

Spatial Correlation of Acceleration Spectral Ordinates from European Data

S. Esposito & I. Iervolino

Università degli Studi di Napoli Federico II, Italy



SUMMARY

Spatial correlation of spectral acceleration (S_a) is required in modelling hazard for risk assessment of spatially distributed systems. The estimation of an appropriate correlation model is still a research task in earthquake engineering, because of several issues related to: data, statistical approaches and estimation tools, and tests to evaluate the estimated models. This paper focuses on the evaluation of the spatial correlation of S_a using records belongs the European Strong motion Database (ESD). The estimation has been performed on multiple earthquakes, considering a ground motion prediction equations (GMPE) fitted to the same records. Results show that the decay rate of correlation, as a function of inter-site distance, tends to increase with structural period. Based on that, a simple linear relationship for the correlation length as a function of structural period is derived, discussed, and compared with previous research on the same topic.

Keywords: spectral acceleration, ground motion residuals, semivariogram.

1. INTRODUCTION

The modeling of spatial correlation of ground motion intensity measures (IMs) has become a relevant topic in seismic risk analysis due to the requirement to extend risk assessment, usually related to site-specific structures, to spatially distributed systems such as building portfolios and lifelines. In particular, on the hazard side, probabilistic seismic hazard analysis (PSHA; McGuire, 2004) refers to ground motion prediction equations (GMPEs), which provide probabilistic distribution of the chosen IM (e.g., elastic spectral acceleration, S_a) conditional on earthquake magnitude, source-to-site distance, and other parameters such as local geological conditions. Since it was demonstrated that GMPE residuals are spatially correlated (e.g., Boore et al., 2003), it may be the case that neglecting such a correlation bias the loss assessment of distributed systems (e.g., Esposito, 2011).

Spatial correlation models available in literature (e.g., Goda and Hong, 2008; Jayaram and Baker, 2009; Sokolov et al., 2010) depend on inter-site separation distance and provide the limit at which correlation may technically considered to be lost (i.e., distance beyond which IMs may be considered uncorrelated). Moreover, if S_a is of concern, the correlation may depend also on the oscillation period S_a refers to.

It is important to note that those models provide different correlation length even if estimating correlation for the same IM. This is supposed to depend on several factors such as the dataset used, the GMPE chosen to compute intraevent residuals, and the working assumptions of the estimation. For example, Sokolov et al. (2010), starting from the strong-motion database of TSMIP network in Taiwan, investigated the dependency of spatial correlation on site classes and geological structures, asserting that a single generalized spatial model may not be adequate for all of Taiwan territory.

In some cases (e.g., Wang and Takada, 2005; Jayaram and Baker, 2009) existing GMPEs are used, while, in others (e.g., Goda and Hong, 2008; Goda and Atkinson, 2009; Sokolov et al., 2010), ad-hoc fit on the chosen dataset is adopted. Generally, regressions analysis used to develop prediction equations does not incorporate the correlation structure of residuals as a hypothesis. Hong et al. (2009) and Jayaram and Baker (2010), evaluated the influence of considering the correlation in fitting a

GMPE, finding a minor influence on regression coefficients and a more significant effect on the variance components. Goda and Atkinson (2010), investigated the influence of the estimation approach, emphasizing its importance when residuals are strongly correlated.

Herein, the evaluation of the spatial correlation of S_a (5% damped) residuals is carried out using the European Strong motion Database (ESD; <http://www.isesd.hi.is/>). The GMPE with respect to which residuals are computed is those of Akkar and Bommer (2010) and only the records used to fit the considered GMPE are employed to estimate spatial correlation.

In the paper, the first part briefly describes the geostatistical framework adopted to estimate the correlation. Subsequently, the working assumptions and a description of the considered dataset, are given. Then, results of the estimation of correlation lengths for spectral acceleration at nine structural periods, ranging from 0.1 s to 2.85 s are discussed. Finally, a simple linear relationship for the correlation length as a function of structural period is derived and compared with previous research on the same topic.

