THE UNIVERSITY
of EDINBURGH

Model selection in earthquake recurrence relationships: *b*-value bias in tectonic, volcanic and induced seismological settings

Gina-Maria Geffers

Thesis submitted for the degree of Doctor of Philosophy

The University of Edinburgh

2022

Acknowledgements

Firstly, I would like to thank my supervisors Ian Main and Mark Naylor for their continued support throughout. Ian always provided ideas, helpful advice and direction at all times and extremely constructive and prompt feedback in the writing stage of this thesis. Thank you for being so patient and understanding with me, especially during all the troubling times and especially during the write-up. Thank you to Mark for helping me through all the coding troubles, the calming words, advice and motivation which always made me want to try again. Thank you also to my advisor, Katriona, who gave me a lot of confidence after a really tough time.

I would like to thank Kirsty for being my personal motivator, ever since I bothered her on Day 1 of this PhD, for leaving me oranges stating that 'everything will *b* ok' and having an open ear all times of any day for so many years. I would have given up a million times if it wasn't for you. Thank you also for proofreading the entire thesis and for all the feedback. You are an absolute star. Thank you to Rachel W. for all those coffee rants we had in our first few years. Thank you to Sophie, for bonding with me over having to share a bed at IUGG and the abundant tennis chat that we have been having for many years now and was the light of many of my days. Thank you for constantly reminding me that I'm capable of this and being my 'rant pad.' Thank you to my flatmate Mylène for making that last lockdown so very bearable and supporting me in this last stretch. I'm very grateful for the attic friends I've made, particularly Berit for being a sane and fellow German who always sent optimism my way and the fun times we had at EGU. Thank you to Becca for turning from a stranger to a friend and bonding with me during our trip of a lifetime at AGU. Thank you also to Elizabeth and Dylan and for all the F1 chat.

Thank you to my pre-PhD friends Imogen for sharing this journey with me and understanding me through all the ups and (mainly) lows and Stefan who was only ever a message away, even at the most dire times.

I would also like to thank the Edinburgh University Boat, Volleyball and Rifle Clubs

for all the crazy, intense, fun, exhausting but always exciting times I've had with each of them, especially to the last two for being a rock during the pandemic and allowing me to unleash my energy.

Lastly and most importantly, I would like to thank my mum and sister for not having a clue what this thesis is about (more a blessing than a curse) and for keeping me grounded throughout all these years. It was never meant to be easy and the two of you have constantly reminded me of this, while always believing in me even when I couldn't. I am so grateful that you are both there waiting for me on the other side of the finish line. Thank you.

Declaration

I declare that this thesis was composed by myself, that the work contained herein is my own except where explicitly stated otherwise in the text, and that this work has not been submitted for any other degree or professional qualification except as specified.

Gina-Maria Geffers

Abstract

The Gutenberg-Richter (GR) law $\log F = a - bm$ expresses the relationship between the magnitude of earthquakes (m) and their frequency of occurrence (F) via two constants a and b, which describe the seismicity rate and the ratio of small to large earthquakes in any given region and time, respectively. The b-value is a key parameter in earthquake statistics because it determines the recurrence relationships used in earthquake hazard assessment and therefore the extrapolated likelihood of extreme events that may be larger than those that have occurred so far in the instrumental or historical record.

In practice, b-value estimates are easily affected by estimates of other prior required parameters such as the magnitude of completeness of a catalogue. Any estimate of b is also conditional on the assumption that the underlying distribution is the GR law, but this may not always be the case. These dependencies can easily lead to bias in estimating the hazard, and therefore ultimately the risk at a certain location. The aim of this research is to quantify these biases, and to understand their potential impact on the seismic hazard, notably in defining earthquake recurrence relationships.

Here I test both the GR law and a modified version, denoted MGR, which introduces a gradual taper to the frequencies at large magnitudes. The MGR has the advantage of maintaining a finite rate of energy release in the population, at the expense of an extra parameter representing a characteristic or 'corner' magnitude characterising the taper. I also examine a variety of existing methods for parameter estimation of the magnitude of completeness to assess how sensitive changes in the b-value are due to amount of data, the dynamic range of magnitudes above the completeness threshold, the model estimation itself, and then apply these to a wide variety of examples of seismicity. This choice of a variety of seismological settings incorporates volcanic and induced earthquakes because they have not yet been as widely researched in terms of the potential biases described above as tectonic seismicity, and to test the common claim that b-values are systematically higher for volcanic and induced seismicity. This

allows a comparative discussion of the extent to which the recurrence relationships are similar or different to those of tectonic seismicity, and of how parameters in tectonic seismicity may be affected by the nature and scale of the study regions chosen. This is done initially on synthetic data sampled from a known underlying distribution, and then by applying the lessons learned to the real data.

The results show that induced and volcanic seismicity do not necessarily have systematically higher b-values, i.e. there are not necessarily relatively more small than large events in such settings. Instead, the b-value is often biased to high values, due to a combination of narrow dynamic range of magnitude and number of events, the application of incorrect methods of model estimation, and the prior assumption of an inappropriate underlying distribution associated with a particular physical process. Furthermore, I show that volcanic settings can have surprisingly low b-values when the data analysed does not just cover specific outbreaks or eruptions, but spans a much longer period of time, thereby increasing both the amount of data and the dynamic range of magnitudes. While it is not simple to identify the true underlying model when investigating new, real catalogues, I suggest how alternate hypotheses may be tested before assuming the default GR law in all situations, and provide a protocol for doing so. I also show that the source of the bias in the b-value can be best explained by the convergence of the mean magnitude from below to an asymptotic limit in random samples. As a direct consequence, the maximum likelihood b-value requires large amounts of data and dynamic range to reach full convergence with high accuracy and precision.

The results are important because they quantify the systematic error in the b-value that may arise in different settings from the different aspects of the primary data, its sampling, the prior assumption of the form of the frequency-magnitude distribution and the methods of analysis in common use at present. In turn, such systematic errors are important because they can have a large effect on forecasting the likelihood of rare, large events, and hence should be considered as a potential source of epistemic uncertainty in seismic hazard analysis.

Lay Summary

To study earthquakes, we generally split earthquake data on event time, location, depth and magnitude into smaller catalogues which represent a set number of earthquakes that have occurred in a given geographical area and during a specific period of time. Magnitudes are particularly important as they tell us how big or small an earthquake is and they generally follow a relationship called the Gutenberg-Richter law. This expresses a link between the magnitude and the total number or frequency of earthquakes in a particular catalogue and is a critical control on estimates of seismic hazard needed to assess and manage the risk of earthquakes to buildings, infrastructure and people. The gradient of the Gutenberg-Richter law is known as the b-value and is important because it provides a ratio of small to large earthquakes in the catalogue. If the b-value is small, there are relatively more large earthquakes than we would expect in a given catalogue; if the b-value is large, there are relatively many more small earthquakes than large ones.

There is little scientific debate about the b-value in settings where earthquakes occur naturally, for example due to plate tectonics, where commonly $b \approx 1.0$. Yet, for areas of seismicity which are related to volcanism or human activity (which we call 'induced seismicity'), there is more controversy. For instance, in these areas, a high b-value suggests that there are normally more smaller earthquakes because the Earth's crust does not possess the same capabilities to rupture as it does at plate boundaries. However, the lack of data is problematic because the magnitudes are often smaller and not as many earthquakes are recorded by monitoring networks as they lack the necessary sensitivity to detect events above the background noise. There are also no particularly large events which would increase the range between the minimum and maximum magnitudes observed in a catalogue, and often large events that do occur come as a surprise.

To test what happens with lack of data, I create synthetic earthquake catalogues and analyse the accuracy and precision of estimates of the b-value where we know the

answer. I apply this understanding to real earthquake catalogues from a variety of geographical locations that show tectonic, volcanic and induced seismicity. To do this, I use a method to reduce catalogues in size (and the range of magnitudes observed) to show what happens to the b-value when the amount of available data is reduced. Furthermore, I take an in-depth look at what happens when standard methods which assume the Gutenberg-Richter distribution are used on earthquake data that actually does not follow this distribution, but maybe a different one. The results suggest that in all cases, too little data results in incorrect b-values which are often too high (but also sometimes too low). When the wrong distribution is used, this can also lead to high b-values. Therefore, when using b-values as a parameter for other calculations (such as associated hazard), one needs to be sure that this really is the correct b-value and not an erroneous one due to undersampling.

