PART 4

Earthquake recurrence parameters and maximum possible earthquake magnitudes

Introduction

The parameter "recurrence time" (also called the "recurrence interval" or "return time") is widely used for hazard assessment in seismology. The recurrence time of an earthquake is usually defined as the average number of years between occurrences of an earthquake greater or equal to a given magnitude or intensity in a particular area. Once a probabilistic model for the distribution is chosen (usually a Poissonian model), the probability of damaging earthquakes in a given time interval can be derived (Saichev and Sornette, 2007).

The "maximum possible earthquake magnitude", m_{max} is another important parameter with several definitions and references. For example, the International Commission for Large Dams (ICOLD,1992) defines the Maximum Credible Earthquake (MCE) as the largest reasonably conceivable earthquake that could occur along a recognized fault or within a geographically defined tectonic province. This value can be assessed deterministically (where earthquake magnitude is associated with length of the fault) or probabilistically (linked e.g. to a 50% probability of exceedance in a infinite time interval). The US Nuclear Regulatory Commission (USNRC) RG-1208 states that "…*to completely characterize the seismic potential for a source, it is also necessary to estimate the largest earthquake magnitude that a seismic source is capable of generating under the current tectonic regime*". This estimated magnitude defines the upper bound of the earthquake recurrence relationship. The International Atomic Energy Agency Safety Guide NS-G-3.3 (IAEA, 2002) provides many references to m_{max} and detailed guidance on its assessment.

There have been several attempts to estimate earthquake recurrence periods and m_{max} for South Africa. Fernandez and Guzman (1978) used the Gumbel type III distribution to determine earthquake recurrence parameters in terms of Modified Mercalli intensity. Only recurrence intervals for Cape Town were reported, indicating a 10% probability of exceedance of intensity 7.2, 7.6, 8.1, 8.3, 8.4, 8.6 in 50, 100, 200, 300, 400 and 500 years respectively. Shapira et al. (1989) provided an in-depth study of frequency-magnitude relationships for the whole of South Africa. The region was arbitrarily subdivided into four large sub-regions. Completeness of the earthquake catalogs was assessed using the method of Stepp (1973). Frequency-magnitude relationships were analysed taking into account incompleteness using the methodology developed by Wiechert (1980). In summary, for the mining regions, the Gutenberg-Richer *b*-value was found to be about 1.5, the level of completeness (m_{min}) to be M_L 3.4, and the maximum possible seismic event magnitude (m_{max}) to be 5.5. In other regions the *b*-value was found to be in the range of 1.0, m_{min} 4.3 and m_{max}, 6.3-7.5.

Fernandez (1994) performed a probabilistic analysis of mine tremor occurrence using both the Weibull (Hagiwara, 1974) and Poissonian distributions and found good agreement for events of magnitude 3.4 and above. He evaluated seismic safety for short and long time periods, which showed that on a daily basis there is a 95% probability that no event above magnitude 3.8 or above will occur (95% safety level). At a semi-annual basis it was found that the Far West Rand region and Klerksdorp districts have lowest safety levels (20%) for events of magnitude 3.8 and above. This level is 50% for the Free State and Rand Districts.

Research funded by Hannover Re (Kijko and Retief, 2001) looked at the potential risk faced by the insurance industry due to seismic activity. Maximum possible earthquake magnitudes were determined for each cresta zones (South Africa has been divided into 16 cresta zones based on our postal code system. Zonation according to postal zones is commonly used by the insurance industry when assessing risk. The allocation of postal codes to cresta zones can be found on www.cresta.com) and this information was translated into a probabilistic damage profile for each cresta zone. Zones that showed highest m_{max} (M_L 7.0-8.0) came from the Cape and KwaZulu Natal regions. Brandt et al (2002) developed a methodology to determine earthquake recurrence and m_{max} from a combined catalogue of micro-seismic data, regional and historical data and used earthquake data from the Cape region in a case study.

In this work, we estimate the earthquake recurrence time using a fundamentally different approach to earlier works as we begin with a detailed delineation of seismotectonic units.

Methodology

The standard assumption adopted is that the distribution of earthquakes, with respect to their size, obeys the classic Gutenberg-Richter relation

$$\log N(m) = a - b \cdot (m - m_{\min}) \tag{1}$$

where $N(m)$ is the number of earthquakes of $m \geq m_{\min}$, occurring within a specified period of time, and a and b are parameters.

By definition, the mean seismic activity rate λ is calculated as the ratio

$$\lambda = \frac{Number\ of\ earthquakes\ with\ m \geq m_{\min}}{Time\ span\ of\ observations}$$

or equivalently as

$$\lambda = \frac{n(m \geq m_{\min})}{\Delta t} \tag{2}$$

where n is the number of earthquakes of magnitude m_{min} and greater that occurred within a specified time interval Δt. Therefore the mean return period for a given earthquake exceeding m_{min} is defined as 1/λ.

For the frequency-magnitude Gutenberg-Richter relation, the respective cumulative distribution function (CDF) of magnitudes, $F_M(m)$, which is bounded from above by $m_{\max}$, is (Page, 1968), where β is bln10:

$$F_M(m) = \begin{cases} 0, & \text{for } m < m_{\min}, \\ \dfrac{1-\exp[-\beta(m-m_{\min})]}{1-\exp[-\beta(m_{\max}-m_{\min})]}, & \text{for } m_{min} \le m \le m_{\max}, \\ 1, & \text{for } m > m_{\max}, \end{cases} \quad (3)$$

The International Atomic Energy Agency (IAEA) Safety Guide NS-G-3.3 provides detailed conceptual guidance on how to assess m_{max}. Broadly two approaches are possible:

1. Statistical analysis of the magnitude–frequency recurrence relationships for earthquakes associated with a particular structure. In this study, several procedures defined in Kijko (2004) will be used to assess m_{max}. In general they are represented by the addition of an extra term to the maximum observed magnitude $m_{\max}^{obs}$:

$$m_{\max} = m_{\max}^{obs} + \int_{m_{\min}}^{m_{\max}} [F_M(m)]^n \mathrm{d}m, \quad (4)$$

Descriptions of these procedures are included in the Appendix. They are listed below:

- i. Tate-Pisarenko Procedure [TP]
- ii. Kijko-Sellevoll Procedure (Cramer's Approximation) [KSA]
- iii. Kijko-Sellevoll Procedure (Exact Solution) [KSE]
- iv. Tate-Pisarenko-Bayes Procedure [TPB]
- v. Kijko-Sellevoll-Bayes Procedure [KSB]
- vi. Robson-Whitlock Procedure [RW]
- vii. Robson-Whitlock-Cooke Procedure [RWC]

viii. Non-Parametric with Gaussian Kernel Procedure [NPG]
ix. Non-Parametric Procedure based on Order statistics [OS]
x. Procedure based on a few Largest Earthquakes [5LM]

2. Direct empirical relationships can be used when sufficient information about the seismological and geological history of the movement of a fault or structure (such as segmentation, average stress drop and fault width) is available to allow estimates to be made of the maximum rupture dimensions and/or displacements of future earthquakes. In the absence of suitably detailed data, the potential maximum magnitude of a seismogenic structure can be estimated from its total dimensions. However, to use this approach, the fraction of the total length of the structure that can move in a single earthquake must be stipulated. The fraction to be used will depend on the characteristics of the fault, in particular on its segmentation. This kind of information is not known to this detail in South Africa and this kind of approach will not be used in this study.

To assess the Gutenberg-Richter parameters, several methods are used:

1. CLASSICAL MAXIMUM LIKELIHOOD ESTIMATE METHOD

If complete magnitude data m_1, m_2,…, m_n, for earthquakes with threshold magnitude m_{min} are available, the *b*-value is typically calculated from the classical maximum likelihood estimation (Aki, 1965)

$$b = \frac{\log_{10} e}{\overline{m} - m_{\min}} \tag{5}$$

where $\overline{m}$ is the sample mean magnitude $\sum_{i=1}^{n} m_i / n$, $\log_{10} e \cong 0.4343$, m_{min} is the threshold magnitude above which the data is complete and n is number of earthquakes with magnitudes equal to or larger than m_{min}.

2. THE L_1-NORM REGRESSION ANALYSIS

Kijko (1994) provides a method to determine the earthquake parameters using the L_1-norm method. Given the sample of main earthquake magnitudes m_i, $i = 1,...,n$. All magnitudes are greater than or equal to m_{min}, where the value of the magnitude m_{min} is known as the threshold of completeness. We assume further that the magnitudes m_i, are independent, identically distributed, random values whose CDF has a known functional form $F_M(m)$. In addition, we assume that magnitudes are arranged in increasing order, that is $m_1 \le m_2 \le \ldots \le m_{n-1} \le m_n$. Let vector θ denote the unknown parameters of the $F_M(m)$, where one of the parameters is unknown maximum regional magnitude m_{max}. If, for example, the CDF of the classical frequency-magnitude of Gutenberg-Richter (eq. 3) is used, vector $\theta = (\beta, m_{\max})$. Our aim is to find the solution vector θ, denoted as $\hat{\theta}$, that minimizes the misfit function

$$\Phi(\theta) = \sum_{i=1}^{n} | F_M(m_i) - \hat{F}_M(m_i) |, \tag{6}$$

where $|\cdot|$ denotes the absolute value and $\hat{F}_M(m_i)$ is the empirical distribution function equal to $i/(n+1)$. The misfit criterion (6) is less sensitive to outliers than the classical least-squares procedure and tends to ignore the effect of large differences between the theoretical, $F_M(m_i)$, and the empirical, $\hat{F}_M(m_i)$ distribution functions. The minimization of the misfit function (6) is accomplished by a direct, (derivative free) Nelder-Mead simplex-based minimization procedure (Press *et al.,* 1994).

THE L_2-NORM REGRESSION ANALYSIS

Conceptually, the procedure is the same as the L_1 Norm regression except that the absolute values of CDF residuals, (6), are replaced by respective residuals taken to the power 2. Therefore, the respective misfit function is of the form

$$\Phi(\theta) = \sum_{i=1}^{n} [F_M(m_i) - \hat{F}_M(m_i)]^2 , \tag{7}$$

which is equivalent to the classical least-squares procedure. One has to be aware of the limitation of the above choice. The use of the least-squares technique is equivalent to the assumption that the distribution of the CDF residuals is of Gaussian nature.

