SPATIAL CORRELATION OF THE SYSTEMATIC SITE- AND PATH-SPECIFIC CORRECTIONS OF A GMPE CALIBRATED IN NORTHERN ITALY

Giovanni Lanzano¹, Sara Sgobba², Francesca Pacor³, Lucia Luzi⁴, Rodolfo Puglia⁵, Maria D’Amico⁶, Chiara Felicetta⁷

ABSTRACT

The probabilistic assessment of the seismic hazard (PSHA) at an individual site is a standard practice, but an ergodic assumption is commonly made: the ground-motion uncertainty computed by a Ground Motion Prediction Equation (GMPE) from a global dataset is assumed to be the same as the variability at a single site. In this paper, the ergodic assumption is relaxed by means of a residual analysis, accounting for the impacts on both the median and aleatory standard deviation of a GMPE. The aleatory variability is separated from the systematic source, path and site effects using a strong motion data set from Northern Italy with multiple recordings at each site and multiple earthquakes within small regions. A local model, specifically tailored for the area, is used as the reference GMPE, which predicts the geometric mean of horizontal response spectral accelerations in the period range 0.04-4s. The spatial covariance of such repeatable effects is modeled, in order to generate spatially correlated fields of path, source and site corrections and their associated variabilities. The results can be used to prepare fully non-ergodic hazard maps of parameters of engineering interest.

Keywords: GMPEs; Northern Italy; Seismic hazard; Non-ergodic PSHA; spatial correlation.

1. INTRODUCTION

When we perform a Probabilistic Seismic Hazard Analysis (PSHA), an ergodic assumption is commonly made (Anderson and Brune, 1999): an ergodic process is a random process in which the distribution of variables in space is assumed to be the same distribution in time at a given site. Hence, PSHA treats that spatial uncertainty of ground motions at many sites as an uncertainty of a single site over time. Different studies have demonstrated however that such assumptions does not account for more specific effects related to the earthquake source, recording site conditions and waveforms source-to-site path, which are not properly modelled by Ground Motion Prediction Equations (GMPEs). As consequence, the associated scatter around the median of the corresponding prediction is still considerable and thus a significant research effort to date is directed towards the development of methods useful to reduce the total variability. The most common approach in this field consists in performing a residual analysis (Al-Atik, 2010) that enables to identify and compute the systematic ground motion site-, location- and path-effects. The variability of these error components could be accounted for “epistemic” uncertainties, thus reducing the “aleatory” variability ones (e.g., Walling, 2009; Al Atik et al., 2010; Anderson and Uchiyama, 2011; Lin et al., 2011; Walling and Abrahamson, 2012; Dawood and Rodriguez-Marek, 2016).

¹Researcher, Istituto Nazionale di Geofisica e Vulcanologia, Milan, Italy, giovanni.lanzano@ingv.it
²Researcher, Istituto Nazionale di Geofisica e Vulcanologia, Milan, Italy, sara.sgobba@ingv.it
³Researcher, Istituto Nazionale di Geofisica e Vulcanologia, Milan, Italy, francesca.pacor@ingv.it
⁴Researcher, Istituto Nazionale di Geofisica e Vulcanologia, Milan, Italy, lucia.luzi@ingv.it
⁵Technologist, Istituto Nazionale di Geofisica e Vulcanologia, Milan, Italy, rodolfo.puglia@ingv.it
⁶Technologist, Istituto Nazionale di Geofisica e Vulcanologia, Milan, Italy, maria.damico@ingv.it
⁷Research Fellow, Istituto Nazionale di Geofisica e Vulcanologia, Milan, Italy, chiara.felicetta@ingv.it
2. METHOD

In a fully-ergodic approach, the prediction value $\bar{\mu}_{es}$ and the standard deviation $\bar{\sigma}$ are:

$$\bar{\mu}_{es} = \mu_{es}$$
$$\bar{\sigma} = \sigma_{tot} = \sqrt{\tau^2 + \phi^2}$$

where $\mu_{es}$ is the median prediction of the reference GMPE and the standard deviation $\sigma$ is calculated from the standard deviation $\tau$ of the between-event term, $\delta B_{e}$, and the standard deviation $\phi$ of the within-event term, $\delta W_{es}$. In the following notation, the subscripts $e$ relates to the seismic event, $s$ to the site, $r$ to the region and $p$ to the path.

