SEISMIC HAZARD ASSESSMENT AND ZONING IN JAVA: NEW AND ALTERNATIVE PROBABILISTIC ASSESSMENT MODELS

PAUL W. BURTON[1], GRAEME WEATHERILL[1], DWIKORITA KARNAWATI[2] and
SUBAGYO PRAMUMIJOYO[2]

[1] *Seismic Risk Group, School of Environmental Sciences, University of East Anglia, Norwich, U.K.*
[2] *Department of Geological Engineering, Gadjah Mada University, Yogyakarta, Indonesia*
Email: p.burton@uea.ac.uk

ABSTRACT: Applying alternative and different approaches to seismic hazard assessment is instructive. It allows learning from the different outcomes of the different approaches. These outcomes may be mutually reinforcing or diverge, suggest further study and research is needed, or provide new insights into old problems. Herein Java island-scale seismic hazard will be considered by applying different probabilistic approaches to hazard assessment. Results from two distinct methods are provided for Java: 1) primary zoning using K-means partitioning of seismicity into spatial clusters (progressed into zones) which are then developed into seismic hazard maps using Monte Carlo earthquake catalogue simulation, and 2) extreme value analysis applied at a matrix of points throughout a zoneless Java. The latter approach has been used before, the former adopts seismicity partitioning into spatial clusters prior to Monte Carlo modelling and is novel. The earthquake catalogue analysed is NEIC (1973-2006). This catalogue is homogenised to the moment magnitude scale M_W and Poisson declustering of fore- and after-shocks applied. The completeness threshold is around 4.9 M_W. Shallow earthquakes down to 80 km depth contribute most to the hazard and are partitioned into 1 to K trial clusters of seismicity by minimising the total within cluster distance from seed centroids. Repeated trials produce an optimum partition. A variety of indices can be invoked to try to quantify cluster quality for a given K; in addition to this, it is decided to seek the best value of K by testing the influence of K on ensuing seismic hazard analyses. Monte Carlo synthesis generates synthetic catalogues for each K value, from which peak ground acceleration (PGA) hazards are calculated and compared against results from the observed catalogue to choose acceptable K values. To summarise the results, seismic hazard maps are constructed for two acceptable values of K (8 and 27) for Java from the Poisson declustered catalogue of shallow earthquakes using the Boore, Joyner, Fumal attenuation law. Not surprisingly the smaller value of K with 8 clusters (progressed to zones) produces the smoother hazard map. All of the maps indicate highest hazard around the Sunda Strait and a general expectation in Java Island of 100-300 cm s^{-2} with one-in-ten chance of exceedance in 50 years.

1. INTRODUCTION

Subduction tectonics controls most of the major seismicity of Java with the Australian plate dipping below Java. Nevertheless the Bantul, Yogyakarta earthquake of May 27, 2006 occurred at shallow depth in the Sunda plate rather than in the subducting Australian plate. So the ensuing seismic hazard can not be attributed to a single type of seismotectonic action. Previous maps of seismic hazard in Indonesia and Java have been provided by Kertapati *et al.* (1999), GSHAP (1999) and Petersen *et al.* (2004) and show broad bands of contoured seismic hazard that largely follow the trend of the island arc system. These maps adopt the arbitrary but conventional seismic hazard statistic of the 475-year average occurrence.

Most probabilistic seismic hazard analyses fundamentally follow the longstanding approach of Cornell that was pioneering development in the 1960s. There are alternatives which fundamentally, or in detail, differ from the Cornell approach. Our purpose here is to present the outcomes to probabilistic seismic hazard analysis of strong ground shaking from two other approaches which are: extreme value analysis; Monte Carlo (MC) modelling. The emphasis here relates to description of the earthquake catalogue used in the analyses, the NEIC catalogue 1973-2006, and presentation for inspection of seismic hazard maps for Java from the latter approach but with comparison to the extreme value approach. These methods are described in detail elsewhere and summarized briefly herein. The arbitrary but conventional seismic hazard statistic of the 475-year occurrence i.e. the ground shaking that has a one-in-ten chance of being exceeded in a 50-year period is adopted for maps that depict the expected peak ground acceleration distribution.

