Automatic calculation of seismicity rates in eastern Queensland

Russell Cuthbertson

Environmental Systems and Services

Abstract

In the probabilistic approach to seismic hazard analysis, earthquake source zones are assigned seismicity rates with the assumption that seismicity is distributed uniformly within each zone. Defining a seismic source zone and then calculating the seismicity rates of that zone has typically required manual calculations and corrections that are time consuming and which ignore significant proportions of the recorded dataset.

An automatically calculated detection threshold has been shown to adequately represent detectability changes with time (even deteriorating changes) and detectability changes in space. Using this detection threshold to correct observed seismicity rates utilises a much larger proportion of the dataset, and produces a linear magnitude-frequency plot over a larger magnitude range, than is possible with existing previous techniques. This in turn provides more accurate estimates of the a and b-values.

The minimal amount of subjective input means the results from one area can be compared directly with the results from another.

Application of the technique to eastern Queensland shows that previous seismicity models have all overestimated the recorded seismicity.

Introduction

Recent attempts to quantify earthquake hazard in Australia have used a probabilistic approach as developed by Cornell (1968) and McGuire (1976). In this approach earthquake source zones are assigned seismicity rates with the assumption that seismicity is distributed uniformly within each zone. The analysis has typically involved a two-step process of first defining a seismic source zone and then calculating the seismicity rates of that zone.

Calculation of the seismicity rates requires corrections to the recorded dataset for non-uniformity. These corrections have typically been performed manually using a technique that neglects large proportions of the dataset. Australian earthquake databases are not large and discounting any large proportion of the database from analysis will compromise the results.

This paper describes a methodology of automatically processing a major proportion of the recorded seismicity to obtain both seismicity rates and source zones. There is no discussion regarding the validity or otherwise of the concept of source zones with uniform seismicity - we are simply describing a technique that can be used to assist in the analysis.

Data to demonstrate the application of the process to areas of eastern Queensland is taken from the Queensland earthquake catalog maintained by Environmental Systems & Services. This data was declustered using the event code (foreshock, mainshock, aftershock, swarm) that has been subjectively assigned by a seismologist.

Seismicity rates - theory

The power-law relationship between earthquake numbers and magnitude is exhibited in plots of the logarithm of the number of earthquakes (or rate of earthquakes) against magnitude as a straight line that is defined by an “a-value” (the earthquake rate at a defined magnitude) and a “b-value” (the gradient). Energy considerations can be used to show that this log-linear relationship must have an upper magnitude bound. There has been considerable discussion about the effect of this upper bound (and its value) but the simplest approach is to have a simple maximum magnitude above which no earthquakes occur. The effect of this on a cumulative magnitude-rate plot, where the number of earthquakes above a given magnitude is plotted, is to deflect the graph as magnitudes increase to the upper limit until it is asymptotic to the maximum magnitude (see Fig. 1).

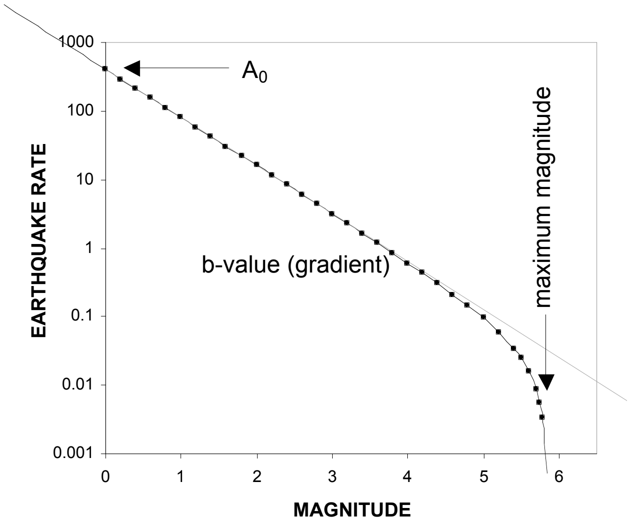

Figure 1: Theoretical cumulative magnitude-rate plot showing A_0, b-value and maximum magnitude.

An improved estimate of the b-value will be obtained if the earthquakes are spread over a large range of magnitudes. A long observation period will provide a larger number of earthquakes which will in turn reduce the uncertainty in the estimate of the rates.

In this paper all rates are expressed in terms of the annual number of earthquakes greater than a particular magnitude in an area 100 km by 100 km, and all a-values are for magnitude 0 (A_0).

