate RV $\mathbf{Z}_i = \{M_i, X_i, Y_i, \mathrm{IM}_{i,1}, \mathrm{IM}_{i,2}, ..., \mathrm{IM}_{i,s}\}$, and $\Omega_{\mathbf{Z}}$ may indicate its domain. (In this context, it may be useful also to denote the domain of the variables as Ω with a subscripted list of the variables; for example, $\Omega_{M,X,Y}$ is the space in which the variable $\{M, X, Y\}$ takes values.)

As per site-specific PSHA, assume that the multivariate variables $\{\mathbf{Z}_1, \mathbf{Z}_2, \ldots\}$, associated with different arrival times of the process $N(t)$, are stochastically independent of each other and identically distributed (i.e., what is observed in one earthquakes does not affect the probabilities related to the next one) and are stochastically independent of $N(t)$. It is useful to remark that this assumption does not prevent the components of $\mathbf{Z}_i$ from being dependent in the same earthquake (e.g., because of spatial correlation among IMs in one event).

On these premises, site-specific hazard in the following description is first contextualized in the light of the properties of the Poisson process, then the dependence among hazards for different sites in multiple earthquakes is addressed, and, finally, the results for multisite hazard are derived.

Note that from now on the variable associated with the generic arrival time will be denoted as $\mathbf{Z}$, without any subscript due to the independent and identically distributed assumption (1), which allows simplifying the notation in the rest of the article, when possible.

Site-Specific Hazard: The Process Counting IM Exceedance at Each Considered Site

The marginal stochastic process counting exceedances at the generic jth site (i.e., site-specific hazard) can be easily formulated by partitioning the space $\Omega_{\mathbf{Z}}$ in the two (disjoint) subsets: (1) $A_{\mathrm{IM}_j > \mathrm{im}_j^*}$, which contains all the values of $\mathbf{Z}_i = \{M_i, X_i, Y_i, \mathrm{IM}_{i,1}, \mathrm{IM}_{i,2}, \ldots, \mathrm{IM}_{i,s}\}$ (earthquake features and the effects at the sites) resulting in $\mathrm{IM}_j > \mathrm{im}_j^*$, and (2) $A_{M_j \leq \mathrm{im}_j^*}$, which contains all the other values that $\mathbf{Z}_i$ can assume (i.e., nonexceedance of the threshold). Because of the recalled properties of the Poisson process, the two filtered processes associated with these disjointed subsets are stochastically independent processes. In particular, the process associated with subset $A_{\mathrm{IM}_j > \mathrm{im}_j^*}$ is the process that counts the number of exceedances at the jth site, $N_{\mathrm{IM}_j > \mathrm{im}_j^*}(t)$. Conversely, the process associated with subset $A_{\mathrm{IM}_j \le \mathrm{im}_j^*}$, namely $N_{\mathrm{IM}_j \le \mathrm{im}_j^*}(t)$, is the process counting all the events that do not cause exceedance at site j. Both these processes are homogeneous Poisson, and their rates are given in equation (5):

$$\begin{cases} \lambda_{\mathrm{IM}_j > \mathrm{im}_j^*} = \nu \cdot P(\mathbf{Z} \in A_{\mathrm{IM}_j > \mathrm{im}_j^*}) = \nu \cdot P(\mathrm{IM}_j > \mathrm{im}_j^*) \\ \lambda_{\mathrm{IM}_j \le \mathrm{im}_j^*} = \nu \cdot P(\mathbf{Z} \in A_{\mathrm{IM}_j \le \mathrm{im}_j^*}) = \nu \cdot P(\mathrm{IM}_j \le \mathrm{im}_j^*) \end{cases} . \quad (5)$$

Being stochastically independent, the sum of these Poisson processes is still a Poisson process, which (obviously) coincides with $N(t)$; that is, the process that counts the earthquakes occurring on the considered seismic source.

Clearly, $\lambda_{\mathrm{IM}_j > \mathrm{im}_j^*}$ in equation (5) coincides with that in equation (2) from PSHA. Indeed, it is possible to write equation (6) to compute it. In the equation, $I_{(\mathrm{IM}_j > \mathrm{im}_j^*)}$ is a function that equals one in the case of threshold exceedance at site j, and zero otherwise, and $f_{M,X,Y,\mathrm{IM}_1,\ldots,\mathrm{IM}_s}(m, x, y, \mathrm{im}_1, \ldots, \mathrm{im}_s)$ is the distribution of $\mathbf{Z}_i = \{M_i, X_i, Y_i, \mathrm{IM}_{i,1}, \mathrm{IM}_{i,2}, \ldots, \mathrm{IM}_{i,s}\}$.

$$\begin{aligned} \lambda_{\mathrm{IM}_j > \mathrm{im}_j^*} &= \nu \cdot P(\mathbf{Z} \in A_{\mathrm{IM}_j > \mathrm{im}_j^*}) = \nu \cdot P(\mathrm{IM}_j > \mathrm{im}_j^*) \\ &= \nu \cdot \iint_{\ldots} \int_{\mathbf{Z} \in \Omega_{\mathbf{Z}}} I_{(\mathrm{IM}_j > \mathrm{im}_j^*)} \\ &\quad \cdot f_{M,X,Y,\mathrm{IM}_1,\ldots,\mathrm{IM}_s}(m, x, y, \mathrm{im}_1, \ldots, \mathrm{im}_s) \\ &\quad \cdot dm \cdot dx \cdot dy \cdot d(\mathrm{im}_1) \cdot \ldots \cdot d(\mathrm{im}_s) \\ &= \nu \cdot \iiint_{M,X,Y \in \Omega_{M,X,Y}} P(\mathrm{IM}_j > \mathrm{im}_j^* | m, x, y) \\ &\quad \cdot f_{M,X,Y}(m, x, y) \cdot dm \cdot dx \cdot dy \end{aligned} \quad (6)$$

imply (1) $(\mathrm{IM}_j > \mathrm{im}_j^*) \cap (\mathrm{IM}_h \le \mathrm{im}_h^*)$, (2) $(\mathrm{IM}_j \le \mathrm{im}_j^*) \cap (\mathrm{IM}_h > \mathrm{im}_h^*)$, and (3) $(\mathrm{IM}_j > \mathrm{im}_j^*) \cap (\mathrm{IM}_h > \mathrm{im}_h^*)$; that is, three out of the four possible combinations of joint exceedance and/or nonexceedance at the two sites in one event.

Hence, because these subsets are disjointed, the filtered processes $N_{(\mathrm{IM}_j > \mathrm{im}_j^*) \cap (\mathrm{IM}_h \le \mathrm{im}_h^*)}(t)$, $N_{(\mathrm{IM}_j \le \mathrm{im}_j^*) \cap (\mathrm{IM}_h > \mathrm{im}_h^*)}(t)$, and $N_{(\mathrm{IM}_j > \mathrm{im}_j^*) \cap (\mathrm{IM}_h > \mathrm{im}_h^*)}(t)$ are stochastically independent HPPs, whose rates are reported in equation (7), which were obtained by filtering the process describing the occurrence of earthquakes.

$$\begin{cases} \lambda_{(\mathrm{IM}_j > \mathrm{im}_j^*) \cap (\mathrm{IM}_h \le \mathrm{im}_h^*)} = \nu \cdot P[\mathbf{Z} \in A_{(\mathrm{IM}_j > \mathrm{im}_j^*) \cap (\mathrm{IM}_h \le \mathrm{im}_h^*)}] \\ \lambda_{(\mathrm{IM}_j \le \mathrm{im}_j^*) \cap (\mathrm{IM}_h > \mathrm{im}_h^*)} = \nu \cdot P[\mathbf{Z} \in A_{(\mathrm{IM}_j \le \mathrm{im}_j^*) \cap (\mathrm{IM}_h > \mathrm{im}_h^*)}] \\ \lambda_{(\mathrm{IM}_j > \mathrm{im}_j^*) \cap (\mathrm{IM}_h > \mathrm{im}_h^*)} = \nu \cdot P[\mathbf{Z} \in A_{(\mathrm{IM}_j > \mathrm{im}_j^*) \cap (\mathrm{IM}_h > \mathrm{im}_h^*)}] \end{cases} \quad (7)$$

These are the rates of earthquakes causing the specific joint exceedance or nonexceedance at the two sites corresponding to the subsets (1), (2), and (3).