Overall, I show that high b-values (suggesting many more small earthquakes than large ones) are expected in small data samples. This then underestimates the number of large earthquakes that are likely to happen in a particular region and consequently results in underestimation of the hazard, resulting in inadequate procedures to manage the risk.

Table of Contents

Glossary
Symbols

log	logarithm to base 10
ln	logarithm to base e
M_0	seismic moment
μ	shear modulus
A	area of rupture
D	average displacement
B	power-law exponent in the gamma distribution for seismic moment
M_c	moment of completeness
M_θ	corner moment
M_i	individual catalogue moments
m	magnitude
m_W	moment magnitude
m_L	local magnitude
Δm	magnitude bin
m_{min}	minimum magnitude
m_{max}	maximum magnitude
m_c	magnitude of completeness
m_θ	corner magnitude
c, d	parameters of moment magnitude relation
N or n	sample size
a	a-value
b	b-value (to base 10) for the Gutenberg-Richter law (magnitudes)
β	b-value (to base e) for the Gutenberg-Richter law (moments)
$\tilde{b}$	estimate of the b-value
F	equivalent to N
I	maximum amplitude
T	dominant period recorded by the seismometer
h	depth of hypocenter
S	correction for the variation of amplitude with the earthquake's depth
Δ	epicentral distance from seismometer (in km)
$\langle x \rangle$	expectation value of a general parameter x
$\bar{x}$	mean value of x
$p(x)$	probability density function of x
$\hat{x}$	estimated value of x
$\Gamma(n)$	gamma function
ω	maximum magnitude observed in a finite sample
λ	underlying b-value (to base e)
$\hat{\lambda}_{Aki}$	Aki's estimate of the underlying λ-(equivalent to b) value
r_a	threshold dynamic range for accuracy at 95% confidence interval
r_p	threshold dynamic range for precision at 95% confidence interval
δx	standard deviation in x
dm	incremental magnitude bin size

Abbreviations

PSHA	probabilistic seismic hazard analysis
FMD	frequency-magnitude distribution
GR	Gutenberg-Richter
MGR	modified Gutenberg-Richter
OEF	operational earthquake forecasting
MLE	maximum likelihood estimate
CI	confidence interval
INGV	Istituto Nazionale di Geofisica e Vulcanologia
USGS	United States Geological Survey
IMO	Icelandic Meteorological Office
SCSN	Southern California Seismic Network
SCEDC	Southern California Earthquake Data Center
NCEDC	Northern California Earthquake Data Center
OGS	Oklahoma Geological Survey
HF	hydraulic fracturing
PNR	Preston New Road

Chapter 1

Introduction

1.1 The research problem

For many decades, the scientific community has been calculating seismic hazard and its associated risk, particularly in regions of regular and large seismicity where the implications to populations could be severe. The level of accuracy of these hazard assessments varies, dependent on the available data, but also on the process and methods used to calculate the hazard. Most research has been focused on the hazard imposed by intermediate to large magnitude events (Main, 1995), as they are the most consequential economically, politically and in human terms (England and Jackson, 2011). Probabilistic seismic hazard analysis (PSHA) generally uses magnitudes greater than ~ 4.5 (Baker, 2013) as smaller magnitude events are not believed to be capable of damaging structures and are hence classified as irrelevant for seismic risk calculations. However, in recent times, we have seen that smaller events can result in damage, or at least significant concern, particularly in cases of induced seismicity (Ellsworth, 2013; McGarr and Barbour, 2017; Foulger et al., 2018; Ellsworth et al., 2019). All studies of seismic hazard involve quantifying the likelihood of earthquake occurrence, expressed as a frequency-magnitude distribution (FMD).

The Gutenberg-Richter law (GR, Gutenberg and Richter (1944)), $logF = a - bm$, is one of the main relationships to describe earthquake occurrence in tectonic regions

(Kagan, 1991). The two constants a and b represent the seismicity rate and the ratio of small to large earthquakes in a given region and time, respectively. The original GR law has no upper bound, so implies a non-physical infinite maximum magnitude. Estimating the b-value relies on an accurate estimate of magnitude of completeness m_c in the earthquake catalogue. There are many methods to estimate m_c (Mignan and Woessner, 2012) and still it often remains difficult to obtain an accurate and precise value, in return affecting the inferred b-value of a given catalogue. In tectonic settings, the b-value is commonly found to be ~ 1.0 (Frohlich and Davis, 1993; Kagan, 1999). In contrast, it is often estimated to be > 1.0 in cases of volcanic and induced seismicity (Roberts et al., 2015; Mousavi et al., 2017; Clarke et al., 2019). However, $b > 1.0$ in other case studies may also be a result of lack of data (in sample size, N, and/or dynamic range, $m_{max} - m_{min}$), both of which are required for accurate and precise estimates of b.

In addition to the GR law, there are further distributions to describe earthquake magnitudes such as the characteristic earthquake distribution (Schwartz and Coppersmith, 1984) or the modified GR distribution (MGR, Kagan (1997)) which introduces an exponential tail or roll-off to the cumulative frequency in the seismic moment distribution. Nonetheless, the 'go-to' assumption for the behaviour of earthquake catalogue magnitudes remains the GR law. If our estimates of the b-value are inaccurate or imprecise, either because of sampling or using the wrong model to estimate its value, then the estimates of seismic hazard will similarly be biased or highly uncertain.

1.2 The claim

This thesis addresses some of the main issues surrounding lack of knowledge (in data or the underlying distribution) as sources of epistemic uncertainty in estimating seismic hazard from earthquake catalogues, as identified above. As a means of doing so, I start by creating and using synthetic earthquake data (magnitudes) drawn randomly from both GR and MGR distributions to form a benchmark where we know the true values of the parameters in the model. The model parameters generating the synthetic data are

designed to mimic those of tectonic, volcanic and induced seismicity. The significance of these synthetic results is evaluated using Monte Carlo simulations and used as a reference for the following analysis on real earthquake catalogues.

To estimate m_c in real catalogues, I use the b-value stability (BVS) method which in its original form assumes the GR law. I have adapted this method to assume MGR in cases where the underlying distribution is an MGR distribution (in synthetic data only) and use both methods applied to real catalogue data, indicating the issues of assuming the potentially incorrect distribution for underlying data.

I further show that the elevated b-values reported in many studies of volcanic and induced seismicity are likely to be biased to high b-values when there is clear lack of data and dynamic range. Furthermore, the analysis on real data shows that more data is required to discriminate between a truly high b-value and a biased high b-value in the case of an underlying GR distribution. This is particularly problematic because any bias in the b-value to high values is likely to result in underestimation of the hazard from rare, large events. Ultimately, a sufficient dynamic range of the data is absolutely crucial, in addition to the number of events within a catalogue, to determine an unbiased b-value.

The assumption of an infinite maximum magnitude remains in the standard GR and MGR methods applied here. However, I go on to test two analytic corrections to these MLE methods which include truncation of the GR and MGR laws in the case of a finite maximum magnitude, thereby meeting the physical constraint of a finite rate of energy or seismic moment release. I also test the hypothesis that the mean magnitude in a finite sample is a good approximation for the expectation value of the magnitude in an infinite sample. I show that this theory explains the observed convergence of the mean magnitude with respect to dynamic range, finding that the mode of the data is well matched to the maximum likelihood expected in a random sample. This theory is propagated into the convergence of the b-value and applied to synthetic and real data examples. The results demonstrate that in many published studies there is insufficient dynamic range to obtain a precise and accurate estimate of the b-value, and

that b-values in these cases are likely to be biased to high values, resulting in a biased estimate of the likelihood of larger events.

