Problems in Parameter Estimation in Probabilistic Seismic Hazard Analysis and some Solutions

By

Petrus Johannes Vermeulen

Submitted in partial fulfilment of the requirements for the degree

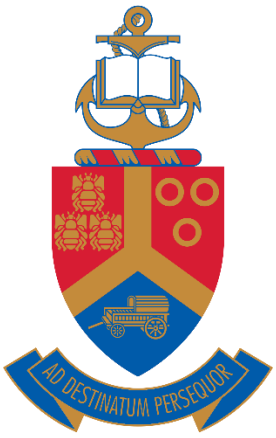
Philosophiae Doctorate (Engineering Geology)

In the Faculty of Natural and Agricultural Sciences

University of Pretoria

Pretoria

February 2020



Declaration

I, Petrus J. Vermeulen, declare that this thesis, which I hereby submit for the degree of Doctor of Philosophy (PhD) in Engineering Geology at the University of Pretoria, is my own work and has not previously been submitted by me for a degree at this or any other tertiary institution.

Signature:

Date:

Ethics Clearance

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Faculty of Natural and Agricultural Sciences
Ethics Committee

E-mail: ethics.nas@up.ac.za

ETHICS SUBMISSION: LETTER OF APPROVAL

Mr PJ Vermeulen
Department of Geology
Faculty of Natural and Agricultural Science
University of Pretoria

Reference number: NAS106/2019
Project title: Problems in parameter estimation in Probabilistic Seismic Hazard Analysis and some solutions.

Dear Mr PJ Vermeulen,

We are pleased to inform you that your submission conforms to the requirements of the Faculty of Natural and Agricultural Sciences Ethics committee.

Note that you are required to submit annual progress reports (no later than two months after the anniversary of this approval) until the project is completed. Completion will be when the data has been analysed and documented in a postgraduate student's thesis or dissertation, or in a paper or a report for publication. The progress report document is accessible on the NAS faculty's website: Research/Ethics Committee.

If you wish to submit an amendment to the application, you can also obtain the amendment form on the NAS faculty's website: Research/Ethics Committee.

The digital archiving of data is a requirement of the University of Pretoria. The data should be accessible in the event of an enquiry or further analysis of the data.

Yours sincerely,

Chairperson: NAS Ethics Committee

Acknowledgements

I sincerely thank the following people. Without your help and support, I could not have completed this study.

Professor Andrzej Kijko, my mentor and supervisor. I really cannot thank you enough for your guidance and support.

Dr Ansie Smit for your support with administration, as well as guidance and moral support.

Jannie and Melinda Vermeulen, my parents, for your love and support — morally, financially, and in countless other respects.

Professor Japie Vermeulen, my late uncle, who was a tremendous inspiration to me and who, together with my father, provided the motivation to do my doctorate.

Adri van Heerden, my editor. You did more than just editing. You guided, supported, and educated me as well.

I am also indebted to the Kreuiter family. Larry, Arja, and Andrea — life was tough during our time together, but I owe you my sincere thanks, nonetheless

I would also like to thank the following organisations that I am indebted to:

The Natural Hazard Centre of the University of Pretoria (Prof. Kijko and Dr Ansie Smit) for support and funding.

The National Research Foundation for the scholarship that funded my doctoral studies to a considerable extent.

Research outputs

- The current thesis

- Journal articles published:

Vermeulen, P.J. and Kijko, A. (2017). More statistical tools for maximum possible earthquake magnitude estimation. *Acta Geophysica* 65(4): 579–587.

Vermeulen, P.J. and Kijko, A. (2018). Discussion of "Joint Maximum Likelihood Estimators for Gutenberg–Richter Parameters λ_0 and β using Sub-Catalogs". *Earthquake Spectra*.

Articles in preparation

I am currently working on an article summarising parameter estimation in PSHA, which comprises a summary of this thesis in article format.

Conferences

Oral presentation entitled *Parameter Estimation in PSHA: Current problems and some solutions*, at the EGU annual conference held in Vienna, April 2016.

Software

I have contributed to the computational software package mmax originally created and owned by Andrzej Kijko.

Data and Resources

The United States Geological Survey Earthquakes (www.earthquake.usgs.gov/earthquakes/search, last accessed August 2019).

Abstract

Probabilistic Seismic Hazard Analysis (PSHA) is not a new study field — indeed, it dates from the late 1960s. However, the original and introductory study paid scant attention to a crucial aspect, namely the estimation of the model parameters. Consequently, over the ensuing five decades, Parameter Estimation in Probabilistic Seismic Hazard Analysis (PE-PSHA) has not gained due recognition as an independent field of study. A review of the relevant body of literature indicates that PE-PSHA is not yet regarded as an entity, a coherent body of literature, or a study field. This study aims to introduce PE-PSHA as a distinct field of study.

In 1968, Cornell introduced what is known today as Probabilistic Seismic Hazard Analysis (PSHA). Although a landmark study, it is peculiar and even astonishing that Cornell (1968) simply ignored the crucial aspect of parameter estimation of models. This aspect and the implications of ignoring the importance of parameter estimation are discussed in detail in this thesis.

Seismicity modelling in general and the classic Cornell–McGuire procedure are introduced, which provides the platform for the introduction of the parameters typically associated with it, usually referred to as *seismicity parameters*.

Subsequently, each parameter is discussed in detail, clarifying the development of estimation techniques, as well as the problem areas that could be identified. In some instances, solutions are put forward, either as own research by the author or gleaned from the literature.

A discussion is presented on the magnitude of completeness (m_c) of seismic catalogues, along with a critical analysis of the estimation techniques currently employed. Concerns about some of these methods are discussed comprehensively and clarified by detailed argument.

The two principal model parameters are discussed, namely the Gutenberg–Richter b-value and the rate of seismicity (RoS). A review of the estimation techniques of these parameters is presented, as well as the problems encountered. This review also serves as an overview of the historical development of the estimation of the two parameters. Various solutions have been put forward to some of the problems encountered; however, these solutions are not being employed. Subsequently, some estimators for the b-value for incomplete catalogues are compared.

The maximum possible earthquake magnitude for a given area (m_{max}) from the seismic catalogue data is discussed. A few procedures (or estimators) have been proposed, although only by a few researchers. The estimators are discussed in some detail and are analysed critically, among which are methods newly investigated by the author.

The concept of seismic zones is discussed, as, although seismic zones are not parameters, the delineation of seismic sources is a modelling procedure that requires estimation from the catalogue data similar to estimating parameters (this can be regarded as a generalised part of parameter estimation). The practice of seismic zoning based largely on expert opinion is analysed critically, and a number of alternatives are discussed.

In the conclusion to the study, the need for PE-PSHA to be regarded as an entity, or separate field of study, is highlighted. In addition, the discussed problems and solutions are reviewed, and recommendations are made. Finally, possible future research areas are pointed out.

Contents

i. Summary

Probabilistic Seismic Hazard Analysis (PSHA) is not a new study field — indeed, it dates from the late 1960s. However, the original and introductory study paid scant attention to a crucial aspect, namely the estimation of the model parameters. Consequently, over the ensuing five decades, Parameter Estimation in Probabilistic Seismic Hazard Analysis (PE-PSHA) has not gained due recognition as an independent field of study. A review of the relevant body of literature indicates that although numerous studies deal with the many aspects of PE-PSHA, a lack of suitable keywords, among other problems, hampers retrieval of the relevant information. This is indicative of the fact that PE-PSHA is not regarded yet as an entity, a coherent body of literature, or a distinct study field. Consequently, the problems related to parameter estimation remain obscure and, in some instances, somewhat vague. Considerable skill and determination are, therefore, required from researchers attempting to become acquainted with the field, and even experts might find it difficult to trace relevant information. Conceivably, researchers could even lose scope of the field they are making contributions to.

This study aims to introduce PE-PSHA as a distinct field of study. The thesis is formatted unconventionally, in typical monograph style. The subject is set out systematically throughout the text, and the associated problems and solutions are presented as part of the text flow without specific emphasis. However, this is not a problem, as the novel research contributions of the author are listed in the preliminary section (iii).

In 1968, Cornell introduced what is known today as Probabilistic Seismic Hazard Analysis. Although a landmark study, it is peculiar and even astonishing that Cornell (1968) simply ignored the crucial aspect of parameter estimation of models. This aspect and the implications of ignoring the importance of parameter estimation are discussed in detail in this thesis. In Chapter 1, the following are introduced, namely, seismicity modelling in general, the Cornell-McGuire procedure, and the parameters typically associated with it. Seismic catalogues are discussed briefly, as these are the data sources from which the parameters have to be estimated. Subsequently, parameter estimation and, finally, the problem, or research frontier, are discussed.

In subsequent chapters, each parameter is discussed in detail, clarifying the development of estimation techniques, as well as the problem areas that could be identified. In some instances, solutions are put forward, either as own research by the author or gleaned from the literature.

In Chapter 2, a discussion is presented on the magnitude of completeness of seismic catalogues, along with critical analysis of the estimation methods and techniques currently employed to estimate it. Concerns about some of these methods are discussed comprehensively and clarified by detailed argument.

The two principal model parameters, the Gutenberg–Richter b-value and the RoS, are discussed in Chapter 3. This chapter serves as an overview of the historical development of the estimation of the b-value and related problem areas, which have been reviewed and deliberated in several studies, although the solutions proposed are used rarely. In addition, an introduction is presented to the concept of time-varying RoS, which is crucial to PSHA but difficult to implement. Again, various solutions have been put

forward but, unfortunately, are not being used. Finally, some estimators for the b-value for incomplete catalogues are compared.

In Chapter 4, estimation of the maximum possible earthquake magnitude for a given area (m_{max}) from the seismic catalogue data is discussed. A few procedures (or estimators) have been proposed, albeit only by a few researchers. The estimators are discussed in some detail and analysed critically. Among these are methods newly investigated by the author. In Appendices I and II, as related material to this text, links are drawn between certain estimators, and the problem of the divergence of the Tate–Pisarenko estimator is addressed, along with proof of the asymptotic equivalence of Cooke's estimator and the Tate–Pisarenko estimator.

The concept of seismic zones is discussed in Chapter 5. This is not parameter estimation *per se*, but, similar to parameters, the delineation of seismic sources is a modelling procedure that requires estimation from the catalogue data. The practice of seismic zoning based largely on expert opinion is analysed critically, and a number of alternatives are discussed.

In Chapter 6, PSHA is discussed in general, from a wide perspective, i.e. beyond the Cornell–McGuire procedure. The reason for this discussion is that parameter estimation and source delineation can differ among the various approaches.

The conclusion to this work is presented in Chapter 7, Chapter 8 provides a final summary and conclusion, and the possibilities for further research are discussed in Chapter 9.

ii. Research Statement

Seismic Hazard Assessment, by definition, is the key to preventing loss of life, destruction of infrastructure, and economic losses caused by seismic activity. Probabilistic Seismic Hazard Analysis is currently the preferred method to conduct quantified, precise seismic hazard assessments. Therefore, the importance of properly carrying out PSHA is clear and can hardly be exaggerated. Yet, the numerous problems and gaps still existing in parameter estimation in PSHA appear to indicate that this study field is still in its infancy. This is an underappreciated fact, as the original work, i.e. the landmark study by Cornell (1968), did not even consider parameter estimation as part of the analysis — or, at least, that is the general impression held by practitioners. Cornell implicitly wrote only about the last part of the analysis, which could be conducted by engineers after the parameter values had been provided to them by seismologists, i.e. this implies that parameter estimation is outsourced. Today, complete hazard analysis, including parameter estimation, is typically carried out by a single team; however, the importance of parameter estimation is still being neglected and values are accepted and used with the only proviso being that they appear reasonable. Nevertheless, several studies have found that the outcome of a PSHA could be sensitive, or even extremely sensitive, to variations in parameter values (e.g. Barani *et al.*, 2007; Rabinowitz and Steinberg, 1991; Rabinowitz *et al.*, 1998; Molkenthin *et al.*, 2015; Molkenthin *et al.*, 2017;

Rohmer *et al.*, 2014; Cramer *et al.*, 1996; Joshi and Sharma, 2008; Atkinson and Charlwood, 1983; Bender and Perkins, 1993).

Accordingly, it is essential to obtain a wider perspective on the current state of parameter estimation practice and procedures and, specifically, to identify, as clearly as possible, the different strengths and weaknesses, in terms of precision and implementability, of the methods currently in use, identify the gaps and problems that remain in the study field and, most important, find viable methods and solutions to fill the gaps and solve the problems.

iii. Research Contribution

The author conducted a critical evaluation of the methods employed to estimate the level of completeness (LoC). In addition to presenting a review of such methods, the author has discovered, what appears to be, fundamental problems related to the methods of Mignan (2012), Mignan *et al.* (2011), the Maximum Curvature Method (MAXC), and the Median-Based Analysis of Segment Slope (MBASS). This is a matter of consequence, as deficiencies could detrimentally affect the outcome of hazard analysis. The analysis and resulting considerations of the author are based on concise mathematical and logical argument, i.e. rigorous scientific reasoning, implying that these are more than simply personal opinion.

The author shows that the maximum likelihood estimator of the b-value and seismicity, as presented by Ordaz and Giraldo (2018), is a special case of the Kijko–Sellevoll method (Kijko and Sellevoll, 1989). This points to the practical applicability of the Kijko–Sellevoll method.

The author conducted simulations to compare the performance of the Aki–Utsu method (Utsu, 1965; Aki, 1965), the extended Aki–Utsu method (Kijko and Smit, 2012), and two additional methods presented by Kijko (2017) to estimate the b-value. The simulations were conducted by coding in MATLAB (MathWorks, USA).

The author conducted a critical evaluation and comparison of the application of statistical estimation methods to estimate m_{max} that, to the best of the author's knowledge, although proposed and published in pure statistics, has not been implemented. These are the Generalized Least Squares Method, the Method of Moments, the method by Fraga Alves and Neves (2014), and the maximum likelihood estimate (MLE) for the Extreme Value Theory (EVT). The current study is intended to complement the study by Kijko and Singh (2011), adding additional methods. In the course of this study, MATLAB coding was done by the author, which has been made available as an updated version of the software application provided by Kijko and Singh (2011).

There is no agreed method for estimation, and no specific grounds could be discerned to use any specific method among those already described in the literature and those presented by the author. However, as is discussed later in this work, the advantages and disadvantages of the different methods are clear and

certain methods may be more applicable depending on the situation, such as, among others, the data, and the frequency-magnitude recurrence law (FMRL).

It has become apparent that the solution to the Tate–Pisarenko method sometimes diverges, and the current author has derived convergence criteria for the iterated Tate–Pisarenko method described by Kijko and Singh (2011).

The author noted a connection between Cooke's method and the EVT. Cooke's method precedes the full development and recognition of the EVT as a distinct field. This connection between Cooke's method and the EVT links the Kijko–Sellevoll estimator to the EVT and is an aspect for further investigation. Additionally, the contribution by Vermeulen and Kijko (2017) in a study establishing this link shows mathematically the asymptotic equivalence between Cooke's and the Tate–Pisarenko estimators. Accordingly, this points to the Kijko–Sellevoll method not only linking to the Tate–Pisarenko method but also to the EVT, which explains why the performance of these methods can be expected to be similar.

iv. Notation and Symbols

$P[A]$ – probability of the event (or case) A

m_{max} – maximum possible magnitude

m_c or m_{min} – magnitude of completeness

λ – rate of seismicity

v. Abbreviations

CDF: Cumulative Distribution Function

ETAS: Epidemic Type Aftershock Sequence

EVT: Extreme Value Theory

FMD: Frequency-Magnitude Distribution

FMRL: Frequency-Magnitude Recurrence Law

FMSL: Frequency-Magnitude Scaling Law

GMM: Ground Motion Model

GMPE:	Ground Motion Prediction Equation
GLS:	Generalised Least Squares
GPD:	Generalized Pareto Distribution
GR:	Gutenberg–Richter
LoC:	Level of Completeness
MAXC:	Maximum Curvature
MBASS:	Median-Based Analysis of Segment Slope
MLE:	Maximum Likelihood Estimate
MSE:	Mean Square Error
PDF:	Probability Density Function
PE-PSHA:	Parameter Estimation for Probabilistic Seismic Hazard Analysis
PH procedure:	Parametric-Historic Procedure
PSHA:	Probabilistic Seismic Hazard Analysis
RoS:	Rate of Seismicity

1. Introduction

The objective of this study is to carefully explore the study field of parameter estimation in Probabilistic Seismic Hazard Analysis (PSHA) in relation to its history, current state, and current problems. Furthermore, solutions will be presented to several of the problems that were noted, and recommendations will be suggested about others. An account of the development of the estimation of individual parameters is presented as the history of this field to date. In some instances, where problems were detected, an analysis is presented in the form of a logical or mathematical argument and, in other instances, as a comparative study of contemporary methods. In some instances, the solutions presented to the problems that were detected reflect the contribution of the current author. In addition, discussions are conducted on the methods to estimate the magnitude of completeness, Gutenberg–Richter b-value, and seismic activity rate, as well as the maximum possible magnitude for a region.

In particular, a complication is pointed out related to some of the methods employed to estimate the magnitude of completeness, and rigorous mathematics and logic are employed to prove that such weaknesses are, indeed, of concern. A quite recent closed-form solution for the simultaneous maximum likelihood estimation of the b-value and the seismic activity rate is shown to be equivalent to a special case of the Kijko–Sellevoll procedure. A comparison of several recent methods to estimate the b-value shows that these lead to quite similar results, although differences could possibly occur in extreme cases. Various methods for estimating m_{max} are discussed, including those deriving from the current research. A connection is drawn between Cooke's (1979, 1980) estimator and the EVT. The theory of the Kijko–Sellevoll estimator developed by Haraala and Orosco (2016, 2018a, 2018b) is elaborated on in detail and in terms that seismologists can relate to. Finally, Probabilistic Seismic Hazard Analysis is discussed in broader terms, considering procedures other than that of Cornell–McGuire. In general practice, the classic Cornell–McGuire procedure is conducted after estimation of the parameters but, in other methodologies, such as the Parametric-historic (P-H) procedure, parameter estimation forms an integral part of and cannot be separated from the seismic hazard analysis procedure. It is important to note here that, whichever methodology is used, parameter estimation should be considered a critical element of hazard analysis. In addition, interestingly, the Kijko–Sellevoll procedure, which is relevant to the current research, is an integral part of the P-H procedure for PSHA.

1.1. Probabilistic Seismic Hazard Analysis and the Cornell–McGuire Procedure

Earthquakes constitute some of the most destructive natural forces known. The main cause of loss and fatality is the extensive damage to buildings and other infrastructure because of the ground motion at the specific locality. Undeniably, the adage "*earthquakes don't kill people, buildings do*" is valid. It is important to understand this concept. An earthquake releasing a considerable amount of energy produces substantial ground motion *in a given region*. However, a comparatively small earthquake can also produce considerable ground motion *in the immediate vicinity*. Obviously, in either instance, no loss or fatality

might occur in a mostly unpopulated region. This implies that even severe shaking will not cause damage or loss if there are no humans or infrastructure in the vicinity. In addition, earthquakes usually do not have much effect on ecosystems and, therefore, specifically in sparsely populated areas, the slight natural damage does not affect the inhabitants much. Furthermore, only minor damage will be incurred in a region where structures and infrastructure are *designed* to resist ground motion that does not exceed a certain level. Such design is usually based on analyses of the level of ground motion that should be prepared for. Assessing the probability that a certain level of ground motion will be exceeded (referred to as the probability of exceedance) is vital in engineering design and disaster management, and to the insurance industry. Such an assessment is known as *Probabilistic Seismic Hazard Analysis*. Further, the calculated *hazard* is used to calculate the *design parameters* required for engineering structures to withstand such hazard. Or, the *risk* of the hazard combined with the structures exposed to the hazard could be calculated as the probability of a given loss being exceeded. Note that the term *hazard* is used here specifically to refer to the natural phenomenon of ground shaking unrelated to the presence of any humans or infrastructure that is put at *risk* because of the potential hazard. *Risk* refers to the probability that the existing population and infrastructure would sustain damage because of the potential hazard. Further, in hazard analysis, the term *hazard* is used interchangeably with the probability (or potential) of such hazard occurring.

The classic study by Cornell (1968) (hereinafter referred to as Cor68) on, what he referred to as engineering seismic risk analysis, marked the start of what is known today as PSHA. Practitioners distinguish between seismic hazard, which refers to the probability of ground motion occurring and ground motion levels being exceeded, and seismic risk, which refers to the probability of a specific number of casualties, the extent of damage, or amount of loss in terms of value occurring or being exceeded. This implies that hazard relates specifically to the *cause* of the possible negative effect, whereas risk relates to the negative effect itself (Kijko, 2011). The aim of PSHA is to provide an estimate of the probability of recurrence of a given ground motion parameter at a site of interest, which can be divided into five main components. These are (1) the identification, or estimative identification, of the spatial distribution of areas (or crustal volumes) of the crust of the Earth surrounding the area that is capable of producing earthquakes. Such areas are modelled typically as seismic zones or areas of equal seismic potential. (2) The average rate at which earthquake events occur at potential source spaces (or seismic zones in the typical instance) in time (referred to as the *rate of seismicity* [RoS]) of a spatial element or source zone. (3) The relative frequency with which earthquakes of various sizes occur, which has to be determined quantitatively. It is known that larger earthquakes occur less frequently than smaller earthquakes do (referred to as a *frequency-magnitude scaling law* [FMSL]). A value called the Gutenberg–Richter b-value, which is discussed later, gives an effective quantitative description of the FMSL. (4) The manner in which seismic waves attenuate with distance from the earthquake source for a given type of source mechanism. Attenuation with distance is quite complex and depends on the characteristics of the earthquake source mechanism, as well as the mechanical properties of the crustal volume through which the seismic waves travel. Such attenuation is modelled by what is termed ground motion prediction equations (GMPEs), also known as the ground motion model (GMM). For the sake of completeness, it is worth mentioning that the local amplification effect of the underlying geologic material is incorporated in a GMPE in a final term in the equation. (5) The estimation of uncertainty and variation in the GMPE models. The fact that emphasis

is placed on the estimation of the uncertainty and variation in GMPE models, and not on the quantities determined in the previous component, is arbitrary and a matter of controversy. Undeniably, uncertainty in the estimation of all quantities is of primary importance.

The RoS and the relation describing the FMSL at each source are a function of time that is not necessarily constant. In Cor68 and subsequent developments of the procedures for PSHA, the assumption was made that seismicity and the FMSL remained constant with time. In some instances, this is justified, but it must be recognised that a method is often developed from the simplest form, only to be refined later.

1.2. Rate of Seismicity (λ) and the Frequency-Magnitude Scaling Law

The RoS and FMSL are related closely and are introduced here in more detail. The RoS is merely the average number of earthquakes per unit time. Assuming that, at a probabilistic level, the RoS remains constant in time for a given area or spatial volume element, the modelling of earthquake occurrence is simply a constant, which can be referred to as λ. It appears as if it could be modelled fairly accurately by this seemingly trivial process and, indeed, in many instances, this can be done. Nevertheless, such a constant-rate random-event-producing phenomenon has intriguing and complex but useful properties. It is referred to as a Poisson point process. As regards the RoS (λ_0) for earthquakes above a given reference magnitude level (m_0) (also referred to as the rate of exceedance of m_0), to describe such a Poisson process in time, it is customary to use the probability that an earthquake exceeding m_0 would occur at least once in a time interval, which is given by

$$P[m \geq m_0] = 1 - \exp(\lambda_0 t). \tag{1.1}$$

This is considered the simplest way to capture the process and can be derived logically through fundamental principles.

The FMSL is assumed by Cor68 (and most currently practising seismologists) to follow the Gutenberg–Richter (GR) FMSL (alternately referred to as the GR law or the GR relation), defined by the relation (Gutenberg and Richter, 1944):

$$\log(n) = a - bm, \tag{1.2}$$

where n is the number of earthquakes exceeding a magnitude m within a range $[m_0, m_{max}]$ and within a given time interval, b is a scaling parameter (commonly referred to as the b-value), and a is the logarithm of the number of events greater than or equal to the reference magnitude m_0. Equation (1.2) was first established empirically by Ishimoto and Iida (1939) and popularised by Gutenberg and Richter (1944). The b-value is a crucial parameter and captures much of the information on the FMSL. Cor68, citing Isacks and

Oliver (1964), noted that the b-value typically varies between 0.65 and 1.0; however, as mentioned already, Cor68 does not refer to the issue of estimating the parameter. In the early years, seismologists were satisfied to consider $b = 1$ (Cor68). The equation (1.2) can be written conveniently as the cumulative probability distribution of earthquake sizes (Page, 1968)

$$F(m) = \begin{cases} 0, & m < m_c \\ \dfrac{1 - \exp[-\beta(m - m_c)]}{1 - \exp[-\beta(m_{max} - m_c)]}, & m_c \le m \le m_{max} \\ 1, & m > m_{max} \end{cases}, \tag{1.3}$$

here $\beta = b\ln(10)$. The values of b and β will be used interchangeably, and equation (1.3) does not appear to be simpler, but it is a proper description of the FMSL in probabilistic terms. Parameter a appears to have been lost in the process; however, this is not the case but is actually the point where the link between the RoS and FMSL becomes apparent. Careful consideration reveals that a is merely the theoretical RoS value for a cut-off magnitude of 0. The GR law implicitly incorporates the time factor, which is dropped in this probabilistic description in the formulation (1.3).

As reflected in equation (1.3), the GR FMSL does not hold indefinitely up to arbitrarily large magnitude earthquakes. At some point, there has to be a cut-off magnitude, or a maximum possible regional magnitude (m_{max}), which serves as an upper bound to the FMSL.

Given the RoS λ_{min} (i.e. the *rate of exceedance* of m_{min}), the *rate of exceedance* (RoE) λ_m can be obtained of any given magnitude m by the relation:

$$\lambda_m = \lambda_{min} P[M \ge m] = \lambda_{min}[1 - F(m)]. \tag{1.4}$$

Note the elegance of this relation between the FMSL and RoS/RoE. In addition to being elegant and simple, it is immensely important. In calculating hazard, another RoE value that is associated with ground motion is of interest here and relates to the RoS. Given the GMPE, the rate at which the logarithm of a ground motion parameter y, e.g. peak ground acceleration (pga), will be exceeded is calculated by employing the cumulative probability distribution of the logarithm of the pga, $F(y)$, and is given by

$$\lambda_{pga}(y) = \lambda_{min} P[Y \ge y] = \lambda_{min}[1 - F(y)], \tag{1.5}$$

where λ_{pga} refers to the RoE of a log(pga). Other ground motion parameters frequently used to describe ground motion are peak ground velocity, peak ground displacement, spectral acceleration, and the Modified Mercalli intensity value. Note that the Modified Mercalli intensity behaves like the logarithm of other ground motion parameters for sound physical reasons, but these reasons are beyond the scope of interest of this work. In addition to the Modified Mercalli intensity, the logarithm of ground motion

parameters is used because attenuation with distance and other parameters of the GMPE simplify from an exponential equation to an equation, where each parameter is captured in a summand term. Assuming that the rupture mechanism of the earthquake source and local site conditions are known and that the GMPE is a linear function of magnitude (m), the $F(y)$ can be obtained by a change of variables (Cor68). Consequently, the average annual RoE of a level y of some ground motion parameter takes the form (McGuire, 2008)

$$\lambda_y = \lambda_{min} \int_{m_{min}}^{m_{max}} \int_{R|M} P[Y \leq y|r,m] f_{R|M}(r|m) f_M(m) dr dm. \tag{1.6}$$

Probability density functions (PDFs) involving r (space) as a random variable constitute a description of the spatial model of seismicity, which will be discussed presently. Note the assumption that one FMSL holds for the entire area. If not, $f_M(m)$ would be an inseparable function of r, the order of integration would change, and would take place over functions of the form $f_{M|R}(m|r)f_R(r)$. As this can become quite complicated, it is done only in refined models and is often done by numerical methods rather than analytically. When multiple sources can be identified, (1.6) becomes

$$\lambda_y = \sum_i \lambda_{y,i} = \sum_i \lambda_{min,i} \int_{m_{min}}^{m_{max}} \int_{r_{i,min}}^{r_{i,max}} P[Y \leq y|r,m] f(m|r) f_i(r) dr dm, \tag{1.7}$$

where $\lambda_{min,i}$ is the RoS of the i^{th} source, with the spatial distribution $f_i(r)$, and bounds $r_{i,min}$ and $r_{i,max}$. Equation (1.7) is the most common formulation of PSHA, and its evaluation is known as the Cornell–McGuire procedure (Kijko, 2011). Note that λ_y is the central trend of a probability distribution, as the variability of the GMPE, ε, has to be taken into account. That is, $P[Y \leq y|r,m]$ is assumed to be a complementary cumulative normal distribution. It has become the practice to finally integrate over the uncertainty parameter up to a certain level to give a single probability value. However, this practice has been criticised and an alternative (extreme) value distribution has been proposed (Pavlenko, 2015). The reason for the extreme value distribution is probably that recording of the peak (maximum value) of ground acceleration is equivalent to recording extreme values. Moreover, as Pavlenko (2015) states, the Weibull distribution allows for more realistic modelling of the extremes of peak ground motion values, which are bounded according to sound theory in physics of ground motion and are also clearly apparent in the data in the cases that were investigated.

1.3. Seismic Catalogues

Seismic catalogues are primarily records of earthquakes in time and space, along with any additional information available on the events, including, among others, the source mechanism and faulting style. Earthquakes have been recorded since ancient times, at first orally and, subsequently, by other means as civilisation progressed. As science advanced, the interest in earthquakes increased, with increasingly more events being recorded, albeit, for a long time, only the largest events. Along with the inception and progress of seismology as a study field, the systematic recording of earthquakes was, and is, being done at an ever-increasing scale. The development of more refined methods and superior technology have facilitated the establishment of denser seismic recording networks, enabling improved accuracy and the detection of much smaller earthquakes. Figure 1 illustrates schematically a conceptualisation of a typical earthquake catalogue such as described here.

Over time, the record of earthquakes has become more complete and more reliable. Prehistoric events (or palaeoearthquakes) are events that precede written records and are inferred, albeit with a large margin of error, from geologic evidence on fault lines and effects, such as near-source deformation. Only events of an extremely large magnitude can be inferred in this way and the resulting record is quite incomplete. Along with the population growth at relevant locations, the records on the experiences and effects of earthquakes have increased. From these records, the intensities and magnitudes of the events can be deduced, with a relatively large margin of error. Only the largest and most devastating earthquakes were recorded in this fashion. However, with the inception of modern technology and instruments (seismometers), scientists started to record and catalogue earthquakes systematically in such a fashion that, in a given area, all earthquakes could be recorded above a given magnitude level. With time, the sophistication and sensitivity of seismometers, the extent of the seismometer networks, and the sophistication of mathematical processing and inversion methods have improved. Because of these advances, the level above which all earthquakes are recorded completely (the magnitude level of completeness) has dropped and the margin of error on magnitude estimates has been reduced considerably. From this viewpoint, the level of completeness (LoC) is a monotonically decreasing function of time (Kijko *et al.,* 2016). At times, for instance, during World War II, seismic stations were not in operation, resulting in a time-gap in the catalogues. Accordingly, earthquake catalogues are modelled as consisting of three parts, namely, prehistoric, historical (incomplete), and instrumental. The instrumental part, in turn, is subdivided into subparts with different magnitude levels of completeness and, possibly, time-gaps owing to events, such as, among others, natural disasters, war, and political unrest that damage infrastructure. It is easy to use only the instrumental part of the catalogue, discarding the rest of the data; however, this is not advisable. Several methods that use the full extent of the catalogue do exist and are discussed in subsequent chapters, along with a discussion on the estimation of seismicity parameters.

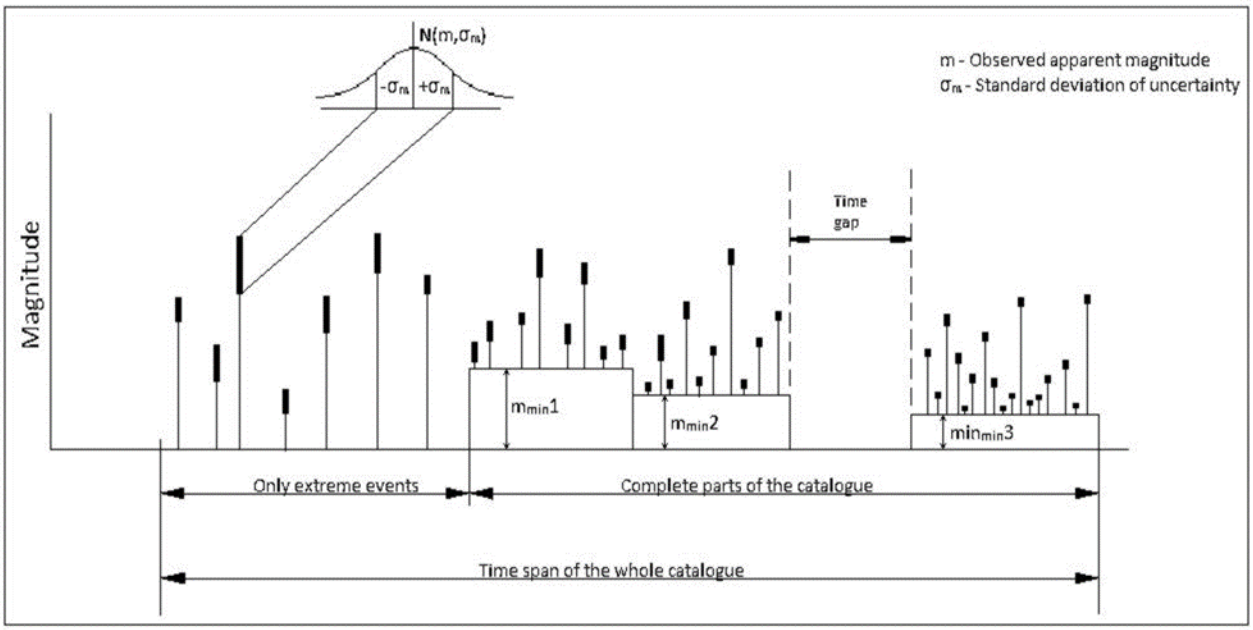


Figure 1. Conceptualisation of a typical earthquake catalogue. The first part consists only of extreme events recorded in a descriptive fashion. The extreme part may constitute part of or the entire historical catalogue. The second part, typically including the entire instrumental part of the catalogue and, possibly, the later part of the historical catalogue, comprise different parts, with the levels of completeness $m_{min1}, m_{min2}, \ldots$ decreasing as time progresses. Time gaps may exist because of sociopolitical factors that caused a lack of records of seismic activity. The historical part of the catalogue typically has substantial uncertainty attached to its recorded magnitude values and is much more difficult to determine (or estimate) than the instrumental part of the catalogue. With time, the uncertainty decreases in both the historical and the instrumental parts of the catalogue. (After Kijko and Sellevoll, 1992).

1.4. Introduction to Parameter Estimation in PSHA

When the logarithm of the number of earthquakes recorded in a given magnitude interval is plotted against the magnitude value, the decreasing trend of the number of observed earthquakes with increasing magnitude typically defines a straight line (GR FMSL). It should be noted that this is the case when a maximum magnitude value is not imposed, i.e. $m_{max} = \infty$. Such a plot is referred to as a Gutenberg–Richter plot. The slope of this value is the value b in equation (1.2). It should be noted that, strictly, a straight line results only when a maximum magnitude value is not imposed, i.e. $m_{max} = \infty$. However, for lower magnitudes, the straight line model is a good approximation. When normalised for time, equation (1.2) becomes

$$\log(\lambda_m) = a - bm. \tag{1.8}$$

The parameter a in this instance would correspond to the seismicity of magnitude 0. The b-value corresponds to the slope of the line. It would be physically impossible for earthquakes of arbitrarily large size to occur. There must be a finite upper bound (or supremum; see Glossary) m_{max}. Accordingly, the straight line is observed on a specific interval $[m_c, m_{max}]$.

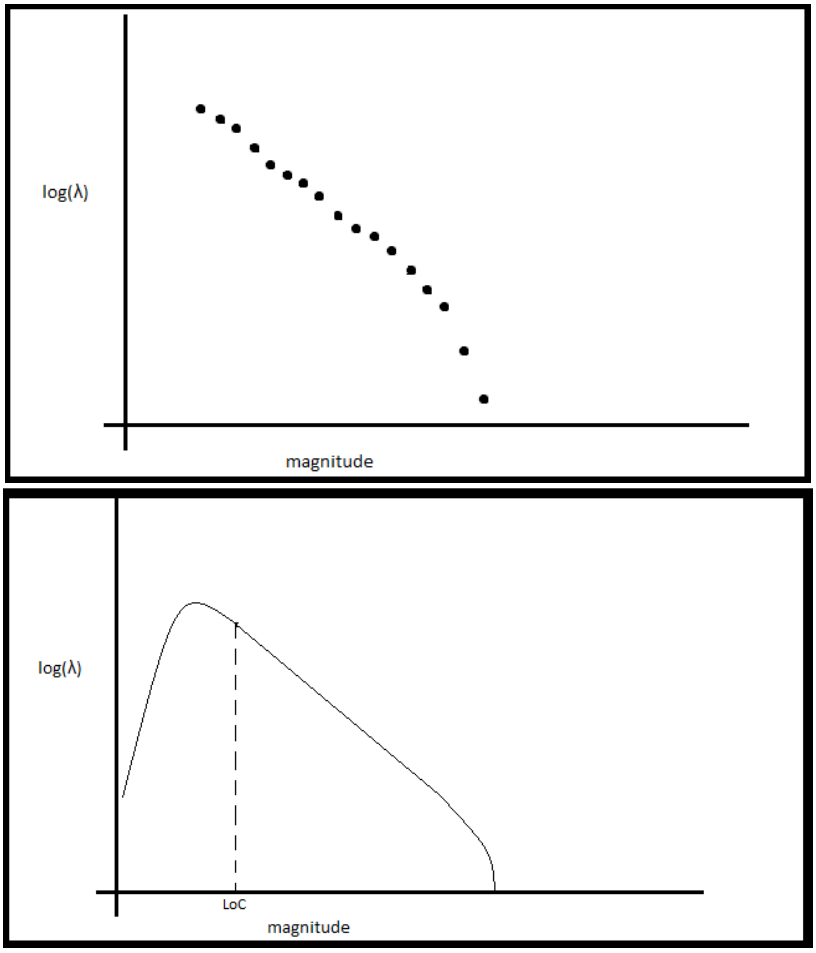


Figure 2. Top: Typical Gutenberg–Richter plot. Bottom: Idealised example of a Gutenberg–Richter plot, with data loss below the magnitude level of completeness (LoC).

Figure 2 shows a typical Gutenberg–Richter Frequency-Magnitude plot. The downward bend at the largest magnitudes is because of the finitude of the GR FMSL; in other words, the fact that earthquakes cannot take on magnitude values beyond a finite value m_{max}. Moreover, when real data are plotted, the line starts to curve below the LoC because of the loss of data, as in Figure 2.

Employing the Cornell–McGuire procedure, PSHA appears relatively simple and, for engineers, this is probably the case. However, for the teams of seismologist and geologists, the Cornell–McGuire procedure is only the last step, as considerable prior effort is required to determine the parameters (β, λ_0, and m_{max},

as well as the GMPEs) to conduct the procedure. Determining these parameters requires meticulous and extensive analysis of the source zones by geologists and seismologists and advanced statistical analysis of seismic catalogues by statistical seismologists. The teams have to determine the extent of the source zones and the size and times of occurrence of palaeoearthquakes to supplement the catalogue data and the estimates of parameters derived from the data (e.g. Kramer, 1996; Wallace, 1981). Statistical analysis of catalogues is required when such catalogues contain ambiguous and uncertain data (e.g. Kijko and Sellevoll, 1992; Kijko *et al.,* 2016; Leptokaropoulos *et al.*, 2018).

GMMs and the estimation of their parameters are beyond the scope of this work. Interestingly, large research teams have been working on refining GMPEs for different regions on a global scale, using empirical data, geologic and geophysical data, and physical simulations (Power *et al.,* 2008). For comprehensive details on the topic of GMPEs, the reader is referred to the information available on the Pacific Earthquake Engineering Research Center web page (PEER, 2018).

1.5. The Problem of Parameter Estimation

As already mentioned, a concerted effort is required to determine the parameters β, λ, and m_{max}. In addition, the LoC, m_c, has to be estimated for each complete sub-catalogue in the earthquake catalogue (which consists of the instrumental part and possibly some of the historical parts of the catalogue). Although not required as a parameter in the Cornell–McGuire procedure, the LoC must be determined to enable estimation of the rest of the parameters from earthquake catalogues. This is because in the methodologies customarily used, only the part of catalogues above the LoC can be used. As, per definition, loss of data occurs at lower magnitude values, Kijko and Smit (2017) devised a method to incorporate all the available data, dispensing with the need for the estimation of the LoC. However, this new research has yet to make its way into customary practice, and possible disadvantages of the method might still be uncovered. Determining the LoC for each sub-catalogue requires intensive analysis and, furthermore, is the subject of considerable research (e.g. Wiemer and Wyss, 2000; Woessner and Wiemer, 2005).

As a parameter, the LoC, m_c , for sub-catalogues in the complete part of the catalogue is of interest, as it indicates which part of the data can be used. As seen in Figure 3, m_c is the point on a Gutenberg–Richter plot below which curvature starts. The most obvious way of dealing with the problem would be by graphically determining this point (''eyeball estimation''). However, because using data below the LoC for parameter estimation results in incorrect estimates and discarding the valuable data above the LoC is undesirable, several methods have been developed in attempts to estimate the LoC more accurately (e.g. Wiemer and Wyss, 2000; Woessner and Wiemer, 2005; Amorèse, 2007).

Determining the most convenient subdivision of the complete part of the catalogue presents another problem, namely, if data from a part with a lower LoC are allocated to a part with a higher LoC, the data that could have been used are discarded because of the higher LoC. Although little has been published on this topic, there are notable studies by Stepp (1972) and Tinti and Mulgaria (1985).