Results and Discussions

Data

The starting point for the evaluation of recurrence time and m_{max} is the creation of seismotectonic units. A multidisciplinary geoscientific database was compiled for use in the creation of a seismotectonic model for South Africa (see Part 3 of this thesis). It was found that although many research milestones have been reached in each earth-science discipline, the science community still needs to fine-tune its work in order to build a seismotectonic model that is entirely data-driven. In spite of the inherent problems with the data (which is common in most regions of the world), an attempt was made to build a regional seismotectonic model with the available information. These seismotectonic units will be used as a basis for this study.

Maximum observed earthquake magnitude $m_{\max}^{obs}$

The maximum observed earthquake magnitude $m_{\max}^{obs}$ is a very important parameter for determining m_{max} according to the techniques that have been adopted in this study (refer to equation 4). The maximum observed earthquake magnitude observed in each unit (zone) is plotted in Figure 1. Overall the largest observed magnitude comes from the background zone (Zone 20). A magnitude 7.2 earthquake occurred in the Manica province in Mozambique on 22 February 2006. Since this work is limited only to the regions of South Africa, this earthquake will not be considered in any further detail. Other significant zones where large earthquakes have occurred are Zones 1, 2 and 5 which include the M_L 6.3 earthquakes of 4 December 1809, 29 September 1969 and 31 October 1919, respectively. The Koffiefontein seismotectonic unit (Zone 8) contains another significant earthquake (M_L 6.2) that occurred on 20 February 1912. In all other zones, the largest earthquake is less than M_L 6.0. In the mining regions the largest observed event was a M_L 5.3 earthquake that occurred in Zone 11 on the 9th March 2005 in Stilfontein.

Estimation of m_{max}

m_{max} was estimated where possible using several statistical procedures (refer to results in Table 1 and Table 2). These results are plotted in Figure 1 for each zone. Note that no solutions were possible for zones 3, 4, 6, 9, 14 and 19 due to insufficient data points. Within each zone, estimates of m_{max} produced by the different techniques are generally similar. Between zones, m_{max} varies significantly. As can be expected, the patterns observed for the m_{max} estimations largely follow that obtained for the $m_{\max}^{obs}$. Zone 20 contains the largest m_{max} estimates, followed by Zones 1, 2 and 5. Zones 7 and 8 also contain larger m_{max} estimates. For the mining regions (Zones 10-12), m_{max} estimates are just less than M_L 6.0.

Interpretation of the *b*-value

The cumulative number of earthquakes as a function of magnitude is plotted in Figure 2. The curve fitted to the data give an indication of the relative levels of seismic activity (looking mainly at height of the curve wrt the x-axis) and the relative rate of occurrence of large and small earthquakes (the gradient or *b*-value) within each zone. The zones that delineate mining regions (Zones 10, 11 and 12) have the highest levels of seismicity for magnitudes below 4.8. The next highest level of seismicity is found in Zone 2 (Ceres region), a zone containing tectonic seismicity, followed by Zone 5 (east coast region), Zone 8 and Zone 15.

Mean recurrence time

Using equation 2 and the maximum likelihood estimation for the *b-value* one can obtain the mean recurrence time for each zone. This is shown in Figure 3. The zones located in the mining regions (Zones 10, 11 and 12) show much steeper gradients when compared to zones located in regions with earthquakes that are tectonically related. Earthquakes of magnitude M_L 3.0-4.4 have recurrence intervals less than 1 year for the mining regions. Not much can be stated about earthquakes larger than magnitude 5.6 because they have simply not been observed in these regions.

The zones located in the regions with earthquakes that are tectonically related, are classified in terms of two features: levels of seismicity and gradient of the recurrence time graph. Two groups are distinct, one with *b*-values of 0.56-0.66 and another with *b*-values in the range of 0.7 to 1.08. The hierarchy observed in terms of seismicity levels for the first group is as follows: Zone 20, 2, 8, 1, 15 and 16. For the second group this is Zone 5, 7, 13, 18 and 17. The second group corresponds to those regions that have high levels of seismic activity but lack adequate station coverage (see Part 1, Figure 1 for distribution of monitoring stations). This could be the reason for the difference in *b*-value where there is a relatively higher number of large magnitude earthquakes compared to a smaller number of low magnitude earthquakes.

In an additional step, equations (6) and (7) above were used to assess Gutenberg-Richter parameters for each zone. This is an independent check for the maximum likelihood estimation and for obtaining additional estimates of parameters for statistical purposes. Note that assessment of these parameters is an essential initial step in the seismic hazard assessment for any given site and proper evaluation can be very beneficial in obtaining a stable result.

Figure 4 shows, where possible, the goodness of fit to the cumulative frequency–magnitude plots using three methods: maximum likelihood estimation (MLE), L1 Norm estimation (L1N) and Least squares estimation (LSQ). In a few instances no solution was obtained. It is difficult, from the results alone, to determine the best method to be used because the goodness of fit depends on quality of the data, which varies from zone to zone. For example, very good fits are obtained with agreement between methods for Zones 2, 8, 11, 12, 15 and 16. However there is some disagreement at larger magnitudes. The MLE method works well for Zone 7 and both the MLE and L1N method work well for Zone 16. Furthermore, both the L1N and the LSQ method could not be used in some cases because the m_{max} estimations were undefined (L1N – Zones 1, 7, 10, 13, 17 and 18, LSQ – Zones 1, 13 and 17).

In some cases m_{max} was estimated at some value below the $m_{\max}^{obs}$ (L1N-Zone 11 and 12, LSQ-Zones 10, 11, 12 and 16, MLE – Zone 17 and 18). This phenomenon occurs primarily in those zones containing earthquake data that deviate largely from the Gutenberg-Richter distribution. Zones 10, 11 and 12 are characterized by mining activity and the occurrence of earthquakes in this region does not follow a pure Gutenberg-Richter distribution. Some bimodal distribution needs to be considered here. The data recorded in zones 1, 7, 13, 17 and 18 are incomplete and hence deviate largely from the Gutenberg-Richter distribution. Results using all three methods are listed in Table 2.

Comparison of methods to model the frequency-magnitude statistics

The findings from this work were compared with results from previous studies: The work of Fernandez and Guzman (1978) cannot be compared directly because their results were expressed in the form of MMI Intensities and here we are only looking at earthquake magnitudes.

When compared to the work of Shapira et al (1989); the m_{max} and *b*-values for the mining regions were estimated at 5.8 (this study) and 5.5 (Shapira et al., 1989), 1 (this study) and 1.5 (Shapira et al., 1989), respectively. For the tectonic regions m_{max} and *b* values were estimated at 6.3-7.8 (this study) and 6.3-7.5 (Shapira et al., 1989), 0.6-1 (this study) and 1 (Shapira et al., 1989), respectively. These differences are largely due to differences in the zonation method and the level of completeness of the catalogue, which has improved significantly since the Shapira et al. (1989) study.

The m_{max} results are comparable with the results from the Kijko and Retief (2001) study owing to the fact that the methodologies for estimation are similar.

Conclusions

From the analysis, the following broad conclusions were obtained:

- The mining regions have higher levels of seismicity, but relatively fewer large earthquakes.
- With regard to tectonic zones, Ceres region, KZN zone and Koffiefontein (in that order) have the highest levels of seismicity
- The earthquake database is highly incomplete and variable for each zone. Consequently no one method is suitable to assess parameters related to the earthquake distribution. Each zone needs to be considered on a case-by-case basis.

- Typically m_{max} is 5.8 for the mining zones and 7.3 for tectonic zones

Future improvements to this study that are recommended include:

- Refinement of the zonation once improvement in the earth science database is made
- Redo the analysis using methods that take the incompleteness and uncertainty of the earthquake catalogs into account
- Perform a seismic hazard assessment for key cities in South Africa
- Map population risk and vulnerability zones in order to prioritize high risk areas

REFERENCES

Aki K (1965), Maximum likelihood estimate of *b* in the formula log N = a – bM and its confidence limits, *Bull. Earthq. Res. Inst.*, Univ Tokyo, **47,** 237-239**.**

Brandt MBC, Saunders I and Graham G (2002), Frequency-magnitude relationship and estimated m_{max} from micro-seismic data, and regional and historical data sets for the southern part of South Africa, *Africa Geoscience Review*, **9**, 199-210.

Fernandez LM and Guzman JA (1978), Earthquake hazard in South Africa. Council for Geoscience, Report no 1978-0102. 28 pp.

Fernandez LM (1994), The Seismic Safety of South African Gold Mines: A Probabilistic Analysis of the Mine Tremors Occurrence, Council for Geoscience. Report No: 1994-0073. 25pp

Hagiwara Y (1974), Probability of earthquake occurrence as obtained from a Weibull Distribution Analysis of crustal strain. *Tectonophysics,* **23,** 313-318.

IAEA (2002), Safety Standards Series, NS-G-3.3, International Atomic Energy Agency, Vienna.

ICOLD (1992), Selecting Seismic Parameters for Large Dams, Bulletin 72. International Commission for Large Dams.

Kijko A and Retief P (2001), Seismic Risk in South Africa, A Study of the Potential Risk faced by the SA Insurance Industry as a result of Seismic Activity, Hannover Re.

Kijko A (2004), Estimation of the maximum earthquake magnitude $m_{max.}$ *Pure Appl. Geophys*, **161**, 1-27.

Page R (1968), Aftershocks and microaftershocks, *Bull. Seism. Soc. Am.* **58**, 1131-1168.

Press, WH, Flannery BP, Teukolski SA and Vetterling WT (1994), *Numerical Recipes: The art of Scientific Programming*, Cambridge University Press, New York.

Saichev A and Sornette D (2007), Theory of Earthquake Recurrence times, *Journal of Geophysical Research*,**112**, B04313, doi: 10.1029/2006JB004536

Shapira A, Fernandez LM and Du Plessis A (1989), Frequency-magnitude relationships of southern African seismicity, *Tectonophysics*, **167**, 261-271.

