Author manuscript, published in "Bulletin of Earthquake Engineering 6, 3 (2008) 389-405"
DOI : 10.1007/s10518-008-9063-8

Investigating strong ground motion variability using analysis of variance and two-way-fit plots

John Douglas & Pierre Gehl
BRGM — ARN/RIS, 3 avenue C. Guillemin, BP 36009,
45060 Orléans Cedex 2, France.

June 11, 2007

Abstract

A simple method to quantitatively assess the relative importance of unmodelled site and source effects on the observed variation in ground motions is presented. The method consists of analysis of variance (ANOVA) using the computed residuals with respect to an empirical ground-motion model for strong-motion records of various earthquakes recorded at a common set of stations. ANOVA divides the overall variance into the components due to site and source effects not modelled by the ground-motion model plus the residual variance not explained by these factors. To test this procedure, four sets of observed strong-motion records: two from Italy (Umbria-Marche and Molise), one from the French Antilles and one from Turkey, are used. It is found that for the data from Italy the vast majority of the observed variance is attributable to unmodelled site effects. In contrast, the variation in ground motions in the French Antilles and Turkey data is largely attributable, especially at short periods, to source effects not modelled by the ground-motion prediction equations used.

Key words: strong-motion data, ground-motion prediction equations (GMPEs), analysis of variance, site effects, source effects, two-way-fit plots

1 Introduction

Analysis of variance (ANOVA) is a powerful technique developed by R.A. Fisher (e.g. Fisher, 1990) in which the total variation within a set of observations is separated into components associated with possible sources of variability (e.g. Moroney, 1990). It is commonly employed when controlled experiments are conducted, such as those undertaken in agriculture and it is difficult to apply when controlled experiments cannot usually be conducted (as in engineering seismology). In an earlier (but different) application of ANOVA in engineering seismology, Douglas (2004a,b) uses this procedure to investigate possible regional dependence of strong ground motions between five regions of Europe and between Europe, California and New Zealand. Although ANOVA is a well-established technique in other domains, its use in engineering seismology is still limited.

ANOVA provides a simple method for investigating the relative contributions of site and source effects to the overall variation in earthquake ground motions. In this technique the contributions to the observed variability in ground motions can be separated into the variability due to the source, that due to the site and that with an unexplained cause, which could be mainly attributable to path effects. Since strong-motion records are associated with a variety of magnitudes, style-of-faulting, source-to-site distances and site classes, whose effects on ground-motion variation are already approximately known, the residuals with respect to ground-motion prediction equations (GMPEs) (e.g. Douglas, 2003) are examined here. The residuals are computed based on the logarithms of observed and predicted ground motions. This procedure approximately removes the first-order effects of magnitude, style-of-faulting, distance and site classification and therefore the variations in ground motions not modelled by the GMPEs are determined. The use of GMPEs is required because the ground-motion data within each group come from earthquakes with differing magnitudes and focal mechanisms and were recorded at various source-to-site distances and at stations of different site classes. Ideally an underlying model would not be required but currently sufficient observations are not available to dispense with the GMPEs. The method is applied here for peak ground acceleration (PGA) and elastic response spectral acceleration for 5% damping (SA) at 0.2, 0.5, 1.0 and 2.0 s, for the horizontal component definitions used by the underlying GMPEs.

Since the landmark study of Joyner and Boore (1981) it has become standard practice when deriving GMPEs to separate the aleatoric variabilities (usually known as σ) into their inter-event and intra-event (plus, sometimes, the inter-site) components. The consideration of these different components is important in deriving unbiased coefficients, due to non-independence of records from a given earthquake or station, and also it is necessary when using the models in certain applications (e.g. Bommer and Crowley, 2006). In addition, it is useful to understand the sources of the variability so that efforts to reduce the large scatter in GMPEs can be prioritised, e.g. large intra-event σs means that improvements in the modelling of local site effects within GMPEs could lead to a significant reduction in the overall aleatoric variability. This improvement in the understanding of the sources of ground-motion variability is the aim of this article.

The technique presented here complements the procedures of Lee et al. (1998) and Chen and Tsai (2002), which have similar aims. However, both these methods rely on large well-distributed datasets in order to obtain robust results. In particular, the method of Chen and Tsai (2002) derives GMPEs, requires the existence of many dozens of records from individual stations and earthquakes. In contrast the procedure proposed here needs significantly less data.

In this article, the proposed technique is applied to four sets of strong-motion data. The first consists of data from five stations that recorded four earthquakes (20 records in total) of the Umbria-Marche (central Italy) 1997–1998 sequence. The second set is from five stations and five earthquakes (25 records in total) within the Kocaeli (Turkey) 1999 sequence. The third set of data comes from four stations that recorded four earthquakes (16 records in total) of the Molise (southern Italy) 2002–2003 sequence. The final set comprises records from six stations that recorded six earthquakes (36 records in total) within the Les Saintes (Guadeloupe, French Antilles) 2004–2005 sequence. The next section introduces the proposed method, using the data from Umbria-Marche as an example. Then the different sets of data and the results obtained are discussed, in turn. The article ends with some discussion and conclusions.

2 Proposed method

Analysis of variance is a useful tool that helps the user to identify sources of variability from one or more potential sources, called 'effects' or 'factors'. However, the application of this simple method has the following prerequisites:

- the population from which the data are obtained must be normally or approximately normally distributed [this has been demonstrated many times using residuals from logarithmically-transformed ground motions (e.g. Bommer et al., 2004)];
- the samples must be random samples of the population [this is satisfied since no preliminary selection of data was performed];
- the variances of the populations must be equal [this is fulfilled since ground motions from different stations and earthquakes are approximately equally scattered (e.g. Atkinson, 2006)];
- the groups must have the same sample size. This is difficult to fulfill because it requires a sequence of earthquakes all recorded by the same stations. This constraint explains the low number of records studied in each set of ground-motion data.

It is important to keep in mind that values studied here are not PGAs nor SAs, but the residuals between predicted and measured logarithms of PGAs and SAs. ANOVA enables the separation of the variability in the residual acceleration into two causes: unmodelled site and source effects. The significance of the contributions of the unmodelled factors can be assessed by computing the ratio between the σ^2 values for each of these effects to the residual σ^2 value and comparing this to the F value for the degrees of freedom and the significance level considered (e.g. Moroney, 1990).

In order to avoid the bias introduced by a particular GMPE, the analysis is repeated for the various models, listed in Table 1. The GMPEs derived using data from broad regions were used for all sequences whereas the local models were only used for the sequence corresponding to their region, i.e.: Zonno and Montaldo (2002) and Bindi et al. (2006) are used only for the Umbria-Marche sequence; Kalkan and Gülkan (2004), Özbey et al. (2004) and Ulusay et al. (2004) are used only for the Kocaeli sequence; and Luzi et al. (2006) is used only for the Molise sequence. Douglas (2007) argues that average ground motions for the same magnitude and source-to-site distance do not show clear evidence for regional variation. In many parts of the world where observational data is limited, it is more defensible to use well-constrained ground-motion models developed using data from other regions than to base ground-motion estimates on local models, which are often less robust.

[Table 1 about here.]