2. EARTHQUAKE CATALOGUE

The first priority of any probabilistic seismic hazard analysis (PSHA) is the foundation of a good earthquake catalogue. Earthquake catalogues that include Indonesia and Java are available from several sources. Well known ones include: the National Earthquake Information Center (NEIC) database, Engdahl and Villasenor (2002), the SEASEE (1985) catalogue and data from other USGS sources. Here the well established NEIC catalogue has been adopted, from which data for Java have been extracted as NEIC (1973-2006).

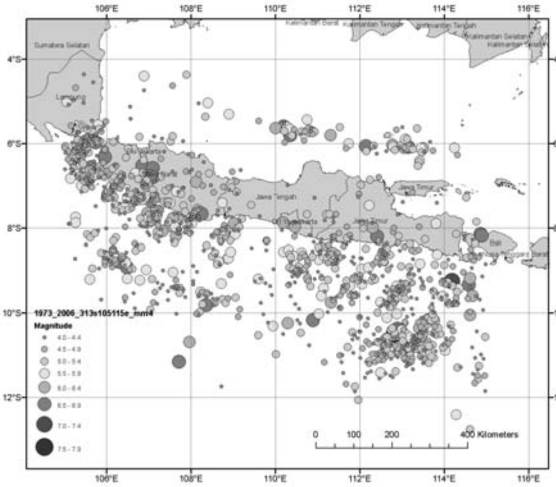

Figure 1 Spatial distribution of earthquakes in the NEIC Java catalogue.

For this regional scale analysis of seismicity in Indonesia, the Java catalogue contains 1993 events in the magnitude range $4.5 \leq M_w \leq 7.8$. The area covered by the catalogue is enclosed by 3.5ºS - 13ºS and 105ºE - 115ºE. It is necessary to homogenize the magnitude data to a uniform scale prior to analysis. The reported magnitude for this catalogue is mostly body-wave magnitude (m_b), with some events recorded in terms of surface wave magnitude (M_S) and moment magnitude (M_w). The NEIC Java catalogue has been homogenised into two magnitude scales, M_w and M_S, and much of the procedure is described in Burton and Cole (2008). The essential equation for conversion to Mw from m_b is from Scordilis (2006) and to Ms from mb is from Rezapour and Pearce (1998) and otherwise it is assumed M_w equals M_S when required. The seismicity distribution throughout Java represented by this homogenized earthquake catalogue is shown in Figure 1 which illustrates the epicentral distribution for Java.

For the purposes of MC seismic hazard analysis, non-Poissonian events are removed from the earthquake catalogue. The algorithm adopted (Musson, 1999) identifies foreshocks and aftershocks around a large event using the fixed distance windows described in Gardner and Knopoff (1974), and a moving time window of 100 days. When purged of non-Poissonian events the NEIC catalogue contains 1511 earthquakes, spanning the same magnitude range as described previously. A completeness analysis is also performed suggesting the catalogue is complete above 4.9 M_w.

3. SEISMICITY CLUSTERING USING K-MEANS AND PROGRESS TO ZONES

The K-means algorithm is a useful technique for partitioning a set of spatially distributed data into clusters (Hartigan, 1975; review by Jain *et al.*, 1999). Where locally robust geological and seismotectonic data are in short supply, this method of cluster analysis can be applied as a means to identify spatial differences in seismicity.