2. MODELING SPATIAL CORRELATION

2.1. Geostatistical framework

The spatial modeling of IMs for earthquake engineering purposes is mostly based on estimation of correlations between residuals of ground motion parameters at two different sites via geostatistical tools. In particular, focusing on spectral acceleration, GMPEs generally model the logarithms of S_a for a specific structural period, T , and related heterogeneity, at a site p due to earthquake j as in Eqn. 2.1:

$$\log Sa(T)_{pj} = \overline{\log Sa(T)_{pj}}(M, R, \underline{\theta}) + \eta_j + \varepsilon_{pj} \quad (2.1)$$

where $\overline{\log Sa(T)_{pj}}(M, R, \underline{\theta})$ is the mean of the logs conditional on parameters such as magnitude (M), source-to-site distance (R), and others ($\underline{\theta}$); η_j denotes the inter-event residual, which is a constant term for all sites in a given earthquake and represents the systematic deviation of the specific seismic event from the overall mean; and ε_{pj} is the intraevent variability of ground motion. ε_{pj} and η_j are usually assumed to be independent random variables, normally distributed with zero mean and standard deviation σ_{intra} and σ_{inter} , respectively. Consequently, $\log Sa(T)_{pj}$ is modeled as a normal random variable with mean $\overline{\log Sa(T)_{pj}}(M, R, \underline{\theta})$ and standard deviation σ_T , where $\sigma_T^2 = \sigma_{inter}^2 + \sigma_{intra}^2$.

If hazard assessment at two or more sites at the same time is of concern, the joint probability density function (PDF) of IMs at all locations can be conveniently modeled with a multivariate normal distribution (Jayaram and Baker, 2008). In this case, it is assumed that the logs of IM form a Gaussian random field (GRF), defined as a set of random variables, one for each site \mathbf{u} in the study area $S \in \mathbb{R}^2$. In particular, ground motion parameters at n spatial locations $\mathbf{u}_p \{p=1, \dots, n\}$ are modeled as a collection of random variables characterized by a mean vector provided by the GMPE and a covariance matrix, Σ , as in Eqn. 2.2, where the first term produces perfectly correlated inter-event residuals (Malhotra, 2008), while the second term (symmetrical) produces spatially correlated intraevent residuals. In Eqn. 2.2, spatial correlation for intraevent residuals is function of the separation distance h (under the hypothesis of second-order stationarity and isotropy of the GRF); i.e., the spatial heterogeneity does not depend on the direction considered.

$$\Sigma = \sigma_{inter}^2 \cdot \begin{bmatrix} 1 & 1 & \cdots & 1 \\ 1 & 1 & \cdots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \cdots & 1 \end{bmatrix} + \sigma_{intra}^2 \cdot \begin{bmatrix} 1 & \rho(h_{12}) & \cdots & \rho(h_{1n}) \\ \rho(h_{21}) & 1 & \cdots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ \rho(h_{n1}) & \rho(h_{n2}) & \cdots & 1 \end{bmatrix} \quad (2.2)$$

The covariance matrix typically depends on parameters obtained from a non-negative definite spatial parametric covariance model (Cressie, 1993) that may be fitted on empirical observations. To quantify spatial variability of geo-referenced data the semivariogram is often used, $\gamma_j(h)$, that under the hypothesis of second-order stationary and isotropy mentioned above, it is strictly related to the correlation coefficient and it is defined as in Eqn. (2.3):

$$\gamma_j(h) = \text{Var}(\varepsilon_j) \cdot [1 - \rho_j(h)] \quad (2.3)$$

where $\text{Var}(\varepsilon_j)$ is the homoscedastic variance of intraevent residuals, and $\rho_j(h)$ denotes the spatial correlation coefficient between intraevent residuals at two generic sites separated by h -distance.

The estimation of correlation based on semivariograms usually deploys in three steps: (1) computing the empirical semivariogram¹ $\hat{\gamma}(h)$; (2) choosing a functional form to fit it; and (3) estimating the parameters of the chosen function fitting empirical data.