Singh M, Kijko A and Durrheim R (2009), Seismotectonic Models for South Africa. Synthesis of Geoscientific Information, Problems and Way Forward, *Seismological Research Letters*, **80** (1), 71-80.

Stepp JC (1973), Analysis of completeness of the earthquake sample in the Puget Sound area, In: Contributions to Seismic Zoning. Natl. Oceanic Atmos. Adm. Tech. Rep. ERL 267-ESL30, 94pp.

Wiechert, DH (1980), Estimation of earthquake recurrence parameters from unequal observation periods for different magnitudes, *Bull. Seism. Soc. Am*, **70**, 1337-1346

U.S. Nuclear Regulatory Commission (2007). *A Performance Based Approach to Define Site Specific Ground Motion*. Regulatory Guide 1.208, U. S. Nuclear Regulatory Commission, March 2007.

Acknowledgements

This work is submitted as an MSc thesis and funded by the Council for Geoscience (CGS) and the National Research Foundation as part of the AfricaArray Project which is hosted by the University of Witwatersrand, South Africa and Penn State University, U.S.A. The CGS Seismology Unit provided the seismological information. Prof. Durrheim provided supervision. The manager of the Seismology Unit of the Council for Geoscience, Michelle Grobbelaar, is thanked for her continued support in making this work possible.

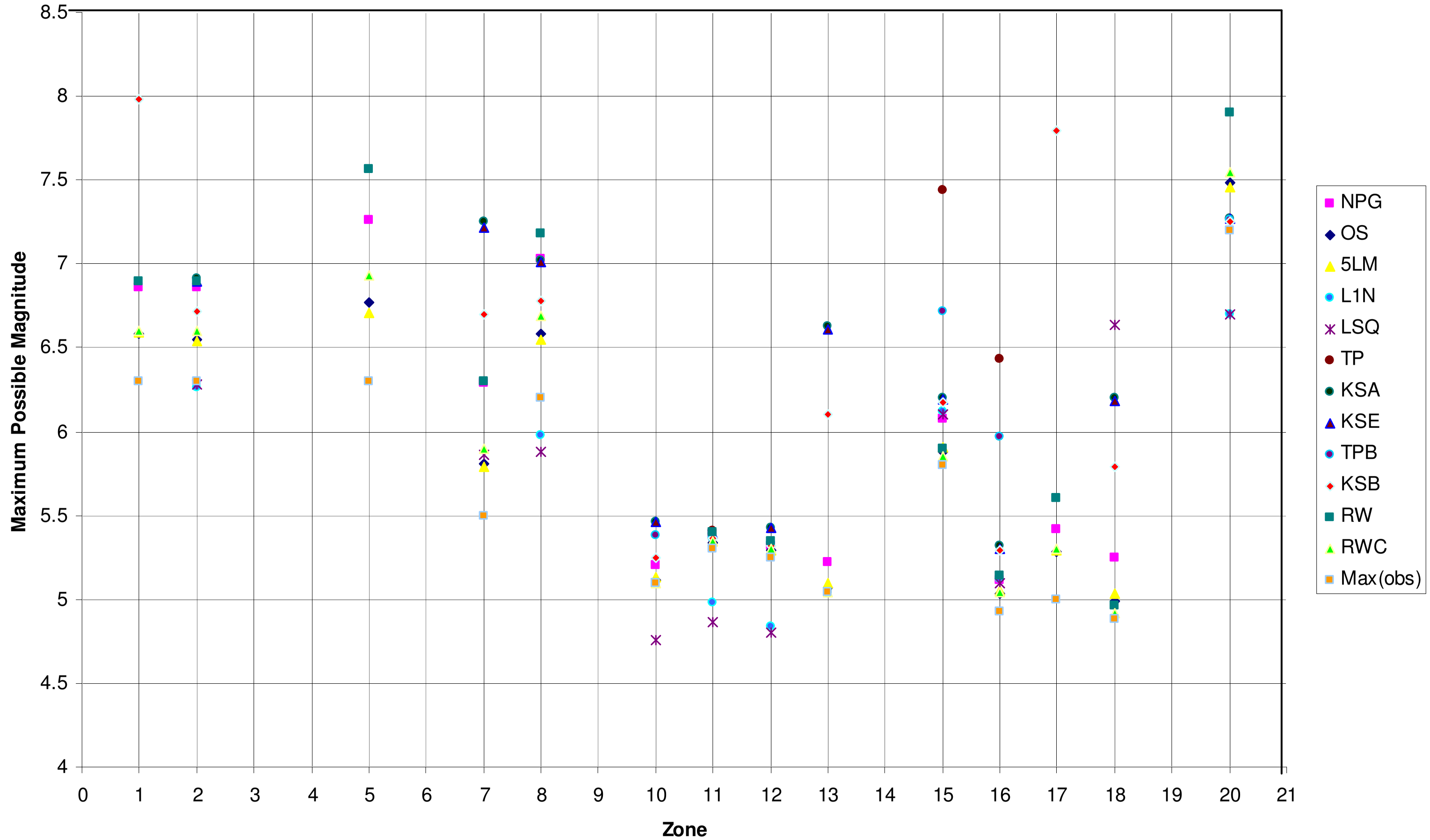

Figure 1 Ranges of maximum possible earthquake magnitude, m_{max}, obtained for seismotectonic zones using various statistical techniques