For a set of K initial estimates of the centroids (m_k) the algorithm partitions the N data by minimizing the total within-cluster sum of squares:

$$TWCSS = \sum_{i=1}^{N}\sum_{k=1}^{K} I\left(x_i \in C_k\right)\left\|x_i - m_k\right\|^2$$

where I(x) is 1 if x is true and 0 otherwise. The partition can be influenced by the choice of initial centroids. This is overcome by performing an ensemble analysis of 100 trials and selecting the optimum partition from the trials. This gives a result near but not necessarily exactly equal to the global optimum. A second issue is that the optimum number of clusters is not known. TWCSS is lower for better partitions but it also decreases with increasing K, making it poor at identifying the optimum number of zones. There are several indices of cluster quality which can be used to overcome this, amongst which that of Krzanowski and Lai (1988), the KL index, produces a robust indication of cluster quality. In principle optimum K maximizes this index. In practice several values of K may stand out as producing reasonable fits and so it is recommended that each of these be tested and explored, rather than simply selecting the numerical optimum. The K-means algorithm can be applied to any spatial point data set. In seismology, the assigning of each earthquake to a point source is not physically ideal. So each earthquake location is weighted according to its rupture length using a relation from Wells and Coppersmith (1994). Full details are provided in Weatherill and Burton (2008).

A further step is to test each value of K in terms of how K influences ensuing seismic hazard analysis. Monte Carlo is used to produce 100 synthetic catalogues of duration equal to the observed catalogue. Maximum ground motion at each point in a grid of NG points separated at $0.5^{\circ} \times 0.5^{\circ}$ is calculated for the observed and each synthetic catalogue. The mean and standard deviation of the logarithm of the maximum ground motion are compared with the observed values and the goodness-of-fit assessed using χ^2 ; partitions producing lower χ^2 suggest better fit to the observed seismicity. Monte Carlo hypocentres are created by either random sampling from a uniform zone encapsulating each cluster or by sampling with replacement from the observations within each cluster. Magnitudes within a cluster are generated by random sampling from the cumulative distribution function of the doubly truncated Gutenberg and Richter relation with b-values determined using the maximum likelihood method. Maximum magnitude is determined in a cluster using the cumulative moment method of Makropolous and Burton (1983). The strong ground motion parameter used for this purpose is Arias intensity (Ia), here defined by the global attenuation relation of Travasarou *et al.* (2003):

$$\ln\left(I_a\right) = 2.800 - 1.981\left(M_W - 6\right) + 20.72\ln\left(M_W / 6\right) - 1.703\ln\left(\sqrt{R_{rupt}^2 + 8.78^2}\right)...$$
$$+ \left(0.454 + 0.101\left(M_W - 6\right)\right)S_C + \left(0.479 + 0.334\left(M_W - 6\right)\right)S_D - 0.166F_N + 0.512F_R + \sigma P$$

where R_{rupt} is the closest distance to the rupture plane, S_C and S_D indicate soil type and F_N and F_R indicate fault type. Site category B and strike slip faulting is assumed; the latter represents a compromise between the lower ground motion values created by normal faulting earthquakes and higher values created by thrust faulting earthquakes, although in reality, the seismicity of Java is typically thrust faulting along the Sunda arc and some shallow extensional faulting along the north of the island and Java Sea. Arias intensity is adopted at this stage because the duration dependence of the parameter means that it often provides a better correlation to Intensity (Cabanas *et al.*, 1997) and other indices of damage than zero-period acceleration. Also Arias intensities are more strongly non-linear at lower magnitude than PGA because of the duration dependence, so smaller earthquakes should contribute less significantly to hazard.

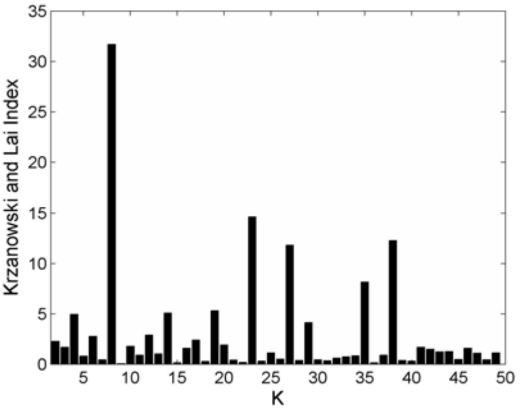

Figure 2 An optimum-K index for the NEIC Java catalogue: shallow and Poisson-declustered.