Note that exceedance of the IM threshold in one earthquake at site j can be seen as $\mathbf{Z}$ belonging to the union of subsets (1) and (3); that is, $A_{\mathrm{IM}_j > \mathrm{im}_j^*} \equiv A_{(\mathrm{IM}_j > \mathrm{im}_j^*) \cap (\mathrm{IM}_h \le \mathrm{im}_h^*)} \cup A_{(\mathrm{IM}_j > \mathrm{im}_j^*) \cap (\mathrm{IM}_h > \mathrm{im}_h^*)}$. For the same reason, exceedance of the IM threshold in one earthquake at site h can be seen as $\mathbf{Z}$ belonging to the union of subsets (2) and (3): $A_{\mathrm{IM}_h > \mathrm{im}_h^*} \equiv A_{(\mathrm{IM}_j \le \mathrm{im}_j^*) \cap (\mathrm{IM}_h > \mathrm{im}_h^*)} \cup A_{(\mathrm{IM}_j > \mathrm{im}_j^*) \cap (\mathrm{IM}_h > \mathrm{im}_h^*)}$. As a consequence, it is possible to write equation (8), which shows that (in general) the processes $N_{\mathrm{IM}_j > \mathrm{im}_j^*}(t)$ and $N_{\mathrm{IM}_h > \mathrm{im}_h^*}(t)$, from site-specific PSHA, are not stochastically independent. This is because both share one component, $N_{(\mathrm{IM}_j > \mathrm{im}_j^*) \cap (\mathrm{IM}_h > \mathrm{im}_h^*)}(t)$. This ultimately proves stochastic dependence in time of hazard at different sites.

$$\begin{cases} N_{\mathrm{IM}_j > \mathrm{im}_j^*}(t) = N_{(\mathrm{IM}_j > \mathrm{im}_j^*) \cap (\mathrm{IM}_h \le \mathrm{im}_h^*)}(t) + N_{(\mathrm{IM}_j > \mathrm{im}_j^*) \cap (\mathrm{IM}_h > \mathrm{im}_h^*)}(t) \\ N_{\mathrm{IM}_h > \mathrm{im}_h^*}(t) = N_{(\mathrm{IM}_j \le \mathrm{im}_j^*) \cap (IM_h > im_h^*)}(t) + N_{(\mathrm{IM}_j > \mathrm{im}_j^*) \cap (\mathrm{IM}_h > \mathrm{im}_h^*)}(t) \end{cases} \quad (8)$$

Using a variable considering multiple sites, $\mathbf{Z}$, to derive the site-specific hazard integral, ultimately shows that in this framework the site-specific PSHA is a particular case of the multisite hazard.

Dependence among the Processes Counting Exceedances over Time at Multiple Sites

If it is possible that one earthquake causes exceedance at two sites j and h (i.e., exceedances are not mutually exclusive), then the processes $N_{\mathrm{IM}_j > \mathrm{im}_j^*}(t)$ and $N_{\mathrm{IM}_h > \mathrm{im}_h^*}(t)$, $j \ne h$, are stochastically dependent. To understand the nature of this form of dependence, let us consider the following three disjoint subsets: (1) $A_{(\mathrm{IM}_j > \mathrm{im}_j^*) \cap (\mathrm{IM}_h \le \mathrm{im}_h^*)}$, (2) $A_{(\mathrm{IM}_j \le \mathrm{im}_j^*) \cap (\mathrm{IM}_h > \mathrm{im}_h^*)}$, and (3) $A_{(\mathrm{IM}_j > \mathrm{im}_j^*) \cap (\mathrm{IM}_h > \mathrm{im}_h^*)}$. These contain all the values of the earthquake features that

On the other hand, note that, in the case in which it is certain that a single earthquake cannot cause joint exceedance at sites j and h (i.e., in the case $A_{\mathrm{IM}_j > \mathrm{im}_j^*}$ and $A_{\mathrm{IM}_h > \mathrm{im}_h^*}$ are disjoint), $N_{\mathrm{IM}_j > \mathrm{im}_j^*}(t)$ and $N_{\mathrm{IM}_h > \mathrm{im}_h^*}(t)$ are stochastically independent. In fact, in this case, the joint exceedance cannot occur (i.e., $P[\mathbf{Z} \in A_{(\mathrm{IM}_j > \mathrm{im}_j^*) \cap (\mathrm{IM}_h > \mathrm{im}_h^*)}] = 0$), and equation (8) can be rewritten in the simplified form of equation (9), in which the counting processes for the two sites do not share any term, because $N_{(\mathrm{IM}_j > \mathrm{im}_j^*) \cap (\mathrm{IM}_h > \mathrm{im}_h^*)}(t)$ is certainly zero.

$$\begin{cases} N_{\mathrm{IM}_j > \mathrm{im}_j^*}(t) = N_{(\mathrm{IM}_j > \mathrm{im}_j^*) \cap (\mathrm{IM}_h \le \mathrm{im}_h^*)}(t) \\ N_{\mathrm{IM}_h > \mathrm{im}_h^*}(t) = N_{(\mathrm{IM}_j \le \mathrm{im}_j^*) \cap (\mathrm{IM}_h > \mathrm{im}_h^*)}(t) \end{cases} \quad (9)$$

Then, it can be argued that the processes $N_{\mathrm{IM}_j > \mathrm{im}_j^*}(t)$ and $N_{\mathrm{IM}_h > \mathrm{im}_h^*}(t)$ are stochastically independent if the integral in equation (10) is zero. In the integral, $I_{[(\mathrm{IM}_j > \mathrm{im}_j^*) \cap (\mathrm{IM}_h > \mathrm{im}_h^*)]}$

equals one when the subscripted event occurs and zero otherwise. (This formulation was introduced in Esposito and Iervolino, 2011, and named the regional hazard integral.)

$$P[(\mathrm{IM}_j > \mathrm{im}_j^*) \cap (\mathrm{IM}_h > \mathrm{im}_h^*)]$$
$$= \iint \cdots \int_{\mathbf{Z} \in \Omega_{\mathbf{Z}}} I_{[(\mathrm{IM}_j > \mathrm{im}_j^*) \cap (\mathrm{IM}_h > \mathrm{im}_h^*)]} \cdot f_{M,X,Y,\mathrm{IM}_1,\ldots,\mathrm{IM}_s}(m, x, y, \mathrm{im}_1, \ldots, \mathrm{im}_s) \cdot dm \cdot dx \cdot dy \cdot d(\mathrm{im}_1) \cdot \ldots \cdot d(\mathrm{im}_s)$$
$$= \iiint_{M,X,Y \in \Omega_{M,X,Y}} P[(\mathrm{IM}_j > \mathrm{im}_j^*) \cap (\mathrm{IM}_h > \mathrm{im}_h^*)|m, x, y] \cdot f_{M,X,Y}(m, x, y) \cdot dm \cdot dx \cdot dy \tag{10}$$

The condition just expressed has an interesting seismic interpretation: the site-specific HPPs, counting exceedances over time of the intensity thresholds at two or more sites, are stochastically independent if, and only if, the probability of exceedance at multiple sites in one earthquake is zero. In other words, the numbers of exceedances observed in the same time interval at different sites are stochastically independent if and only if any earthquake can cause exceedance at one site at the most. It is worth noting that, referring to the Dependence of IMs at Multiple Sites in One Event: Mean and Residuals of GMPEs section, the only way to satisfy this condition occurs when the distance among the considered sites is large enough to assure that exceedance at a site implies nonexceedance at all others. In particular, it is also worth noting that, for the counting processes being independent, it is not sufficient that residuals of the GMPE are uncorrelated.