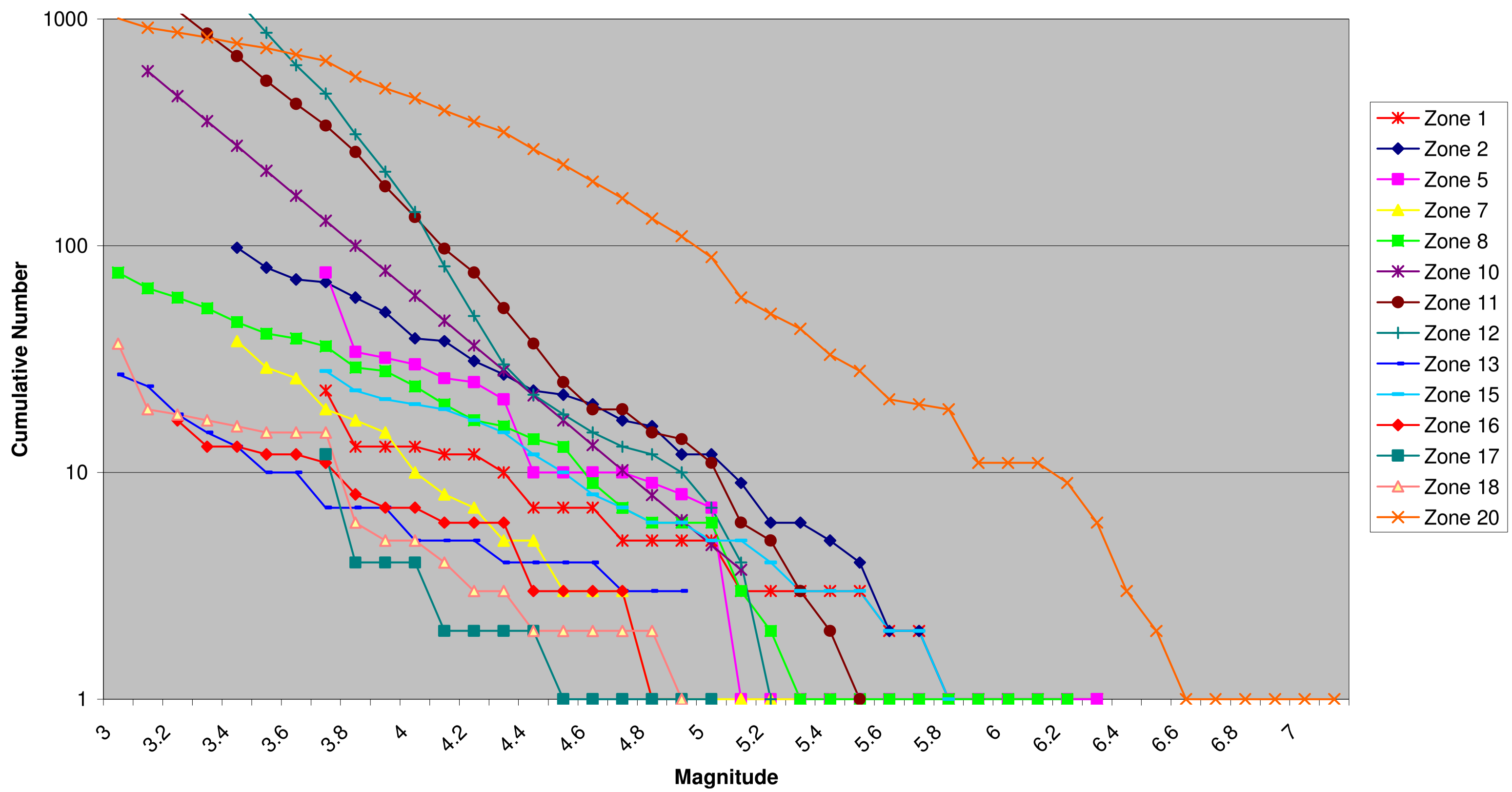

Figure 2 Frequency magnitude curves for seismotectonic zones

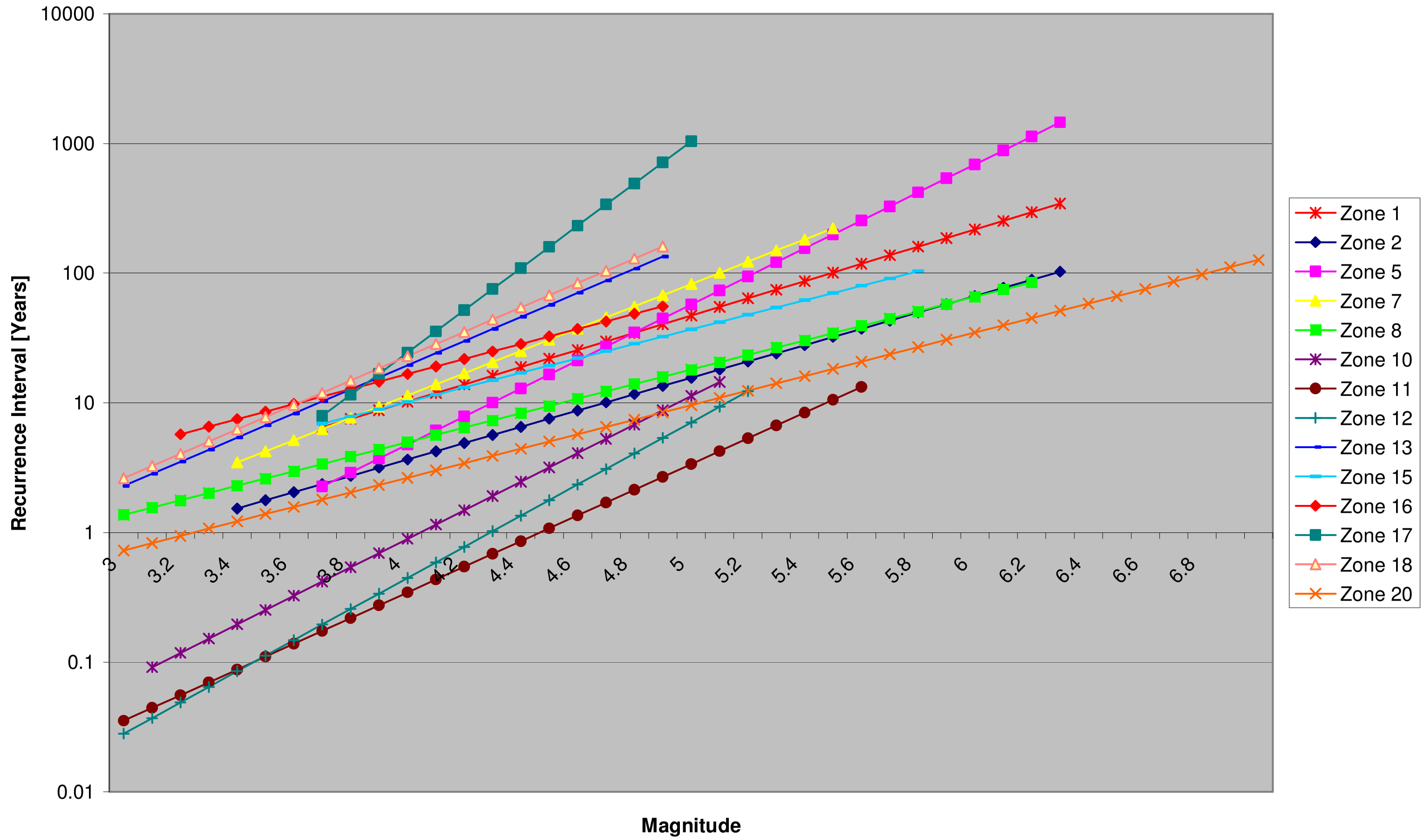

Figure 3 Calculated mean recurrence interval for specified magnitude for respective seismotectonic zones. Note that each zone has its own threshold and maximum magnitude (m_{min} and m_{max}) and the recurrence curve is drawn between these limits.

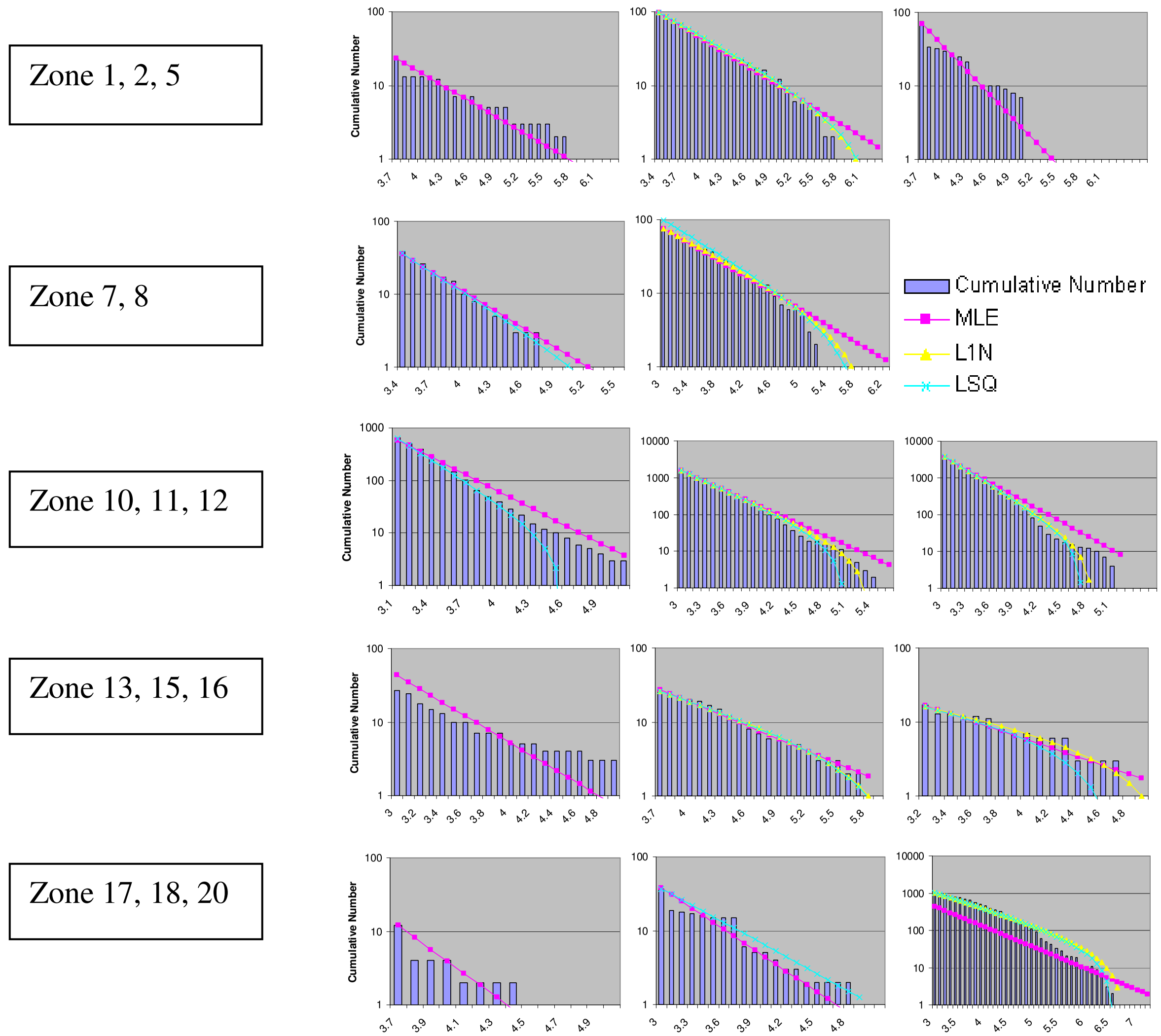

Figure 4 Goodness of fit to the cumulative frequency–magnitude plots using three methods: maximum likelihood estimation (MLE), L1 Norm estimation (L1N) and Least squares estimation (LSQ). The x-axis is magnitude.

Table 1 Estimates of m_{max} for each zone using several statistical techniques. Refer to the text for details of techniques.

	1	2	5	7	8	10	11	12	13	15	16	17	18	20
NPG	6.86±0.6	7.26±0.98	7.26±0.98	6.29±0.81	7.03±0.85	5.20±0.22	5.88±0.33	5.32±0.21	5.22±0.247	6.08±0.34	5.11±0.27	5.42±0.46	5.25±0.42	-
OS	6.58±0.39	7.26±0.98	6.77±0.55	5.81±0.42	6.58±0.47	5.11±0.28	5.73±0.26	5.29±0.28	5.07±0.28	5.88±0.29	5.03±0.3	5.28±0.4	4.99±0.3	7.48±0.39
5LM	6.59±0.38	7.26±0.98	6.71±0.48	5.79±0.38	6.55±0.42	5.13±0.24	5.34±0.26	5.29±0.24	5.10±0.25	5.91±0.26	5.05±0.27	5.29±0.37	5.03±0.28	7.46±0.35
TP	-	-	-	-	-	-	6.72±1.12	-	-	7.44±1.66	6.43±1.52	-	-	7.27±0.21
KSA	-	6.91±0.64	-	7.25±1.76	7.02±0.84	5.46±0.41	5.81±0.27	5.43±0.27	6.63±1.61	6.2±0.45	5.32±0.44	-	6.2±1.33	7.27±0.21
KSE	-	6.90±0.63	-	7.22±1.73	7.01±0.84	5.46±0.41	5.81±0.28	5.43±0.27	6.61±1.58	6.19±0.44	5.30±0.42	-	6.18±1.31	7.27±0.210.21
TPB	-	-	-	-	-	5.38±0.35	5.73±0.23	5.34±0.22	-	6.72±0.94	5.97±1.06	-	-	7.26±0.21
KSB	7.98±1.69	7.26±0.98	-	6.70±1.21	6.78±0.61	5.25±0.25	5.71±0.22	5.32±0.21	6.10±1.08	6.17±0.42	5.29±0.42	7.8±2.8	5.79±0.93	7.25±0.21
RW	6.90±0.75	7.26±0.98	7.56±1.34	6.30±0.92	7.18±1.08	5.10±0.45	5.88±0.52	5.35±0.46	5.04±0.45	5.9±0.46	5.14±0.49	5.6±0.75	4.96±0.45	7.90±0.83
RWC	6.60±0.44	7.26±0.98	6.93±0.7	5.90±0.51	6.69±0.58	5.10±0.32	5.75±0.34	5.3±0.32	5.04±0.32	5.85±0.32	5.04±0.33	5.30±0.44	4.92±0.32	7.55±0.47

Table 2 Parameters for the CDF of earthquake magnitudes obtained for each zone using three methods: maximum likelihood estimation (MLE), L1 Norm estimation (L1N) and Least squares estimation (LSQ)