3 The 1997–1998 Umbria-Marche sequence

These data are a suitable choice to apply the proposed method since there are numerous earthquakes of similar magnitudes recorded at a common set of stations for which site classifications are known. The data and associated parameters used are those previously employed by Ambraseys et al. (2005) to derive ground-motion estimation equations for PGA and SA. Table 2 summarises the data selected and shows one example of the ANOVA technique using the GMPEs of Ambraseys et al. (2005).

[Table 2 about here.]

Table 3 summarises the results of ANOVA for these data for PGA and SA at 0.2 s and at 0.5 s for the different GMPEs selected. Due to long-period noise in some of the records, ANOVA could only be performed up to a period of 0.5 s (see Ambraseys et al. (2005) for details of the record processing procedure applied). All records are required at each period in order to be able to apply the proposed method. Akkar and Bommer (2007) argue that the processing method of Ambraseys et al. (2005) is too conservative hence it may be possible to extend the ANOVA analysis to periods longer than 0.5 s.

[Table 3 about here.]

The ratios of the variances are reported in Table 3 along with the significance level of the effect (in bold if significant) showing that unmodelled site effects are highly significant. The selected set of records contains data from Nocera Umbra (NCR), Gubbio-Piana (GBP) and Rieti (RTI), which were shown by Ambraseys et al. (2005) to display large site-specific amplifications that are poorly modelled by their ground-motion model. Nocera Umbra is located near a sub-vertical fault with highly fractured rocks that amplify high frequency motions (e.g. Marra et al., 2000) and Gubbio-Piana and Rieti stations are located in sedimentary basins that generate large-amplitude surface waves (e.g. Castro et al., 2004). Therefore the conclusion reached here that site effects are important for these data is unsurprising. Interestingly, the analysis shows that unmodelled source effects are not significant for these data meaning that source parameters in addition to those already present (magnitude and style-of-faulting) will not significantly reduce the standard deviation of the ground-motion model.

Figure 1 represents a two-way-fit plot (Tukey, 1972) of the computed residuals (discrepancy between predicted and observed ground motion) for all combinations of earthquake (descending lines) and station (ascending lines): the residual is shown by the vertical coordinate (logarithm of measured acceleration minus logarithm of predicted acceleration). The method to construct this plot is explained by Tukey (1972). For instance, the intersection of the NCR station line and the earthquake A line gives the approximate residual of the difference between the observed and the predicted logarithm of the parameter for Nocera Umbra for the earthquake on 26th September 1997 at 00:33. Note that the vertical coordinate of the intersections do not give the exact residuals since a two-way linear fit does not exactly describe the residuals: according to the plot, the PGA residual that results from the 26th September 1997 Nocera Umbra record is equal to 0.5721, while it is actually 0.6640 (see Table 2). Yet, this graphical method is a good way to show strong tendencies in the distribution of residuals.

[Figure 1 about here.]

4 The 1999 Kocaeli sequence

Following the 17th August 1999 Kocaeli (M_w7.6 HRV CMT) earthquake, many aftershock records were obtained. The data used here were recorded at considerable source-to-site distances by stations of Kandilli Observatory and from moderate earthquakes. Table 4 details the records analysed.

[Table 4 about here.]

Table 5 presents the results obtained using several GMPEs. The results show that both unmodelled source and site effects are very significant for these data. The observed high ratios result from low unexplained residuals, meaning that other effects have very little impact on the accuracy of predicted ground motion. Even if both effects are significant, it is noticeable that source effects are more important than site effects at short periods. Baturay and Stewart (2003) show that for soft soil sites (like at least two of the sites considered here) individual site response analysis can be particularly beneficial for the reduction of ground-motion prediction uncertainties. The analysis of variance results confirms the trend suggested by the two-way-fit plots shown in Figure 2: all lines, whether they represent individual stations or earthquakes, show wide dispersion.

In order to test the sensitivity of the results obtained on the selection of the ground-motion model used to compute the residuals, the analysis was repeated using the following regional models: Ulusay et al. (2004), Kalkan and Gülkan (2004) and Özbey et al. (2004). Although the values of σ^2 and F varied slightly when different models were used, overall the results were similar and the same significance levels were obtained (see Table 5).

[Table 5 about here.]

[Figure 2 about here.]

5 The 2002–2003 Molise sequence

These data from the Molise earthquake sequence of 2002–2003 were generated by a sequence of several earthquakes of similar magnitude, whose fault rupture mechanisms were mostly strike-slip. Table 6 summarises the data used for the analysis.

[Table 6 about here.]

Like the Umbria-Marche sequence, the Molise sequence is located within Italy. A regional model from Luzi et al. (2006) has been added to the list of GMPEs in order to test the effect of including a regional model. Table 7 contains the results of ANOVA for these data.

[Table 7 about here.]

As for the Umbria-Marche sequence, the results show that insufficiently described site effects are mainly responsible for the discrepancy between predicted and measured data. The F test reveals that unmodelled site effects are significant (at less than 5%), whereas unmodelled source effects do not contribute significantly to the unmodelled variation in ground motions.

This means that additional source parameters would not significantly reduce the observed scatter. On the contrary, more efforts should be focused on site effect modelling. The Molise aftershocks used in this study are all close in magnitude and faulting mechanism (strike-slip), which could account for the absence of significant unmodelled source effects. Figure 3 shows that the average residuals with respect to earthquakes are similar while the average residuals for each station are much more scattered.

[Figure 3 about here.]

6 The 2004–2005 Les Saintes sequence

Similarly to the other events studied here, the Les Saintes sequence consisted of a series of shocks of similar magnitudes recorded at comparable distances by a set of common stations. The data and associated parameters were assessed by Douglas et al. (2006). Table 8 summarises the data used.

[Table 8 about here.]

Douglas et al. (2006) quantitatively examine the ability of nine recent sets of GMPEs to predict ground motions from shallow crustal earthquakes (like these events) recorded on the French Antilles. They find that none of the examined models closely predicts the observed ground motions, which are generally of lower amplitude and are more variable than predicted by the models. Like for the other sequences, several GMPEs are tested here in order to avoid any bias coming from the underlying model. Unlike for the other sequences, there are no peer-reviewed GMPEs available derived using data from the French Antilles therefore the effect of using a regional model could not be tested. Table 9 gives the results of ANOVA for the data from the Les Saintes sequence.

[Table 9 about here.]

Unlike for the Umbria-Marche or the Molise data, the Les Saintes results show that unmodelled source effects contribute most to the overall variance and that this effect is highly significant. Douglas et al. (2006) investigate ground motions recorded during two pairs of Les Saintes aftershocks in terms of variabilities in ground motions due to the source. This analysis was performed by calculating the ratios between response spectra at various common stations from two aftershocks after having corrected for minor differences in magnitudes and distances. The ratios therefore show the effect of source variability on ground motions since the site effects have been removed through the computation of the ratio. They found that for one pair of events (those of 21st November 2004 at 13:37 and 18:53) the source variability caused differences in ground motions up to ten times for some periods and some stations. However, for the other pair of events (those of 27th November 2004 at 23:44 and 2nd December at 14:47) the ground-motion variability (and hence source variability) was considerably less. Hence, the highly significant source effects found here confirm this result. For these data, unmodelled site effects are not important at short periods while at long periods they contribute significantly to the overall variability in ground motions. Five out of the six selected stations are classified as being on rock therefore site effects could be expected to be less significant for these data than for the other data sets studied here, where stations are located on more heterogeneous geological formations. In Figure 4, all station lines seem very close except one (GJYA): showing that, in general, site effects are well modelled. On the contrary, there is wide dispersion between the earthquake lines, confirming the important source effects revealed by ANOVA.