Joint Distribution of the Numbers of Exceedances at Multiple Sites in Any Time Interval

Because of the stochastic dependence just discussed, the joint distribution of the number of exceedances at multiple sites in the $(0, t]$ time interval is not easy to calculate. For example, considering two different sites, say j and h, in general the joint distribution cannot be obtained as the product of the (Poisson) probabilities computed using the marginal processes $N_{\mathrm{IM}_j > \mathrm{im}_j^*}(t)$ and $N_{\mathrm{IM}_h > \mathrm{im}_h^*}(t)$ from the site-specific hazard. In fact, to calculate the probability that, in $(0, t]$, exactly n_j and exactly n_h exceedances are observed at sites j and h $(j \neq h, j, h = 1, 2, \ldots, s)$, respectively, equation (11) must be adopted.

$$P\{[N_{\mathrm{IM}_j > \mathrm{im}_j^*}(t) = n_j] \cap [N_{\mathrm{IM}_h > \mathrm{im}_h^*}(t) = n_h]\}$$
$$= P\Bigg\{\bigcup_{n_{j,h}=0}^{\min(n_j,n_h)} [(N_{(\mathrm{IM}_j > \mathrm{im}_j^*) \cap (\mathrm{IM}_h \le \mathrm{im}_h^*)}(t) = n_j - n_{j,h}) \cap (N_{(\mathrm{IM}_j \le \mathrm{im}_j^*) \cap (\mathrm{IM}_h > \mathrm{im}_h^*)}(t) = n_h - n_{j,h}) \cap (N_{(\mathrm{IM}_j > \mathrm{im}_j^*) \cap (\mathrm{IM}_h > \mathrm{im}_h^*)}(t) = n_{j,h})]\Bigg\}$$
$$= \sum_{n_{j,h}=0}^{\min(n_j,n_h)} P[N_{(\mathrm{IM}_j > \mathrm{im}_j^*) \cap (\mathrm{IM}_h \le \mathrm{im}_h^*)}(t) = n_j - n_{j,h}] \cdot P[N_{(\mathrm{IM}_j \le \mathrm{im}_j^*) \cap (\mathrm{IM}_h > \mathrm{im}_h^*)}(t) = n_h - n_{j,h}] \cdot P[N_{(\mathrm{IM}_j > \mathrm{im}_j^*) \cap (\mathrm{IM}_h > \mathrm{im}_h^*)}(t) = n_{j,h}] =$$
$$= e^{-[\lambda_{(\mathrm{IM}_j > \mathrm{im}_j^*) \cap (\mathrm{IM}_h \le \mathrm{im}_h^*)} + \lambda_{(\mathrm{IM}_j \le \mathrm{im}_j^*) \cap (\mathrm{IM}_h > \mathrm{im}_h^*)} + \lambda_{(\mathrm{IM}_j > \mathrm{im}_j^*) \cap (\mathrm{IM}_h > \mathrm{im}_h^*)}] \cdot t} \cdot \sum_{n_{j,h}=0}^{\min(n_j,n_h)} \frac{[\lambda_{(\mathrm{IM}_j > \mathrm{im}_j^*) \cap (\mathrm{IM}_h \le \mathrm{im}_h^*)} \cdot t]^{n_j - n_{j,h}}}{(n_j - n_{j,h})!} \cdot \frac{[\lambda_{(\mathrm{IM}_j \le \mathrm{im}_j^*) \cap (\mathrm{IM}_j > \mathrm{im}_j^*)} \cdot t]^{n_h - n_{j,h}}}{(n_h - n_{j,h})!} \cdot \frac{[\lambda_{(\mathrm{IM}_j > \mathrm{im}_j^*) \cap (\mathrm{IM}_h > \mathrm{im}_h^*)} \cdot t]^{n_{j,h}}}{n_{j,h}!} \tag{11}$$

Even if the equation looks complicated, its derivation directly follows from the fact that the processes $N_{(\mathrm{IM}_j > \mathrm{im}_j^*) \cap (\mathrm{IM}_h \le \mathrm{im}_h^*)}(t)$, $N_{(\mathrm{IM}_j \le \mathrm{im}_j^*) \cap (\mathrm{IM}_h > \mathrm{im}_h^*)}(t)$, and $N_{(\mathrm{IM}_j > \mathrm{im}_j^*) \cap (\mathrm{IM}_h > \mathrm{im}_h^*)}(t)$ are stochastically independent HPPs, because the subsets they refer to (subsets 1–3 of the previous section) are mutually exclusive. In fact, their mean functions are those in equation (12):

$$\begin{cases} E[N_{(\mathrm{IM}_j > \mathrm{im}_j^*) \cap (\mathrm{IM}_h \le \mathrm{im}_h^*)}(t)] = \lambda_{(\mathrm{IM}_j > \mathrm{im}_j^*) \cap (\mathrm{IM}_h \le \mathrm{im}_h^*)} \cdot t = \nu \cdot t \cdot P[\mathbf{Z} \in A_{(\mathrm{IM}_j > \mathrm{im}_j^*) \cap (\mathrm{IM}_h \le \mathrm{im}_h^*)}] \\ E[N_{(\mathrm{IM}_j \le \mathrm{im}_j^*) \cap (\mathrm{IM}_h > \mathrm{im}_h^*)}(t)] = \lambda_{(\mathrm{IM}_j \le \mathrm{im}_j^*) \cap (\mathrm{IM}_h > \mathrm{im}_h^*)} \cdot t = \nu \cdot t \cdot P[\mathbf{Z} \in A_{(\mathrm{IM}_j \le \mathrm{im}_j^*) \cap (\mathrm{IM}_h > \mathrm{im}_h^*)}] \\ E[N_{(\mathrm{IM}_j > \mathrm{im}_j^*) \cap (\mathrm{IM}_h > \mathrm{im}_h^*)}(t)] = \lambda_{(\mathrm{IM}_j > \mathrm{im}_j^*) \cap (\mathrm{IM}_h > \mathrm{im}_h^*)} \cdot t = \nu \cdot t \cdot P[\mathbf{Z} \in A_{(\mathrm{IM}_j > \mathrm{im}_j^*) \cap (\mathrm{IM}_h > \mathrm{im}_h^*)}] \end{cases}. \tag{12}$$

Moreover, the probability of the union of events in equation (11) results in the summation at the last line of the same equation, because the events of the kind

$$[N_{(\mathrm{IM}_j > \mathrm{im}_j^*) \cap (\mathrm{IM}_h \le \mathrm{im}_h^*)}(t) = n_j - n_{j,h}] \cap [N_{(\mathrm{IM}_j \le \mathrm{im}_j^*) \cap (\mathrm{IM}_h > \mathrm{im}_h^*)}(t) = n_h - n_{j,h}] \cap [N_{(\mathrm{IM}_j > \mathrm{im}_j^*) \cap (\mathrm{IM}_h > \mathrm{im}_h^*)}(t) = n_{j,h}]$$

obtained using different values of $n_{j,h}$, are mutually exclusive.

In summary, equation (11) represents the bivariate (i.e., joint) probability mass function (PMF) of the RV, counting the numbers of exceedances in $(0, t]$ at two arbitrarily selected sites. Along the exact same line of reasoning and procedure, the PMF for the joint exceedance at a larger (arbitrary) number of sites may be computed, yet the resulting equation would be more elaborate.