Our investigations show that the KL index produces good partitions at K = 9, 11, 13, 24 and 37 for the full shallow catalogue and K = 8, 23, 27 and 38 for the Poisson-declustered catalogue (Figure 2). The declustered catalogue produces the lower K value. Several of these estimates are consistent (K = 8-9, 23-24 and 37-38) and these values are good candidates for optimum number of zones. The χ^2 method on the declustered catalogue generates a larger amount of variability; amongst the better performing values are K = 3, 9, 21, 38 and 45 for the inhomogeneous seismicity (non-zoned) method and K = 8, 17, 34 and 47 for homogeneous seismicity (zoned). These values of K are less consistent than those from the clustering indices because the χ^2 values are dependent on many parameters of each cluster (b, db, M_{max}, rate, cluster compactness i.e. zone size), rather than a single index. Where particular indices stand out they define zonation schemes worthy of investigation. Spatial partitions for K = 8 and 27 are indicated by ellipses in Figure 3.

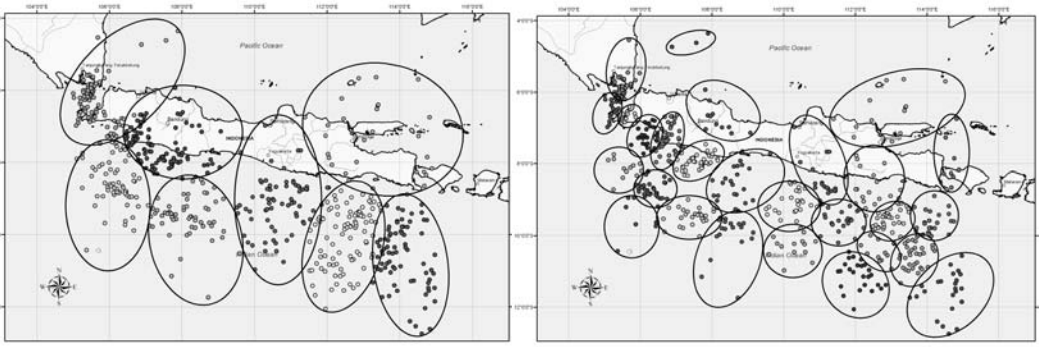

Figure 3 Partitions of the NEIC Poisson-declustered catalogue for shallow earthquakes: K = 8 and K = 27.

4. SEISMIC HAZARD: EXTREME VALUE AND MONTE CARLO ANALYSES

The next step is to use these results in combination with the Monte Carlo approach to seismic hazard analysis (e.g. Ebel and Kafka, 1999; Musson, 1999; Weatherill and Burton, 2006, 2008) to produce seismic hazard maps for Java. Progressing the K = 8 clusters to geometric "zones" produces the set in Figure 4. The degree of homogeneous versus inhomogeneous seismicity in "zones" is apparent; indeed these "zones" can be treated as inhomogeneous seismicity (non-zoned) or traditional homogeneous seismicity (zoned). Only the declustered catalogue is used because the Monte Carlo approach considers seismicity to be Poissonian. K = 8, 27 and 45 (not illustrated) is adopted for shallow earthquakes in Java, and although not necessarily with the best KL indices, these performed well and give an indication of the impact with increasing number of cluster zones.