As regards the first parameter of direct interest, an optimal estimate of the value of β is easy in the simplest case and relatively simple for a single complete catalogue of superior quality. If the range $[m_c, m_{max}]$ is relatively large (in practice this has to be only a few orders of magnitude), m_{max} can be assumed to be at infinity. Accordingly, the MLE of β is given by the simple equation (Aki, 1965)

$$\beta = \frac{1}{\overline{m} - m_c}. \tag{1.9}$$

Taking m_{max} into account, the MLE of β is (Page, 1968)

$$\frac{1}{\beta} = \overline{m} - m_{min} + \frac{(m_{max} - m_c)\exp[-\beta(m_{max} - m_c)]}{1 - \exp[-\beta(m_{max} - m_c)]}. \tag{1.10}$$

In addition, the β value is often estimated by a linear least squares regression on the Gutenberg–Richter plot for single complete catalogues (e.g. Bender, 1983; Guttorp, 1987). However, estimating β is not nearly as simple when this has to be done from the historical and complete subparts of the catalogue.

Estimating the RoS, λ_{min}, is linked closely with the parameter β. The MLE for a single complete catalogue that contains n observations and spans a time interval t is

$$\lambda_{min} = \frac{n}{t}. \tag{1.11}$$

Again, this is rather simple. The task is more complex, however, as, for different sub-catalogues with different LoCs, i^{th} sub-catalogue, the RoS $\lambda_{min,i}$ will be estimated for magnitudes above the LoC $m_{min,i}$ of the sub-catalogue, as λ_0 has to be estimated for a lower bound, which serves as a general lower bound in the analysis of the catalogue, m_0. It is a vital aspect of the statistical analysis to determine λ_0 from the complete sub-catalogues and, if possible, the historical part of the catalogue. The estimation of λ in the case of incomplete catalogues is discussed in Chapter 3.

Estimating m_{max} from the seismic catalogue, data is not a straightforward task, even in the instance of a simple complete catalogue, and the solutions to the problem are not quite intuitive either. Essentially, it is an attempt at estimating the maximum possible earthquake magnitude. Several studies have been published on this topic (Kijko, 2004; Kijko and Singh, 2011; Vermeulen and Kijko, 2017; Beirlant *et al.,* 2018; Pisarenko, 1991; Pisarenko *et al.,* 1996; Pisarenko *et al.*, 2014 and references therein; Pisarenko and Rodkin, 2017); however, some controversy remains. Some authors (e.g. Holshneider *et al.*, 2011) consider the parameter ill defined, whereas others (e.g. Raschke, 2015; Kagan, 2002a and references therein; Kagan, 2002b) have specifically proposed modifications to the tail of the distribution (1.3) for the purpose of estimating m_{max}. Probably, the most practical way to deal with the problem is to obtain a single value (point estimate) with its variance and possibly higher moments or to obtain confidence

bounds for the value. The problem associated with using confidence bounds is that the level of confidence is arbitrary and confidence bounds are prone to misinterpretation in such unconventional applications. Engineers tend to turn to and simply accept expert opinion; however, in view of the controversy surrounding the subject, there is very often little reason to believe that expert opinion would be any more reliable than statistical estimates are. In cases where there is reason to value expert opinion on the same level, or more highly than data, Bayesian formalisms should be resorted to.

Determining or modelling the spatial distribution of earthquake occurrence is the final problem in parameter estimation, or ''parameter estimation in a broader (generalised) sense''. It is another aspect of the problem of PSHA that Co68 did not consider, or had taken for granted, or as a given. Usually, it is modelled non-parametrically, as no specific functional model can be used as a generic law for the spatial distribution of seismicity. The RoS gives the relative values for the spatial distribution — a link of which the importance cannot be overemphasised. Actually, such a non-parametric, almost undefined model requires an infinite number of descriptive parameters, which are modelled empirically with an acceptably high resolution (i.e. a grid-wise approach). Spatial distribution is usually modelled by quite simple, rather trivial seismic zones based mainly on expert opinion. Typically, the RoS is assumed to be uniform over a seismic zone and the boundaries are arbitrary and geometrically simple. Further, assuming the FMSL as constant over a seismic zone is a gross oversimplification. In other words, the spatial model of seismicity involves variations both in the RoS and the FMSL.

The ambiguity, disagreement, and unresolved problems related to parameter estimation point to a problematic gap in PSHA, i.e. reliable parameter estimation from seismic catalogues. Although numerous authors have made attempts at filling this gap, much remains to be done. This work attempts to address this problem to a large extent.

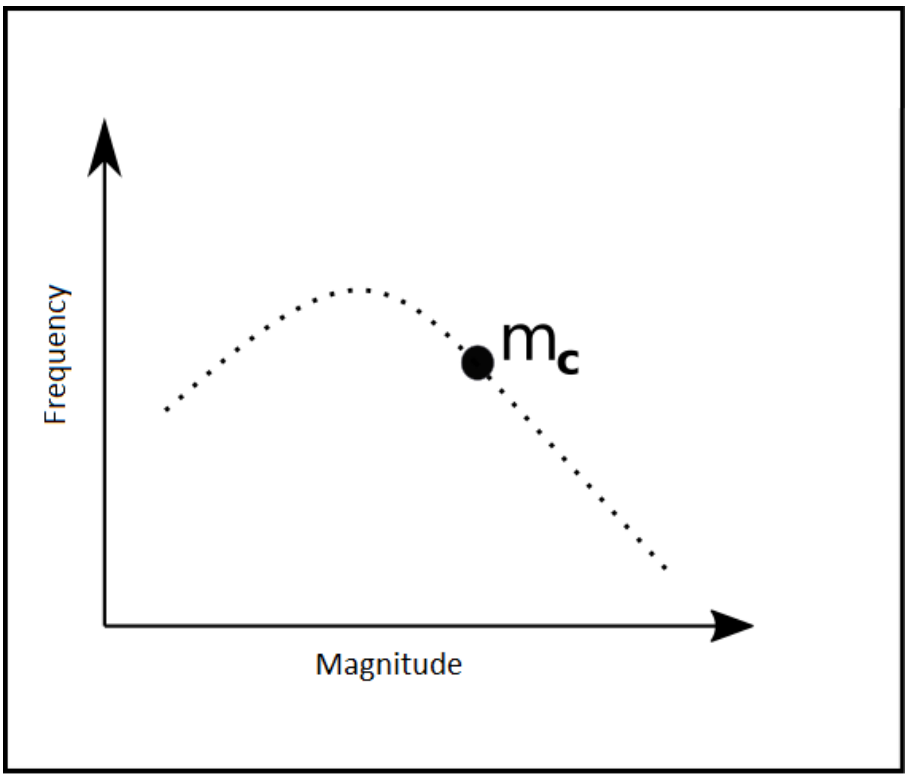


Figure 3. Illustration of a Gutenberg–Richter plot with m_c emphasised. It is the point below which curvature ensues.

2. Magnitude of Completeness

A precise definition of the magnitude LoC is given by Rydelek and Sacks (1989), "*$[m_c]$ is defined as the lowest magnitude at which 100% of the events in a space-time volume are detected* " (in Amorèse, 2007).

An earthquake catalogue will be incomplete below a specific LoC depending on the detection capabilities of the recording instrumentation, seismic noise, and decisions to only record magnitudes above a certain cut-off level, among other aspects (Mignan and Woessner, 2012). The result is that the frequency of the recorded events does not reflect the true frequency, i.e. it would be smaller. The simplest, most intuitive model for a frequency-magnitude relation affected by data loss would be a "true" magnitude distribution $f(m)$, multiplied by a data loss probability function $q(m)$. The resulting frequency-magnitude distribution takes the form (Ringdal, 1975)

$$p(m) = q(m)f(m)/C, \tag{2.1}$$

where $C = \int_{-\infty}^{\infty} q(m)f(m)\,dm$ is a normalising constant. Ringdal (1975) presents an elegant justification for $q(m)$ to be a cumulative Gaussian distribution function, namely $q(m) \sim \mathrm{erf}(\mu, \sigma)$. Other authors who use this form for $q(m)$ in accordance with the work by Ringdal (1975) include Iwata (2012), Ogata and Katsura (1993), Alamilla *et al.* (2014), and Woessner and Wiemer (2005).

Mignan (2012) introduced five different models for frequency-magnitude distribution (FMD), where $q(m)$ in equation (2.1) differs for each model. Model I is an angular-shaped function with $q(m) = \exp(\kappa(m - m_c))$ for magnitudes below m_c, and unity above m_c (in which case it might be described more appropriately as wedge-shaped, although Mignan does not use this terminology). Mignan contends that Model I is the most elementary FMD that would result when no spatial and temporal variations are present. Model II (intermediary FMD) results from local space–time variations, with the FMD resulting from the superposition (or summation) of several angular FMDs with varying m_c, but with the variation being so small as not to result in multiple maxima. Model III (intermediary FMD with multiple maxima) is the result of the superposition of angular FMDs, with m_c varying in such a way that the resulting FMD would have multiple maxima. Model IV (gradually curved FMD) is the result of the superposition of angular FMDs in such a way that it can be approximated by equation (2.1). It is viewed as the result of variations in m_c because of the regional distribution of seismic networks. Model V (gradually curved FMD with multiple maxima) is the result of the superposition of several angular FMDs, leading to an FMD with several maxima, and it can be approximated by the superposition of several gradually curved FMDs. It is viewed as the result of large variations in m_c because of the superposition of local and regional seismic networks. According to the literature, despite the proposal of the angular FMD for improved detection capability, the Gaussian model appears to prevail. Furthermore, the work by García-Hernández *et al.* (2019) appears to indicate that the Gaussian distribution is the most appropriate model in most instances.

To compute the Gutenberg–Richter b-value from the commonly used Aki–Utsu estimator (Aki 1965, Utsu 1965) and its variants (Molchan *et al.*, 1970; Weichert, 1980; Rosenblueth and Ordaz, 1987; Kijko and Sellevoll, 1989, 1992; Kijko and Smit, 2012, 2017), it is necessary to use a part of the catalogue for which the GR relation holds. This would correspond to a so-called LoC (m_c), above which $q(m) = 1$. Data below m_c are usually discarded, as the estimation of the b-value will produce incorrect results if the Aki–Utsu method were used. However, when the data above M_c are discarded, valuable information is lost. A correct estimate of m_c is, therefore, of vital importance. Methods not requiring the parameter m_c have been developed, such as those by Ogata and Katsura (1993), Alamilla *et al.* (2014), and Kijko and Smit (2017). However, these methods are not being used frequently yet. Accordingly, m_c is a crucial parameter to be estimated. In the Gaussian data loss model, there is a chance of data loss for all magnitudes. The model of Woessner and Wiemer (2005) and the angular FMD have a well-defined value of m_c. Although Mignan and Chen (2016) argue against using the Gaussian model for data loss, a gradually curved FMD is usually observed, in contrast with the angular FMD they propose.

The first estimator of m_c was probably developed by Stepp in 1972 and is still in use today. This method is based on the assumption that earthquakes follow a Poissonian distribution in time. It utilises the fact that, for a stationary Poisson process, the standard deviation of the mean number of earthquakes is

inversely proportional to the number of earthquakes. Because the expected number of earthquakes is directly proportional to the elapsed time, the number of earthquakes can be replaced by the extent of the time interval. For a catalogue divided into different magnitude classes, the time before the present, for which a given magnitude class is complete, could be estimated by the point where a plot of the standard deviation of the mean number of earthquakes against the time before present deviates from a linear trend, with a slope of negative unity. Such a plot is known as a completeness plot.

The study by Tinti and Mulgaria (1985) presents an interesting approach, as they assign a completeness index to different magnitude classes over time periods, in which the apparent rate of earthquake occurrence is considered homogeneous. For the j^{th} magnitude class, the Poisson mean $\mu_j(t_i)$ is estimated from the catalogue. The relative and absolute completeness indexes are given, respectively, by

$$\gamma_j(t) = \frac{\mu_j(t)}{\tilde{\mu}_j}, \qquad C_j(t) = \frac{\mu_j(t)}{\sigma_j}, \tag{2.2}$$

where $\tilde{\mu}_j$ is the highest rate of occurrence found in the j^{th} magnitude class, and σ_j is the true Poisson rate of occurrence for the j^{th} class. The rate σ_j is estimated for large magnitude classes for very recent time periods and is extrapolated for lower magnitude classes from the Gutenberg–Richter relation.

Wyss *et al.* (1999) and Wiemer and Wyss (2000) introduced the widely used maximum curvature (MAXC) method, where m_c is estimated as the point of MAXC of the cumulative frequency-magnitude plot. This corresponds to the highest value of the non-cumulative frequency-magnitude plot.

Wiemer and Wyss (2000) introduced the goodness of fit (GFT) method, which compares the observed FMD with synthetic distributions. The a- and b-values for the synthetic distribution are computed from the data above a given cut-off magnitude m_{co}. The statistic R is defined as

$$R(a, b, m_c) = 100 - \left(\frac{\sum_{m_{co}}^{m_{max}} |B_i - S_i|}{\sum_i B_i} 100 \right) \tag{2.3}$$

where B_i is the predicted value of the cumulative number of magnitude counts in each bin, and S_i the actually observed count. Starting with low values of m_{co} and increasing incrementally, m_c is estimated as the first value m_{co} for which the data are modelled at a pre-specified accuracy by a straight line. The accuracy of the model is given by R (as a percentage).

Cao and Gao (2002) estimate m_c by considering the stability of the estimated b-value. The GR relation has a constant b-value; therefore, fluctuation (typically a steady increase) of the b-value with increasing cut-off magnitude indicates that the cut-off value is too low. These authors regard m_c as the first cut-off magnitude for which two consecutive estimated b-values differ by less than a predefined constant.

Woessner and Wiemer (2005) introduced a different criterion based on the uncertainty δb of the estimated b-value of Shi and Bolt (1982)

$$\delta b = 2.3b^2 \sqrt{\frac{\sum_{i=1}^{n}(m_i - \bar{m})^2}{n(n-1)}}, \tag{2.4}$$

where n is the number of earthquakes, and $\bar{m}$ is the mean value of the earthquake magnitudes. The variable m_c is taken as the first cut-off magnitude for which $|b_{ave} - b| < \delta b$, where b_{ave} is the average of b-values calculated for consecutive cut-off values.

Marsan (2003) defines the log-likelihood of completeness as the logarithm of the probability that the model GR law calculated with the given cut-off magnitude could predict the number of earthquakes if there were a small magnitude increment just below it. m_c is chosen so that the log-likelihood of completeness drops for $m = m_c$, and the calculated b-value drops for $m < m_c$.

Woessner and Wiemer (2005) use a model to estimate m_c that incorporates the entire magnitude range (EMR), including values below m_c . This model is based on equation (2.1), with a slightly modified Gaussian model for the probability of detection. The part of the FMD below m_c is given by a truncated Gaussian distribution, and the part above m_c by a GR relation. This model is fitted to the observed FMD using the maximum likelihood method.

Mignan and Woessner (2012) consider the model used in the EMR method as an incorrect description of the natural process, as it does not conform to equation (1.1). A logical alternative to the model is a truncated Gaussian distribution for $q(m)$, truncated at m_c.

Amorèse (2007) applies the multiple change-point method, as developed by Lazante (1996). It is referred to as the Median-Based Analysis of Segment Slope (MBASS). The logic behind this is to calculate a theoretical "change-point" in the slope of the non-cumulative FMD. First, the segment slope at magnitude value m_2 is defined as

$$s(m_2) = \frac{\log(n(m_1)) - \log(n(m_2))}{m_1 - m_2}. \tag{2.5}$$

The values s are assigned a rank, and SR_i is assigned the sum of the ranks up to the i^{th} segment slope. The adjusted sum is computed:

$$SA_i = |2SR_i - i(n+1)|. \tag{2.6}$$

The maximum of the adjusted sum is identified as a possible change-point m_{ch} and marked as point n_1. To test whether m_{ch} qualifies as a change-point, the Wilcoxon–Mann–Whitney non-parametric test is employed. The test statistic, z, for this test is computed as follows (Lazante, 1996)

$$\begin{gathered} n_2 = n - n_1 \\ W_{crit} = n_1(n+1)/2 \\ s_W = \sqrt{n_1 n_2 (n+1)/12} \end{gathered}$$

$$\delta = \begin{cases} 0.5 & if\ W < W_{crit} \\ -0.5 & if\ W > W_{crit} \\ 0 & if\ W = W_{crit} \end{cases} \tag{2.7}$$

$$z = (W - W_{crit} + \delta)/s_w.$$

The hypothesis that m_{ch} is not a change-point is then not rejected ("accepted" for practical purposes) at some chosen level of confidence. The first auxiliary change-point is set at the next largest value of SA_i.

Rydelek and Sacks (1989) noted that seismic noise increases during the day compared with the night, causing incompleteness. These authors developed an easy and effective way to estimate the magnitude level below which such diurnal variation in seismic noise causes incompleteness. For a given magnitude class, starting at the centre of a circle, a unit vector (phasor) is added to a path in the direction of the hour on a 24-hour clock. It is assumed that earthquakes follow a homogeneous Poisson process (i.e. random in time). The probability of the resulting vector (or so-called walkout) to exceed a given radius R is $P_R = \exp(-R^2/n)$, where n is the number of earthquakes. Therefore, the null hypothesis of the recorded earthquakes occurring randomly in time is rejected below some confidence level $(1 - P_R)$. For the 95% confidence interval, this would be when $R = 1.73\sqrt{n}$. If the resulting walkout is larger than R, the catalogue is biased and, consequently, incomplete at the given magnitude level.

Referring to the work of Godano *et al*. (2014), Godano (2017) notes that the b-value changes continuously with a magnitude threshold m_c, even at large values of m_c, which can be ascribed to the spatial and/or temporal non-homogeneity of the b-value. According to Godano (2017), this view is supported by, among others, Kagan (2004), Helmstetter *et al.* (2006), Enescu *et al*. (2007), Lippiello *et al.* (2007), Peng *et al.* (2007), Lippiello *et al.* (2012), Omi *et al.* (2013), de Arcangelis *et al*. (2016), and Lippiello *et al.* (2016). It is clear that part of the scientific community agrees on this. Obviously, this complicates the estimation of m_c even more (Godano, 2017).

Godano (2017) proposes a method based on the calculation of the ''harmonic mean of the magnitudes'' larger than a given threshold

$$\nu = \frac{n}{\sum_{i=1}^{n} 1/m_i}, \tag{2.8}$$

where $m_i > m_{th}$, and m_{th} is the threshold magnitude. The harmonic mean is related linearly to m_{th} for $m_{th} > m_c$, and is higher than the theoretical value in the case of a GR law. In addition, it has been observed to be practically constant for values less than m_c. The difference Δ between the harmonic mean ν and the theoretical value of the linear trend has to be investigated. Godano (2017) proposes an estimate $\widehat{m}_c$ of m_c when $\Delta < \delta_{th}$ for some prechosen value δ_{th}. Judging from the results, the method of Godano (2017) does seem to be superior to the method of b-value stability. For the sake of comparison, this estimator shall be referred to as m_c^{Harm}.

2.1. Consideration of Methods

Mignan and Woessner (2012) found that $m_c^{MAXC} \sim \mu$ and that $m_c^{MBS} > \mu + \sigma$, and m_c^{GFT} and m_c^{MBASS} yield intermediate results. This formulation was done under the assumption that $q(m) \sim \mathrm{erf}(\mu, \sigma)$. In a study by Huang *et al.* (2016), it was found that m_c^{MAXC} consistently underestimates $m_{,c}$, as does m_c^{MBS} and m_c^{GFT}, but m_c^{EMR} yields intermediate results. Huang *et al.* (2016) used three models, namely (1) $= 0.9$, $\mu = 1.5$, and $\sigma = 0.2$; (2) as in model 1, except $\sigma = 0.4$; (3) model 1 is combined with an equal number of events from a catalogue for which all events (100%) above magnitude 1.5 are recorded. As Huang *et al.* (2016) note, a larger σ indicates a slower change in the detection capability with a change in magnitude. Huang *et al.* (2016) assume that the correct value of m_c would be where one in every 500 events is missed. This means m_c^{MAXC} underestimates m_c in the instance of gradually curved FMDs (e.g. Woessner and Wiemer, 2005; Mignan and Woessner, 2012). Godano (2017) shows that the proposed estimator m_c^{Harm} is superior to methods based on b-value stability.

2.2. Critical Analysis of the Arguments of Mignan (2012) and Mignan *et al.* (2011)

Mignan *et al.* (2011) use the difference $\Delta = m_c^{MBASS} - m_c^{MAXC}$ as a proxy to the gradual curving of the FMD and note that Δ becomes smaller as spatial and temporal heterogeneities decrease. Mignan and Woessner (2012) claim that it has been shown in Mignan *et al.* (2011) that Δ tends to zero as heterogeneities decrease to zero; however, in fact, this is only a suggestion. Mignan (2012) proposes the angular FMD model for homogeneous catalogues but does not demonstrate that this model is applicable in reality. The argument by Ringdal (1975) that $q(m) \sim \mathrm{erf}(\mu, \sigma)$ still appears to be an exceptionally good physical justification of the model. Indeed, as Mignan and Chen (2016) state, the graduality with which the FMD curves appear to decrease with the increase in inhomogeneity of aggregated local catalogues (in other words, many local catalogues, each with their own LoC, are superimposed in the larger catalogue); however, it has not been proven nor shown by the authors that $q(m) \sim \exp(m)$, as is required by the angular FMD model. The obvious alternative view would be that σ greatly decreases in the model $q(m) \sim \mathrm{erf}(\mu, \sigma)$. The physical interpretation of the angular FMD model is that the probability of detection

decreases exceptionally rapidly immediately below the value of m_c. It can be argued equally well that σ decreases with decreasing homogeneity, resulting in a more gradual curvature of the FMD, as noted by Huang *et al.* (2016).

2.3. Critical Analysis of the Maximum Curvature Method

In several studies published over the last fifteen years, the magnitude LoC was estimated by, what is termed, the Method of Maximum Curvature and, notably, Woessner and Wiemer (2005) consider this a promising method. The magnitude LoC of an earthquake catalogue is defined as the magnitude above which all earthquakes are recorded in space–time volume (Amorèse, 2007).

The current author wishes to point out that the Method of Maximum Curvature (Woessner and Wiemer, 2005) produces erroneous estimates of the LoC (m_c) of earthquake catalogues.

Recall that the GR law produces a theoretically linear logarithmic plot, such as in Figure 4.

The gradual data loss can be seen in Figure 5, which is a Gutenberg–Richter plot of the Myanmar earthquake catalogue from 1973 to 2015. It can be observed that at some point a deviation occurs from linearity in the Gutenberg–Richter plot. This is the LoC m_c.

As the name implies, this method estimates m_c as the point of MAXC on the Gutenberg–Richter plot. This, as the current author will show, is an erroneous estimate of m_c, when certain specific conditions are not met, as it does not correspond to the point where the curvature starts.

Woessner and Wiemer (2005) do point out that the method tends to underestimate m_c, but they appear to imply that this only happens for catalogues where the loss of data is gradual. It is true that the underestimation is more pronounced in instances of gradual loss; however, the limitation of this estimator is inherent and will always occur. Although m_c can appear attractive for use as a lower bound for m_c, it will only provide an accurate estimate in an instance where data loss is sudden, such as the exponential model of Mignan (2012). The results obtained by García-Hernández *et al.* (2019) illustrate that this is not a typical case, rather, from the data they used, it appears that this is the most unlikely case.

Estimating m_c correctly requires estimating the point where curvature starts, i.e. where deviation from a power law starts, as pointed out by Woessner and Wiemer (2005). The value of m_c does not necessarily correspond to the point of MAXC, except that it will always be at a larger or equal value.

2.3.1. Argument by example

As an example, the current author simulated a catalogue for which the probability of detection below m_c is (Woessner and Wiemer, 2005)

$$q(m|\mu,\sigma) = \begin{cases} \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{m} exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) dx, & m < m_c \\ 1, & m \geq m_c \end{cases}, \quad (2.9)$$

with $\mu = 1$ and $\sigma = 0.25$, in accordance with theoretical examples constructed by Woessner and Wiemer (2005). This closely mimics the model of Ogata and Katsura (1993), which was found to be the typical case in the data investigated by García-Hernández *et al.* (2019). The value of m_c was set at $m_c = 3$, which implies a more gradual loss of data below m_c to illustrate the point more clearly. Woessner and Wiemer (2005) note that underestimation when using the MAXC is more pronounced for gradual data loss. The resulting Gutenberg–Richter plot of simulated data is shown in Figure 6. This instance is extreme, i.e. the data loss is so gradual that m_c would be hardly detectable by any method; however, this extreme example illustrates the argument clearly.

As noted in Woessner and Wiemer (2005), an uncomplicated way to identify the point of MAXC is to identify the bin with the highest frequency of events, which, in this case, is 1.03. This is well below the true value of m_c.

2.3.2. Theoretical argument

If, for the sake of simplicity, the minimum recorded magnitude were considered at $m = 0$, the GR relation brings about the probability distribution function for magnitudes

$$w(m|\beta) = \exp(-\beta m). \quad (2.10)$$

Subsequently, combining equations (2.9) and (2.10) and normalising, the PDF for the model of the recorded magnitudes with data loss is obtained (Iwata, 2012)

$$f(m|\beta,\mu,\sigma) = \frac{w(m|\beta)q(m|\mu,\beta,\sigma)}{\int_{-\infty}^{\infty} w(m|\beta)q(m|\,\mu,\beta,\sigma)dm}. \quad (2.11)$$

The point of MAXC can be obtained from equation (2.11) by setting the derivative of the numerator of the left-hand side of equation (2.11) equal to zero

$$\frac{d}{dm}[w(m|\beta)q(m|\mu,\beta,\sigma)] = 0, \quad (2.12)$$

which results in

$$w(m|\beta) = \exp(-\beta m), \tag{2.13}$$

which is independent of m_c.

An instance can be modelled where the point of MAXC (the bin with the highest frequency of events) is at m_c by putting m_c at a point smaller than that satisfying equation (2.13). However, this implies a sudden and almost complete data loss corresponding to the exponential model of Mignan (2012).

It must be pointed out that at m_c the derivative of $f(m|\beta, \mu, \sigma)$ does not exist (is infinite), which can, in a certain way, be considered a point of MAXC. However, this is an artefact of the model, and it is also realistic to assume that $f(m|\beta, \mu, \sigma)$ is smooth at m_c, which, however, was not the case with the model used in the current research. Furthermore, when the change in slope at m_c is quite small (corresponding to exceptionally gradual data loss), this point would be hardly detectable.

Furthermore, even an estimate that seeks to estimate the point where deviation from linearity starts will theoretically be biased toward too low values of m_c if the starting point of this deviation were extremely gradual and, at the actual value of m_c, the deviation will be hardly detectable. The reasoning of the current author is consistent with that of Amorèse (2007) and Iwata (2012).

2.3.3. Conclusion

The Method of Maximum Curvature to estimate m_c has been shown both by example and by theoretical argument to have a weak point. Therefore, it is recommended that this method not be used when data loss is gradual below m_c.

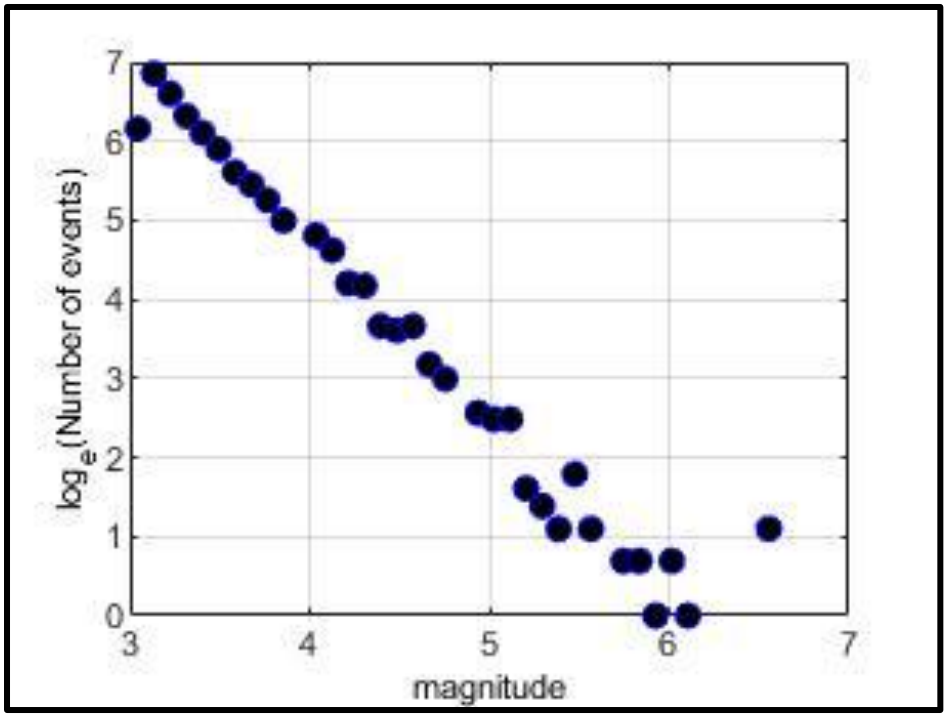


Figure 4. Plot of magnitude frequency for a simulated earthquake catalogue.

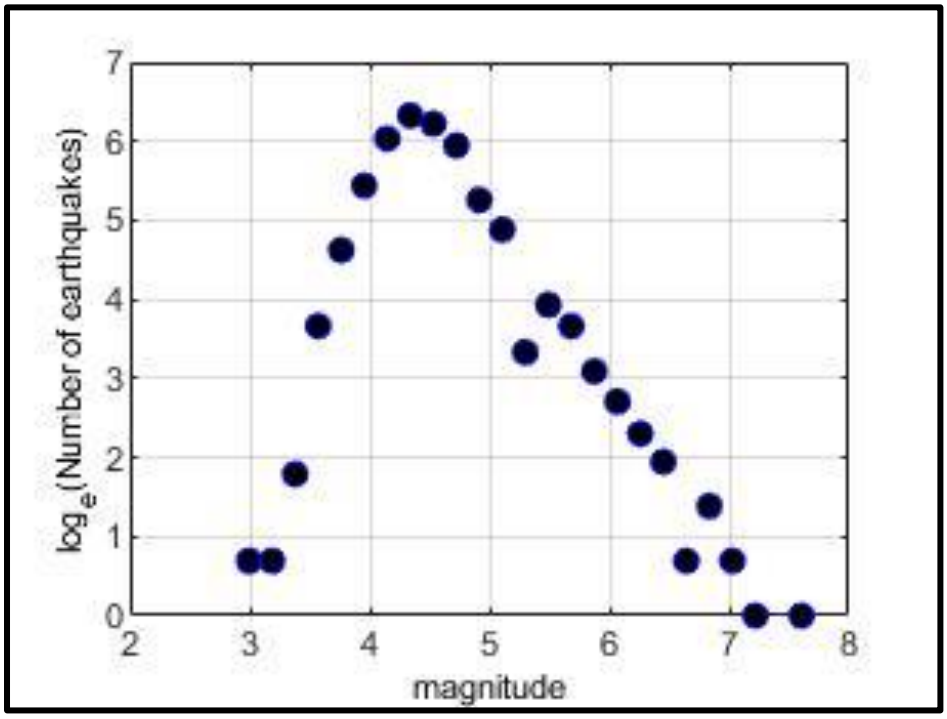


Figure 5. Plot of magnitude frequencies of an earthquake catalogue from Myanmar.

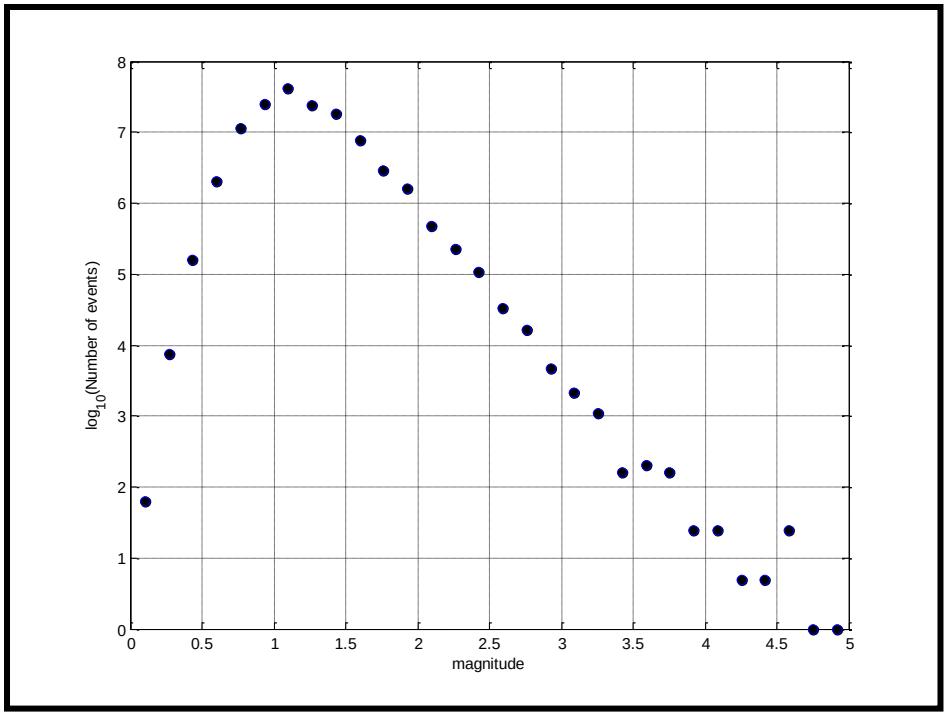


Figure 6 Simulated earthquake catalogue with data loss following a GR law, with data loss according to equation (2), with μ = 1, σ = 0.25, and $\boldsymbol{m_c}$ = 3.

2.4. Critical Analysis of the Median-Based Analysis of Segment Slope

The multiple change-point method, developed by Lanzante (1996), was designed to detect clear changes in a time series distribution, as, e.g. the location of discrete jumps. It may well be applied to a magnitude series, as Amorése (2007) applies the change-point method in the MBASS if a discrete change in the median magnitude value were expected. However, in the model where $q(m) \sim \mathrm{erf}(\mu, \sigma)$, the slope changes gradually. Therefore, the empirical value of the median of the segment slopes is expected to change gradually as well. A change can be detected in the median value of the segment slope only from and below a value of magnitude where the change in the b-value is detectable over random fluctuations. From a theoretical point of view, for a small incremental change ε that the MBASS method will fail to detect in the slope value, a small magnitude increment δ will be found in the magnitude below m_c, such that the slope takes on that change in value. Therefore, the value m_c^{MBASS} will be below the value of m_c. This finding is in agreement with that of Huang *et al.* (2016).

3. Seismicity Parameters b and λ

In this section, the estimation of the Gutenberg–Richter b-value and the RoS (λ) is discussed in detail. The b-value is, besides m_{max}, the essential parameter in the GR relation, which is the FMD, and λ is the essential parameter in the temporal distribution. These two values are related closely, with λ featuring implicitly in parameter a of the GR relation. Theoretically, a is the RoS for m, i.e. $\lambda(0) = a$, which is the extrapolation of the value of λ to magnitude 0.

3.1. Estimation of the b-value

The GR relation has been introduced already

$$log(n) = a - bm. \tag{3.1}$$

The GR FMSL is ubiquitous in nature, and an exceptional number of academic reports have been published on the topic (Marzocchi and Sandri, 2003). Furthermore, it is observed in various tectonic settings and, in many instances, even in induced seismicity (Marzocchi and Sandri, 2003) and in small-scale laboratory experiments (e.g. Scholz, 1968). In the GR relation, the b-value is a particularly important parameter, as it describes the ratio of small to large earthquakes. It has been assigned various physical meanings, including the phenomenon of Self-Organized Criticality (e.g. Bak and Tang, 1989; Sornette and Sornette, 1989). Mogi (1962) finds that the heterogeneity of rock affects the b-value (Scholz, 1968). According to Scholtz (1968), the b-value depends on a characteristic way on stress, as an increase in stress causes a decrease in the b-value. Subsequent studies on the b-value include those by Rabinovitch *et al.* (2001), Turcotte *et al.* (2003), and Ben-Zion (2008), among others. It has been noted, however, that deviations do occur (see particularly Youngs and Coppersmith, 1985). Nevertheless, the GR law, with its essential b-value, remains the basis for describing the FMSL for natural tectonic earthquakes. However, other studies (e.g. Wesnousky, 1994) hold a contradictory view, namely that the assumption of the general applicability of the GR law and the characteristic earthquake model is obsolete.

The b-value was calculated by Gutenberg and Richter (1944) using the least squares technique (Guttorp and Hopkins, 1986) but the use of the least squares technique in calculating the b-value has been criticised strongly, as will be explained later. However, the b-value was convenient and, probably, the most obvious way to obtain an approximation with accuracy that was adequate for their purposes. Currently, most estimation methods are based on the estimator derived by Utsu (1965) employing the Method of Moments, and by Aki (1965) using the maximum likelihood method. Subsequently, numerous variations for this popular estimator have been developed, which will be discussed later. The formula for the estimator of Utsu and Aki (hereinafter the Aki–Utsu estimator) is given by

$$\beta = \frac{1}{\overline{m} - m_{min}}, \tag{3.2}$$

where $\beta = \ln(10)\, b$, $\overline{m}$ is the mean of the magnitudes in the catalogue above m_{min} , and m_{min} is the minimum considered magnitude (typically the LoC m_c). Actually, this is the estimator of the value β for a shifted exponential distribution, which is well-known in pure statistics. The b-value and the value β will be used interchangeably. Aki (1965) presents a simple approximate variance of the estimator (3.2)

$$\sigma(\beta) = \frac{\beta}{\sqrt{n}}. \tag{3.3}$$

As mentioned in the introduction, each spatial element should have its own β-value, although such changes can usually be assumed to be quite gradual. Accommodating the spatial and slow temporal changes in β, Shi and Bolt (1982) proposed the standard deviation of the Aki–Utsu estimator for large samples

$$\sigma\left(\hat{\beta}\right) = \beta^2 \sigma(\overline{m}), \tag{3.4}$$

where

$$\sigma(\overline{m}) = \frac{1}{n}\sum_{i}^{n}(m_i - \overline{m})^2. \tag{3.5}$$

It is most unfortunate that, although such advanced theoretical tools are available, they are not being used in practice. Ogata and Yamashina (1986) note that, as is true for many statistical estimators, the Aki–Utsu estimator is biased, particularly when a small number of earthquakes (or data points) are involved. These authors propose an alternative, adapted formula

$$\beta = \frac{(n-1)}{\sum_{i=1}^{n}(m_i - m_c)}. \tag{3.6}$$

This formula also maximises entropy, which is a desirable property for an estimator. Nevertheless, the Aki–Utsu estimator in its original form is still predominant in practice, as formulas (3.2) and (3.6) are practically equivalent with regard to a large number of data points. As a theoretical consideration to explore the properties of the value β, Ogata and Yamashina (1986) propose a distribution of the estimated β-value arising from the Bayes' theorem by employing the non-informative prior (deriving from the Fisher information measure), which is given by

$$p(\beta) = \frac{[\sum(m_i - m_c)]^n}{(n-1)!}\beta^{N-1}\exp\left[-\beta\left(\sum(m_i - m_c)\right)\right]. \quad (3.7)$$

Notably, the mean of the distribution (3.7) is given by (3.2), and its mode by (3.6).