1.3 Thesis layout

Chapter 2 summarises the scientific context for the problems addressed, including a review of common methods used for estimating b-values and associated uncertainties and explains how crucial obtaining robust estimates are for seismic hazard analysis.

In Chapter 3, I discuss the two main frequency-magnitude distributions which I use in this thesis, current existing methods used to estimate the b and corner moment (M_θ) parameters and some modifications I have made so the methods are appropriate for the true underlying distribution. After discussion on the model selection criteria I use here, I then introduce the computational process used to create synthetic earthquake data (magnitudes) for the ideal case of random sampling from a known underlying distribution, and finally I introduce the real earthquake catalogues analysed in Chapters 4 and 5.

Chapter 4 is a research paper which has been published (Geffers et al., 2022). Here, I present the methods used to develop and analyse synthetic and real catalogues. Chapter 5 then presents a more in depth analysis of the b-value in synthetic catalogues, particularly when changing single or multiple parameters simultaneously and a different approach to 'thinning' of catalogues as well as further investigation using other real catalogue data which was not included in the research paper.

In Chapter 6, I test the analytical corrections for the case of a finite maximum magnitude in an infinite sample and compare these to synthetic data previously created. I also test an alternate hypothesis for the convergence of the b-value. Finally, Chapter 7 offers a comprehensive discussion of the results obtained in all research chapters, focusing in particular on the bias of the b-value and its uncertainty, both in terms of accuracy and precision, in varied settings and the implications of these. I then outline further limitations encountered throughout the work performed in this thesis and ideas

for future work. Finally, I summarise the main conclusions in Chapter 8.

Chapter 2

Literature Review

2.1 Motivation

The consequences of large earthquakes often lead to huge impacts on a society and its economy, even in more developed parts of the world (e.g. the 2011 Tohoku earthquake which resulted in almost 16 000 deaths and economic losses of around USD 230 billion (World Bank, 2012; Koshimura et al., 2014)). For this reason, it is crucial to be able to assess the hazard which these events bring with them, allowing risk assessments required for land use, infrastructure planning and widespread preparation. While the seismic risk has been increasing due to increased exposure, the seismic hazard has remained fairly constant when averaged over long time periods, due to the stability of global dynamics (Main, 1996). A common technique used as a first step in hazard mitigation is probabilistic seismic hazard analysis (section 2.3). Much research has been focused on the hazard imposed by intermediate to large magnitude events (Main, 1995), as they are ultimately the most consequential economically, politically and in human terms. However, assessing the hazard of smaller to medium-sized events, especially within volcanic and induced seismicity, can be just as important in many applications. A major part of this thesis is rooted in the understanding of small to intermediate earthquakes, which dominate the statistics, how and if they scale the same as large events, and whether it is physically acceptable to apply existing models to such smaller

events.

This chapter will examine the current understanding of hazard associated with seismic activity (section 2.3) and the use of common models for earthquake occurrence (section 2.4). In addition, I will summarise the frequently used methods for parameter estimation of these models (section 2.5) and how they may change in a variety of scenarios and different types of seismicity (section 2.6).

2.2 Earthquake phenomenology

The effects of large earthquakes are many and varied, often classified as primary and secondary impacts. Primary effects will occur as a direct result of the ground shaking, such as collapse of buildings. Secondary effects are then results of primary effects, including build up of tsunamis, fires, liquefaction, landslides or spread of disease. In large events, primary and secondary effects can combine to exacerbate loss of lives and decline in economy, for example in the 2004 Sumatra-Andaman earthquake and subsequent tsunami, resulting in more than 225 000 deaths in 14 different countries and huge economic recession due to dwindling tourism, heightened poverty, displacement, inequality and conflict (Telford et al., 2006). Understandably, seismologists are doing their utmost to better understand the nature of these events, in particular the associated hazard and resulting risk. Here, I introduce the basic principles behind the occurrence of seismic events and how they are quantified and understood.

2.2.1 Magnitude scales and seismic moment

The magnitude of an earthquake was the first quantitative measure of the 'size' of an event – based on the maximum amplitude of the waves recorded on a seismogram, corrected for distance from the source. There are a variety of magnitude scales, all of the general form

$$m = \log \frac{I}{T} + S(h, \Delta) + C$$

where I is the maximum amplitude, T is the dominant period recorded by the instrument, S is a correction for the variation of amplitude with the earthquake's depth h and epicentral distance Δ from the seismometer. C is a calibration factor that depends on the instrument and sometimes on the region of interest (Stein and Wysession, 2003). A major advantage of magnitude is that it can be determined from analogue records, and hence there is a longer record of past events.

The earliest magnitude scale was introduced by Charles Richter in 1935, based on records of local shocks in southern California. It is therefore known as a local magnitude, denoted as m_L. This is a logarithmic measure of the peak amplitude of horizontal ground displacement measured by a Wood-Anderson seismometer at a standard distance of 100km, where a magnitude zero event has 1mm of net displacement recorded by this type of seismograph in California (Richter, 1935). Richter introduced magnitude as a logarithmic measure of radiated source energy, by analogy with the logarithmic stellar (apparent) magnitude scale. Since then, various local and global magnitude scales have been developed and put in use, however, the main limitations of magnitudes remain: firstly, they are empirical and therefore have no direct connection to the physics of the earthquake source. Secondly, magnitude estimates vary significantly with azimuth due to the source amplitude radiation patterns, and thirdly, all magnitude scales saturate as the dominant frequencies of larger events fall below those capable of being recorded by the instrument (Figure 2.1). Today, the standard magnitude scale for teleseismic earthquakes is the moment magnitude scale. It is based on the seismic moment M_0 which is a true physical source parameter and does not saturate for large events as it is measured from the radiated energy (the zero-frequency asymptote of the wave spectrum after correcting for radiation pattern) or from geodetic measurements of displacement at the surface measuring ground motion and hence not affected by the instrumentation used (Hanks and Kanamori, 1979). The seismic moment is defined by

$$M_0 = \mu A D \tag{2.1}$$

where μ is the shear modulus, A is the area of rupture and D is is the average

displacement along A. It is a direct scalar measure of the size of the force couple at the earthquake source, in SI units of Nm. For earthquake records, the inversion from the seismograms requires digital data, so seismic moments have only been available routinely since about 1976 (Ekström et al., 2012). This seismic moment is the key parameter in the moment magnitude scale developed by Hanks and Kanamori (1979):

$$m_W = \frac{2}{3}\log M_0 - 10.7. \tag{2.2}$$

However, inferring the scalar seismic moment, and hence the moment magnitude of an event involves deconvolution of recording site, path effects and source radiation patterns from digital data, leading many network operators to still prefer using magnitudes that are more simply computed, and can be inferred from analogue data, such as m_L. These are then converted to moment magnitudes for seismic hazard analyses using empirically derived regression formulae (Goertz-Allmann et al., 2011). The reasons for this are that journalists and the public are used to magnitude estimates – magnitude captures the dynamic range of earthquake size better – and that such regressions allow us to calibrate past magnitudes in terms of seismic moment, and hence homogenise earthquake catalogues.

Many of the magnitude scales shown in Figure 2.1, such as m_L (denoted M_L on the graph), m_S (surface magnitude, denoted M_S on the graph) and m_B (long-period body-wave magnitude) saturate for large teleseismic earthquakes. The m_L scale, for example, saturates around a magnitude of 7.0 (Howell, 1981). This means that magnitude estimates are too small for a given energy release at the source. Only m_W does not saturate for large teleseismic earthquakes. For smaller earthquakes, the seismic moment is not regularly reported and m_L is commonly used.

