13th World Conference on Earthquake Engineering
Vancouver, B.C., Canada
August 1-6, 2004
Paper No. 29

USE OF ANALYSIS OF VARIANCE FOR THE INVESTIGATION OF REGIONAL DEPENDENCE OF STRONG GROUND MOTIONS

John DOUGLAS[1]

SUMMARY

The statistical technique known as analysis of variance is applied to a large set of strong-motion data to investigate whether strong ground motions show a regional dependence. This question is important when selecting strong-motion records for the derivation of ground motion prediction equations and also when choosing strong-motion records from one geographical region for design purposes in another region. Three regions with much strong-motion data (California, New Zealand and active parts of Europe) are investigated here. The regional variability of four strong-motion parameters investigated, namely peak ground acceleration (PGA) and response spectral acceleration (SA) for 5% damping at 0.2, 0.5 and 1s. An analysis of ground motions in California and New Zealand and Europe and New Zealand shows that there is little evidence for significant differences in ground motions between the two regions considered, although this conclusion is based on limited data. A comparison of ground motions in California and Europe, for which there is more data available, shows that there is evidence for significantly higher ground motions in California than in Europe although some of this difference may be attributable to uncertainties in the magnitude conversion formulae adopted.

INTRODUCTION

One important problem in the derivation of equations for the estimation of earthquake ground motions is the selection of records based on their geographical origin [e.g. 1]. To derive equations for which the coefficients are robust and which can be used for a wide range of magnitudes and distances it is desired that the set of records used be as large as possible. However, some previous studies [e.g. 2, 3, 4] have found that strong ground motions seem to have a regional dependence. The consequence of this finding is that data from different regions should not be combined because it would increase the standard deviation of the derived equations and could lead to biased predictions when the equations are used. This, however, can lead to small sets of data for which the derived equations can be less well constrained than equations derived with data from larger regions, even though the associated standard deviations of the equations can be lower. Choosing sets of records based on political borders, which do not usually follow tectonic boundaries, is not justified, especially in Europe where countries are small and earthquakes often occur in border regions. An example of such a region is the southern Alps, where accelerograms of the same strong earthquake can be recorded in Austria, Italy, Slovenia and Switzerland. For example, important strong-

[1] Department of Civil & Environmental Engineering, Imperial College London, UK. Email: john.douglas@ukonline.co.uk.

motion records were obtained in both Italy and Slovenia of the Friuli earthquake sequence of 1976 and the Bovec (12th April 1998) earthquake. Within some countries there are two or more areas where differing crustal stress states mean that ground motions could be different. For example, central Italy is undergoing extension whereas Friuli, in northeastern Italy, is undergoing compression. Therefore simply using a country's borders to select strong-motion records is often not adequate.

In view of this, it is necessary to investigate whether earthquake ground motions in certain regions of the world are sufficiently similar so that these data can be combined into one set or whether motions in certain regions differ so that the data should be kept separate.

In the past such investigations have been undertaken by a number of authors. For example, Free [4] finds large differences in ground motions between different stable continental regions and between ground motions in stable continental regions and those in active regions in Europe as predicted by the equations of Ambraseys et al. [5]; Lee [3] finds differences between ground motions in the former Yugoslavia and western USA; and Sigbjörnsson and Baldvinsson [2] find that observed PGAs in Iceland are lower than predicted by equations based on data from western USA or Europe.

These studies have been based on assessing how similar predicted ground motions are using previously derived equations. This method would be adequate if the equations used were derived using the same data selection method, the same distribution of data with respect to magnitude, distance and other independent variables, and the same functional form and regression techniques. However, many different techniques have been adopted in the past for the derivation of such equations [e.g. 1] and consequently it is difficult to find equations for different regions that have been derived using similar techniques. Even if two equations for two different regions were derived using identical techniques, the distribution of data with respect to the independent parameters is likely to be different [e.g. 6]. These differing distributions are likely to affect the predicted ground motions from the equations.

Therefore a different technique to investigate the problem of regional dependence of earthquake ground motions is attempted here. This technique is based on the analysis of variance method developed by R. A. Fisher in 1918 [7] and often applied in other areas of science, such as genetics and agriculture, although not commonly used in engineering seismology or earthquake engineering. Analysis of variance is often applied to the results of controlled experiments. However, because engineering seismology is an observational science, where controlled experiments cannot be made, it is used here to investigate observational data.