Database completeness

The level of completeness of an earthquake catalog can be defined as the minimum magnitude above which all of the earthquakes in a given space-time volume have been recorded. While some earthquakes below the level of completeness may be recorded, not every one will be. This magnitude threshold will vary both in time and in space.

Spatial variation will occur because a seismograph network will only be able to record small earthquakes if they occur close to a seismograph. Earthquakes from outside a network or where the network has a large station spacing will only be recorded if they are of a larger magnitude.

Temporal variation of the magnitude detection level will occur when networks are installed and removed, when instrumentation is changed and even when operating or analysis procedures are altered.

Non-uniformity of dataset

A magnitude threshold that varies in both space and time makes the mapping of source zones and assigning of seismicity rates difficult. Variation of the detection magnitude in space will make identifying a source zone with a constant seismicity level difficult. Variation of the detection magnitude with time will make the determination of seismicity rates difficult.

Note that we are talking about non-uniformity introduced by observation methodology. The database may well contain natural variations in seismicity with time but this is a situation we choose to ignore. It is assumed that the natural variations in space are supposedly taken into account in the choice of source zones (if such a concept is indeed valid!).

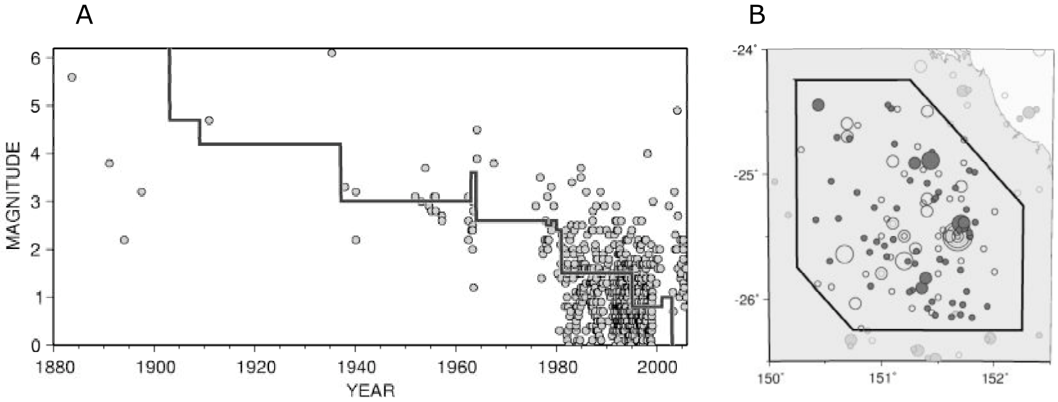

Figure 2: (A) Earthquake date plotted against magnitude for events from database within zone in southeast Queensland (shown in B). Solid line (in A) is calculated detection limit for central point of zone.

Figure 2(A) is a plot of the time of occurrence of earthquakes in a region of southeast Queensland plotted against magnitude. The area is indicated in Figure 2(B). Note that the configuration of the zone was chosen simply as a demonstration of the technique – it is not meant to represent what would be considered a final source zone. The lack of lower magnitude earthquakes in the period 1880 to 1980 indicates that the detection threshold in this region has changed over time.

Also plotted in Figure 2(A) is the theoretical detection threshold for a central point in the zone (see Network Detectability section for a description of the algorithm). The detection threshold theoretically separates a region of "constant" seismicity (in time) (above) from an area where earthquakes are less well recorded or not recorded at all (below). Note that the detection level generally improves with time but that there are instances where it has degraded.

The classical approach to analysing this dataset would be to select a series of horizontal time windows that are entirely above the detection threshold. Rates of recorded earthquakes within each band are then adjusted according to the duration. This simple approach would not include data from periods prior to any increase in the detection threshold magnitude.