These PMFs are evidently relevant, among other possible applications, for those studies devoted to constrain hazard at sites with poor earthquake observations, based on observations collected in a certain period of time at other sites. To complete the framework, in the following section the process counting the total number of exceedances at an arbitrary number of sites is discussed.

The Distribution of the Number of Exceedances Collectively Observed at Multiple Sites

This section targets the distribution (the PMF) of the random variable counting the total number of exceedances $\#Ex$ observed at the sites (s in number) in the $(0, t]$ interval. This result is obtained via the process $N_{\#Ex}(t)$, which rigorously counts the total number of exceedances. It is the sum of the processes counting the exceedances at the specific sites: $N_{\#Ex}(t) = \sum_{j=1}^{s} N_{\mathrm{IM}_j > \mathrm{im}_j^*}(t)$. However, as demonstrated, the terms of the sum are (in general) not independent, and then $N_{\#Ex}(t)$ is not a Poisson process.

At this point, consider $s+1$ disjoint subsets A_k, $k = 0, 1, 2, \ldots, s$ partitioning $\Omega_{\mathbf{Z}}$, the domain of the multivariate mark, which describes the characteristics of each possible earthquake and its possible effects at the sites. In particular, assume that A_k contains all the values of the features of the earthquake $\mathbf{Z}$ for which the exceedance of the thresholds is observed at exactly k sites in one earthquake. This is why the A_k subsets are $s+1$ in number: k has zero as the minimum, whereas s, corresponding to all sites experiencing exceedance in the same event, is the maximum. For example, the subset A_0 contains the values of $\mathbf{Z}$ associated with the earthquakes that do not cause exceedances at any of the sites. Similarly, A_3 is the subset that contains the values of $\mathbf{Z}$ associated with all the earthquakes causing exceedances at three sites. In particular, if $\mathbf{Z}_i \in A_3$, it is possible to say that the ith earthquake has caused exceedances at exactly three sites, but it is not possible to distinguish (based on this information alone) which are the sites where these exceedances occurred.

Once again, the filtered Poisson process associated with the generic A_k subset is still an HPP. In other words, the process $N_{A_k}(t)$ counting how many times exactly k exceedances are observed in $(0, t]$ among the sites is an HPP with rate determined as shown in equation (13):

$$\lambda_k(t) = \nu \cdot P(\mathbf{Z} \in A_k), \quad k = 0, 1, \ldots, s \quad . \tag{13}$$

In the equation, $P(\mathbf{Z} \in A_k)$ (i.e., the probability of observing exactly k exceedances, given the occurrence of one earthquake) can be calculated by solving the integral in equation (14), in which $I_{(Z \in A_k)}$ is 1 if $\mathbf{Z} \in A_k$ and 0 otherwise.

$$\begin{aligned} P(\mathbf{Z} \in A_k) = \iint \cdots \int_{\mathbf{Z} \in \Omega_{\mathbf{Z}}} & I_{(\mathbf{Z} \in A_k)} \\ & \cdot f_{M,X,Y,\mathrm{IM}_1,\ldots,\mathrm{IM}_s}(m, x, y, \mathrm{im}_1, \ldots, \mathrm{im}_s) \\ & \cdot dm \cdot dx \cdot dy \cdot d(\mathrm{im}_1) \cdot \ldots \cdot d(\mathrm{im}_s) \end{aligned} \tag{14}$$

Moreover, all these HPPs are stochastically independent, because the A_k subsets are mutually exclusive; that is, one earthquake cannot cause exactly k and exactly m exceedances ($k \neq m$). Because the $N_{A_k}(t)$ processes are independent HPPs, it is advantageous to express $N_{\#Ex}(t)$ as the linear combination of them. In other words, $N_{\#Ex}(t)$ can be seen as the linear combination of the processes ($s+1$ in number), each of which is an HPP counting the number of earthquakes causing exactly k exceedances collectively at the sites:

$$N_{\#Ex}(t) = \sum_{k=0}^{s} k \cdot N_{A_k}(t) \quad . \tag{15}$$

The mean and variance functions of such a process are

$$\begin{cases} E[N_{\#Ex}(t)] = \sum_{k=0}^{s} k \cdot E[N_{A_k}(t)] = v \cdot t \cdot \sum_{k=0}^{s} k \cdot P(\mathbf{Z} \in A_k) \\ \mathrm{Var}[N_{\#Ex}(t)] = \sum_{k=0}^{s} k^2 \cdot \mathrm{Var}[N_{A_k}(t)] = v \cdot t \cdot \sum_{k=0}^{s} k^2 \cdot P(\mathbf{Z} \in A_k) \end{cases} \quad . \tag{16}$$

These equations were derived immediately because they benefit from the independence of the combined $N_{A_k}(t)$ processes. This is why it was more convenient for the variance to express $N_{\#Ex}(t)$ as in equation (15) rather than as $N_{\#Ex}(t) = \sum_{j=1}^{s} N_{\mathrm{IM}_j > \mathrm{im}_j^*}(t)$. For the mean, equation (17) also could be used due to linearity of the expectation holding regardless of the dependence of the site-specific counting processes involved.

$$\begin{aligned} E[N_{\#Ex}(t)] &= E\left[\sum_{j=1}^{s} N_{\mathrm{IM}_j > \mathrm{im}_j^*}(t)\right] = \sum_{j=1}^{s} E[N_{\mathrm{IM}_j > \mathrm{im}_j^*}(t)] \\ &= v \cdot t \cdot \sum_{j=1}^{s} P(\mathbf{Z} \in A_{\mathrm{IM}_j > \mathrm{im}_j^*}) \\ &= v \cdot t \cdot \sum_{j=1}^{s} P(\mathrm{IM}_j > \mathrm{im}_j^*) \quad . \end{aligned} \tag{17}$$

At this point, equation (16) allows us to remark on three results relevant to multisite hazard analysis.