The procedure is similar to that undertaken by Douglas and Smit [8], who assessed limits on the accuracy of ground motion predictions using strong-motion records independently of the functional form adopted for the equations.

This technique was recently tried by Douglas [9] for the analysis of strong-motion data from five small regions in Europe. In this paper the technique is tested on data from three larger regions: California, New Zealand and the five parts of Europe where the method was originally tested. There is often not enough observed strong-motion data of ground motions of engineering significance (i.e. close to the source of moderate and large earthquakes) to enable ground motion estimation equations to be derived using only data from a particular region. Consequently near-source data from other regions is often combined to better constrain the behaviour of the derived equations in the near field [e.g. 10]. It is common practice to combine all available data from shallow crustal earthquakes in: all parts of Europe [e.g. 5], all parts of California [e.g. 11], and all parts of New Zealand [e.g. 12]. It would be beneficial to be able to combine all these data together because it would allow the derivations of more robust equations. Therefore these three areas are chosen as the study areas in this paper. Recently Boatwright et al. [13] have found differences in the attenuation rate of PGA and peak ground velocity between four regions within California (the Eureka area, the extended Bay Area, the Sierra and western Mojave desert and the San Juan Bautista area). These possible differences have not been taken into account because it is not yet common practice to separate data from different parts of California.

The null hypothesis in this article is that median ground motions are equal in the three regions considered. This hypothesis will not be rejected unless it is shown that there are significant differences in the recorded ground motions between the different regions.

DATA USED

Three large regions where large shallow earthquakes occur were selected. These regions are California, New Zealand and parts of Europe. By applying the method used here, Douglas [9] showed that ground motions from five regions (the Caucasus region, central Italy, Friuli, Greece and south Iceland) in Europe did not seem to significantly differ. Therefore these five regions were combined in this study for comparison with California and New Zealand. Only records from earthquakes with focal depths less than or equal to 30km were used.

The data was selected from the Imperial College London strong-motion data archive as it stood on 6th January 2004. All records from these regions were visually inspected and those identified as being of too low quality were rejected.

The magnitude scale chosen for this study is surface-wave magnitude (M_s), which has been uniformly reassessed for most of the moderate and large earthquakes in the strong-motion database. For earthquakes with no M_s estimate, m_b estimates or M_L estimates were converted to M_s through conversion formulae. The conversion from m_b to M_s was made using the formula: M_s=1.74m_b-3.82 which was derived by Ambraseys & Bommer [14]. M_L estimates for earthquakes occurring in the five parts of Europe were converted to M_s using conversion formulae derived for each subregion by Douglas [9]. M_L for earthquakes in New Zealand and California were converted to M_s using the formulae of Dowrick [15].

Distance to the surface projection of rupture (d_f) [11] was used if available (records from most earthquakes with M_s>6 have such a estimate). If d_f is not available epicentral distance (d_e) is used instead; for events of M_s < 6 d_e and d_f are similar because the fault length of M_s<6 earthquakes is usually less than 10 km.

Local site conditions at the strong-motion stations are classified using the categories of Boore et al. [16] based on average shear-wave velocity in the top 30 m ($V_{s,30}$) (most strong-motion stations used do not have an estimated $V_{s,30}$ value so the classification was done using descriptions of the sites). The site classes are: soft soil (C) with 180<$V_{s,30}$=360 ms^{-1} , stiff soil (B) with 360<$V_{s,30}$=750 ms^{-1} and rock (A) with $V_{s,30}$ >750 ms^{-1} . Table 1 gives the distribution of the strong-motion data used with respect to local site class.

Table 1 also gives the distribution of records with respect to faulting mechanism for each of the three regions. It shows that most records from California come from strike-slip and reverse faulting earthquakes with very few from normal faulting earthquakes which are rare in California, records from New Zealand are mostly associated with normal and reverse earthquakes while those from Europe are reasonably uniform with respect to style of faulting but with a significant proportion from normal faulting earthquakes, which are common in central Italy and Greece, and from reverse faulting earthquakes, which are common in Friuli and the Caucasus region.