Note that distribution (3.7) is the probability distribution that arises from multiple estimations of β from multiple random samples from an exponential distribution, such as that deriving from the GR law, as it would, for instance, be observed in a Monte Carlo experiment. A considerable weakness of the Aki–Utsu estimator and its derivatives has been discussed, i.e. it does not take into account the maximum possible magnitude m_{max}. It is based on the assumption that the FMD is completely unbounded, which is common in theoretical statistics but completely unrealistic in the physical application to seismicity. In an attempt to resolve this, Page (1968) presents the maximum likelihood estimator of β in an instance where a maximum magnitude is imposed. The estimator is given by

$$\beta = \left[\bar{m} - \frac{m_c - m_{max}\exp[-\beta(m_{max} - m_c)]}{1 - \exp[-\beta(m_{max} - m_c)]}\right]^{-1}, \quad (3.8)$$

where m_{max} is the maximum possible magnitude. If the magnitude range of the data were sufficient (typically $m_{max} - m_c \geq 3$), the Aki–Utsu estimator would still be an adequate estimator (Marzocchi and Sandri, 2003). Marzocchi and Sandri (2003) warn strongly against the bias introduced by the assumption of continuity of binned (or grouped) magnitudes. This implies that the assumption is that the distribution is continuous; however, the sample employed in the current research is in the form of a histogram, which requires groups, or bins, of values. Formulas (3.2)–(3.10) are all based on the assumption of continuity, which is a natural but, unfortunately, erroneous approximation. The smaller the bins, the more negligible would be the error. The bias is not considerable for the usual bin width of $\Delta m = 0.1$ of instrumental catalogues, but it is definitely of significance for historical data, of which the bin widths are typically above $\Delta m = 0.5$ (Marzocchi and Sandri, 2003; Bender, 1983; Guttorp and Hopkins, 1986). Accordingly, Guttorp and Hopkins (1986) deem it necessary to derive the following MLE for β for the discretised (binned) model

$$\hat{\beta} = \frac{1}{\Delta}\log\left(1 + \frac{\Delta}{\bar{m} - m_c}\right), \quad (3.9)$$

where Δ is the bin width. The difference between estimators (3.2) and (3.9) is of the order Δ^2. For $\Delta m = 0.1$, the error would be of the order 0.01, which, in the current author's opinion is still considerable. However, this important theoretical result is not used in practice. Guttorp and Hopkins (1986) give the following approximate confidence bands for their maximum likelihood estimator

$$\left(\hat{\beta}-\frac{z_\alpha}{\sqrt{n}}\left(\bar{k}\left(\bar{k}+1\right)\right)^{\frac{1}{2}},\hat{\beta}+\frac{z_\alpha}{\sqrt{n}}\left(\bar{k}\left(\bar{k}+1\right)\right)^{1/2}\right), \tag{3.10}$$

where $\bar{k}=\sum k_i/n$, k_i is number of earthquakes in the i^{th} bin and n denotes number of bins. A standard deviation is not proposed, unfortunately. Quite astonishingly, a second, different maximum likelihood estimator was proposed by Bender (1983) for binned magnitudes, with the formula

$$\frac{q}{1-q}-\frac{nq^n}{1-q^n}=\sum_{i=1}^{n}\frac{(i-1)k_i}{N}, \tag{3.11}$$

where $q=\exp(-\beta\Delta m)$, n is the number of bins each with k_i earthquakes and N is the number of earthquakes where $N=\sum_{i=1}^{n}k_i$. (Intuitively it might be expected that only a single maximum likelihood estimator exists). It appears that equation (3.11) does not take m_{max} into account; however, it is done implicitly by taking the number n of magnitude groups into account, although, for some groups, k_i could be zero. In view of such difficulties, Bender (1983) conducted a meticulous and comprehensive study (that is highly undervalued) to derive the distribution of the estimated b-values for different numbers of intervals, numbers of earthquakes, and the like. Because magnitude data are grouped, only a finite number of possible b-values are possible, and the distribution is discrete. Bender compares the distributions of different estimators for combinations of continuous or grouped data and uses the maximum observed value and the true maximum value, such as b=1.0, over a range of sample sizes. The estimators that Bender (1983) compares are interval least squares, cumulative least squares, minimum χ^2, the maximum likelihood formula for continuous data, and the maximum likelihood formula for grouped data. Note that, for all techniques, only a discrete set of estimated b-values is possible in the instance of grouped data. The maximum likelihood formula for grouped data is that given in equation (3.11). The maximum likelihood formula for continuous data is given by formula (3.11). In the minimum χ^2 procedure, the quantity $\sum_{i=1}^{n}\frac{[k_i-\exp(\alpha)\exp(-\beta m_i)]^2}{\exp(\alpha)\exp(-\beta m_i)}$ is minimised. In the interval least squares procedure, Bender (1983) weights the number of observations by $\frac{1}{i}$ (the author notes that this is arbitrary; however, Guttorp (1987) lends support and explains the applicability of the weighting scheme). Presumably, Bender (1983) must have used some intuitive reasoning. Bender (1983) uses the classic cumulative least squares formula, and also gives a corrected version of the estimators by Page (1968) and the Aki–Utsu estimators. The corrected version of the Page (1968) formula is given by Guttorp (1986)

$$\frac{\sum_{i=1}^{n}k_im_i}{\sum_{i=1}^{n}k_i}=m_c+\frac{n\Delta m}{2}\left[1-\frac{1+\exp(-\beta n\Delta m)}{1-\exp(-\beta n\Delta m)}\right]+\frac{\Delta m}{2}\frac{1+\exp(-\beta\Delta m)}{1-\exp(-\beta\Delta m)}. \tag{3.12}$$

The corrected version of the Aki–Utsu estimator is

$$\lim_{n\to\infty}\frac{\sum_{i=1}^{n}k_i m_i}{\sum_{i=1}^{n}k_i}=m_c+\frac{\Delta m}{2}\frac{1+\exp(-\beta\Delta m)}{1-\exp(-\beta\Delta m)}. \tag{3.13}$$

This implies that formulas (3.12) and (3.13) are adjusted to suite the discretised, or binned, instance. Tinti and Mulgaria (1987) derive yet another estimator for grouped data

$$\hat{\beta}=\frac{\ln(\hat{p})}{\Delta m} \tag{3.14}$$

where

$$\hat{p}=z/(z+1) \tag{3.15}$$

and

$$z=\frac{\overline{m}-m_c}{\Delta m}-0.5. \tag{3.16}$$

Confidence intervals are given for this estimator by Tinti and Mulgaria (1987), as follows

$$r=zN, \tag{3.17}$$

where r follows the negative binomial distribution, given by

$$P_N(r,p)=\binom{N+r-1}{N-1}(1-p)^N p^r, \quad r=0,1,2,\ldots \tag{3.18}$$

Then, the cumulative probability distribution associated with r is given by

$$L_N(r,p)=\sum_{k=0}^{r}P_N(k,p). \tag{3.19}$$

Tinti and Mulgaria (1987) evaluate the confidence intervals for the parameter p by, "finding the lower end $p_{l,\alpha}$ and the upper end $p_{u,\alpha}$ of the interval for which the probability of containing the true value of parameter p is larger than and simultaneously as close as possible to the given amount $1-\alpha$, called the confidence coefficient." (Tinti and Mulgaria, 1987). Accordingly

$$P\left(p_{l,\alpha} \leq p \leq p_{u,\alpha}\right) \geq\approx 1-\alpha, \tag{3.20}$$

where $\geq\approx$ denotes "the best approximation in excess." (Tinti and Mulgaria, 1987).

The end points of the confidence intervals are then given by the implicit expressions

$$\begin{cases} L_N\left(r-1, p_{l,\alpha}\right) = 1-\alpha/2 \\ L_N\left(r, p_{u,\alpha}\right) = \alpha/2 \end{cases}, \tag{3.21}$$

that constitute a formal solution to the problem. Inverting equation (3.14) produces the corresponding end points of the confidence interval for parameter β, as follows

$$\begin{cases} \beta_{l,\alpha} = -\ln\left(p_{u,\alpha}\right)/\Delta \mathrm{m} \\ \beta_{u,\alpha} = -\ln\left(p_{l,\alpha}\right)/\Delta m \end{cases}. \tag{3.22}$$

As the estimator by Tinti and Mulgaria (1987) is not straightforward, the estimators by Guttorp (1987) and Bender (1983) are preferred for their simplicity. Actually, a study could be conducted on a comparison of the performance of the three estimators.

Nava *et al.* (2017) noted another source of bias in the Aki–Utsu estimator, as large magnitudes are often under- or over-sampled. As a solution, they propose choosing an upper threshold m_u magnitude, below which this under- and over-representation should not have an effect. Using only values between m_c and m_u, they propose a corrected estimate of the sample mean, given by

$$\hat{\bar{m}} = \frac{m_c + \frac{1}{\beta} - e^{-\beta(m_u - m_c)}\left(m_2 + \frac{1}{\beta}\right)}{1 - e^{-\beta(m_u - m_c)}}, \tag{3.23}$$

with β being unknown, the first estimate of β is from the Aki–Utsu estimator using the usual sample mean. The estimator for the sample mean and β are then solved iteratively by using the above equation for the mean and

$$\hat{\beta} = \frac{1}{\hat{\bar{m}} - m_c}. \tag{3.24}$$

It is important to note that m_u is not at all equivalent to m_{max}, as m_u is a value above which data values do exist but are discarded for the purpose of estimation. A related study by Nava *et al.* (2017), not as yet subjected to further investigation, points out seemingly valid concerns regarding histograms and

sampling. That is, the values in the tail often do not appear and, if they do, they do not take on integer values, which can be much higher than the theoretical values. Actually, this is a general problem encountered with distribution tails.

Further to the previously raised concerns about the least squares method, which was the initial method used by Gutenberg and Richter (1944), the general point of view is that this technique does not have a statistical foundation, according to the arguments of Page (1968) and Bender (1983). To the best of the current author's knowledge, a straightforward argument against the least squares method has not been presented explicitly in the literature, except by Sandri and Marzocchi (2006). They contend that the logarithmic transformation on the GR plot, which is a histogram, distorts the errors, most likely not producing a normal distribution of scattering, which, actually, is a fundamental assumption in employing the least squares method. Page (1968) refers to a study by Suzuki (1958), which shows that the technique is not applicable. Bender (1983) shows the extreme bias of applying the technique empirically and presents a brief explanation. In contrast, according to Shi and Bolt (1982), the least squares technique is applicable in some instances. Kijko (1994) notes that the L_1 norm, i.e. the sum of the absolute residuals, is less sensitive to outliers. In fact, Kijko (1994) presents an algorithm for a norm L_p to be minimised, which is optimal. The L_p norm is defined as $|y_i - f(x_i)|^p$. Although this L_p norm remains subject to scrutiny, it does represent a definite improvement. Guttorp (1987) notes that the ordinary least squares fit on the cumulative frequency-magnitude plot is inappropriate, as the summing number of earthquakes leads to dependence in the observations. This is probably the principal argument against the ordinary least squares method. Guttorp (1987) developed a generalized least squares approach, noting that the suggestion for weights of least squares (Bender, 1983) is asymptotically correct, as proven by the calculations. Therefore, if the least squares procedure were to be used at all, it should be the generalized least squares procedure. Bender (1983) presents a comparison of the performance of the least squares procedure and the other procedures, including the generalized least squares procedure. Sandri and Marzocchi (2006) note additional problems with the least squares estimation. First, fitting to the cumulative form of the distribution leads to an underestimation of the uncertainty of the estimate, which is ascribed to the filtering effect of high frequency noise in the cumulation operation. This relates to the problem Guttorp (1987) noted when he applied the generalized least squares approach. Second, the logarithmic transformation produces a bias in the estimation of the b-value, most probably because of the nature of an asymmetrical scatter under the logarithmic transformation. The bias depends strongly on sample size, as shown by Bender (1983). Given the superiority of the Guttorp (1987) generalised least squares (GLS) approach, it is sensible and prudent to explore this method.

This generalised least squares procedure by Guttorp (1987) is described as follows. Define $F_n(t) = \#\{j: X_j \leq t\}/n$, where X_j is rank ordered? Define $Y_i = \log(1 - F(m_i))$, $i = 1,2, \dots, n$.

The ordinary least squares line has a slope

$$\hat{\beta}_{LS} = \frac{\sum_{i=2}^{n} Y_{i-1} m_i}{\sum_{i=2}^{n} {m_i}^2}. \tag{3.25}$$

Another estimator by Guttorp (1987) used often, is called the slope-intercept estimate and is given by

$$\hat{\beta}_{SI} = \frac{\sum_{i=2}^{n} Y_{i-1}(m_i - \bar{m})}{\sum_{i=2}^{n}(m_i - \bar{m})^2}, \tag{3.26}$$

where $\bar{m}$ is the arithmetic mean of the magnitude values. Guttorp (1987) notes that $\hat{\beta}_{LS}$ has a smaller variance than does $\hat{\beta}_{SI}$. Nevertheless, the above two estimates are considered inappropriate by Guttorp (1987), whereas the weights proposed by Bender (1983) improve the situation. These weights are $\frac{\exp(Y_i)}{1-\exp(Y_i)}$. Guttorp (1987) presents a generalised least squares estimator, as follows. Let $\boldsymbol{Y} = (Y_1, \dots, Y_{n-1})^T$ and $\boldsymbol{m} = (m_2, \dots, m_n)^T$. Denote the covariance matrix of Y by $\boldsymbol{W}$. The generalised least squares estimator is then given by

$$\hat{\beta}_{GLS} = \frac{\boldsymbol{Y}^T \boldsymbol{W}^{-1} \boldsymbol{m}}{\boldsymbol{m}^T \boldsymbol{W}^{-1} \boldsymbol{m}}. \tag{3.27}$$

Guttorp (1987) uses an asymptotic approach that is valid for large n to estimate $\boldsymbol{W}$. Let $F(m_i) = \sum_1^i f_x(m_j)$ and $\bar{F}(m_i) = 1 - F(m_i)$. Now, let $w_i = \frac{F(m_i)}{n\bar{F}(m_i)}$. Then, the approximation holds

$$\boldsymbol{W} = \begin{bmatrix} w_1 & w_1 & \dots & w_1 \\ w_1 & w_2 & \dots & w_2 \\ \dots & \dots & \dots & \dots \\ w_1 & w_2 & \dots & w_k \end{bmatrix}. \tag{3.28}$$

Numerical experiments showed that this approximation was too crude. Although this leaves a gap in the theory, jack-knife or bootstrap procedures are robust numerical procedures to obtain an accurate approximation of the variance. The inverse of this is a band-diagonal matrix with elements W^{ij}, with band elements

$$\begin{aligned} W^{11} &= \frac{w_2}{w_1(w_2 - w_1)} \\ W^{12} &= \frac{1}{w_2 - w_1} \\ W^{jj-1} &= \frac{1}{w_j - w_{j-1}} \\ W^{jj} &= \frac{w_{j+1} - w_{j-1}}{(w_{j+1} - w_j)(w_j - w_{j-1})} \end{aligned} \tag{3.29}$$

$$W^{jj+1} = \frac{1}{w_{j+1} - w_j}, 2 \leq j \leq k-2$$

$$W^{k-2k-1} = \frac{1}{w_{k-1} - w_{k-2}}$$

$$W^{k-1k-1} = \frac{1}{w_{k-1} - w_{k-2}}.$$

As the function F is unknown, it is approximated by F_n. The variance is estimated by

$$Var\ \hat{\beta}_{GLS} = (\boldsymbol{m}^T \boldsymbol{W}^{-1} \boldsymbol{m})^{-1}. \tag{3.30}$$

The variance of the least squares estimate without intercept $\hat{\beta}_{LS}$ is estimated by

$$Var\hat{\beta}_{LS} = \frac{\boldsymbol{m}^T \boldsymbol{W} \boldsymbol{m}}{(\boldsymbol{m}^T \boldsymbol{m})^2}. \tag{3.31}$$

For the variance of least squares with intercept, it is replaced by $m - \overline{m}$, where $\overline{m} = 1^T \boldsymbol{m}/1^T 1$ and 1 is a $n-1$ vector of ones.

3.2. Estimation of the *b*-value for Catalogues with Time Intervals with Different Levels of Completeness

Hitherto, the discussion has been about the b-value alone. However, as has been noted earlier, the b-value and the RoS are related closely. These factors can be seen together in the GR relation, noting that the number of earthquakes occurs within a certain time interval. This can well be taken as, or rescaled to, a unit time interval, thereby transforming it into the RoS. The interrelation is clarified in this section of the thesis.

An early attempt at dealing with the estimation of the b-value from catalogues with time intervals with different levels of completeness was that by Molchan and co-workers (1970). Molchan *et al.* (1970; in Weichert [1980]) divide magnitudes into classes and use $\frac{n_i}{T_i}$, "event numbers divided by the time interval of completeness for each magnitude interval, as maximum likelihood estimator [of the annual seismicity rate]" (Weichert, 1980).

Two years after the publication of the work by Molchan *et al.* (1970), Stepp (1972) attempted to assess the seismicity parameters; however, he deals mainly with the incompleteness in his own work.

Nevertheless, the method by Stepp (1972) to assess incompleteness was the first to be implemented generally and it is still being used today. Stepp (1972) calculates the annual seismicity rate for each magnitude class with the data at hand. Afterward, he proceeds to estimate β and λ by linear least squares regression, assuming a Gutenberg–Richter FMSL law (this practice has already been criticised in the current study, but it is not essential to the steps of assessing incompleteness). However, the procedure is criticised by Weichert (1980) for failing to implement a maximum magnitude. It would be elucidating to conduct a study on the effects of the negligence of maximum magnitude in implementing the Stepp (1972) method.

Weichert (1980) proposes an estimation method for the b-value in an instance where the LoC changes with time. This is probably the most popular method, as it is rigorous but still understandable and reliable. Weichert (1980) divides magnitude range into magnitude classes, each with n_i events occurred with the time period of completeness t_i, and provides a maximum likelihood estimator, given by

$$\frac{\sum_i t_i \exp(-\beta m_i)}{\sum_j \exp(-\beta m_j)} = \frac{\sum n_i m_i}{N}, \tag{3.32}$$

that can be solved via iteration. The variance of this estimator is given by

$$\mathrm{var}(\beta) = \frac{1}{N} \frac{[\sum_i \exp(-\beta m_i)]^2}{[\sum_i t_i m_i \exp(-\beta m_i)]^2 - \sum_i t_i \exp(-\beta m_i) \sum_i \sum_i t_i {m_i}^2 \exp(-\beta m_i)}. \tag{3.33}$$

The accompanying annual event rate is given by

$$\lambda_0 = \frac{N \sum_i \exp(-\beta m_i)}{\sum_j t_j \exp(-\beta m_i)}, \tag{3.34}$$

and the variance of λ_0 is $\frac{\lambda_0}{N}$. For large enough total number of events N, λ_0 can be considered as distributed normally, and confidence intervals can be established.

Rosenbleuth (1986) derives a joint likelihood equation for β and λ. Dividing the magnitude range into intervals with size Δm, and assuming a Poisson process, the probability that an earthquake of magnitude m_i will occurr within the time interval t is

$$-\lambda_i' \Delta m e^{\lambda_i' t \Delta m}. \tag{3.35}$$

The probability that no earthquake would occur in the magnitude segment containing m_j is

$$e^{\lambda'_j t\Delta m}. \tag{3.36}$$

Finally, the likelihood of the occurrence of earthquakes m_i is

$$\mathcal{L}(\boldsymbol{M}|\beta,\lambda) = \prod_i (-\lambda'_i) exp\left(\sum_k \lambda'_k t\Delta m\right), \qquad k = i,j, \tag{3.37}$$

and by means of taking the limit as $\Delta m \to 0$ as an idealised case,

$$\prod_i (-\lambda'_i)\, exp(-\lambda_0 t). \tag{3.38}$$

For sub-catalogues with different levels of completeness, this becomes

$$\prod_i (-\lambda'_i)\, exp \sum_k (-\lambda_{0k} t_k). \tag{3.39}$$

Noting that $-\lambda'_i = \frac{d\lambda_i}{dm} = \frac{d}{dm}\lambda_0(1 - F(m)) = \lambda_0 f(m)$, and elaborating, it is found that

$$\mathcal{L}(\boldsymbol{M}|\beta,\lambda) = const.\, {\lambda_0}^n \beta^n \prod_i \exp(\beta(m_i - m_{min})) \sum_k exp(-\alpha_k \tau_k), \tag{3.40}$$

where $\alpha = \frac{\lambda_0}{(e^{-\beta m_i} - e^{-\beta m_{max}})}$ and $\tau = \left(e^{-\beta m_{0k}} - e^{-\beta m_{max}}\right) t_k$.

Weichert (1980) and Rosenblueth (1986) prefer to decouple equations for the sake of using prior distributions in a Bayesian scheme; however, this is not actually possible because of the presence of the term $\alpha_k \tau_k$. Details on the approximate decoupling are available in Rosenblueth (1986); however, the Rosenblueth method is not popular. Nevertheless, an investigation of the method can be considered worthwhile.

Realising the need to include historical data, Kijko and Sellevoll (1989) propose another procedure to simultaneously determine the MLE of β and λ_0. The importance of this work can hardly be overstated. Historical values are often discarded or not used to their full potential because of the lack of methodology to incorporate them. The procedure by Kijko and Sellevoll (1989) allows for historical data to be included in the estimation, the so-called extreme part of the catalogue — so-called because there is good reason to assume that the distribution of these values can be modelled by an extreme value distribution. In

addition, Kijko and Sellevoll (1989) steer readers to including other forms of information by using the Bayesian formalism, noting that the likelihood function extends easily to a Bayesian formalism if required. Smit *et al.* (2019) elaborate on the applicable method for such an instance, and this landmark work should serve as a guideline for the use of the Bayesian formalism for including information on seismicity parameters in PE-PSHA.

The probability that the largest magnitude in a time span of t years will be less than a given level x is given by

$$g(x,t) = \exp\left[-\nu_0 t\left(\frac{\exp(-\beta m_{max}) - \exp(-\beta x)}{\exp(-\beta m_{max}) - \exp(-\beta m_0)}\right)\right], \tag{3.41}$$

where $\nu_0 = \lambda(1 - F(m_0))$, $F(.)$ is the cumulative Gutenberg–Richter distribution, and m_0 is the threshold magnitude for the extreme part of the catalogue. If the extreme part of the catalogue consists of n_0 events with magnitudes $\boldsymbol{m_0} = (m_{01,}m_{02}, \dots, m_{n_0})$ occurring within time intervals $\boldsymbol{t_0} = (t_{01}, t_{02}, \dots, t_{n_0})$, the likelihood function for the extreme part of the catalogue is

$$L_0(\boldsymbol{m_0}, \boldsymbol{t_0}|\beta, \lambda) = \prod_{i=1}^{n_0} g(m_{0i}, t_{0i}|\beta, \lambda). \tag{3.42}$$

The likelihood function for β for the i^{th} complete sub-catalogue of magnitudes $\boldsymbol{m_i}$, each with a number of observations n_i and completeness m_{ci}, is given by

$$L_{i\beta}(\boldsymbol{m_i}|\beta) = \frac{\beta^{n_i}\exp\left(-\beta\sum_{j=1}^{n_i} m_{ij}\right)}{[\exp(-\beta m_{ci}) - \exp(-\beta m_{max})]^{n_i}}. \tag{3.43}$$

Assuming a homogeneous Poisson process, the likelihood function for λ is given by

$$L_{i\lambda}(T_i, n_i|\lambda) = const\,\exp(-\nu_i T_i)\,(\nu_i T_i)^{n_i}, \tag{3.44}$$

where $const$ is a normalising constant, T_i is the time length of the i^{th} complete sub-catalogue, and $\nu_i = \lambda(1 - F(m_i))$. Therefore, for the i^{th} complete sub-catalogue, the likelihood function for the parameter pair (β, λ) is

$$L_i(\boldsymbol{m}_i, T_i, n_i|\beta, \lambda) = L_{i\beta}(\boldsymbol{m}_i|\beta)\; L_{i\lambda}(T_i, n_i|\lambda), \tag{3.45}$$

and, for the entire catalogue, the combined likelihood function is

$$L(\beta,\lambda) = L_0(\boldsymbol{m_0}, \boldsymbol{t_0}|\beta,\lambda)\prod_{i=1} L_i(\boldsymbol{m}_i, T_i, n_i|\beta,\lambda). \tag{3.46}$$

Kijko and Sellevoll (1989) present an analytic solution for the likelihood function (3.46), but it can be readily solved numerically. For the standard error of this MLE, the approximate variance–covariance matrix is calculated as $D(\boldsymbol{\Theta}) = A^{-1}$, where A is described by

$$A = \{a_{ij}\} = -\frac{\partial^2 \ln(L)}{\partial\Theta_i \partial\Theta_j}, \tag{3.47}$$

where $\boldsymbol{\Theta} = (\lambda, \beta)$.

The maximum likelihood equations of Weichert (1980) differ from those of Kijko and Sellevoll (1989). Those of Weichert (1980) are based on magnitude classes, whereas those of Kijko and Sellevoll (1989) are based on the assumption of continuous magnitude values. However, it should be possible to prove that a discretised version of the Kijko and Sellevoll (1989) solution is the same as that of Weichert (1980). This is an important topic for future research.

Kijko and Smit (2012) realised the need for a simpler, generalised estimator that could accommodate catalogue sub-intervals of varying levels of completeness based on the Aki–Utsu estimator. These authors extended the Aki–Utsu maximum likelihood estimator of β for catalogues with time intervals with different levels of completeness in a most elegant form. Suppose there are s sub-catalogues and the i^{th} sub-catalogue containing n_i earthquakes have an LoC m_c^i, and spans a period of t_i years. Accordingly, maximising the likelihood function

$$L\left(\beta\middle|m_j^i\right) = \prod_{i=1}^{s}\prod_{j=1}^{n_i} f(m_j^i|\beta) = \prod_{i=1}^{s}\prod_{j=1}^{n_i} \beta\exp(-\beta(m_j^i - m_c^i)), \tag{3.48}$$

is done through the simple formula

$$\hat{\beta} = \left(\frac{r_1}{\hat{\beta}_1} + \frac{r_2}{\hat{\beta}_2} + \cdots + \frac{r_s}{\hat{\beta}_s}\right)^{-1}, \tag{3.49}$$

where $r_i = \frac{n_i}{\sum_{j=1}^{s} n_j}$ and $\hat{\beta}_i$ is the Aki–Utsu estimator applied to the i^{th} sub-catalogue. The estimator is approximately normally distributed, with a standard deviation

$$\sigma_{\hat{\beta}} = \frac{\hat{\beta}}{\sqrt{N}}, \tag{3.50}$$

where $N = \sum_{i=1}^{s} n_i$. This is the same formula as that derived for the classic Aki–Utsu estimator and applied to one complete catalogue. The confidence intervals of the $\hat{\beta}$ estimators are given by

$$\left(\hat{\beta} - z_{\frac{\alpha}{2}}\sigma_{\hat{\beta}}^2, \hat{\beta} + z_{\frac{\alpha}{2}}\sigma_{\hat{\beta}}^2\right), \tag{3.51}$$

where $z_{\frac{\alpha}{2}}$ is the $(1 - \frac{\alpha}{2})$ quantile of the standard normal distribution. The activity rate λ corresponding to m_{min} is given by

$$\lambda = \frac{N}{\sum_{i=1}^{s} t_i \exp(-\beta(m_c^i - m_{min})}. \tag{3.52}$$

3.2.1. Two additional estimators by Kijko (2017)

Kijko (2017) presents two more estimators for β, on which discussions follow here. Choose an arbitrary lower limit m_0 for magnitudes. The mean value of the magnitude for any period of time is $\overline{m}^{(0)} = \sum_{j=1}^{n_0} m_j{}^{(0)}/n_0$, where $j = 1, \dots, n_0$. Here, the value m_0 is reference magnitude, specifically less than or equal to all levels of completeness of sub-catalogues. If it were known how many events n_0 of events $m_j^0 \geq m_0$ have occurred during the time interval of the incomplete catalogue, the Aki–Utsu estimator (3.2) can be applied

$$\beta = \frac{1}{\overline{m}^{(0)} - m_0}. \tag{3.53}$$

The next logical step is to estimate n_0. The number of events in the interval $[m_0, \infty)$ in the i^{th} sub-catalogue can be estimated as

$$n_{0i} = \frac{n_i}{\left[1 - F(m_c^{(i)})\right]}, \tag{3.54}$$

where $F\left(m_c^{(i)}\right) = 1 - \exp\left[-\beta\left(m_c^{(i)} - m_0\right)\right]$. Following (3.54), the sum of magnitudes in the i^{th} sub-catalogue, the interval $[m_0, \infty)$, is

$$\sum_{j=1}^{n_{0i}} m_{i,j}^{(0)} = \frac{n_i\left[\bar{m}_i - \left(m_c^{(i)} - m_0\right)\right]}{\left[1 - F(m_c^{(i)})\right]}, \tag{3.55}$$

and for the entire catalogue, where $m_{i,j}^{(0)}$ is the magnitude of the j^{th} event in the interval $[m_0, \infty)$. For the entire catalogue, consisting of s sub-catalogues, the sum of the magnitudes in $[m_0, \infty)$ is

$$\sum_{j=1}^{n_0} m_j^{(0)} = \sum_{i}^{s}\sum_{j=1}^{n_{0i}} m_{i,j}^{(0)} = \sum_{j=1}^{s} \frac{n_i\left[\bar{m}_i - \left(m_c^{(i)} - m_0\right)\right]}{\left[1 - F(m_c^{(i)})\right]} \tag{3.56}$$

Finally, it is possible to estimate the mean value of all the magnitudes in the interval $[m_0, \infty)$ that occurred over the time span of the entire catalogue

$$\begin{aligned}\bar{m}^{(0)} = \sum_{j=1}^{n_0} m_j^{(0)}/n_0 &= \sum_{i}^{s}\left(\sum_{j=1}^{n_{0i}} m_{i,j}^{(0)}\right)/n_0 \\ &= \sum_{j=1}^{s} \frac{n_i\left[\bar{m}_i - \left(m_c^{(i)} - m_0\right)\right]}{\left[1 - F(m_c^{(i)})\right]} \Big/ \sum_{i=1}^{s}\left(\frac{n_i}{\left[1 - F(m_c^{(i)})\right]}\right).\end{aligned} \tag{3.57}$$

Now, it is possible to substitute equation (3.57) into estimator (3.53), the result of which is

$$\hat{\beta} = 1\Big/\left(\frac{\sum_{i=0}^{s} \frac{n_i\left[\bar{m}_i - \left(m_c^{i}\right)\right]}{1 - F\left(m_c^{(i)}\right)}}{\sum_{i=0}^{s} \frac{n_i}{1 - F\left(m_c^{(i)}\right)}} - m_0\right). \tag{3.58}$$

As β occurs in the distribution $F(m)$ it appears on both sides of the equation. In numerical experiments, Kijko (2017) finds that iteration of the equation quickly converges. This was also the finding of the current author. A rough estimate of the standard deviation of the estimator (3.58) is merely that of the Aki–Utsu estimator, which is approximately (Aki, 1965)

$$\sigma_\beta = \frac{\beta}{\sqrt{n_0}} \tag{3.59}$$

In the second estimator by Kijko and Smit (2017), iteration is avoided, and their derivation starts from the same point as their derivation of the estimator (3.58). However, it will not be derived here, as it is quite intuitive, taking a form resembling estimator (3.49). Kijko (2017) merely substitutes the time fraction $\tau_i = \frac{t_i}{t}$ for the fraction of events $r_i = \frac{n_i}{n}$.

$$\hat{\beta} = \left(\frac{\tau_1}{\hat{\beta}_{01}} + \frac{\tau_2}{\hat{\beta}_{02}} + \cdots + \frac{\tau_s}{\hat{\beta}_{0s}} \right)^{-1}, \tag{3.60}$$

and $\hat{\beta}_{0i}$ is estimated using the standard Aki–Utsu estimator over the respective sub-catalogue. Note that these methods present "weights" that differ from the original extended Aki–Utsu estimator (3.49). The Kijko (2017) method disregards the fact that there are fewer data points in some catalogues because of different levels of completeness, and only considers the time span of the catalogue, corresponding to the estimated ratio of the number of events in a Poisson process with the same intensity λ. Intuitively, this appears to render it inferior to estimator (3.49). However, such assumption disregards the fact that the relative number of events stems from a significantly more complex derivation of this method, where a hypothetical number of events in a sub-catalogue are assumed above the chosen reference level m_0.

3.2.2. Estimator by Ordaz and Giraldo (2018) as a special case of the Kijko–Sellevoll procedure

Ordaz and Giraldo (2018) attempt to improve on both the Aki–Utsu maximum likelihood estimator and that of Kijko and Smit (2012) for β and λ by proposing a joint maximum likelihood estimation of the pair (β, λ). This is considered a theoretical improvement, as it uses not only the marginal likelihoods of β and λ but also their simultaneous likelihood. In addition, Ordaz and Giraldo (2017) show numerically that this estimation is superior to separate, mariginal maximum likelihood estimators. However, the current author showed that, surprisingly, the Ordaz and Giraldo (2018) equations present only a special case in the scheme developed by Kijko and Sellevoll (1989). When no extreme part of the catalogue is used, and it is supposed that $m_{max} = \infty$, the equations look exactly the same as those derived from the scheme by Kijko and Sellevoll (1989). The derivation of the likelihood functions differs slightly from that of Kijko and Sellevoll (1989) in that Ordaz and Giraldo (2018) use every single interval between consequtive earthquakes, whereas Kijko and Sellevoll (1989) use the total time span of each complete sub-catalogue. It turns out that these differences in the derivation lead to equivalent likelihood functions, as could be

expected because the likelihood functions derive from the same initial distributions of magnitude and inter-event time distribution. However, note that Kijko and Sellevoll (1989) use the probability of observing n_i earthquakes in a time T_i equation (3.44), whereas Ordaz and Giraldo (2017) use the the inter-event time distribution. From a logical point of view, however, the equivalence of the characterisations in the Poisson process (as a counting process, or distribution of inter-event times) dictates that the outcome should be equivalent. The likelihood function that Ordaz and Giraldo (2017) arrive at is

$$\hat{\beta} = \left(\frac{\tau_1}{\hat{\beta}_{01}} + \frac{\tau_2}{\hat{\beta}_{02}} + \cdots + \frac{\tau_s}{\hat{\beta}_{0s}}\right)^{-1}, \tag{3.61}$$

$$\mathcal{L}(\lambda_0, \beta) = \prod_{j=1}^{s}\prod_{i=1}^{n_j} {\lambda_{01}}^N \exp\left[-\beta \sum_{j=1}^{s} n_j\,(m_{0j} - m_0)\right] \exp\left[-\lambda_0 \sum_{j=1}^{L} T_j \exp\left[-\beta(m_{0j} - m_0)\right]\right] \prod_{j=1}^{L} \beta^{n_j} \exp\left[-\beta^{n_j} \sum_{i=1}^{s}\sum_{j=1}^{n_j} (m_{i,j} - m_{0j})\right], \tag{3.62}$$

where λ_{0j} is the seismic rate corresponding to earthquakes above the LoC m_{0j}. This corresponds exactly to the equations of Kijko and Sellevoll (1989) without the inclusion of the extreme part of the catalogue. The explicit equations given by Ordaz and Giraldo (2018) correspond to those obtained from the general closed-form solutions of Kijko and Sellevoll (1989). The explicit equations of Ordaz and Giraldo (2017) maximise the likelihood equation (3.46), whereas those of Kijko and Sellevoll (1989) span a wider range of cases (i.e. an imposed maximum magnitude, and an extreme [historical] part of the catalogue, which are ignored by Ordaz and Giraldo [2018]); however, the results are the same for the special case.

The nonlinear system of two equations is

$$\frac{N}{\hat{\lambda}_0} - T^* = 0, \tag{3.63}$$

$$\frac{N}{T^*}\sum_{j=1}^{s} T_j^* (m_{0j} - m_0) + \frac{N}{\hat{\beta}} - (Q + S) = 0 \tag{3.64}$$

where

$$T_j^* = T_j \exp\left[-\hat{\beta}(m_{0j} - m_0)\right], \tag{3.65}$$

$$T^* = \sum_{j=1}^{L} T_j^*, \tag{3.66}$$

$$Q = \sum_{j=1}^{s} n_j (m_{0j} - m_0), \tag{3.67}$$

$$S = \sum_{j=1}^{s} S_j, \tag{3.68}$$

$$S_j = \sum_{i=1}^{n_j} (m_{i,j} - m_0). \tag{3.69}$$

The exact details of the derivation of equivalence of the two estimators are given in Vermeulen and Kijko (2018).

It has to be acknowledged that the b-value varies through space and time. Probably the first researchers to deal with this phenomenon are Shi and Bolt (1982), Guttorp and Hopkins (1986) and Ogata and Katsura (1993). Ogata and Katsura (1993) compose a likelihood function for parameters including $\beta(t, x, y)$, and use a cubic spline method to interpolate $\beta(t, x, y)$. Shi and Bolt present their general result, as discussed in the previous section, but they note that their results are applicable to the b-value varying in space and time. Guttorp and Hopkins (1986) use a running time window to estimate $\beta(t)$, showing that

$$E[\hat{\beta}(t)] = \beta(t) + O\left(\frac{1}{Kp}\right) + O(K), \tag{3.70}$$

and

$$Var[\hat{\beta}(t)] = \left(\frac{\sinh(\frac{\Delta\beta}{2})}{\frac{\Delta}{2}}\right)^2 + o\left(\frac{1}{Kp}\right), \tag{3.71}$$

for the estimator

$$\hat{\beta} = \frac{1}{\Delta}\log\left(1 + \frac{\Delta}{\bar{M} - M_0}\right), \tag{3.72}$$

for a Poissonian rate p and time window $[t - K, t + K]$, and bin width Δ.

Cao and Gao (2002) observe temporal changes in the b-value beneath the northeastern Japan Island Arc over the period 1991–1995 relative to that over the period 1984–1990. According to Hutton *et al.* (2010), the b-value did not change much in southern California over the period 1932–2008. Wyss *et al.* (2000) note that asperities on faults have the lowest b-values. However, Wyss and Wiemer (2000) show that the 7.3 magnitude Landers earthquake in 1992 caused a significant change in the b-value in much of southern California.

Godano *et al.* (2014) note that several authors (listed in Godano *et al.*, 2014) have investigated the spatial and temporal variability of the b-value and proposed various physical mechanisms for this variability (assuming that it is not caused by variability and/or bias in the statistical estimation). According to some other authors (also listed in Godano *et al.*, 2014), the b-value does not vary significantly for tectonic earthquakes, based on the observation that the distribution of the seismic moment is quite stable in space and time.

Godano *et al.* (2014) investigated the distribution of b-values. They consider three different regimes for the variability of the b-value, which are (1) it stems from the inclusion of the magnitudes below the magnitude of completeness, (2) the slope changes toward larger magnitudes, (3) the b-value does not take on a single value but varies stochastically in terms of space–time. Chan *et al.* (2012) estimated the spatial distribution of b-values in Taiwan one year prior to the occurrence of several large earthquakes, as shown in Figures 7 (a) and (b). The earthquakes were located in areas with slightly lower b-values than those of the surrounding areas. Furthermore, these anomalies in the b-values appear to evolve with time until the actual earthquake occurs. Similarly, in investigating the San Andreas Fault, Wiemer and Wyss (1997) hypothesise that "highly stressed asperities", which are potential regions for earthquakes, can be identified by low b-value anomalies.

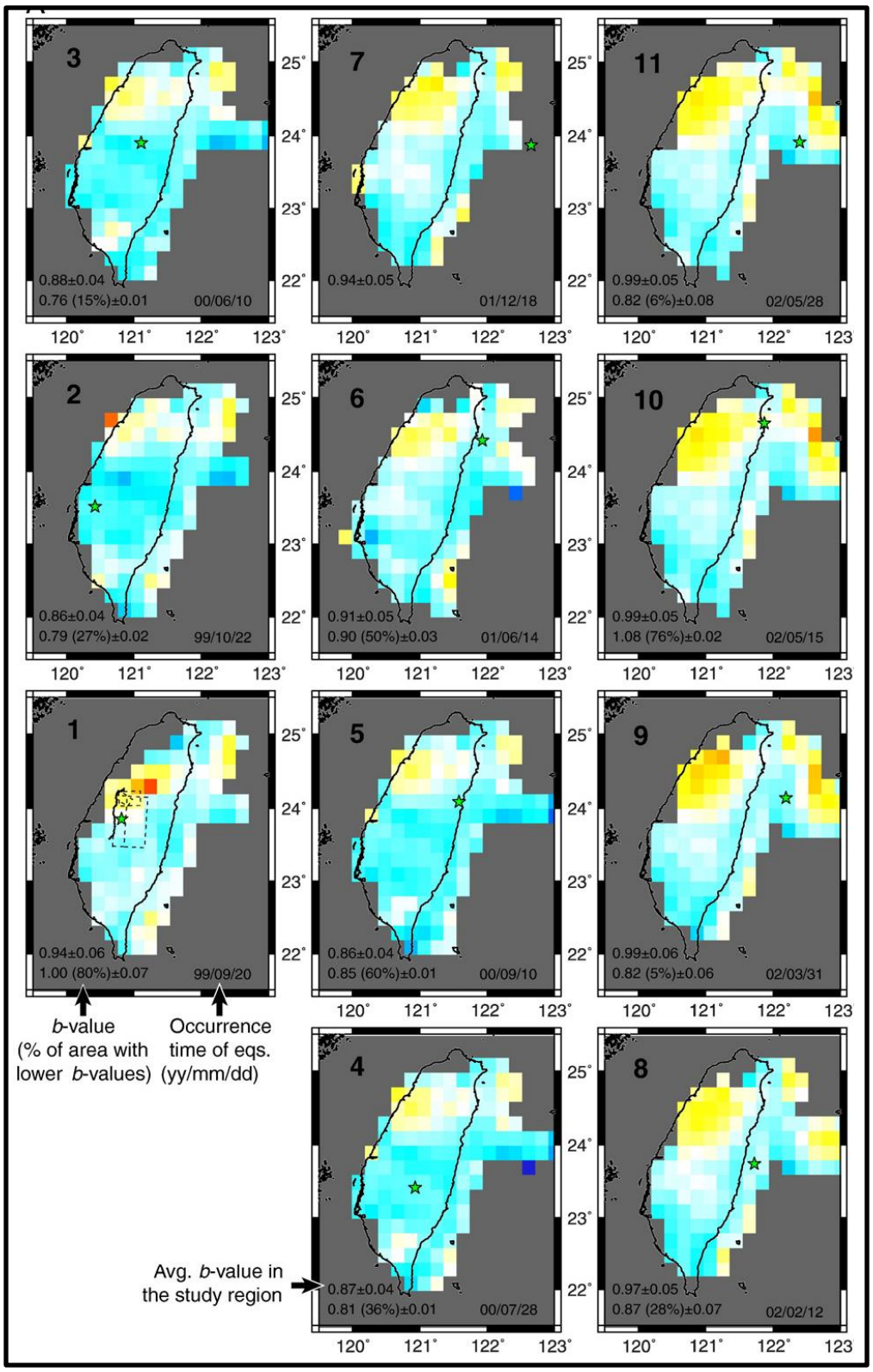


Figure 7(a). Spatial distribution of *b*-values in Taiwan one year prior to the occurrence of several large earthquakes (from Chan *et al.,* 2012).

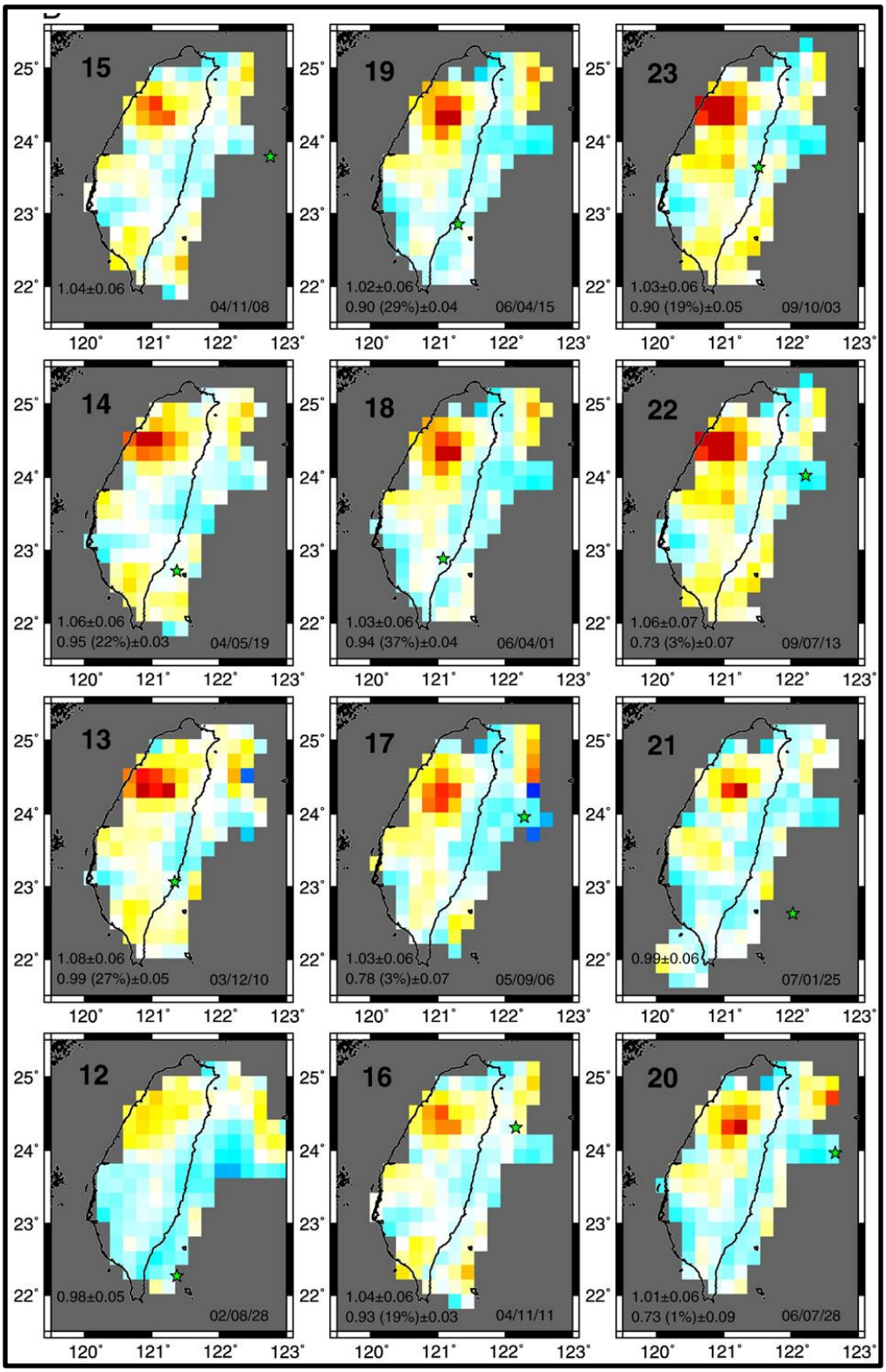


Figure 7(b). Spatial distribution of b-values in Taiwan one year prior to the occurrence of several large earthquakes (from Chan *et al.,* 2012).