[Figure 4 about here.]

7 Conclusions

This article proposed a simple quantitative method to investigate and separate the variability in earthquake ground motions into that attributable to site effects and that due to source effects. The method is based on analysis of variance of the residuals of ground-motion intensity parameters computing using ground-motion models that approximately remove the effects of magnitude, style-of-faulting, source-to-site distance and simple site classification. The technique was then applied to four sets of strong-motion data. It is found that for two of the sets (those from the Umbria-Marche 1997-1998 sequence and the Molise 2002-2003 sequence) unmodelled site effects are much more important than source effects in explaining the observed variability within the residuals. However, for the other sets of examined data (those from the 1999 Kocaeli sequence and the Les Saintes 2004–2005 sequence) unmodelled source effects are the largest contributor to overall variability, confirming the findings of Lee et al. (1998) using a different approach and Californian data. GMPEs produced within the PEER Next Generation Attenuation of Ground Motions (NGA) project, which are currently being finalised, have modelled the effects on ground motions of other source characteristics, than magnitude and style of faulting, in order to reduce the inter-event σs obtained. For example, the GMPEs of Chiou and Youngs (2006) include, in addition to the standard independent parameters, the effects of the dip, width and depth of the rupture plane on ground motions. The need to improve the modelling of different source effects within GMPEs is demonstrated by the results obtained here since unmodelled source effects have been shown to contribute a large proportion of the overall variability. Finally, two-way-fit plots introduced by Tukey (1972) provide a useful way of visually demonstrating the two sources of unmodelled variabilities in ground motions.

8 Acknowledgements

Some of the work presented in this article was funded by the Agence Nationale de la Recherche (ANR) project Quantitative Seismic Hazard Assessment (QSHA). The rest was funded by internal BRGM research projects. Data from the strong-motion networks operated by ENEL (now operated by SSN), ENEA, Kandilli Observatory (Bogazici University), BRGM and RAP-IPGP were used for this study. The strong-motion networks on Guadeloupe are operated by BRGM and the Institut de Physique du Globe de Paris (IPGP), which is under the aegis of the Réseau Accélérometrique Permanent (RAP) of France. The RAP data centre is based at Laboratoire de Géophysique Interne et de Tectonophysique, Grenoble. We are very grateful to the personnel of the organisations that operate the strong-motion networks that provided data for this study, without which this study would have been impossible.

References

N. A. Abrahamson and W. J. Silva. Empirical response spectral attenuation relations for shallow crustal earthquakes. *Seismological Research Letters*, 68(1):94–127, Jan/Feb 1997.

S. Akkar and J. J. Bommer. Empirical prediction equations for peak ground velocity derived from strong-motion records from Europe and the Middle East. *Bulletin of the Seismological Society of America*, 97(2):511–530, Apr 2007. doi: 10.1785/0120060141.

N. N. Ambraseys and M. W. Free. Surface-wave magnitude calibration for European region earthquakes. *Journal of Earthquake Engineering*, 1(1):1–22, 1997.

N. N. Ambraseys, J. Douglas, S. K. Sarma, and P. M. Smit. Equations for the estimation of strong ground motions from shallow crustal earthquakes using data from Europe and the Middle East: Horizontal peak ground acceleration and spectral acceleration. *Bulletin of Earthquake Engineering*, 3(1):1–53, 2005. doi: 10.1007/s10518-005-0183-0.

G. M. Atkinson. Single-station sigma. *Bulletin of the Seismological Society of America*, 96(2): 446–455, Apr 2006. doi: 10.1785/0120050137.

M. B. Baturay and J. P. Stewart. Uncertainty and bias in ground-motion estimates from ground response analyses. *Bulletin of the Seismological Society of America*, 93(5):2025–2042, Oct 2003.

D. Bindi, L. Luzi, F. Pacor, G. Franceshina, and R. R. Castro. Ground-motion predictions from empirical attenuation relationships versus recorded data: The case of the 1997-1998 Umbria-Marche, central Italy, strong-motion data set. *Bulletin of the Seismological Society of America*, 96(3):984–1002, 2006. doi: 10.1785/0120050102.

J. J. Bommer and H. Crowley. The influence of ground-motion variability in earthquake loss modelling. *Bulletin of Earthquake Engineering*, 4(3):231–248, 2006. doi: 10.1007/s10518-006-9008-z.

J. J. Bommer, N. A. Abrahamson, F. O. Strasser, A. Pecker, P.-Y. Bard, H. Bungum, F. Cotton, D. Fäh, F. Sabetta, F. Scherbaum, and J. Studer. The challenge of defining upper bounds on earthquake ground motions. *Seismological Research Letters*, 75(1):82–95, 2004.

D. M. Boore, W. B. Joyner, and T. E. Fumal. Equations for estimating horizontal response spectra and peak acceleration from western North American earthquakes: A summary of recent work. *Seismological Research Letters*, 68(1):128–153, Jan/Feb 1997.

K. W. Campbell and Y. Bozorgnia. Updated near-source ground-motion (attenuation) relations for the horizontal and vertical components of peak ground acceleration and acceleration response spectra. *Bulletin of the Seismological Society of America*, 93(1):314–331, 2003.

S. Castellaro, F. Mulargia, and Y. Y. Kagan. Regression problems for magnitudes. *Geophysical Journal International*, 165(3):913–930, 2006. doi: 10.1111/j.1365-246X.2006.02955.x.

R. R. Castro, F. Pacor, D. Bindi, G. Franceshina, and L. Luzi. Site response of strong ground motion stations in the Umbria, central Italy, region. *Bulletin of the Seismological Society of America*, 94(2):576–590, 2004.

Y.-H. Chen and C.-C. P. Tsai. A new method for estimation of the attenuation relationship with variance components. *Bulletin of the Seismological Society of America*, 92(5):1984–1991, 2002.

(B.) S.-J. Chiou and R. R. Youngs. Chiou and Youngs PEER-NGA empirical ground motion model for the average horizontal component of peak acceleration and pseudo-spectral acceleration for spectral periods of 0.01 to 10 seconds. Interim report for USGS review, Jul 2006. URL `http://peer.berkeley.edu/lifelines/nga_docs/nov_13_06/Chiou_Youngs_NGA_2006.html`.

J. Douglas. Earthquake ground motion estimation using strong-motion records: A review of equations for the estimation of peak ground acceleration and response spectral ordinates. *Earth-Science Reviews*, 61(1–2):43–104, 2003.

J. Douglas. An investigation of analysis of variance as a tool for exploring regional differences in strong ground motions. *Journal of Seismology*, 8(4):485–496, Oct 2004a.