Table 1: Data used for assessing regional dependence of ground motions. %N is the percentage of records from normal faulting earthquakes, %S is the percentage of records from strike-slip faulting earthquakes, %R is the percentage of records from reverse faulting earthquakes, %O is the percentage of records from oblique faulting earthquakes, %A is the number of records from stations classified as rock, %B is the number of records from stations classified as stiff soil, %C is the number of records from stations classified as soft soil and N is the total number of records used (the focal mechanism of some earthquakes is unknown and the local site conditions at some stations are unknown).

Region	%N	%S	%R	%O	%A	%B	%C	N
Parts of Europe	32	14	24	3	42	40	13	1280
California	7	32	36	16	8	35	38	959
New Zealand	24	4	32	22	12	61	24	233

In this article discussion will be limited to ground motions defined in terms of PGA and SA for natural periods 0.2, 0.5 and 1s at 5% damping using the larger horizontal component of each record for each parameter.

METHOD

In this study the data space was divided into small intervals within which an analysis of variance was performed. Intervals of 5 km × 0.25 M_s units were used for this analysis so that there was a sufficiently high number of records within each bin. This is a larger interval size than used by Douglas and Smit [8], who used 2 km × 0.2 M_s units, because when records are split by regions there is not enough data to use such small bins.

The common (base 10) logarithm of the ground motion amplitudes is taken before the analysis of variance is performed since it has been demonstrated [e.g. 8] that this transformation is justified because the standard deviations of the untransformed ground motions are proportional to the mean of the ground motions. A logarithmic transformation removes this dependence [e.g. 17].

In each interval a one-way analysis of variance calculation is made to assess whether the means of the transformed ground motion amplitudes from the different regions are significantly different. A key assumption in analysis of variance is that the variances of each subset are equal. This seems justified because most regression analyses for the prediction of ground motions have found similar standard deviations even when data from different regions of the world is used. For example, most equations for the prediction of peak ground accelerations are associated with standard deviations of about 0.25 to 0.30 in terms of common logarithms [e.g. 1].

In analysis of variance two estimates of the variance of the ground motions are calculated. One estimate is the between-region variance (with n-1 degrees of freedom, where n is the number of regions) and the other is the within-region variation (with N-n degrees of freedom, where N is the total number of records within the bin). Whether or not the medians of the ground motions for the different regions differ, the within-region variation will be an unbiased estimator of the true variance, σ^2 ; the between-region estimator, however, will only be unbiased if the medians of the ground motions are equal, otherwise its expectation will be larger than σ^2. The ratio of the two estimates of the variance of the ground motions is compared to the critical value of F using an F test. The null hypothesis that the median ground motions are equal is rejected if this ratio is greater than the critical value of F for the significance level used (in this study 5%) [e.g. 18].

Correction for site class

The lack of data within each bin militates against the possibility of splitting the data further into site classes. If the data was sufficient one technique to simultaneously analyze both the regional and site dependence effects would be via a two-way analysis of variance [e.g. 18]. Local site effects should be included in the analysis to avoid their effect obscuring possible regional effects or suggesting that there is a regional dependence when in fact the observed differences is due to differing average site conditions in the different regions. To try to reduce the effect of local site conditions the empirical site coefficients of Ambraseys et al. [5] were used to convert the observed ground motions to the expected ground motions on rock. This means PGAs from soft sites were divided by 1.33 and those from stiff sites were divided by 1.31; spectral accelerations at 0.2 s from soft sites were divided by 1.39 and those from stiff sites were divided by 1.36; spectral accelerations at 0.5 s on soft sites were divided by 1.59 and those from stiff sites were divided by 1.40; and spectral accelerations at 1.0 s on soft sites were divided by 1.66 and those from stiff sites were divided by 1.34. An analysis was also performed using all the data irrespective of local site conditions; this lead to similar results to those reported below.

