Estimation of the Frequency-Magnitude Gutenberg-Richter *b*-value without Knowledge of the Time-Varying Level of Completeness

Andrzej Kijko[1] and Ansie Smit[2]

[1]University of Pretoria Natural Hazard Centre, Department of Geology, University of Pretoria, Private Bag X20, Hatfield, Pretoria, 0028, South Africa. E-mail: andrzej.kijko@up.ac.za
[2]University of Pretoria Natural Hazard Centre, Department of Geology, University of Pretoria, Private Bag X20, Hatfield, Pretoria, 0028, South Africa. E-mail: ansie.smit@up.ac.za

Abstract: One of the disadvantages of the classic Gutenberg-Richter *b*-value estimator is its dependence on the applied level of completeness m_C (Aki 1965; Utsu 1965). Although several techniques have been developed for its assessment, the determination of m_C remains problematic. Kijko and Smit (2017) have attempted to provide a simple estimate for the *b*-value that is free from an assumed level of completeness. The authors have provided estimators by employing both the Method of Moments (MM) and the Maximum Likelihood Estimation (ML) for the most commonly observed shape of the apparent magnitude distribution, i.e., when it is curved gradually, as described by category IV of distributions in Mignan (2012). The methodology is extended in this paper to take into account a seismic event catalog that exhibits time-varying completeness. The proposed procedure is not restricted to any particular shape of the apparent frequency-magnitude distribution. Additionally, independent information can be incorporated using the Bayesian formalism. In instances where the applied sample of earthquake magnitudes is complete, the newly derived MM and ML *b*-value estimators take the form of the classic Utsu (1965) and Aki (1965) solutions. The inclusion of weak seismic events, and accounting for time variation in the data, can provide more reliable input parameters for earthquake hazard and risk assessments. The procedure is applied to the earthquake catalog for the Ceres–Tulbagh area, which is the most seismic active region of South Africa. This catalog spans the period 1620 to 2017, is highly incomplete, and exhibits an unknown time-varying level of completeness.

Keywords: Incompleteness; seismic source parameters; *b*-value of Gutenberg-Richter; time-varying level of completeness.

1 Introduction

The assessment of seismic hazard source (area-specific) parameters as the mean activity rate λ and the frequency-magnitude Gutenberg-Richter *b*-value, became a crucial factor after the introduction of the probabilistic seismic hazard analysis technique by Cornell (1968). However, one disadvantage of estimating these seismic hazard source parameters is that the existing methodologies depend on the level of completeness $m_{\min} \le m_C$. In Kijko and Smit (2017), the authors addressed the question of how these parameters should be assessed when no level of completeness is assumed. Derivations were provided for both the Method of Moments (MM) and the Maximum Likelihood Estimation (MLE) procedures. However, Kijko and Smit (2017) assumed that the level of completeness m_C was unknown, but was constant over the entire time span of the catalog. This paper extends the approach for an instance where m_C is unknown, but is not stationary over time.

The most often applied approach to account for the incompleteness of a seismic event catalog is by assuming that seismic events are subject to random deletions that can be accounted for by the introduction of the probability of completeness $P_C(m)$. The inclusion of $P_C(m)$ is equivalent to the statement that the apparent (observed) probability density function (PDF) of earthquake magnitude m is

$$f_A(m) = cP_C(m) f_M(m) \tag{1}$$

where $f_A(m)$ is the apparent PDF, $f_M(m)$ is the 'true' PDF of earthquake magnitude, c is a normalizing coefficient that equals

$$c = 1 / \int_{m_{\min}}^{\infty} P_C(m) f_M(m) dm \tag{2}$$

and $m_{\min}$ is the smallest observed magnitude in the catalog. Let us assume that $f_M(m)$ is described by the Gutenberg-Richter relation of the form (Aki 1965)

$$f_M(m) = \beta \exp\left[-\beta\left(m - m_C\right)\right] \qquad m \ge m_C \tag{3}$$

Proceedings of the 7th International Symposium on Geotechnical Safety and Risk (ISGSR)
Editors: Jianye Ching, Dian-Qing Li and Jie Zhang

Published by Research Publishing, Singapore.
ISBN: 978-981-11-2725-0; doi:10.3850/978-981-11-2725-0_IS11-2-cd

with $\beta = b / \ln 10$. For most seismic event catalogs, the m_C is generally higher than $m_{\min}$ because of spatial and temporal variations and/or non-uniform sensitivity in the seismograph array.