J. Douglas. On the regional dependence of earthquake response spectra. *ISET Journal of Earthquake Technology*, 44(1), 2007. In press.

J. Douglas. Use of analysis of variance for the investigation of regional dependence of strong ground motions. In *Proceedings of Thirteenth World Conference on Earthquake Engineering*, 2004b. Paper no. 29.

J. Douglas, D. Bertil, A. Roullé, P. Dominique, and P. Jousset. A preliminary investigation of strong-motion data from the French Antilles. *Journal of Seismology*, 10(3):271–299, 2006. doi: 10.1007/s10950-006-9016-0.

R. A. Fisher. *Statistical Methods, Experimental Design, and Scientific Inference.* Oxford University Press, Oxford, United Kingdom, 1990.

W. B. Joyner and D. M. Boore. Peak horizontal acceleration and velocity from strong-motion records including records from the 1979 Imperial Valley, California, earthquake. *Bulletin of the Seismological Society of America*, 71(6):2011–2038, Dec 1981.

E. Kalkan and P. Gülkan. Site-dependent spectra derived from ground motion records in Turkey. *Earthquake Spectra*, 20(4):1111–1138, Nov 2004.

Y. Lee, Y. Zeng, and J. G. Anderson. A simple strategy to examine the sources of errors in attenuation relations. *Bulletin of the Seismological Society of America*, 88(1):291–296, Feb 1998.

L. Luzi, P. Morasca, F. Zolezzi, D. Bindi, F. Pacor, D. Spallarossa, and G. Franceschina. Ground motion models for Molise region (southern Italy). In *Proceedings of First European Conference on Earthquake Engineering and Seismology (a joint event of the 13th ECEE & 30th General Assembly of the ESC)*, 2006. Paper number 938.

F. Marra, R. Azzara, F. Bellucci, A. Caserta, G. Cultrera, G. Mele, B. Palombo, A. Rovelli, and E. Boschi. Large amplification of ground motion at rock sites within a fault zone in Nocera Umbra (central Italy). *Journal of Seismology*, 4(4):543–554, 2000.

M. J. Moroney. *Facts from Figures.* Penguin Books, 2nd edition, 1990.

C. Özbey, A. Sari, L. Manuel, M. Erdik, and Y. Fahjan. An empirical attenuation relationship for northwestern Turkey ground motion using a random effects approach. *Soil Dynamics and Earthquake Engineering*, 24(2):115–125, 2004.

K. Sadigh, C.-Y. Chang, J. A. Egan, F. Makdisi, and R. R. Youngs. Attenuation relationships for shallow crustal earthquakes based on California strong motion data. *Seismological Research Letters*, 68(1):180–189, Jan/Feb 1997.

P. Spudich, W. B. Joyner, A. G. Lindh, D. M. Boore, B. M. Margaris, and J. B. Fletcher. SEA99: A revised ground motion prediction relation for use in extensional tectonic regimes. *Bulletin of the Seismological Society of America*, 89(5):1156–1170, Oct 1999.

J. W. Tukey. Some graphic and semigraphic displays. In T. A. Bancroft, editor, *Statistical Papers in Honor of George W. Snedecor*, pages 293–316. The Iowa State University Press, Ames, Iowa, USA, 1972.

R. Ulusay, E. Tuncay, H. Sonmez, and C. Gokceoglu. An attenuation relationship based on Turkish strong motion data and iso-acceleration map of Turkey. *Engineering Geology*, 74 (3-4):265–291, 2004.

G. Zonno and V. Montaldo. Analysis of strong ground motions to evaluate regional attenuation relationships. *Annals of Geophysics*, 45(3–4):439–454, Jun–Aug 2002.

List of Figures

1 Two-way-fit plot (Tukey, 1972) for data from the 1997-1998 Umbria-Marche sequence. The numbers on the ordinate are the approximate residuals with respect to the GMPEs of Ambraseys et al. (2005). 14

2 Two-way-fit plot for data from the 1999 Kocaeli sequence. The numbers on the ordinate are the approximate residuals with respect to the GMPEs of Ambraseys et al. (2005). 15

3 Two-way-fit plot for data from the 2002–2003 Molise sequence. The numbers on the ordinate are the approximate residuals with respect to the GMPEs of Ambraseys et al. (2005). 16

4 Two-way-fit plot for data from the 2004–2005 Les Saintes sequence. The numbers on the ordinate are the approximate residuals with respect to the GMPEs of Ambraseys et al. (2005). 17

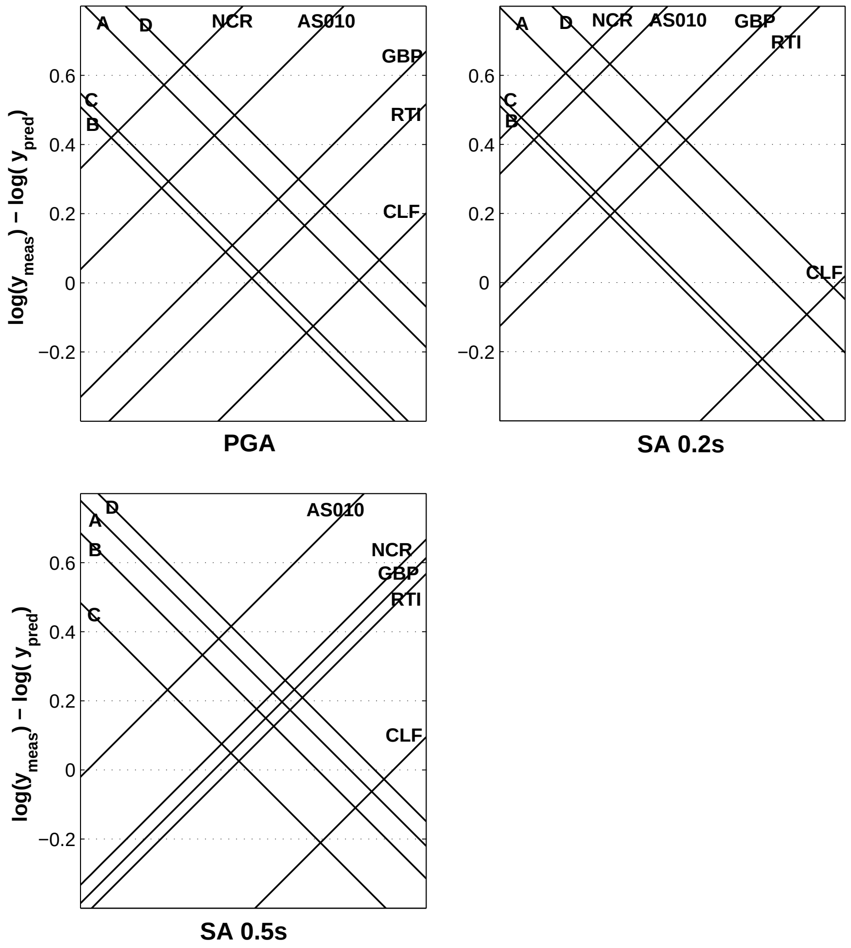
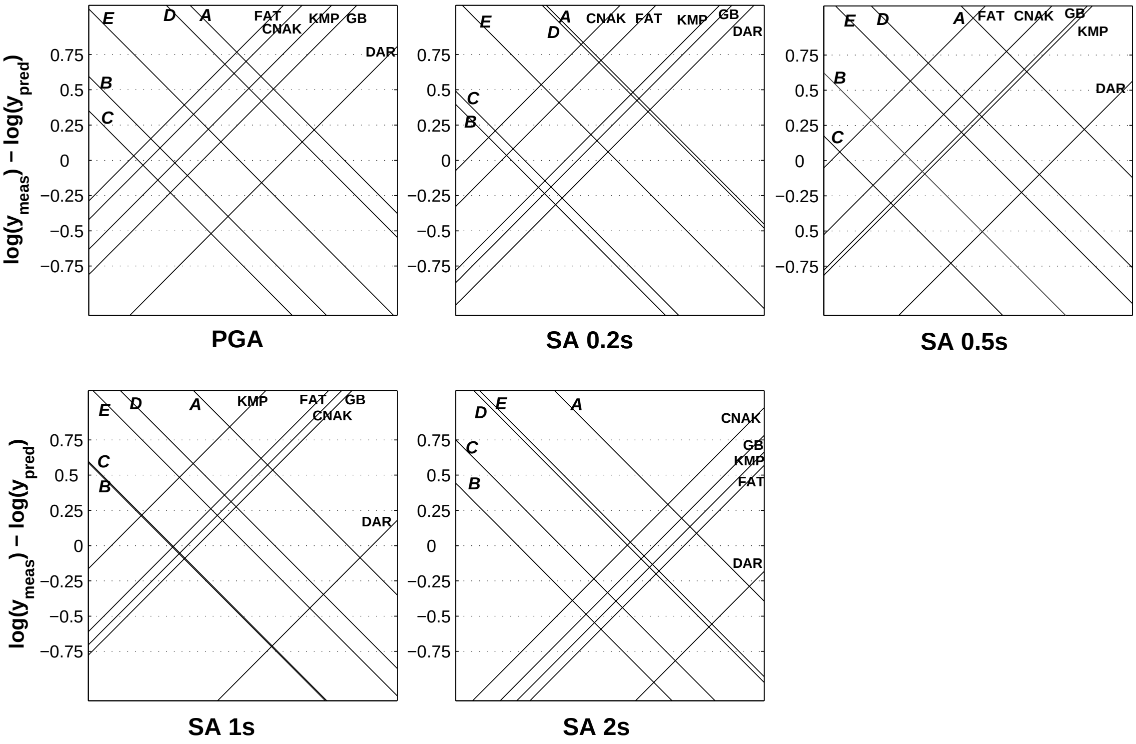