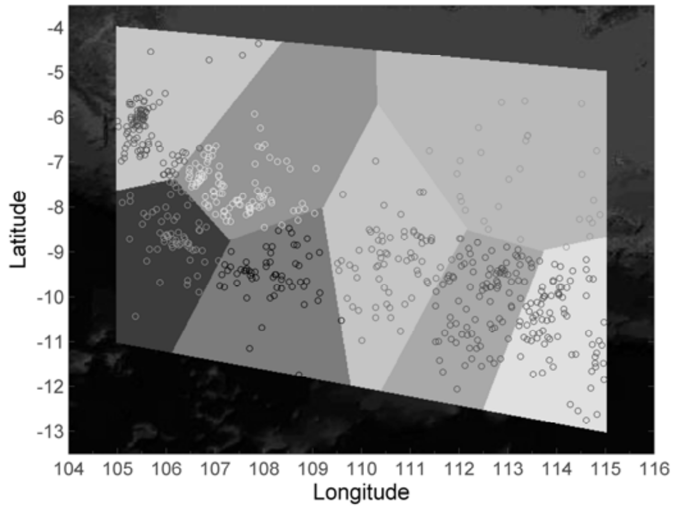

Figure 4 Uniform zones created by the K-means partition of the NEIC shallow declustered dataset, K = 8.

The attenuation relation for PGA used to produce the Monte Carlo seismic hazard maps is Boore *et al.'s* (1997):

$$\ln PGA = -0.242 + 0.527\left(M_W - 6\right) - 0.778\ln\left(\sqrt{r_{jb}^2 + 5.57^2}\right) - 0.371\ln\left(\frac{V_{S30}}{1396}\right) + 0.52P$$

where Vs30 = 1070 m/s is recommended for a rock site. The geometric mean PGA from the synthetic catalogue results is shown in Figure 6. Prior to this Figure 5 provides the results of an extreme value analysis, the development of which is not described here. This uses Patwardhan *et al.* (1978) attenuation. We have described the technique elsewhere (Burton *et al.*, 2003) and this specific result is taken from Burton and Cole (2008) for comparison purposes.

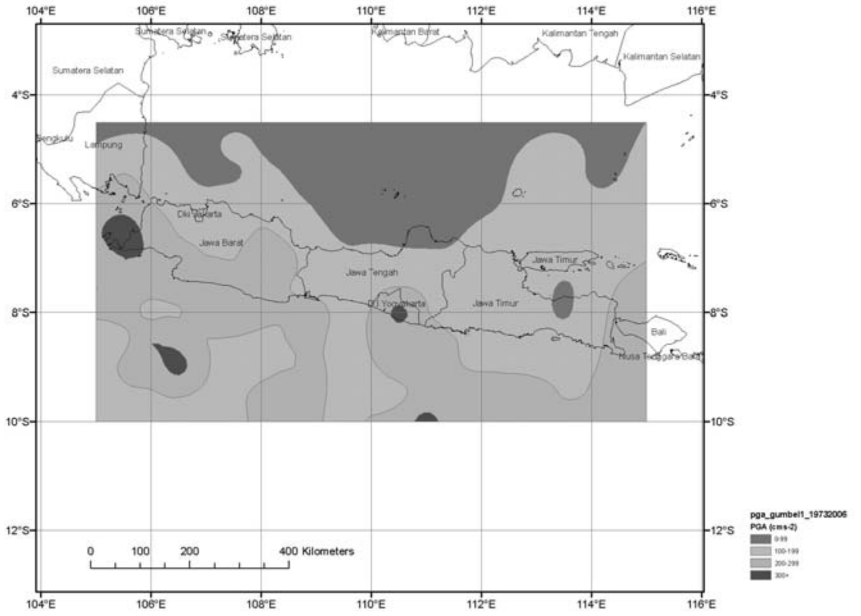

Figure 5 Extreme value map of seismic hazard showing PGA with 10% probability of exceedance in 50 years (after Burton and Cole, 2008).