Classic PSHA assumes that earthquakes follow a Poissonian distribution in time. The Poisson process is the result of events occurring independently. For a given time interval, the probability of having n events is (Kramer, 1996)

$$P[N = n] = \frac{(\lambda t)^n e^{-\lambda t}}{n!}. \tag{3.73}$$

The probability that at least one event occurs is

$$P[N \geq 1] = 1 - e^{-\lambda t}. \tag{3.74}$$

For small probabilities in a unit time interval, the following approximation can be made (Kijko, 2011)

$$P[N \geq 1; t = 1] = 1 - e^{-\lambda} = 1 - (1 - \lambda + \frac{1}{2}\lambda^2 - \cdots) \approx \lambda \tag{3.75}$$

If the RoE of earthquakes with magnitude m were λ_m, the RoE λ_{m1} of m_1, with $m_1 > m$ is related to λ_m through the GR relation from the relationship

$$\lambda_{m1} = \lambda_m(1 - F(m_1|m_c = m)), \tag{3.76}$$

where $F(m_1|m_c = m)$ is the cumulative GR relation in distribution form. Characteristic of the Poisson process is that λ stays constant through time and, in many instances, this assumption is sufficient for hazard analysis. However, there is ample reason to believe that $\lambda \equiv \lambda(t)$ varies with time. In the following section, this topic is considered in more detail. Most of the models dealt with in the next section are not readily implementable yet; however, they are discussed not only for the sake of completeness but also because research is required to incorporate these models and methods into readily implementable PSHA practice.

3.3. Non-Poissonian Seismicity Models

3.3.1. Renewal processes

The inter-event time distribution defines the renewal process completely (Polidoro *et al*., 2013). The renewal process is a generalisation of the stationary Poisson process, where the inter-event time distribution may take on different distributions from the exponential distribution. The result is a process where the probability of the next event is dependent on the time occurrence of the previous event. In

other words, the process is not ''memory-less'' as in the case of the Poisson process. For the property of ''memory-lessness'' to hold, the inter-event time distribution has to be exponential. The term ''renewal process'' alludes to the property that the probability of the next event occurring in a time interval from the present depends on the time of occurrence of the previous event, but events prior to that have no effect. Therefore, the process, or more specifically, the ''memory'' of the process is reset. The most popular of such methods is probably the Brownian Relaxation Oscillator model introduced by Matthews *et al*. (2002). This model features a deterministic oscillator, moving between the minimum and maximum rupture potential, with a stochastic Brownian motion added to the motion of the oscillator. The authors of this method specifically decided not to attach any physical state variable to the rupture potential, as they view the total rupture potential that correlates with, but is not completely described by, such variables of the state as strain, stress, moment deficit, and the Coulomb stress history on a fault. The inter-event time distribution resulting from this model is an inverse Gaussian distribution (Matthews *et al*., 2002)

$$f(t;\mu,\alpha) = \left(\frac{\mu}{2\pi\alpha^2 t^3}\right)^{1/2} exp\left[-\frac{(t-\mu)^2}{2\mu\alpha^2 t}\right], \tag{3.77}$$

where α is the coefficient of variation and μ is the mean recurrence interval length. Matthews *et al*. (2002) view the coefficient of variation as of greater importance than the standard deviation because of its role in the interpretation and use of the model. The coefficient of variation is also called the aperiodicity, as it describes how many inter-event times are clustered around the mean value (which may be seen as the period). The model by Matthews *et al*. (2002) is presented as a theoretical model that competes well with models chosen based on empirical analysis, such as the lognormal, Weibull, and gamma distributions. This model is attractive because it is not only derived theoretically but also competes well with the empirical models.

The lognormal distribution is desirable for its wide application and familiarity, and because it relates closely to the normal distribution, whereas the Weibull and gamma models are desirable because of their wide application in failure analysis (Matthews *et al*., 2002), and because they include the exponential distribution of the Poisson process as a special case (Matthews *et al.*, 2002).

Pandey and van Noortwijk (2004) use the Erlang distribution for time-dependent structure reliability analysis. The substantiation is analogous to the argument used by Matthews *et al.* (2002) for the Brownian Passage-time model; however, in this instance, the added stochastic motion does not take on a normal distribution. Instead, a gamma distribution is used. Polidoro *et al*. (2013) note the applicability of the model to earthquake inter-event times, and that, compared with the Brownian Passage-time model, it allows the analytic formulation of the counting process.

3.3.2. Markov renewal processes

These processes allow for dependence between the magnitude and arrival times of earthquakes while maintaining the renewal property (properties of the next arrival are independent of all but the last arrival) (Polidoro *et al*., 2013). The two models of note in this category (also the only two listed by Polidoro *et al*., 2013) are the slip-predictable and the time-predictable models.

The slip-predictable model represents a regime in which stress accumulates from a ground state for a random period of time, after which it reverts to the ground state again. The magnitude of the earthquake is determined by such a reduction in stress and is, therefore, dependent in a deterministic fashion on the time that has elapsed since the last event. The inter-event times are modelled by a Weibull distribution. Note that the distribution of inter-event times uniquely determines the distribution of magnitude (Polidoro *et al*., 2013).

The time-predictable model, on the other hand, starts by specifying a fixed load state above the ground state at which rupture will take place. The rate at which the load increases is an exact function of time. The energy release, however, is random when the rupture takes place. The inter-event time between the last event and the next event is, therefore, dependent on the magnitude of the previous event (Polidoro *et al*., 2013).

3.3.3. Earthquake clusters and aftershocks

Earthquakes tend to occur as clearly recognisable clusters in time and space in some regions, such as Japan (e.g. Ogata, 1988) and California (USA) (e.g. Reasenberg and Jones, 1989). Usually, one earthquake is considered the main shock, those before it as foreshocks, and those after it as aftershocks. It is commonly assumed that the main shock should be the largest in the cluster (or sequence) of earthquakes. Omori (1894) proposes a mathematical relation, later generalised by Utsu and Ogata (1995), to describe the decay of aftershock rates with time after the main shock. The generalised version by Utsu and Ogata (1995) is given by

$$\lambda_A(t) = \frac{K}{(t-c)^p}, \tag{3.78}$$

where λ_A is the rate of aftershock occurrence, and K, c, and p are constants. Equation (3.78) is often called the modified Omori's law or the Omori–Utsu formula. The modified Omori's law appears to be accepted generally.

Yeo and Cornell (2005) devised a method to assess the hazard of aftershocks for a given main shock in a time-dependent fashion. They call their method aftershock PSHA (or APSHA). The essence of the method

is integration over a given time interval of a Gutenberg–Richter-type frequency-magnitude law that is scaled according to the modified Omori's law. Specifically, the mean RoS over a given time is given by

$$\tilde{\mu}(y,t,T;m_m) = \left\{\frac{10^{a+b(m_m-m_{min})}-10^{a}}{p-1}[(t+c)^{1-p}-(t+T+c)^{1-p}]\right\} \times \iint_{R}^{m_m}\int_{m_{min}} P[Y>y|m,r]f_{R|M}(r|m)f_M(m;m_m)drdm, \quad (3.79)$$

where a and b are the constants of the GR relation, m_m is the magnitude of the main shock, and m_{min} is the lower limit of magnitudes of interest. The term in curly brackets is the Gutenberg–Richter-type FMSL that is scaled according to the modified Omori's law, and the integral term corresponds to the integral term on a typical PSHA used to rescale the seismicity rate (see, e.g. Kramer, 1996).

Iervolino *et al*. (2014) devised a methodology to include aftershock sequences in a long-term PSHA using a modelling paradigm, where the main shock and the aftershocks are considered to occur in a single instant of time. They call this method sequence-PSHA (or SPSHA). They show that the effect of aftershocks on the exceedance of some measure of intensity over the long term can be accounted for by multiplying the recurrence rate of main shocks by an appropriate term.

3.3.4. Seismic hazard analysis with the ETAS model

The Epidemic Type Aftershock Sequence (ETAS) model, introduced by Ogata (1988), is not based on the assumption that all main shocks are greater than their aftershocks. Moreover, it allows for aftershocks to have aftershocks themselves, and an aftershock can have more than one main shock, or parent (Zhuang *et al.*, 2012). In fact, this model does not distinguish clearly between main shocks and aftershocks. The recurrence rate of the ETAS model is given by:

$$\lambda(t) = \mu + \sum_{t_i<t} c(m_i)g(t-t_i), \quad (3.80)$$

where μ is the main shock recurrence rate, m_i is the magnitude of the i^{th} earthquake, and $c(\cdot)$ and $g(\cdot)$ specify the dependence of the aftershock recurrence rate depending on magnitude and time, respectively. In equation (3.80), the modified Omori's law is used often for the function $g(t-t_i)$, and the PDF corresponding to the Gutenberg–Richter relation is used for $c(m_i)$.

Beauval *et al.* (2006) account for aftershocks by using Monte Carlo simulations of the ETAS model. Iervolino *et al*. (2014) mention the possibility for their model to be adapted to account for aftershocks occurring according to the ETAS model.

3.3.5. Seismic hazard analysis with other non-Poissonian models

For a Poisson point process, the following approximation can be made

$$\Pr[T \leq t] = 1 - e^{\int_0^t \lambda(s)ds} = 1 - \left(1 - \int_0^t \lambda(s)ds + \frac{1}{2}\left[\int_0^t \lambda(s)ds\right]^2 - \cdots\right)1 \approx \int_0^t \lambda(s)ds, \qquad (3.81)$$

where the last approximation holds if $\max_{s\in[t0,t)}\{\lambda(s) \times t\}$ were small. This is a generalisation of, e.g. Cornell and Winterstein (1988). Polidoro *et al.* (2013) provide the elegant and simple approximation

$$\begin{aligned}\Pr[X \geq a|N(t_0)\ \&\ last\ event\ at\ t_0] &\approx \left[1 - \int_0^t f_t(s|\mathrm{H_t})ds\right] \\ &\times \int_{m_{min}}^{m_{max}} P[IM > a|m,r] \cdot f\big(m \,\big|\, N(t_0) = 0\ \&\ N(t_0 + t) = 1\big)dm.\end{aligned} \qquad (3.82)$$

Approximations (3.81) and (3.82) hold for several models used in practice; however, the general case is history dependent and, therefore, the analysis must be formulated carefully for the present time t_0 , as well as the projected time.

It is important to note that there are time-dependent models that are renewal point processes, i.e. the hazard rate depends merely on the time elapsed since the previous event, and fully history-dependent models, in which the complete history (or a large part of it) plays a role.

A general history-dependent model, such as the ETAS model, can become almost overly complex, as $\lambda(u|\mathrm{H_t}) = \lambda\big(u|\mathrm{H_{t_0}}\big) + \lambda\big(u|\mathrm{H_{t_0}}\big)$ must be accounted for. This implies that the hazard rate at the time $\mathrm{t_0} + \mathrm{t}$ (i.e. t time units from now) depends on both the history $\mathrm{H_{t_0}}$ up until the present and the $\mathrm{H_{t_0+t}}|\mathrm{H_{t_0}}$, i.e. what happens in t time units from now. This can render approximations (3.81) and (3.82) void, unfortunately. In such an instance, the sum has to be evaluated

$$P[IM > im] = \sum_{n=1}^{\infty} P[IM > im|N(t_0) \cap N(t) = n] \times P[N(t) = n|N(t_0) = 0], \qquad (3.83)$$

where IM is some intensity measure, and N(t) is the counting process. The second term is an integral over a random process (Daley and Vere-Jones, 2007), of which the theory is beyond the scope of the current work. However, Beauval *et al.* (2006) did conduct an investigation by using Monte Carlo simulations.

As regards the models that do allow practically applicable PSHA, such as those discussed in section 2, equation (3.82) holds. It is worth noting the series of approximations from which equation (3.82) follows (Polidoro *et al.*, 2013)

$$\begin{aligned} P[IM > im|\, N(t_0) = 0] &= \sum_{n=1}^{\infty} P[IM > im|N(t_0) = 0 \cap N(t) = n] \times P[N(t) = n|N(t_0) = 0] \\ &\approx P[IM > im|N(t_0) = 0 \cap N(t) = 1] \times P[N(t) = 1|N(t_0) = 0] \\ &\approx P[IM > im|N(t_0) = 0 \cap N(t) = 1] \times P[N(t) \geq 1|N(t_0) = 0] \\ &= P[IM > im|N(t_0) = 0 \cap N(t) = 1] \times \{1 - P[N(t) = 0|N(t_0) = 0] \\ &= P[IM > im|N(t_0) = 0 \cap N(t) = 1] \times \int_m P[IM > im|m, r] \cdot f_{M|N(t_0)=0 \cap N(t)=1}(m) dm. \end{aligned} \tag{3.84}$$

The physical interpretation of these approximations is that the hazard can be approximated by the probability of one event occurring, as the rest of the events have a negligible effect on probability in typical situations, such as long recurrence intervals of earthquakes. This is the so-called first-event approximation (Cornell and Winterstein, 1988).

The first-event approximation is useful in all the instances described in Polidoro *et al.* (2013). These include the Brownian Passage-time model (or inverse Gaussian renewal process), the inverse Erlang renewal process, the inverse gamma renewal process, the homogeneous Poisson process, the slip-predictable model, and the time-predictable model. Note, however, that it only holds for the first-event approximation and the hazard function $h_t(\cdot)$ effectively resets when a new event occurs.

Once the overhead seismicity rate becomes large, the probability of the actual renewal of the process has to be dealt with and integrated over all the possible courses the process might take. No analytical approximations appear to have been reported in the literature for an instance of the overhead seismicity not being small enough to allow for the first-event approximation. Cornell and Winterstein (1988) use numerical models (Monte Carlo models) to determine hazard. Therefore, the use of analytic approximations is constrained to instances with low seismicity, unless new analytic forms can be developed for more general cases. Nevertheless, the first-event approximation is probably applicable in most instances (see, e.g. Cornell and Winterstein, 1988).

A useful property that holds in many instances (exceptions include the Markov renewal processes) is that both λ and h for a lower magnitude can be rescaled for higher magnitudes as (derived from Cornell and Winterstein, 1988)

$$\lambda_{m2} = \lambda_{m1}(1 - F_M(m2|m_{min} = m1)), \tag{3.85}$$

$$h_{m2} = h_{m1}\big(1 - F_M(m2|m_{min} = m1)\big),$$

where $m2 > m1$, and F_M is the distribution function of magnitude.

Unfortunately, one of the fundamental equations of classical PSHA is refuted in the general case of a point process, which is that the probability of exceeding an event at least once in a given time period is

$$P[Y > y; t] = \ 1 - \exp[-\lambda(y) \cdot t] \tag{3.86}$$

For a non-homogeneous Poisson process $\lambda,\cdot\ t$ can be replaced by $\Lambda(t) \coloneqq \int_{t0}^{t} \lambda(t)dt$, but the exponential distribution in equation (3.86) can usually be replaced exactly by the inter-event time cumulative distribution function (CDF). The disadvantage is that much of the simplicity of equation (3.86) is lost, as well as the ''memory-lessness'' property of the exponential distribution.

An ideal generalisation of the classic PSHA would probably be that in which locally "flat" approximations can keep the desirable properties of the homogeneous Poisson process. Because the deviation is often not far from the idealised case of the homogeneous Poisson process, such approximations might not be too farfetched. Two relevant studies in this regard are the Weibull distribution used by Cornell and Winterstein (1988) and the introduction of the compound/Bayesian distribution, mentioned by Cornell and Winterstein (1988) and used by Kijko and Graham (1999).

Cornell and Winterstein (1988) introduced a Weibull model for both inter-event time and magnitude. This is an elegant generalisation of the exponential distribution, including the exponential distribution as a special case. The Weibull model allows more weight in the tail, enabling "characteristic" inter-event times and magnitudes. Further, the Weibull model allows the assessment of the traditional hazard parameters.

3.3.6. Weibull model of Cornell and Winterstein (1988).

Cornel and Winterstein (1988) generalised the Poisson process by extending the inter-event time distribution to a Weibull model that includes the exponential distribution of the Poisson process as a special case

$$1 - F_\tau(t) = P[\tau > t] \approx \exp\left[-\left(\frac{V_\tau!\, t}{E[\tau]}\right)^{\frac{1}{V_\tau}}\right], \tag{3.87}$$

where $V_\tau = (var[T])^{\frac{1}{2}}/E[\tau]$ is the coefficient of variation, also called aperiodicity (Zhuang *et al.*, 2012.) The value of V_τ is commonly between 1 and 0, for which the approximation (3.87) holds for deterministic, periodic behaviour $V_\tau = 0$ (i.e. no random scatter), and for the Poisson model $V_\tau = 1$ (standard deviation

of inter-event time is equal to its expected value). The hazard function associated with approximation (3.87) takes the form

$$h(t_0) = \frac{f_\tau(t_0)}{1 - F_\tau(t)} \approx \frac{(V_\tau - 1)!}{E[\tau]} \left(\frac{V_\tau!\, t_0}{E[\tau]} \right)^{(1-V_\tau)/V_\tau}. \tag{3.88}$$

A similar model is used for magnitude so that the extended Gutenberg–Richter FMSL is

$$\ln[N(m)] \approx \alpha - [\beta(m - m_0)]^{1/V_M}, \;\; m \geq m_0 \tag{3.89}$$

with $V_M = (var[M])^{\frac{1}{2}}/E[M]$.

Nemati (2015) investigated the varying rates and migration of seismicity in Northern Iran, including the Alborz seismotectonic province, which is the area south of the Caspian sea, the Kopeh Dagh (also referred as the Kopet Dagh) seismotectonic province, which lies directly east of the Kopeh Dagh province, and the Azerbaijan seismotectonic province just west of the Alborz province. Nemati (2015) hypothesises that the local migration of seismicity in the Alborz seismotectonic province, evident in aftershocks, could be related to the regional migration of seismicity. The study by Nemati (2015) indicates that periods of higher and lower seismicity in this area could be distinguished clearly. In Azerbaijan, it appears that the east of the region was seismically active in the period 1801–1895, but the west was active during 1930–1934, and recent seismicity is mostly uniform in the north. A cumulative plot of events clearly shows that different, discrete periods of varying seismicity occur. In the Kopeh Dagh area, large seismic events appear to have migrated from east to west over the last two centuries. A cumulative plot of this clearly shows discrete periods of varying seismicity. According to Nemati (2015), there is a clear westward migration of seismicity in the north of Iran, which is logical in the context of the tectonic dynamics of the area.

Goldfinger *et al.* (2013) propose that exceptionally long cycles can occur in fault activity, which can only be investigated by palaeoseismic studies. However, unless these cycles include abrupt changes, they have no effect on seismic hazard assessments and the b-value, as the period over which the cycles occur render them nearly stationary. However, acceptance of the concept of very long cycles of seismicity does contribute to the assessment of the maximum possible earthquake m_{max}, as is discussed in the next chapter.

There appears to be a global increase in seismicity. However, Omnetti (2013) dismisses this as an artefact of a random time series, whereas Bufe and Perkins (2005) believe it is an indication of the Earth "behaving as a coherent tectonic system".

Awoyemi *et al.* (2017) studied the variation of the b-value with space and time for the African and parts of the Eurasian plates. In their study, as regards temporal variations, a certain number of events were

selected for a moving window (Figure 8, for example) and, for spatial variations, grid points were calculated for a constant radius.

3.4. Solution to Simultaneous b-value and Seismicity Variations

Kijko *et al.* (2016) present an elegant solution to variations in the b-value and seismicity, both spatially and temporally. This is the compound distribution, in which a parameter in the distribution of a random variable is also a random variable (DeGroot, 2005). Benjamin (1968) and Campbell (1982, 1983) are probably the first to implement compound distributions in seismic hazard analysis. They employ the gamma distribution for both the b-value and the mean seismic activity λ. The PDF of the gamma distribution has a form

$$f_X(x) = x^{(q-1)} \frac{p^q}{\Gamma(q)} \exp(-px), \quad x, p, q > 0, \tag{3.90}$$

in which $\Gamma(q)$ is the gamma function

$$\Gamma(q) = \int_0^\infty y^{q-1} \exp(-y)\, dy, \tag{3.91}$$

where the parameters p and q are related to the mean μ_x and variance σ_x^2 by the relations

$$\mu_x = \frac{q}{p}, \tag{3.92}$$

$$\sigma_x^2 = \frac{q}{p^2}. \tag{3.93}$$

Mixing the GR distribution with the gamma distribution, where $p_\beta = \frac{\mu_\beta}{\sigma_\beta^2}$ and $q_\beta = \left(\frac{\mu_\beta}{\sigma_\beta}\right)^2$, the density distribution is obtained (denoting μ_β by β)

$$f_M\left(m \middle| p_\beta, q_\beta\right) = C_\beta \beta \left(\frac{p_\beta}{p_\beta + m_{max} - m_c}\right)^{q_\beta + 1} \tag{3.94}$$

where

$$C_\beta = \left[1 - \left(\frac{p_\beta}{p_\beta + m_{max} - m_c}\right)^{q_\beta}\right]^{-1} \tag{3.95}$$

If, in the same way, the variation of λ with a gamma distribution were modelled, with $p_\lambda = \frac{\mu_\lambda}{\sigma_\lambda^2}$ and $q_\lambda = \left(\frac{\mu_\lambda}{\sigma_\lambda}\right)^2$, the compound Gamma–Poisson distribution is obtained (Benjamin, 1968)

$$P_n(\lambda, t, p_\lambda, q_\lambda) = \frac{\Gamma(n + q_\lambda)}{n!\,\Gamma(q_\lambda)}\left(\frac{p_\lambda}{1 + p_\lambda}\right)^{q_\lambda}\left(\frac{t}{t + p_\lambda}\right)^n. \tag{3.96}$$

The gamma distribution is dynamic and versatile, i.e. it is ideal to model the distribution of β and λ.

3.5. Comparative Study

The current author conducted a comparison on a number of selected estimators [(3.32), (3.46), (3.49), (3.58), and (3.60)] by employing Monte Carlo simulations. Estimator (3.32) is the Weichert (1986) maximum likelihood estimator based on magnitude intervals (classes). Estimator (3.46) is the Kijko and Sellevoll (1989) likelihood equation, maximised to estimate the parameters. Estimator (3.49) is the extended Aki–Utsu estimator (Kijko and Smit, 2012). This is a pure maximum-likelihood-based solution, of which the basis is the Aki–Utsu estimator [equation (1)], extended to accommodate incomplete catalogues. Estimators (3.58) and (3.60) are based on and motivated by the extended Aki–Utsu estimator. The number of earthquakes is extrapolated to a lower reference magnitude, on which the estimates are subsequently based. The current author is critical of the estimator (3.60); however, his reservations are based merely on intuition.

The particulars of the inputs for the Monte Carlo simulations and the results are reported in Tables 1 (a) and (b).

Synthetic catalogue:

β =2.303, (b=1), m_{max} = 7.0, m_0 = 3.0; λ_0 = 100;

Complete part 1	50 years duration. m_{c1} = 4.2
Complete part 2	50 years duration. m_{c2} = 4.0
Complete part 3	50 years duration. m_{c3} = 3.6
Complete part 4	50 years duration. m_{c4} = 3.0

Table 1 (a). Results of β assessments by different procedures. The assessments are based on 10 000 simulated catalogues.

	Kijko–Smit, 2012	**Kijko, 2017 (1)**	**Kijko, 2017 (2)**	**Kijko–Sellevoll, 1989**	**Weichert, 1986**
Mean	2.1903	2.0947	2.0866	2.1821	2.0046
Standard deviation	0.0265	0.0395	0.0432	0.0269	0.0337
MSE (mean square error)	0.0133	0.0448	0.0485	0.0153	0.0899
Bias	-0.1125	-0.2080	-0.2160	-0.1207	-0.2980
99% confidence interval	[2.1289, 2.2538] (span: 0.1249)	[2.0060, 2.1917] (span: 0.1857)	[1.9884, 2.1900] (span: 0.2061)	[2.1196, 2.2465] (span: 0.1269)	[1.9271, 2.0857] (span: 0.1586)
95% confidence interval	[2.1460, 2.2335] (span: 0.0875)	[2.0314, 2.1614] (span: 0.1300)	[2.0170, 2.1585] (span: 0.1415)	[2.1370, 2.2260] (span: 0.0890)	[1.9503, 2.0609] (span:0.1106)

Synthetic catalogue:

β = 2.303, (b=1), m_{max} = 9.0, m_0 = 3.0; λ_0 = 100;

Complete part 1	50 years duration. m_{c1} = 4.2
Complete part 2	50 years duration. m_{c2} = 4.0
Complete part 3	50 years duration. m_{c3} = 3.6
Complete part 4	50 years duration. m_{c4} = 3.0

Table 1 (b). Results of β assessments by different procedures. The assessments are based on 10 000 simulated catalogues.

	Kijko–Smit, 2012	**Kijko, 2017 (1)**	**Kijko, 2017 (2)**	**Kijko–Sellevoll, 1989**	**Weichert, 1986**
Mean	2.1903	2.0947	2.0866	2.1821	2.0046
Standard deviation	0.0265	0.0395	0.0432	0.0269	0.0337
MSE	0.0133	0.0448	0.0485	0.0153	0.0899
Bias	-0.1125	-0.2080	-0.2160	-0.1207	-0.2980
99% confidence interval	[2.1289, 2.2538] (span: 0.1249)	[2.0060, 2.1917] (span: 0.1857)	[1.9884, 2.1900] (span: 0.2016)	[2.1196, 2.2465] (span: 1.269)	[1.9271, 2.0857] (span: 0.1586)
95% confidence interval	[2.1460, 2.2335] (span: 0.0875)	[2.0314, 2.1614] (span: 0.1300)	[2.0170, 2.1585] (span: 0.1415)	[2.1370, 2.2260] (span: 0.0890)	[1.9503, 2.0609] (span: 0.1106)

As the results indicate, the performance of all the estimators is quite similar, with the performance of the Kijko–Smit (2012) procedure and the Kijko–Sellevoll procedure appearing slightly superior. However, more variation in parameters m_{ci}, λ_0, and m_{max} are required to arrive at any firm conclusions. Some estimators perform better in one part of (m_{ci}, λ_0, m_{max}) space, whereas others can perform better in other parts of the space. A study on their performance with varying b-values should also be considered to enhance the results, as, in reality, the b-value does vary and is uncertain. The findings of this brief study do not facilitate any clear conclusion about the superiority of any of the methods investigated, or their applicability for use in PSHA.

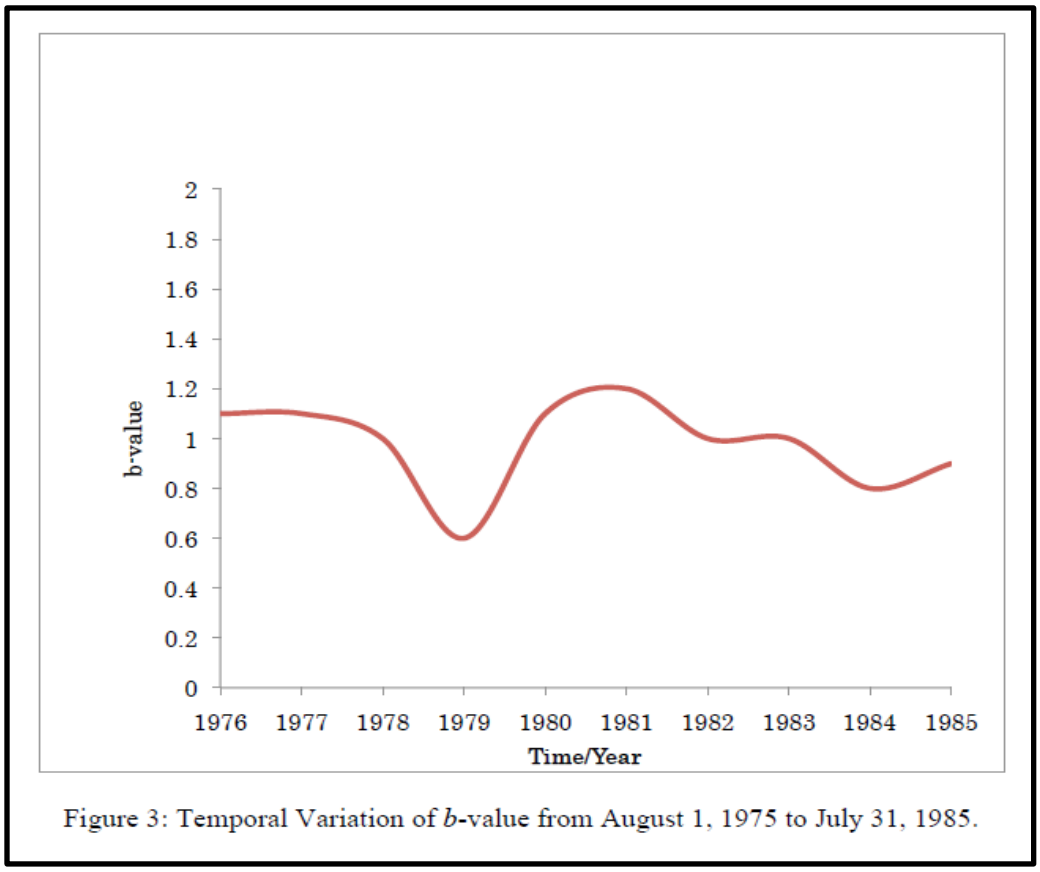


Figure 8. Variation of b-value with space and time for the African and parts of the Eurasian plates (from Awoyemi *et al.,* 2017).

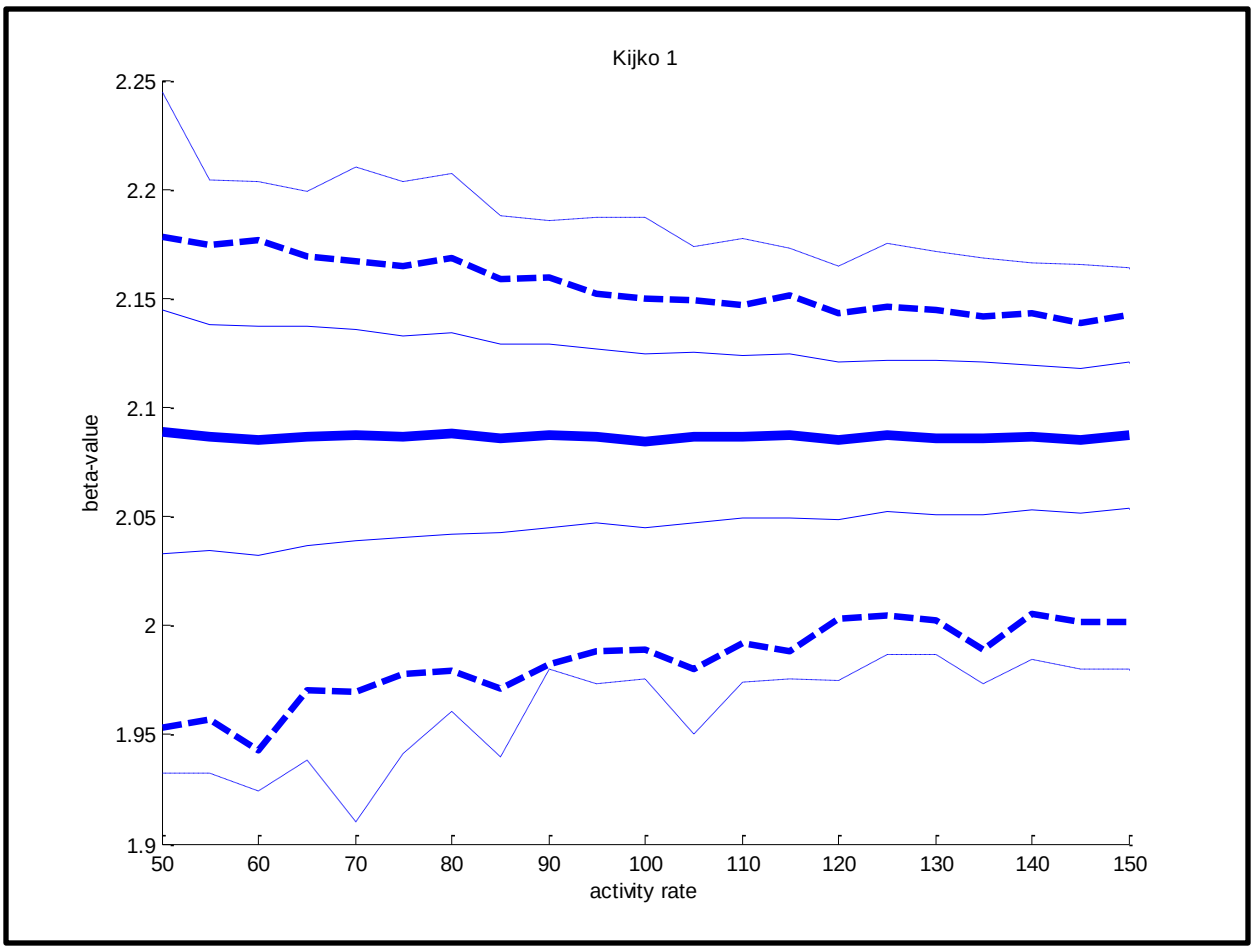


Figure 9(a). Performance of the first estimator of Kijko (2016). The synthetic catalogue consists of four sub-catalogues, each of 50 years duration, and at levels of completeness 4.2, 4.0, 3.6, and 3.0. The thick solid line indicates the value of the estimator, thin solid lines indicate 1 standard deviation, and the dashed lines indicate the 95% and 99% confidence limits.

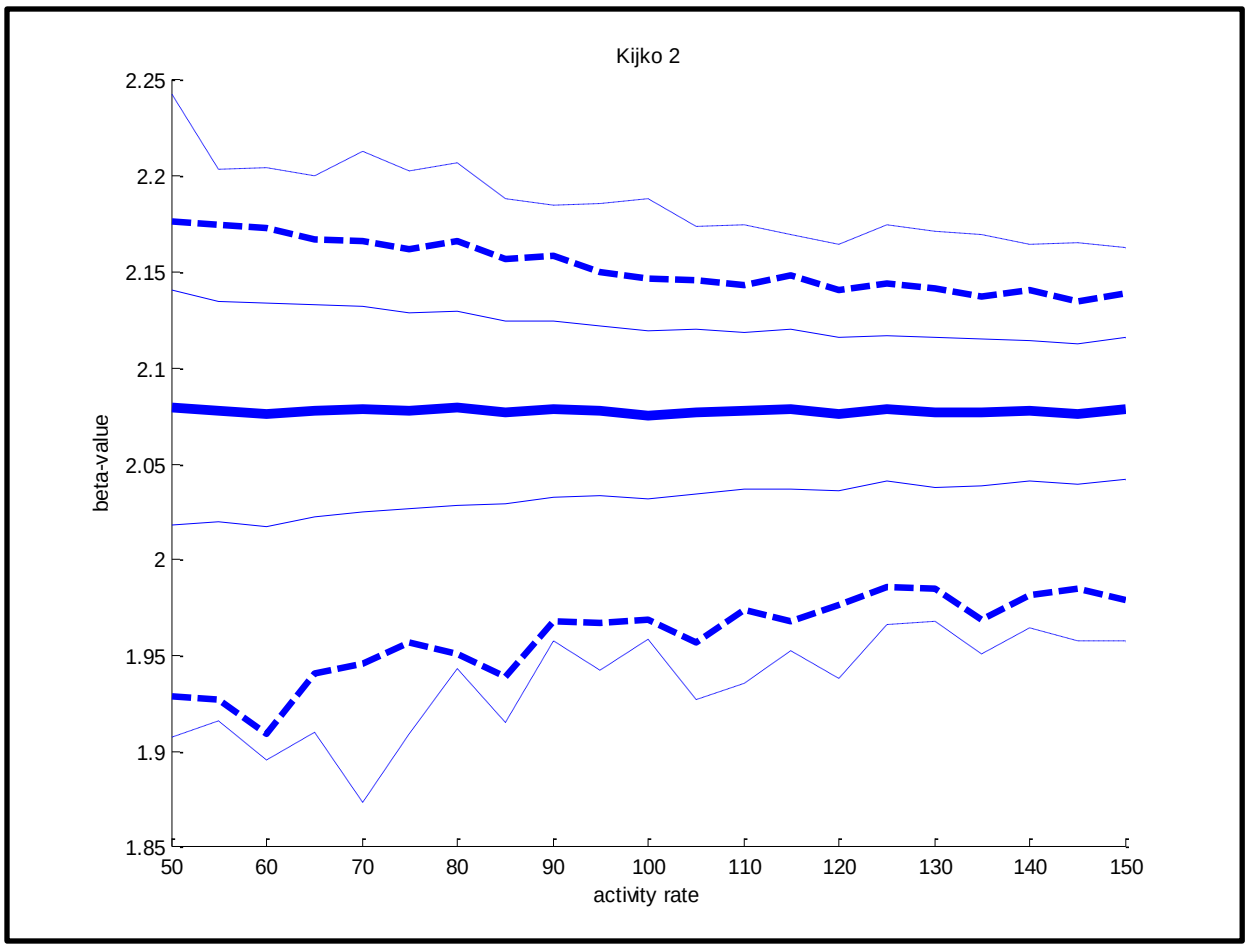


Figure 9(b). Performance of the Kijko (2016) second estimator. The synthetic catalogue consists of four sub-catalogues, each of duration of 50 years at levels of completeness of 4.2, 4.0, 3.6, and 3.0. The thick solid line indicates the value of the estimator, thin solid lines indicate 1 standard deviation, and the dashed lines indicate the 95% and 99% confidence limits.

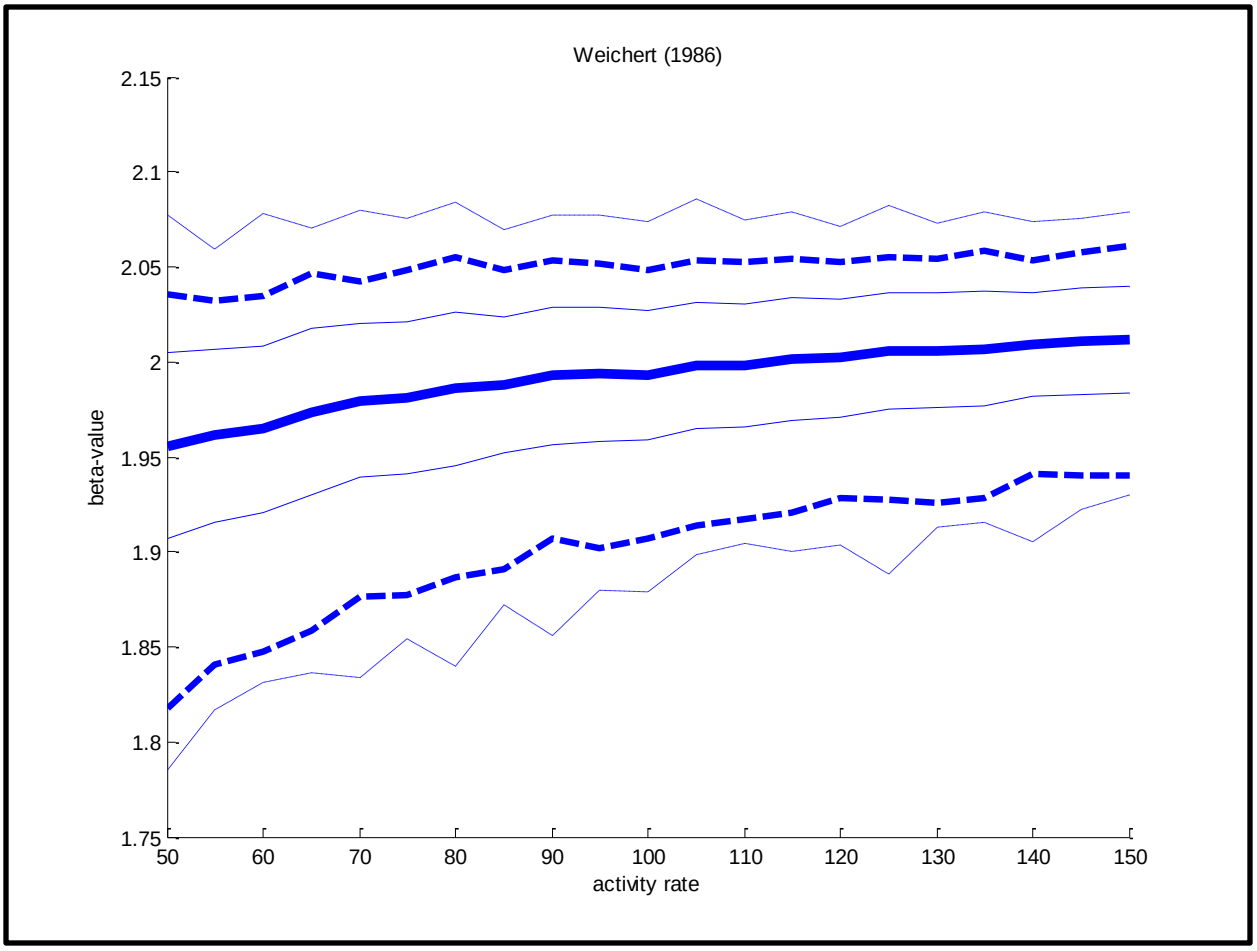


Figure 9(c). Performance of the Weichert (1986) estimator. The synthetic catalogue consists of four sub-catalogues, each of 50 years duration at levels of completeness of 4.2, 4.0, 3.6, and 3.0. The thick solid line indicates the value of the estimator, thin solid lines indicate 1 standard deviation, and the dashed lines indicate the 95% and 99% confidence limits.