	1	2	5	7	8	10	11	12	13	15	16	17	18	20
Number of events	23	106	137	58	76	760	1604	3588	43	35	19	16	37	1007
Tectonic/Mining region	T	T	T	T	T	M	M	M	M	T	T	T	T	T
m_{min}	3.7	3.4	3.7	3.4	3	3.1	3	3.1	3	3.7	3.2	3.7	3	3
$m_{\max}^{obs}$	6.3	6.3	6.3	5.5	6.2	5.1	5.3	5.0	5.0	5.8	4.9	5.0	4.9	7.2
MLE – *b*-value	0.66±0.2	0.63±0.06	1.08±0.2	0.86±0.1	0.56±0.06	1.10±0.03	0.99±0.02	1.2±0.02	0.93±0.2	0.56±0.09	0.58±0.1	1.63±0.7	0.94±0.2	0.56±0.02
MLE – *a* –value	3.82	4.13	5.84	4.47	3.56	6.18	6.17	7.15	4.43	3.51	3.08	7.11	4.4	4.32
L1N – *b*-value	-	0.56	0.8	0.9	0.49	1.35	1.00	1.35	0.72	0.42	0.28	0.75	0.76	0.42
L1N – m_{max} -value	-	6.26±0.2	-	-	5.98±0.3	-	5.34±0.34	4.84±0.46	-	6.12±0.38	5.12±0.28	-	-	6.70±0.54
LSQ – *b*-value	-	0.56	0.95	0.87	0.47	1.22	0.97	1.29	0.85	0.42	0.22	1.58	0.73	0.40
LSQ – m_{max} -value	-	6.28±0.2	-	5.86±0.41	5.88±0.38	4.76±0.39	5.10±0.56	4.80±0.49	-	6.10±0.36	5.10±0.26	-	6.64±1.77	6.70±0.54

Appendix: Procedures Considered For Determination Of Maximum Regional Earthquake Magnitude, m_{max}.

This work was first published in Kijko (2004). This work is included here for ease of reference.

Suppose that in the area of concern, within a specified time interval T, all n of the main earthquakes that occurred with a magnitude greater than or equal to $m_{\min}$ are recorded. Let us assume that the value of the magnitude $m_{\min}$ is known and is denoted as the threshold of completeness. We assume further that the magnitudes are independent, identically distributed, random values with probability density function (PDF), $f_M(m)$, and the cumulative distribution function (CDF), $F_M(m)$. The unknown parameter $m_{\max}$ is the upper limit of the range of magnitudes and is thus termed the maximum regional magnitude, and is to be estimated.

1. Tate-Pisarenko Procedure

We observe, that after the transformation $y = F_M(m)$, the CDF of the largest among (Y_1, … Y_n), that is Y_n, is equal to y^n, and its expected value $E(Y_n)$ = $n/(n+1)$. One of the possibilities for obtaining the estimator of $\hat{m}_{\max}$ is to introduce the condition $E(Y_n) = y_n$, from where we obtain the equation

$$F_M(m_{\max}^{obs}) = \frac{n}{n+1}. \qquad (1)$$

Thus, the estimator of $m_{\max}$ becomes a function of the known observations $m_{\max}^{obs}$ and n, and is obtained as a root of equation (1). The above result is valid for any CDF of earthquake magnitude $F_M(m)$, describing the complete as well as largest events, and does not require the fulfilment of the truncation condition. From (1), it is easy to derive an alternative equation, which is approximate, but shows the required value of $m_{\max}$ in a more “explicit” way

$$m_{\max} = m_{\max}^{obs} + \frac{1}{n\, f_M\,(m_{\max}^{obs})}\,. \qquad (2)$$

It should be noted that in equation (2), the desired $m_{\max}$ appears on both sides. However, from this equation an estimated value of $m_{\max}$ (denoted as $\hat{m}_{\max}$) can be obtained only by iteration.

Equation (2) is, by its nature, very general and has several interesting properties. For example, it is valid for each PDF, $f_M(m)$, and does not require the fulfilment of any additional conditions. It may also be used when the exact number of earthquakes, n, is not known. In this case, the number of earthquakes can be replaced by λT. Such a replacement is equivalent to the assumption that the number of earthquakes occurring in unit time conforms to a Poisson distribution with parameter λ, with T the span of the seismic catalogue. It is also important to note that, since the value of the correction term $\Delta = \left[n\, f_M\,(m_{\max}^{obs})\right]^{-1}$ is never negative, equation (2) provides a value of $\hat{m}_{\max}$, which is never less than the largest magnitude already observed.

The end-point estimator (2) was probably first derived by Tate (1959). If applied to the Gutenberg-Richter magnitude distribution with PDF (Page, 1968),

$$f_M\,(m) = \begin{cases} 0, & \text{for } m < m_{\min}, \\ \dfrac{\beta \exp[-\beta(m - m_{\min})]}{1 - \exp[-\beta(m_{\max} - m_{\min})]}, & \text{for } m_{min} \le m \le m_{\max}, \\ 1, & \text{for } m > m_{\max}, \end{cases} \qquad (3)$$

it takes the simple form

$$m_{\max} = m_{\max}^{obs} + \frac{1 - \exp[-\beta(m_{\max}^{obs} - m_{\min})]}{n \cdot \exp[-\beta(m_{\max} - m_{\min})]} \qquad (4),$$

where $\beta = b\ln(10)$, and b is the parameter of the frequency-magnitude Gutenberg-Richter distribution. Equation (4), with small modifications was first used first by Pisarenko *et al.* (1996). The solution of equation (4) provides the estimated value of m_{max} (denoted as $\hat{m}_{\max}$) and is called the Tate-Pisarenko estimator of m_{max}

The approximate variance of the T-P estimator of m_{max} for the frequency-magnitude Gutenberg-Richter distribution is of the form (Kijko and Graham, 1998)

$$VAR(\hat{m}_{\max}) = \sigma_M^2 + \frac{n+1}{n^3}\left[\frac{1-\exp[-\beta(m_{\max}^{obs} - m_{\min})]}{\beta\exp[-\beta(m_{\max}^{obs} - m_{\min})]}\right]^2, \tag{5}$$

where σ_M denotes the standard error in the determination of the largest observed magnitude $m_{\max}^{obs}$.

2. Kijko-Sellevoll Procedure (Cramer's Approximation)

We observe that m_n, which is the largest observed magnitude (denoted also as $m_{\max}^{obs}$), has a CDF

$$F_{M_n}(m) = \begin{cases} 0, & \text{for } m < m_{\min}, \\ \left[F_M(m)\right]^n, & \text{for } m_{\min} \le m \le m_{\max}, \\ 1, & \text{for } m > m_{\max}. \end{cases} \tag{6}$$

After integrating by parts, the expected value of M_n, $E(M_n)$, is

$$E(M_n) = \int_{m_{\min}}^{m_{\max}} m\,\mathrm{d}F_{M_n}(m) = m_{\max} - \int_{m_{\min}}^{m_{\max}} F_{M_n}(m)\,\mathrm{d}m. \tag{7}$$

Hence

$$m_{\max} = E(M_n) + \int_{m_{\min}}^{m_{\max}} \left[F_M(m)\right]^n \mathrm{d}m . \tag{8}$$

This expression, after replacement of the expected value of the largest observed magnitude, $E(M_n)$, by the largest magnitude already observed, $m_{\max}^{obs}$, provides the equation

$$m_{\max} = m_{\max}^{obs} + \int_{m_{\min}}^{m_{\max}} \left[F_M(m)\right]^n \mathrm{d}m , \tag{9}$$

in which the desired $m_{\max}$ appears on both sides. However, from this equation an estimated value of $m_{\max}$, $\hat{m}_{\max}$, can be obtained only by iteration.

Cooke (1979) was probably the first to obtain this estimator (9) of the upper bound of a random variable[1]. If applied to the assessment of the maximum regional magnitude, equation (9) states that the maximum regional magnitude $m_{\max}$ is equal to the largest magnitude already observed, $m_{\max}^{obs}$, increased by an amount $\Delta = \int_{m_{\min}}^{m_{\max}} \left[F_M(m)\right]^n \mathrm{d}m$. Similar to (2), equation (9) is, by its nature, very general and is valid for each CDF, $F_M(m)$. Of course, the drawback of the formula is that it requires integration. For some of the magnitude distribution functions the analytical expression for the integral does not exist or, if it does, it requires awkward calculations. This is, however, not a real hindrance, since numerical integration with today's high-speed computer platforms is both very fast and accurate.

[1] It should be noted that in his original paper, Cooke (1979) gave an equation in which the upper limit of integration is $m_{\max}^{obs}$ rather than $m_{\max}$. Clearly, for large n, when the value of $m_{\max}^{obs}$ and $m_{\max}$ are close to each other, the two solutions are virtually equivalent.

For the frequency-magnitude Gutenberg-Richter relation, the respective CDF of magnitudes, which are bounded from above by $m_{\max}$, is (Page, 1968):

$$F_M(m) = \begin{cases} 0, & \text{for } m < m_{\min}, \\ \dfrac{1-\exp[-\beta(m-m_{\min})]}{1-\exp[-\beta(m_{\max}-m_{\min})]}, & \text{for } m_{min} \le m \le m_{\max}, \\ 1, & \text{for } m > m_{\max}, \end{cases} \tag{10}$$

Following equation (9), the estimator of $m_{\max}$ requires the calculation of the integral

$$\Delta = \int_{m_{\min}}^{m_{\max}} \left[\frac{1-\exp\left[-\beta\left(m_{\max}^{obs}-m_{\min}\right)\right]}{1-\exp\left[-\beta\left(m_{\max}-m_{\min}\right)\right]} \right]^n \mathrm{d}m, \tag{11}$$

This integral (11) does not have a simple evaluation. It can be shown that an approximate**,** straightforward estimator of $m_{\max}$, can be obtained through the application of Cramér's approximation. According to Cramér (1961), for large n, the value of $[F_M(m)]^n$ is approximately equal to $\exp\{-n[1-F_M(m)]\}$. Simple calculations show that after replacement of $[F_M(m)]^n$ by its Cramér approximate value, integral (11) takes the form

$$\Delta = \frac{E_1(n_2)-E_1(n_1)}{\beta \exp(-n_2)} + m_{\min}\exp(-n), \tag{12}$$

where $n_1 = n/\{1-\exp[-\beta(m_{\max}-m_{\min})]\}$, $n_2 = n_1 \exp[-\beta(m_{\max}-m_{\min})]$, and $m_{\max}^{obs}$ denotes an exponential integral function. The function $E_1(\cdot)$ is defined as $E_1(z) = \int_z^{\infty} \exp(-\zeta)/\zeta \,\mathrm{d}\zeta$, and can be conveniently approximated as $E_1(z) = \dfrac{z^2 + a_1 z + a_2}{z(z^2 + b_1 z + b_2)}\exp(-z)$, where a_1 = 2.334733, a_2 = 0.250621, b_1 = 3.330657, and b_2 = 1.681534 (Abramowitz and Stegun, 1970). Hence, following equation (9), for the Gutenberg-Richter frequency-magnitude distribution, the estimator of m_{max} is obtained as a solution of the equation

$$m_{\max} = m_{\max}^{obs} + \frac{E_1(n_2) - E_1(n_1)}{\beta \exp(-n_2)} + m_{\min} \exp(-n). \tag{13}$$

It must be noted that in its current form, equation (13) does not constitute an estimator for m_{max}, since expressions n_1 and n_2, which appear on the right hand side of the equation, also contain m_{max}. In the general case, the assessment of m_{max} is obtained by the iterative solution of equation (13). However, numerical tests based on simulated data show that when m_{max} - m_{min} ≤ 2, and n ≥ 100, the parameter m_{max} in n_1 and n_2 can be replaced by $m_{\max}^{obs}$, thus providing an m_{max} estimator which can be obtained without iterations.