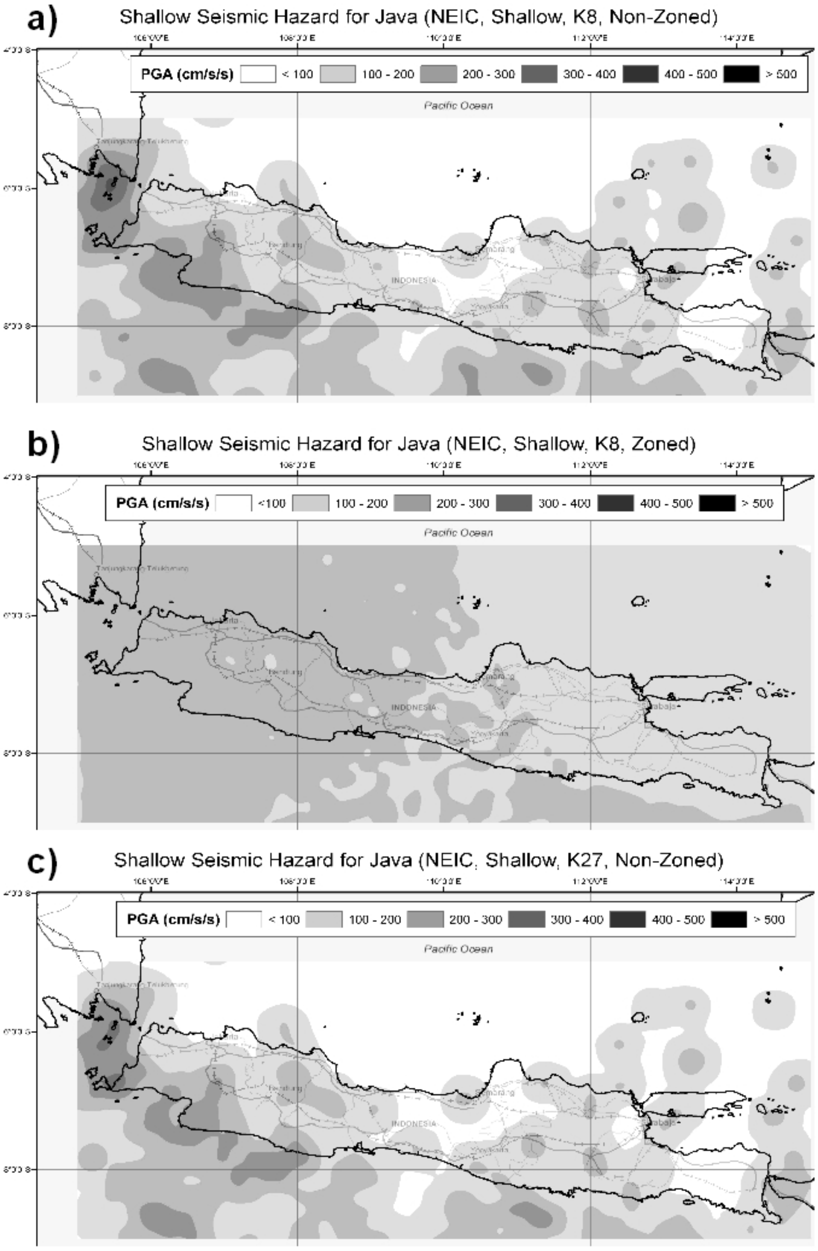

Figure 6 Monte Carlo maps of seismic hazard showing PGA with 10% probability of exceedance in 50 years with: a) K = 8 (non-zoned), b) K = 8 (zoned) and c) K = 27 (non-zoned).

The most obvious contrast between the zoned and non-zoned methods in Figure 6 is that hazard is smoothed out to a greater degree for the zoned analysis. This effect is more obvious when fewer zones are used (Figures 6a-b). All of the models, whether extreme value or the variety of MC approaches, generally agree that the highest hazard is associated with the Sunda Strait, which is clearly consistent with the high density of shallow events found in this region. Hazard also generally decreases from the highest values in the southwest of Java and Sunda arc, to the lowest values in the east and north of the region. This pattern of seismic hazard reflects the increasing depths and decreasing density (per unit area on the Earth's surface) of earthquakes deeper in the crust and upper mantle originating from the Wadati-Benioff zone. For much of the mainland of Java the PGA with a 10 % probability of being exceeded in 50 years lies between 100 and 300 cm s^{-2}. One might reasonably expect this value to increase when hazard from intermediate depth earthquakes is included in the analysis.