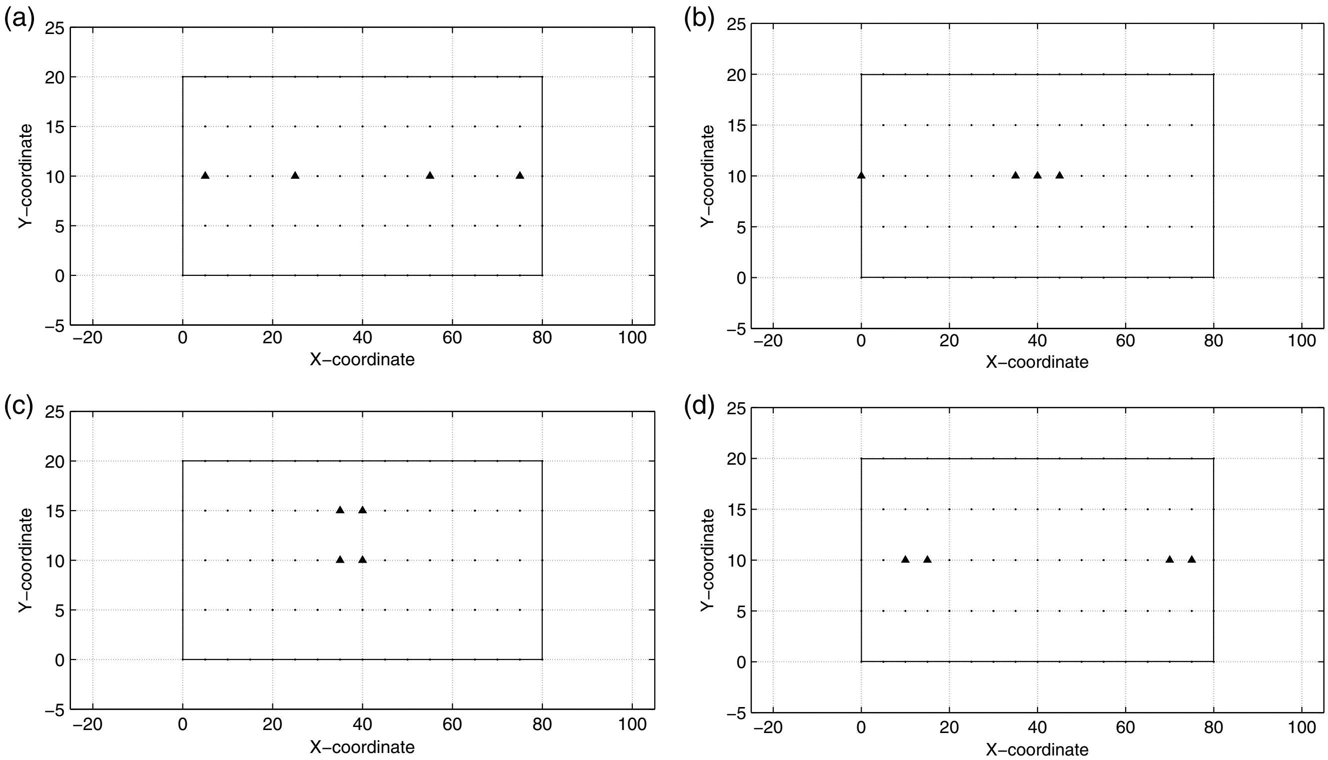


Figure 1. (a–d) Each panel reports a configuration for the four sites considered in the first application (triangles) and the ideal seismic source zone (distances in kilometers).

1. Because, in general, its mean and variance functions differ from each other, the process counting the total number of exceedances over time in the region cannot be a Poisson process.
2. If the probability that a single earthquake can cause exceedances in more than one site is zero [i.e., $P(\mathbf{Z} \in A_k) = 0$ for any $k \geq 2$], then the mean and variance coincide, and the $N_{\#Ex}(t)$ process is still an HPP. This is because, as discussed in previous sections, in this case the marginal processes $N_{\mathrm{IM}_1 > \mathrm{im}_1^*}(t), N_{\mathrm{IM}_2 > \mathrm{im}_2^*}(t), \ldots,$ and $N_{\mathrm{IM}_s > \mathrm{im}_s^*}(t)$ are stochastically independent.
3. Given the mean, the variance of the process counting the total number of exceedances over time is at its minimum if the marginal processes counting exceedances at the sites are independent.

It is also worthwhile to focus on the PMF of the random variable counting the total number of exceedances in $(0, t]$, which stems from the discussed process looking at a specific time interval. Because the counting process is not an HPP, the PMF is not a Poisson distribution. In fact, it can be formulated starting from

$$P[N_{\#Ex}(t) = x] = P\left[\sum_{k=1}^{s} k \cdot N_{A_k}(t) = x\right] = \sum_{n_1=0,\ldots,n_k=0}^{x} I_{\left(\sum_{k=1}^{s} k \cdot n_k = x\right)} \cdot \bigcap_{k=1}^{s} P[N_{A_k}(t) = n_k], \tag{18}$$

in which the terms

$$\bigcap_{k=1}^{s} P[N_{A_k}(t) = n_k]$$

are products of the probabilities of observing, in $(0, t]$, exactly n_k events, each causing exactly k exceedances collectively at the sites. The equation sums the products that it is possible to obtain by considering values of $(n_1, n_2, \ldots, n_s = 0, 1, 2, \ldots)$, which satisfy the condition

$$\sum_{k=1}^{s} k \cdot n_k = x \quad .$$

Therefore, $I_{\left(\sum_{k=1}^{s} k \cdot n_k = x\right)}$ is one in the case $\sum_{k=1}^{s} k \cdot n_k = x$ and zero otherwise. In other words, the equation considers all possible combinations of number and kind of earthquakes (in terms of the number of exceedances collectively caused at the sites) that can occur in $(0, t]$. For example, it considers that the following can occur in $(0, t]$: exactly two earthquakes that produce zero exceedances, exactly three earthquakes that produce one exceedance, exactly one earthquake that produces two exceedances, and so on. Starting from these combinations, the equation retains only the probabilities of those cases for which the number of occurring earthquakes, combined with the number of sites at which each of them causes exceedance, gives exactly x total exceedances in $(0, t]$.

Illustrative Applications

To prove the importance of the arguments discussed so far, it is worthwhile to illustrate the implications they can have in practical applications of current interest to earthquake engineering and engineering seismology. In particular, two applications are developed: one referring to the seismic loss assessment for an infrastructure made of a distributed portfolio of four buildings, and one referring to an exercise aimed at validating the probabilistic (site-specific) seismic hazard pooling observations of real earthquakes at several sites.

To this aim, an ideal $20 \times 80\ \mathrm{km}^2$ areal seismic source is considered. This area is discretized in point-like earthquake sources represented by the dots in the panels of Figure 1 (this number of point-like sources for the discretization of the source is for the first application, and will be changed in the second one). The event rate of the earthquakes is assumed to be $\nu = 1$ event/yr, globally over the source zone and it is uniformly partitioned among the point-like sources. The distribution of magnitude is a truncated exponential one, defined in the [4.5,7] range. The b-value of the Gutenberg–Richter relationship (Gutenberg and Richter, 1944) is equal to one. The considered IM is the peak ground acceleration (PGA), and the GMPE is that of Ambraseys *et al.* (1996).

A Loss Assessment Study for a Portfolio of Buildings

It is assumed that a company has four facilities located in the region; that is, $s = 4$ sites of interest. These facilities constitute a supply chain for the company. For the purposes of the application, four possible spatial arrangements of the facilities are considered. Each configuration of the four sites corresponds to a panel in Figure 1, in which the sites are indicated as triangles and the seismic source is within the rectangular frame.

The objective of the application is to compute the mean and variance of the seismic annual loss for each of the four configurations of the sites. To this aim, for each of the 85 (equally likely) point-like sources in the zone (Fig. 1), 10^5 Monte Carlo simulations of magnitude were performed. For the four sites in each of the four configurations, the simulations computed the PGAs that have a marginal probability of exceedance in one year equal to 0.0035. This probability was arbitrarily selected to determine the components of the threshold vector $\{\mathrm{im}_1^*, \mathrm{im}_2^*, \mathrm{im}_3^*, \mathrm{im}_4^*\}$. In fact, in the context of the application, these are considered as the accelerations that, if exceeded in one event at each of the sites, cause the expected losses described previously. (Of course, a more refined representation of the fragility of the facilities could be used instead, but this simplification does not affect the value of this illustrative application.)

Because PGAs at the four sites were also simulated to have realizations of the random field of IMs, they were also used to compute the probabilities $P(\mathbf{Z} \in A_k)$, that is, the probability that an earthquake causes exceedance of im_j^*, $j = 1,2,3,4$ exactly at k sites ($k = 0,1,2,3,4$). These probabilities, which can be formalized via an integral of the type in equation (19), were computed by means of the observed frequencies using the IMs from the Monte Carlo runs. In the equation, $I_{(\mathbf{Z} \in A_k)}$ is one in the case in which the subscripted event occurs and zero otherwise. Values obtained for the four configurations of the sites are reported in Table 1.

$$P(\mathbf{Z}\in A_k) = \iint \dots \int_{\mathbf{Z}\in\Omega_{\mathbf{Z}}} I_{(\mathbf{Z}\in A)} \cdot f_{M,X,Y,\mathrm{IM}_1,\dots,\mathrm{IM}_4}(m,x,y,\mathrm{im}_1,\dots,\mathrm{im}_4) \cdot dm \cdot dx \cdot dy \cdot d(\mathrm{im}_1) \cdot \dots \cdot d(\mathrm{im}_4) \approx$$
$$= \frac{\text{number of simulated earthquakes for which it resulted } \mathbf{Z}\in A_k}{\text{total number of simulated earthquakes}} \quad (19)$$