Regarding point (1): empirical semivariograms are computed as a function of site-to-site separation distance, with different possible estimators. The classical estimator is the method-of-moments (Matheron, 1962), which is defined for an isotropic random field in Eqn. (2.4), where $\varepsilon(\mathbf{u} + \mathbf{h}) - \varepsilon(\mathbf{u})$ is the difference between intraevent residuals at sites separated by h -distance value, $N(h)$ is the set of pairs of sites, separated by the same h -distance in the dataset, at which the residual of interest are measured, and $|N(h)|$ is the cardinal of $N(h)$.

$$\hat{\gamma}(h) = \frac{1}{2 \cdot |N(h)|} \cdot \sum_{N(h)} [\varepsilon(\mathbf{u} + \mathbf{h}) - \varepsilon(\mathbf{u})]^2 \quad (2.4)$$

Since this estimator can be badly affected by atypical observations (Cressie, 1993), Cressie and Hawkins (1980) proposed a more *robust*, in the statistical sense (i.e., less sensitive to outliers) estimator. Consistent with Esposito and Iervolino (2011), both estimators are used in the following for the evaluation of intraevent spatial correlation of $Sa(T)$.

Because earthquake records are not available on a regular grid, to compute the empirical semivariogram it is necessary to define tolerance bins around each possible h -value and, hence, to divide data in classes. For the selection of distance a rule of thumb is to choose the maximum bin size as a half of the maximum distance between sites in the dataset, and to set the number of bins so that there are at least thirty pairs per bin (Journel and Huijbregts, 1978).

Point (2), concerning the interpretation of experimental semivariograms, consists in the identification of a model among the family of functions able to capture and emulate its trend. The three basic stationary and isotropic models are: exponential, spherical, and Gaussian (Goovartes, 1997). In

¹ As assumed in Esposito and Iervolino (2011), in this work the estimation of spatial correlation was performed using all data available from different earthquakes and regions (deemed homogeneous), differently from studies focusing on earthquake-specific data (e.g., Jayaram and Baker, 2009). This assumption allows to neglect the subscript j in the following equations assuming a common semivariogram for different events (i.e. invariant through earthquakes).

particular, the exponential model which is the most common one in literature, is described in Eqn. (2.5):

$$\gamma(h) = c_0 + c_e \cdot (1 - e^{-3 \cdot h/b}) \quad (2.5)$$

where c_0 is defined *nugget*, that is, the limit value of the semivariogram when h tends to zero, c_e is the *sill*, or the population variance of the random field (Barnes, 1991), and b is the *practical range*, or correlation length, defined as the inter-site distance at which $\gamma(h)$ equals 95% of the sill, or in other words, the distance at which the correlation may be considered technically lost and residuals of ground motion intensity may be considered independent (under the GRF assumption).

Regarding point (3): several goodness-of-fit criteria to find the best parametric model have been proposed in geostatistical literature. Studies dealing with earthquake data sometimes use visual or trial and error approaches in order to appropriately model the semivariogram structure at short site-to-site distances, where it is significant (Jayaram and Baker, 2009). In this work experimental semivariograms are fitted using recursively the least squares method (LSM) on data obtained with the robust estimator.

2.2. Estimating correlation on multievent data

As assumed in Esposito and Iervolino (2011), in this work the estimation of spatial correlation was performed using all data available from different earthquakes and regions. Moreover in order to estimate only one parameter, that is the range, empirical semivariograms were computed starting from normalized intraevent residuals obtained, expressed for a single earthquake j and a generic site p , as

$\varepsilon_{pj}^* = \varepsilon_{pj} / \sigma_p$, where σ_p is the standard deviation of the intraevent residual at the site p (in the study, the intraevent standard deviation is common for all sites consistent with GMPEs used to compute residuals²). In fact the standardization enables to not estimate the sill, as it should be equal to one. Therefore, equation (2.3) becomes equation (2.6), where the superscript represents an empirical estimate.

$$\hat{\gamma}(h) = 1 - \hat{\rho}(h) \quad (2.6)$$

Normalized intraevent residuals from multiple events (and regions) are, then, pooled to fit a unique correlation model. This because geostatistical estimation needs a relatively large number of data to model the semivariogram (i.e., such to have more than thirty pairs in each h bin), which are not available for individual events in the chosen datasets (to follow).

The use of all data implies the assumption of the same isotropic semivariogram with the same parameters for all earthquakes; therefore the experimental semivariogram becomes that of Eqn. (2.7) where $|N(h)|$ is the number of pairs in the specific h bin from all earthquakes and the differences, $\varepsilon_{pj}^* - \varepsilon_{qj}^*$, are taken considering only pairs from the same earthquake without mixing up residuals from different events. Therefore individual events are kept separated, and pairs in the specific bin are then pooled to obtain empirical semivariogram points. This is formally shown in Eqn. (2.8), where n_j represents the number of records for the j -th event.