5. DISCUSSION AND CONCLUSIONS

K-means partitioning into clusters with K = 8 and 27, both supported by good KL index values, were selected to inspect a variety of results from this methodology. When clusters are progressed into geometric "zones" from which seismic hazard is calculated and mapped, treating the "zones" as inhomogeneous or homogeneous seismicity produces contrasting maps (Figures 6a-b for K = 8); the latter, not surprisingly, are much smoother. Inspection of epicentral patterns in "zones" (Figure 4) treated as homogeneous emphasizes what we know to be true; it is difficult to view many of these as truly homogeneous, there are long-standing issues of stationarity of the process remaining unconsidered and Figure 6a is preferred at this time. Introducing more clusters (K = 27) progressed to zones produces a similar hazard map (compare Figures 6a and c); Figure 6c is slightly less smooth than 6a but retains very similar features. Recent earthquake hazard zone maps opt for five zones for Java (Supartoyo *et al.*, 2006), there does not seem to be good reason to opt for 27. Figure 6a remains preferred at this time with 8 inhomogeneous areas of seismicity (non-zoned).

The project has its roots in the Bantul, Yogyakarta 2006 earthquake catastrophe (more than 5,000 fatalities and about hundred thousand houses collapsed or damaged locally). An eventual aim is to analyse seismic hazard on three scales: local (Bantul, in situ study), provincial (Yogyakarta Province) and national (island) scale. This will provide an opportunity in local populations to understand their potential exposure within the relative hierarchy of earthquake hazard across Indonesia. These new island-scale maps indicate a PGA range in Yogyakarta Province of 100-200 cm s^{-2} with local high 200-300 cm s^{-2} for Bantul area. The extreme value approach, using a different attenuation law, shows a similar contoured hazard configuration in Yogyakarta Province with values 200-300 cm s^{-2} and local high at 300+ cm s^{-2} around Bantul.

6. ACKNOWLEDGEMENTS

We are grateful to several colleagues in the seismological community for supplying and indicating sources of data. The project is funded partially by the British Council.

7. REFERENCES

Boore, D.M., Joyner, W.B. and Fumal, T.E. (1997). "Equations for Estimating Horizontal Response Spectra and Peak Acceleration from Western North American Earthquakes: A Summary of Recent Work", *Seismological Research Letters*, Vol. 68, 128 - 153.

Burton, P.W. and Cole, S.W. (2008). "Development towards the seismic hazard mapping of Java", in "*The Yogyakarta Earthquake of May 27, 2006*", eds. D. Karnawati, S. Pramumijoyo, R. Anderson and S. Husein, pp. 10-1–10-13, Star Publishing Company, Inc., Belmont, CA.

Burton, P.W., Xu, Y., Tselentis, G.-A., Sokos, E. and Aspinall, W. (2003). "Strong ground acceleration seismic hazard in Greece and neighbouring regions", *Soil Dynamics & Earthquake Eng.,* Vol. 23**,** 159-181.

Cabanas, L., Benito, B. and Herraiz, M. (1997). "An approach to the measurement of the potential structural damage of earthquake ground motions", *Earthquake Engineering and Structural Dynamics*, Vol. 26, 79 -92.

Ebel, J.E. and Kafka, A.L. (1999). "A Monte Carlo Approach to Seismic Hazard Analysis", *Bulletin of the Seismological Society of America,* Vol. 89, 854-866.

Engdahl, E.R. and Villasenor, A. (2002). "Global Seismicity 1900 – 1999", in "*Intenational Handbook of Engineering Seismology",* edited by W.H.K. Lee, H. Kanamori, P.C. Jennings, and C. Kisslinger, Part A, Chapter 41, pp. 665 - 690, Academic Press.

Gardner, J.K. and Knopoff, L. (1974). "Is the sequence of earthquakes in Southern California, with aftershocks removed, Poissonian?", *Bulletin of the Seismological Society of America*, Vol. 64, 1363 - 1367.