2.2.2 Earthquake catalogues

Earthquake catalogues are a database of earthquakes and their properties tabulated for a certain region in space and time. The standard properties are origin time, hypocenter coordinates, depth and magnitudes. Many catalogues provide further information

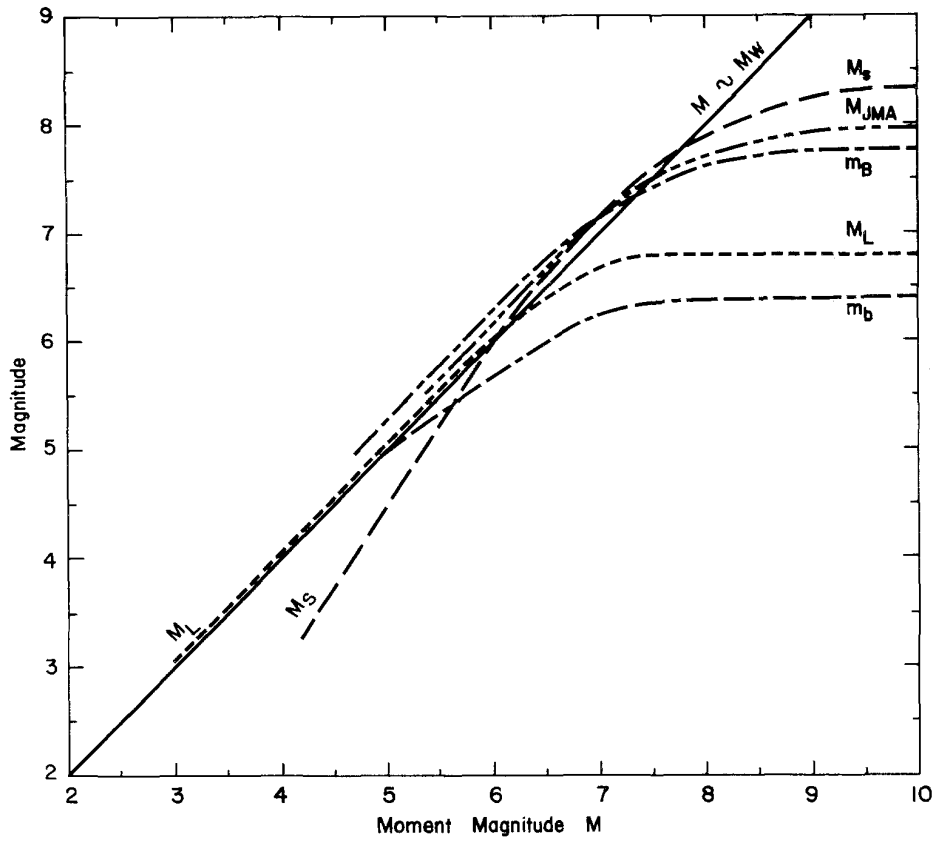

Figure 2.1: A comparison of the moment magnitude scale with other magnitude scales, indicating saturation of each. Note the different nomenclature on the figure, where both M and m are magnitude indicators. M_w (moment magnitudes), M_S (surface magnitudes), M_{JMA} (Japan Meteorological Agency magnitudes), m_B (long-period body-wave magnitude), M_L (Richter magnitude), m_b (short-period body-wave magnitude) after Heaton et al. (1986).

such as focal mechanisms, seismic moment, or more than one magnitude (on different scales) for each event. This data is the basis for extensive analyses and development of statistical models and identification of patterns in seismicity which can aid in characterising and understanding earthquake behaviour. In this thesis, I also make use of synthetic catalogues which are based on random computational sampling from a known underlying distribution of events.

2.2.3 Scale-free geometry

Many geological phenomena are scale invariant – including aspects related to the frequency-size distributions of faults and earthquakes (Turcotte, 1997). Scale invariant phenomena can be described in terms of a fractal distribution of objects, where, for example, faults of different sizes cannot be distinguished without a scale bar, and the number of faults larger than a specified length follow a power-law distribution (Turcotte, 1997). The same appears to be true for earthquake ruptures, implying a power-law distribution of seismic moment in a population of earthquakes (Turcotte, 1997; Clauset et al., 2009). Other commonly reported scale invariances in earthquake populations include the spatial distribution of hypocenters and the frequency of aftershocks (Meng et al., 2019). While there have been other statistical distributions which have been applied to specific aspects of seismic distributions (such as the gamma distribution for inter-event time (Touati et al., 2009)), the power-law distribution is the only one with no characteristic length scale, diagnostic of scale-free behaviour (section 2.4.1).

2.3 Probabilistic seismic hazard analysis (PSHA)

Probabilistic seismic hazard analysis (PSHA) involves the quantitative estimation of ground motion hazards which are likely to occur within a given time window and geographical area at a given probability, typically 2% or 10% in 50 years (Cornell, 1968; Wesson et al., 2007; Stein et al., 2012; Field et al., 2014; Petersen et al., 2020). It has been widely used for half a century by governments and industry in applications regarding

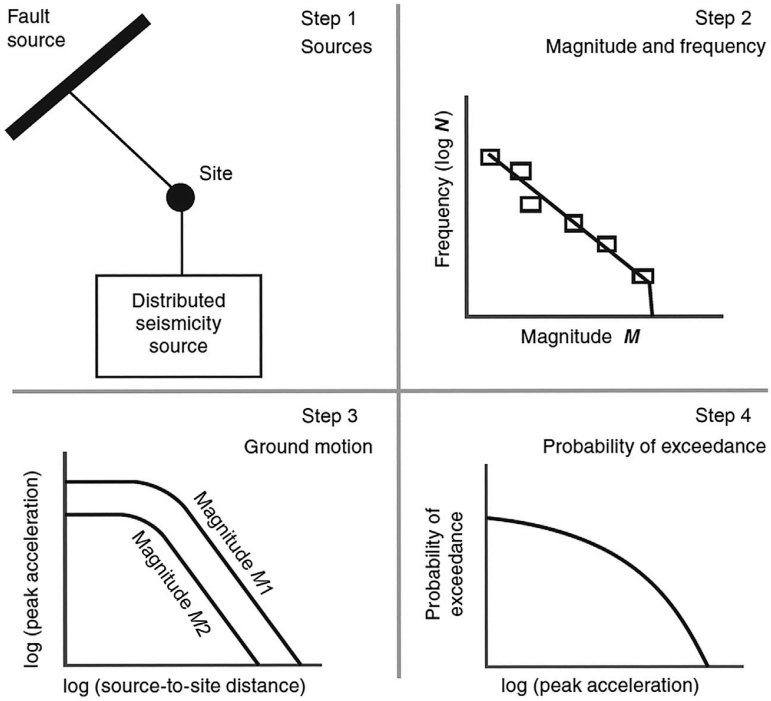

Figure 2.2: The four steps of probabilistic seismic hazard analysis (Stirling, 2014).

lives and property as well as in developing safety criteria for nuclear power plants, roads, railways, dams, establishing building regulation and design and determining earthquake insurance rates (Mulargia et al., 2017). The definition of earthquake hazard is not made in terms of a simple physical quantity that can be measured purely objectively, but also includes political and economic considerations, such as the choice of the level of hazard to publish in the resultant maps.

PSHA consists of four basic steps: a) the definition of earthquake sources, b) the estimation of the probability of recurrence of seismicity and the maximum magnitude, c) attenuation of ground motion characteristics with respect to distance and d) the probability of exceedance of a given level of ground motion from combining the first three steps. These steps are represented in Figure 2.2. By repeating this for a grid of sites, a map of the probability of occurrence of a given level of ground shaking can be made for application in building design and regulation.

A main limitation in this process is the assumption that the recurrence relationship (section 2.4) and the attenuation laws for strong ground motion are known accurately,

which is not always the case (Reiter, 1991). Knowledge of recurrence is crucial to identify where and when large earthquakes may occur and how large they will be. To attack this problem, a variety of past earthquake locations and magnitudes, active faults, geological recurrence times and fault slip rates are combined to constrain future long-term seismic hazard (Reiter, 1991). In this thesis, my goal is to improve the estimates of earthquake recurrence over a broad range of magnitudes.