At this point in addressing the goals of the application, the mean and the variance of the annual expected loss can be retrieved from equation (16), replacing k with L_k. For example, for the configuration in Figure 1a, these results are

$$\begin{cases} E[L(1)] = \nu \cdot t \cdot \sum_{k=0}^{4} L_k \cdot P(\mathbf{Z} \in A_k) = 1 \cdot 1 \cdot (10 \cdot 0.01396 + 100 \cdot 0.00005) = 0.1446 \\ \mathrm{Var}[L(1)] = \nu \cdot t \cdot \sum_{k=0}^{4} L_k^2 \cdot P(\mathbf{Z} \in A_k) = 1 \cdot 1 \cdot (10^2 \cdot 0.01396 + 100^2 \cdot 0.00005) = 1.896 \end{cases} . \quad (20)$$

Given the occurrence of one earthquake and the possibility of contemporary failure of multiple facilities in the same seismic event, it is assumed that the losses in case of failure of k facilities in the portfolio (i.e., L_k, with values in monetary units) are as follows: $L_k = 0$ for $k = 0$, $L_k = 10$ for $k = 1$, $L_k = 100$ for $k = 2$, $L_k = 1000$ for $k = 3$, and $L_k = 10{,}000$ for $k = 4$. The loss is not additive in this case; that is, it is assumed that, because of the domino effect in the chain, the loss if k sites fail is larger than k times that referring to the failure of one facility only.

In fact, note that although these results were easy to obtain with the formalization of this study, they could not be computed via the marginal exceedance processes (i.e., site-specific hazards) at the sites (due to the discussed supply-chain effect and the presence of stochastic dependence). This is apparent from Table 1, in which the values of mean and variance of the annual loss for all the four configurations represented in Figure 1 are given. The relative differences among reported values also quantify how much the loss depends on the spatial configuration of the portfolio.

A Hazard Validation Study

This example is in the context of studies such as Albarello and D'Amico (2008), in which an interesting problem is discussed. Consider a region for which a hazard map is available from PSHA; that is, there is a map of IM values that have a given probability of exceedance in a certain time interval at each of the sites in the region. In particular, assume that the map reports the site-specific PGA with 10% exceedance probability in 30 years for each site.

The data of the problems are the following. Suppose we have gathered earthquake recordings from 68 seismic stations, deployed in the considered region and that have continuously operated for 30 years. Suppose also that, according to the earthquake data collected, in 13 cases, the PGA with 10% exceedance probability in 30 years was exceeded. The scope is to formally establish whether the observations are statistically consistent with the hazard map or the latter is underestimating the real seismic hazard in the region.

A viable approach to address this problem is that of hypothesis testing. In fact, consistent with PSHA, the probability of exceedance of the PGA threshold is computed via the Poisson distribution. This means that, according to equation (1), to obtain 0.1 as the probability of exceedance in 30 years, the rate must be $\lambda_{\mathrm{IM}>\mathrm{im}^*} = 0.0035$ for each of the sites, because these Poisson variables at all the sites are equally distributed. Obviously, the corresponding PGA thresholds (i.e., im*) are expected to be different among the sites.

Assuming that the processes counting exceedances observed at the 68 sites are independent, then the total number of exceedances is still a Poisson RV, being the sum of independent Poisson RVs. Consequently, the probability of observing k exceedances in 30 years over the 68 stations is given by

$$
\begin{aligned}
&P(k\,\text{exceedances of im}^*\,\text{at}\,68\,\text{ sites in}\,30\,\text{ yr}) \\
&= \frac{(\lambda_{\mathrm{IM}>\mathrm{im}^*}\cdot s\cdot t)^k}{k!}\cdot e^{-\lambda_{\mathrm{IM}>\mathrm{im}^*}\cdot s\cdot t} \\
&= \frac{(0.0035\cdot 68\cdot 30)^k}{k!}\cdot e^{-0.0035\cdot 68\cdot 30} \quad . \qquad (21)
\end{aligned}
$$

In other words, the random variable counting the total number of exceedances has the following mean and variance: $E[N_{\#Ex}(t)] = \mathrm{Var}[N_{\#Ex}(t)] = 0.0035\cdot 68\cdot 30 = 7.14$.

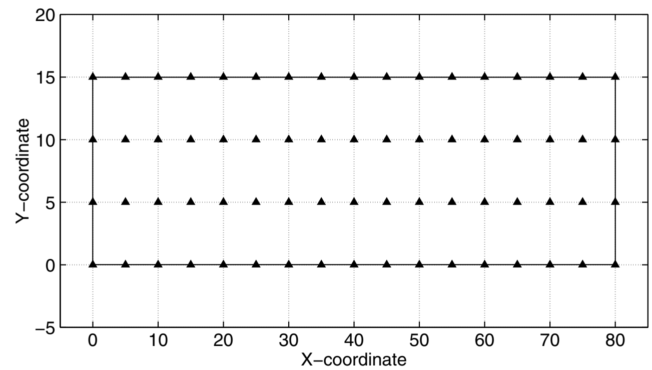


Figure 2. Ideal seismic source zone for the second application and recording sites (triangles). Each site is also a point-like seismic source (distances in kilometers).

At this point, the hazard may be tested against observations, assuming, for example, a significance level equal to 0.05. In particular, the (null) hypothesis is that the hazard map is correct (i.e., the mean number of exceedances in 30 years is 7.14), whereas the alternative hypothesis is that the number of observed exceedances is inconsistent with what is suggested by the map (i.e., the true mean is larger). The test can be carried out by computing the probability of observing at least 13 exceedances at 68 stations in 30 years via equation (21). This probability is 0.03, which is lower than 0.05; therefore, the analyst would conclude that the PSHA is biased with respect to the actual seismicity of the region, at the significance level 0.05.

On the other hand, it was demonstrated in the previous sections that the Poisson distributions counting exceedances of im* at the different sites, although equally distributed, are not independent if one earthquake can cause more than one exceedance in the region. Therefore, the use of the Poisson distribution in equation (21) may be erroneous. In fact, here it is shown how the presence of stochastic dependence among processes counting exceedances at different sites can change the conclusion of the test. To this aim, for each of the 68 (equally likely) point-like sources in the zone (Fig. 2), 10^4 Monte Carlo simulations of magnitude were performed. In each of these runs, the PGAs at the recording sites were also simulated. The obtained set of 68×10^4 observations was used to compute the PMF of the number of sites that may experience exceedance of the PGAs from the map in a single

Table 1

Probabilities of Observing from Zero to Four Exceedances of $\mathrm{im}_j^*, j = 1,2,3,4$ at the Four Sites, Given the Occurrence of One Earthquake for the Configurations of Figure 1

	$P(\mathbf{Z}\in A_k)$					Mean	Variance
Configuration in Figure 1	$k=0$	$k=1$	$k=2$	$k=3$	$k=4$	$E[L(1)]$	$\mathrm{Var}[L(1)]$
Figure 1a	0.98600	0.01396	0.00005	0.00000	0.00000	0.144	1.846
Figure 1b	0.98661	0.01277	0.00059	0.00004	0.00000	0.223	43.723
Figure 1c	0.98735	0.01143	0.00107	0.00015	0.00001	0.494	1439.452
Figure 1d	0.98652	0.01291	0.00057	0.00000	0.00000	0.186	7.104