The approximate variance of the K-S estimator of m_{max} for the frequency-magnitude Gutenberg-Richter distribution is of the form

$$VAR(\hat{m}_{\max}) = \sigma_M^2 + \left[\frac{E_1(n_2) - E_1(n_1)}{\beta \exp(-n_2)} + m_{\min} \exp(-n) \right]^2, \tag{14}$$

where σ_M denotes standard error in the determination of the largest observed magnitude $m_{\max}^{obs}$.

A significant shortcoming of the K-S equation for m_{max} estimation comes from the implicit assumptions that (i) seismic activity remains constant in time, (ii) the selected functional form of magnitude distribution properly describes the observations, and (iii) the parameters of the assumed distribution functions are known without error.

3. Kijko-Sellevoll Procedure (Exact Solution)

Following the generic equation (9), the estimator of m_{max} requires the calculation of integral (11), an integral that does not have a simple solution. It can be shown (Dwight, 1961), under the condition that n is a positive integer, that integral (11) can be expressed as

$$\varDelta = \frac{m_{\max} - m_{\min} + \frac{1}{\beta}\sum_{i=1}^{n}\frac{(-1)^i}{i}\binom{n}{i}\left[1-\exp(-i\beta(m_{\max}-m_{\min})\right]}{\left[1-\exp(-\beta(m_{\max}-m_{\min})\right]^n}. \tag{15}$$

It follows therefore from (9) and (15), that the estimator of m_{max} is obtained as a solution of the equation

$$m_{\max} = m_{\max}^{obs} + \frac{m_{\max} - m_{\min} + \frac{1}{\beta}\sum_{i=1}^{n}\frac{(-1)^i}{i}\binom{n}{i}\left[1-\exp(-i\beta(m_{\max}-m_{\min})\right]}{\left[1-\exp(-\beta(m_{\max}-m_{\min})\right]^n}. \tag{16}$$

Thus, solution of equation (16) describes the **exact** estimator of m_{max}, when the magnitudes are distributed according to the Gutenberg-Richter relation.

The approximate variance of the exact K-S estimator of m_{max} for the frequency-magnitude Gutenberg-Richter distribution is of the form

$$VAR(\hat{m}_{\max}) = \sigma_M^2 + \left[\frac{m_{\max} - m_{\min} + \frac{1}{\beta}\sum_{i=1}^{n}\frac{(-1)^i}{i}\binom{n}{i}\left[1-\exp(-i\beta(m_{\max}-m_{\min})\right]}{\left[1-\exp(-\beta(m_{\max}-m_{\min})\right]^n}\right]^2, \tag{14}$$

where σ_M denotes standard error in the determination of the largest observed magnitude $m_{\max}^{obs}$.

4. Tate-Pisarenko-Bayes Procedure

Many studies of seismic activity suggest that the seismic process can be composed of temporal trends, cycles, short-term oscillations and pure random fluctuations. A list of some well-documented cases of temporal variation of seismic activity from all over the world is given in Kijko and Graham (1998). When the variation of seismic activity is a random process, the Bayesian formalism, in which the model parameters are treated as random variables, provides the most efficient tool in accounting for the uncertainties considered above. In this section, a Bayesian-based Tate-Pisarenko equation for the assessment of the maximum regional magnitude will be presented in which the uncertainty of the Gutenberg-Richter parameter *b* is taken into account.

Following the assumption that the variation of the *b*-value in the frequency-magnitude Gutenberg-Richter distribution may be represented by a Gamma distribution with parameters *p* and *q*, the Bayesian (also known as compound or mixed) PDF and CDF of earthquake magnitude takes the form (Campbell, 1982):

$$f_M(m) = \begin{cases} 0, & \text{for } m < m_{\min}, \\ \beta C_\beta \left(\dfrac{p}{p + m - m_{\min}} \right)^{q+1}, & \text{for } m_{min} \le m \le m_{\max}, \\ 1, & \text{for } m > m_{\max}, \end{cases} \qquad (15)$$

and

$$F_M(m) = \begin{cases} 0, & \text{for } m < m_{\min}, \\ C_\beta \left[1 - \left(\dfrac{p}{p + m - m_{\min}} \right)^{q} \right], & \text{for } m_{min} \le m \le m_{\max}, \\ 1, & \text{for } m > m_{\max}, \end{cases} \qquad (16)$$

where C_β is a normalizing coefficient. It is not difficult to show that *p* and *q* can be expressed in terms of the mean and variance of the β-value, where

$p = \overline{\beta} / (\sigma_\beta)^2$ and $q = (\overline{\beta} / \sigma_\beta)^2$. The symbol $\overline{\beta}$ denotes the known, mean value of the parameter β, σ_β is the known standard deviation of β describing its uncertainty, and C_β is equal to $\{1 - [p/(p + m_{\max} - m_{\min})]^q\}^{-1}$.

Knowledge of the Bayesian, Gutenberg-Richter PDF of earthquake magnitude (15), makes it possible to construct the Bayesian version of the T-P estimator of $m_{\max}$. Following the Tate equation (2), the Bayesian version of the T-P equation becomes

$$m_{\max} = m_{\max}^{obs} + \frac{1}{n\beta C_\beta}\left(\frac{p}{p + m_{\max}^{obs} - m_{\min}}\right)^{-(q+1)} . \qquad (17)$$

The value of $m_{\max}$ obtained from the solution of equation (17) will be termed the Tate-Pisarenko-Bayes estimator of $m_{\max}$, or, in short, T-P-B.

The approximate variance of the T-P-B estimator of $m_{\max}$ for the frequency-magnitude Gutenberg-Richter distribution is of the form

$$VAR(\hat{m}_{\max}) = \sigma_M^2 + \left[\frac{n+1}{n^3}\right]\frac{1}{(n\beta C_\beta)^2}\left(\frac{p}{p + m_{\max}^{obs} - m_{\min}}\right)^{-2(q+1)}, \qquad (18)$$

where σ_M denotes the standard error in the determination of the largest observed magnitude $m_{\max}^{obs}$.

5. Kijko-Sellevoll-Bayes Procedure

Knowledge of the Bayesian, Gutenberg-Richter PDF of earthquake magnitude (16), makes it possible to construct the Bayesian version of the K-S estimator of $m_{\max}$. Following the generic equation (9), the estimation of $m_{\max}$ requires calculation of the integral

$$\Delta = (C_\beta)^n \int_{m_{\min}}^{m_{\max}} \left[1 - \left(\frac{p}{p + m - m_{\min}} \right)^q \right]^n \mathrm{d}m, \tag{19}$$

which, after application of Cramér's approximation, can be expressed as

$$\Delta = \frac{\delta^{1/q+2} \exp[nr^q/(1-r^q)}{\beta} \left[\Gamma(-1/q, \delta r^q) - \Gamma(-1/q, \delta) \right], \tag{20}$$

where $r = p/(p + m_{\max} - m_{\min})$, $c_1 = \exp[-n(1 - C_\beta)]$, $\delta = nC_\beta$, and $\Gamma(\cdot,\cdot)$ is the Incomplete Gamma Function. Following the generic equation (9), the Bayesian version of the K-S equation determining $m_{\max}$ takes the form

$$\hat{m}_{\max} = m_{\max}^{obs} + \frac{\delta^{1/q+2} \exp[nr^q/(1-r^q)}{\beta} \left[\Gamma(-1/q, \delta r^q) - \Gamma(-1/q, \delta) \right]. \tag{21}$$

Again, as in all previous cases (equation 4, 13 and 16), equation (21) does not provide an estimator for $m_{\max}$, since some terms on the right hand side also contain $m_{\max}$. Thus, the estimator of $m_{\max}$, when the uncertainty of the Gutenberg-Richter parameter b is taken into account, can be calculated only by iteration. So the evaluated value of $m_{\max}$ will be denoted as the Kijko-Sellevoll-Bayes estimator of $m_{\max}$, or, in short, K-S-B. An extensive comparison of performances of K-S and K-S-B estimators is given in Kijko and Graham (1998).

The approximate variance of the K-S-B estimator of $m_{\max}$ for the frequency-magnitude Gutenberg-Richter distribution is of the form

$$VAR(\hat{m}_{\max}) = \sigma_M^2 + \left[\frac{\delta^{1/q+2} \exp[nr^q/(1-r^q)}{\beta} \left[\Gamma(-1/q, \delta r^q) - \Gamma(-1/q, \delta) \right] \right]^2 \tag{22}$$

where σ_M denotes standard error in the determination of the largest observed magnitude $m_{\max}^{obs}$.