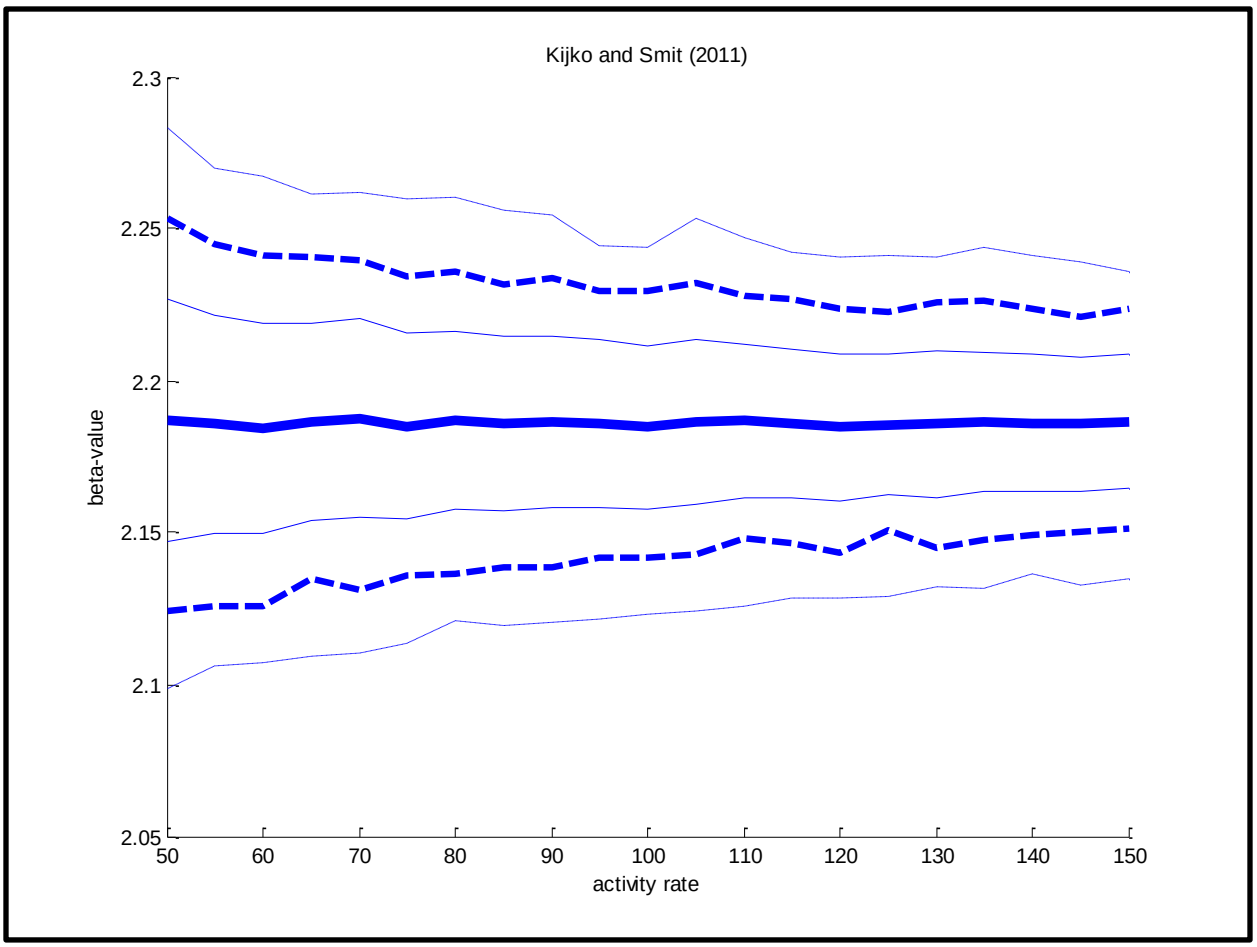


Figure 9(d). Performance of the extended Aki–Utsu estimator (Kijko and Smit, 2016). The synthetic catalogue consists of four sub-catalogues, each of 50 years duration, at levels of completeness of 4.2, 4.0, 3.6, and 3.0. The thick solid line indicates the value of the estimator, thin solid lines indicate 1 standard deviation, and the dashed lines indicate the 95% and 99% confidence limits.

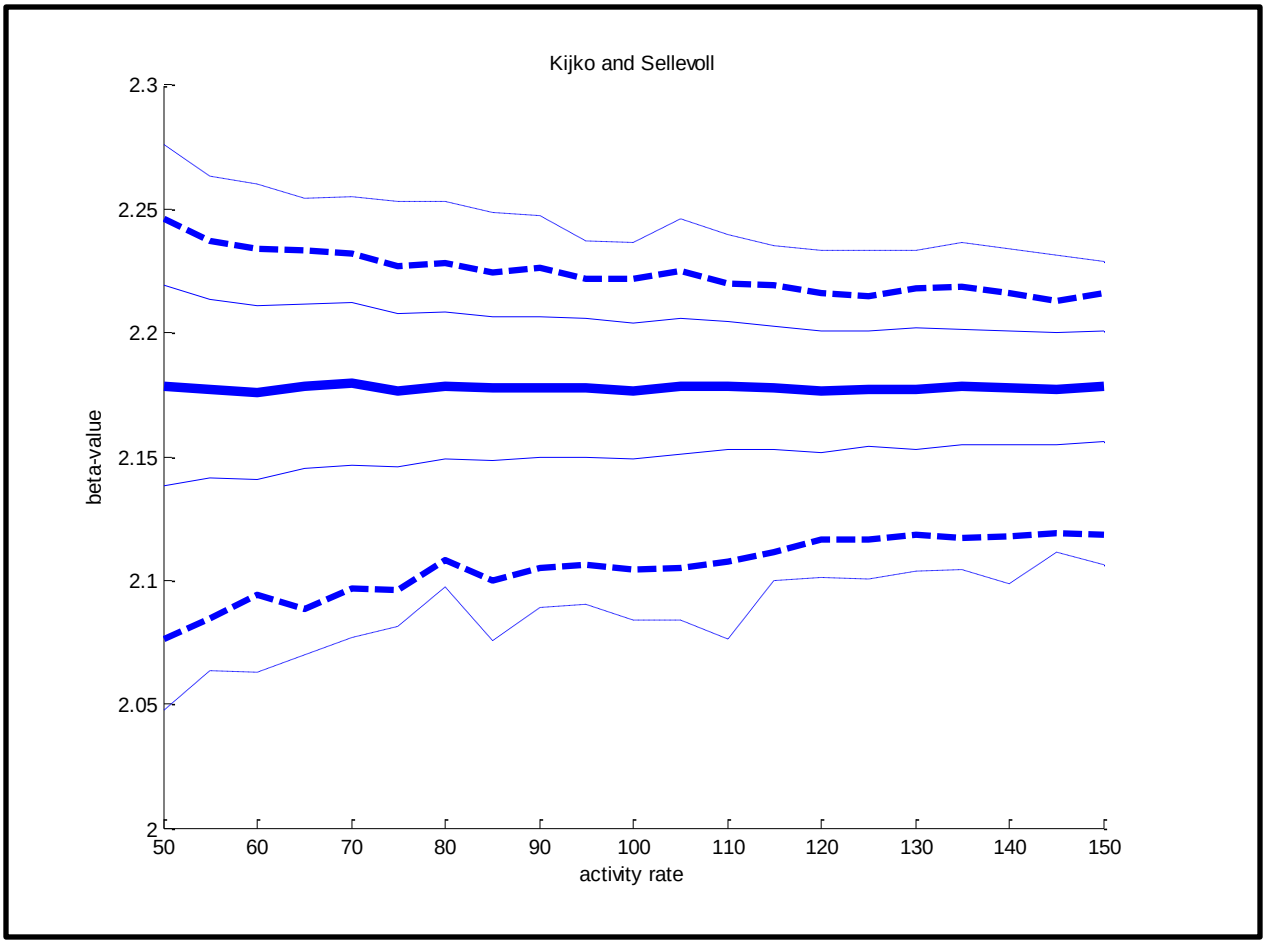


Figure 9(e). Performance of the Kijko–Sellevoll (Kijko and Sellevoll, 1989) estimator. The synthetic catalogue consists of four sub-catalogues, each of 50 years duration, at levels of completeness of 4.2, 4.0, 3.6, and 3.0. The thick solid line indicates the value of the estimator, thin solid lines indicate 1 standard deviation, and the dashed lines indicate the 95% and 99% confidence limits.

However, the results of this study do appear to indicate that the Weichert (1980) estimator is biased at lower activity rates (relating to fewer observations). As indicated by the results (Tables 1[a] and [b]), the other estimators to some extent outperform the Weichert (1980) estimator. Similar deductions can be made from the simulations, of which the results are shown in Figures 9(a–e). The performance of the other estimators is extremely similar. Using estimator (3.58) is not recommended, because of its complexity and the need for iteration. Either estimator (3.46) or estimator (3.60) can be used, depending on which is more applicable.

4. Area-Characteristic, Maximum Possible Earthquake Magnitude m_{max}

"*The problem of evaluating the maximum earthquake possible for a particular seismic region has always occupied one of the central places in seismology. This is attributable, on the one hand, to the fact that the maximum possible magnitude M_{max} is one of the parameters of the earthquake recurrence law [i.e. the Gutenberg–Richter law], and on the other hand, to the fact that the possible destructive consequences of earthquakes in a particular region, as well as the seismic resistant construction norms determined by them, are quite dependent on M_{max}.*" (Pisarenko, 1991).

4.1. Parametric Estimators

The first notable estimator for m_{max} was probably that derived by Pisarenko (1991) and Pisarenko *et al.* (1996). The latter is a statistically optimal estimate of m_{max} , i.e. it is unbiased and has minimum variance in an ideal case, where the magnitude is defined well and no uncertainty exists in the parameters. The estimator derives from the Rao–Blackwell theorem, which states that, given an unbiased estimator for a parameter θ, conditional to a complete sufficient statistic, it gives the minimum variance unbiased estimator $\bar{\hat{\theta}}$ of θ. Now, m_{max}^{obs} is a completely sufficient statistic. An unbiased estimator for m_{max} is (Pisarenko, 1996)

$$\hat{m}_{max} = m_{max}^{obs} + \frac{F(m_{max}^{obs})}{nf(m_{max}^{obs})}, \tag{4.1}$$

where *n* is a number of seismic events. By using (4.1.) and the sufficient statistic m_{max}^{obs} , the minimum variance unbiased estimator is derived, given by

$$\bar{\hat{m}}_{max} = m_{max}^{obs} + \frac{1}{nf(m_{max}^{obs}|m_{max}^{obs})}. \tag{4.2}$$

The best unbiased estimator of the variance of $\bar{\hat{m}}_{max}$ is

$$\hat{var}\left[\bar{\hat{m}}_{max}\right] = \frac{1}{n^2 f(m_{max}^{obs}|m_{max}^{obs})^2}. \tag{4.3}$$

Pisarenko *et al.* (1996) present estimations of m_{max} when the apparent magnitude is considered when parameters are uncertain, and when the LoC changes with time (i.e. incomplete catalogues).

Pisarenko (1991) proposes a method to determine the $1-\alpha$ upper confidence limit of m_{max}. In the case of the GR relation, it is given by

$$m_{max,1-\alpha} = m_{max}^{obs} - \frac{1}{\beta}\ln\left[\frac{1-(1-\alpha^{\frac{1}{n}})\exp[\beta(m_{max}^{obs}-m_c)]}{\alpha^{\frac{1}{n}}}\right], \tag{4.4}$$

or, in a more general instance of a frequency-magnitude distribution $F(m)$

$$\Pr\left[m_{max} < F^{-1}\left(m_{max}^{obs};\alpha^{\frac{1}{n}}\right)\right] = 1-\alpha, \tag{4.5}$$

which leads to

$$\Pr[m_{max} < z] = 1-[F(m_{max}^{obs};z)]^n. \tag{4.6}$$

This is known as the fiducial distribution function (Pisarenko, 1991; Kijko and Singh, 2011).

"*The fiducial distribution function $Pr[m_{max} < z]$ has the following sense: for any true value of the m_{max} parameter and for the sample statistics $[m_{max}^{obs}]$ corresponding to it the inequality $m_{max} < z$ is satisfied with the probability*." (Pisarenko, 2001).

$$\Pr[m_{max} < z] = 1-[F(m_{max}^{obs};z)]^n, \qquad m_{max}^{obs} \le z < \infty, \tag{4.7}$$

as $z \to \infty$, $\Pr[m_{max} < z]$ tends to a value less than unity. According to Pisarenko (2001), "it can be assumed formally that in this case the upper limit for m_{max} is equal to ∞." Pisarenko elaborates on the instances where $\Pr[m_{max} < z]$ can be considered to have interpretable meaning and, as regards the other instances, corresponds to instances when inadequate data are available to assess m_{max}. "To be sure, the upper limit ∞ provides no information on the maximum possible magnitude, but formally it is convenient to assume that with some positive probability $[\alpha]$ the upper limit assumes the value ∞." (Pisarenko, 1991). Although this interpretation appears convenient, the meaning of this statistic is controversial. Kijko and Graham (1998) derived the same estimator using a different approach. Given a uniform distribution, with observations $(x_1, x_2, x_3, \dots x_n)$, the unbiased estimator of the upper limit of the distribution, x_{max} has a form (Gibowicz and Kijko, 1994)

$$\hat{x}_{max} = \frac{n+1}{n}x_{max}^{obs}. \tag{4.8}$$

The relation (4.8) is based on the fact that any cumulative distribution $F(x)$ follows a uniform distribution in the interval (0,1), which results in the following relation

$$1 = \frac{n+1}{n} F(x_{\max}^{obs} | x_{max}). \tag{4.9}$$

By assuming that $F(x_{\max}^{obs} | x_{max})$ is the bounded GR relation, Gibowicz and Kijko (1994) solve equation (4.9) for m_{max} to obtain the following estimator for the maximum magnitude

$$\widehat{m}_{max} = \frac{1}{\beta} ln \left\{ e^{-\beta m_c} - \left[e^{-\beta m_c} - e^{-\beta m_{max}^{obs}} \right] \frac{n+1}{n} \right\}. \tag{4.10}$$

Kijko and Graham (1998) obtain a version of the Pisarenko estimator (4.2) from relation (4.9) by approximating a solution to m_{max} through a Taylor expansion, with the further approximation that $F(m_{\max}^{obs} | m_{max}) \cong 1$

$$\widehat{m}_{max} = m_{max}^{obs} + \frac{1}{n f(m_{max}; m_{max})}. \tag{4.11}$$

Because m_{max} appears on both sides of equation (4.11), it is solved through iteration with the first approximation of m_{max} as m_{max}^{obs}, which renders the first iteration of (4.10) the same as an estimator (4.2). Kijko (2004) states that estimator (4.11) was probably first derived by Tate (1959). For the GR relation (4.11), the first iteration takes the form

$$\widehat{m}_{max} = m_{max}^{obs} + \frac{1 - \exp[-\beta(m_{max}^{obs} - m_c)]}{n\beta \exp[-\beta(m_{max}^{obs} - m_c)]} \tag{4.12}$$

Iteration of equation (4.12) is not guaranteed to converge, but it is more likely to do so with a larger number of observations. Exact conditions for conversion of this iteration procedure are derived by Vermeulen and Kijko (2017). The variance that Kijko and Graham (1998) give for estimator (4.12) is

$$Var[\widehat{m}_{max}] = var[m_{max}^{obs}] + \frac{n+1}{n^3} \left[\frac{1 - \exp[-\beta(m_{max}^{obs} - m_c)}{\beta \exp[-\beta(m_{max}^{obs} - m_c)} \right]^2. \tag{4.13}$$

The variance given by Pisarenko *et al.* (1996) applied to the GR relation would be

$$Var[\hat{m}_{max}] = \frac{1}{n^2}\left[\frac{1-\exp[-\beta(m_{max}^{obs}-m_c)}{\beta\exp[-\beta(m_{max}^{obs}-m_c)}\right]^2. \quad (4.14)$$

Equations (4.13) and (4.14) are almost equivalent, although equation (4.14) does not take the standard error of m_{max}^{obs} into consideration.

Kijko and Graham (1998) present another estimator for m_{max} , derived from that of Cooke (1979). From this derives the estimator

$$\hat{m}_{max} = m_{max}^{obs} + \int_{m_c}^{m_{max}} [F(m)]^n dm, \quad (4.15)$$

where n denotes the number of events. Using the empirical distribution and an asymptotic approximation for a large number n of observations, Cooke (1979) derives the following estimator from (4.15)

$$\hat{m}_{max} = 2m_{max}^{obs} - (1-e^{-1})\sum_{i=0}^{n-1} e^{-i} m_{rank:n-i} \quad (4.16)$$

However, Cooke (1979) does not discuss the variance of the estimators.

Kijko and Sellevoll (1989) derive the following approximation for the integral term in (4.15) if a GR relation were assumed

$$\Delta = \frac{E_1(n_2) - E_1(n_1)}{\beta\exp(-n_2)} + m_c\exp(-n), \quad (4.17)$$

where $n_1 = n/\{1-\exp[-\beta(m_{max}-m_c)]\}$, $n_2 = n_1\exp[-\beta(m_{max}-m_c)]$, and $E_1(z) = \int_z^\infty \exp(\zeta)/\zeta\, d\zeta$. The approximate variance of the estimator is

$$Var[\hat{m}_{max}] = Var[m_{max}^{obs}] + \left[\frac{E_1(n_2) - E_1(n_1)}{\beta\exp(-n_2)} + m_c\exp(-n)\right]^2. \quad (4.18)$$

Kijko and Singh (2011), in addition, provide an exact solution for equation (4.15) in the instance of a GR distribution

$$\widehat{m}_{max} = m_{max}^{obs} + \frac{(m_{max} - m_c) + \frac{1}{\beta}\sum_{i=1}^{n} \frac{(-1)^n}{i}\binom{n}{i}(1 - \exp[-i\beta(m_{max} - m_c)])}{(1 - \exp[-\beta(m_{max} - m_c])^n}. \quad (4.19)$$

The approximate variance of this estimator is

$$\widehat{m}_{max} = var[m_{max}^{obs}] + \left[\frac{(m_{max} - m_c) + \frac{1}{\beta}\sum_{i=1}^{n} \frac{(-1)^n}{i}\binom{n}{i}(1 - \exp[-i\beta(m_{max} - m_c)])}{(1 - \exp[-\beta(m_{max} - m_c])^n}\right]^2. \quad (4.20)$$

Based on extensive numerical tests (Kijko and Smit, 2017), it appears that the exact solution is not superior to its approximate counterpart (4.18). In the current author's experience, one problem is the calculation of factorials that derive from binomial expansion. This quickly produces extremely large numbers that make scaling and exact calculation a most challenging task. In addition, the approximate solution is much simpler and more elegant.

Acknowledging the fact that β and seismicity could vary in time, Kijko and Graham (1998) provide a mixed (known also as a compound or Bayesian) distribution (Kijko and Graham, 1998), which can be applied to estimators (4.1), (4.11), and (4.15). This involves the assumption that β varies according to a gamma distribution, with parameters $p = \beta/(\sigma_\beta)^2$ and $q = \left(\beta/\sigma_\beta\right)^2$. The distribution of β is treated as a prior distribution and the resulting posterior is used as the frequency-magnitude relation. This relation is given by (Campbell, 1982)

$$f_M(m) = \begin{cases} 0 & for\ m < m_c \\ \beta C_\beta \left(\frac{p}{p + m - m_c}\right)^{q+1} & for\ m_c \le m \le m_{max}, \\ 0 & for\ m > m_{max} \end{cases} \quad (4.21)$$

and

$$F_M(m) = \begin{cases} 0 & for\ m < m_c \\ C_\beta \left[1 - \left(\frac{p}{p + m - m_c}\right)^q\right] & for\ m_c \le m \le m_{max}\,. \\ 1 & for\ m > m_{max} \end{cases} \quad (4.22)$$

Raschke (2012) develops an estimator for m_{max} from the fiducial distribution in equation (4.6), used by Pisarenko (1991), and discussed by Kijko (2004) and Kijko and Singh (2011). Raschke eliminates the problem noted by Kijko and Singh (2011) that $P[m_{max} < \infty] < 1$ by its normalisation renders it a point

estimate. Specifically, the normalised distribution for $\widehat{m}_{max}$ for n observations and a given m_{max}^{obs} is given, on its support, by

$$F_{m_{max}}(\widehat{m}_{max}) = \frac{1 - \left(\frac{F^*(m_{max}^{obs})}{F^*(\widehat{m}_{max})}\right)^n}{1 - F^*(m_{max}^{obs})^n}, \qquad m_{max} \geq \widehat{m}_{max}^{obs}, \tag{4.23}$$

where $F^*(\cdot)$ is the original untruncated cumulative distribution. If the exponential distribution were substituted by the original GR CDF, the resulting CDF takes the form

$$F_{m_{max}}(\widehat{m}_{max}) = \frac{1 - \left(\frac{1 - \exp[-\beta(m_{max}^{obs} - m_c)]}{1 - \exp[-\beta(\widehat{m}_{max} - m_c)]}\right)^n}{1 - \left(1 - \frac{1 - \exp(m_{max}^{obs})}{1 - \exp(m_c)}\right)^{\mathrm{n}}}, \qquad m_{max} \geq \widehat{m}_{max}^{obs}, \tag{4.24}$$

and the PDF on its support is

$$f_{m_{max}}(\widehat{m}_{max}) = \frac{\frac{\left\{1 - \exp\left[-\beta\left(m_{max}^{obs} - m_c\right)\right]\right\}^n n\beta \exp[-\beta(\widehat{m}_{max} - m_c)]}{\{1 - \exp[-\beta(\widehat{m}_{max} - m_c)]\}^{\mathrm{n}+1}\left(1 - \left\{1 - \exp\left[-\beta\left(m_{max}^{obs} - m_c\right)\right]\right\}^n\right)}}{F_{m_{max}}\left(m_{max}^{obs} - m_c\right)}. \tag{4.25}$$

The support of the two distributions are $[m_{max}^{obs}, \infty)$. Raschke (2012) proposes the expectation of $f_{m_{max}}(\widehat{m}_{max})$ as a point estimate for m_{max}. Obviously, the median is also a possible estimator, but this has not been investigated. The expected value of $f_{m_{max}}(\widehat{m}_{max})$ has been tested thoroughly by Raschke (2012) and the results obtained indicate superior performance. This expected value is asymptotically mean and median unbiased, as the density becomes a delta function for $n \to \infty$.

Vermeulen and Kijko (2017) investigated the estimation of m_{max} using the Method of Moments (Vermeulen and Kijko, 2017), as described by Dixit and Nasiri (2008). Moments of the sample are used to calculate the truncation point of an exponential distribution. However, according to the literature, using this method appears to have become uncommon, and the maximum likelihood method is often preferred. Nevertheless, this method is still worth investigating. Using this method requires the shift transformation $x_i = m_i - m_c$. As shown by Vermeulen and Kijko (2017), the estimate of the truncation point of an exponential distribution given by Dixit and Nasiri (2008) takes the form

$$x_{max} = \frac{-3WY + nZ + D^{\frac{1}{2}}}{2(nY - 2W^2)}, \tag{4.26}$$

where $D = n^2Z^2 - 15W^2Y^2 - 14nWYZ + 12nY^3 + 16W^3Z$, $W = nr_1$, $Y = nr_2$, $Z = nr_3$, $r_1 = \sum_i x_i$, $r_2 = \frac{1}{n}\sum_i x_i^2$, and $r_3 = \frac{1}{n}\sum_i x_i^3$. That is, r_i is the i^{th} moment.

The following analytic expression gives the approximate variance of the moment estimator

$$var[\hat{x}_{max}(\boldsymbol{r})] = [\nabla\hat{x}_{max}(\boldsymbol{r})]^T * \boldsymbol{D}(\boldsymbol{r}) * [\nabla\hat{x}_{max}(\boldsymbol{r})]. \tag{4.27}$$

If the correlation between moments were ignored, $\boldsymbol{D}(\boldsymbol{r})$ takes the form

$$\boldsymbol{D}(\boldsymbol{r}) = \begin{bmatrix} \frac{r_2 - r_1^2}{N} & 0 & 0 \\ 0 & \frac{r_2 - r_1^2}{N} & 0 \\ 0 & 0 & \frac{r_2 - r_1^2}{N} \end{bmatrix}, \tag{4.28}$$

$$[\nabla\hat{x}_{max}] = \begin{bmatrix} -4r_1 & 3r_2 & 2r_1 \\ 1 & r_1 & -6r_2 \\ 0 & -1 & 2r_1 \end{bmatrix} * \begin{bmatrix} \frac{A_2A_3(A_2^2 - 4A_1A_3)^{-\frac{1}{2}} + (A_2^2 - 4A_1A_3)^{\frac{1}{2}}}{2A_2^2} \\ \hline \frac{-1 + A_2(A_2 - 4A_1A_3)^{\frac{1}{2}}}{2A_1} \\ \hline -\frac{2A_1(A_2^2 - 4A_1A_3)}{2A_1} \end{bmatrix}, \tag{4.29}$$

where $A_1 = r_2 - 2r_1^2$, $A_2 = 3r_1r_2 - r_3$, and $A_3 = 2r_1r_2 - r_3$. Investigation using numerical experiments has shown that this approximation is generally too large by up to an order of magnitude, indicating room for improvement. Practically, however, it is feasible to resort to jackknife or bootstrap procedures.

Furthermore, it is a sensible option to use bootstrap or jackknife resampling to determine approximate variance and as a check for the approximate variance in the analytic form given above. Vermeulen and Kijko (2017) find that the moment estimator generally provides estimates of m_{max} lower than the maximum observed magnitude and, at times, extremely unrealistically large values. From bootstrap resampling, it appears that the distribution of the estimator is heavy tailed, or nearly so, which explains the behaviour of the estimator. Furthermore, using a jackknife approach for comparison, the use of estimator (4.26), when ignoring the correlation between moments, yields incorrect results, although of the same order as the correct value.

4.2. Non-parametric Estimators

Robson and Whitlock (1964) employ the theory of bias for estimation, originally developed by Quenouille (1956). In particular, an estimate of the bias of the m_{max}^{obs} is subtracted as an estimate of m_{max} according to the method by Quenouille (1956) to obtain an improved estimator of m_{max}. Robson and Whitlock (1964) specifically note that the estimator

$$\hat{m}_{max} = 2m_{max}^{obs} - m_{n-1}, \tag{4.30}$$

has variance similar to $\hat{m}_{max} = m_{max}^{obs}$, but has the advantage that it is mean unbiased to the order n^{-3} and asymptotically median unbiased. In equation (4.30), m_{n-1} denotes the second largest observed magnitude. The approximate variance of this estimator is (Kijko and Singh, 2011)

$$Var(\hat{m}_{\max}) = 5\sigma_M^2 + (m_{max}^{obs} - m_{n-1})^2, \tag{4.31}$$

where a general estimator is derived

$$\hat{m}_{max} = \sum_{j=0}^{k} (-1)^j \binom{k+1}{j+1} m_{n-j}, \tag{4.32}$$

where the maximum value of k = n-1. However, it was shown (Robson and Whitlock, 1964) that the efficiency of the estimator (4.32) actually decreases with increasing k. Furthermore, the estimator (4.32) is not guaranteed to be greater than or equal to m_{max}^{obs}. The mean squared error of estimator (4.32) is given by

$$E(\hat{m}_{max} - m_{max})^2 = \binom{2k}{k} n^{-2} \left[\frac{F(m_{max})}{F'(m_{max})}\right]^2 + O(n^{-3}), \tag{4.33}$$

where $F(m)$ is the untruncated version of the probability distribution of magnitude.

The statistical theory of extreme values is a relatively new independent field of study. Here, referring to extreme values implies that they are the largest values in given periods of time or among a given number of observations. The initial theory was developed systematically, in chronological order, by Fréchet (1927), Fisher and Tippett (1928), Von Mises (1936), and Gnedenko (1943). Gnedenko (1943) was the first to rigorously lay out the theory developed by the previous authors, i.e. the work by Gnedenko (1943) is the culmination of the earlier studies. Pickands (1975) initiated the study of the theory of extreme values as a study field in its own right.

The basis of the EVT is that the distributions of extreme values all tend to the same generalised shape of distribution when rescaled and shifted, except for the anomalous theoretical instances where this distribution does not exist. This distribution is given by the formula

$$\Phi_\gamma(x) = \exp\left[-(1-\gamma x)^{-\frac{1}{\gamma}}\right]. \tag{4.34}$$

For the special case, when the distribution parameter $\gamma = 0$, i.e. the distribution (4.34) takes the simple form

$$\Phi_0(x) = \exp(-e^{-x}). \tag{4.35}$$

In mathematical terms, there exist sequences a_n and b_n, such that

$$\lim_{n\to\infty} F^n(a_n x + b_n) = \Phi_\gamma(x), \tag{4.36}$$

for some γ. According to the definition, $F(x)$ belongs to *the domain of attraction of* $\Phi_\gamma(x)$. The parameter γ is referred to as the shape parameter. Only the instance $\gamma \leq 0$ is of interest here, as finite distributions belong to the domain of attraction of $\Phi_\gamma(x)$ under this condition (De Haan and Ferreira, 2007 p. 19, Theorem 1.2.1.).

Two functions of interest will be defined. Define $U(n) \coloneqq INV(\frac{1}{1-F})$ where $INV(\cdot)$ refers to the inversion operator.

Suppose that for some positive function a and some positive or negative function A with $\lim_{t\to\infty} A(t) = 0$,

$$D(x) \coloneqq \lim_{t\to\infty} \frac{\frac{U(tx)-U(t)}{a(t)} - \frac{x^\gamma - 1}{\gamma}}{A(t)}, \qquad x > 0, \tag{4.37}$$

is some function that is not a multiple of $\frac{1-x^\gamma}{\gamma}$ and not identically zero. The functions a and A are referred to, respectively, as *first order* and *second order* auxiliary functions.

An alternative formulation of EVT is that, given a distribution function $F(x)$

$$\lim_{t\to\infty}\frac{1-F(t+x)}{1-F(t)}=H(x|t)=1-\left(1+\gamma\frac{x}{s}\right)^{-\frac{1}{\gamma}}, \tag{4.38}$$

where s is called the scale parameter. The instance where $\gamma=0$ is interpreted as the limiting case $1-\exp(\frac{x}{s})$. This is interpreted as the shape that the tail of the distribution takes on. Equation (4.38) is referred to as the Generalized Pareto Distribution (GPD).

Kijko and Singh (2011) note that combining the results from Cooke (1979) and Robson and Whitlock (1964) leads to the estimator

$$\widehat{m}_{max}=m_{max}^{obs}+(2\xi)^{-1}\left(m_{max}^{obs}-m_{n-1}\right). \tag{4.39}$$

Close examination reveals that $-\xi$ is, in fact, the shape parameter of the extreme value distribution to which domain of attraction the distribution of magnitude values belongs (see De Haan and Ferreira, 2007, p.19, equation 1.2.3, and p.12, corollary 2.1.4.2). The case considered by Cooke (1979, 1980) is that for which $\xi>0$. Cooke (1980) also presents the numerical values of a_i, for different values of ξ in the equation

$$\widehat{m}_{max}=\sum_{i=1}^{n}a_i m_{n-i+1}. \tag{4.40}$$

It is noted in Cooke (1979, 1980) and Kijko and Singh (2011) that, for a truncated distribution, $\xi=-1$. In this instance

$$\widehat{m}_{max}=m_{max}^{obs}+\frac{1}{2}\left(m_{max}^{obs}-m_{n-1}\right). \tag{4.41}$$

The approximate variance for this estimator is (Kijko and Singh, 2011)

$$Var(\widehat{m}_{\max})=0.5[3\sigma_M^2+0.5(m_{max}^{obs}-m_{n-1})^2]. \tag{4.42}$$

Cooke (1979 and 1980) presents coefficients a_i in equation (4.40) for different values of ξ. The assumption throughout the current work up to now was that the distribution of magnitudes is truncated at m_{max}. However, possibly, the cut-off of the distribution might not be a sharp truncation, such that $f(m_{max})$ is some finite positive number and $f(m)=0$ for $m>m_{max}$. It is quite possible, for instance, that in the very end tail of the distribution, $f(m)$ deviates slightly from the GR law in that it tends to zero in a

continuous fashion and $f(m_{max}) = 0$. In such an instance, ξ can differ from -1, and the results obtained by Cooke could be useful. Coefficients for a_i in equation (3.40) are given in Tables 3 (a–d) for different values of γ, where γ is the parameter defined in the above discussions on the EVT and the GPD, and $\gamma = -\xi$.

Table 3 (a). Coefficients for equation 4.40; $\gamma = \frac{1}{2}$.

R	A1	A2	A3	A4	A5	A6	A7	Square error
2	2	-1						0.667
3	1.636	0.273	-0.909					0.545
4	1.440	0.240	0.160	-0.840				0.480
5	1.314	0.129	0.146	0.109	-0.788			0.438
6	1.224	0.204	0.136	0.102	0.082	-0.748		0.408
7	1.257	0.293	0.239	0.096	0.077	0.064	-0.716	0.386

Table 3 (b). Coefficients for equation 4.40; $\gamma = \frac{1}{3}$.

R	A1	A2	A3	A4	A5	A6	A7	Square error
2	2.5	-1.5						0.564
3	1.951	0.585	-1.537					0.440
4	1.654	0.496	0.372	-1.523				0.373
5	1.463	0.439	0.329	0.269	-1.501			0.330
6	1.328	0.398	0.299	0.244	0.210	-1.479		0.300
7	1.226	0.368	0.276	0.226	0.193	0.171	-1.459	0.277

Table 3 (c). Coefficients for equation 4.40; $\gamma = \frac{1}{4}$.

R	A1	A2	A3	A4	A5	A6	A7	Square error
2	3	-2						0.532
3	2.273	0.909	-2.182					0.403
4	1.882	0.753	0.602	-2.237				0.334
5	1.632	0.653	0.522	0.448	-2.255			0.289
6	1.456	0.583	0.466	0.399	0.355	-2.260		0.258
7	1.325	0.530	0.424	0.363	0.323	0.294	-2.259	0.235

Table 3 (d). Coefficients for equation 4.40; $\gamma = \frac{1}{5}$.

R	A1	A2	A3	A4	A5	A6	A7	Square error
2	3.5	-2.5						0.518
3	2.598	1.237	-2.855					0.384
4	2.117	1.008	0.840	-2.964				0.313
5	1.811	0.863	0.719	0.634	-3.027			0.268
6	1.598	0.761	0.634	0.560	0.509	-3.062		0.236
7	1.439	0.685	0.571	0.504	0.458	0.424	-3.082	0.213

The value of ξ can be determined by employing the techniques described in De Haan and Ferreira (2007). These include the Pickands estimator, the maximum likelihood estimator, the moment estimator, and the negative Hill estimator.

Suppose $(m_1, m_2, m_3, \dots, m_n)$ are observations ordered from smallest to largest. It is assumed that $\gamma = 0$. Fraga Alves and Neves (2014) propose the estimator, as applied to the estimation of m_{max} by Vermeulen and Kijko (2017)

$$\widehat{m}_{max} = m_n + \sum_{i=0}^{k-1} a_{i,k}(m_{n-k} - m_{n-k-i}), \tag{4.43}$$

with $a_{i,k} \coloneqq -\ln(2)^{-1}[\ln(k+i) - \ln(k+i+1)]$ ($a_{i,k} > 0$ is a requirement), $i = 1,2,\dots,k$, such that $\sum_{i=0}^{k-1} a_{i,k} = 1$. The choice of k is made from a sequence such that $\lim_{n\to\infty}(k_n) = \infty$, and $\lim_{n\to\infty} \frac{k_n}{n} = 0$. Values $k = [(\ln(n)]^r$ with $r \in (0,2]$ are recommended by Fraga Alves and Neves (2014). The value of k might not be an integer, as it should be. However, it is practical to merely round it to the nearest integer. The asymptotic distribution of the quantity $\frac{1}{\hat{a}(n/k)}(\widehat{m}_{max} - m_{max})$ can be used to estimate confidence intervals to the estimate, where $\hat{a}(n/k)$ is an empirical *auxiliary* function. The asymptotic distribution of $\frac{1}{\hat{a}(n/k)}(\widehat{m}_{max} - m_{max})$ is given by

$$F(z) = \exp\{\exp(z)\} - \frac{\ln 2}{2} - \frac{\lambda}{\ln 2}, \tag{4.44}$$

where $\lambda \coloneqq \frac{(\ln(2))^2}{2}$. To obtain $\hat{a}(n/k)$, Fraga Alves and Neves (2014) propose the maximum likelihood estimator

$$\hat{a}(n/k) = \frac{1}{k}\sum_{i=0}^{k-1}(m_{n-i,n} - m_{n-k,n}). \tag{4.45}$$

Accordingly, the distribution (4.44) can be used to estimate the variance and confidence intervals of the estimator (4.44).

Vermeulen and Kijko (2017) note that the assumption $\gamma = 0$ can be controversial. The case where an extreme value distribution with $\gamma = 0$ is bounded is a relatively singular instance. It can be shown that all the derivatives at the end point must be zero. However, the results from Vermeulen and Kijko (2017) obtained in Monte Carlo simulations using this method are satisfactory.

In the instance where $\gamma < 0$, the tail end can be estimated by a routine maximum likelihood estimation. Vermeulen and Kijko (2017) investigate this estimator. From Coles (2001), the joint likelihood equation of γ and s in the GPD is derived [see equation (4.32)]

$$l(s,\gamma) = -k\log s - \left(1 - \frac{1}{\gamma}\right)\sum_{i=1}^{k}\log\left(1 + \frac{\gamma m_i}{s}\right), \tag{4.46}$$

which should be maximised to obtain the maximum likelihood estimator $(\hat{\gamma}_{MLE}, \hat{s}_{MLE})$ of (γ, s). The maximum likelihood estimator of the end point of the distribution is then given by (De Haan and Ferreira, 2007)

$$\widehat{m^*} = m_n - \frac{\hat{s}_{MLE}}{\hat{\gamma}_{MLE}}. \tag{4.47}$$

Note that the term $\frac{\hat{s}_{MLE}}{\hat{\gamma}_{MLE}}$ is negative and is still of the form $m_{max} = m_{max}^{obs} + \Delta$. The asymptotic distribution of the variable $\zeta = \sqrt{k}\gamma^2\frac{\widehat{m^*} - m^*}{\gamma}$ is given by De Haan and Ferreira (2006) to determine the variance and confidence limits of $\widehat{m^*}$. Vermeulen and Kijko (2017) replace the quite complex statement of the variance by

$$\begin{aligned} var[\widehat{m^*}] &= var\left[m_n - \frac{\hat{s}_{MLE}}{\hat{\gamma}_{MLE}}\right] \\ &\approx var[m_n] - \left(\frac{\hat{s}_{MLE}}{\hat{\gamma}_{MLE}}\right)^2\left(\frac{var[\hat{s}_{MLE}]}{\hat{s}_{MLE}{}^2} + \frac{var[\hat{\gamma}_{MLE}]}{\hat{\gamma}_{MLE}{}^2} - 2\frac{cov[\hat{s}_{MLE}, \hat{\gamma}_{MLE}]}{\hat{s}_{MLE}, \hat{\gamma}_{MLE}}\right), \end{aligned} \tag{4.48}$$

''where $var[\hat{s}_{MLE}]$, var$[\hat{\gamma}_{MLE}]$, and $cov[\hat{s}_{MLE}, \hat{\gamma}_{MLE}]$ are readily estimated by calculating the inverse information matrix of $(\hat{s}_{MLE}, \hat{\gamma}_{MLE})$, and $var[x_n]$ corresponds to the standard error of the determination of the largest value x_n'' (Vermeulen and Kijko, 2017).

The results obtained by Vermeulen and Kijko (2017) using the maximum likelihood method are somewhat biased toward larger values; in other words, the estimate is conservative. However, the bias vanishes

asymptotically. The method is quite generic, and the assumptions about the shape of the tail are minimal. This is a positive aspect, as there is considerable controversy about the shape of the upper tail (see e.g. Pisarenko and Sornette, 2004; Kagan and Schoenberg, 2001; Pisarenko *et al*., 2003; Pisarenko and Sornette, 2006). (Digressing slightly, the EVT has been found ideal to analyse the properties of the upper tail of the FMD.)

Another direct consequence of the EVT is that

$$\lim_{t\to\infty}\frac{1-F(t+g(t)x)}{1-F(t)}=1-(1+\gamma x)^{-\frac{1}{\gamma}}, \tag{4.49}$$

for some function $g(t)$. Beirlant *et al.* (2017) propose, what they call, the Truncated GPD Estimator. Fix t, the threshold above which magnitude values in the tail of the distribution are considered. Their estimator of m_{max} is given by

$$\widehat{m}_{max}=m_{n-k,n}+\frac{\widehat{g(t)}_k}{\hat{\gamma}_k}\left[\left(\frac{1-k}{\left(1-\frac{\widehat{g(t)}_k}{\hat{\gamma}_k}(m_n-m_{n-k})^{\hat{\gamma}}\right)}\right)-1\right], \tag{4.50}$$

where k is such that $t=m_{n-k}$. The estimation $(\hat{\gamma}_k,\widehat{g(t)}_k)$ of $(\gamma,g(t))$ is done using the maximum likelihood method (see Beirlant *et al.,* 2017 for more detail). The estimator is applied to different values of k (equivalently t). It can be represented in a diagram similar to that of the Hill plot. An approximate $100(1-\alpha)\%$ upper confidence bound for $\widehat{m}_{max}$ is given by

$$\begin{aligned}\widehat{m}_{max}-(\ln\alpha+1)&\frac{(k+1)^2}{(n+1)\{[1-F_t(\widehat{m}_{max})]/F_t(\widehat{m}_{max})\}}\\&\times\left(1+\frac{(k+1)}{(n+1)\{[1-F_t(\widehat{m}_{max})]/F_t(\widehat{m}_{max})\}}\right)^{\hat{\gamma}_k}\frac{{\hat{\gamma}_k}^2}{\widehat{g(t)}_k},\end{aligned} \tag{4.51}$$

where $F_t(m)=\frac{F(m+t)}{F(t)}$.