² There are different alternatives for standardization (e.g., Jayaram and Baker, 2010; Goda and Atkinson, 2010), but they lead to results which seem to be not significantly affected by a choice with respect to another (Esposito and Iervolino, 2011).

$$\hat{\gamma}(h) = \frac{1}{2 \cdot |N(h)|} \sum_{N(h)} [\varepsilon_{pj}^* - \varepsilon_{qj}^*]^2 \quad (2.7)$$

$$N(h) = \left\{ \left(j, \varepsilon_{pj}^*, \varepsilon_{qj}^* \right) : \left\| \varepsilon_{pj}^* - \varepsilon_{qj}^* \right\| = h; p, q = 1, \dots, n_j; j = 1, \dots, k \right\} \quad (2.8)$$

3. DATA AND ANALYSIS

3.1. Dataset

The considered recordings from ESD dataset exactly correspond to data used in Esposito and Iervolino (2011); i.e., records used to fit the Akkar and Bommer (2010) GMPE (except that, only earthquakes for which more than one record was available were considered). The experimental semivariograms were obtained using a bin width of 4 km. Characteristics of the dataset, and distribution of data pairs as a function of separation distance bins, are shown in Fig. 3.1.

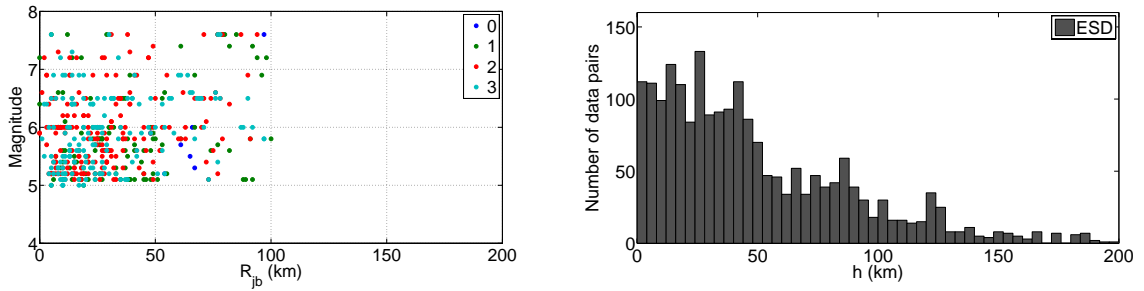


Figure 3.1. a) The ESD subset with respect to M , R_{jb} (i.e. the closest horizontal distance to the vertical projection of the rupture, or Joyner-Boore distance) and local site conditions: rock (3), stiff soil (2), soft soil (1), and very soft soil (0); b) histograms of the number of data pairs as a function of site-to-site separation distance bin (4 km).

3.2. Results

The estimation was performed for nine structural periods ranging between 0.1 s and 2.85 s. Because the GMPE used to obtain intraevent residuals refers to geometric mean of horizontal components, the correlation was estimated for this IM. Both estimators (classical and robust discussed above) were used; no significant differences, as expected, were found in the shape of the experimental semivariogram. The robust estimator was, then, considered to fit the correlation function.

The exponential model was chosen to fit empirical points, since this model is widely adopted in literature. Moreover, the choice of using the same model for all periods allows to compare results and to investigate possible dependency of the model parameters on the fundamental period. Assuming that there is no nugget effect (as this study does not investigate variations at a smaller scale with respect to that of the tolerance); the only parameter to estimate is the range, b in Eqn. (2.5).

To this aim, LSM was used iteratively to fit the model on the empirical semivariogram. In particular, the range was defined starting from the value estimated by the LSM applied until half of the maximum separation distance of residuals in the dataset. The range obtained from this step was, then, used as a lower bound for the limit distance to apply LSM again on the semivariogram. This latter step allowed to adjust the fitting and to obtain the final practical range values, that is, the distances at which correlation may be considered to be lost for each spectral ordinate.

Following this procedure, the estimated ranges for the estimated exponential models were retrieved for ESD. Fig. 3.2 shows empirical semivariograms with both estimations and fitted models and, as a

summary, practical ranges are given in Table 3.1.