A schematic illustration of the apparent frequency-magnitude distribution $f_A(m)$ and the theoretical Gutenberg-Richter frequency-magnitude relation (Eq. 3) is provided in Figure 1. From a physical point of view, the probability of completeness $P_C(m)$ describes the departure of the apparent distribution $f_A(m)$ from the exponential nature of the Gutenberg-Richter relation (Eq. 3).

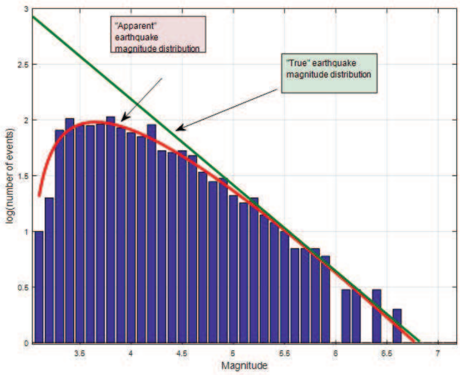

Figure 1. Schematic illustration of the apparent frequency-magnitude distribution (Eq. 1) and the theoretical Gutenberg-Richter frequency-magnitude relation (Eq. 3) (Kijko and Smit 2017).

Different authors have used various functional forms of the probability of completeness $P_C(m)$, e.g., the exponential, cumulative logistic, Pareto II, normal, and log-normal distributions. A comparison of the most often applied parametrizations of $P_C(m)$ is provided in the recent work by García-Hernández et al. (2018). Interestingly, most of these parametrizations assume that within the selected time span of the catalog, the seismicity fulfils the condition of stationarity. In other words, if a seismic event catalog is incomplete, the degree of completeness for the entire time span is the same. Only a few researchers have gone beyond this limitation and considered the assessment of seismic occurrence parameters from catalogs with an unknown level of and space–time-varying completeness. This includes work by Lee and Brillinger (1979) and Van Dyck (1985), with implementation by Stepp and King (1986) and Ogata and Katsura (1993).

In this paper, we introduce a simple and straightforward alternative method to account for the time-varying level of completeness. The parametrization $P_C(m)$ is presented in Section 2. Section 3 follows, with the extension of the original Utsu (1965) and Aki (1965) estimators, as discussed by Kijko and Smit (2017). The efficiency of the two proposed approaches is illustrated in Section 4, where the MM and MLE estimators of the mean activity rate λ and the b-value of Gutenberg-Richter are tested based on synthetic data, and the seismic event catalog for the Ceres–Tulbagh area, South Africa.

2 Parametrization of Time-Varying Incompleteness

Let us assume that the incomplete seismic event catalog can be described by an apparent set of n independent magnitude events M_i (i = 1,…n) that are all equal to, or are larger than the lowest observed magnitude $m_{\min}$ ($M_i \geq m_{\min}$). Furthermore, let us assume that the earthquake magnitudes M_i are distributed according to the apparent probability density function (Eq. 1).

As shown by Kijko and Smit (2017), if, within a specified segment of a seismic event catalog $<t, t+\Delta t>$, the probability of completeness is the same, i.e., $P_C(m, <t, t+\Delta t>) \equiv P_C(m)$, a simple and straightforward form of the b-value estimator can be obtained by

$$P_C(m, <t, t+\Delta t>|) \propto (m-\gamma)^{\alpha-1} \tag{4}$$

where the shape parameter $\alpha > 0$ and the location parameter $\gamma < m$. For simplicity of notation, we are omitting the fact that both parameters α and γ are functions of time, i.e., the correct notations would be $\alpha(t, t+\Delta t)$ and

$\gamma(t,t+\Delta t)$. As shown by García-Hernández *et al.* (2018), this parametrized probability (Eq. 4), in conjunction with the 'true' magnitude distribution $f_M(m)$, can describe a broad class of apparent frequency-magnitude distributions $f_M(m)$. Some possible shapes $P_C(m)$ are shown in Figure 2.

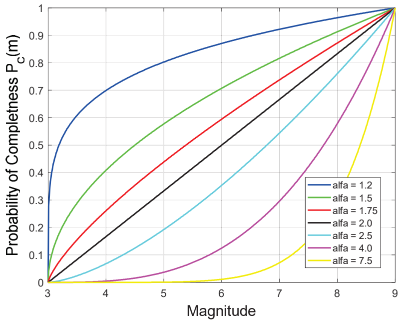

Figure 2. The probability of completeness function $P_C(m)$ can take on various shapes. For $\alpha = 2$, $P_C(m)$ is a straight line. The function converges faster for $\alpha < 2$. Highly incomplete and uncertain catalogs are characterized by $\alpha > 2$ (Kijko and Smit 2017).