$\delta B_{e}$ is the mean of the total residual for each event and represents the average misfit of the event ground-motion with respect to the median GMPE prediction. As a result, the within-event term is calculated as $\delta W_{es} = R_{es} - \delta B_{e}$ where $R_{es}$ is the logarithmic difference between observations and predictions.

In order to modify the GMPE median prediction and the related standard deviation for PSHA calculation, a non-ergodic approach is here proposed, following the work by Lanzano et al. (2017) in which the correction terms are inferred from the decomposition of the total residuals $R_{es}$. In a fully non-ergodic approach, the prediction value $\bar{\mu}_{es}$ and the standard deviation $\bar{\sigma}$ from equation (1) and (2), respectively, become as follows:

$$\bar{\mu}_{es} = \mu_{es} + \delta S2S_{e} + \delta P2P_{r} + \delta L2L_{r}$$
$$\bar{\sigma} = \sigma_{0} = \sqrt{\tau_{0}^2 + \phi_{0}^2}$$

where $\delta S2S_{e}$ is the site-to-site term, $\delta P2P_{sp}$ is the path-to-path term and $\delta L2L_{r}$ is the location-to-location term; $\tau_{0}$ is the standard deviation of the region- corrected between-event residuals $\delta B_{0,er}$, (defined as $\delta B_{0,er} = \delta B_{er} - \delta L2L_{r}$), and $\phi_{0}$ is the standard deviation of the event-, site- and path-corrected term $\delta W_{0,esp}$ (defined as $\delta W_{0,esp} = \delta W_{esp} - \delta P2P_{sp}$).
The site-to-site term $\delta S2S_s$ is computed as mean of the within-event residuals for each recording site. It quantifies the average misfit of the station ground-motion, predicted by the reference GMPE, with respect to the event-corrected median value. $\phi_{WS}$ is the standard deviation of the event- and site-corrected term $\delta WS_s$, over all stations, calculated as $\delta WS_s = \delta W_{es} - \delta S2S_s$.

The standard deviation of the event- and site-corrected term at an individual station $s$, $\phi_{WS,s}$, can be calculated as:

$$\phi_{WS,s} = \sqrt{\frac{\sum_{e=1}^{NE_s} \delta WS_{es}}{NE_s - 1}}$$

where $NE_s$ is the number of earthquakes recorded by the station $s$. Then, the single-station sigma at individual site $\sigma_{WS,s}$ can be calculated similarly to Eq. (4), replacing $\phi_{WS}$ with $\phi_{WS,s}$.

The location-to-location $\delta L2L_r$ term indicates how the ground motion of an event recorded in a given small region differs from the mean prediction of the events in the entire source regions and can be calculated as the mean of the between-event residuals $\delta B_{sr}$ for each source region. The standard deviation of $\delta L2L_r$ over all the regions is $\tau_{L2L}$. $\tau_{0,r}$ is the standard deviation of region-corrected between-event in a single-region is calculated:

$$\tau_{0,r} = \sqrt{\frac{\sum_{e=1}^{NE_r} \delta B_{sr}^2}{NE_r - 1}}$$

The path-to-path term $\delta P2P_{sp}$ represents how the specific characteristics of a travel path lead to ground-motions that are systematically different from the ground-motion predicted by the GMPE and is calculated as the mean of the event- and site-corrected residuals from earthquakes recorded at the station $s$ relative to a path $p$. Hence, for each station we can define a number of $\delta P2P_{sp}$ depending on the sampled source-to-site paths. The standard deviation of the $\delta P2P_{sp}$ over all the stations and regions is $\phi_{P2P}$. The event-, site- and path-corrected standard deviation for a station $s$ and path $p$, $\phi_{0,sp}$, can be calculated as:

$$\phi_{0,sp} = \sqrt{\frac{\sum_{e=1}^{NE_r} \delta W_{es}^2}{NE_r - 1}}$$

The single-station single-path sigma at individual site $\sigma_{0,sp}$ can be calculated similarly to Eq. (4), replacing $\phi_0$ with $\phi_{0,sp}$ and $\tau_0$ with $\tau_{0,r}$.