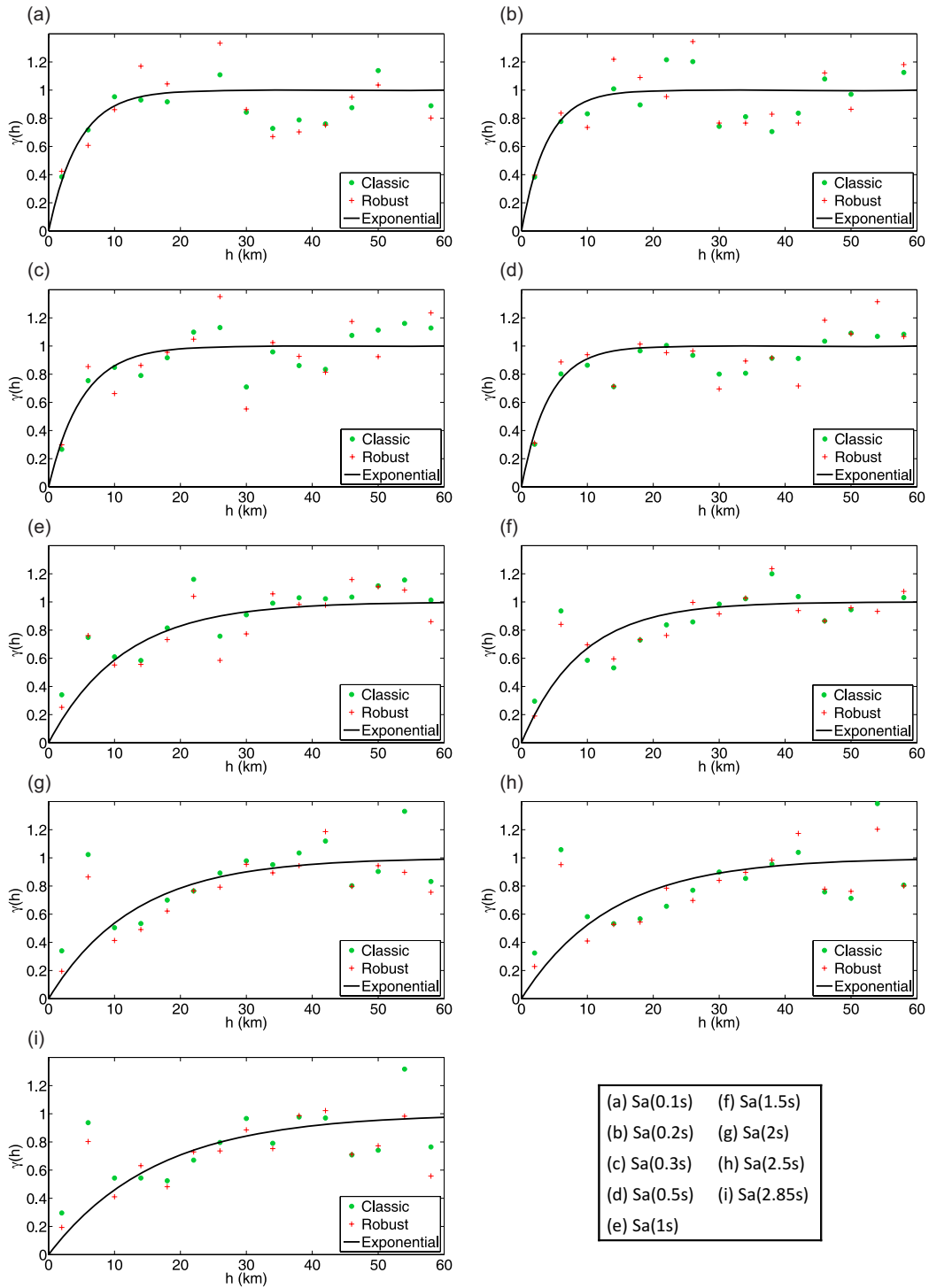


Figure 3.2. Empirical semivariograms and fitted exponential models.

3.3. Predictive equation

Empirical results demonstrate that correlation length tends to increase with period. Except for high frequencies at which there is no clear trend. This seems to be consistent with past studies of ground motion coherency (Zerva and Zervas, 2002). In fact, the coherency describes the degree of correlation between amplitudes and phases angles of two time histories at each of their component frequencies.