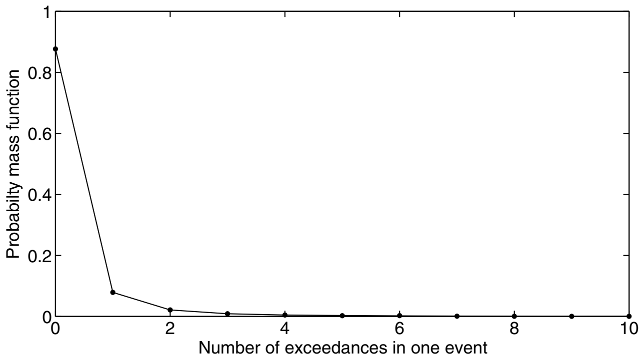
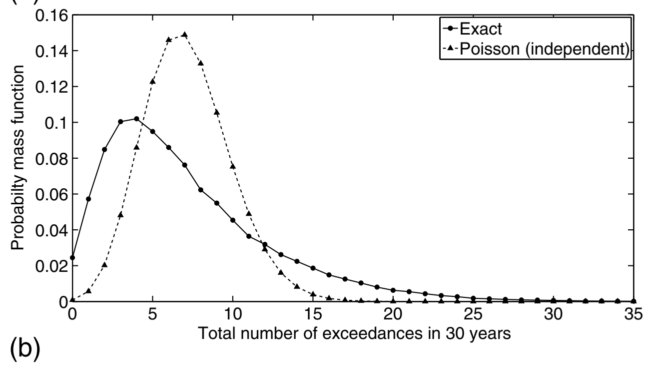


Figure 3. Probability mass function $P(\mathbf{Z}\in A_k)$ of the collective number of exceedances in one seismic event in the region with 68 sites (values up to 10 collective exceedances out of 68).

(a)



(b)

$P[N_{\#Ex}(30)=k]$;

k	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
PMF	0.02	0.06	0.09	0.10	0.10	0.09	0.09	0.08	0.06	0.05	0.05	0.04	0.03	0.03	0.02	0.02

Figure 4. (a) Distribution of total number of exceedances at 68 sites in 30 years; comparison with the independent (wrong) Poisson assumption. (b) Probability mass function of the total number of exceedances in 30 years, $P[N_{\#Ex}(30) = k]$, in a region made of 68 sites and discretized by means of 68 seismic sources. Only some k values are reported in (b).

earthquake, similar to what was done with equation (19); the only difference is that now there are 68 sites. Indeed, the obtained distribution, given in Figure 3, is the PMF of the collective number of exceedances in one event (from 0 to 68). Note that, in the figure, only the probabilities up to 10 joint exceedances (out of 68) are reported.

The distribution of the variable counting the total number of exceedances at the sites in 30 years, $N_{\#Ex}(30)$, is now explored. The mean and the variance are computed via equation (16) and given in equation (22). The variance in the independent case is equal to the mean, that is, 7.14. The difference with the (correct) dependent case is apparent:

$$\begin{cases} E[N_{\#Ex}(30)] = 7.14 \\ \mathrm{Var}[N_{\#Ex}(30)] = 27.68 \end{cases} . \tag{22}$$

The PMF of the total number of exceedances in 30 years, for the region made of 68 sites and discretized by means of 68 seismic sources, is pictured in Figure 4a, whereas its values, $P[N_{\#Ex}(30) = k]$ for $k = 0, 1, \ldots, 15$, are given in Figure 3b. Figure 3b can be used to compute the probability of observing at least 13 exceedances in 30 years, which is 1 minus the sum of the values until 12. This probability is equal to 0.14, which is larger than 0.05. Therefore, at the significance level 0.05, the null hypothesis that the map hazard is accurate (statistically consistent with observations) cannot be rejected, which is the contrary conclusion with respect to the case of assuming independence of site-specific processes.

Finally, to further picture the effect of the dependence among site-specific processes $N_{\mathrm{IM}_j > \mathrm{im}_j^*}(t), j = 1, 2, \ldots, 68$, on the PMF of the variable counting the total number of exceedances in 30 years, Figure 4 also shows how different such a distribution is from the Poissonian one considered in the case of independence assumption.

Conclusions

PSHA for multiple sites (i.e., computing the probability that a set of sites is going to collectively experience the exceedance of arbitrary ground-motion intensity thresholds in a given period of time) is a topic of current interest to engineering seismology and earthquake engineering for at least two reasons: (1) seismic risk assessment of multisite systems and (2) hazard validation studies.

Although site-specific hazard analysis is consolidated, the study provided a probabilistic formalization of hazard for multiple sites based on the theory of stochastic point processes. The approach considered working hypotheses consistent with those of site-specific (classical) PSHA. As a consequence, it was demonstrated that the random variable counting the total number of exceedances over time at the sites in a given period is, in general, not Poisson, even when each process counting exceedance of an IM threshold at a single site is Poisson (see point 6 below). Starting from this, a number of results were derived.

1. The key ingredients for regional hazard assessment are the rate of occurrence of earthquakes on the earthquake source(s) and a probabilistic description of the effects (in terms of exceedance or nonexceedance) that single earthquakes can produce at the considered sites.
2. When multiple sites are concerned, the only condition allowing the hazards for the sites to be independent is that a single earthquake can produce IM exceedance at one site at the most.
3. Based on point 2, it is apparent that, even if the residuals of the GMPE are uncorrelated, the processes counting exceedances over time at multiple sites are still stochastically dependent if the distance among them is not such to determine mutually exclusive exceedances in one earthquake event.
4. The joint distribution of the number of exceedances for a vector of two IMs at two sites was derived for an arbitrary time interval (the same scheme may be used to extend it to any number of sites).

5. The probability distribution of the total number of exceedances occurring in a given time for an arbitrary number of sites was also formulated, and closed-form expressions for the mean and the variance of the process counting the total number of exceedances were also obtained.
6. Multisite hazard is not regulated by a Poisson process, and its variance is underestimated in the case in which the presence of the discussed stochastic dependence is neglected; if only one site is considered, classical PSHA results and HPP are obtained.

Finally, illustrative applications have demonstrated the applicative value of the results and of the closed-form formulations.

Data and Resources

All data and resources are from the listed references.

Acknowledgments

The study presented in this article was developed partially within the activities of Rete dei Laboratori Universitari di Ingegneria Sismica (ReLUIS) for the research program funded by the Dipartimento della Protezione Civile (2014–2018) and partially in the framework of Analisi e Monitoraggio dei Rischi Ambientali (AMRA, S.c. a r.l.) for the Harmonized Approach to Stress Tests for Critical Infrastructures Against Natural Hazards (STREST), project funded by the European Community (Seventh Framework Programme for Research); Contract Number 603389. Helpful comments from Manuela Villani are also acknowledged.

References

Albarello, D., and V. D'Amico (2008). Testing probabilistic seismic hazard estimates by comparison with observations: An example in Italy, *Geophys. J. Int.* **175,** 1088–1094.

Ambraseys, N. N., K. U. Simpson, and J. J. Bommer (1996). Prediction of horizontal response spectra in Europe, *Earthq. Eng. Struct. Dynam.* **25,** 371–400.

Cornell, C. A. (1968). Engineering seismic risk analysis, *Bull. Seismol. Soc. Am.* **58,** 1583–1606.

Cornell, C. A., and H. Krawinkler (2000). Progress and challenges in seismic performance assessment, *PEER Center News* **3,** 1–3.

Eguchi, R. (1991). Seismic hazard input for lifeline systems, *Struct. Saf.* **10,** 193–198.

Esposito, S., and I. Iervolino (2011). PGA and PGV spatial correlation models based on European multievent datasets, *Bull. Seismol. Soc. Am.* **101,** 2532–2541.

Esposito, S., and I. Iervolino (2012). Spatial correlation of spectral acceleration in European data, *Bull. Seismol. Soc. Am.* **102,** 2781–2788.

Esposito, S., I. Iervolino, A. d'Onofrio, A. Santo, P. Franchin, and F. Cavalieri (2015). Simulation-based seismic risk assessment of gas distribution networks, *Comp.-Aid. Civ. Inf.* **30,** 508–523.