Following the procedure describe above, it is possible to compute the path-specific GMPE median correction as sum of the site-to-site $\delta S2S_s$, location-to-location $\delta L2L_r$ and path-to-path term $\delta P2P_{sp}$ terms and also evaluate the associated variability single-station single-path sigma $\sigma_{0,sp}$ at individual site.

3. CASE-STUDY FRAMEWORK

The strong-motion dataset here used is relative to the study area of Northern Italy and it was used by Lanzano et. al. (2017) to evaluate the systematic components of ground motion variability. The set of data was selected from the ITalian ACcelerometric Archive (ITACA 2.0: Luzi et al., 2008; Pacor et al., 2011a) and Engineering Strong Motion Database (ESM: Luzi et al. 2016). The selected recordings are 2241, relative to 88 events and 168 stations (for more details see Lanzano et al., 2017). The Mw 6.4 Friuli 1976–1977 and the Mw 6.1 Emilia 2012 seismic sequences are included, both characterized by the reverse style of faulting.
Moreover, in the same area a region-specific GMPE has been calibrated by Lanzano et al., (2016) and used as reference model for the residual analysis computation. Figure 1 shows the geographic distribution of the seismic events and recording stations in the study area defined in the latitude range 44.0°–46.3° N and longitude range 8.00°–13.5° E. The seismogenic zonation (ZS9, Meletti et al. 2008) was used to define the source regions from which the path-to-path and the location-to-locations terms were computed. As reported by Lanzano et al. (2017), about one-third of the events belongs to the 2012 Emilia sequence, within zone ZS9 912, corresponding to 70% of records in the selected dataset. About 1/3 of the stations have a number of records between 5 and 10, while about 90 stations (54%) sampled only one source-to-site path, mainly relative to events occurred in the zone 912. As discussed in Lanzano et al. (2016), stations belonging to class C according to Eurocode 8 (EC8) classification (i.e., shear wave velocity \(V_{S,30}\) between 180 and 360 m/s), located in the deepest part of the Po Plain or in smaller basins in the Apennines, have been isolated and grouped into a new site class named C1. The rationale for the modification of the site classification is that for C1 complex 2D-3D site effects are expected due to the presence of surface waves generated at the basin edges, with remarkable soil amplification at frequencies lower than 1 Hz. Since 1/3 of the data events are generated below the Po plain, recordings at short distances are mainly associated to sites C and C1 (Figure 2a). Figure 2b shows the trends of the region site amplifications with respect to rock site as a function of period, compared to the C site amplifications, predicted by the GMPE calibrated for Italy (ITA10 – Bindi et al., 2011). It can be noted that sites located inside sedimentary basins (denoted with C1) show a very different trend compared to site located outside (class C). Furthermore, at short periods (< 0.4s) this correction accounting for basin effects gives coefficients about 1.3 times lower than that predicted for ITA10 class C. This might be explained with a strong attenuation of short period phases propagating through thick sedimentary covers. Conversely, at long periods (> 0.4s), sites located inside sedimentary basins show an increase of amplification, up to 1.5 times compared to ITA10 class C.
In this study, we exploited the results of Lanzano et al. (2017) to compute the site-to-site term, $\delta S_S$, the path-to-path term, $\delta P_{SP}$, the location-to-location term, $\delta L_{LR}$, and the single-site single-path variability, $\sigma_{0,rs}$. To decompose the event- and site-corrected residuals, each station had to sample more than one source-to-site path. Indeed, if a single station records multiple events from a single region, the site and path contributions cannot be separated: the site term will include both the effects, while the path-corrected variability will coincide with the event- and site-corrected sigma at single-site.

As main outcome, the standard deviation $\sigma_0$ obtained by removing the systematic site- and path-components of variability exhibits an average reduction of about 37% with respect to total variability $\sigma_{tot}$ (Lanzano et al., 2017). The authors also found that the variability of the location-to-location term is quite low and smaller than the other studies. Indeed, the term $\delta L_{LR}$ is almost negligible for the majority of the ZS9 regions, probably as a consequence that a large part of the waveforms of the dataset belongs to a single seismic sequence (2012 Emilia), thus reducing the variability among the seismic sources.