Considering that coherency decreases with increasing distance between measuring points and with increasing frequency, it may be reasonable to expect more coherent ground motion, as spectral acceleration evaluated at longer periods exhibits more correlated peak amplitudes. This same issue aspect was also discussed in Jayaram and Baker (2009) where, generally, the estimated ranges increased with period.

Table 3.1. Estimated ranges for the ESD dataset.

Database	Period [s]	Range [km]
ESD	0.1	13.7
	0.2	11.6
	0.3	15.3
	0.5	12.5
	1	33.9
	1.5	27.0
	2	39.0
	2.5	40.5
	2.85	48.8

A simple linear predictive model was fitted³ using LSM in order to capture the trend of the range as a function of structural period, as expressed in Eqn. 3.1 and shown in Fig. 3.3a. Thus, based on the results provided, the correlation between normalized intraevent residuals separated by h may be obtained as $\rho(h, T) = \exp(-3 \cdot h/b(T))$.

$$b(T) = 11.7 + 12.7 \cdot T \quad (3.1)$$

Finally, to compare results shown herein with existing studies, in Fig. 3.3b estimated ranges are compared with some correlation lengths available in literature; those models were chosen considering: (i) Californian dataset from Goda and Hong (2008), (ii) *all earthquakes* model from Hong et al. (2009), and (iii) the *predictive model* based on all earthquakes from Jayaram and Baker (2009a). Results of this study seem to be comparable with ranges estimated in literature in terms of both trend as a function of T , and the estimated values of correlation length. It is worth to mention, for completeness, that an exception in this respect is represented by the Goda and Atkinson (2009, 2010) models, in which ranges are larger (never below about 60 km) and the dependence of the correlation on period is not significant.

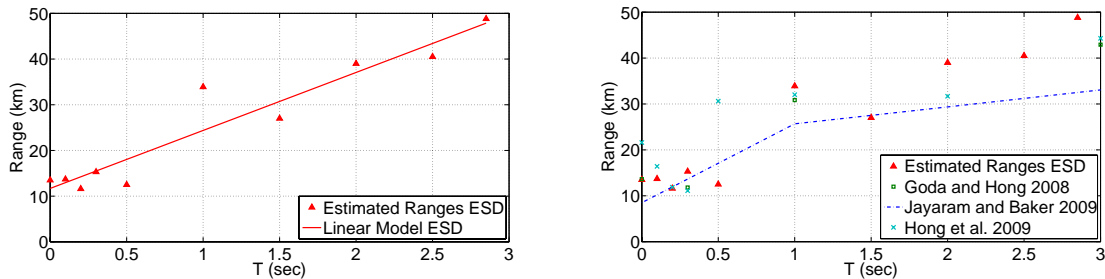


Figure 3.3 (a) Linear model and estimated; (b) comparison with available models from literature.

³ Note that for the fitting of the predictive model the range obtained in Esposito and Iervolino (2011) for PGA has been included. Therefore the estimation was performed based on ten points.

4. CONCLUSIONS

The study presented focus on the assessment of intraevent spatial correlation of elastic spectral acceleration at periods ranging between 0.1 s and 2.85 s. A subset of the European Strong Motion Database was used to compute residuals starting from a GMPE calibrated on the same data.

Consistent with the available literature on the topic, hypotheses of stationarity and isotropy of the random fields were retained to compute experimental semivariograms of standardized intraevent residuals (with respect to the standard deviation estimated by the GMPE).

Because a relatively small number of records for each earthquake was available, records from multiple events and regions were pooled to develop models based on large numbers of observations.

Exponential correlation models were calibrated on empirical data findings. The choice of using the same functional form (exponential) for all periods allowed to compare results and to investigate possible dependency of the parameters on the period the spectral ordinate refers to. In fact, it was found that, as expected to some extent, practical ranges tend to increase as structural period increase, and simple linear functions were fitted to capture this trend.

Results were also compared with results from main references about the same topic finding, generally, consistency.

Starting from the correlation models provided, it is possible to get the joint probability density function for $Sa(T)$ at all locations in a region of interest, for earthquake engineering and seismic risk assessment applications.