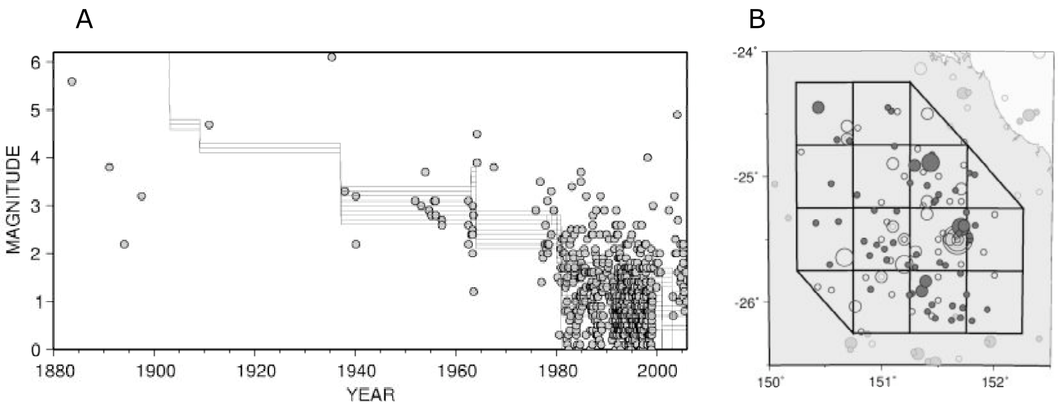

Figure 3: (A) Earthquake date plotted against magnitude for events from database within zone in southeast Queensland (shown in B). Multiple lines (in A) are calculated detection limits for central point of each sub-zone in B.

Figure 3 shows data from the same zone as Figure 2 but with the detection level versus time for a grid of points within that zone. The variation of the detectability across that zone would mean that any simple analysis would have to pick the maximum of all the possible detection curves to ensure a uniform dataset is being analysed.

Australia, and especially Queensland, has low seismicity and a short recording history so the earthquake databases are not large. Discounting any large proportion of the database from analysis will reduce the number of events available to statistically insignificant levels.

Completeness periods

There have been numerous approaches designed to automatically determine completeness periods from the shape of the magnitude-frequency plot (see Woessner and Wiemer, 2005). These techniques require a considerable amount of data and will only provide a detection threshold for a single time period. To obtain a detection threshold over time requires numerous subsets to be analysed – each one having to contain a significant number of events to be statistically useful.

The Stepp Test (Stepp, 1972) has been used in numerous studies to obtain completeness periods from recorded data. This test relies on the statistical property of the Poisson distribution – looking for periods during which the recorded earthquake rate is uniform. Figure 4 shows two Stepp Test plots (for the period up to 1999 and up to 2005) that have been plotted with earthquake rate (number / time) on the vertical axis rather than the more traditional standard deviation (square root of the rate divided by time).

This method of plotting Stepp Test results has two distinct advantages over the more traditional method. Firstly, the fitted lines are horizontal rather than sloping, allowing for easier interpretation. Secondly, the need to artificially adjust the observed rates to get estimated rates is obviated as the rates are read directly from the plot.

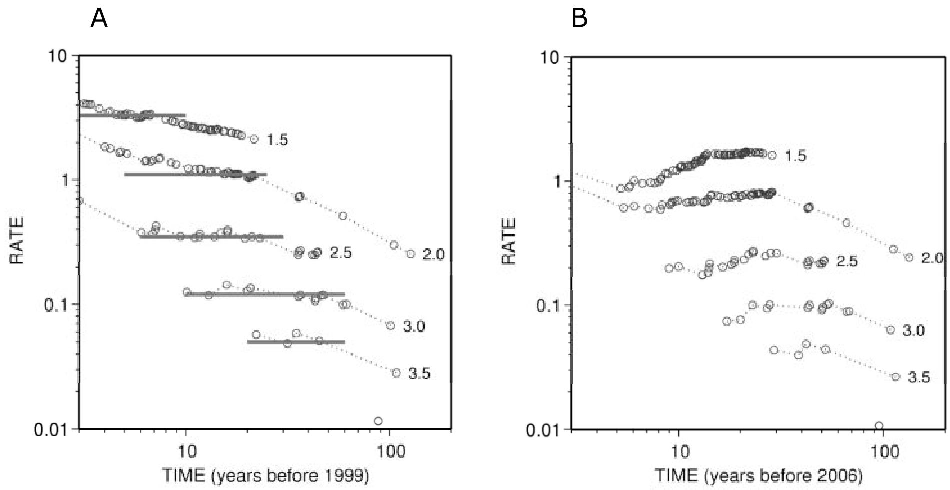

Figure 4: Modified Stepp Test plots for earthquakes in zone depicted in Figures 2 and 3. Base year for calculation is 2000 for A and 2006 for B. Labels are magnitude units. Cumulative seismicity rates (horizontal lines) have been interpreted for A.

Figure 4(A), showing data up until the end of 1999, can be fitted with horizontal lines but the data in Figure 4(B) shows no time periods where horizontal lines can be confidently drawn. During the period 2000-2003 there was a decrease in the detection capability of the network, leading to a reduction in the number of earthquakes recorded. The Stepp Test does not handle this situation at all. The only way to use the Stepp Test would be to only include data up until the end of 1999.