In its simplest form, the finite size of the Earth and the finite deformation rate (Main, 1995; Sornette and Sornette, 1999) imply a maximum magnitude of tectonic earthquakes. However, despite much ongoing scientific debate, it is yet to be constrained (Reiter, 1991; Kijko, 2004; Holschneider et al., 2014; Shcherbakov et al., 2019) and quantifying this estimate is often complicated and requires a significant degree of expert judgement (Reiter, 1991). However, it is commonly known that the largest earthquakes so far have been generated by subduction zones, including the 1960 Valdivia earthquake (m_W = 9.6), the 1964 Alaska earthquake (m_W = 9.2) and the more recent 2004 Sumatra (m_W = 9.3) and 2011 Tohoku (m_W = 9.0) earthquakes. The idea of a maximum earthquake magnitude is an important factor in seismic hazard analysis, as understanding this and constraining the upper bound is crucial in considering effects and consequences on societies. Reiter (1991) defines the maximum possible earthquake magnitude as the upper bound determined by earthquake processes, however unlikely, associated with a specific earthquake source. Yet, instrumental and historical records are almost always too short to reflect the full potential of faults that would be required for an accurate calculation of the maximum earthquake magnitude. On the other hand, Vanneste et al. (2018) have suggested that there may be exceptions when hindcasting historical earthquake catalogues, if the time window used contains at least one of the largest events or if it has converged to the long-term average.

2.3.1 Time-independent hazard

Standard PSHA assumes that earthquakes are a random, time-independent Poisson process described by the following criteria:

1. Earthquakes are independent (memoryless)

2. Seismicity is stationary

3. Earthquakes can't occur simultaneously.

This implies that earthquakes are equally likely directly after the last event and much later, leading to a degree of non-causal clustering in time and suggesting that there is no such thing as an 'overdue' earthquake (Stein et al., 2012). Hence, for long-term seismic hazard assessment, time-independent forecasts commonly use earthquake data where the catalogue is declustered to remove dependent events (van Stiphout et al., 2012) – a process where foreshocks, mainshocks and aftershocks are separated and only mainshocks retained.

2.3.2 Time-dependent hazard

Over the past years, time-dependent PSHA has gained popularity. It includes a degree of predictability above that of a Poisson process, assuming that the seismic hazard varies with time. This requires extensive knowledge of earthquake probabilities in space, time and magnitude (Stirling, 2014). In one scenario, the probability distribution for the time between earthquakes is quasi-periodic, with small standard deviation of recurrence times compared to their mean (Stein et al., 2012). This means the 'renewal' process required to generate seismicity is not solely stochastic, and includes physical constraints such as the increase and release of stress on the fault (Matthews et al., 2002). Hence, the probability of the next large earthquake varies with time, suggesting that the probability of an event is small directly after the last earthquake, but increases with time. In this scenario there is a finite memory of past events. While this idea is intuitively appealing, the authors state that in practice it remains unresolved which format (stochastic or physical renewal) better describes earthquake recurrence.

Time-dependent physical renewal models suggest that the probability of a major earthquake in the next hundred years may be much smaller than that of the time-independent model (Stein et al., 2012). However, the use of time-dependent renewal

models also requires a larger number of parameters that may be poorly constrained by the available earthquake history (Parsons, 2008). Hence, the resulting time-varying models for seismicity are generally associated with large epistemic uncertainty and random variability, making them less useful for operational forecasting (Stirling, 2014). Stein et al. (2012) suggest that approximately 8000 to 11 000 years of seismic record would be needed to show conclusively that seismicity is the result of a time-dependent physical renewal process or a time-independent stochastic one.

2.3.2.1 Operational Earthquake Forecasting (OEF)

A major current goal of time-dependent PSHA is to develop operational earthquake forecasting (OEF), which focuses on the dissemination of authoritative information about regional seismic hazards in space, time, magnitude, and ideally in terms of ground motion (Jordan et al., 2011). The aim is to provide real-time forecasts to help societies and communities prepare for earthquakes on a range of time scales, including during earthquake sequences. Forecasts here differ from predictions in that they specify probabilities that future target events will occur within certain space-time domains, while predictions cast a deterministic alarm for an event to occur, or not to occur, within a specified location, time interval and magnitude range. Various OEF systems are currently in place, particularly in Italy, New Zealand and California (for example `http://terremoti.ingv.it/en/help` or `https://www.geonet.org.nz/`). The Italian system forecasts event probabilities and ground-motion hazard for a variety of time intervals, whereas in the US, a variety of models and combinations are currently in use to provide alerts following moderate and large events and their aftershocks (Field et al., 2016). Efforts are ongoing to develop an approach which provides consistent forecasts over a range of magnitudes and over a wide range of time scales (Field et al., 2016). OEF in New Zealand has been developed for the public and government decision makers, using a combination of short-term, medium-term and long-term models to forecast event probabilities and ground-motion hazard for time scales of 1 day to 50 years (Field et al., 2016). The main focus has been on long-term hazard models as they are the most

important forecasting tools for civil protection against earthquake damage, employed to mitigate earthquake risk by informing building codes. A good understanding of the background seismicity rate is crucial, however, this requires long-term catalogues which we do not yet have access to due to geologically young records (Zechar et al., 2016).

PSHA has been much criticised for models failing to anticipate levels of acceleration, in particularly the 2011 Christchurch and Tohoku earthquakes (Stein et al., 2011). In hindsight, when major earthquakes have occurred in areas of low hazard on hazard maps, such as the Christchurch example, it is often concluded that the models are wrong. However, these could simply be events that would normally occur extremely rarely, but happen to have occurred in the recorded time period (Stirling, 2014). The standard limitations for OEF include the choice of return period and the reliability of scenario earthquakes from poorly defined sources. Data quality, quantity and uncertainty also play a major role, which provided a strong motivation for undertaking the work described in this thesis. Through the use of the past as a proxy for the future, the uncertainty in short-term forecasting remains large as it reflects the long-term distribution of hazard but does not identify where and when large earthquakes are likely to occur (Stein et al., 2012). Any statements about future seismic events involve inherent uncertainties which future research aims to minimise, but is unlikely to avoid entirely on timescales of a human lifetime or that of a building.

The work in this thesis focuses on the second step of time-independent earthquake hazard – the estimation of earthquake recurrence expressed by the frequency-magnitude relation and its uncertainty. The different forms of frequency-magnitude distributions are discussed in detail in the following section.

2.4 Earthquake frequency-magnitude distributions

Frequency-magnitude distributions (FMD) are a visual representation of the distribution of earthquake magnitudes that have occurred within a catalogue or other frame of time

or space, expressed in the form of a histogram. They are the basis for establishing the rate of recurrence which is the second step of PSHA (section 2.3). By collecting a range of magnitudes (section 2.2.1) within 'bins', they can be plotted as a histogram against their logarithmic frequency of occurrence indicating the relationship between the two variables. Nowadays, the magnitudes are most often grouped into Δm intervals of 0.1, occasionally even $\Delta m = 0.01$ (Marzocchi et al., 2020). The latter binning is likely to be optimistic as magnitude uncertainties are typically ± 0.2 - 0.3 for instrumental catalogues (Musson, 2012; González, 2017; Leptokaropoulos et al., 2018b). One of the reasons bin sizes of 0.1 are most commonly used is due to the sampling error on the frequency which is equivalent to $\sqrt{N}$ (where N is the number of events) at one standard deviation for Poisson processes, as assumed in seismic events. Too narrow a bin width runs the risk of a large number of bins with low or even zero N, and hence large counting errors, and too wide a bin loses resolution.

Reported magnitudes in earthquake catalogues are likely to be associated with random and systematic uncertainties. Tinti and Mulargia (1985) and Tinti et al. (1987) showed that parameter estimates based on best fits of models to the FMD strongly depend on magnitude uncertainties, hence suggesting the correction of bias in the best fit parameters, whereas Leptokaropoulos et al. (2018b) suggest correcting the bias on the FMD and then performing the best fit assuming no residual bias. Furthermore, uncertainties may arise from measurements leading to significant distortion in the estimation of other derived parameters such as the magnitude of completeness (section 2.5.1.1, Leptokaropoulos et al. (2018b)). Magnitude uncertainties can be due to a variety of other sources such as random errors and catalogue heterogeneities (conversion among different magnitude scales), instrumentation, methodology, seismological network implementation and modification or from geophysical phenomena including attenuation, site effects and fore- and aftershock sequences. This is of particular importance in catalogues where multiple magnitude scales are used to report events (section 2.2.1). Leptokaropoulos et al. (2018b) further suggest that a correction needs to be applied to the total number of events belonging into each magnitude bin in order

to compensate for the effect of the magnitude uncertainty. Therefore, the binning of magnitudes also leads to inherent bias unless corrected for (Tinti and Mulargia, 1985). This bias is further exacerbated if model parameters are estimated by linear regression, where for example the standard least-squares method explicitly requires an assumption of a Gaussian distribution of errors, which is inconsistent with a Poisson process (Greenhough and Main, 2008).