Figure 1: Two-way-fit plot (Tukey, 1972) for data from the 1997-1998 Umbria-Marche sequence. The numbers on the ordinate are the approximate residuals with respect to the GMPEs of Ambraseys et al. (2005).



PGA

SA 0.2s

SA 0.5s

SA 1s

SA 2s

Figure 2: Two-way-fit plot for data from the 1999 Kocaeli sequence. The numbers on the ordinate are the approximate residuals with respect to the GMPEs of Ambraseys et al. (2005).

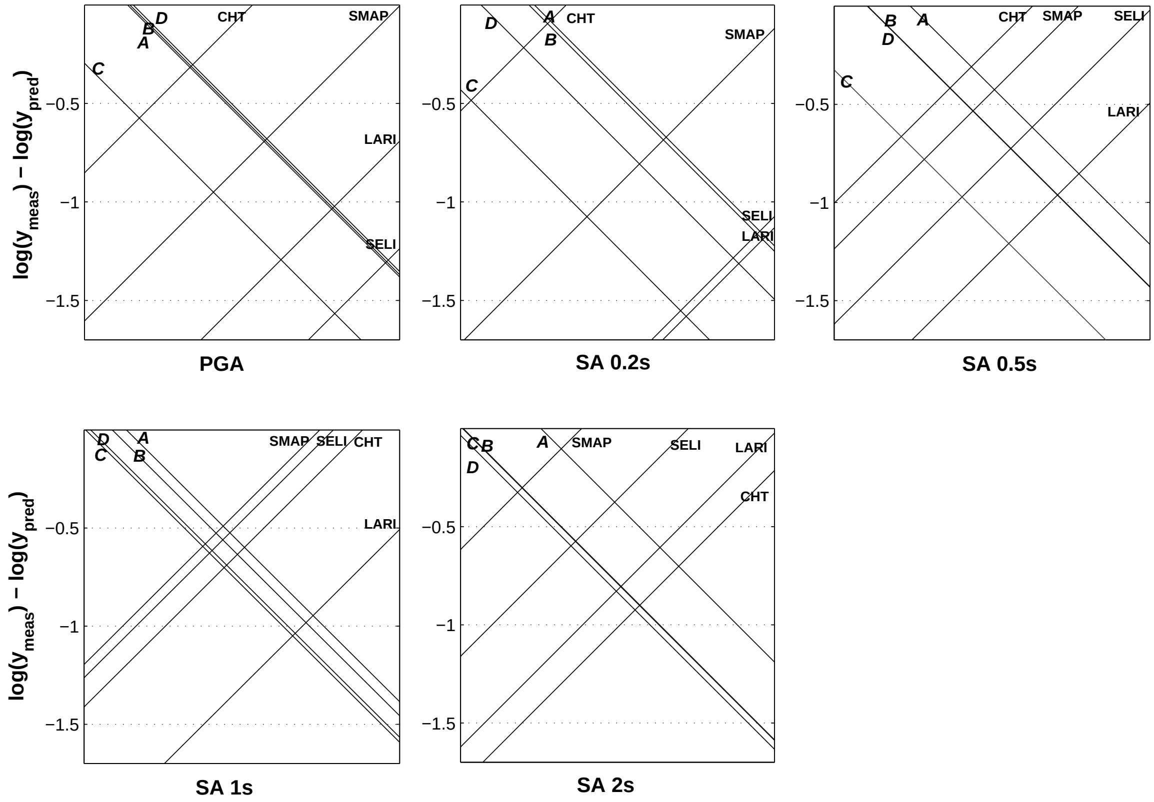


Figure 3: Two-way-fit plot for data from the 2002–2003 Molise sequence. The numbers on the ordinate are the approximate residuals with respect to the GMPEs of Ambraseys et al. (2005).