6. Robson-Whitlock Procedure

Let us assume that the analytical form of the magnitude distribution $F_M(m)$ is not known and we want to estimate the right end point of the magnitude, viz. the maximum earthquake magnitude $m_{\max}$. This can be achieved in several ways. One of them is to apply the classical, (Quenouille, 1956) technique of successive bias reduction, modified to fit the factorial series rather than the power series in $1/n$. Robson and Whitlock (1964) showed that, under very general conditions, and when the data are arranged in ascending order of magnitude, viz. $m_1 \le m_2 \le ... \le m_{n-1} \le m_{\max}^{obs}$, Quenouille's approach leads to the following rule in estimation of $m_{\max}$

$$\hat{m}_{\max} = m_{\max}^{obs} + (m_{\max}^{obs} - m_{n-1}) . \tag{23}$$

Equation (23) was probably first derived by Robson and Whitlock (1964), and so it is often called the Robson and Whitlock (R-W) estimator. It can be shown that the above estimator is mean-unbiased to order n^{-2} and asymptotically median-unbiased.

The simplicity of the equation (23) makes it very attractive. It can be applied in cases of limited and/or doubtful seismic data, when one wants to get quick results without going into sophisticated analysis. Unfortunately, it can be shown, that the bias reduction is achieved at the expense of a high value of the mean squared error. In fact, Robson and Whitlock (1964) derived a general formula for an estimator of truncation point, with mean unbiased to any order n^{-k}

$$\hat{m}_{\max} = \sum_{j=0}^{k} (-1)^j \binom{k+1}{j+1} m_{n-j} \ , \tag{24}$$

where $0 < k < n$. Unfortunately, this formula does not provide a guarantee that $\hat{m}_{\max} \geq m_{\max}^{obs}$, or that the estimated maximum magnitude $\hat{m}_{\max}$ is equal to or exceeds the maximum magnitude $m_{\max}^{obs}$, already observed.

The approximate variance of the R-W estimator of $m_{\max}$ for the frequency-magnitude Gutenberg-Richter distribution is of the form

$$VAR(\hat{m}_{\max}) = 5\sigma_M^2 + (m_{\max}^{obs} - m_{n-1})^2 \ , \tag{25}$$

where σ_M denotes standard error in the determination of the largest observed magnitude $m_{\max}^{obs}$.

7. Robson-Whitlock-Cooke Procedure

Cooke (1979) showed that a significant reduction of the mean squared error of the R-W estimator is possible when some information about the shape of the upper tail of the probability distribution function $f_M(m)$ is known. Assuming that the observed magnitudes are sampled from distribution, which is truncated (as is the double truncated Gutenberg-Richter relation (3)), the improved version of the Robson-Whitlock estimator (23) is

$$\hat{m}_{\max} = m_{\max}^{obs} + 0.5(m_{\max}^{obs} - m_{n-1}) \tag{26}$$

and its variance

$$VAR(\hat{m}_{\max}) = 0.5\ [3\sigma_M^2 + 0.5(m_{\max}^{obs} - m_{n-1})^2] \ , \tag{27}$$

8. Non-Parametric with Gaussian Kernel Procedure

Most of the procedures presented in the previous sections are parametric and are applicable when the empirical log-frequency-magnitude graph for the seismic series exhibits apparent linearity, starting from a certain $m_{\min}$ value. However, many studies of seismicity show that, in some cases, (i) the empirical distributions of earthquake magnitudes are of bi- or multi-modal character, (ii) the log-frequency-magnitude relation has a strong non-linear component or (iii) the presence of "characteristic" events (Schwartz and Coppersmith, 1984) is evident.

In order to use the generic equation (9) in such cases, the analytical, parametric models of the frequency-magnitude distributions should be replaced by a non-parametric counterpart. It can be done in several ways. In this report two different approaches are presented. In this section the earthquake magnitude distribution is replaced by its approximation, constructed as a sum of the kernel functions using sample data. In the following section the magnitude distribution is replaced by it empirical counterpart based on the formalism of the order statistics (David, 1981).

The non-parametric, kernel-based approximation of PDF is an approach that deals with the direct summation of certain type of functions using sample data. Given the sample data m_i, $i = 1, \ldots, n$, and the kernel function $K(\bullet)$, the kernel estimator $\hat{f}_M(m)$ of an actual, and unknown PDF $f_M(m)$, is

$$\hat{f}_M(m) = \frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{m - m_i}{h}\right), \tag{28}$$

where h is a positive smoothing factor (Parzen, 1962). The kernel function $K(\bullet)$ is a PDF, symmetric about zero. The specific choice of it is not so important for the performance of the method; many unimodal distribution functions ensure similar efficiencies. In this work the Gaussian kernel function,

$$K(\xi) = \frac{1}{\sqrt{2\pi}} \exp\left(-\xi^2/2\right), \tag{29}$$

is used. However, the choice of the smoothing factor *h* is crucial because it affects the trade-off between random and systematic errors. Several procedures exist for the estimation of the value of this parameter, none of them being distinctly better for all varieties of real data (Silverman, 1986). For purposes of this report the least-squares cross-validation (Hall, 1983; Stone, 1984) was used. The details of the procedure are given by Kijko *et al*. (2001).

Following the functional form of a selected kernel (29) and the fact that the data comes from a finite interval $[m_{\min},\ m_{\max}]$, the respective estimators of PDF and CDF of seismic event magnitude are

$$\hat{f}_M(m) = \begin{cases} 0, & \text{for } m < m_{\min}, \\ \dfrac{\left(h\sqrt{2\pi}\right)^{-1}\sum\limits_{i=1}^{n}\exp\left[-0.5\left(\dfrac{m-m_i}{h}\right)^2\right]}{\sum\limits_{i=1}^{n}\left[\Phi\left(\dfrac{m_{\max}-m_i}{h}\right)-\Phi\left(\dfrac{m_{\min}-m_i}{h}\right)\right]}, & \text{for } m_{\min} \le m \le m_{\max}, \\ 0, & \text{for } m > m_{\max}, \end{cases}$$

(30)

and

$$\hat{F}_M(m) = \begin{cases} 0, & \text{for } m < m_{\min}, \\ \dfrac{\sum\limits_{i=1}^{n}\left[\Phi\left(\dfrac{m-m_i}{h}\right)-\Phi\left(\dfrac{m_{\min}-m_i}{h}\right)\right]}{\sum\limits_{i=1}^{n}\left[\Phi\left(\dfrac{m_{\max}-m_i}{h}\right)-\Phi\left(\dfrac{m_{\min}-m_i}{h}\right)\right]}, & \text{for } m_{\min} \le m \le m_{\max}, \\ 1, & \text{for } m > m_{\max}. \end{cases}$$

(31)

where $\Phi(\xi)$ denotes the standard Gaussian cumulative distribution function.

Despite its flexibility, a model-free technique such as the one above has been used only occasionally in seismology. One of the first uses was in the estimation of the conditional failure rates from successive recurrence times of micro-earthquakes (Rice, 1975). The non-parametric CDF of seismic event occurrence time was also employed by Sólnes *et al.* (1994). Another application involved the estimation of spatial distribution of seismic sources (Woo, 1996; Bommer *et al.,* 1997; Jackson and Kagan, 1999, and the references therein) and the non-parametric estimation of temporal variations of magnitude distributions in mines (Lasocki and Weglarczyk, 1998).

By applying the non-parametric, Gaussian-based assessment of the CDF as given by equation (9), the approximate value of the integral for Δ is

$$\Delta \cong \int_{m_{\min}}^{m_{\max}} [\hat{F}_M(m)]^n \,\mathrm{d}m = \int_{m_{\min}}^{m_{\max}} \left[\frac{\sum_{i=1}^{n} \left[\Phi\left(\frac{m-m_i}{h}\right) - \Phi\left(\frac{m_{\min}-m_i}{h}\right) \right]}{\sum_{i=1}^{n} \left[\Phi\left(\frac{m_{\max}-m_i}{h}\right) - \Phi\left(\frac{m_{\min}-m_i}{h}\right) \right]} \right]^n \mathrm{d}m. \tag{32}$$

Therefore, the equation for $m_{\max}$ based on the non-parametric Gaussian estimation of PDF takes the form

$$m_{\max} = m_{\max}^{obs} + \int_{m_{\min}}^{m_{\max}} \left[\frac{\sum_{i=1}^{n} \left[\Phi\left(\frac{m-m_i}{h}\right) - \Phi\left(\frac{m_{\min}-m_i}{h}\right) \right]}{\sum_{i=1}^{n} \left[\Phi\left(\frac{m_{\max}-m_i}{h}\right) - \Phi\left(\frac{m_{\min}-m_i}{h}\right) \right]} \right]^n \mathrm{d}m. \tag{33}$$

The value of $m_{\max}$ obtained from equation (33) will be denoted as the non-parametric, Gaussian-based estimator or, in short, N-P-G. Properties of the above estimator are discussed in the following section, where an alternative non-parametric estimator of $m_{\max}$ is presented.

Simple computations show that the approximate variance of the N-P-G estimator of maximum regional magnitude $\hat{m}_{\max}$ (33) is given by

$$Var(\hat{m}_{\max}) = \sigma_M^2 + \Delta^2, \tag{34}$$

where the correction factor Δ is described by equation (32), and the upper limit of integration, $m_{\max}$, is replaced by its estimate, $\hat{m}_{\max}$.

The drawback of formula (33) is that it requires numerical integration. However, it need not be a real obstacle, since numerical integration with today's PC's is both very fast and accurate.