RESULTS

California and New Zealand

Data from New Zealand is limited and there is not a sufficient quantity to enable the analysis of variance technique proposed here to be used for many magnitude-distance intervals, because an interval needs to contain at least two or more records from each region so that an estimate of the variances required can be made. The technique could only be applied in seven intervals, namely: 20-25km, 5.00-5.25M_s; 55-60km, 5.25-5.50M_s; 5-10km, 5.50-5.75M_s; 20-25km, 5.50-5.75M_s; 50-55km, 6.00-6.25M_s; 55-60km, 6.50-6.75M_s; and 170-175km, 7.25-7.50M_s. Of these intervals only for the interval 50-55km, 6.00-6.25M_s for PGA was a significant difference in ground motions in the two regions found and this was based on only two records in the interval from each region. Therefore this technique suggests there is no significant difference in ground motions from shallow earthquakes in California and New Zealand although there is little available data.

Europe and New Zealand

As for the comparison between California and New Zealand there is insufficient data from New Zealand to enable the technique proposed here to be used for many magnitude-distance intervals. For only ten intervals is there sufficient data to enable an analysis of variance to be performed, these are: 0-5km, 2.00-2.25M_s; 25-30km, 4.00-4.25M_s; 5-10km, 4.75-5.00M_s; 20-25km, 5.00-5.25M_s; 30-35km, 5.25-5.50M_s; 5-10km, 5.50-5.75M_s; 20-25km, 5.50-5.75M_s; 60-65km, 5.50-5.75M_s; 60-65km, 6.25-6.50M_s; and 55-60km, 6.50-6.75M_s. For only one of these intervals (25-30km, 4.00-4.25M_s) is there a significant difference in the ground motions (PGA and SA at 1s) between the two regions. Therefore from this analysis there is little evidence for differences in ground motions between Europe and New Zealand although this is based on limited data.

California and Europe

Table 1 shows that a large amount of data from California and parts of Europe was collected for this study. Therefore there is enough data to allow the analysis of variance method proposed here to be used for a large number of magnitude-distance intervals. In order for only accurate estimates of the medians and variances in each bin to be studied further only bins with three or more records from each region were considered. In total 47 magnitude-distance intervals contained enough data from each region to enable an analysis of variance to be performed. To display the results in a concise form a reasonably complicated type of graph had to be used, see Figure 1. Figure 1 displays a series of 47 subplots arranged in an overall plot showing the magnitude (on the y-axis) and distance (on the x-axis) ranges of the bins. Each small graph displays the medians of the ground motions for each of the four strong-motion parameters considered (the first two points are PGA, the second two points are spectral acceleration at 0.2s, the third two points are spectral acceleration at 0.5s and the final two points are spectral acceleration at 1.0s). The ordinate of the small graphs is logarithm of acceleration in ms^{-2}. Therefore they can be thought of as response spectra with only four ordinates. The left point in each pair is for California and the right point is for Europe. If the difference in the medians was found to be significant at the 5% significance level using the F-test then the marker is a cross rather than a dot. The two numbers in the top right corner are the total number of records in the bin from each region (the left number is for California and the right number is for Europe).

Figure 1 shows that for 26 of the bins, covering most of the magnitude and distance space where the analysis could be performed, there is no significant difference between ground motions in California and Europe at any of the periods considered. For the rest of the 47 bins there are significant differences in the ground motions in the two regions at least one period. For all of these 21 bins the ground motions in California are significantly higher than those in Europe except for the interval 0-5km and 6.50-6.75M_s where European ground motions are significantly higher than those in California for PGA and SA at 0.5s and 1s. All the data in this bin from Europe comes from the South Iceland earthquake of 21st June 2000 with magnitude 6.6, which was recorded by three stations within 5km of the fault. The PGAs at these three stations were 8.2ms^{-2}, 7.1ms^{-2} and 5.6ms^{-2}, all of which are significantly higher than would be expected from such a sized earthquake recorded at such distances. For example, the predicted PGAs using the equation of Ambraseys et al. [5] at rock sites at the same distances (2, 4 and 3km) as these stations from a 6.6M_s earthquake are 5.1ms^{-2}, 4.0ms^{-2} and 4.5ms^{-2} respectively. Spectral accelerations of these three records are also larger than would be expected. One possible reason for the larger ground motions at these stations from this earthquake is directivity because the stations are located at the end of the fault that ruptured during the earthquake and the rupture propagated towards them. Consequently since the data within this bin from Europe comes from a single earthquake and the ground motions from this earthquake may not be typical of European ground motions the finding that ground motions in Europe are higher than those in California for this interval should not be consider to be proven.