The results of a Stepp Test will also be compromised if there has been any natural variation in seismicity. A Stepp Test would also not be able to discern the effect of detection levels varying across a source zone.

The fitting of lines to each plot in a Stepp Test is a time-consuming process if done manually and prone to gross errors if attempted automatically.

Network detectability

Rather than analysing the recorded data in an effort to obtain detection thresholds, we used a network detection magnitude that was calculated automatically from the network configuration (in space and time), instrumentation type and crustal attenuation. Software was written so that the magnitude detection level at any location and time could be determined from these inputs.

A quality factor was subjectively assigned to each seismic station based on the instrumentation that was installed (seismometer or accelerometer), the noise level at the site and the recording technique (continuous or triggered). While a more sophisticated algorithm could be used the current simple method appears adequate.

The number of stations required to detect an earthquake was an additional parameter that could be varied. For this paper a value of one was used and shown to be suitable - i.e. an earthquake had to be only within detection distance of one station for it to be deemed detected.

For the subsequent analysis, detection levels were calculated for each one-month interval from 1900 to 2006 using the seismograph history. The attenuation function used was that used in Rynn (1987).

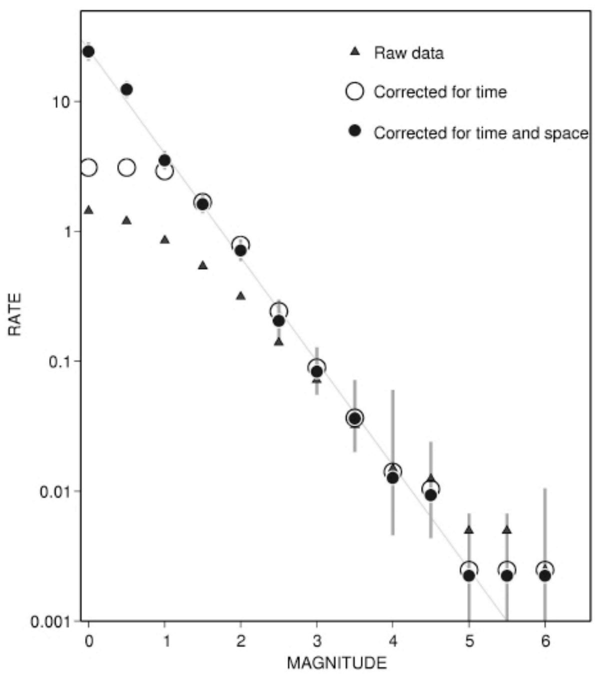

Figure 5: Magnitude-Rate plot for earthquakes in the zone depicted in Figures 2 and 3. Rates are annual numbers of earthquakes above a given magnitude per 10,000 km^2. Reference line has a b-value of 0.8. 90% confidence limits are shown for rates that have been corrected for time and space.

Calculating rates using detectability curve

Once a source zone has been mapped then rates can be calculated by only considering events that occurred above the detection threshold (and correcting for the length of time that the magnitude being considered was above the detection level).

Figure 5 shows the earthquake rates for the same area of southeast Queensland as depicted in Figures 2 and 3. Raw rates from earthquakes within the zone are shown as solid triangles while the rates corrected using the detectability curve for a central point of the zone are shown as solid black circles (The solid circles are described in the following section).

The raw data shows a linear trend for only a small section of the magnitude range but this trend extends down to magnitude ~1.5 for the data that has been corrected for detectability variations in time. The correction has also caused a marked steepening of the curve – an increase in the b-value to ~0.8 (see reference line).

Improving estimates by adding results

The process described in the previous section accounts for the variability of detection threshold with time but it does not take into account variability in space. As shown in Figure 3 there is a significant variation in detection level across the zone. This variation can be handled by dividing the source zone into multiple sub-zones (as in Figure 3) such that there is essentially no variation in detectability across each sub-zone. The same analysis as described above is then performed on each sub-zone.

Various sizes of square were chosen but the optimal trade-off (between spatial resolution and number of earthquakes) was for a square measuring ~50 x 50 km. Results from the individual sub-zones have large errors (due to the small number of events) but if they are added together (using weights based on the period of time above the detection threshold) then the errors are significantly reduced.

The results of subdividing the source zone in Figure 2(B) into 15 sub-zones (as shown in Figure 3(B)) and then summing the results are shown in Figure 5 as the solid circles (along with 90% confidence limits). The range over which a linear trend is observed is now extended down to magnitude 0.