GSHAP, (1999). Giardini, D. (ed.) "The Global Seismic Hazard Assessment Program (GSHAP) 1992-1999, summary volume", *Annali di Geofisica*, Vol. 42, 957-1230. (Giardini, D., Grunthal, G., Shedlock, K.M. and Zhand, P., "The GSHAP global seismic hazard map", pp. 1225-1230.)

Hartigan, J.A. (1975). "*Clustering Algorithms*", John Wiley and Sons, New York.

Jain, A.K., Murty, M.N. and Flynn, P.J. (1999). "Data Clustering: A Review", *ACM Computing Surveys*, Vol. 31, 264 - 322.

Kertapati, E.K., Setiawu, Y.B. and Ipranta (1999). "Earthquake ground shaking hazard map of Indonesia", 1:5,000,000, Transverse Mercator Projection, Geological Research and Development Centre, Jl. Diponegoro 57, Bandung 40122.

Krzanowski, W.J. and Lai, Y.T. (1988). "A criterion for determining the number of groups in a data set using sum-of-squares clustering", *Biometrics*, Vol. 44, 23 - 34.

Makropoulos, K.C. and Burton, P.W. (1983). "Seismic risk of circum-Pacific earthquakes I: Strain-energy release", *Pure and Applied Geophysics*, Vol. 121, 247-267.

Musson, R.M.W. (1999). "Probabilistic seismic hazard maps for the North Balkan region", *Annali di Geofisica*, Vol. 42, 1109-1124.

Patwardhan, A., Sadigh, K., Idriss, I.M. and Youngs, R. (1978). Reported in Idriss, I.M. (1978) "Characteristics of earthquake ground motions", *Proc. ASCE Geotechnical Engineering Division Speciality Conference: Earthquake Engineering and Soil Dynamics*, Vol. III (June), 1151-1265.

Petersen, M.D., Dewey, J., Hartzell, S., Mueller, C., Harmsen, S., Frankel, A.D. and Rukstales, K. (2004). "Probabilistic seismic hazard analysis for Sumatra, Indonesia and across the Southern Malaysian Peninsula", *Tectonophysics,* Vol. 390*,* 141-158.

Rezapour, M. and Pearce, R.G. (1998). "Bias in surface-wave magnitude Ms due to inadequate distance corrections", *Bulletin of the Seismological Society of America*, Vol. 88, 43 - 61.

Scordilis, E.M. (2006). "Empirical global relations converting Ms and mb to moment magnitude", *Journal of Seismology*, Vol. 10, 225 - 236.

SEASEE (1985). Soetardjo, Untung, M., Arnold, E.P., Soetadi, R., Ismail, S. and Kertapati, E.K. "A catalogue of destructive earthquakes in Indonesia for the period 1821-1984", in: E.P. Arnold (editor) "Southeast Asia Assoc. of Seismol. and Earthquake Eng., Series on Seismology", Vol. V – "Indonesia", Part A, 42 pages.

Supartoyo, Putranto, E.T. and Djadja (2006). "Earthquake hazard zone map of Indonesia", 3rd edition, Center for Volcanology and Geological Hazard Mitigation, Jl. Diponegoro, Bandung.

Travasarou, T., Bray, J.D. and Abrahamson, N.A. (2003). "Empirical attenuation relationship for Arias Intensity", *Earthquake Engineering and Structural Dynamics*, Vol. 32, 1133 - 1155.

Weatherill, G.A. and Burton, P.W. (2006). "The application of multiple random earthquake simulations to probabilistic seismic hazard assessment in the Aegean region", in *Proc. First European Conference on Earthquake Engineering and Seismology,* Paper No. 583, Geneva, Switzerland, 3-8 September 2006.

Weatherill, G. and Burton, P.W. (2008). "Delineation of shallow seismic source zones using K-means cluster analysis, with application to the Aegean region", *Geophysical J. Int.* [submitted].

Wells, D.L. and Coppersmith, K.J. (1994). "New empirical relationships among magnitude, rupture length, rupture width, rupture area, and surface displacement", *Bulletin of the Seismological Society of America*, Vol. 84, 974-1002.