Significant differences in ground motions for small magnitude (M_s<5) earthquakes may be attributable to problems with the magnitude conversion formulae. M_s estimates are not commonly given for such small earthquakes and therefore a magnitude conversion from m_b or M_L needs to be performed. For the European earthquakes individual conversion formulae for M_L to M_s for each of the five subregions were derived using the available data (see Douglas [9] for details) and consequently are likely to be appropriate for all the records. For the M_L to M_s conversion for Californian earthquakes the formula presented by Dowrick [15] was used. This equation was derived using the M_L data presented by Joyner & Boore [11] for moderate and large magnitude (M_s>5) earthquakes and consequently may not be appropriate for smaller magnitude earthquakes. Since however, a significant proportion of the intervals for small magnitude earthquakes do not show a significant difference in ground motions it is unlikely that the problem of magnitude conversion is significantly affecting the results.

Differences in ground motions between the two regions at moderate distances (i.e. d>50km) could be attributable to differences in the crustal structure in the two regions [e.g. 19]. Suhadolc and Chiaruttini [19] compute theoretical decay curves for peak ground accelerations using a modal summation technique for four regions (Imperial Valley in California, Apennines in central Italy, Irpina in southern Italy and Friuli in north-east Italy) with differing crustal structures. They find that the decay curves differ by over a factor of ten particularly for d>100km because of the domination of certain seismic phases at different distances [19, Fig. 9]. Crustal structure in the different regions of Europe included in the European set (Caucasus region, central Italy, Friuli, Greece and south Iceland) is significantly different and also crustal structure in California also varies considerably. Therefore it is unlikely that the average structures in the two large regions differ by more than the variation in structures within the regions and consequently systematic differences due to crustal structure are considered to be improbable.

One probable reason why ground motions in California seem to be significantly higher than those in Europe in certain magnitude and distance intervals is due to differences in ground motions due to style of faulting. The predominant faulting mechanisms in California are strike-slip (32%) and reverse (36%), with few records from normal faulting earthquakes (only 7%), whereas about a third of the records in the European set are from normal faulting earthquakes and fewer from reverse faulting earthquakes (24%). Bommer et al. [20] have investigated the effect of faulting mechanism on observed strong ground motions and have shown that there is considerable evidence that PGAs and SAs from reverse faulting earthquakes are about 20 to 40% higher than PGAs and SAs from strike-slip faulting earthquakes and that there is some evidence for slightly smaller ground motions from normal faulting earthquakes compared with strike-slip earthquakes. Therefore this effect may contribute to the observed differences in ground motions between California and Europe.