9. Non-Parametric Procedure based on Order statistics

When the analytical form of the magnitude distribution is not known, it can be replaced by empirical distributions and the formalism of the order statistics can be employed. For *n* earthquake magnitudes arranged in increasing order, that is $m_1 \le m_2 \le \ldots \le m_{n-1} \le m_n$, any empirical distribution function $\hat{F}_M(m)$ can be approximated as

$$\hat{F}_M(m) = \begin{cases} 0, & \text{for } m < m_1, \\ \dfrac{i}{n}, & \text{for } \mathrm{m_i} \le m \le m_{i+1},\ i = 1, \ldots, n-1, \\ 1, & \text{for } m > m_n. \end{cases} \tag{35}$$

The approximate value of integral Δ is then

$$\begin{aligned} \varDelta \equiv \int_{m_{\min}}^{m_{\max}^{obs}} \left[\hat{F}_M(m)\right]^n \mathrm{d}m &= \sum_{i=1}^{n-1} \left(\frac{i}{n}\right)^n \left(m_{i+1} - m_i\right) \\ &= m_{\max}^{obs} - \sum_{i=0}^{n-1} \left[\left(1 - \frac{i}{n}\right)^n - \left(1 - \frac{i+1}{n}\right)^n\right] m_{n-i}. \end{aligned} \tag{36}$$

Since for large *n,* the value of $(1+1/n)^n \cong e,$ after simple algebraic rearrangements, the correction factor $\varDelta = m_{\max}^{obs} - (1-e^{-1})\sum_{i=0}^{n-1} e^{-i} m_{n-i}$, and the order-statistics based estimator of $m_{\max}$ takes the form

$$\hat{m}_{\max} = m_{\max}^{obs} + \left(m_{\max}^{obs} - (1-e^{-1})\sum_{i=0}^{n-1} e^{-i} m_{n-i}\right). \tag{37}$$

The value of $m_{\max}$ obtained from equation (37) will be denoted as the non-parametric, order statistics based estimator or, in short, N-P-OS.

Assuming that the standard error in determination of magnitude $m_1 \ldots m_n$ is known and equal to σ_M , for large *n*, the approximate variance of the order statistics based estimator (37) is equal to

$$Var(\hat{m}_{\max}) = c_0 \sigma_M^2 + \Delta^2, \tag{38}$$

where $c_0 = (1+e^{-1})^2 + e^{-2}(1-e^{-1})/(1+e^{-1}) \cong 1.93.$

Both nonparametric estimators, (eq. 33 and 37) are very useful. The great attraction of the non-parametric approach is that it does not require specification of functional form for the magnitude distribution $F_M(m)$. Therefore, by its nature, it is capable to deal with cases of empirical distributions of any complexity, distributions which considerably violate the log-linearity for e.g multimodal and/or when the "characteristic" earthquakes are present. The drawback of the estimators is that, formally they require knowledge of all events with magnitude above the specified level of completeness m_{min}, though in practice, this can reduce to knowledge of a few (say 5) of the largest events. Such a reduction is possible because the contribution of weak events with increasing index i decrease very rapidly. The procedure that provides the value of m_{max}, based on a few largest magnitudes, is discussed in the following section.

10. Procedure based on a few Largest Earthquakes

In this section a simple formula for estimation of the maximum regional magnitude is given, which can be applied in the case when only several largest earthquake magnitudes are available.

In language of mathematical statistics, the case when a known number of observations is missing from either end of the distribution is known as (single) data censoring (David, 1981). The problem of estimating the bounds of random variables when the data is censored, so only the m largest (or smallest) of the observations are available, has been discussed by Cooke (1980). Theoretical results expressed in terms of problem of the maximum regional earthquake magnitude determination, can be summarised as follows.

Suppose that in the area of concern within a specified time interval T, from n occurred main seismic events with magnitudes $m_i \geq m_{min}$ (i = 1, …, n), only the n_0 largest magnitudes are known. Following Gnedenko's condition (Gnedenko, 1943), which suggests that for very broad class of cumulative distribution functions $F_M(m)$, for m near to its upper endpoint, $F_M(m)$ is linear in m, one can assume that the estimator of m_{max} can be expressed as linear functions of the order statistics

$$\hat{m}_{\max} = \sum_{i=1}^{n_0} a_i m_{n-i+1}, \tag{39}$$

where a_i (i = 1,…, n_0) are the coefficients to be determined. Let us concentrate on the most important case from a practical point of view, viz. when the distribution of earthquake magnitudes is truncated from the top. Cooke (1980) has found that for truncated distributions, the minimization of the mean squared error of estimator (39) can be obtained when $a_1 = 1 + 1/n_0$, $a_2 = ... = a_{n_0 - 1} = 0,$ and $a_{n_0} = -1/n_0$, that is $\Delta = \left(m_{\max}^{obs} - m_{n-n_0+1}\right)/n_0$, and

$$\hat{m}_{\max} = m_{\max}^{obs} + \frac{1}{n_0}\left(m_{\max}^{obs} - m_{n-n_0+1}\right). \tag{40}$$

Probably, the greatest attraction of estimator (40) relies on its simplicity and that even for a small number of observations n_0, the estimator is nearly optimal (in sense of its mean squared errors). It comes from the fact that for large n, the largest few observations carry most of the information about its endpoint. It is interesting to notice that the value of m_{max} estimated according to formula (40) is based only on two observations: the n_0^{th} largest magnitude m_{n-n_0+1}, and the largest observed magnitude $m_{\max}^{obs}$. Clearly, the better estimator of m_{max} can be obtained by inclusion of the remaining $n_0 - 2$ largest observations. This can be done by application of the Quenouille's technique, originally developed for averaging of the bias of an estimator (Quenouille, 1956). Averaging correcting factor $\Delta = \left(m_{\max}^{obs} - m_{n-i+1}\right)/n_0$ over the $n_0 - 1$ possible choices of *i* produces

$$\Delta = \frac{1}{n_0}\left(m_{\max}^{obs} - \frac{1}{n_0 - 1}\sum_{i=2}^{n_0} m_{n-i+1}\right), \tag{31}$$

and the estimator of $m_{\max}$ equal to

$$\hat{m}_{\max} = m_{\max}^{obs} + \frac{1}{n_0}\left(m_{\max}^{obs} - \frac{1}{n_0 - 1}\sum_{i=2}^{n_0} m_{n-i+1}\right). \tag{32}$$

Assuming that the standard errors in determination of the magnitudes m_{n-i+1} are the same and equal to σ_M, the approximate variance of the estimator (32) is

$$Var(\hat{m}_{\max}) = c_0\sigma_M^2 + \Delta^2, \tag{33}$$

where $c_0 = (n_0^2 + n_0 - 1)/[n_0(n_0 - 1)]$.

The greatest attraction of estimator (32) relies on its simplicity and that it requires knowledge of magnitudes of only a few of the largest events.

11. REFERENCES

Abramowitz M and Stegun IA (1970), *Handbook of Mathematical Functions*, 9th ed. Dover Publ., New York.

Bommer J, McQueen C, Salazar W, Scott S and Woo G (1997), Spatial distribution of seismic hazard in El Salvador. In *Proc. Seminario sobre Evaluación y Mitigación del Riesgo Sísmico en el Area Centroamericana* (Universidad Centroamericana, San Salvador 1997).

Campbell KW (1982), Bayesian analysis of extreme earthquake occurrences. Part I. Probabilistic hazard model, *Bull. Seism. Soc. Am*. **72**, 1689-1705.

Cooke P (1979), Statistical inference for bounds of random variables, *Biometrika*, **66**, 367-374.

Cooke P (1980), Optimal linear estimation of bounds of random variables, *Biometrika*, **67**, 257-258.

Cramér H (1961), *Mathematical Methods of Statistics*, Princeton University Press. Princeton.

David HA (1981), *Order Statistics,* John Wiley and Sons, New York.

Dwight HB (1961), *Tables of Integrals and Other Mathematical Data,* 3d edition, The Macmillan Co., New York.

Gnedenko B (1943), Sur la distribution limite du terme maximum d'une série aléatoire, *Ann. Math.* **44**, 423-254.

Hall P (1983), *Large Sample Optimality of Least Squares Cross-validation in Density Estimation*, Ann. Statist. **11**, 1156-1174

Kagan YY (1991), Seismic moment distribution, *Geophys. J. Int.,* **106**, 123-134.

Kijko A (1994) Seismological outliers: L_1 or adaptive L_p norm application, *Bull. Seism. Soc. Am.*, **84**, 473–477

Kijko A (2004), Estimation of the maximum earthquake magnitude $m_{max.}$ *Pure Appl. Geophys*, **161**, 1-27.

Kijko A and Graham G (1998), "Parametric-Historic" procedure for probabilistic seismic hazard analysis. Part I: Assessment of maximum regional magnitude m_{max}, *Pure App. Geophys*, **152,** 413-442.

Kijko A, Lasocki S and G. Graham (2001), Nonparametric seismic hazard analysis in mines. *Pure App. Geophys*, **158,** 1655-1675.

Lasocki S and Weglarczyk S (1998), *Complex and variable magnitude distribution of microtremors induced by mining: An example from a gold mine, South Africa*. In *Abstracts XXVI General Assemb. ESC* (Tel Aviv 1998) p. 37.

Page R (1968), Aftershocks and microaftershocks, *Bull. Seism. Soc. Am*. **58**, 1131-1168.

Parzen E (1962), *On Estimation of Probability Density Function and Mode,* Annals of Mathematical Statistics **33**, 1065-1076.

Pisarenko VF (1991), Statistical evaluation of maximum possible magnitude, *Izvestiya, Earth Physics*, **27**, 757-763.

Quenouille MH (1956), Note on bias in estimation, *Biometrica*, **43**, 353-360.

Rice J (1975), *Statistical Methods of Use in Analysing Sequences of Earthquakes*, Geophys. J. R. Astr. Soc. **42**, 671-683.

Robson DS and Whitlock JH (1964), Estimation of a truncation point, *Biometrika*, **51**, 33-39.

Schwartz DP and Coppersmith KJ (1984), Fault behavior and characteristic earthquakes: examples from the Wasatch and San Andreas fault zones, *J. Geophys. Res.*, **89**, 5681-5698.

Silverman BW (1986), *Density Estimation for Statistics and Data Analysis*, Chapman and Hall, London.

Sólnes J, Halldórsson B and Sigbjörnsson R (1994), *Assessment of seismic risk based on synthetic and upgraded earthquake catalogs of Iceland*, In *Proc. 9th Int. Seminar on Earthquake Prognosites* (San José, Costa Rica 1994).

Stone CJ (1984), *An Asymptotically Optimal Window Selection Rule for Kernel Density Estimates,* Ann. Statist. **12**, 1285-1297.

Tate RF (1959), *Unbiased Estimation: Function of Location and Scale Parameters,* Ann. Math. Statist. **30**, 331-366.

Woo G (1996), *Kernel Estimation Methods for Seismic Hazard Area Source Modeling*, Bull. Seism. Soc. Am. **86**, 353-362.