Another estimator that Beirlant *et al.* (2017) propose is, what they call, the Truncated Pareto Estimator. Because magnitude has been established empirically to follow an exponential distribution, the tail does not tend to a Pareto distribution, as heavy tails do. However, the exponential distribution of magnitude implies a Pareto distribution of energy E. It is easy to show that if a random variable X that follows an exponential distribution, $Y=e^X$, follows a Pareto distribution. Beirlant *et al.* (2016) propose a method for estimating the end point of a truncated Pareto distribution, to which the Truncated Pareto Estimator

refers. Accordingly, the Truncated Pareto Estimator can be applied to seismic energy values that are related to the moment magnitude through the relation (Lay and Wallace, 1995)

$$E = 2 \times 10^{1.5(M-1)} = \exp[\ln 2 + (M-1)1.5 \ln 10].$$

Denote the random variable energy by E, and the value under investigation by e. Using the method by Beirlant *et al.* (2016), the truncation point of the Pareto distribution for the model of E is then estimated by

$$\hat{e}_{max} = 2 \times 10^{(m_{n-k}-1)} \left(\frac{\left(\frac{2 \times 10^{1.5(m_{n-k}-1)}}{2 \times 10^{1.5(m_n-1)}} \right)^{1/\hat{\gamma}_{E,k}} - \frac{1}{k+1}}{1 - \frac{1}{k+1}} \right)^{-\hat{\gamma}_{E,k}}, \tag{4.52}$$

then

$$\hat{m}_{max} = \frac{\log_{10}\left(\frac{\hat{e}_{max}}{2}\right)}{1.5} + 1. \tag{4.53}$$

Again, the estimation should be done for different values of k and plotted with k on the abscissa and $\hat{m}_{max}$ on the ordinate. An approximate upper $100(1-\alpha)\%$ confidence bound is given by

$$\hat{m}_{max} - \frac{(k+1)^2 \hat{\gamma}_{E,k}(\ln \alpha + 1)}{(n+1)\{[1 - F_{E,k}(\hat{e}_{max})]/F_{E,k}(\hat{e}_{max})\}} / 1.5 \ln 10. \tag{4.54}$$

Leaving the EVT aside, another non-parametric estimator for m_{max} was developed by Kijko *et al.* (2001) for instances where the FMD is not modelled easily by a specific distribution. This is typical for mining areas and seismicity associated with volcanic activity. In this instance, the FMD is modelled by a sum of Gaussian kernel distributions. The distribution function takes the form

$$F_M(m) = \begin{cases} 0 & for\ m < m_c \\ \frac{\sum_{i=1}^{n} \left[\Phi\left(\frac{m - m_i}{h}\right) - \Phi\left(\frac{m_c - m_i}{h}\right) \right]}{\sum_{i=1}^{n} \left[\Phi\left(\frac{m_{max} - m_i}{h}\right) - \Phi\left(\frac{m_c - m_i}{h}\right) \right]} & for\ m_c \le m \le m_{max}, \\ 1 & for\ m > m_{max} \end{cases} \tag{4.55}$$

and the density function

$$f_M(M) = \begin{cases} 0 & for\ m < m_c \\ \dfrac{\left(h\sqrt{2\pi}\right)^{-1} \sum_{i=1}^{n} \exp\left[-\frac{1}{2}\left(\frac{m_c - m_i}{h}\right)^2\right]}{\sum_{i=1}^{n}\left[\Phi\left(\frac{m_{max} - m_i}{h}\right) - \Phi\left(\frac{m_c - m_i}{h}\right)\right]} & for\ m_c \leq m \leq m_{max}, \\ 1 & for\ m > m_{max} \end{cases} \tag{4.56}$$

where $\Phi(\cdot)$ is the error function and h is a smoothing factor. The smoothing factor h is of paramount importance. Kijko *et al.* (2001) used what is called *least squares cross-validation*, described in Silverman (2018). Without going through the derivation, which can be found in Kijko *et al.* (2001), h turns out to be the solution of the equation

$$\sum_{i,j}\left\{\frac{1}{\sqrt{2}}\left[\frac{(m_i - m_j)^2}{2h^2} - 1\right]\exp\left[-\frac{(m_i - m_j)^2}{4h^2}\right] - 2\left[\frac{(m_i - m_j)^2}{h^2} - 1\right] \times \exp\left[-\frac{(m_i - m_j)^2}{2h^2}\right]\right\} - 2n = 0. \tag{4.57}$$

Estimators (4.2), (4.9), or (4.15) can be applied subsequently to $F_M(m)$ and $F = f_M(m)$ in equations (4.5) and (4.5), respectively. Note that although a trial has not yet been conducted on the application of estimators (4.3) to (4.9), they are clearly valid estimators.

4.3. Least Squares and the L_p Norm

The least squares approach has also been applied to the estimation of the pair (β, m_{max}), but the validity of this estimator is questionable, as is the application of the least squares approach to determine the b-value. If the least squares approach can be applied, the minimisation of the L_1 norm is also a suitable alternative, as it is less sensitive to outliers. The L_1 norm is defined as

$$\Phi(\boldsymbol{\theta}) = \sum_{\mathbf{i=1}}^{\mathbf{n}} |F_M(m_i) - \hat{F}_M(m_i)|, \tag{4.58}$$

where $\boldsymbol{\theta}$ is varied, so as to minimise $\Phi(\boldsymbol{\theta})$. In this instance, $= (\beta, m_{max})$. The function $\hat{F}_M(m_i)$ is the empirical distribution function and is given by $\hat{F}_M(m_i) = \frac{i}{n+1}$. Further, Kijko (1994) describes a procedure to determine the best value for p in the L_p norm, where the L_p norm is similarly defined as

$$\Phi(\boldsymbol{\theta}) = \sum_{\mathbf{i=1}}^{\mathbf{n}} |F_M(m_i) - \hat{F}_M(m_i)|^p. \tag{4.59}$$

As mentioned in Chapter 3, the generalised least squares method is more suitable when determining the b-value on an unbounded exponential distribution, as done by Guttorp (1987). Vermeulen and Kijko (2017) attempted to introduce m_{max} and add it as a parameter to minimise the Mahalanobis distance, i.e. $\boldsymbol{\Theta} = (\beta, m_{max})$ is to be varied. The details of the procedure are exactly the same as those for the case in the method of Guttorp (1987), described in Chapter 3. However, in numerical experiments, Vermeulen and Kijko (2017) found that this procedure was not suitable, as ''the generalised least squares (GLS) method implemented with the results of Guttorp (1987) does not always provide a (unique) solution... The sensitivity of the inversion procedure to the parameter $\mathrm{m_{max}}$ causes a number of problems. In some instances, it is possible for the minimum to be obtained with $\mathrm{m_{max}}$ at infinity, which can be described intuitively as an inability to detect the required magnitude endpoint." (Vermeulen and Kijko, 2017). "On other occasions, a minimum of residuals would be reached at a value of $\mathrm{m_{max}}$ that is close to but less than the maximum observed magnitude. Attempts to find a minimum to the generalised squared residual function were often unsuccessful, which probably indicates that the method is not able to detect $\mathrm{m_{max}}$, as it tends to put the value of $\mathrm{m_{max}}$ at infinity. Note that employing a catalogue with a larger range does not improve the situation; that is, the method does not perform well with a [larger] magnitude range of $m_{min} - m_c$. Consequently, in practice, it appears that a range of one order of magnitude is too large. The reason for this counter-intuitive result is that although the effect of the truncation is mostly detected in the larger magnitude values, the smaller values are allocated the largest weight." (Vermeulen and Kijko, 2017).

4.4. Toward a Theory of the Kijko-Sellevoll Estimator

The Kijko-Sellevoll estimator was initially formulated by Kijko (1988) by the development of an estimator of the expected value of the maximum earthquake that could possibly occur within the time span of the catalogue, given the maximum observed magnitude within the catalogue. By using the moment generating function, Kijko (1988) arrived at a formulation that turns out to be an excellent approximation of and almost equivalent to the estimator commonly known as the Kijko–Sellevoll estimator (equation 4.60 below). In the meantime, Kijko (2004) had established a link with the work of Cooke (1979, 1980). The estimator is now used often but little mathematical formalism behind it is established. Here, the work of Haraala and Orosco (2016, 2018a, 2018b) is presented as a mathematical formalism and a step toward a theory behind the Kijko–Sellevoll estimator.

Haraala and Orosco (2016, 2018a, 2018b) established some essential and elegant mathematical formalism behind the Kijko–Sellevoll estimator, stated here, for the purpose of clarity, as

$$m_{max} = E(M_n|m_{max}) + \int_{m_c}^{m_{max}} \left\{ \frac{1 - \exp[-\beta(x - m_c)]}{1 - \exp[-\beta(m_{max} - m_c)]} \right\}^n dx. \tag{4.60}$$

Central to the theory are the functions that are referred to by Haraala and Orosco (2016, 2018a, 2018b) as the Kijko–Sellevoll functions, f_n^{ks1}, f_n^{ks2}, and f_n^{ks3} , which satisfy

$$E(M_n|m_{max}) = m_{max} - \frac{1}{\beta} f_n^{KS1}(z), \tag{4.61}$$

$$E(M_n|m_{max}) = m_{min} + \frac{1}{\beta} f_n^{KS2}(z), \tag{4.62}$$

$$Var(M_n|m_{max}) = \frac{1}{\beta^2} f_n^{KS3}(z), \tag{4.63}$$

where $z = \beta(m_{max} - m_c)$ and are given in closed (analytical) form by

$$f_n^{KS1}(x) = \sum_{k=1}^{\infty} \frac{(1 - e^{-x})^k}{k + n}, \tag{4.64}$$

$$f_n^{KS2}(x) = \sum_{k=1}^{\infty} \frac{n}{k} \frac{(1 - e^{-x})^k}{k + n}, \tag{4.65}$$

$$f_n^{KS3}(x) = \sum_{k=1}^{\infty} \frac{2n}{2n + k} \left\{ \sum_{j=1}^{k-1} \frac{1}{n + j} \right\} \frac{(1 - e^{-x})^k}{k + n}. \tag{4.66}$$

To show that (4.59) has a unique solution, the function g is defined as

$$g(\mathfrak{m}) = \mathfrak{m} - E(M_n|m_{max}) - \int_{m_{min}}^{\mathfrak{m}} F_{M_n}(x) dx. \tag{4.67}$$

Differentiating g (without going into much detail), the following is obtained

$$g'(\mathfrak{m}) = -n \int_{m_{min}}^{\mathfrak{m}} [F_m(x|\mathfrak{m})]^{n-1} \frac{\partial}{\partial \mathfrak{m}} F_m(x|\mathfrak{m}) dx \geq 0, \tag{4.68}$$

for all $\mathfrak{m} \in [m_c, \infty)$, i.e. the function is monotone decreasing. Furthermore, the point $\mathfrak{m} = E(M_n|m_{max})$ is negative. Therefore, equation (4.67) has a unique solution. This is extremely important, as it is proof that the estimator, if it did exist, will always have a unique solution. This is indeed an important result. Lasocki and Urban (2011) found that the simple iterative scheme diverges, which, under close examination, is also apparent from the study by Haraala and Orosco (2016).

Attempting to find an exact solution of the integral in (4.60), Haraala and Orosco (2016) apply the so-called "integration by parts" technique iteratively until the integral terms no longer appear. This results in the sum

$$\int_{m_{min}}^{m_{max}} \left\{\frac{1-\exp[-\beta(x-m_c)]}{1-\exp[-\beta(m_{max}-m_c)]}\right\}^n dx = \frac{1}{\beta}\frac{\beta(m_{max}-m_c) - \sum_{k=1}^{n}\frac{\{1-\exp[-\beta(m_{max}-m_c)]\}^k}{k}}{1-exp[-\beta(m_{max}-m_c)]}. \tag{4.69}$$

Using the identity (Abramowitz and Stegun, 1964)

$$-\log(1-z) = \sum_{k=1}^{\infty}\frac{z^k}{k}, \tag{4.70}$$

and writing

$$\beta(m_{max}-m_c) = -\log(1-(1-\exp[-\beta(m_{max}-m_c)]) = \sum_{k=1}^{\infty}\frac{(1-\exp[-\beta(m_{max}-m_c)])^k}{k}, \tag{4.71}$$

and setting $z = \exp[-\beta(m_{max}-m_c)]$, it is possible to express(4.69) in the form

$$\int_{m_{min}}^{m_{max}} \left\{\frac{1-\exp[-\beta(x-m_c)]}{1-\exp[-\beta(m_{max}-m_c)]}\right\}^n dx = \frac{1}{\beta}\frac{\sum_{k=n+1}^{\infty}\frac{\{1-\exp[-\beta(m_{max}-m_c)]\}^k}{k}}{1-exp[-\beta(m_{max}-m_c)]}. \tag{4.72}$$

By dividing each term by the denominator and re-indexing, the following is obtained

$$\int_{m_{min}}^{m_{max}} \left\{\frac{1-\exp[-\beta(x-m_c)]}{1-\exp[-\beta(m_{max}-m_c)]}\right\}^n dx = \frac{1}{\beta}\sum_{k=1}^{\infty}\frac{\{1-\exp[-\beta(m_{max}-m_c)]\}^k}{k+n}. \tag{4.73}$$

Therefore, the series in equation (4.73) is a representation of the function f_n^{KS1}, as seen in equation (4.61).

Equation (4.73) is not only a more compact and more elegant representation of the series in question but also it is numerically much more stable. However, the rate of convergence is extremely slow. Sequence transformations for conversion acceleration can be used to address this problem effectively. Haraala and Orosco (2016) present an algorithm for such a transformation that is most effective and improves the rate of convergence by approximately 16 orders of magnitude.

From the function f_n^{KS2} derives another representation of the expected maximum value in terms of the true maximum value

$$E(M_n|m_{max}) = m_c + \int_{m_c}^{m_{max}} [1 - F_{M_{(n)}}(m|m_{max})]\, dm \tag{4.74}$$

where

$$\int_{m_c}^{m_{max}} [1 - F_{M_{(n)}}(m|m_{max})]\, dm \tag{4.75}$$

is the first representation of the function f_n^{KS2} [compare with equation (4.62)] multiplied by the coefficient $\frac{1}{\beta}$. Expression (4.75) can be written as

$$(m_{max} - m_c) - \int_{m_c}^{m_{max}} F_{M_{(n)}}(m|m_{max})\, dm. \tag{4.76}$$

Replacing $(m_{max} - m_c)$ and $\int_{m_c}^{m_{max}} F_{M_{(n)}}(m|m_{max})\, dm$ by the corresponding series in equations (4.71) and (4.73), respectively, the expression is obtained

$$\frac{1}{\beta}\sum_{k=1}^{\infty}\frac{(1-\exp[-\beta(m_{max}-m_c)])^k}{k} - \frac{1}{\beta}\sum_{k=1}^{\infty}\frac{\{1-\exp[-\beta(m_{max}-m_c)]\}^k}{k+n} = \frac{n}{\beta}\sum_{k=1}^{\infty}\frac{(1-\exp[-\beta(m_{max}-m_c)])^k}{k(k+n)}. \tag{4.77}$$

Accordingly, the function f_n^{KS2} can be represented by the series

$$n\sum_{k=1}^{\infty}\frac{(1-\exp[-\beta(m_{max}-m_c)])^k}{k(k+n)}. \tag{4.78}$$

Note that the relation between the functions f_n^{KS1} and f_n^{KS2} is

$$\beta(m_c-m_{max})=f_n^{KS-1}\big(\beta(m_c-m_{max})\big)+f_n^{KS-2}\big(\beta(m_c-m_{max})\big). \tag{4.79}$$

Equation (4.79) provides a way of expressing the KS-2 function as a formula involving a finite sum

$$f_n^{KS-2}\big(\beta(m_c-m_{max})\big) = \frac{[(1-exp[-\beta(m_{max}-m_c)])^n-1]\beta(m_c-m_{max})+\sum_{k=1}^{n}\frac{(1-exp[-\beta(m_{max}-m_c)])^k}{k}}{(1-exp[-\beta(m_{max}-m_c)])^n}. \tag{4.80}$$

However, Haraala and Orosco (2016) do not recommend the use of equation (4.80) to calculate f_n^{KS2} independently because of numerical instabilities. Instead, KS-1 is to be calculated first and equation 4.78 can then be used to calculate the value of the KS-2 function. The reason for having a KS-2 function as well is that it presents a natural way of estimating m_{max}, in that KS-2 has an inverse function satisfying

$$f^{-1}(\beta[E(M_n|m_{max})-m_c])=\beta(m_{max}-m_c), \tag{4.81}$$

and it does not diverge as m_{max} approaches infinity, whereas the KS-1 function does. Haraala and Orosco (2018a) show that the KS-2 function maps the interval $[m_c,\infty]$ onto $[m_c,m_c+1/H_\eta]$. Accordingly, using equation (4.81), any quantity above m_c+1/H_η will map m_{max} beyond infinity.

Haraala and Orosco (2018a) prove that the KS functions exist for all real numbers. Haraala and Orosco (2018b) present a superior algorithm to calculate generalised harmonic numbers using the Ramanujan harmonic expansion.

In the process, two more special functions are defined that are related closely to well-known special functions. These are the Incomplete Beta Function of the Second Kind and the Incomplete Psi Function of the Second Kind (Abramowitz and Stegun, 1964). The Kijko–Sellevoll functions are related to these special functions, which, in turn, relate to the well-known Beta and Psi functions.

The Incomplete Beta Function $B_x(a,b)$ is given by (Abramowitz and Stegun, 1964)

$$B_x(a,b)=\int_0^x\frac{t^{a-1}}{(1-t)^{b-1}}dt. \tag{4.82}$$

Analogous with the Incomplete Beta Function, Haraala and Orosco (2018a) define the Incomplete Beta Function $\mathcal{B}_x(p)$ of the Second Kind as

$$\mathcal{B}_x(p) = \int_0^x \left(\frac{t}{x}\right)^{p-1} \frac{1}{1-t}\, dt. \tag{4.83}$$

The functions B_x and $\mathcal{B}_x$ are related as

$$\mathcal{B}_x(p) = \frac{B_x(p,0)}{x^{p-1}}. \tag{4.84}$$

It is found that

$$E(\beta(m_{max} - m)|m_{max}) = \mathcal{B}_{t_{max}}(\eta + 1), \tag{4.85}$$

where $t_{max} = 1 - \exp[\beta(m_{max} - m_c)]$, and η is the frequency of observation (usually a whole number).

Analogous to the Incomplete Beta Function, Haraala and Orosco (2018a) define the Incomplete Psi Function as

$$\psi_x(z) = \int_0^x \frac{1 - t^{z-1}}{1-t}\, dt - \gamma, \tag{4.86}$$

where γ is a Euler–Mascheroni constant. The Incomplete Psi Function of the Second Kind is defined as

$$\Psi_x(z) = \int_0^x \left(1 - \left(\frac{t}{x}\right)^{z-1}\right) \frac{1}{t-1}\, dt - \gamma. \tag{4.87}$$

Then, it is found that

$$E(m|m_{max}) = \Psi_{t_{max}}(\eta + 1). \tag{4.88}$$

It may be seen that $\Psi_{t_{max}}(\eta + 1)$ behaves as a harmonic number H_η when $m_{max} \to \infty$, i.e. when $t_{max} \to 1$. In other words, for integral values of η,

$$E(m|m_{max}) \sim \sum_{k=1}^{\eta} \frac{1}{\eta}. \tag{4.89}$$

Note that, for the positive real values, H_η is defined as

$$H_\eta = \int_0^1 \frac{1-t^\eta}{1-t} dt. \tag{4.90}$$

In Haraala and Orosco (2018b), as a corollary to determining a practical way of solving the harmonic series in f_n^{KS3}, the Ramanujan expansion of harmonic numbers is revisited and is generalised naturally to all generalised harmonic numbers (defined over the positive real numbers and not only the integer values). The study by Haraala and Orosco (2018b) is devoted in large part to proving that this generalisation holds. Most important to note is that the Ramanujan expansion employs triangular numbers $m = \frac{\eta(\eta+1)}{2}$ for the harmonic number H_η and the Bernoulli numbers. The Bernoulli numbers are discussed in some detail because of their importance in the Ramanujan expansion. Equally important to consider is the error term when using a finite portion of the Ramanujan expansion because of the consequences when applying this expansion in the numerical calculation of harmonic numbers. It has been found that this error quickly becomes exceedingly small as the number of expansion terms increases. The implication is that using the Ramanujan expansion would be more efficient than the algorithms used generally to compute harmonic numbers, e.g. those implemented in MATLAB, a widely used computer programming language (MathWorks, USA). Accordingly, resulting from the study by Haraala and Orosco (2018b), the Ramanujan expansion is found most useful in reducing computational complexity and, although a formal comparison of the computational complexity has not been conducted, this opinion is based on empirical evidence.

5. Seismic Source Characterisation and Seismic Zones

5.1. Seismic Zoning

It is customary practice in PSHA to define seismic source features geometrically and in terms of location for use in the hazard analysis, with each feature (zone or fault line) assigned its specific, constant seismicity parameters. The Senior Seismic Hazard Analysis Committee (SHHAC; Budnitz *et al.,* 1997) document *Recommendations for Probabilistic Seismic Hazard Analysis: Guidance on uncertainty and use of experts* (Budnitz *et al.*, 1997) remains the main reference for constructing seismic source zone models. Each zone or fault line, defined by its geometry, shape, and location (a line, or feature, or area in space), is assigned the parameter combination (m_{max}, b, λ) (or, more generally, an FMRL and a temporal recurrence model).

Budnitz *et al.* (1997) define four different types or categories of seismic sources, namely, Type 1, faults (represented as linear or planar features), Type 2 "Area sources enclosing concentrated zones of seismicity", Type 3 "Regional area sources", and Type 4 "Background area sources".

Budnitz *et al.* (1997) define a seismic source as a region of the crust of the Earth with quite uniform seismic characteristics and which is dissimilar to those of adjacent sources. These authors contend that some variation in the seismicity parameters (a- and b-values) within a given seismic source is permissible. However, typically, some factors are considered uniform in a seismic source, such as the distribution of m_{max} and the probability of activity (Budnitz *et al.,* 1997).

Fault geometry, Type 1, is necessarily carried out in detail, and the dip and extent of the fault plane are determined. This is a well-established, routine task for geologists and geophysicists and the accuracy is quite acceptable. Type 2 becomes somewhat more difficult when multiple possible sources are identified but specific source locations cannot be delineated. As shown in Figure 10, the boundaries of concentrated seismicity can be constructed artificially, excluding a small number of earthquakes; therefore, boundaries are uncertain. Bender (1986) proposes a distributed model, of which the boundaries are referred to as "fuzzy" boundaries by Budnitz *et al.* (1997).

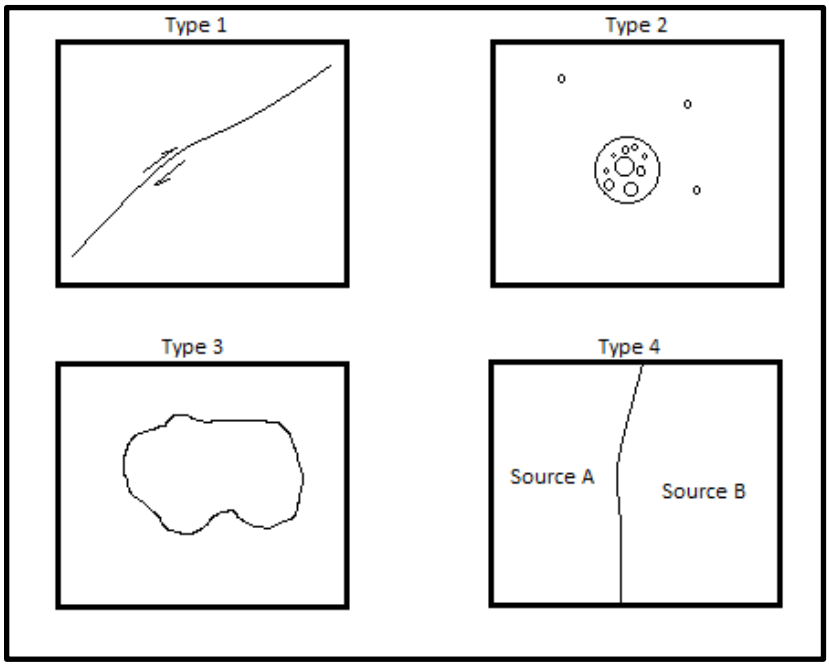


Figure 10. A seismic source [zone] is a construct developed for seismic hazard analysis as a means of approximating the locations of earthquake occurrences. (Figure and caption from Budnitz *et al.*, 1997).

According to the guidelines of Budnitz *et al*. (1997), the first task is to identify the seismic source zones. However, these authors refer to Thenhaus (1983), who is of the opinion that ill-defined procedures are

used to delineate seismic source zones. Furthermore, according to Thenhaus (1983), there is no single standard to delineate seismic source zones across a vast area, such as that of the USA, owing to the non-uniform appropriate seismological, geological, and geophysical information available for areas with such differing tectonic and geologic settings. Thenhaus (1983) states, as noted in Budnitz *et al*. (1997), that seismicity in the USA is associated ambiguously with a geologic structure, which only serves to compound the problem. Therefore, according to Budnitz *et al.* (1997), as the evaluation and interpretation of the available information depend strongly on individual expert judgment or opinion, different source zones can be defined differently by different experts.

Budnitz *et al.* (1997) define the procedures used in delineating seismic source zones somewhat more rigorously; however, arguably, there is still room for improvement. Various guidelines have been proposed for the interpretation of the different types of information; however, this still, "depend[s] strongly on the individual judgment or the opinion" (Bender, 1986).

Seismic sources of Types 3 and 4 suffer even more from the problems pointed out by Bender (1986). "The area-source boundaries enclose regions that earth scientists believe are relatively uniform with respect to the PSHA application" (Budnitz *et al.,* 1997).

The "fuzzy" boundaries suggested by Bender (1986) are generally not used. According to Budnitz *et al.* (1997), the uncertainties in source boundaries are employed in hazard analysis by identifying alternative source configurations, each having its own relative weight or credibility. This weighting scheme adds another dimension of expert opinion, with inherent, possibly subjective, bias.

Table (4) summarises the data types employed to define the types of seismic sources and the relative usefulness of each data type by Budnitz *et al.* (1997). According to Budnitz *et al*. (1997), ''relative usefulness'' refers to how sound the technical basis is that the data provide for the source definition. This table does provide a set standard; however, it remains an arbitrary choice and there is room for a substantial difference of opinion.

5.2. Frankel Alternative to Seismic Source Zoning

An alternative to the Cornell (1968) concept of zoning is proposed by Frankel (1995). This method does not utilise expert opinion at all and is purely data driven; in other words, the seismic source characteristics are deduced from the data only by using statistical methods. This method contrasts with the traditional seismic zoning method.

"*One of the motivations for directly using the smoothed historical seismicity is to get away from the judgments involved in drawing seismic source zones in a region where the causative structures of seismicity are largely unknown, such as the central and eastern U.S. In some respects, our approach goes against a recent trend in seismic hazard analysis for using several experts to choose separate sets of source zones*." Frankel (1995).

Table 4. Data Used to Assess Seismic Source Locations and Geometries and Their Relative Usefulness" (Table caption and data from Budnitz *et al.*, 1997).

TYPE OF SOURCE	DATA/BASIS FOR SOURCE	RELATIVE USEFULNESS/CREDIBILITY
Type 1: Faults	Mapped fault with historical rupture	1
	Mapped Quaternary fault at surface	1
	Mapped localised Quaternary deformation, inferred fault at depth	2
	Borehole evidence for fault, especially in young units	2
	Geophysical evidence (e.g. seismic reflection) of fault at depth	2
	Map of pre-Quaternary faults	3
Type 2: Concentrated Zone	Concentrated zone of well-located instrumental seismicity	1
	Mapped fault(s) at surface or subsurface in proximity to seismicity	1
	Zone of historical/poorly located seismicity	2
	Structural features/trends parallel to seismicity zone	2
	Focal mechanisms/stress orientation	3
	Rapid lateral changes in structures/tectonic features	3
Type 3: Regional Zone	Changes in spatial distribution/concentration/density of seismicity	1
	Regions of genetically related tectonic history	1
	Regions of similar structural styles	2
	Changes in crustal thickness or crustal composition	2
	Regions of different geophysical signature	3
	Changes in regional stress	3
	Changes in regional physiography	3
Type 4: Background Zones	Regional differences in structural styles/tectonic history	1
	Major physiographic/geologic provinces	1
	Changes in character of seismicity	3

Frankel (1995) explains that directly using the smoothed historical seismicity is based on an effort to do away with the subjective decisions related to having to delineate seismic source zones with mostly unknown causative structures of seismicity, e.g. the central and eastern U.S.A. The author affirms that the simple methodology used by Frankel (1995) produces values similar to the mean probability of exceedance obtained by a more complex EPRI study that employs teams of experts to determine sets of area source zones.

Frankel cites Jacob *et al.* (1994) who used spatial smoothing on seismicity rates for low-magnitude earthquakes. Frankel (1995) expands on this approach, using spatially smoothed seismicity rates supplemented with a heuristic approach to incorporate the historical part of the catalogue. This incorporates the background seismicity that is supposed to incorporate prehistoric, geologically inferred seismicity (see Frankel, 1995, for more detail). The next chapter describes a more sophisticated, data-driven approach to determining non-uniform spatial seismicity parameters.

Although expert opinion has the potential to introduce bias and epistemic uncertainty, it must be recognised that, in many instances, it is still valuable and even necessary. In the following chapter, alternative methods for PSHA are discussed. Several of these methods allow for data-driven models, as well as including expert opinion in a rigorous manner that is acceptable in current scientific practice.

6. Alternative Methods for PSHA

As mentioned already, the seminal Cor68 paper serves as the inception of the study field known as Probabilistic Seismic Hazard Analysis. The Cor68 procedure has become an integral part of PSHA and the author (Cor68) makes a considerable contribution to the framework for PSHAs currently conducted. Cor68 provides a method for determining the return periods of the level of strong ground motion at a particular site. It is important to note that Cor68 indicates the generality of the procedure, which the author applied in a specific instance, and which served as a case study in some sense.

Originally, the procedure Cor68 developed was intended for engineers, and the author pertinently mentions the sources from which data could be obtained. Clearly, Cor68 realises the need for careful and accurate estimation of the data that seismologists and geologists can provide. As engineers are not expected or advised to employ such data outside their field of expertise, Cor68 does not discuss how these data should be recorded and processed into the form required by them.

It is important to note that modern PSHAs include procedures regarded by Cor68 as solely the province of seismologists and geologists; however, with regard to the information required by engineers. Although engineers are also involved in the processing of the data, geoscientists (seismologists and geologists) often carry out entire PSHA procedures. It must be noted here that such interchanging of the roles of experts can become problematic without proper protocol and, in the opinion of the current author, often no protocol is followed at all.

A typical PSHA is conducted by employing the Cornell–McGuire procedure, as follows:

(1) Identification of the seismic sources (or seismic zones).
(2) Estimating the largest possible earthquake m_{max} that could occur in each zone.
(3) Assuming a Gutenberg–Richter FMD, determining the b-value (or equivalently the β) and, assuming a Poisson process, determining the activity rate of each zone. Accordingly, the parameter vector $\boldsymbol{\Theta} = (\beta, \lambda)$ is introduced. A uniform distribution for $\boldsymbol{\Theta}$ is usually assigned to each source. Although this might not reflect reality at all, at the time of publication of Cor68, this was the extent of information available, as well as the methods allowed. However, this practice has persisted and is still employed today. Therefore, in practice, determining this value has evolved into a combined effort from engineers and geoscientists, with the teams of specialists collaborating to integrate different types of information into estimates of $\boldsymbol{\Theta}$.
(4) Developing (or choosing) a ground motion prediction equation to estimate the actual strong ground motion parameters at the site, given an earthquake of magnitude m at a distance r. Conveniently, the simplest form of GMPE is $s(m, r) := s = c_1 + c_2 m + c_3 \ln(r)$, resulting in an exponential frequency-ground-shaking relation. Currently, the foremost practitioners of PSHA insist on extremely specific GMPEs.
(5) Using the information from the previous steps to obtain a function/graph/curve reflecting the probability of exceedance at the site under consideration. Using the Total Probability Theorem, the probability of exceedance of the strong ground motion parameter is calculated as a function of the distribution of seismic source FMD. The marginal distributions of magnitude and distance are summed over to produce a combined distribution function of the strong ground motion parameter. This is referred to by McGuire (1993) as a *deductive* procedure.

Veneziano *et al.* (1984) propose a different approach, using only catalogue data, which McGuire (1993) refers to as the *historical* procedure. A method to fit a model to catalogue data is referred to by McGuire (1993) as a *parametric-historic* procedure. Veneziano *et al.* (1984) consider their proposed method complementary to what they refer to as traditional methods. As catalogue data are used, the explicit identification of source zones is bypassed. These authors present two different approaches, the first of which is entirely non-parametric and does not deal with the uncertainty of the estimate. The second approach derives from a vastly different paradigm, i.e. considering the probability of exceedance as a random variable. This latter method does make use of models that require the estimation of parameters. It appears from the available literature that the *parametric-historic* procedure is the same as the widely known "Parametric-historic" procedure of Kijko and Graham (1998, 1999), but the current author cannot find the procedure that Veneziano *et al.* (1984) allude to in the abstract of their report. In their idea of the *parametric-historic* procedure, Veneziano *et al*. (1984) propose a method to deal with catalogue incompleteness by using a computer program that is able to deal with incompleteness, as well as different types of data inputs and different GMPEs. However, despite a thorough literature search, the current author was unable to locate a description of this method. It appears to be practically lost. The idea summarised by Veneziano *et al.* (1984) is realised in the Parametric-historic procedure of Kijko and Graham (1998, 1999).

The deductive-type procedures are attractive, as they are able to account for all the types of deviations from the "standard" model, i.e., "it accounts for phenomena such as migration of seismicity, seismic 'gaps' or, in general, any nonstationary properties of seismicity. This is possible because the procedure is parametric by nature." (Kijko and Graham, 1998, 1999). The tool that is fundamental to the *deductive* methods is the logic tree, which allows for the incorporation of different numbers, models, and parameter values, among others, by a simple "tree-node" structure. Although the logic tree is not discussed in detail here, as various subtleties complicate understanding it, employing the tool does influence the outcome because it allows for a considerable subjective representation of the weights or probabilities. A hazard curve is subsequently calculated for each path by employing the Total Probability Theorem.

As they were unsatisfied with the way non-quantitative data were handled, obscuring valuable catalogue data, Kijko and Graham (1998, 1999) resolved to reduce the "blurring effect" of bias on the parameter and resulting hazard. The procedure they developed is referred to as the "Parametric-historic" (P-H) procedure (Kijko and Graham, 1998, 1999), and this name is also used in the current research. The primary aim of developing the P-H procedure was to solve two particular deficiencies in the existing methods. These are the inability to employ the available data ''obscured'' in incomplete and uncertain catalogues, and overcoming the complication presented by the need to specify parameters. Many catalogues can be incomplete and uncertain, but the information they contain needs to be extracted, and tools are developed to solve the problem. Also, "a procedure that accepts the varying quality of different parts of the catalogue and does not require specification of seismic source zones would be an ideal tool for analysing and assessing the seismic hazard.'' (Kijko and Graham, 1998, 1999). The P-H procedure does require parameters to be specified; however, it is based on a model that assumes a Gutenberg–Richter FMSL and a Poisson process in time, which simplifies the setting of the parameters to the (b, λ, m_{max}) (the parameters considered in detail in the previous chapters). It is most important to note here that the P-H procedure is inherently data driven.

It might not be commonly known that the P-H procedure is capable of accommodating geological and geophysical information, the effect of active faults, predominant directivity of raptures, mixed information on known faults and less defined areas, different types of distances, and other information. This ability is enabled by the Bayesian formalism, because "...in each step, all the parameters are estimated by the maximum likelihood procedure." (Kijko and Graham, 1999). Therefore, the P-H procedure does not exclude the *deductive* paradigm and information — a thoroughly underappreciated attribute. In fact, the capacity of this procedure appears to be completely unknown and underutilised for reasons unknown but, probably, because the authors chose not to emphasise this outstanding feature, but only to allude to it with an obscure technical statement. Therefore, to promote this versatile feature, it is highlighted in the current study. However, the Bayesian formalism to incorporate information from relatively diverse types of data is not discussed here, as it is well represented in the literature. It is useful to note that the Bayesian update procedures are not difficult, although the Bayesian theory can be extremely complicated. A comprehensive comparison of the Cornell–McGuire and P-H procedures is provided in a recent study by Pavlenko (2016).

A description follows of only the data-driven (or catalogue-driven) part of the procedure, which comprises two steps. In the first, an area around the site of interest is isolated as the area where any earthquakes

produced could still influence the risk. For this area, the model parameters of the FMD and the activity rate parameter, which describes the rate at which earthquakes are produced in the area, are applicable to the area. In the second step, the focus is on the specific site, and the most applicable GMPE is implemented to calculate the distribution of the strong ground motion parameter, given the magnitude and distance that were selected.

In the first step, a catalogue of seismic events is used, including the prehistoric, geologically induced events (for the sole purpose of comparison with the estimation of m_{max}). The catalogue can be visualised conceptually, as in Figure 11. The historical part of the catalogue is considered the part that contains only extremely large events recorded before the existence of recording instrumentation such as seismographs. The magnitudes of the earthquakes are inferred from historical analysis of the effects of the earthquakes. Modified Mercalli intensity values (Modified Mercalli Intensity Scale) are used to determine and locate such earthquakes as accurately as possible; however, the events obtained are subject to high estimation error. The second part of the catalogue, recorded instrumentally, contains sections with decreasing levels of completeness, along with an increase in the sensitivity of the seismic network, which is attributed to improved instrumentation, expanded networks, and advanced analysis techniques, among others. The catalogue entries within a distance $R_L = \max\{r\}$ are isolated and used as the working catalogue. Representative estimates (in this instance, mostly MLEs) of the parameters (discussed in the previous sections) are obtained. Each parameter is introduced in an ideal case, where the representation of the underlying process is a stationary Poisson process, and the catalogue is completely homogeneous, with the LoC not being time dependent. Progressively, generalisations of estimation methods are introduced for the purpose of dealing with non-homogeneous, *incomplete* catalogues, which derive from a more complete and realistic representation of what practitioners of PSHA have to contend with. In addition to the need to include in the estimation methods the capability to deal with an *incomplete* instrumental catalogue, it is desirable and ideal to incorporate the historical part that is derived from Mercalli Intensity estimates from historical records (no instrumental records are available for this part). A solution to achieving such incorporation is proposed by Kijko and Sellevoll (1989). The fundamental assumption (which validates, invalidates, or justifies the use of an approximation) is the following hypothesis (*hypothesis 1): that, in historical times, only the strongest earthquakes in a time period would have been noted, with the rest remaining unnoted.* It cannot be overemphasised how important it is that the methodology of analysis of the extreme part of the catalogues is based on *hypothesis 1.*

The methodology of the P-H procedure of Kijko and Graham (1998, 1999) is presented formally as follows. Denote by $F(m)$ the upper bounded Gutenberg–Richter distribution. The notation used by Kijko and Sellevoll (1989) is convenient and compact (in addition, the formality of using capital letters for random variables is used); therefore, their notation is adopted here. Following the adoption by Kijko and Sellevoll (1989) of the representation of $F(m)$ used by Page (1968), Cosentino *et al.* (1977) derive

$$A_1 = e^{-\beta m_0},$$
$$A_2 = e^{-\beta m_{max}}, \tag{6.1}$$

$$A(x) = e^{-\beta x}.$$

Then, $F(m)$ is expressed as

$$F(x) = \frac{A_1 - A(x)}{A_1 - A_2}. \tag{6.2}$$

The objective here is to construct a joint likelihood function over the entire catalogue. Therefore, the innovative, but often neglected and underappreciated likelihood function for the historical part of the catalogue developed by Kijko and Sellevoll (1989) is implemented here. The likelihood function is constructed as follows.

It is customary practice to model seismic activity as a Poisson process with intensity λ, which results in the probability that an event will occur within a time t, given by

$$P[\#X \geq 1] = exp[-\lambda t] \tag{6.3}$$

To obtain the probability that an event will occur with a size lower than the specified threshold magnitude m_0, it is necessary to rescale λ to $\nu_0 = 1 - F(m_0)$, such

$$P[X \leq m_0] = exp[-\nu_0 t] \tag{6.4}$$

And, once again rescaling for a specific or arbitrary value of x, the desired likelihood becomes

$$l_0 = exp\left[-\nu_0 t\left(\frac{A_2 - A(x)}{A_2 - A_{10}}\right)\right] \tag{6.5}$$

where $A_{10} = \exp(-\beta x_0)$. This comprises the first part of the PDF likelihood function for the historical part of the catalogue. Astonishingly, this PDF captures the likelihood properties of observation in terms of both time and magnitude.