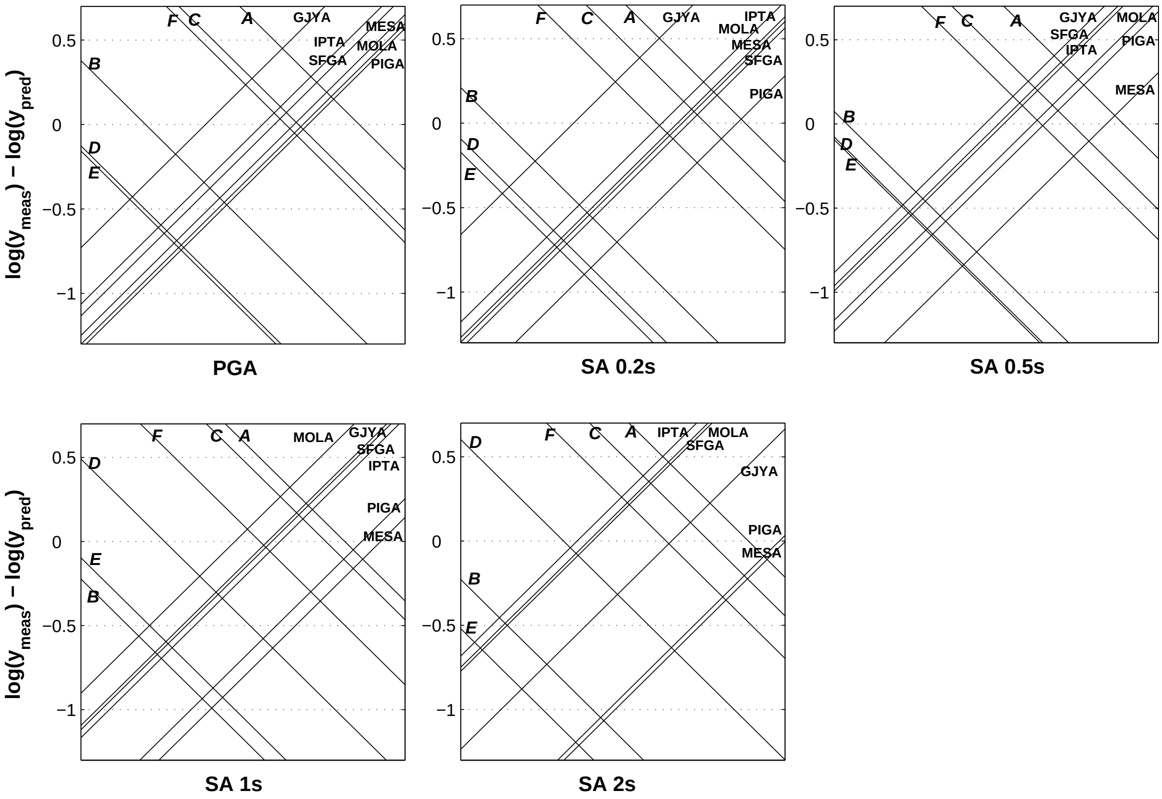


Figure 4: Two-way-fit plot for data from the 2004–2005 Les Saintes sequence. The numbers on the ordinate are the approximate residuals with respect to the GMPEs of Ambraseys et al. (2005).

List of Tables

1 GMPEs selected for this study, the regions used as sources of accelerograms and the magnitude and distance ranges (R_{epi} is epicentral distance, R_{jb} is distance to surface projection of rupture (Joyner and Boore, 1981), R_{hypo} is hypocentral distance, R_{rup} is distance to rupture and R_{seis} is distance to seismogenic rupture). 20

2 Data from the Umbria-Marche 1997-1998 sequence used in this study (distances in italics are R_{jb}). $\log(\mathrm{PGA}_{\mathrm{meas}}) - \log(\mathrm{PGA}_{\mathrm{pred}})$ is the residual with respect to the GMPE of Ambraseys et al. (2005), i.e. logarithm of measured PGA minus logarithm of predicted PGA. m_e and m_s respectively represent the mean value of the residuals for each earthquake and for each station. 21

3 Summary of the results for the Umbria-Marche sequence for the different GMPEs. The numbers give the ratio of inter-event or inter-site σ^2 to the residual σ^2_R. A bold number means the effect is significant at 0.1% or less using the F test. $\mathrm{df}_{\mathrm{S}} = 4$, $\mathrm{df}_{\mathrm{E}} = 3$, $\mathrm{df}_{\mathrm{R}} = 12$ and $\mathrm{df}_{\mathrm{T}} = 19$. 22

4 Data from the 1999 Kocaeli sequence. See caption of Table 2 for abbreviations used. * indicates that surface-wave magnitude (M_s) was converted to M_w using Equation 6.2 of Ambraseys and Free (1997). 23

5 Summary of the results for the Kocaeli sequence obtained with several models. The numbers give the ratio of inter-event or inter-site σ^2 on the residual σ^2. A bold number means the effect is significant at 0.005% or less, using the F test. The GMPE of Ulusay et al. (2004) predicts only PGA. $\mathrm{df}_{\mathrm{S}} = 4$, $\mathrm{df}_{\mathrm{E}} = 4$, $\mathrm{df}_{\mathrm{R}} = 16$ and $\mathrm{df}_{\mathrm{T}} = 24$. 24

6 Data from the Molise 2002–2003 sequence. See caption of Table 2 for abbreviations used. * indicates that body-wave magnitude (m_b) was converted to M_w using equation of Castellaro et al. (2006). 25

7 Summary of the results for the Molise sequence obtained with several models. The numbers give the ratio of inter-event or inter-site σ^2 on the residual σ^2. A bold number means the effect is significant at 5.0% or less, using the F test. The GMPE of Luzi et al. (2006) only predicts PGA. $\mathrm{df}_{\mathrm{S}} = 3$, $\mathrm{df}_{\mathrm{E}} = 3$, $\mathrm{df}_{\mathrm{R}} = 9$ and $\mathrm{df}_{\mathrm{T}} = 15$. 26

8 Data from the Les Saintes 2004–2005 sequence. See caption of Table 2 for abbreviations used. 27

9 Summary of the results for the Les Saintes sequence obtained with several models. The numbers give the ratio of inter-event or inter-site σ^2 on the residual σ^2. A bold number means the effect is significant at 0.001% or less, using the F test. $\mathrm{df}_\mathrm{S} = 5$, $\mathrm{df}_\mathrm{E} = 5$, $\mathrm{df}_\mathrm{R} = 25$ and $\mathrm{df}_\mathrm{T} = 35$. 28

Table 1: GMPEs selected for this study, the regions used as sources of accelerograms and the magnitude and distance ranges (R_{epi} is epicentral distance, R_{jb} is distance to surface projection of rupture (Joyner and Boore, 1981), R_{hypo} is hypocentral distance, R_{rup} is distance to rupture and R_{seis} is distance to seismogenic rupture).