4. SPATIAL CORRELATION OF SYSTEMATIC TERMS

We consider a sub-area of study that includes the epicenters of the Mw 6.1 Emilia 2012 seismic sequences that is defined in the latitude range 44.6°–45.1° N (equivalent to 55.6 km) and longitude range 10.6°–11.8° E (equivalent to 95 km), for a total area of about 5000 km$^2$ (Figure 3).

The study area includes the largest alluvial basin in Europe, which is also the densest area of temporary seismic stations installed in the Po Plain area soon after that 2012 Emilia sequences started. As consequence, multiple seismic recordings were available at several stations in near-source conditions (for the most part referred to the 2012 Emilia seismic sequence).

The sub-set of data is composed by 49 recording stations, characterized by at least 5 recordings, belonging to C1 site category (see § 3). The evidence that all the stations in the sub-set are located on the same site class has a relevant impact on the evaluation of local site effects in the residual analysis since the soil amplification is intended to be referred to the specific site class C1, not to traditional bedrock conditions.

In this study, as in Lanzano et al. (2017), the number of source regions $r$ corresponds to the maximum number of the paths $p$ sampled by each station included in the dataset. Within this subset, the most of stations have only recorded the Emilia 2012 sequence (one region and one path), thus for them the site- and path- contributions cannot be separated and the path-to-path $\delta P_{SP}$ term is null.
Figure 3 – Spatial distribution of stations (triangles) in the sub-area of study. Yellow stars and red rectangles represent the epicentres and the surface projection of the faults of the two main shocks of the 2012 Emilia seismic sequence. Black lines delimit the ZS9 seismic source zone 912.

As an example, Figure 4 shows the median correction as a function of period for two nearby stations. IV.MODE (Figure 4a) is a permanent station of the RSN network (managed by INGV) and sampled four paths, with the main contribution (26 records over 37) from the ZS9 912 source zone, which corresponds to the area source of the 2012 Emilia seismic sequence. In this latter case, the absolute $\delta P2P_s$ term is significantly lower than the site-to-site term (black line) for periods longer than $T=0.2s$. IV.RAV0 (Figure 4b) is a temporary station installed by DPC (Italian Department of Civil Defence) after the first mainshock and recorded events of the 2012 seismic sequence only (ZS9 912). For this reason, the $\delta P2P_s$ is null and $\delta S2S_s$ is representative of a single-site single-path correction of the median.

Table 1 provides the codes of the selected stations with the number of records, the site- and path-correction term to the median prediction of the GMPE and the corresponding variability at the site. Table 1 includes the two stations (IV.MODE and IT.MRN) with the largest number of records; two stations with extreme median correction terms (positive for IV.T0827 and negative for IV.T0821); and two stations with the lowest (TV.MIR07) and highest (IT.SAN0) variabilities.
Table 1. Selected stations and corresponding site- and path- specific variability terms at different periods.

<table>
<thead>
<tr>
<th>Net code</th>
<th>Station code</th>
<th># recs</th>
<th>PGA</th>
<th>T = 0.2s</th>
<th>T = 0.3s</th>
<th>T = 1s</th>
<th>T = 2s</th>
<th>T = 4s</th>
</tr>
</thead>
<tbody>
<tr>
<td>IV</td>
<td>MODE</td>
<td>26</td>
<td>0.177</td>
<td>-0.371</td>
<td>-0.040</td>
<td>0.351</td>
<td>0.315</td>
<td>0.351</td>
</tr>
<tr>
<td>IT</td>
<td>MRN</td>
<td>21</td>
<td>0.103</td>
<td>0.109</td>
<td>0.021</td>
<td>-0.085</td>
<td>-0.045</td>
<td>-0.077</td>
</tr>
<tr>
<td>IV</td>
<td>T0821</td>
<td>17</td>
<td>0.330</td>
<td>-0.306</td>
<td>-0.281</td>
<td>-0.451</td>
<td>-0.371</td>
<td>-0.514</td>
</tr>
<tr>
<td>IT</td>
<td>SAN0</td>
<td>16</td>
<td>0.053</td>
<td>-0.087</td>
<td>-0.068</td>
<td>-0.196</td>
<td>-0.146</td>
<td>-0.111</td>
</tr>
<tr>
<td>TV</td>
<td>MIR07</td>
<td>12</td>
<td>0.008</td>
<td>-0.189</td>
<td>-0.076</td>
<td>0.492</td>
<td>0.389</td>
<td>0.323</td>
</tr>
<tr>
<td>IV</td>
<td>T0827</td>
<td>5</td>
<td>0.514</td>
<td>0.380</td>
<td>0.462</td>
<td>0.396</td>
<td>0.482</td>
<td>0.543</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Variability at individual site</th>
</tr>
</thead>
<tbody>
<tr>
<td>IV</td>
</tr>
<tr>
<td>IT</td>
</tr>
<tr>
<td>IV</td>
</tr>
<tr>
<td>IT</td>
</tr>
<tr>
<td>TV</td>
</tr>
<tr>
<td>IV</td>
</tr>
</tbody>
</table>