For the rest of the catalogue, the likelihood function is uncomplicated, requiring only consideration of the change in the magnitude of completeness m_i for each sub-catalogue. The relevant equations, already introduced, are repeated here, with that capturing parameter β, as follows

$$l_{i\beta} = \beta^{n_i} \exp\left(-\beta \sum_{j=1}^{n_i} X_{ij}\right) \Big/ (A_{1i} - A_2)^{n_i} \tag{6.7}$$

where $A_{1i} = \exp(-\beta m_i)$, and the likelihood function capturing the time of observation

$$\begin{aligned} l_{i\lambda} &= const \exp(-\nu_i T_i)(\nu_i T_i)^{n_i}, \\ \nu_i &= \lambda(1 - F(m_i)) \end{aligned} \tag{6.8}$$

denoting

$$l_i = l_{i\lambda} \cdot l_\beta. \tag{6.9}$$

The full joint likelihood function is

$$L(\beta, \lambda | X) = \prod_{i=0}^{s} L_i\,(\beta, \lambda | X_i) \tag{6.10}$$

This is the full likelihood function for the catalogue around the site of interest, which must be maximised to obtain an estimate of (λ, β). In most instances, the magnitude values in the catalogue are also uncertain, and it is common to assume that the error in magnitude values are distributed normally. In such instance, the combined density and distribution functions, respectively, of "apparent" magnitude following the GR law and, having a normally distributed error, becomes

$$f_M(m | m_c, m_{max}, \sigma_M) = f_M(m | m_c, m_{max}) C(m, \sigma_M), \tag{6.11}$$

$$F_M(m | m_c, m_{max}, \sigma_M) = F_M(m | m_c, m_{max}) D(m, \sigma_M), \tag{6.12}$$

$$C(m, \sigma_m) = \frac{\chi^2}{2}\left[\operatorname{erf}\left(\frac{m_{max} - m}{\sqrt{2}\sigma_M}\right) + \operatorname{erf}\left(\frac{m - m_{max}}{\sqrt{2}\sigma_M} - \chi\right)\right], \tag{6.13}$$

$$D(m,\sigma_M)=\frac{\left\{A_1\left[erf\left(\frac{m-m_c}{\sqrt{2}\sigma_M}+1\right)\right]+A_2\left[erf\left(\frac{m_{max}-m}{\sqrt{2}\sigma_M}-1\right)\right]-2C(m,\sigma_M)A(m)\right\}}{2[A(m)]}, \quad (6.14)$$

where σ_M is the standard deviation of earthquake magnitude determination, $A(m)=\exp(-\beta m)$, $A_1=\exp(-\beta m_c)$, $A_2=\exp(-\beta m_{max})$, erf(·) is the error function, with $\chi=\beta\sigma_M/\sqrt{2}$, being unbounded from both ends.

The functions are scaled as follows

$$\tilde{f}_M(m|m_c,m_{max},\sigma_M)=f_M(m|m_c,m_{max},\sigma_M)/[1-F_M(m_c|m_c,m_{max},\sigma_M)] \quad (6.15)$$

and

$$\tilde{F}_M(m|m_c,m_{max},\sigma_M)=\frac{F_M(m|m_c,m_{max},\sigma_M)-F_M(m_c|m_c,m_{max},\sigma_M)}{[1-F_M(m|m_c,m_{max},\sigma_M)]} \quad (6.16)$$

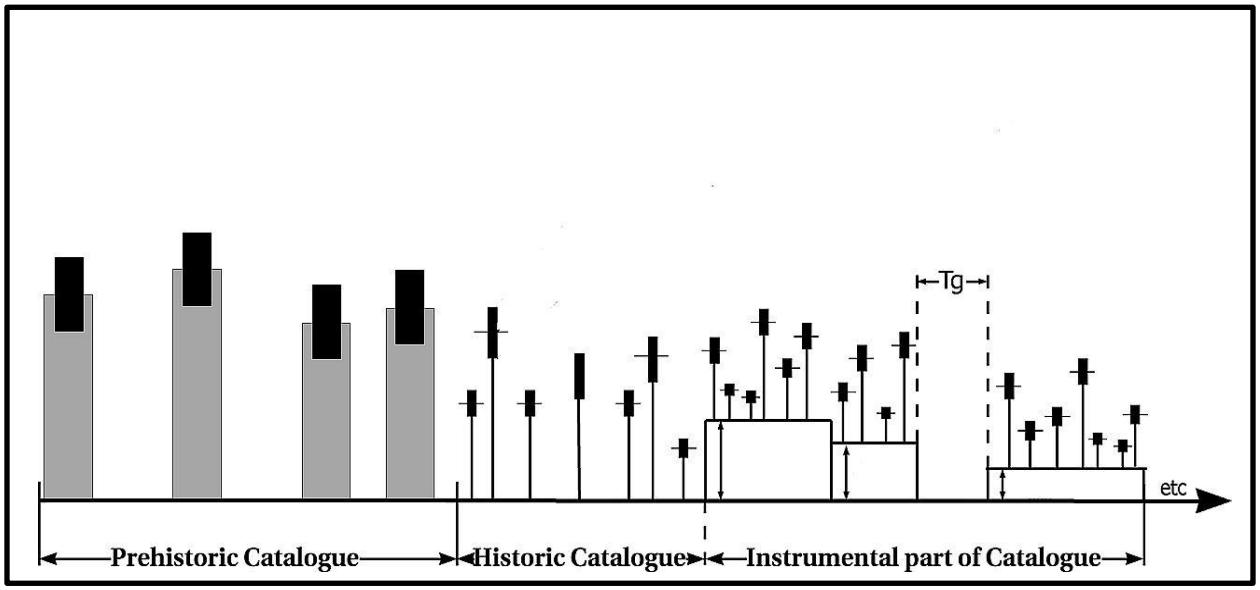


Figure 11. Conceptualisation of an earthquake catalogue including the prehistoric, historic, and instrumental parts of the catalogue, where T_g denotes gaps in observations (Figure from Smit *et al.*, 2019).

The "apparent" activity rate is

$$\tilde{\lambda}(m) = \lambda(m)\exp(\chi^2) \tag{6.17}$$

The likelihood functions have to be adjusted accordingly.

The extreme value distribution, the distribution of the strongest earthquake in magnitude range $<m_0, m_{max}>$ and within a time span *t* becomes

$$\tilde{f}_M^{max}(m|m_0, m_{max}, t) = \frac{\tilde{\lambda}_0 t \tilde{f}_M(m|m_0, m_{max}, \sigma_M) exp\left[-\tilde{\lambda}_0 t\{1 - \tilde{F}_M(m|m_0, m_{max}, t, \sigma_M)\}\right]}{1 - \exp(-\tilde{\lambda}_0 t)} \tag{6.18}$$

and the likelihood function for the extreme part of the catalogue takes the form

$$l_0 = const \prod_{j=1}^{n_0} \tilde{f}_M^{max}\left(m_{0j}\middle|m_0, m_{max}, \sigma_{M_{0j}}\right). \tag{6.19}$$

For the instrumental part of the catalogue, the likelihood functions take the form

$$l_{i\beta}(\beta) = \prod_{j=1}^{n_i} \tilde{f}_M\left(m_{ij}\middle|m_c^i, m_{max}, \sigma_{M_{ij}}\right), \tag{6.20}$$

and

$$l_{i\lambda}(\lambda, \beta) = const\left(\tilde{\lambda}_i t_i\right)^{n_i} \exp\left(-\tilde{\lambda}_i t_i\right), \tag{6.21}$$

where

$$\tilde{\lambda}_i = \tilde{\lambda}\left(m_c^i\right) = v_i \exp(\chi^2), \tag{6.22}$$

and $v_i = \lambda(1 - F(m_i))$, and, denoting the full likelihood function for a sub-catalogue by

$$l_i(\beta,\lambda) = l_{i\lambda}(\lambda,\beta) \cdot l_{i\beta}(\beta), \tag{6.23}$$

the full likelihood function to be maximised is

$$\mathcal{L}(\beta,\lambda) = \prod_{i=0}^{s} l_i\,(\lambda,\beta). \tag{6.24}$$

Having obtained the parameters, Kijko and Graham (1999) show that the strong ground motion parameters are distributed according to an exponential law. It must be noted (as Cor1968 did) that complex integration limits are being ignored. However, in most instances, exponential distribution has been found an excellent model of the distribution of ground motion at a particular site. In the current author's opinion, supported by previous research (Vermeulen, 2014), it is advisable not to use the parameter $\gamma = \frac{\beta}{c_2}$ or, if used, it should be done with caution. The maximum possible ground motion will be

$$\ln(a_{max}) = GMPE(r_{min}, m_{max}). \tag{6.25}$$

Given the uncertainties about the parameters, with the assumption that the parameters are Gaussian distributed, and that the parameters of the GMPE can be used to transform the distribution of magnitude and distance to strong ground motion, the standard deviation on $ln(a_{max})$ is

$$\sigma_{total} = \sqrt{\sigma^2_{\ln(a_{max})} + c_2^2\sigma_M^2 + \sigma_R^2(c_3 + c_4/r)^2}, \tag{6.26}$$

which allows evaluating $\ln(a_{max})$ at a certain confidence level.

The P-H procedure has been found most useful in generating the estimated spatial distribution of $\Theta = (\beta, \lambda, m_{max})$. This eliminates the need for subjective "zoning" when data are used. Certainly, it is often desirable to include extra information, which can be done simply by a Bayesian formalism.

Kijko and Graham (1998, 1999) list several studies on the application of a parametric-historic procedure, with probably the most well-known being that by Frankel (1995). The basic concept derives from determining area-characteristic parameters for a grid of points by employing statistical analysis. The most advanced and formal of these are those of Kijko and Graham (1998, 1999), with possible variations if applied rigorously. Methods usually advance continually, and the P-H procedure is mostly open to

extension. The methods mentioned in Vermeulen and Kijko (2017) and in Beirlant *et al.* (2017a, 2017b) can be used to estimate the maximum magnitude. Naturally, recent tools developed by Kijko *et al.* (2017) are utilised, as these represent the sequel to the methodology the P-H procedure is based on.

The P-H procedure has the advantage of employing nearly all the catalogue data in a mathematically rigorous fashion. However, it should be emphasised that the Bayesian formalism enables the incorporation of data that differ from the catalogue data. Furthermore, the Bayesian formalism can allocate as much weight to the other data as to the catalogue data. The P-H procedure allows the mathematics for using the catalogue data to be made formal and provides a way for the information from the catalogue to not be distorted by quasi-quantitative incorporation into the analysis in such a way that the effect of such quasi-quantitative information on the analysis cannot be traced. The Bayesian formalism is defined and briefly described in the Glossary. This is done because it is not widely known how easily the maximum likelihood procedure can be extended by a simple Bayesian technique without having to consider the full details of Bayesian statistics.

Alternatives to the P-H procedure have been proposed, but none as powerful for incorporating catalogue data. However, such procedures are worth mentioning here as a comparison with and possible extensions to the P-H procedure.

The approach of Wang (2011) is considered a combination between PSHA and determinative seismic hazard analysis (DSHA), as, for a given location, a hazard curve is constructed employing the relation between magnitude, distance, the GMPE, and variability. The function

$$m = g(R, GMPE, \sigma\epsilon), \tag{6.27}$$

is merely substituted into the GR scaling law

$$\frac{1}{\lambda} = e^{\alpha - \beta g(R, GMPE, \sigma\epsilon)}. \tag{6.28}$$

Subsequently, Wang (2011) determines the parameters from the catalogue but he does not have the tools that the P-H procedure provides. However, these tools can be used easily. If a grid of points were constructed determining α and β using the likelihood methods in the P-H procedure, this procedure is essentially no different from the P-H procedure. The method by Wang (2011) provides only another, different perspective by placing emphasis on a specific step in the procedure. It is clear that this leads to full deaggregation, and it is interesting to see here the actual log-linearity of the hazard curve for a given site.

Although Araya and Der Kiureghian (1988) claim to propose a new method for PSHA, their method is actually not new but presents PSHA in its most generalised mathematical formulation. However, their paper is considered a landmark, as they discuss the fact that increasingly refined variables and models can be introduced in this general formulation. This is simply the Total Probability Theorem formulated to

emphasise explicitly that any number of factors can be accommodated. These authors also keep uncertainty separate throughout the analysis. The difference in the detail of their methodology is that they transform the distribution of variables to a Gaussian form. Such a practice could be useful but is beyond the scope of this work.

The approach of Raschke (2014), referred to as the delta method, is strongly recommended to keep track of uncertainties (this is the same approach used by Araya and Der Kiureghian [1988]). Mulargia *et al.* (2017) criticise PSHA for not being verifiable, and an easy philosophical argument turns this into a problem related to the uncertainty of seismic hazard estimates and, particularly, the uncertainty of our knowledge about the uncertainty in hazard estimates. Given this situation, the certainty of our knowledge needs to be scrutinised, and our epistemic uncertainty about aleatory uncertainty must be handled conservatively. It is important to communicate epistemic uncertainty, whether it is incorporated in the uncertainty bounds of the hazard curve or as an additional figure. It is unfortunate that this is seldom done.

7. Final Summary

This thesis introduced (or revealed) Parameter Estimation in Probabilistic Seismic Hazard Analysis (PE-PSHA) as a field of study in its own right, which, at the same time, is also an essential part of any PSHA. PSHA was introduced, as done by the Cornell–McGuire procedure, and the parameters for the models used in the Cornell–McGuire procedure were discussed briefly. These essential parameters are the Gutenberg–Richter b-value, RoS, and m_{max}. Substantiation was presented for considering PE-PSHA, a necessary but neglected part of PSHA, as a distinct study field. Information related to this field of study is scattered throughout the relevant body of literature and is, herewith, brought together in view of it being considered an integrated, essential part of PSHA.

As the source of data, seismic catalogues are discussed briefly. Typically, an earthquake catalogue consists of a prehistoric part, a historical part, and an instrumental part, with subparts for which the minimum magnitude that can be detected reliably decreases with time as seismic instruments, networks, and processing methodologies become more advanced and, therefore, more sensitive. It is important to note that, ideally, all the data should be used despite it being easier to use only the instrumental part of a catalogue — a viewpoint that is reflected throughout the text.

An essential parameter that is not a model parameter in PSHA but has to be determined to use catalogue data efficiently and obtain accurate estimates of model parameters, is the LoC of each subpart of the earthquake catalogue. The LoC is the point above which all earthquakes are recorded. The relevant methods discussed are those by Stepp (1972), Tinti and Mulgaria (1985), Maximum Curvature Method of Wiemer and Wyss (2000), goodness of fit method of Wiemer and Wyss (2000), method of Cao and Gao (2002) using b-value stability, and an alternative by Woessner and Wiemer (2005), Marsan (2003) method using b-value stability, Entire Magnitude Range method of Woessner and Wiemer (2005) (including the current author's proposed slight modification), MBASS method by Amorèse (2007), and the method of Godano (2017) using the harmonic mean. A critical analysis, with some important considerations, is put

forward in rigorous logical and mathematical arguments. These include the proxy method of Mignan (2012) and Mignan *et al.* (2011), the MAXC, and the MBASS method. These critical considerations of estimation emphasise the fact that estimating the LoC is no simple matter.

Next, the estimation of the b-value was discussed. The discussion started with the classic Aki–Utsu estimator, which is the maximum likelihood estimator (also derivable by the Method of Moments), in relation to the GR law holding for arbitrarily large magnitudes. This estimator assumes a homogeneous, complete instrumental catalogue. An approximate variance for the estimator is given. Shi and Bolt (1982) note that the estimator is biased for a limited amount of data (a typical problem with estimators) and they propose a correction. In addition, Shi and Bolt (1982) provide an estimate of the variance that takes into account slow spatial and temporal variation. Ogata and Yashima (1986) use a Bayesian approach to arrive at a probability distribution for the estimated b-value. The first estimator that takes m_{max} into account is probably the modified Aki–Utsu estimator given by Page (1968). Marzocchi and Sandri (2003) note that estimators for the b-value, assuming arbitrarily large magnitudes, produce satisfactory results when $(m_{max} - m_{\min}) \geq 3$ is a guideline.

The next advancement in estimating the b-values is a correction for bias introduced because the data are not continuous but are grouped in intervals of approximately 0.01 for recent instrumental values and 0.1 for older data. Guttorp and Hopkins (1986), Bender (1983), and Tinti and Mulgaria (1987) all derive maximum likelihood estimators for the b-value that account for magnitude grouping.

The least squares approach was the first to be used by scientists because of its natural appeal to the linear trend of the GR plot. However, several authors have been critical for several reasons, which are justified rigorously (see references in section 4.3). Kijko (1994) proposes the minimisation of an optimal norm instead of the least squares method. However, the problem of asymmetry of the scatter of logarithmic transformation of data still poses a problem, as noted by Guttorp (1987), who derives a GLS estimator to solve the problem.

The above-mentioned estimators still have the shortcoming of assuming a single homogeneous catalogue with one LoC. Stepp (1972) was probably the first to provide a way to divide a catalogue into its different subparts, with different LoC values. Stepp (1972) used a least squares approach to estimate b-values for each subpart, which is not optimal. Weichert (1980) and Rosenblueth (1986) both derived maximum likelihood estimators that simultaneously estimate the b-value and the RoS, and their methods take magnitude grouping into account. Kijko and Sellevoll (1989) provide a simultaneous maximum likelihood estimator for the b-value and the RoS with the incorporation of historical data. Unfortunately, their estimator does not take magnitude grouping into account.

Kijko and Smit (2012) recognised the need for a simpler generalised estimator of the b-value for incomplete catalogues. One of the estimators they propose is an extension of the Aki–Utsu estimator, which has an extremely simple expression after the simplification of the expression. Ordaz and Giraldo (2017) derive a simultaneous maximum likelihood estimator for the b-value and RoS. However, as shown in the current study, this is only a special case of estimation by the method of Kijko and Sellevoll (1989), where no historical data are included.

A problem from the past that still needs attention is the variation of the b-value in space and time. This problem is addressed by Shi and Bolt (1982), Guttorp and Hopkins (1986), and Ogata and Katsura (1993).

PSHA is mostly conducted assuming a constant (stationary) RoS (λ), and, symbolically, $\lambda(t) \equiv K$, with K some constant with units of (seismic events)/(unit time). In other words, the variation in RoS with time is assumed to be constant. This is completely justified in many instances, with the variation in the RoS being negligible. However, in many other instances, the variation in RoS with time is large enough to have an effect on the seismic hazard. In this thesis, prominent models were discussed, such as renewal processes, Markov renewal processes, the Omori-Utsu formula for earthquake clustering (specifically with aftershocks), and the ETAS model.

For nonstationary Poissonian models, two reasonable approximations were given that could simplify the incorporation of time-varying RoS into seismic hazards, such as equations (3.82) and (3.83). Unfortunately, the history-dependent models make such elegant approximation much harder. For any time in the future, the RoS does not depend only on the time up to the present, but also on what happens from the present up to that point in the future. For approximation, the possibilities with their probabilities must be captured of what is to happen from the present until the point in time in future when the hazard is to be estimated. Up to the present, the Monte Carlo simulation has been almost the only way to analyse the hazard, if such models held in reality (and history dependence in seismicity is a reality). Polidoro *et al.* (2013) and Iervolino *et al.* (2014), in notable work, manage to incorporate aftershocks into an elegant integral formula in their sequence-PSHA. This formulation depends on the so-called first-event approximation, which means that hazard can be approximated by the first event occurring, of which Cornell and Winterstein (1988) present an explanation. Polidoro *et al.* (2013) describe instances where this theory will hold. No analytical approximations appear to exist in the literature where the overhead seismicity is not small enough for the first-event approximation to hold, in which case the Monte Carlo methods are resorted to. Under mild assumptions, where the RoS is not exactly Poisson and where the hazard depends on the time and magnitude of the last event (renewal model), Cornell and Winterstein (1988) propose a Weibull model that is a generalisation of the GR law. In this model, the RoS behaviour deviates somewhat from that of the Poissonian model and its ''memory-lessness'' property.

Kijko *et al.* (2016) present a most elegant and notable model that allows for variation in both the b-value and the RoS. This is done in a manner similar to Bayesian incorporation of the variability of the parameters that are related to epistemic uncertainty; however, in this instance, the variability accounts primarily for physical variability, although it can be used to account for epistemic variability as well. The variation in both the b-value and the RoS is assumed to follow a gamma distribution. The reason for this is the versatile properties of this distribution, i.e. the many forms of distribution it is able to capture. This model is easily applied in traditional PSHA.

Four promising estimators of the b-value were compared, which are those by Weichert (1980), Kijko and Sellevoll (1989), Kijko and Smit (2012), and Kijko and Smit (2016). A comparison was conducted with Monte Carlo simulation. Estimator (3.32) by Weichert (1980) appeared somewhat inferior in the cases modelled, and the performance of the others was quite similar. Estimator (3.47) by Kijko and Sellevoll (1989) and (4.61) by Kijko and Smit (2012) are preferred for their simplicity.

The parameter m_{max} plays a vital role in PSHA but has received relatively little research attention. In this thesis, the estimation of m_{max} from seismic catalogue, data was considered specifically. The focus was on finite (point) estimates, noting that giving the second moment or variation is essential. The following estimators and features were mentioned and discussed:

- Tate–Pisarenko estimator (4.2) (Pisarenko, 1991; Pisarenko *et al.*, 1996), which is the minimum variance unbiased estimator, and its elaborated form (4.11) given by Kijko (2004).
- The fiducial distribution (4.7), used by Pisarenko (1991).
- Cooke's (1979) estimators (4.14) and (4.15).
- Kijko–Sellevoll estimator derived from (4.14) for the GR law and the derived cases presented in Kijko and Graham (1998), Kijko (2004), and Kijko and Singh (2011). These estimators include equations (4.16), (4.18), (4.20), and (4.21).
- The Raschke (2012) estimator, which takes the form of the mean or median of the distribution (4.24).
- The Method of Moments estimator (4.25), with a discussion of its variance, as done by the current author in Vermeulen and Kijko (2017).
- Robson and Whitlock estimators (4.29) and (4.31).
- Estimator (4.38), and variations and essential properties from Robson and Whitlock (1964), Cooke (1979), and Kijko and Singh (2011).
- The estimator (4.40) from Cooke (1979, 1980) and Kijko and Sigh (2011), and estimator (4.39) from Cooke (1979, 1980).
- A note by Vermeulen and Kijko (2017) on the connection between estimators (4.40), (4.39), and the modern EVT.
- The method by Fraga Alves and Neves (2014) applied to the problem of estimating m_{max} , and investigated by Vermeulen and Kijko (2017).
- A maximum likelihood estimator based on the EVT (4.46) applied and investigated by Vermeulen and Kijko (2017).
- Estimator (4.49) by Beirlandt *et al.* (2017).
- Estimator (4.52) combined with (4.51) by Beirlandt *et al.* (2017).
- Gaussian kernel method by Kijko *et al.* (2001).
- The least squares procedure, the L_1 norm (4.57), and the optimised L_p norm (Kijko, 1994). The disadvantage of all of these error-norm-minimisation methods is that they do not function optimally for non-symmetrically distributed scatter.
- Application by the current author, as in Vermeulen and Kijko (2017), of the Guttorp (1987) generalised least squares procedure to simultaneously estimate the b-value and m_{max}. Although the application of the generalised least squares problem did not prove to be quite practical, it is worth further investigation.

Please refer to Chapter 4 for more detail in the discussion of these methods.

The contribution by the current author is presented in the final section of Chapter 4, where the theory of the Kijko–Sellevoll estimator, as developed by Haraala and Orosco (2016, 2018a, 2018b), is summarised

in a less theoretical form and in a manner that clarifies the scientific applicability of the theory and makes this outstanding and novel work accessible, even to researchers less conversant with the mathematics.

Seismic source characterisation was discussed and addressed in view of the Budnidz *et al.* (1997) guidelines for seismic zoning. The discussion takes the most critical view of normal zoning procedures because of the related oversimplification and non-optimal use of data. Alternative approaches are available but are not regularly employed, such as those of Frankel (1995) and Kijko and Graham (1999). The main problem with the zoning procedure is that it introduces a large amount of bias and epistemic uncertainty by relying directly, and solely, on expert opinion, which is left completely unaccounted for, or, at best, accounted for vaguely. The Bayesian formalism is suggested as a superior alternative, after employing the data optimally to delineate seismic source distribution.

Finally, alternative approaches to PSHA were discussed. These include the historical approach of Veneziano *et al.* (1984) (of which the Earthquake Engineering Research Institute [EERI] unfortunately refuses to release copies), the P-H procedure of Kijko and Graham (1998 and 1999), and that introduced by Wang (2011).

8. Conclusion

An earthquake magnitude catalogue for a given region (the part of interest here) consists of a series of earthquake event times, magnitude values, and (preferably) the estimated uncertainty of the magnitude values. The catalogue can be divided into two main parts, namely a historical part where only the largest events were recorded, and an instrumental part where networks of seismic stations record events quite accurately and, for practical purposes, record all events above the LoC within the region. With time, networks are upgraded and the LoC drops. This divides the instrumental part of the catalogue, once again, into smaller sub-catalogues with a unique LoC m_c^i. Gaps could be present in the instrumental part, with one particular gap common to many seismic catalogues being the period of the Second World War. The catalogue can be viewed as a marked point process, i.e. a point process where each point (or event) is assigned a value (magnitude). To fully characterise the marked point process, specification is needed for distribution of points on the real line (in time) and the distribution of the marking values (magnitude). In the current work, taking a perspective different from that of Reiter (1990), the process of PSHA is regarded as constructing a marked point process model for earthquake occurrences. This is traditionally done for a list of seismic source zones but can be done in an "area-characteristic" manner as well. Accordingly, from this perspective, the model construction of the marked point process is central to the analysis. Having chosen the Poisson process and a Gutenberg–Richter FMD, the problem is reduced to that of parameter estimation. For a complete PSHA, the ground motion has to be calculated for a given site; however, the perspective can be changed to view the calculation of ground motion probabilities as a "simple" transformation (the development of GMPEs is by no means simple but is viewed here as a given). This view is justified and useful, as it clearly places the focus on the primary task of seismologists related to an earthquake catalogue (input).

A typical PSHA is carried out using the assumption that earthquake occurrences can be modelled by a Poisson process in time and the bounded Gutenberg–Richter FMSL. Therefore, the essential parameters to be estimated for an earthquake hazard model are the Poissonian intensity (or seismic activity) rate λ, the Gutenberg–Richter b-value (alternatively the β-value), and the maximum possible earthquake magnitude m_{max}. This is denoted by a parameter vector $\mathbf{\Phi} = (\lambda, \beta, m_{max})$. Accordingly, the determination of $\mathbf{\Phi}$ is vital in a typical PSHA study and is the main focus of the current work. It should be noted that other models can certainly be developed for use in modelling the temporal distribution and frequency-magnitude scaling law, in which $\mathbf{\Phi}$ takes a different form. However, the Poisson process is the simplest "marked" point process model, with seismic activity λ, which, naturally, has an analogous parameter in more complex models. In addition, to a great degree, the bounded Gutenberg–Richter distribution is the most natural model to assume, for various reasons, e.g. it is simple and fulfils the maximum entropy criterion under exceptionally natural assumptions (Berrill and Davis, 1980). Other models also typically have an analogous value to β. A topic of intensive research is whether an upper bound m_{max} can be attributed to the frequency-magnitude law. There is an intuitive appeal to assign m_{max} to the range of possible magnitudes; however, strong arguments by early seismologists demonstrate that it is necessary to impose m_{max} for a realistic model (Kijko and Graham, 1998; and references therein). An upper bound is, indeed, necessary for any realistic model. The current author questions the validity of a sharp truncation in favour of a continuous decrease to zero from the left. The only difference is that the tail would take a shape that differs from the exponential model. However, if using the EVT to estimate m_{max} and the value of the shape parameter was different from -1, the implication is that the end point is not a truncation (Cooke, 1979). Because of the nature of catalogues, i.e. missing data at lower magnitude values, another parameter, the auxiliary parameter, is of importance and is estimated for different subparts of an earthquake catalogue. This is called the level of completeness m_c, defined as the magnitude level above which a hundred percent of all earthquakes are recorded. Below this level, some events are missed, typically at random. In time, m_c usually decreases as instrumental networks and processing techniques become available. This implies that for a single catalogue, the different parts each have their own LoC m_c^i. A typical catalogue is conceptualised in Figures 10 and 11.

To date, there is no accepted, robust method to estimate m_c , although various good methods based on sound arguments have been proposed. The first widely used method is that developed by Stepp (1972), and it is still used widely today. Subsequent methods have been developed, typically aimed at detecting the empirical distribution from linearity. Closely related to this are methods inspecting the stability of the estimated b-value, excluding consecutively larger magnitudes. The current author is critical of various workers not being conservative when using these methods. For example, in Wiemer and Wyss (2000), an accuracy measure is used in many of these methods, and m_c is chosen to be the point where a statistic reaches a certain level of accuracy. However, often, toward higher magnitudes, a clear monotonous trend to even greater accuracy is present, which the current author regards as an indication that the desired value has not been reached yet. This opinion appears to be reflected in the results obtained by Godano *et al.* (2017). The current author does not claim to have any rigorous argument to substantiate such criticism but, intuitively, it simply appears more sensible to him. However, admittedly, this remains a matter of personal opinion. The estimator proposed by Godano *et al.* (2017), based on the value of the harmonic mean for magnitudes over a given threshold, appears to be the most robust method.

Furthermore, the current author is critical, but, here, on a sound basis, of the popular Maximum Curvature Method. In fact, the current author has shown rigorously and mathematically that this method is flawed, and it does not serve the purpose it is intended for. The exception is a theoretical case where no curvature is present over any interval, but the trend changes at a point (i.e. infinite curvature at one point). Kijko and Smit (2017a) derived an estimator (with alternative variations using the Method of Moments and the maximum likelihood method) of the b-value that does not explicitly require the estimation of m_c. The current author considers this a breakthrough, and this method should replace the customary practice of the explicit estimation of m_c. The opinion of the current author is based on Aki (1995, pers. comm. with A. Kijko) being so dissatisfied with the sensitivity to this parameter that he did not even include it in his landmark book *Quantitative Seismology* (Aki and Richards, 2002). For the sake of completeness, the estimation of m_c as a parameter was discussed. Ideally, this can be viewed as data preparation and the actual parameter the estimation pertains to is defined earlier.

The β-value and the seismicity rate λ are interlinked closely and are, therefore, often estimated as a pair $\boldsymbol{\Psi} = (\beta, \lambda)$. Most often, for practical reasons, β is estimated first and subsequently λ. In some instances, however, they are estimated simultaneously. It must be emphasised that these two parameters are essential to the distributions defining the marked point process used to model seismicity, as they relate directly to parameters a and b in the GR relation. Gutenberg and Richter (1944) use a least squares regression on the cumulative distribution curve to determine the b-value, but the literature indicates that this method is criticised widely. It is not an adequate method for use because the assumption of normally, or even symmetrically distributed errors is not met. The Aki–Utsu estimator equation (1.9) is one of the first proposed methods and remains one of the most popular. It is effective for a single complete catalogue, for which the magnitude of completeness m_c is known. Numerous modifications or alternatives to the Aki–Utsu estimator have been proposed, taking into account grouping (binning) of magnitudes in regular intervals (Bender, 1983; Guttorp and Hopkins, 1986), instances where a maximum likelihood is imposed (Page, 1968), and instances where incomplete catalogues are used (Weichert, 1980; Kijko and Sellevoll, 1989; Kijko and Smit, 2012; Ordaz and Giraldo, 2017). The current author has shown that the equations given by Ordaz and Giraldo (2017) indicate a special case of the estimator by Kijko and Sellevoll (1989). Guttorp (1987) partially solved the problem of the use of least squares regression by introducing the generalised least squares regression. However, because of the transformation on frequency data, the applicability of the method remains questionable. Another alternative is the least absolute value regression, which is less sensitive to outliers, or even an optimised L_p norm proposed by Kijko (1994). The most popular methods are probably those of Weichert (1980) and Kijko and Sellevoll (2012) and, in the current study, they have been shown to provide quite similar results and, therefore, no distinction can be made as to which is superior or more appropriate.

Models other than the stationary point process have been proposed, and they appear quite plausible. One of the major implications is that $\lambda = \lambda(t)$ varies with time. Various explanations of the models have been presented; however, the only way to incorporate these models into hazard analysis is by considering λ as varying stochastically (Kijko *et al.*, 2017), or by employing one of the Monte Carlo simulations. In seismic hazard studies, the assumption of a stationary point process remains prominent, and the current

author foresees that the Kijko–Smit–Sellevoll–Bayes procedure will probably be of substantial value in the future.

The estimation of m_{max} is the final estimation of importance mentioned in the current study. Various methods are available, with none of them having been shown to be superior to the others. The estimators can be categorised into parametric estimators, semi- and non-parametric estimators, and EVT-based estimators. A minor, but possibly crucial aspect, is that, according to Cooke (1979), a sharply truncated distribution should have a shape parameter equal to $\gamma = -1$. However, this is often not the outcome when EVT-based methods are used. Nevertheless, there is no firm basis to reject slightly different shapes from $\gamma = -1$, i.e. this does not render these estimates invalid.

Zones have been introduced as a primitive/preliminary method of delineating the spatial distribution of seismic sources. Zones are usually assigned a constant $\mathbf{\Phi}$ based on geological, geophysical, and seismic catalogue data. This method, however, does not use seismic catalogue data optimally and relies heavily on arbitrary boundaries set by teams of experts (geoscientists and engineers).

The current author wishes to emphasise that the problem with an expert opinion is that it is highly subjective and has an epistemic uncertainty. Epistemic uncertainty in itself can become most problematic in that it is extremely hard to quantify, and the problem is compounded if the expert judgement were to quantify this. (The advanced topic of the handling of epistemic uncertainty is an active field of research [e.g., Khakzad, 2019; Knutti et al., 2019; Pedde et al. 2019; Randle et al., 2019] but is not applied in the field of parameter estimation in PSHA.) Bayesian formalisms have the ability to enable an expert opinion to supplement data with a lesser effect of subjective bias on epistemic uncertainty, as they allow the optimal use of data in a rigorous way before factoring in expert opinion. The concept of epistemic uncertainty is well summarised by McGuire (2004), and implementation of Bayesian formalism to the problem of estimating m_{max} is done by Cornell (1994). Although this is an early application of the Bayesian formalism, it is still not standard practice in PSHA.

Some researchers have realised the need to optimise the use of seismic catalogue data, such as Frankel (1995), who recognised this deficiency and attempted to solve it. Later, Kijko and Graham (1999) developed a refined method, which allows the incorporation of geological and geophysical data, as well as expert opinion. This method allows for the determination of *area-characteristic* parameters $\mathbf{\Phi}$, describing a "floating earthquake" scenario, and a type of deterministic *worst case scenario* estimation. Applied to a grid of points, an estimation of the spatial distribution of $\mathbf{\Phi}$ can be obtained, avoiding traditional arbitrarily delineated zones.

9. Further Research

Further research is required to find optimal procedures for estimating the LoC, as, as was pointed out in Chapter 2, shortcomings in the methods remain. It is also obvious that finding a limiting point, the point where curvature commences (or stops if viewed from the opposite direction) on a histogram, is not

possible to accomplish without uncertainty. However, it remains an open question of what such an optimal procedure to estimate the LoC might be.

Methods to estimate the b-value and RoS for incomplete catalogues simultaneously should be adapted for grouped magnitudes. Weichert (1980) does present a method for estimating both parameters, but the estimation of the RoS is based on a prior calculation of the b-value.

A link could possibly be established between the maximum likelihood estimators of Guttorp and Hopkins (1986) and Bender (1983). This would enable combining their research and findings, and not having to choose between them, assuming that they are unrelated (if, in fact, they are related as hypothesised). A link could also possibly be established between the maximum likelihood estimators of Weichert (1980), Rosenblueth (1986), and Kijko and Sellevoll (1989).

Further usable analytic/closed-form solutions can be derived for special instances of the Kijko–Sellevoll procedure. The ideal would be a closed-form solution for the instance where incomplete catalogues with a historical part can be incorporated. Closed-form solutions are preferable for the insight they afford into estimators, and for enabling further development of the estimators.

Methods for the addition of different types of data into the Kijko–Sellevoll method by employing the Bayesian formalism have not been established properly. Because of the maximum likelihood formulation of the procedure, it is clear that this is possible, but the procedure is not established yet.

Other alternatives to seismic zoning. The Bayesian formalism provides a method to incorporate expert opinion, as well as geological and geophysical data into a spatial distribution derived from earthquake catalogues.

Investigation of possible modifications to the GLS procedure that would guarantee a solution. Because this procedure appears to be a valid method of estimating the b-value, incorporating m_{max} into such calculation is definitely an aspect to consider. Modifications to the least squares procedure do exist, as do modifications to the GLS procedure.

Study on the effects of uncertain or varying additional parameters in minimum variance unbiased models. As a possible starting point, compound distribution has been proposed as a model that incorporates b-value uncertainty. A useful estimator could be a minimum variance unbiased estimator of these distribution parameters.

Detailed investigation of the EVT for point estimators for m_{max}. The assumption of a truncated exponential model is inconsistent with some of the assumptions (it is an irregular case and has to be treated with particular care). The main question here is why these inconsistent assumptions do not affect the variation in the results.

The extent of the applicability of expert opinion should be investigated. Neglecting this aspect would not be prudent and allocating it the main priority and using it as the main estimate to be modified by the data is not good practice, although this is the current practice. Bayesian procedures are the state-of-the-art, most applicable methodology to incorporate expert opinion, as it is easy to build a base model or estimate

and include expert opinion afterward using these formalisms. Furthermore, the subject of epistemic uncertainty in expert opinion has to be addressed in totality.

The Kijko–Sellevoll method prompts speculation on whether it would be possible to devise a technique to determine whether the catalogue data have been used optimally in a formal mathematical or probabilistic context. The aim of the Kijko–Sellevoll formalism appears to be to exhaust the information that is extractable from the data to estimate the parameters. However, questions that arise are, e.g. is there a theoretically maximum amount of information that can be extracted from earthquake catalogue data? Is this limiting maximum attainable?

Some excellent non-parametric estimators for m_{max} are discussed in Chapter 4. The question is to what extent non-parametric methods can be used in other areas of parameter estimation.

The problem of the convenient subdivision of the complete part of the catalogue has to be investigated. This implies that if the data from a part with a lower LoC were allocated to a part with a higher LoC, data that could have been used are discarded because of the higher LoC. Two notable publications dealing with this subject are by Stepp (1972) and Tinti and Mulgaria (1985); however, this is not sufficient, i.e. further research is required.

Kijko and Sellevoll (1989) assume, with sound argument, that the historical part of the catalogue can be modelled by an extreme value distribution. However, this assumption has not been tested.

The performance of the b-value estimators of Tinti and Mulgaria (1987), Guttorp (1986), and Bender (1983) for grouped magnitude data could be compared.

The incorporation of time-varying models into regular PSHA practice should be investigated; for instance, further development relevant to the work of Cornell and Winterstein (1988). These authors noted their intention to initiate such research; however, this appears not to have been done.

Techniques could be developed to incorporate and utilise the prehistoric part of earthquake catalogues.

10. References

Abramowitz, M. and Stegun, I.A. (1964). *Handbook of mathematical functions: with formulas, graphs, and mathematical tables*. Courier Corporation.

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 University 43: 237–239.

Aki, K. and Richards, P.G. (2002). *Quantitative seismology*.

Alamilla, J.L., Vai, R. and Esteva, L. (2014). Seismicity assessment using earthquake catalogues with uncertain and incomplete data: probabilistic formulation. *Journal of Seismology* 18(4): 715–729.

Amorèse, D. (2007). Applying a change-point detection method on frequency-magnitude distributions. *Bulletin of the Seismological Society of America* 97(5): 1742–1749.

Araya, R. and Der Kiureghian, A. (1988). *Seismic hazard analysis: improved models, uncertainties and sensitivities*. Earthquake Engineering Research Center, College of Engineering, University of California.

Atkinson, G.M. and Charlwood, R.G. (1983). Uncertainties in probabilistic seismic hazard assessment as a function of probability level: a case history for Vancouver, British Columbia. *Bulletin of the Seismological Society of America* 73(4): 1225–1241.

Awoyemi, M.O., Hammed, O.S., Shode, O.H., Olurin, O.T., Igboama, W.N. and Fatoba, J.O. (2017). Investigation of *b*-value variations in the African and parts of the Eurasian plates. *Science of Tsunami Hazards 36*(2).

Bak, P. and C. Tang (1989). Earthquakes as a self-organized critical phenomenon. *Journal of Geophysical Research: Solid Earth* 94(B11): 15635–15637.

Barani, S., Spallarossa, D., Bazzurro, P. and Eva, C. (2007). Sensitivity analysis of seismic hazard for Western Liguria (North Western Italy): a first attempt towards the understanding and quantification of hazard uncertainty. *Tectonophysics* 435(1-4): 13–35.

Beauval, C., Hainzl, S. and Scherbaum, F. (2006). Probabilistic seismic hazard estimation in low-seismicity regions considering non-Poissonian seismic occurrence. *Geophysical Journal International* 164(3): 543–550.

Beirlant, J., Fraga Alves, I. and Gomes, I. (2016). Tail fitting for truncated and non-truncated Pareto-type distributions. *Extremes* 19(3): 429–462.

Beirlant, J., Fraga Alves, I. and Reynkens, T. (2017). Fitting tails affected by truncation. *Electronic Journal of Statistics* 11(1): 2026–2065.

Beirlant, J., Kijko, A., Reynkens, T. and Einmahl, J.H. (2017). Estimating the maximum possible earthquake magnitude using extreme value methodology: the Groningen case. *Natural Hazards* 1–23

Bender, B. (1983). Maximum likelihood estimation of *b*-values for magnitude grouped data. *Bulletin of the Seismological Society of America* 73(3): 831–851.

Bender, B. (1986). Modeling source zone boundary uncertainty in seismic hazard analysis. *Bulletin of the Seismological Society of America* 76(2): 329–341.

Bender, B.K. and Perkins, D.M. 1993. Treatment of parameter uncertainty and variability for a single seismic hazard map. *Earthquake Spectra*, 9(2): 165–195.

Benjamin, J. R. (1968). Probabilistic models for seismic force design. *Journal of the Structural Division* 94(5): 1175–1196.

Ben-Zion, Y. (2008). Collective behavior of earthquakes and faults: continuum-discrete transitions, progressive evolutionary changes, and different dynamic regimes. *Reviews of Geophysics*, 46(4).

Berrill, J. and R. Davis (1980). Maximum entropy and the magnitude distribution. *Bulletin of the Seismological Society of America* 70(5): 1823–1831.

Budnitz, R.J., Apostolakis, G. and Boore, D.M. (1997). *Recommendations for probabilistic seismic hazard analysis: guidance on uncertainty and use of experts* (No. NUREG/CR-6372-Vol. 1; UCRL-ID-122160). Nuclear Regulatory Commission, Washington, DC (United States). Div. of Engineering Technology; Lawrence Livermore National Lab., CA (United States); Electric Power Research Inst., Palo Alto, CA (United States); USDOE, Washington, DC (United States).

Bufe, C. G. and D. M. Perkins (2005). Evidence for a global seismic-moment release sequence. *Bulletin of the Seismological Society of America*, 95(3): 833–843.

Berger, J.O. and Smith, L.A. (2019). On the statistical formalism of uncertainty quantification. *Annual Review of Statistics and its Application*.

Campbell, K. W. (1982). Bayesian analysis of extreme earthquake occurrences. Part I. Probabilistic hazard model. *Bulletin of the Seismological Society of America* 72(5): 1689–1705.