Reference	Region	M range	d range (km)
	Small regions		
Bindi et al. (2006)	Umbria-Marche	$4.0 \leq M_L \leq 5.9$	$1 \leq R_{\text{epi}} \leq 100$
Kalkan and Gülkan (2004)	Mainly NW Turkey	$4.0 \leq M_w \leq 7.4$	$1 \leq R_{\text{jb}} \leq 250$
Luzi et al. (2006)	Molise	$2.6 \leq M_L \leq 5.7$	$5 \leq R_{\text{hypo}} \leq 55$
Özbey et al. (2004)	NW Turkey	$5.0 \leq M_w \leq 7.4$	$5 \leq R_{\text{jb}} \leq 300$
Zonno and Montaldo (2002)	Umbria-Marche	$4.5 \leq M_L \leq 5.9$	$2 \leq R_{\text{epi}} \leq 100$
Ulusay et al. (2004)	Mainly NW Turkey	$4.1 \leq M_L \leq 7.5$	$5.1 \leq R_{\text{epi}} \leq 99.7$
	Broad regions		
Abrahamson and Silva (1997)	Mainly California	$4.4 \leq M_w \leq 7.4$	$0 \leq R_{\text{rup}} \leq 220$
Ambraseys et al. (2005)	Europe & Middle East	$5.0 \leq M_w \leq 7.6$	$0 \leq R_{\text{jb}} \leq 99$
Boore et al. (1997)	Mainly California	$5.1 \leq M_w \leq 7.7$	$0 \leq R_{\text{jb}} \leq 118$
Campbell and Bozorgnia (2003)	Mainly California	$4.7 \leq M_w \leq 7.7$	$2 \leq R_{\text{seis}} \leq 60$
Sadigh et al. (1997)	Mainly California	$3.8 \leq M_w \leq 7.4$	$0 \leq R_{\text{rup}} \leq 305$
Spudich et al. (1999)	Worldwide extensional regimes	$5.1 \leq M_w \leq 7.2$	$0 \leq R_{\text{jb}} \leq 99$

Table 2: Data from the Umbria-Marche 1997-1998 sequence used in this study (distances in italics are R_{jb}). $\log(\text{PGA}_{\text{meas}}) - \log(\text{PGA}_{\text{pred}})$ is the residual with respect to the GMPE of Ambraseys et al. (2005), i.e. logarithm of measured PGA minus logarithm of predicted PGA. m_e and m_s respectively represent the mean value of the residuals for each earthquake and for each station.

			A	B	C	D		
DD/MM/YY			26/09/97	26/09/97	03/10/97	06/10/97		
HH:MM			00:33	09:40	08:55	23:24		
M_w			5.7	6.0	5.3	5.5		
Station	Code	Site class	R_{epi}					
			$\log(\text{PGA}_{\text{meas}}) - \log(\text{PGA}_{\text{pred}})$				Total	m_s
Assisi Stallone	AS010	Rock	*21*	*14*	19	20		
			0.4325	0.2396	0.2607	0.5445	1.4773	0.3693
Colfiorito	CLF	Stiff soil	*0*	*5*	7	7		
			0.0772	-0.0508	-0.1019	-0.1210	-0.1965	-0.0491
Gubbio Piana	GBP	Soft soil	*38*	*27*	37	38		
			-0.0327	0.1369	0.2616	0.3729	0.7387	0.1847
Nocera Umbra	NCR	Rock	*11*	*1*	10	11		
			0.6640	0.2733	0.4305	0.6933	2.0611	0.5153
Rieti	RTI	Very soft soil	*61*	*66*	67	65		
			0.2713	0.0511	-0.1025	0.2143	0.4342	0.1086
		Total	1.4123	0.6501	0.7484	1.7040	4.5148	
		m_e	0.2825	0.1300	0.1497	0.3408		0.2257

The total sum of squares SS_T is evaluated by adding the squares of all the elements and subtracting the correction factor (square of the sum of elements divided by the number of elements). SS_E and SS_S are calculated the same way, except the squares of elements are not added but the squares of columns or rows. The interaction or residual sum of squares can be deduced from these values: $\text{SS}_R = \text{SS}_T - \text{SS}_E - \text{SS}_S$.

The degrees of freedom, df, are based on the number of elements:

- $\text{df}_\text{E} = N_{\text{Earthquakes}} - 1$;
- $\text{df}_\text{S} = N_{\text{Sites}} - 1$;
- $\text{df}_\text{R} = \text{df}_\text{E}\text{df}_\text{S}$;
- $\text{df}_\text{T} = \text{df}_\text{E} + \text{df}_\text{S} + \text{df}_\text{R}$.

The mean square values σ^2 are computed by dividing the sum of squares by the corresponding degree of freedom. Finally, the ratios of inter-event σ_E^2 to residual σ_R^2 and intra-event σ_S^2 to residual σ_R^2 are obtained. As these ratios are F-distributed with degrees of freedom df_E and df_R for inter-event ratio and df_S and df_R for intra-event ratio, they can be compared to threshold values for the F-test and the significance level of each effect can be evaluated.

Source	Sum of squares		Degrees of freedom		Mean square		Ratio	Significance (F-test)
Sites	SS_S	0.7817	df_S	4	σ_S^2	0.1954	10.4	***
Earthquakes	SS_E	0.1570	df_E	3	σ_E^2	0.0314	1.7	
Residual (Interaction)	SS_R	0.2262	df_R	12	σ_R^2	0.0188		
Total	SS_T	1.1649	df_T	19				

The three asterisks in the last column for sites means that the site effect is significant at 0.1% or less and no asterisks for earthquakes means that this effect is not significant at 5%.

Table 3: Summary of the results for the Umbria-Marche sequence for the different GMPEs. The numbers give the ratio of inter-event or inter-site σ^2 to the residual σ_R^2. A bold number means the effect is significant at 0.1% or less using the F test. $\mathrm{df_S} = 4$, $\mathrm{df_E} = 3$, $\mathrm{df_R} = 12$ and $\mathrm{df_T} = 19$.

	PGA		SA(0.2 s)		SA(0.5 s)		SA(1.0 s)		SA(2.0 s)	
GMPEs	Source	Site	Source	Site	Source	Site	Source	Site	Source	Site
Ambraseys et al. (2005)	2.8	**10.4**	1.5	8.0	1.3	4.0	3.3	3.1	0.8	6.5
Abrahamson and Silva (1997)	1.9	**14.2**	1.1	**10.7**	0.7	2.8	2.4	8.4	1.3	**17.6**
Boore et al. (1997)	2.0	**19.5**	1.4	**10.7**	1.3	4.8	3.7	9.2	3.1	**20.9**
Campbell and Bozorgnia (2003)	3.1	**22.3**	1.5	**13.1**	1.8	5.6	3.0	3.7	1.3	**11.8**
Sadigh et al. (1997)	2.4	**10.9**	1.1	**9.6**	1.0	2.1	2.8	**9.9**	1.8	**20.7**
Spudich et al. (1999)	1.4	4.0	2.2	6.3	1.2	5.9	2.4	3.3	0.8	8.9
Bindi et al. (2006)	2.1	**49.9**	2.1	**28.9**	0.5	**10.0**	2.2	7.5	0.5	**15.8**
Zonno and Montaldo (2002)	2.3	**22.6**	1.1	**17.4**	0.5	4.3	2.1	6.8	0.9	**19.6**

Table 4: Data from the 1999 Kocaeli sequence. See caption of Table 2 for abbreviations used. * indicates that surface-wave magnitude (M_s) was converted to M_w using Equation 6.2 of Ambraseys and Free (1997).

			A	B	C	D	E
DD/MM/YY			13/09/99	22/08/99	31/08/99	20/09/99	11/11/99
HH:MM			11:55	14:31	08:10	14:41	21:28
M_w			5.8	5.3*	5.1*	5.6	4.8
Station	Code	Site class	R_{epi}				
Cekmece Kucuk	CNAK	Stiff soil	115	168	104	129	107
Aslan Cimento	DAR	Soft soil	61	113	48	74	152
Istanbul K.M. Pasa	KMP	Soft soil	101	154	89	115	120
Sirkeci	GB	Unknown	96	149	85	110	125
Fatih Tomb	FAT	Soft soil	100	153	88	114	122

Table 5: Summary of the results for the Kocaeli sequence obtained with several models. The numbers give the ratio of inter-event or inter-site σ^2 on the residual σ^2. A bold number means the effect is significant at 0.005% or less, using the F test. The GMPE of Ulusay et al. (2004) predicts only PGA. $\mathrm{df_S} = 4$, $\mathrm{df_E} = 4$, $\mathrm{df_R} = 16$ and $\mathrm{df_T} = 24$.