Earthquake faults show a remarkable degree of scale invariant behaviour, implying a power-law distribution of source length (section 2.2.3). However, physical models for earthquakes show this behaviour is restricted to a finite range, and large events may be more or less common than expected from the power law frequency-length (and hence frequency-moment) distributions. In physical models for earthquakes, the power law occurs in systems that are precisely tuned to the critical point, where the failure of one element can trigger system-sized failure. Sub-critical models result in an exponential taper at high magnitude, and supercritical models have a characteristic 'bump' at high magnitudes (Main, 1996). These represent three separate 'universality classes' in complex systems models for earthquake recurrence, which I now explore in the following section.

2.4.1 Gutenberg-Richter law

The Gutenberg-Richter (GR) law (Gutenberg and Richter, 1944, 1956) is one of the most fundamental empirical scaling relationships of seismology, relating the frequency of earthquakes of a certain magnitude occurring in a specific space-time window according to

$$\log F = a - bm \tag{2.3}$$

where F is the number of events at a magnitude greater than m with constants a and b describing the rate of occurrence of seismicity and proportion of large to small events, respectively. This is normally visualised on a FMD as discussed in section 2.4 and shown for some earthquake data from New Zealand in Figure 2.3. The incremental

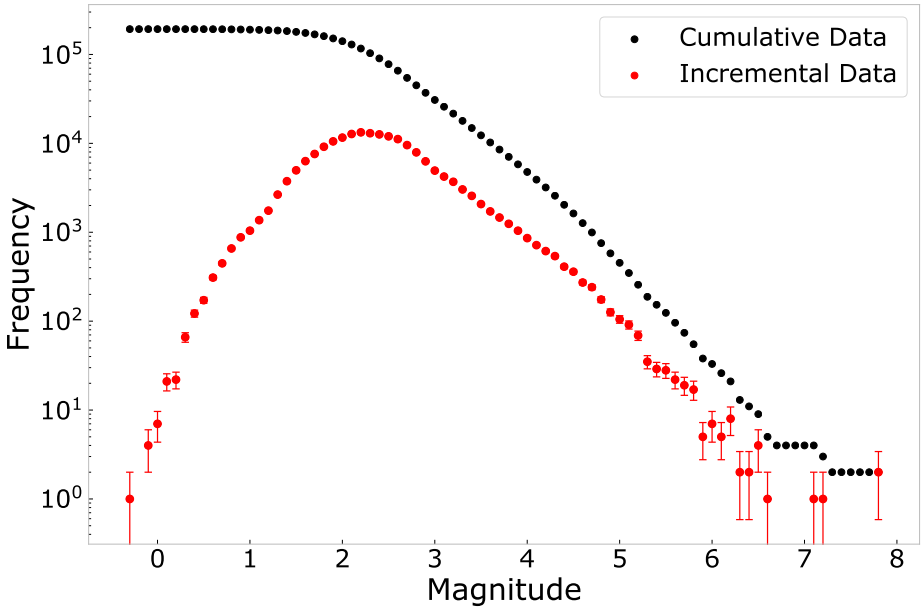

Figure 2.3: FMD for New Zealand events between 2009 and 2017 with N = 192 750, -0.32 < m < 7.82. Red indicates incremental data, black cumulative data. Error bars are indicated as frequency counting errors ($\sqrt{N}$) to one standard deviation.

data shows events within each bin of size 0.1, whereas the cumulative data counts all events greater than a given magnitude, equivalent to summing the incremental data. Frequency errors in the incremental data are shown to one standard deviation, calculated as $\sqrt{N}$. The largest errors (at largest frequency) do not show due to the nature of the logarithmic y-axis.

The GR law is known to occur in a variety of tectonic, volcanic and induced seismicity settings (Kagan, 1991; Pisarenko and Sornette, 2003; Corral, 2004; Boettcher et al., 2009; Clarke et al., 2019), in laboratory experiments (Scholz, 1968) and in theoretical models (Burridge and Knopoff, 1967; Kanamori and Anderson, 1975; Tinti and Mulargia, 1985). While the magnitude uncertainties (section 2.4) can bias the estimate of the b-value, they do not appear to affect the seismicity rate parameter (a-value) significantly (Musson, 2012).

Physically, there must be an upper bound to the GR law as this is limited by factors such as the finite tectonic deformation rate (Main, 1996), the size of regional stress regimes and fault zone geometry (Kagan, 1997). Kagan (1999) also showed that

identifying all parameters within the GR distribution allows assessment of seismic hazard (in particular the return period and the maximum expected magnitude in a given time window). However, physically it is not possible simply to extrapolate the frequency of the smaller events to determine the threat of the rare large events. An open-ended GR distribution would ultimately result in an infinite energy or seismic moment release rate if there is no maximum possible magnitude (Main, 1996).

The Gutenberg-Richter distribution (equation 2.3), when combined with the moment-magnitude relation (equation 2.2), results in a frequency-moment distribution of the form:

$$F(M_0) \sim M_0^{-\beta} \tag{2.4}$$

where, according to Kagan (2002a,b)

$$\beta = \frac{2}{3}b. \tag{2.5}$$

Thus, the GR law is equivalent to a power-law distribution of seismic moments, describing a fractal (scale-free) set as introduced in section 2.2.3 (Main, 1996; Turcotte, 1997). From equation 2.4, the moment sum in a given bin is then $M_0 F(M_0) \sim M_0^{(1-\beta)}$. For $\beta < 1$, or $b < 1.5$ which is the general case for tectonic seismicity, the moment sum over all bins becomes infinite unless a finite maximum moment is introduced. Equally, in the case of $\beta > 1$ or $b > 1.5$, infinite seismic moment is released as magnitudes tend to -∞. This is also physically impossible and therefore a physical limit of minimum magnitude is required. Scientifically, this is the actual smallest magnitude earthquake, rather than what we are able to detect to be the smallest.

In the general case, the GR law exhibits non-universal behaviour and breaks down at certain scales of earthquake magnitudes. Main (1996, 2000) identified that the GR law applies only within a finite scale range (small to intermediate earthquakes) both in model and natural seismicity due to natural fractal sets being constrained to lower and upper scale-lengths, challenging the assumption of the GR law operating the same at all scales. At low magnitudes, the scale-free behaviour of the GR breaks down due to censoring of the data below the magnitude of completeness (section 2.5.1.1). At

high magnitudes, the scale-free behaviour must also break down (when $\beta < 1$) because natural tectonic and volcano-tectonic processes are incapable of an infinite release of energy, as proven above. This, combined with the poor or even absent sampling of the largest events, makes it difficult to estimate the rate of occurrence of the largest events accurately.

There can also be issues at smaller magnitudes, even above the magnitude of completeness. For example, Staudenmaier et al. (2018) showed that only data based on the moment magnitude scale follow the GR law with a constant slope, while local magnitude (m_L) data is bilinear due to scale breakdown at small magnitudes. They further suggest that for small events ($m_L < 2$) a scaling of $m_L \approx 1.5 m_W$, whereas for events $m_L > 3.0$, $m_L \approx 1.0 m_W$, as predicted by the dislocation model for the seismic source of Kanamori and Anderson (1975).