Campbell, K. W. (1983). Bayesian analysis of extreme earthquake occurrences. Part II. Application to the San Jacinto fault zone of Southern California. *Bulletin of the Seismological Society of America* 73(4): 1099–1115.

Cao, A. and S. S. Gao (2002). Temporal variation of seismic *b*-values beneath northeastern Japan island arc. *Geophysical Research Letters* 29(9): 48-1–48-3

Chan, C.H., Wu, Y.M., Tseng, T.L., Lin, T.L. and Chen, C.C. (2012). Spatial and temporal evolution of *b*-values before large earthquakes in Taiwan. *Tectonophysics* 532: 215–222.

Coles, S. (2001). *An introduction to statistical modeling of extreme values*. London: Springer.

Cooke, P. (1979). Statistical inference for bounds of random variables. *Biometrika* 66(2): 367–374.

Cooke, P. (1980). Optimal linear estimation of bounds of random variables. *Biometrika* 67(1): 257–258.

Cornell, C. A. (1968). Engineering seismic risk analysis. *Bulletin of the Seismological Society of America* 58(5): 1583–1606.

Cornell, C. and Winterstein, S.R. (1988). Temporal and magnitude dependence in earthquake recurrence models. *Bulletin of the Seismological Society of America*, *78*(4): 1522–1537.

Cornell, C.A. (1994). Statistical analysis of maximum magnitudes. *The earthquakes of stable continental regions* 1: 5-1.

Cosentino, P., Ficarra, V. and Luzio, D. (1977). Truncated exponential frequency-magnitude relationship in earthquake statistics. *Bulletin of the Seismological Society of America*, 67(6): 1615–1623.

Cramer, C.H., Petersen, M.D. and Reichle, M.S. (1996). A Monte Carlo approach in estimating uncertainty for a seismic hazard assessment of Los Angeles, Ventura, and Orange Counties, California. *Bulletin of the Seismological Society of America*, 86(6): 1681–1691.

Daley, D.J. and Vere-Jones, D. (2007). *An introduction to the theory of point processes: volume II: General theory and structure*. Springer Science & Business Media.

de Arcangelis, L., Godano, C., Grasso, J.R. and Lippiello, E. 2016. Statistical physics approach to earthquake occurrence and forecasting. *Physics Reports*, 628: 1–91.

De Haan, L. and Ferreira, A. (2007). *Extreme value theory: an introduction*. Springer Science & Business Media.

DeGroot, M.H. (2005). *Optimal statistical decisions* (Vol. 82). John Wiley & Sons.

Dixit, U. J. and Nasiri, P.N. (2008). Estimation of parameters of a right truncated exponential distribution. *Statistical Papers*. 49(2): 225–236.

Eadie, W.T., Drijard, D. and James, F.E. (1982). *Statistical methods in experimental physics*. Amsterdam: North-Holland.

Enescu, B., Mori, J. and Miyazawa, M., 2007. Quantifying early aftershock activity of the 2004 mid-Niigata Prefecture earthquake (Mw6. 6). *Journal of Geophysical Research: Solid Earth*, 112(B4).

Fisher, R.A. and Tippett, L.H.C. (1928). Limiting forms of the frequency distribution of the largest or smallest member of a sample. In *Mathematical Proceedings of the Cambridge Philosophical Society* 24(2): 180–190.

Frankel, A. (1995). Mapping seismic hazard in the central and eastern United States. *Seismological Research Letters* 66(4): 8–21.

Fraga Alves, I.F. and Neves, C. (2014). Estimation of the finite right endpoint in the Gumbel domain. *Statistica Sinica* 1811–1835.

Fréchet, M. (1927). Sur la loi de probabilité de l'écart maximum. *Annales de la Société Polonaise de Mathématique* 6: 93–116.

Gibowicz, S.J. and Kijko, A. (1994). *An introduction to mining seismology* (Vol. 55). Elsevier.

García-Hernández, R., D'Auria, L., Barrancos, J., Padilla, G.D. (2019). On the functional expression of frequency-magnitude distribution: a comprehensive statistical examination. *Bulletin of the Seismological Society of America* 109 (1): 482–486. Doi: 10.1785/012180197

Gnedenko, B. (1943). Sur la distribution limite du terme maximum d'une série aléatoire. *Annals of Mathematics* 423–453.

Godano, C. (2017). A new method for the estimation of the completeness magnitude. *Physics of the Earth and Planetary Interiors* 263: 7–11.

Godano, C., Lippiello, E. and de Arcangelis, L. (2014). Variability of the *b* value in the Gutenberg–Richter distribution. *Geophysical Journal International* 199(3): 1765–1771

Goldfinger, C., Ikeda, Y., Yeats, R.S. and Ren, J. (2013). Superquakes and supercycles. *Seismological Research Letters* 84(1): 24–32.

Gutenberg, B. and Richter, C.F. (1944). Frequency of earthquakes in California. *Bulletin of the Seismological Society of America* 34(4): 185–188.

Guttorp, P. (1987). On least-squares estimation of *b* values. *Bulletin of the Seismological Society of America* 77(6): 2115–2124.

Guttorp, P. and D. Hopkins (1986). On estimating varying *b* values. *Bulletin of the Seismological Society of America* 76(3): 889–895.

Helmstetter, A., Kagan, Y.Y. and Jackson, D.D. (2006). Comparison of short-term and time-independent earthquake forecast models for southern California. *Bulletin of the Seismological Society of America*, 96(1): 90–106.

Haraala, M., Orosco, L. (2016). Analysis of the Gutenberg-Richter *b*-value and the m_{max}. Part I: Exact solution of the Kijko–Sellevoll estimator m_{max}. *Cuandos de Ingeneria. Nuova Serie* (Notes in Engineering. New Series). Academic Publications. Faculty of Engineering. Catholic University of Salta. Vol 9: 51–77.

Haraala, M., Orosco, L. (2018a). *Generalized proofs of the Kijko–Sellevoll functions*. Original Manuscript.

Haraala, M., Orosco, L. (2018b). *Ramanujan's Expansion for the real valued Harmonic Numbers*. Original Manuscript.

Holschneider, M., Zöller, G. and Hainzl, S. (2011). Estimation of the maximum possible magnitude in the framework of a doubly truncated Gutenberg–Richter model. *Bulletin of the Seismological Society of America*, 101(4): 1649–1659.

Huang, Y., Zhou, S. and Zhuang., J., (2016). Numerical tests on catalog-based methods to estimate magnitude of completeness. *Chinese Journal of Geophysics*, 59(3): 266–275.

Hutton, K., Woessner, J. and Hauksson, E. (2010). Earthquake monitoring in southern California for seventy-seven years (1932–2008). *Bulletin of the Seismological Society of America*, 100(2): 423–446

Iervolino, I., Giorgio, M. and Chioccarelli, E. (2014). Closed-form aftershock reliability of damage-cumulating elastic-perfectly-plastic systems. *Earthquake Engineering & Structural Dynamics*, 43(4): 613–625.

Ishimoto, M., and K. Iida (1939). Observations sur les séismes enregistrés par le micro-séismographe construit dernièrement (1). *Bulletin of the Earthquake Research Institute University of* Tokyo 17: 443–478 (in Japanese with French abstract).

Isacks, B. and Oliver, J. (1964). Seismic waves with frequencies from 1 to 100 cycles per second recorded in a deep mine in northern New Jersey. *Bulletin of the Seismological Society of America* 54 (6A):1941–1979.

Iwata, T. (2012). Revisiting the global detection capability of earthquakes during the period immediately after a large earthquake: considering the influence of intermediate-depth and deep earthquakes. *Research in Geophysics* 2(1): 4.

Jacob, K., Armbruster, J., Barstow, N. and Horton, S. (1994). Probabilistic ground motion estimates for New York: comparison with design ground motions in national and local codes. In *Proceedings of Fifth US National Conference on Earthquake Engineering, Chicago* (Vol. 3: 119–128).

Joshi, G.C. and Sharma, M.L. (2008). Uncertainties in the estimation of M_{max}. *Journal of Earth System Science*, 117(2): 671–682.

Kagan, Y. Y., Schoenberg, F. (2001). Estimation of the upper cutoff parameter for the tapered Pareto distribution. *Journal of Applied Probability*, 38: 158-175

Kagan, Y.Y., (2004). Short-term properties of earthquake catalogs and models of earthquake source. *Bulletin of the Seismological Society of America*, 94(4): 1207–1228.

Kagan, Y. Y. (2002a). Seismic moment distribution revisited: I. Statistical results. *Geophysical Journal International* 148(3): 520–541.

Kagan, Y. Y. (2002b). *Seismic moment distribution revisited: II.* Moment conservation principle. *Geophysical Journal International* 149(3): 731–754.

Khakzad, N., 2019. System safety assessment under epistemic uncertainty: Using imprecise probabilities in Bayesian network. *Safety Science* 116: 149–160.

Kijko, A. (1994). Seismological outliers: L_1 or adaptive L_p norm application. *Bulletin of the Seismological Society of America* 84(2): 473–477.

Kijko, A. (2004). Estimation of the maximum earthquake magnitude, m_{max}. *Pure and Applied Geophysics* 161(8): 1655–1681.

Kijko, A. (2011). Seismic hazard. *Encyclopedia of Solid Earth Geophysics*, Springer: 1107-–121.

Kijko, A. and G. Graham (1998). Parametric-historic procedure for probabilistic seismic hazard analysis Part I: Estimation of maximum regional magnitude m_{max}. *Pure and Applied Geophysics* 152(3): 413–442.

Kijko, A. and G. Graham (1999). Parametric-historic procedure for probabilistic seismic hazard analysis Part II: Assessment of seismic hazard at specified site. *Pure and Applied Geophysics* 154(1): 1–22.

Kijko, A., Lasocki, S. and Graham, G. (2001). Non-parametric seismic hazard in mines. *Pure and Applied Geophysics*, 158(9-10):1655–1675.

Kijko, A. and M.A. Sellevoll (1989). Estimation of earthquake hazard parameters from incomplete data files. Part I. Utilization of extreme and complete catalogs with different threshold magnitudes. *Bulletin of the Seismological Society of America* 79(3): 645–654.

Kijko, A. and M. A. Sellevoll (1992). Estimation of earthquake hazard parameters from incomplete data files. Part II. Incorporation of magnitude heterogeneity. *Bulletin of the Seismological Society of America* 82(1): 120–134.

Kijko, A. and M. Singh (2011). Statistical tools for maximum possible earthquake magnitude estimation. *Acta Geophysica* 59(4): 674–700.

Kijko, A. and A. Smit (2012). Extension of the Aki–Utsu *b*-value estimator for incomplete catalogs. *Bulletin of the Seismological Society of America*. 102(3): 1283–1287.

Kijko, A. and A. Smit (2017). Estimation of the Frequency–Magnitude Gutenberg-Richter *b*-Value without Making Assumptions on Levels of Completeness. *Seismological Research Letters* 88(2A): 311–318.

Kijko, A. (2017). *Three More b-Value Maximum Likelihood Estimators Calculated from Incomplete Catalogs*. Unpublished Manuscript. U.P. Natural Hazard Centre. University of Pretoria.

Kijko, A., Smit, A. and Sellevoll, M.A. (2016). Estimation of earthquake hazard parameters from incomplete data files. Part III. Incorporation of uncertainty of earthquake-occurrence model. *Bulletin of the Seismological Society of America*, *106*(3): 1210–1222.

Knutti, R., Baumberger, C. and Hadorn, G.H. (2019). Uncertainty Quantification Using Multiple Models—Prospects and Challenges. In *Computer Simulation Validation* (pp. 835–855). Springer, Cham.

Kramer, S. L. (1996). *Geotechnical Earthquake Engineering*. Prentice Hall. Upper Saddle River, NJ.

Lasocki, S. and Urban, P. (2011). Bias, variance and computational properties of Kijko's estimators of the upper limit of magnitude distribution, M_{max}. *Acta Geophysica*, 59(4): 659–673.

Lanzante, J. R. (1996). Resistant, robust and non-parametric techniques for the analysis of climate data: theory and examples, including applications to historical radiosonde station data. *International Journal of Climatology* 16(11): 1197–1226.

Lay, T. and T. C. Wallace (1995). *Modern Global Seismology*, Academic Press, San Diego.

Lee, W. and D. Brillinger (1979). On Chinese earthquake history—an attempt to model an incomplete data set by point process analysis. In *Earthquake Prediction and Seismicity Patterns*: 1229–1257. Springer

Leptokaropoulos, K.M., Adamaki, A.K., Roberts, R.G., Gkarlaouni, C.G. and Paradisopoulou, P.M. (2018). Impact of magnitude uncertainties on seismic catalogue properties. *Geophysical Journal International*, 213(2): 940–951.

Lippiello, E., Bottiglieri, M., Godano, C. and de Arcangelis, L. (2007). Dynamical scaling and generalized Omori law. *Geophysical Research Letters*, *34*(23).

Lippiello, E., Godano, C. and de Arcangelis, L. (2012). The earthquake magnitude is influenced by previous seismicity. *Geophysical Research Letters*, 39(5).

Lippiello, E., Cirillo, A., Godano, G., Papadimitriou, E. and Karakostas, V., 2016. Real-time forecast of aftershocks from a single seismic station signal. *Geophysical Research Letters*, 43(12): 6252–6258.

Marsan, D. (2003). Triggering of seismicity at short timescales following Californian earthquakes. *Journal of Geophysical Research: Solid Earth* 108(B5).

Marzocchi, W. and L. Sandri (2003). A review and new insights on the estimation of the b-value and its uncertainty. *Annals of Geophysics* 46(6).

Matthews, M.V., Ellsworth, W.L. and Reasenberg, P.A. (2002). A Brownian model for recurrent earthquakes. *Bulletin of the Seismological Society of America*, 92(6): 2233–2250.

McGuire, R.K. (1993). Computations of seismic hazard. *Annals of Geophysics*, 36(3-4).

McGuire, R.K. (2004). *Seismic hazard and risk analysis*. Earthquake Engineering Research Institute.

Mignan, A. (2012). Functional shape of the earthquake frequency-magnitude distribution and completeness magnitude. *Journal of Geophysical Research: Solid Earth* 117(B8).

Mignan, A. and C. C. Chen (2016). The spatial scale of detected seismicity. *Pure and Applied Geophysics* 173(1): 117–124.

Mignan, A., Werner, M.J., Wiemer, S., Chen, C.C. and Wu, Y.M. (2011). Bayesian estimation of the spatially varying completeness magnitude of earthquake catalogs. *Bulletin of the Seismological Society of America*, 101(3): 1371–1385.

Mignan, A. and J. Woessner (2012). Estimating the magnitude of completeness for earthquake catalogs. *Community Online Resource for Statistical Seismicity Analysis.*

McGuire, R.K., 2008. Probabilistic seismic hazard analysis: Early history. *Earthquake Engineering & Structural Dynamics*, *37*(3): 329–338.

Mogi, K. (1962). Study of elastic shocks caused by the fracture of heterogeneous materials and its relation to earthquake phenomena. *Bulletin of the Earthquake Research Institute University of Tokyo* 40: 125–173.

Molchan, G., Keilis-Borok, V.I., and Vilkovich, G.V. (1970). Seismicity and principal seismic effects. *Geophysical Journal International* 21(3): 323–335.

Molkenthin, C., Scherbaum, F., Griewank, A., Kuehn, N., Stafford, P.J. and Leovey, H. (2015). Sensitivity of probabilistic seismic hazard obtained by algorithmic differentiation: a feasibility study. *Bulletin of the Seismological Society of America*, 105(3): 1810–1822.

Molkenthin, C., Scherbaum, F., Griewank, A., Leovey, H., Kucherenko, S. and Cotton, F., 2017. Derivative-based global sensitivity analysis: upper bounding of sensitivities in seismic-hazard assessment using automatic differentiation. *Bulletin of the Seismological Society of America*, *107*(2): 984–1004.

Mulargia, F., Stark, P.B. and Geller, R.J. (2017). Why is probabilistic seismic hazard analysis (PSHA) still used? *Physics of the Earth and Planetary Interiors*, *264*: 63–75.

Nava, F.A., Márquez-Ramírez, V.H., Zúñiga, F.R. and Lomnitz, C. (2017). Gutenberg–Richter b-value determination and large-magnitudes sampling. *Natural Hazards*, *87*(1): 1–11.

Nemati, M. (2015). Intermediate-term variations in 200 years seismicity at the north of Iran. *Journal of Seismology* 19(2): 585–605.

Ogata, Y. (1988). Statistical models for earthquake occurrences and residual analysis for point processes. *Journal of the American Statistical Association*, 83(401): 9027.

Ogata, Y. and Katsura, K. (1993). Analysis of temporal and spatial heterogeneity of magnitude frequency distribution inferred from earthquake catalogues. *Geophysical Journal International* 113(3): 727–738.

Ogata, Y. and Yamashina, K.I. (1986). Unbiased estimate for *b*-value of magnitude frequency. *Journal of Physics of the Earth* 34(2): 187–194.

Omi, T., Ogata, Y., Hirata, Y. and Aihara, K. (2013). Forecasting large aftershocks within one day after the main shock. *Scientific Reports* 3: 2218.

Ominetti, G. (2013). *Are great earthquakes clustered?* Scor Papers #27. Last accessed July 2019 <https://pdfs.semanticscholar.org/d21a/e81befc6000db5f8ac9c9dd51b630bfa3ce9.pdf>

Omori, F. (1894). On the after-shocks of earthquakes. *Journal of the Faculty of Science, Imperial University of Tokyo.*7: 111–200.

Ordaz, M. and Giraldo, S. (2018). Joint Maximum Likelihood Estimators for Gutenberg–Richter Parameters λ_0 and β Using Subcatalogs. *Earthquake Spectra* 34(1): 301–312.

Page, R. (1968). Aftershocks and microaftershocks of the great Alaska earthquake of 1964. *Bulletin of the Seismological Society of America* 58(3): 1131–1168.

Pandey, M.D. and Van Noortwijk, J.M. (2004) Gamma process model for time-dependent structural analysis. In *Proceedings of the Second International Conference on Bridge Maintenance, Safety and Management (IABMAS)*

PEER (2018). Last viewed 12 December 2018. < https://apps.peer.berkeley.ed >

Pedde, S., Kok, K., Onigkeit, J., Brown, C., Holman, I. and Harrison, P.A. (2019). Bridging uncertainty concepts across narratives and simulations in environmental scenarios. *Regional Environmental Change*, *19*(3): 655–666.

Pavlenko, V.A. 2015. Effect of alternative distributions of ground motion variability on results of probabilistic seismic hazard analysis. *Natural Hazards*, *78*(3): 1917–1930.

Pavlenko, V. 2016. Ground motion variability and its effect on the Probabilistic Seismic Hazard Analysis. PhD Thesis. University of Pretoria, 2016, pp. 83.

Peng, Z., Vidale, J.E., Ishii, M. and Helmstetter, A. (2007). Seismicity rate immediately before and after main shock rupture from high-frequency waveforms in Japan. *Journal of Geophysical Research: Solid Earth*, 112(B3).

Pisarenko, V.F. (1991). Statistical evaluation of maximum possible earthquakes. *Phys. Solid Earth* 27(9): 757–763.

Pisarenko, V.F., Lyubushin, A.A., Lysenko, V.B. and Golubeva, T.V. (1996). Statistical estimation of seismic hazard parameters: maximum possible magnitude and related parameters. *Bulletin of the Seismological Society of America* 86(3): 691–700.

Pisarenko, V., Sornette, D., Rodkin, M. (2003). Deviations of the distributions of seismic energies from the Gutenberg–Richter law. *arXiv preprint physics/0312020*.

Pisarenko, V. F., Sornette, D. (2004). Statistical detection and characterization of a deviation from the Gutenberg–Richter distribution above magnitude 8. *Pure and Applied Geophysics*, *161*(4): 839–864.

Pisarenko, V., Sornette, D. (2006). New statistic for financial return distributions: power-law or exponential? *Physical and Statistical Mechanics and its Applications*, 366: 387–400.

Pisarenko, V.F., Sornette, A., Sornette, D. and Rodkin, M.V. (2014). Characterization of the tail of the distribution of earthquake magnitudes by combining the GEV and GPD descriptions of extreme value theory. *Pure and Applied Geophysics*, 171(8): 1599–1624.

Pisarenko, V. and M. Rodkin (2017). The estimation of probability of extreme events for small samples. *Pure and Applied Geophysics* 174(4): 1547–1560.

Polidoro, B., Iervolino, I., Giorgio, M. and Chioccarelli, E. (2013, June). Models and issues in history-dependent mainshock hazard. In *Proceedings of 11th Conference on Structural Safety and Reliability, ICOSSAR* (Vol. 13).

Power, M., Chiou, B., Abrahamson, N., Bozorgnia, Y., Shantz, T. and Roblee, C. (2008). An overview of the NGA project. *Earthquake Spectra*, 24(1): 3–21.

Quenouille, M.H. (1956). Notes on bias in estimation. *Biometrika*, 43(3,4): 353–360.

Rabinovitch, A., Frid, V. and Bahat, D. (2001). Gutenberg–Richter-type relation for laboratory fracture-induced electromagnetic radiation. *Physical Review E*, *65*(1): 011401.

Rabinowitz, N. and Steinberg, D.M. (1991). Seismic hazard sensitivity analysis: a multi-parameter approach. *Bulletin of the Seismological Society of America*, 81(3): 796–817.

Rabinowitz, N., Steinberg, D.M. and Leonard, G. (1998). Logic trees, sensitivity analyses, and data reduction in probabilistic seismic hazard assessment. *Earthquake Spectra* 14(1): 189–201.

Randle, C.H., Bond, C.E., Lark, R.M. and Monaghan, A.A. (2019). Uncertainty in geological interpretations: effectiveness of expert elicitations. *Geosphere*, 15(1): 108–118.

Raschke, M. (2012). Inference for the truncated exponential distribution. *Stochastic Environmental Research and Risk Assessment* 26(1): 127–138.

Raschke, M. (2014). Estimation of uncertainties in probabilistic seismic hazard analysis. In *Proceedings of the 12th International Probabilistic Workshop*.

Raschke, M. (2015). Modeling of magnitude distributions by the generalized truncated exponential distribution. *Journal of Seismology* 19(1): 265–271.

Reasenberg, P.A. and Jones, L.M. (1989). Earthquake hazard after a mainshock in California. *Science*, 243(4895): 1173–1176.

Ringdal, F. (1975). On the estimation of seismic detection thresholds. *Bulletin of the Seismological Society of America* 65(6): 1631–1642.

Robson, D. and Whitlock, J. (1964). Estimation of a truncation point. *Biometrika* 51(1/2): 33–39.

Rohmer, J., Douglas, J., Bertil, D., Monfort, D. and Sedan, O. (2014). Weighing the importance of model uncertainty against parameter uncertainty in earthquake loss assessments. *Soil Dynamics and Earthquake Engineering*, 58: 1–9.

Rosenblueth, E. (1986). Use of statistical data in assessing local seismicity. *Earthquake Engineering & Structural Dynamics* 14(3): 325–337.

Rydelek, P. A. and I. S. Sacks (1989). Testing the completeness of earthquake catalogues and the hypothesis of self-similarity. *Nature* 337(6204): 251–253.

Sandri, L. and W. Marzocchi (2007). A technical note on the bias in the estimation of the *b*-value and its uncertainty through the least squares technique. *Annals of Geophysics* 50(3).

Scholz, C. (1968). The frequency-magnitude relation of microfracturing in rock and its relation to earthquakes. *Bulletin of the Seismological Society of America* 58(1): 399–415.

Shi, Y. and B. A. Bolt (1982). The standard error of the magnitude-frequency *b* value. *Bulletin of the Seismological Society of America* 72(5): 1677–1687.

Silverman, B. W. (2018). *Density estimation for statistics and data analysis*, Routledge.

Smit, A., Stein, A. and Kijko, A. (2019). Bayesian inference in natural hazard analysis for incomplete and uncertain data. *Environmetrics*, p. e2566.

Sornette, A. and Sornette, D. (1989). Self-organized criticality and earthquakes. *Europhysics Letters* 9(3): 197.

Stepp, J. (1972). Analysis of completeness of the earthquake sample in the Puget Sound area and its effect on statistical estimates of earthquake hazard. In *Proc. of the 1st International Conference on Microzonation*, Seattle.

Suzuki, Z. (1958). A statistical study on the occurrence of small earthquakes. Chapter 3, *Sci Rept. Tohoku Univ. Ser. 5, Geophys.,* 10: 15–27. In Page, R. (1968). Aftershocks and microaftershocks of the great Alaska earthquake of 1964. *Bulletin of the Seismological Society of America,* 58(3): 1131–1168.

Tate, R.F. (1959). Unbiased estimation: functions of location and scale parameters, *Annals of Mathematical Statistics* 30: 341–366.

Thenhaus, P. C. (1983). *Summary of workshops concerning regional seismic source zones of parts of the conterminous United States convened by the US Geological Survey 1979-1980*, Golden, Colorado. Unknown.

Tinti, S. and Mulargia, F. (1985). Completeness analysis of a seismic catalog. *Annales Geophysicae* 3(3): 407–414.

Tinti, S. and F. Mulargia (1987). Confidence intervals of *b* values for grouped magnitudes. *Bulletin of the Seismological Society of America* 77(6): 2125–2134.

Turcotte, D.L., Newman, W.I. and Shcherbakov, R., (2003). Micro and macroscopic models of rock fracture. *Geophysical Journal International*, *152*(3), 718–728.

Utsu, T. (1965). A method for determining the value of *b* in a formula *logN= a-bM* showing the magnitude-frequency relation for earthquakes. *Geophys. Bill. Hokkaido Univ.* 13: 99–103.

Utsu, T. and Ogata, Y. (1995). The centenary of the Omori formula for a decay law of aftershock activity. *Journal of Physics of the Earth*, 43(1): 1–33.

Van Noortwijk, J. (2004). Gamma process model for time-dependent structural reliability analysis. *Numerical Modelling of Discrete Materials in Geotechnical Engineering, Civil Engineering and Earth Sciences: Proceedings of the First International UDEC/3DEC Symposium*, Bochum, Germany, 29 September–1 October 2004, CRC Press.

Veneziano, D., Cornell, C.A. and O'Hara, T. (1984). Historic method for seismic hazard analysis. *Electrical Power Research Institute, Rep. NP-3438, Palo Alto, CA*.

Veneziano, D. and J. Van Dyck (1987). Statistical analysis of earthquake catalogs for seismic hazard. *Stochastic Approaches in Earthquake Engineering*, Springer: 385–427.

Vermeulen, P.J. (2014). *Extension of the Parametric-Historic Procedure for Probabilistic Seismic Hazard Analysis* (Master's Thesis, University of Pretoria).

Vermeulen, P.J. and Kijko, A. (2017). More statistical tools for maximum possible earthquake magnitude estimation. *Acta Geophysica* 65(4): 579–587.

Vermeulen, P.J. and Kijko, A. (2018). Discussion of ''Joint Maximum Likelihood Estimators for Gutenberg–Richter Parameters λ_0 and β using Sub-Catalogs''. *Earthquake Spectra*.

Von Mises, R. (1936). La distribution de la plus grande de n valeurs. *Rev. Math. Union Interbalcanique* 1: 141–160.

Wang, Z. (2011). Seismic hazard assessment: issues and alternatives. *Pure and Applied Geophysics*, 168(1-2): 11–25.

Wallace, R. E. (1981). Active faults, paleoseismology, and earthquake hazards in the western United States. *Earthquake Prediction: An International Review* 4: 209–216.

Weichert, D. H. (1980). Estimation of the earthquake recurrence parameters for unequal observation periods for different magnitudes. *Bulletin of the Seismological Society of America* 70(4): 1337–1346.

Wesnousky, S.G. (1994). The Gutenberg–Richter or characteristic earthquake distribution, which is it? *Bulletin of the Seismological Society of America*, *84*(6): 1940–1959

Wiemer, S. and S. R. McNutt (1997). Variations in the frequency-magnitude distribution with depth in two volcanic areas: Mount St. Helens, Washington, and Mt. Spurr, Alaska. *Geophysical Research Letters* 24(2): 189–192.

Wiemer, S. and M. Wyss (1997). Mapping the frequency-magnitude distribution in asperities: an improved technique to calculate recurrence times? *Journal of Geophysical Research: Solid Earth* 102(B7): 15115–15128.

Wiemer, S. and M. Wyss (2000). Minimum magnitude of completeness in earthquake catalogs: examples from Alaska, the western United States, and Japan. *Bulletin of the Seismological Society of America* 90(4): 859–869.

Wiemer, S. and M. Wyss (2002). Mapping spatial variability of the frequency-magnitude distribution of earthquakes. *Advances in Geophysics*, Elsevier. 45: 259-V.

Woessner, J. and S. Wiemer (2005). Assessing the quality of earthquake catalogues: estimating the magnitude of completeness and its uncertainty. *Bulletin of the Seismological Society of America* 95(2): 684–698.

Wyss, M., McNutt, S.R. and Wyss, M. (1998). Temporal and three-dimensional spatial analyses of the frequency–magnitude distribution near Long Valley Caldera, California. *Geophysical Journal International* 134(2): 409–421

Wyss, M., Hasegawa, A., Wiemer, S. and Umino, N. (1999). Quantitative mapping of precursory seismic quiescence before the 1989, M 7.1 off-Sanriku earthquake, Japan. *Annals of Geophysics*, 42(5).

Wyss, M., Schorlemmer, D. and Wiemer, S. (2000). Mapping asperities by minima of local recurrence time: San Jacinto-Elsinore fault zones. *Journal of Geophysical Research: Solid Earth* 105(B4): 7829–784.

Yeo, G. L. and Cornell, C.A. (2005). *Stochastic characterization and decision bases under time-dependent aftershock risk in performance-based earthquake engineering*. Pacific Earthquake Engineering Research Center, Berkeley, California.

Youngs, R.R. and Coppersmith, K.J. (1985). Implications of fault slip rates and earthquake recurrence models to probabilistic seismic hazard estimates. *Bulletin of the Seismological Society of America*, *75*(4): 939–964.

Zhuang, J., Harte, D., Werner, M.J., Hainzl, S. and Zhou, S. (2012). Basic models of seismicity: temporal models. *Community Online Resource for Statistical Seismicity Analysis*.

11. Glossary

Incomplete catalogue: an earthquake catalogue that does not have one level of completeness over its entire time span. (Catalogues typically consist of, but are not restricted to, catalogues that contain a historical or extreme part in early years and an instrumental part, of which the level of completeness typically decreases with time).

Support: the finite interval of a probability distribution function between which its values are neither zero nor one. In other words, the interval on the real line for which the values of the distribution falls within the interval $(0,1)$ combined with its end points to make it a closed interval. In symbolic form, the support S of the distribution F is the interval defined by $\{x \epsilon : 0 < F(x) < 1\} \cup \{supremum\{x \epsilon : 0 < F(x) < 1\}]\} \cup \{infimum[\{x \epsilon : 0 < F(x) < 1\}]\}$. This happens to be a single interval.

Supremum: the least (smallest) upper bound.

Bayesian formalism: Statistical inference (such as parameter estimation and model estimation in this work) that is based on the Bayes Theorem. Specific to the purposes of parameter estimation, the Bayes Theorem states that the probability density $p(\theta|\boldsymbol{x})$ of a parameter θ, given the data vector $\boldsymbol{x}$, may be estimated from a prior belief of the values θ in the form of a prior distribution $p(\theta)$, along with the probability distribution $p(\theta|\boldsymbol{x})$ that the data are observed for a given value of θ, as well as the total probability $p(\boldsymbol{x})$ of observing the data $\boldsymbol{x}$. This is done by the formula

$$p(\theta|x) = \frac{p(\theta)p(x|\theta)}{p(\theta)}$$

where $p(\theta)$ is calculated by evaluating the integral $\int_{-\infty}^{\infty} p(\theta)p(x|\theta)d\theta$.

Appendix I: Condition of Existence of the Tate–Pisarenko Solution of m_{max}

This appendix shows the derivation of the conditions for the Tate–Pisarenko estimator *verbatim* as it appears in: Vermeulen, P.J. and Kijko, A. (2017). More statistical tools for maximum possible earthquake magnitude estimation. *Acta Geophysica* 65(4): 579-587.

In practice, it was noticed that when the Tate–Pisarenko method was applied in an iterative fashion, it diverged on some occasions. In this Appendix, a condition is presented for the convergence (and divergence) of this method.

Let m denote the unknown. Assume that m is estimated by means of iteration, which means

$$m_{i+1} = m_0 - \frac{1 - exp[-\beta(m_i - m_{min})]}{n\beta\, exp[-\beta(m_i - m_{min})]}, \tag{A1.1}$$

or

$$m_{i+1} = m_0 - \frac{1}{n\beta} + \frac{1}{n\beta} exp[-\beta(m_i - m_{min})]. \tag{A1.2}$$

Claim: There exists m, such that

$$m = m_0 - \frac{1}{n\beta} + \frac{exp[-\beta(m_i - m_{min})]}{n\beta}. \tag{A1.3}$$

Now let $m' = \beta(m - m_{min})$, therefore, $m' = m'_0 - 1/n + exp(m')/n$ for the iteration $m'_{i+1} = m'_0 - 1/n + exp(m'_i)/n$.

Consider subsequently

$$m' = m'_0 - \frac{1}{n} + exp(m'). \tag{A1.4}$$

Taking the shape of LHS and RHS into consideration (Figure A1), note that for equation (A1.4) to have exactly one root, it must be that LHS = RHS precisely, where the first derivatives of LHS and RHS coincide, that is, where the derivative of RHS = 1. This turns out to be where $m' = \ln(n)$. Therefore, for one root it must be the case that

$$ln(n) = m_0^{'} - \frac{1}{n} + \frac{exp[ln(n)]}{n}, \tag{A1.5}$$

i.e.

$$ln(n) = m_0{}' - \frac{1}{n} + 1, \tag{A1.6}$$

After further graphical consideration, it is found that the equation has no roots if (the curve moved vertically upward in Figure A1)

$$ln(n) < m_0{}' - \frac{1}{n} + 1, \tag{A1.7}$$

and

$$\therefore ln(n) < \beta\left(m_{max}^{obs_{min}} - \frac{1}{n} + 1\right). \tag{A1.8}$$

For large n,

$$ln(n) < \beta\left(m_{max}^{obs_{min}} + 1\right). \tag{A1.9}$$

Equation (A1.4) has exactly two roots when (the curve moves vertically downward in Figure A1)

$$ln(n) > m_0{}' - \frac{1}{n} + 1, \tag{A1.10}$$

or equivalently

$$\therefore \ln(n) > \beta\left(m_{max}^{obs} - m_{min}\right) - \frac{1}{n} + 1. \tag{A1.11}$$

For large *n*,

$$\ln(n) > \beta\left(m_{max}^{obs} - m_{min}\right) + 1. \tag{A1.12}$$

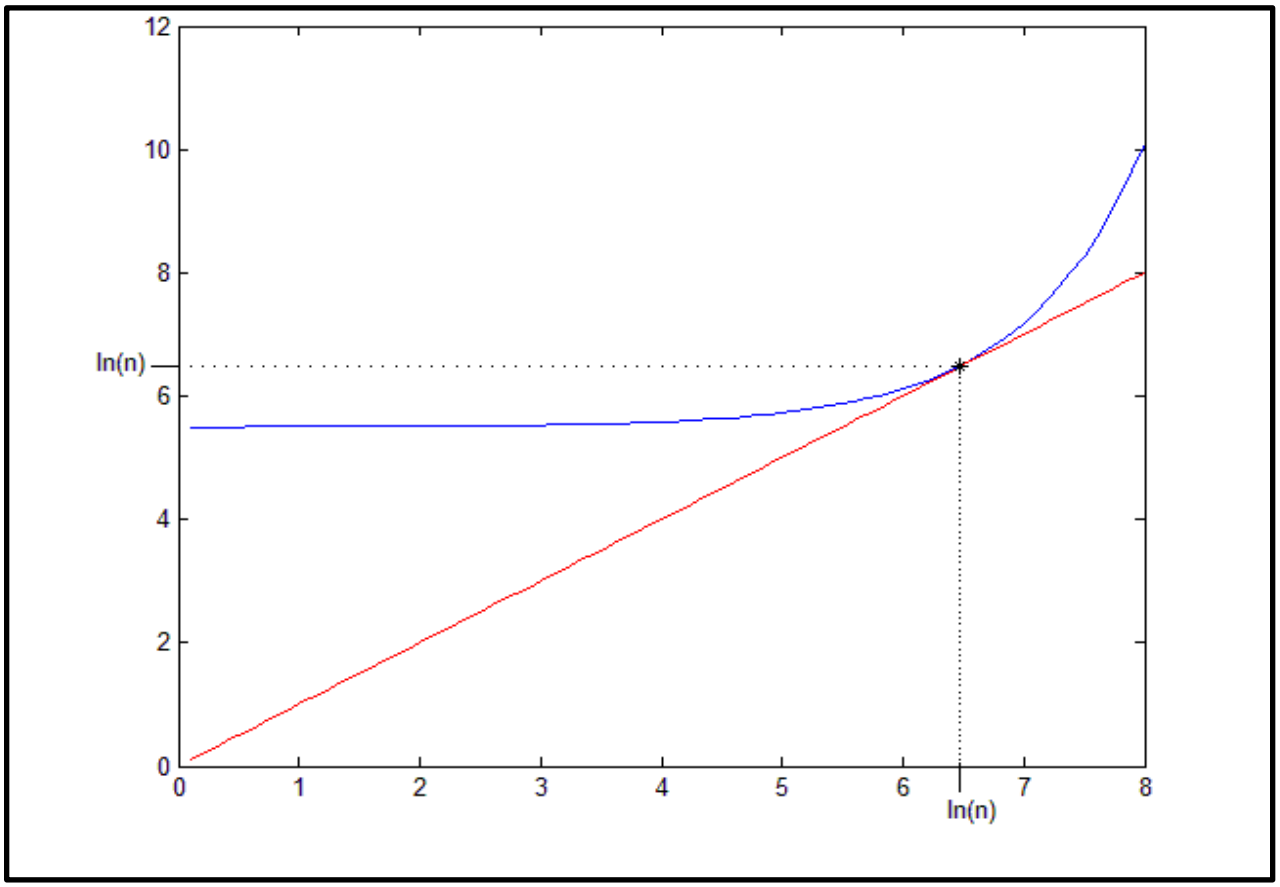


Figure A1. The blue curve represents the curve$(m') = m_0 - 1/n + exp(m')/n$, and the red line the unit function $g(m') = m'$.

If a fixed point (solution) did exist, it would usually indicate two fixed points. In addition, the following can be stated:

(1) It is possible for the smaller fixed point to be less than zero, and, therefore, smaller than m_{max}^{obs}. In either event, the smaller fixed point is neither appropriate nor can it be verified easily, which is a possible point of convergence for iteration. If convergence does take place, it must converge to the larger fixed point.

(2) If the larger fixed point were less than $\mathrm{m}_{\mathrm{max}}^{\mathrm{obs}}$., iteration would not converge, as it is clear that a point of convergence for the iteration would have to be larger than $\mathrm{m}_{\mathrm{max}}^{\mathrm{obs}}$.

(3) It can be verified graphically that if the smaller fixed point were much smaller than $\ln(n)$, the larger fixed point would be much larger than $\ln(n)$.

Appendix II: Asymptotic Equivalence of the Cooke-Kijko and Tate–Pisarenko Estimators

This appendix shows the derivation the asymptotic equivalence of the Cooke–Kijko estimator (also referred to as the Kijko–Sellevoll estimator in the text) and the Tate–Pisarenko estimator *verbatim* as it appears in: Vermeulen, P.J. and Kijko, A. (2017). More statistical tools for maximum possible earthquake magnitude estimation. *Acta Geophysica* 65(4): 579-587.

It is known (Kijko and Singh, 2011) that both the generic equation of Cooke and the Tate–Pisarenko estimates produce asymptotically unbiased estimates, which strongly suggests a measure of equivalence between the two estimators. In this section, we show that Cooke's so-called generic equation, used with the truncated Gutenberg–Richter frequency-magnitude relation, and the Tate–Pisarenko method from Kijko and Singh (2011) are asymptotically equivalent for a large number of earthquakes. In most of the estimators that Kijko and Singh (2011) introduce, they make use of an equation in the form

$$m_{max} = m_{max}^{obs} + \Delta. \tag{A2.1}$$

Specifically, Cooke's generic equation has

$$\Delta = \int_{m_c}^{m_{max}} F(x, m_{max})dx, \tag{A2.2}$$

and the Tate–Pisarenko estimate has

$$\Delta = \frac{1}{n \cdot f(m_{max})}, \tag{A2.3}$$

where $f(\cdot\,|m_{max})$ and $F(\cdot\,|m_{max})$, respectively, are the density and distribution functions of earthquake magnitudes, given the maximum magnitude.

Let $\mathrm{F(m|m_{max})}$ and $\mathrm{f(m|m_{max})}$ take the form of the Gutenberg–Richter frequency-magnitude distribution

$$F(m|m_{max}) = \begin{cases} 0, & m < m_c \\ \dfrac{1 - \exp[-\beta(m - m_c)]}{1 - \exp[-\beta(m_{max} - m_c)]}, & m_c \le m \le m_{max}, \\ 1, & m > m_{max} \end{cases} \tag{A2.4}$$

$$f(m|m_{max}) = \begin{cases} 0, & m < m_c \\ \dfrac{1-\exp[-\beta(m-m_c)]}{1-\exp[-\beta(m_{max}-m_c)]}, & m_c \le m \le m_{max}, \\ 1, & m > m_{max} \end{cases} \quad \text{(A2.5)}$$

Subsequently, using equation (A2.2) with Cramer's approximation for large values of n, we have (Kijko and Singh, 2011)

$$\Delta = \int_{m_c}^{m_{mqx}} \exp\left[-n\left(1-F(x|m_{max})\right)\right]dx\,. \quad \text{(A2.6)}$$

Laplace's approximation (Compson, 2004) leads to an asymptotic evaluation of the integral (39),

$$\Delta = \frac{\exp[-n(1-F(m_{max}|m_{max})]}{nf(m_{max}|m_{max})} = \frac{1}{nf(m_{max}|m_{max})}, \quad \text{(A2.7)}$$

which is the Tate–Pisarenko estimate in equation (A2.3).

References:

Kijko, A., and Singh, M. (2011), Statistical tools for maximum possible earthquake magnitude estimation, *Acta Geophysica* 59(4): 674–700.

Copson, E.T. (2004), *Asymptotic expansions* (No. 55). Cambridge University Press. 36–47