Figure 5a shows the maps of the PGA median correction term for each selected station. Red circles represent positive corrections with respect to mean value, while blue circles negative corrections. Figure 5b shows the maps of the single-site and single-path variability for the same stations. In both figures, the size of the circles is proportional to their absolute value. The extreme values are reported in the title. Data seem spatially clustered since groups of close stations with similar (positive red circles/negative blue circles) values can be clearly recognized. In particular, positive values (red circles) occur at the recording sites nearby the epicentral locations, especially for the second mainshock. Large negative values (blue circles) are observed at the right of the first mainshock. No clear patterns are evident for the non-ergodic variability (Figure 5b): in general, larger variabilities are found in the epicentral area, while smaller values are evident toward the North direction. These evidences lead us to try to build up a spatial correlation model for the median correction terms and for the associated variability, in order to estimate the elements for non-ergodic PSHA where recording stations are not present.
Spatial correlation modelling

It is known that the bias from GMPE median prediction between two sites is dependent on their spatial proximity and thus the correction of a GMPE at a given site without recording data can be inferred by the surrounding stations due to systematic site- and path-effects. All above premised, we applied the classical geostatistic technique to model the spatial covariance of the median correction term and related variability through the semi-variogram function, which provides a measure of the strength of the spatial correlation.

A weighted least square regression method was used to fit a semi-variogram exponential model to the empirical data. Specifically, the best fitting to data was obtained by using an exponential function with

\[ \gamma(h) = a - \exp\left(-\frac{3h}{b}\right) \]

in which the “sill” represents the variance and the “range” is the distance beyond which data are assumed to be spatially uncorrelated. In Figure 6 the empirical model of semi-variogram computed as a function of the inter-site separation distance \(d\) is shown. In detail, the latter has been divided into 30 segments (bins) of 700 m length. Each data point in Figure 6a and 6b represents the semi-variance of the GMPEs median correction while points in Figure 6c and 6d are the semi-variance of the single-station single-path sigma, both plotted in the common distance bins. The semi-variogram \(\gamma^*(h)\) plots of the median correction terms is shown in Figure 6a and 6c at period \(T=0.2\) s and in Figure 6b and 6d at period \(T=4.0\) s.

As for typical semi-variogram plots, the observed trends in Figure 6 are characterized by an increasing degree of spatial correlation with decreasing separation distance. Data appear more scattered and poorly constrained at large separation distances, particularly for variability data at higher periods. Larger scatter is also observed for non-ergodic sigma \(\sigma_{0,ST}\) particularly at longer periods. For the latter, no clear pattern is recognized, thus confirming a poor spatial correlation for the variability. This evidence is probably related to the intrinsic aleatory nature of the non-reducible site- and path- corrected variability, which is expected to be more randomly distributed around space.
Figure 6 – Semi-variogram model of the median correction (a, b) and non-ergodic standard deviation (c, d) at periods $T=0.2$ s (a, c) and $T=4.0$ s (b, d).