2.4.2 Modified Gutenberg-Richter law

As proven in section 2.4.1, a practical upper bound to the GR law must exist in the common case $\beta < 1$ (or equivalently $b < 1.5$) and a lower bound must exist in the case $\beta > 1$ (or $b > 1.5$). One solution to this problem is to apply an exponential taper at large seismic moments, modifying the power-law form of equation 2.4, derived by Main (1996) from the constraint of finite moment release rate. This is commonly termed the 'tapered GR', or the 'modified Gutenberg-Richter' (MGR) law. This applies a taper to the FMD at a characteristic, or 'corner' moment (M_θ) at which the roll-off is initiated:

$$F(M_0) \sim M_0^{-\beta} \exp\left(\frac{-M_0}{M_\theta}\right). \tag{2.6}$$

This distribution was first derived for the incremental frequency (Main and Burton, 1984), though it is more common now to apply a maximum likelihood fit to a version where equation 2.6 applies to the cumulative frequency (Kagan, 1997). In any case, Kagan (2002a) shows that the reliable determination of M_θ requires a catalogue to include some earthquakes that are close in magnitude to that of the corner moment (or equivalent corner magnitude). Therefore, the upper moment bound can be estimated

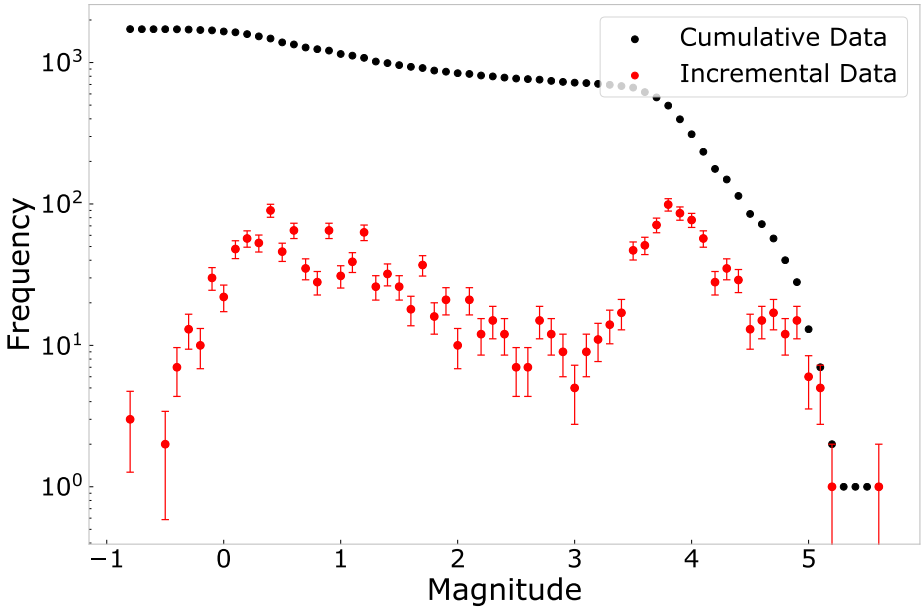

Figure 2.4: FMD for Mount St. Helens events in 1980 with N = 1724, -0.8 < m < 5.7. Red indicates incremental data, black cumulative data. Error bars are indicated as frequency counting errors $\sqrt{N}$.

adequately only for global seismicity or for large sub-catalogues of this. In the case of smaller data sets, a lower limit for the corner moment can be obtained but requires critical assessment of its practicality. This is particularly important when investigating volcanic or induced seismicity, which are generally smaller magnitude data but also often lack data availability.

2.4.3 The 'characteristic' earthquake distribution

As indicated in the previous sections, the FMD is most often fitted to the GR law (section 2.4.1) and sometimes to the modified GR law (section 2.4.2). However, not all data exhibits either of these distributions, there are many more that have been suggested as possibilities. Another common distribution that is observed in complex systems models for earthquakes is that of the 'characteristic' earthquake. This was first suggested from geological recurrence models involving the repetition of earthquakes which have similar (characteristic) properties, such as similar surface displacements and similar seismic moments (Schwartz and Coppersmith, 1984). In the ideal case this would then result

in a periodic earthquake cycle with a constant recurrence interval, but a characteristic 'bump' in the frequency-magnitude data can also occur in data and models that do not exhibit the periodicity of the model developed by Schwartz and Coppersmith (1984). An example of such data is that of Mount St. Helens in Washington, USA. Prior to the 1980 volcanic eruption, there was a great extent of seismic activity at shallow depths below the carapace, shown in Figure 2.4. This catalogue contains 1724 events with magnitudes reported on a variety of magnitude scales. Unlike the GR distribution, the characteristic distribution does not allow extrapolation of small events to estimate the rates of large events. Frequency counting errors (uneven, due to the logarithmic frequency axis) are indicated on the plot, suggesting that the counting errors do not account for the 'bump' in the Mount St. Helens data and consequently the 'bump' is real (Main, 1987). Main (1995) has suggested that this 'bump' in incremental data is due to supercritical behaviour with a characteristic earthquake distribution; only regions which are perfectly critical will exhibit the GR law over the entire magnitude range. In nature, such supercritical behaviour can only occur when a) there is a finite upper limit and b) when the system is driven to a metastable state with an inverted population, and hence may not be long-lived.

Nonetheless, the GR law remains the standard null hypothesis for the majority of methods used on earthquake data today. In this thesis, I will use the GR model as a null hypothesis and additionally use the MGR distribution to test for a suitable alternative.

2.5 Model testing and selection

So far, I have described only the theoretical models for the FMD and the possible effects of uncertainties in magnitude and sampling on the estimation of their parameters. This section describes how we might estimate the model parameters, and decide between competing forms of the FMD described in section 2.4. This will allow the most robust results when applied to real earthquake catalogues in the later stages of this thesis.

2.5.1 Parameter estimation

To be able to estimate any FMD parameters in real data as accurately as possible, a complete earthquake catalogue is required. To obtain this, we must estimate the magnitude of completeness which will then allow estimation of other parameters such as the b-value and M_θ.

2.5.1.1 Magnitude of completeness m_c

The magnitude of completeness, m_c, is the lowest magnitude at which 100% of the earthquakes in a space-time volume are detected (Rydelek and Sacks, 1989; Woessner and Wiemer, 2005; Mignan and Woessner, 2012). This parameter determines the range over which the data can be analysed without artefacts associated with incompleteness and is used as a fundamental constraint on the data used for the determination of relevant model parameters. A finite m_c occurs because smaller events can be masked by larger events, ambient background noise or could be too small to register on a seismometer, immediately rendering the catalogue incomplete below m_c. The a-value i.e. the number or earthquakes above magnitude zero, can be estimated from the best fit b-value and the total number of events above the completeness threshold, equivalent to an extrapolation of the best fit to the data below m_c (Gasperini et al., 2013). A robust estimate of m_c is crucial to avoid under-sampling (if the estimate is too high), or biased estimations of parameter values (if the estimate is too low and uses incomplete data). Generally, two different approaches exist for the determination of m_c. These are a) the network-based approach (Schorlemmer and Woessner, 2008; Mignan et al., 2011) which relies on the probability of the detection of an earthquake, given the sensitivity and distribution of the seismic network and b) the catalogue-based approach which uses the day-to-night ratio of earthquake frequency (Rydelek and Sacks, 1989) or alternately assumes self-similarity in the earthquake generation process for all of the data (Woessner and Wiemer, 2005). In this thesis, I will use only methods within the catalogue-based approach, discussed in section 3.2.1.

2.5.1.2 b-value

The estimation of the b-value requires a robust estimate of m_c, but the precision of the b-value estimate also plays a role. Commonly, b-values are given to 2 significant figures, occasionally more. The sample size and dynamic range needed to obtain a b-value estimate are associated with the desired precision of the b-value itself. Regional variation of b-values is often analysed with much smaller datasets, and hence the variability may be an artefact of this sampling process, as opposed to any inferred non-universality of earthquake scaling relationships (Kagan and Jackson, 1991).