Based on Eqs (1), (3), and (4), after normalization, the density distribution of the apparent magnitude $f_M(m)$ can be written as a three-parameter gamma distribution (Johnson et al, 1994)

$$f_A(x,\boldsymbol{\theta}) = \frac{\beta^{\alpha}(x-\gamma)^{\alpha-1}\exp\left[-\beta\left(x-\gamma\right)\right]}{\Gamma(\alpha)} \tag{5}$$

where $\boldsymbol{\theta} = (\alpha,\beta,\gamma)^T$, $\alpha > 0$, $\beta > 0$, $m > \gamma$, $\Gamma(\alpha)$ denotes the gamma function, and $x = m - m_{\min}$ (Abramowitz and Stegun, 1970). The presence of time interval Δt serves as a reminder that the parameters of the probability of completeness, α and γ, are a function of time *t*, and that they preserve the same values only within the segment time span $< t, t+\Delta t >$.

3 Estimation of Parameters

First, we define the estimates of the parameters in Eq. (5) using the MM and MLE methodologies, under the assumption that within a specific time segment Δt_i, the probability of incompleteness does not change (Kijko and Smit 2017). The MM estimators $\boldsymbol{\theta} = (\alpha,\beta,\gamma)^T$ are presented as

$$\begin{cases} \hat{\alpha} = 4\left[\sum_{i=1}^{n}\frac{\left(M_i-\overline{m}\right)^2}{n}\right]^3 \Bigg/ \left[\sum_{i=1}^{n}\frac{\left(M_i-\overline{m}\right)^3}{n}\right]^2 \\ \hat{\beta} = 2\left[\sum_{i=1}^{n}\left(M_i-\overline{m}\right)^2\right] \Big/ \left[\sum_{i=1}^{n}\left(M_i-\overline{m}\right)^3\right] \\ \hat{\gamma} = \sum_{i=1}^{n}\frac{\left(M_i-\overline{m}\right)^2}{n} - 2\left[\sum_{i=1}^{n}\frac{\left(M_i-\overline{m}\right)^2}{n}\right]^3 \Bigg/ \left[\sum_{i=1}^{n}\frac{\left(M_i-\overline{m}\right)^3}{n}\right]^2. \end{cases} \tag{6}$$

The corresponding MLE estimators are obtained by maximizing the likelihood function

$$L(\boldsymbol{\theta}) = \prod_{i=1}^{n}\left\{\beta^{\alpha}(x_i-\gamma)^{\alpha-1}\exp\left[-\beta\left(x_i-\gamma\right)\right]\right\} \Big/ \Gamma(\alpha) \tag{7}$$

with $\alpha > 0$, $\beta > 0$, $x > \gamma$ and $x - m - m_{\min}$. This results in

$$\begin{cases} \sum_{i=1}^{n} \ln\left(x_i - \hat{\gamma}\right) / n - \ln\left(\hat{\beta}\right) - \Gamma'\left(\hat{\alpha}\right) / \Gamma\left(\alpha\right) = 0 \\ \sum_{i=1}^{n} \left(x_i - \hat{\gamma}\right) / n - \hat{\alpha}\hat{\beta} = 0 \\ \sum_{i=1}^{n} \left(x_i - \hat{\gamma}\right)^{-1} / n - \left[\hat{\beta}\left(\hat{\alpha} - 1\right)\right]^{-1} = 0 \end{cases} . \quad (8)$$

The system of equations (Eq. 8) is solved by iterations.

Now, let us consider the instance of time-varying completeness, as illustrated in Figure 3. Let us assume that the seismic event catalog can be divided into n_S segments, each with length Δt_i, and number of events n_{Ei} (i=1,..., n_S). Within each time segment, the degree of incompleteness controlled by $P_C(x)$ is stationary. In this way, the apparent magnitude distribution, $f_A(x, \Delta t_i)$ is described by the two segment-characteristic parameters of $P_C(m)$, namely, (α_i, γ_i) and β the parameter of $f_M(m)$ (Eq. 3). Therefore, in the instance of time-varying completeness, the vector of the parameters to be estimated is of the form $\boldsymbol{\theta} = (\boldsymbol{\alpha}, \beta, \boldsymbol{\gamma})^T$, where $\boldsymbol{\alpha} = (\alpha_1,..,\alpha_{n_S})$ and $\boldsymbol{\gamma} = (\gamma_1,..,\gamma_{n_S})$.