Gutenberg, B., and C. F. Richter (1944). Frequency of earthquakes in California, *Bull. Seismol. Soc. Am.* **34,** 185–188.

Iervolino, I. (2013). Probabilities and fallacies: Why hazard maps cannot be validated by individual earthquakes, *Earthq. Spectra* **29,** 1125–1136.

Iervolino, I., and M. Giorgio (2015). The effect of dependence of observations on hazard validation studies, *Proc. of CSNI Workshop on testing PSHA results and benefit of Bayesian techniques for seismic hazard assessment*, Eucentre Foundation, Pavia, Italy, 4–6 February 2015, doi: 10.13140/2.1.4860.4960.

Jayaram, N., and J. W. Baker (2009). Correlation model for spatially distributed ground motion intensities, *Earthq. Eng. Struct. Dynam.* **38,** 1687–1708.

Kingman, J. F. C. (1993). *Poisson Processes*, Oxford University Press, New York, New York.

Malhotra, P. K. (2008). Seismic design loads from site-specific and aggregate hazard analyses, *Bull. Seismol. Soc. Am.* **98,** 1849–1862.

McGuire, R. (1988). Seismic risk to lifeline systems: Critical variables and sensitivities, *Proc. 9th World Conference on Earthquake Engineering*, Tokyo, Japan, 2–9 August 1988.

McGuire, R. K. (2004). *Seismic Hazard and Risk Analysis*, Monograph MNO-10, Earthquake Engineering Research Institute, Oakland, California.

Park, J., P. Bazzurro, and J. W. Baker (2007). Modeling spatial correlation of ground motion intensity measures for regional seismic hazard and portfolio loss estimation, in *Applications of Statistics and Probability in Civil Engineering*, Taylor and Francis, London, United Kingdom, 1–8.

Resnick, S. I. (2002). *Adventures in Stochastic Processes*, Third Ed., Birkhauser, Boston, Massachusetts.

Rodriguez-Marek, A., G. A. Montalva, F. Cotton, and F. Bonilla (2011). Analysis of single-station standard deviation using the KiK-net data, *Bull. Seismol. Soc. Am.* **101,** 1242–1258.

Snyder, D. L. (1975). *Random Point Processes*, J. Wiley and Sons, New York, New York.

Weatherill, G., S. Esposito, I. Iervolino, P. Franchin, and F. Cavalieri (2014). Framework for seismic hazard analysis of spatially distributed systems, in *SYNER-G: Systemic Seismic Vulnerability and Risk Assessment of Complex Urban, Utility, Lifeline Systems and Critical Facilities*, K. Pitilakis, P. Franchin, B. Khazai, and H. Wenzel (Editors), Springer, The Netherlands, ISBN 978-94-017-8834-2.

Dipartimento di Ingegneria Industriale e dell'Informazione
Seconda Università degli Studi di Napoli
Via Roma 29
81031 Aversa (CE)
Italy
massimiliano.giorgio@unina2.it
(M.G.)

Dipartimento di Strutture per l'Ingegneria e l'Architettura
Università degli Studi di Napoli Federico II
Via Claudio 21
80125 Naples
Italy
iunio.iervolino@unina.it
(I.I.)

Manuscript received 19 December 2015;
Published Online 9 June 2016;
Corrected Online 17 June 2016

Bulletin of the Seismological Society of America, Vol. 107, No. 5, pp. 2540–2540, October 2017, doi: 10.1785/0120170167

Erratum to
On Multisite Probabilistic Seismic Hazard Analysis

by Massimiliano Giorgio and Iunio Iervolino

The article by Giorgio and Iervolino (2016) provides the basis for other studies, such as the development of a software tool (i.e., Iervolino *et al.*, 2016) and related applications of multisite hazard analysis (e.g., Iervolino *et al.*, 2017). During these further developments, a bug was found in the program used to carry out the last example in the article. In particular, figures 3 and 4 in Giorgio and Iervolino (2016) have to be replaced with Figures 3 and 4 here, respectively, and equation (22) should read as

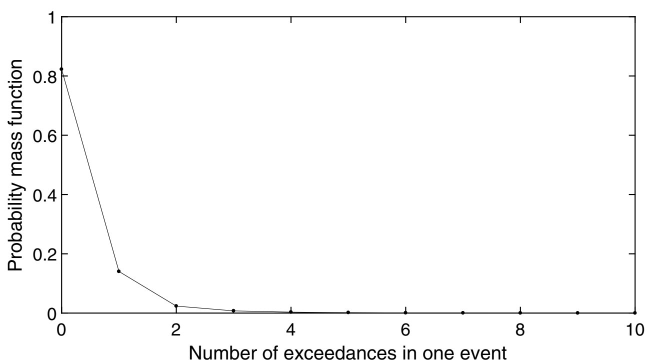
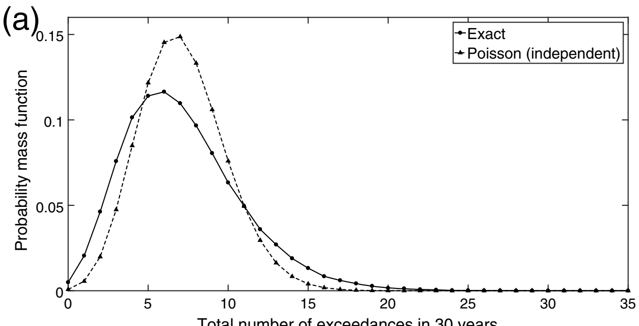


Figure 3. Probability mass function $P(\mathbf{Z} \in A_k)$ of the collective number of exceedances in one seismic event in a region with 68 sites (values up to 10 collective exceedances out of 68).

(a)



(b)

$P[N_{\#Ex}(30) = k]$;

k	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
PMF	0.00	0.02	0.05	0.08	0.10	0.11	0.12	0.11	0.10	0.08	0.06	0.05	0.04	0.03	0.02	0.01

Figure 4. (a) Distribution of the total number of exceedances at 68 sites in 30 years; comparison with the independent (wrong) Poisson assumption. (b) Probability mass function (PMF) of the total number of exceedances in 30 years, $P[N_{\#Ex}(30) = k]$, in a region with 68 sites and discretized by means of 68 seismic sources. Only some k values are reported in (b).

$$\begin{cases} E\left[N_{\#Ex}(30)\right] = 7.14 \\ \mathrm{Var}\left[N_{\#Ex}(30)\right] = 13.46 \end{cases} . \tag{22}$$

It is emphasized that neither the qualitative results of the example nor the discussions and conclusions of the study are affected by this bug, which is pointed out with this erratum to allow the interested reader to fully reproduce all the examples.

References

Giorgio, M., and I. Iervolino (2016). On multisite probabilistic seismic hazard analysis, *Bull. Seismol. Soc. Am.* **106,** 1223–1234, doi: 10.1785/0120150369.

Iervolino, I., E. Chioccarelli, and P. Cito (2016). REASSESS V1.0: A computationally-efficient software for probabilistic seismic hazard analysis, *VII European Congress on Computational Methods in Applied Sciences and Engineering, ECCOMAS*, Crete Island, Greece, 5–10 June.

Iervolino, I., M. Giorgio, and P. Cito (2017). The effect of spatial dependence on hazard validation, *Geophys. J. Int.* **209,** 1363–1368.

Dipartimento di Ingegneria Industriale e dell'Informazione
Università degli Studi della Campania Luigi Vanvitelli
via Roma 29
80131 Aversa (CE), Italy
massimiliano.giorgio@unicampania.it
(M.G.)

Dipartimento di Strutture per l'Ingegneria e l'Architettura
Università degli Studi di Napoli Federico II
via Claudio 21
80125 Naples, Italy
iunio.iervolino@unina.it
(I.I.)

Manuscript received 5 June 2017;
Published Online 25 September 2017