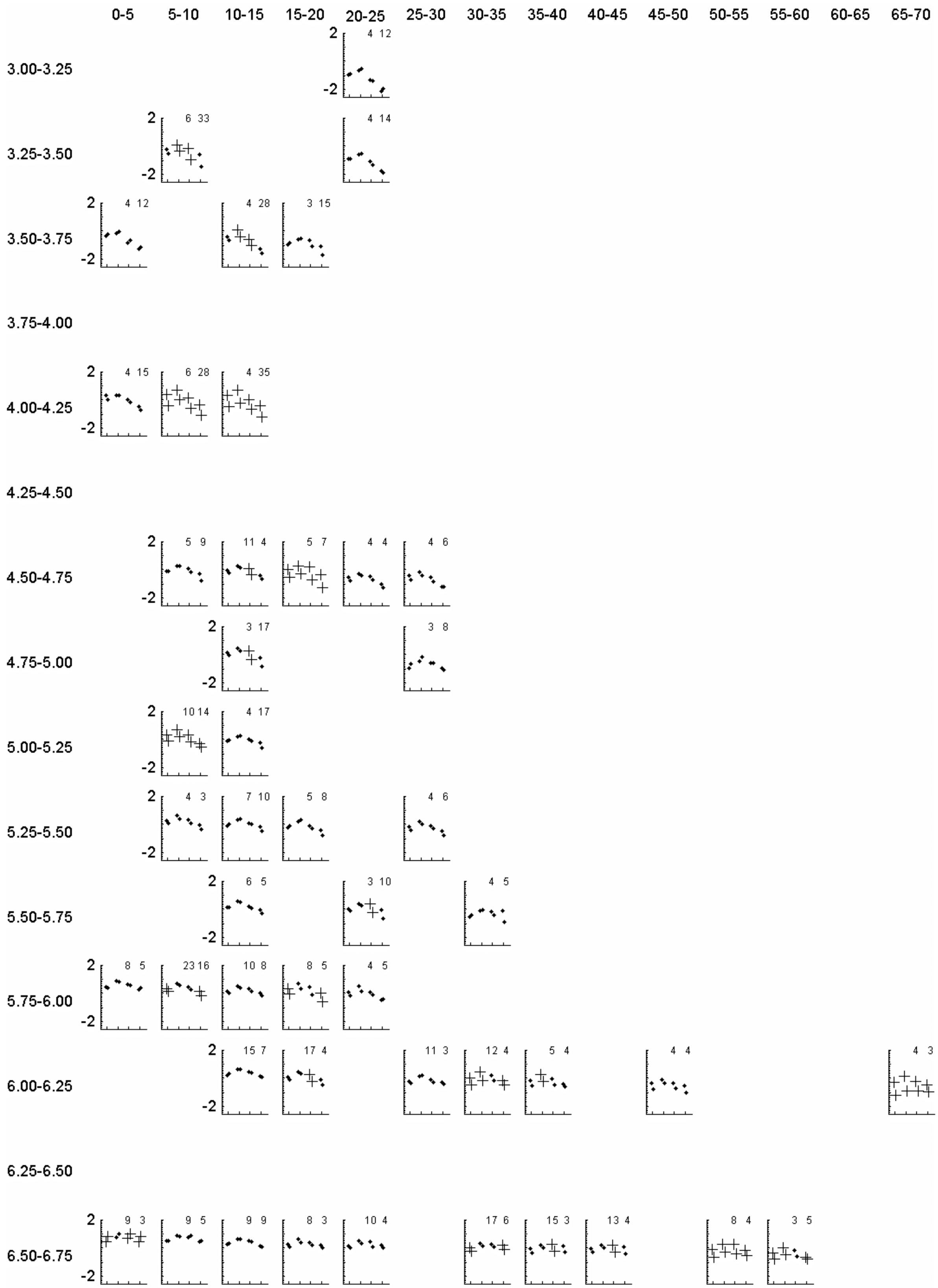

Figure 1: Graphs for each bin where analysis of variance was performed to compare ground motions in California and Europe. See main text for explanation of figure.

DISCUSSION AND CONCLUSIONS

One problem with this method is that comparison cannot be made for the ground motions of most engineering interest, i.e. those from close to large magnitude earthquakes, because of the sparsity of the data. As the amount of data increases, however, the method of analysis presented here can be repeated to assess the similarities between ground motions of engineering significance between different regions. More strong-motion data from large magnitude earthquakes is required to ascertain whether the conclusions reached here, mainly based on ground motions with low engineering significance, hold for larger ground motions.

Analysis of variance has the ability to complement other techniques for assessment of regional dependence of ground motions. Examples of such techniques are comparison of source spectra [21], analysis of residuals from previously derived ground motion estimation equations, comparisons of macroseismic intensity relations and computations of synthetic ground motions considering the regional crustal structure [19].

From the analysis presented here there seems to be little evidence for regional differences in ground motions between California and New Zealand and Europe and New Zealand but this is based on limited data. An analysis for the two regions with large amounts of data (California and Europe) shows that there is evidence for significantly higher ground motions in California than in Europe although this is only true for about half of the magnitude-distance intervals investigated and therefore needs further investigation. One possible reason for these observed differences could be different predominant styles of faulting in the two regions.

ACKNOWLEDGEMENTS

I thank my Dad for the discussion that led to the idea behind this paper, Prof. N. N. Ambraseys for his constant encouragement and suggestions, Prof. R. Sigbjörnsson for his advice regarding this study and information on the local site conditions at Icelandic strong-motion stations and Dr J. J. Bommer, Dr S. K. Sarma and Ms F. O. Strasser for thorough critical reviews of the submitted Journal of Seismology paper on which this article is based. This work was funded by a grant from the Engineering and Physical Sciences Research Council (GR/R52114/01); I thank them for their support.