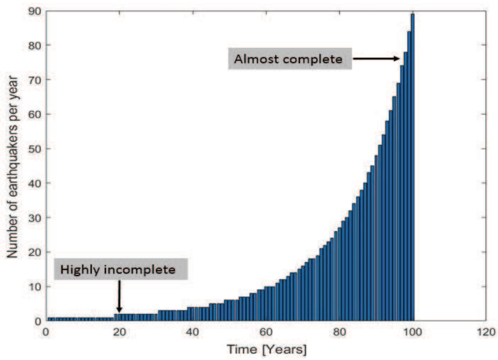

Figure 3. Example of seismic event catalog with time-varying completeness.

The MM procedure for parameter estimation is applied as follows. For each segment i (i=1,..., n_S), parameters (α_i, γ_i) are estimated by employing Eq. (6). The final estimate $\hat{\beta}$ is calculated as the weighted mean of n_S estimates of $\hat{\beta}_i$

$$\hat{\beta} = \frac{\sum_i w_i \hat{\beta}_i}{\sum_i w_i} \quad (9)$$

where w_i is the sample variance of $\hat{\beta}_i$, calculated by the bootstrap method. Similarly, the MLE estimation $\boldsymbol{\theta}$ is provided through the maximization of the likelihood function

$$L(\boldsymbol{\theta}) = \prod_{i=1}^{n_S} L(\boldsymbol{\theta}_i) \quad (10)$$

where

$$L(\boldsymbol{\theta}_i) = \prod_{j=1}^{n_{Ei}} f_A(x_j, \boldsymbol{\theta}_i) \quad (11)$$

with $f_A(x_j,\theta_i) = \beta^{\alpha_i}(x_j-\gamma_j)^{\alpha_i-1}\exp\left[-\beta\left(x_j-\gamma_j\right)\right]/\Gamma(\alpha_i)$, $\theta_i=(\alpha_i,\beta,\gamma_i)^T$, $i=1,\ldots,n_S$, and j represents the earthquake magnitude number within segment i, $j = (1,\ldots,\ n_{Ei})$.

4 Examples

4.1 Parameter estimation using synthetic catalogs

The methodology was first tested on 100 Monte-Carlo simulated catalogs, each with a time span of 100 years. The 'true' b-value was set equal to 1.0 ($\beta \cong 2.30$), and the annual mean activity rate $\lambda(m=4.0)=10$ was generated. Parameter α , which controls the degree of completeness of the seismic event catalog, at the beginning of each catalog, was set as $\alpha = 3.0$. The final value of α , at the end of the catalog, was set as $\alpha = 1.025$. The instance where $\alpha = 1.0$ corresponds with the instance where the catalog is complete. This means that the generated catalogs are incomplete at the beginning (α = 3.0), with the degree of completeness increasing over time to α = 1.025, where the catalogs are almost complete. If the catalogs were complete, they would contain, on average, $\lambda(m=4.0)$ events/year × 100 years = 1,000 events. As the catalogs were generated for $\alpha > 1.0$, they are incomplete, and, on average, each of them contains only 1700 events. Each catalog was divided into 17 segments, each with ca. 100 events. Both MM and MLE procedures were used to estimate their parameters. The respective estimates of the b-value of Gutenberg-Richter, the mean activity rate λ , and their standard errors are shown in Table 1.

Table 1. Comparison of the b-value of Gutenberg-Richter and the mean activity rate λ estimates, obtained by the moment (MM) and maximum likelihood (MLE) methods. The 'true' values of b and λ used during the generation of the synthetic catalogs are b =1.0 and λ = 10.

	$\hat{b} \pm SE^*$	$\hat{\lambda}(m=4.0) \pm SE^*$
Moment Method (MM)	1.04 ± 0.06	8.3 ± 3.7
Maximum Likelihood Method (ML)	1.09 ± 0.09	10.8 ± 1.7

* Standard error

4.2 Parameter estimation for the Ceres–Tulbagh area in South Africa

The south-western part of South Africa, particularly the area within 100 km of Cape Town, experience the highest levels of tectonic seismicity. The two largest events observed in this area are the M_L 6.3 magnitude event of 4 December 1809 located at the Milnerton Fault, and the 29 September 1969 magnitude M_L 6.3 event in the Ceres-Tulbagh area, less than 100 km from Cape Town (Green and McGarr 1972). The seismic event catalog used in this study consists of 282 seismic events, with magnitudes in the range <3.0, 6.3>. The catalog starts on 1 January 1690 and ends on 1 January 2017. Events within a radius of 450 km from the anticipated epicentre of the Ceres–Tulbagh earthquake 33.28˚S and 19.70˚E, (Kijko et al. 2003) were included in the analysis.

Table 2 provides the b-value and the mean activity rate λ estimates using the MM and MLE procedures, as applied to the Ceres–Tulbagh area. The results for the b-value, provided by both methods, are close to 0.96 and relate well with the expected b-value characteristic for stable cratonic areas located at a distance from current and recent plate boundaries (Fenton et al. 2006).