Comparison with model estimates

Three models for the seismicity in eastern Queensland have been produced; Rynn (1987), QUAKES (Cuthbertson and Jaumé, 1996), and AUS5 (Brown and Gibson, 2000). The seismicity predicted from these models along with the results of the technique presented in this paper are shown in Figure 6(a).

As in Figure 5 there are three rates depicted, taken from; raw, recorded data with no correction for detection threshold; data corrected for detectability (but only for detectability at a single point in the centre of the region); and data corrected for detectability at every 0.5 degree within the region. While the linear trend for the data corrected for detection variability in time extends over several magnitude increments, the trend is extended all the way to magnitude 0 when the data is corrected for detection variability in space.

To accentuate the differences between the calculated rates and those predicated from the models Figure 6(B) shows exactly the same data but reduced assuming a b-value of 0.8. The Rynn and AUS5 models have b-values that do not fit the data. The b-value used in the QUAKES model appears to be more consistent with the data but all three models seriously overestimate the seismicity rates (a-values). Future hazard calculations in Queensland should ensure that there is a closer match between the corrected rates and those predicted from the source zone model.

To gain a better understanding of the distribution of seismicity the information from the multiple sub-zones can be plotted on a map. Figure 7 shows the average A_0 rate for each 0.5 degree square that have been smoothed with adjacent grid points and contoured. Only areas for which data is available are shaded and contoured. While the results are for A_0 the map is based on data from all magnitudes recorded (with the appropriate corrections).

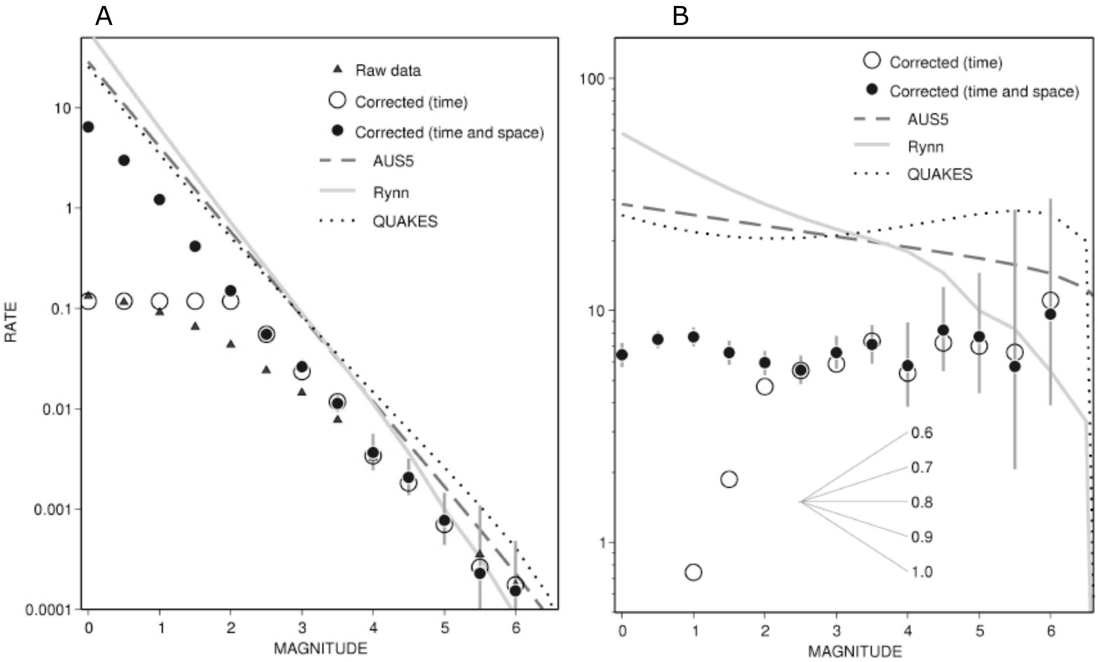

Figure 6: Magnitude-Rate plot for earthquakes in eastern Queensland together with rates predicted from seismicity models. Rates are annual numbers of earthquakes above a given magnitude per 10,000km^2. Data in Figure B has been reduced using a b-value of 0.8. Representative b-values shown for reference. 90% confidence limits are shown for the rates corrected for time and space.