REFERENCES

1. Douglas J. "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: 43–104, 2003.
2. Sigbjörnsson R., Baldvinsson G. I. "Seismic hazard and recordings of strong ground motion in Iceland", Proceedings of Tenth World Conference on Earthquake Engineering; 1:419–424, 1992.
3. Lee V. W. "Pseudo relative velocity spectra in former Yugoslavia", European Earthquake Engineering; IX: 12–22, 1995.
4. Free M. W. "The Attenuation of Earthquake Strong-Motion in Intraplate regions", Ph.D. thesis, University of London, 1996.
5. Ambraseys N. N., Simpson K. A., Bommer J. J. "Prediction of horizontal response spectra in Europe", Earthquake Eng. & Structural Dyn.; 25: 371–400, 1996.
6. Ambraseys N. N., Simpson K. A., Bommer J. J. "Authors' reply", Earthquake Eng. & Structural Dyn.; 26: 295–300, 1997.

7. Fisher R. A. “Statistical Methods, Experimental Design, and Scientific Inference”, Oxford University Press, Oxford, United Kingdom, 1990.
8. Douglas J., Smit P. M. “How accurate can strong ground motion attenuation relations be?” Bull. Seism. Soc. Am.; 91(6): 1917–1923, 2001.
9. Douglas J. “An investigation of analysis of variance as a tool for exploring regional differences in strong ground motions”, J. of Seismology. Submitted, 2003.
10. Fukushima Y., Berge-Thierry C., Volant P., Griot-Pommera D.-A., Cotton F. “Attenuation relation for western Eurasia determined with recent near-fault records from California, Japan and Turkey”, J. of Earthquake Eng.; 7(4): 573–598.
11. Joyner W. B., Boore D. M. “Peak horizontal acceleration and velocity from strong-motion records including records from the 1979 Imperial Valley, California, earthquake” Bull. Seism. Soc. Am.; 71: 2011–2038, 1981.
12. Cousins W. J., Zhao J. X., Perrin N. D. “A model for the attenuation of peak ground acceleration in New Zealand earthquakes based on seismograph and accelerograph data”, Bull. of the New Zealand Soc. for Earthquake Eng.; 32(4): 193–220, 1999.
13. Boatwright J., Bundock H., Luetgert J., Seekins L., Gee L., Lombard P. “The dependence of PGA and PGV on distance and magnitude inferred from northern California ShakeMap data”, Bull. Seism. Soc. Am.; 93(5): 2043–2055, 2003.
14. Ambraseys N. N., Bommer J. J., “Uniform magnitude re-evaluation for the strong-motion database of Europe and adjacent areas”, European Earthquake Eng.; IV: 3–16, 1990.
15. Dowrick D.J. “Magnitude reassessment of New Zealand earthquakes”, Earthquake Eng. & Structural Dyn., 20(6): 577-596, 1991.
16. Boore D. M. Joyner W. B., Fumal T. E. “Estimation of response spectra and peak accelerations from western North American earthquakes: An interim report”, Open-File Report 93-509, U.S. Geological Survey, 1993.
17. Draper N. R., Smith H. “Applied Regression Analysis”, 2nd edition. John Wiley & Sons, 1981.
18. Green J. R., Margerison D. “Statistical Treatment of Experimental Data”, Physical Sciences Data 2, Second Impression edition. Elsevier Scientific Publishing Company, Amsterdam, The Netherlands, 1979.
19. Suhadolc P., Chiaruttini C. “A theoretical study of the dependence of the peak ground acceleration on source and structure parameters”. In “Strong Ground Motion Seismology”, M. Erdik and M. Toksöz, editors, 143–183. D. Reidel Publishing Company, 1987.
20. Bommer J. J., Douglas J., Strasser F. O., “Style-of-faulting in ground-motion prediction equations”, Bull. Of Earthquake Eng.; 1(2):171-203, 2003.
21. Chen S.-Z., Atkinson G. M. Global comparison of earthquake source spectra. Bull. Seism. Soc. Am.; 92: 885–895, 2002.