Table 2. Comparison of the b-value of Gutenberg-Richter and the mean activity rate λ estimates, obtained by moment (MM) and maximum likelihood (ML) methods and applied to the Ceres–Tulbagh area in South Africa.

	$\hat{b} \pm SE^*$	$\hat{\lambda}(m=3.0) \pm SE^*$
Moment Method (MM)	1.01 ± 0.04	4.82 ± 0.02
Maximum Likelihood Method (ML)	0.95 ± 0.09	4.16 ± 0.35

* Standard error

5 Conclusions

Two methods to estimate the frequency-magnitude Gutenberg-Richter b-value are presented, without knowledge of the time-varying level of completeness. The two estimators, MM and MLE, are particularly effective when the apparent incomplete frequency-magnitude distribution is curved gradually, i.e., it belongs to the distributions of category IV of the classification by Mignan (2012). For the purpose of illustration, the procedure is applied to synthetic Monte-Carlo simulated data and to the highly incomplete seismic event catalog of the Ceres–Tulbagh area in South Africa, with an unknown time-varying level of completeness. For both examples, the resulting estimates for both MM and MLE are close. The results can be improved when additional, independent *a priori* information regarding the estimated parameters is available. This can be done by applying the Bayesian formalism.

Acknowledgments

This work is based on the research supported by the National Research Foundation.

References

Abramowitz, M. and Stegun, L.A. (1970). *Handbook of Mathematical Functions*, 9th edn, Dover Publishers, New York, 1046.

Aki, K. (1965). Maximum likelihood estimate of b in the formula log N=a-bM and its confidence limits. *Bulletin of the Earthquake Research Institute*, Tokyo Univ., 43, 237-239.

Brillinger, D.R. (1979). Analyzing point processes subjected to random deletions, *Canadian Journal of Statistics*, 7, 1, 21–27.

Cornell, C.A. (1968). Engineering seismic risk analysis. *Bulletin of the Seismological Society of America*, 58, 1583–1606.

Fenton C., Adams, J., and Halchuk, S. (2006). Seismic hazards assessment for radioactive waste disposal sites in regions of low seismic activity. *Geotechnical and Geological Engineering*, 24, 579–592.

García-Hernández, R., D'Auria, L., Barrancos, J., and Padilla, G.D. (2018). On the functional expression of frequency–magnitude distributions: A comprehensive statistical examination. *Bulletin of the Seismological Society of America*, 108, DOI: 10.1785/0120180197

Green, R.W.E. and McGarr, A. (1972). A comparison of the focal mechanism and aftershock distribution of the Ceres, South Africa, earthquake of September 29, 1969. *Bulletin of the Seismological Society of America*, 62, 869–871.

Johnson, N.L., Kotz, S., and Balakrishnan, N. (1994). *Continuous Univariate Distributions*, Second Edition, 1, 784, John Willey & Sons, Inc., New York.

Kijko, A., Retief, S.J.P., and Graham, G. (2003). Seismic hazard and risk assessment for Tulbagh, South Africa: Part II. Assessment of seismic risk. *Natural Hazards*, 30, 25–41, DOI: 10.1023/A:1025036409750

Kijko, A. and Smit, A. (2017). Estimation of the frequency–magnitude Gutenberg–Richter b-value without making assumptions on levels of completeness. *Seismological Research Letters*, 88(2A), 311–318.

Lee, W.H.K. and Brillinger, D.R. (1979). On Chinese earthquake history – An attempt to model an incomplete data set by point process analysis. *Pure and Applied Geophysics*, 117, 1229–1257.

Mignan, A. (2012). Functional shape of the earthquake frequency-magnitude distribution and completeness magnitude. *Journal of Geophysical Research.*, 117, B08302, DOI: 10.1029/ 2012 JB009347

Ogata, Y. and Katsura, K. (1993). Analysis of temporal and spatial heterogeneity of magnitude frequency distribution inferred from earthquake catalogues. *Geophysical Journal International*, 113, 727–738.

Stepp, J.C. and King, J.L. (1986). *Seismic Hazard Methodology for the Central and Eastern United States: Methodology,* Technical Report NP-4726, Electric Power Research Institute.

Utsu, T. (1965). A method for determining the value of b in the formula log n = a–bM showing the magnitude-frequency relation for earthquakes (with English summary). *Geophys. Bull. Hokkaido Univ*. 13, 99–103.

Van Dyck, J.F.M. (1985). Statistical analysis of earthquake catalogs. *PhD Thesis*, Massachusetts Institute of Technology, December 1985, 269.