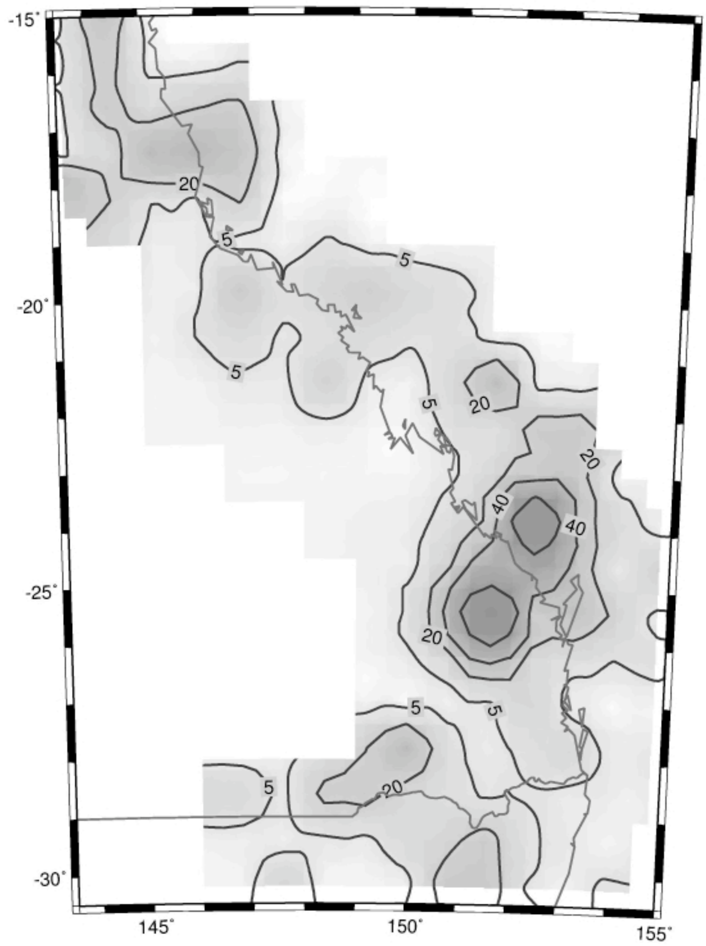

Figure 7: Map of A_0 values for eastern Queensland calculated on a 0.5 degree grid.

This map could be used as a simple, visual aid in deciding how source zones may be mapped. This would be much easier than the approach of: deciding on a zone, calculating a rate and then perhaps having to redraw the zone based on the results. It would also be much better than simply using the larger magnitude earthquakes to decide zone boundaries as it would use data from a much larger range of magnitudes.

Conclusion

An automatically calculated detection threshold has been shown to adequately represent detectability changes with time (even deteriorating changes) and detectability changes in space. Using this detection threshold to correct observed seismicity rates utilises a much larger proportion of the dataset, and produces a linear magnitude-frequency plot over a larger magnitude range, than is possible with existing previous techniques. This in turn provides more accurate estimates of the a and b-values.

The minimal amount of subjective input means the results from one area can be compared directly with the results from another.

While the techniques described in this paper appears to provide useful results, additional work needs to be done on the detection level algorithm and inputs before the technique can be applied universally. In some instances variations in the observed seismicity are not matched by variations in the detection level. This may mean a change is required in the attenuation function, the station quality code or perhaps the number of stations used.

The network history could more closely represent reality by using days instead of months and by including periods when a site was non-operational. The detection algorithm could also be modified to include a “probability of detection” rather than a simple magnitude cut-off. However these added sophistications do not seem to be required as the existing algorithm (once parameters have been properly determined) should provide useful results.

References

Brown, A. and G. Gibson (2000). Reassessment of Earthquake Hazard in Australia, 12th World Conference on Earthquake Engineering, Auckland, paper 751.

Cornell, C. A., (1968). Engineering seismic risk analysis, Bull. Seis. Soc Amer. 58, 1583-1606.

Cuthbertson, R. J. and S. C. Jaumé, (1996): Earthquake hazard in Queensland, in QUAKES Report #1, Section II, 13-39.

McGuire, R. K. (1976). FORTRAN computer program for seismic risk analysis. U. S. Geol. Surv. Open File Report, 303.

Rynn, J. M. W. (1987). Queensland Seismic Risk Study, Queensland Department of Mines, 191 pp.

Stepp, J. C. (1972). Analysis of completeness of the earthquake sample in the Puget Sound area and its effect on statistical estimates of earthquake hazard. 1st Microzonation Conference, Seattle, 1972, 897-909.

Woessner, J. & Wiemer, S. (2005). Accessing the quality of earthquake catalogs: Estimating the magnitude of completeness and its uncertainty. Bull. Seis. Soc Amer. 95, 684-698.