	PGA		SA(0.2 s)		SA(0.5 s)		SA(1.0 s)		SA(2.0 s)	
GMPEs	Source	Site	Source	Site	Source	Site	Source	Site	Source	Site
Ambraseys et al. (2005)	**63.7**	**28.8**	**59.5**	**22.8**	**27.6**	**16.9**	11.7	**20.0**	**21.4**	**15.3**
Abrahamson and Silva (1997)	**85.2**	**35.8**	**68.0**	**21.5**	**31.2**	**20.6**	13.2	**22.3**	**26.5**	**14.3**
Boore et al. (1997)	**102.3**	**23.7**	**76.0**	**17.9**	**43.1**	**19.9**	**21.9**	**21.3**	**47.6**	**13.9**
Campbell and Bozorgnia (2003)	**108.6**	**33.6**	**79.9**	**25.9**	**41.4**	**20.1**	**16.7**	**22.2**	**28.7**	**15.5**
Sadigh et al. (1997)	**70.4**	**35.4**	**52.3**	**22.0**	**28.0**	**20.3**	12.3	**23.7**	**22.7**	**17.3**
Spudich et al. (1999)	**110.4**	**31.4**	**75.4**	**19.5**	**41.0**	**20.5**	**20.4**	**23.1**	**44.9**	**16.3**
Ulusay et al. (2004)	**54.6**	**29.8**								
Kalkan and Gülkan (2004)	**70.2**	**17.3**	**64.6**	**15.1**	**33.7**	**13.9**	**14.9**	**17.4**	**26.1**	8.8
Özbey et al. (2004)	**86.7**	**39.8**	**72.8**	**29.7**	**32.9**	**19.1**	**15.5**	**19.9**	**41.3**	**15.4**

Table 6: Data from the Molise 2002–2003 sequence. See caption of Table 2 for abbreviations used. * indicates that body-wave magnitude (m_b) was converted to M_w using equation of Castellaro et al. (2006).

			A	B	C	D
DD/MM/YY			04/11/02	12/11/02	02/12/02	01/06/03
HH:MM			00:35	09:27	20:52	15:45
M_w			4.3	4.6	3.9*	4.4
Station	Code	Site class	R_{epi}			
Chieti University	CHT	Soft soil	93	91	99	96
San Martino in Pensilis	SMAP	Stiff soil	23	25	22	28
S. Elia a Pianisi	SELI	Stiff soil	10	11	8	6
Larino	LARI	Stiff soil	13	15	13	18

Table 7: Summary of the results for the Molise sequence obtained with several models. The numbers give the ratio of inter-event or inter-site σ^2 on the residual σ^2. A bold number means the effect is significant at 5.0% or less, using the F test. The GMPE of Luzi et al. (2006) only predicts PGA. $\mathrm{df}_\mathrm{S} = 3$, $\mathrm{df}_\mathrm{E} = 3$, $\mathrm{df}_\mathrm{R} = 9$ and $\mathrm{df}_\mathrm{T} = 15$.

	PGA		SA(0.2 s)		SA(0.5 s)		SA(1.0 s)		SA(2.0 s)	
GMPEs	Source	Site	Source	Site	Source	Site	Source	Site	Source	Site
Ambraseys et al. (2005)	2.4	**24.9**	3.7	**28.1**	2.0	**5.0**	0.6	**10.9**	2.1	**13.9**
Abrahamson and Silva (1997)	2.5	**6.9**	**4.4**	**14.2**	1.9	3.2	0.3	6.1	2.5	**6.1**
Boore et al. (1997)	3.5	2.8	1.8	**10.7**	1.6	2.7	0.6	**13.8**	3.7	**19.9**
Campbell and Bozorgnia (2003)	**5.4**	2.0	**9.2**	**8.2**	3.1	2.6	0.4	**8.7**	3.3	**10.8**
Sadigh et al. (1997)	3.5	**10.9**	**5.8**	**21.0**	1.7	**5.7**	0.3	**8.7**	3.1	**6.1**
Spudich et al. (1999)	3.3	**9.6**	1.7	**16.5**	1.7	3.7	0.8	**11.9**	3.1	**11.8**
Luzi et al. (2006)	1.0	2.7								

Table 8: Data from the Les Saintes 2004–2005 sequence. See caption of Table 2 for abbreviations used.

			A	B	C	D	E	F
DD/MM/YY			21/11/04	21/11/04	21/11/04	27/11/04	02/12/04	14/02/05
HH:MM			11:41	13:37	18:53	23:44	14:47	18:05
M_w			6.3	5.3	5.4	4.9	5.0	5.8
Station	Code	Site class	R_{epi}					
Saint Claude Belfond	GJYA	Rock	34	30	24	39	35	26
Ecole Pigeon	PIGA	Rock	50	46	41	55	51	42
Institut Pasteur Abymes	IPTA	Rock	51	48	45	58	53	47
Stade Morne á l'eau	MESA	Soft soil	62	59	57	69	65	59
Radar Meteo-France	MOLA	Rock	62	60	59	70	65	61
Saint Francois	SFGA	Rock	62	61	62	69	65	63

Table 9: Summary of the results for the Les Saintes sequence obtained with several models. The numbers give the ratio of inter-event or inter-site σ^2 on the residual σ^2. A bold number means the effect is significant at 0.001% or less, using the F test. $\mathrm{df_S} = 5$, $\mathrm{df_E} = 5$, $\mathrm{df_R} = 25$ and $\mathrm{df_T} = 35$.

	PGA		SA(0.2 s)		SA(0.5 s)		SA(1.0 s)		SA(2.0 s)	
GMPEs	Source	Site	Source	Site	Source	Site	Source	Site	Source	Site
Ambraseys et al. (2005)	**22.1**	2.0	**24.6**	3.8	**28.2**	3.8	**42.0**	10.9	**74.6**	**34.7**
Abrahamson and Silva (1997)	**17.3**	8.6	**17.7**	9.8	**17.9**	5.0	**29.8**	6.9	**64.9**	**19.5**
Boore et al. (1997)	**40.6**	5.4	**30.1**	6.0	**37.3**	2.0	**69.7**	5.3	**147.2**	**25.5**
Campbell and Bozorgnia (2003)	**35.7**	9.7	**36.0**	10.8	**35.0**	7.4	**46.4**	11.0	**78.8**	**28.0**
Sadigh et al. (1997)	**25.4**	7.9	**25.9**	9.1	**22.6**	4.4	**37.6**	6.5	**80.6**	**20.1**
Spudich et al. (1999)	**41.5**	4.3	**30.5**	5.4	**37.4**	2.7	**72.7**	8.8	**153.0**	**35.1**