ACKNOWLEDGEMENT

This study was supported by AMRA scarl (<http://www.amracenter.com>) under the frame of SYNER-G (7th framework programme of the European Community for research, technological development and demonstration activities; project contract no. 244061). Authors want to thank Sinan Akkar (Middle East Technical University, Turkey) for kindly providing us with datasets used in this study.

REFERENCES

- Akkar, S., and Bommer, J. J. (2010). Empirical Equations for the Prediction of PGA, PGV and Spectral Accelerations in Europe, the Mediterranean Region and the Middle East. *Seismological Research Letters* **81**:2, 195 – 206.
- Barnes, R. J. (1991). The Variogram Sill and the Sample Variance. *Mathematical Geology* **23**:4, 673-678.
- Boore, D. M., Gibbs, J. F., Joyner, W. B., Tinsley, J. C., and Ponti, D.J. (2003). Estimated ground motion from the 1994 Northridge, California, earthquake at the site of the Interstate 10 and La Cienega Boulevard bridge collapse, West Los Angeles, California. *Bulletin of Seismological Society of America* **93**:6, 2737-2751.
- Cressie, N. (1993). Statistics for Spatial Data. Revised edition. Wiley, New York.
- Cressie, N., and Hawkins, D. M. (1980). Robust estimation of variogram. *Mathematical Geology* **12**:2, 115-125.
- Esposito, S. and Iervolino, I. (2011). PGA and PGV spatial correlation models based on European multievent datasets. *Bulletin of Seismological Society of America* **101**:5, 2532-2541.
- Esposito, S. (2011). Systemic seismic risk analysis of gas distribution networks. PhD Thesis, University of Naples Federico II, Naples, Italy. Advisor: I. Iervolino. Available at: wpagina.unina.it/iuniervo
- Goda, K., and Hong, H. P. (2008). Spatial correlation of peak ground motions and response spectra. *Bulletin of Seismological Society of America* **98**:1, 354-365.
- Goda, K., and Atkinson, G. M. (2009). Probabilistic characterization of spatially correlated response spectra for earthquakes in Japan. *Bulletin of Seismological Society of America* **99**:5, 3003-3020.
- Goda, K., and Atkinson, G. M. (2010). Intraevent spatial correlation of ground-motion parameters using SK-net data. *Bulletin of Seismological Society of America* **100**:6, 3055-3067.
- Goovartes, P. (1997). Geostatistics for Natural Resources Evaluation. Oxford University Press, New York.
- Hong, H.P., Zhang, Y., and Goda, K. (2009). Effect of spatial correlation on estimated ground motion prediction equations. *Bulletin of Seismological Society of America* **99**:2A, 928-934.
- Jayaram, N., and Baker, J.W. (2008). Statistical tests of the joint distribution of spectral acceleration values. *Bulletin of Seismological Society of America* **98**:5, 2231-2243.

- Jayaram, N., and Baker, J.W. (2009). Correlation model for spatially distributed ground-motion intensities. *Earthquake Engineering and Structural Dynamics* **38:15**, 1687–1708.
- Jayaram, N., and Baker, J.W. (2010). Considering spatial correlation in mixed-effects regression and impact on ground-motion models. *Bulletin of Seismological Society of America* **100:6**, 3295-3303.
- Journel, A.G., and Huijbregts, C. J. (1978). Mining Geostatistics. Academic Press, London..
- Malhotra, P. (2008). Seismic design loads from site-specific and aggregate hazard analyses. *Bulletin of Seismological Society of America* **98:4**, 1849-1862.
- Matheron, G. (1962). *Traité de géostatistique appliquée*. Editions Technip, Paris, France (in French).
- McGuire, R. K. (2004). Seismic Hazard and Risk Analysis. Earthquake Engineering Research Institute, MNO-10, Oakland, California.
- Sokolov, V., Wenzel, F., Jean, WY., and Wen, K. L. (2010). Uncertainty and spatial correlation of earthquake ground motion in Taiwan. *Terrestrial Atmospheric and Oceanic Sciences* **21:6**, 905-921.
- Wang, M and Takada, T. (2005). Macrospatial correlation model of seismic ground motions. *Earthquake Spectra* **21:4**, 1137-1156.
- Zerva, A. and Zervas , V. (2002). Spatial variation of seismic ground motion, *Applied Mechanics Reviewers* **55:3**, 271-297.