Figure 6 also shows that at longer period (b, d), data are best fitted by a linear model and consequently the coefficients $a$ and $b$ tend to increase. In particular, larger $b$ value indicate a smaller rate of decay of correlation with separation distance (Baker, 2008). It can be thus inferred that median correction and the associated variability computed at long periods show larger correlations than those computed at short periods. This has already observed in the literature since short period waves tend to be more affected by the heterogeneities of the propagation path and as a result tend to be less coherent than long period ground motions (Zerva and Zervas, 2002).

A more direct comparison between the observed trends at different periods can be performed in Figure 7 on the basis of the coefficients $a$ and $b$ of the best-fitting models. In the range graph (Figure 7b) we also reported the maximum fitting inter-distance $d$ (about 20km), which is significantly lower than the $b$ values calibrated by the correlation model at several periods. It means that, in the study area, the investigated parameters are still significantly correlated, especially for the median correction. It is pointed out that in Figure 7 the largest values of the sill and range coefficients for larger periods are not shown as they have been calibrated on the linear model, thus exhibiting unrealistic overestimation.
4.2 Spatial interpolation by kriging method

Based on the spatial correlation models, obtained in the previous section, we apply the kriging algorithm (Davis, 1973) to obtain maps of the non-ergodic parameters in the study area. Figure 8 shows some kriging maps of the median correction and of single-station single-path sigma at two example periods 0.2 s (Figure 8a, c) and 4 s (Figure 8b, d), respectively.

In Figure 8 clear pattern are recognizable in the median correction term (Figure 8a, b) and the site- and path-corrected standard deviation term at T=0.2 s (c) and T=4.0 s (d). Specifically, red contour in Figure 8a looks as a homogeneous area of the map around the Emilia 1-st and 2-nd shocks where the reference GMPE tends to systematically underestimate the effect of shaking at short periods (0.2 s). Differently, the blue
contour area indicates an attenuation in ground motion effects predicted by the model. The observed red area appears to elongate mainly in the North-South direction with period increasing (4s in Figure 8b). This evidence could be physically interpreted as an effect of the complex geomorphological structure beneath the plain and more specifically to the propagation of surface waves for the Emilia earthquake sequences, according to the findings of some studies (Luzi et al., 2013; Massa and Augliera, 2013). These authors found that at relatively long period range (mainly between 3 and 20 s), basin resonance, amplification effects, and long shaking duration appeared at stations within and bordering the basin, due to the presence of thick sediments. Also in the maps related to the corrected variability (Figure 8c, d) a specific trend is recognizable. In Figure 8c, yellow-to-red graded contour area is dominant, particularly in the bottom part of the map. This evidence indicates that residual variability $\sigma_{0,\text{sr}}$ is higher in this zone at short period ($T = 0.2s$), while at relatively long periods ($T=4.0$ s, Figure 8d) this area shows an overall reduction (predominance of blue-graded countour are in Figure 8d compared to Figure 8a), indicating an higher degree of variability at shorter periods. Moreover, it can be observed that in Figure 8b the red gradations tend to elongate mainly in the East-West direction, just below the location of the two epicenters. A possible physical interpretation to this evidence could be related to some directional effect at longer periods in the near source region not accounted by the reference GMPE. These physical-related aspects need to be more carefully investigated in further developments of the present study.

5. CONCLUSIONS

In this study a fully non-ergodic approach has been proposed to compute the median correction of ground-motion prediction and its variability with the aim to improve the assessment of PSHA. The results of residual analysis have been shown to study the overall trend of the repeatable site-specific, location-specific and path-specific correction terms and their deviation from the median model. The residual decomposition has been performed on a reliable dataset related to the Northern Italy region, which is the largest alluvial basin in Europe (i.e., Po Plain area) and also the densest area of seismic stations installed soon after that the first- and second-2012 Emilia sequences started. This “ideal” dataset mainly resembles a single-path dataset, including repeated observations of earthquakes located in a small region (zone ZS9-912, relative to the Emilia sequence), thus allowing to properly evaluate specific source-, path- and site- effects on the ground motion model in this zone.

The spatial correlation of the computed systematic residual components modelled by means of semi-variogram function and kriging interpolation allowed to gather some relevant information about the regional effects of ground motion not accounted by the reference GMPE, even if more general conclusions will be drawn out in further developments.

6. REFERENCES