The GR b-value can give insight into the underlying causes of earthquakes. It is typically close to unity for shallow tectonic earthquakes (Frohlich and Davis, 1993; Kagan, 1999; Bird and Kagan, 2004), within its uncertainty. However, seismicity in volcanic areas or seismicity induced by fluid-related processes, such as dam impoundment or hydrocarbon production, often has elevated b-values (Suckale, 2010; Foulger et al., 2018). Table 1 in Roberts et al. (2015) is an extensive, but by no means exhaustive, summary of many studies which have claimed b-values to be elevated in active volcanic settings. Table 1 within Mousavi et al. (2017) provides a similar summary for b-values in induced seismicity settings showing a large range of fluctuation.

Once a satisfactory value of m_c has been determined by the methods outlined in section 2.5.1.1, the b-value can be inferred – most appropriately using the maximum likelihood estimate (MLE) first obtained by Aki (1965), and usually applied in modified form:

$$\tilde{b} = \frac{log(e)}{\bar{m} - (m_c - \Delta m/2)} \tag{2.7}$$

where $\tilde{b}$ is the estimate of the b-value, $\bar{m}$ is the mean magnitude and Δm is the magnitude bin width as stated in the catalogue (section 2.4).

A common pitfall when estimating the frequency-magnitude (or moment) parameters is to use a least-squares regression to fit a straight line through raw frequency-magnitude data (section 2.4). For example, the application of least-squares makes the b-value estimate overly sensitive to the effects of binning and fluctuations in the tail

of the distribution as not all of the points carry the same weight – it is much more likely for the tail of the distribution to contain only a few events in a single point whereas the points at lower magnitudes may include hundreds of events. Furthermore, the sampling or 'counting' error for a Poisson process with small numbers of events (Naylor et al., 2010) is indicated by the error bars in Figures 2.3 and 2.4, whereas the least-squares method assumes a log-normal (Gaussian) distribution of errors. The misuse of least-squares for b-value estimation is common in the literature, as is the omittance of information regarding the method applied to calculate the b-value, leading to biases in uncertainties and difficulties in replication of results (Roberts et al., 2015).

The moment-magnitude relation for m_L can show a break of slope at around $m_L \approx 2.5$ (section 2.4.1). This propagates into a systematic break of slope in the FMD for catalogues with completeness of $m_c < 2.5$ which is often the case in volcanic and induced seismicity catalogues (Staudenmaier et al., 2018). As a consequence, the estimate of the underlying b-value may be biased. To avoid this problem, the authors argue that it is essential to calculate moment magnitudes for such data, from which it is possible to determine the appropriate scaling at the study site and accordingly correct the m_L values. This correction is essential for seismic hazard analysis, because it allows an unbiased extrapolation of the size distribution for large, damaging events.

García-Hernández et al. (2021) recently provided a novel method for the estimation of the b-value, based on the FMD suggested by Ogata et al. (1993) which covers the entire magnitude range and therefore avoids a direct determination of m_c. They suggest that this method is particularly reliable for the determination of changes in b-values in temporal or spatial calculations but has the disadvantage of being computationally expensive.

2.5.1.3 Corner moment M_θ

The 'corner moment' (section 2.4.2) is a characteristic moment at which the roll-off in the MGR distribution is initiated. Here, the power law changes to exponential decay. The parameter M_θ can also be inferred from the maximum likelihood method, accounting

for any covariance between this parameter and the b-value but the uncertainty in M_θ is asymmetric and often large, and may not exclude an infinite value. Explicit formulae for this are given in (Kagan, 1991, 2002a) and section 4.3.2.2.

2.5.2 Model selection criteria

The same catalogue data may be fitted by multiple models (i.e. GR or MGR), and so it is crucial to identify which model fits that particular data best. A variety of model selection criteria exist for this purpose. In this thesis I am using the Bayesian Information Criterion (BIC) defined by Schwarz (1978) and my rationale for this is further described in section 3.3. Information criteria quantify the relative (log) likelihood of competing hypotheses such as equations 2.4 and 2.6. The BIC quantitatively accounts for the trade-off between the best fit of the model and the simplicity of the model, acting as a modern form of 'Occam's razor' and is therefore a popular option.

2.5.3 Epistemic and aleatory uncertainties

A crucial part in using FMDs to evaluate specific parameters (section 2.5.1) is the understanding and incorporation of uncertainty estimates. In any model and its associated parameter estimation, there are commonly two types of uncertainty – firstly, the epistemic (systematic) uncertainty which arises due to the lack of knowledge of an unknown parameter. In the GR and MGR cases described above this includes the b-value, the a-value and M_θ estimates as well as our lack of knowledge of the correct underlying distribution. This is the case particularly in single cases such as real earthquake catalogues which I will discuss in sections 4.5.2, 5.2 and 6.4. These uncertainties are often exacerbated by using parameters earlier on in the calculation process which are themselves uncertain, for example, magnitude estimates and binning of magnitudes (Tinti and Mulargia, 1985; Marzocchi and Sandri, 2003; Leptokaropoulos et al., 2018b; Marzocchi et al., 2020) or the estimate of m_c (Mignan and Woessner, 2012; Roberts et al., 2015). Secondly, the aleatory (statistical) uncertainty stems from random variations and is irreducible and therefore is observed in synthetic realisations (sections

4.5.1, 5.1 and 6.3).

The most commonly used estimate for the aleatory uncertainty of the b-value is given by (Aki, 1965)

$$\sigma_b = \frac{b}{\sqrt{N}} \tag{2.8}$$

where N is the number of earthquakes in a given sample. Generally this aleatory uncertainty is approximated well by a Poisson distribution which explains both the sampling bias as well as that scatter (i.e. precision) at small and large magnitudes, within confidence limits (Greenhough and Main, 2008; Main et al., 2011). This bias represents a systematic difference between the true parameters and the statistics which are used to estimate those parameters and arises due to the source of data and the methods applied. It can be particularly large when estimated from small samples of data, which in turn will have an effect on the accuracy (measure of the statistical bias) and precision (measure of the statistical variability) of such samples (Schorlemmer et al., 2004; Roberts et al., 2015; Yaghmaei-Sabegh and Ostadi-Asl, 2021). The accuracy can only be measured if the true, underlying value of a certain parameter is known (section 6.5). This is not normally the case in real data and synthetic data is often used to obtain uncertainties and bias estimates of parameters which are then translated to real data. It is important to acknowledge that accuracy and precision can be mutually exclusive, i.e. data can be either accurate, or precise, or both or neither. This will be examined particularly in Chapter 6.

Main et al. (2011) further suggest that the largest source of systematic bias is the slow rate of the process so that the large, damaging events are always insufficiently sampled by historical, instrumental and palaeoseismic data. The uncertainties will be large when N is small and describe biased b-values. Kijko and Smit (2012) showed that this bias may be removed using a reduction bias formula by Ogata and Yamashina (1986) but that in the real, physical world, the uncertainties within the model and its parameters return a more significant effect on estimated b-values than is accounted for by this bias reduction method. The epistemic uncertainty is reducible, for example by

more data becoming available and can in such cases be accounted for but Roberts et al. (2015) found that this offset is often neglected in studies of volcanic seismicity.

Main et al. (2011) conclude that there are multiple reasons for the sample bias which affects accuracy, mainly a) the power-law nature of the ideal GR distribution, b) the Poisson counting errors in small sample sizes and c) the correlation in the data due to earthquake triggering. They conclude that even larger samples of $N = 1000$, which is the commonly stated sample size required to estimate b-values, can show significant uncertainties.

2.6 Applications to seismicity data

The hazard associated with any earthquakes requires an estimate of the likelihood of occurrence of earthquakes of different sizes. The very largest earthquakes always dominate the total seismic moment released, but, depending on the b-value and the type of FMD, the hazard from strong ground motion may be greatest for intermediate to large sized events (Main, 1996). In this section, I will examine the current discussion surrounding recurrence relations for natural, volcanic and induced seismicity, and discuss the effectiveness of measures to control induced seismicity, such as the 'traffic light system'.

2.6.1 Natural and Laboratory Seismicity

The vast majority of studies on FMDs have been based on catalogues of natural seismicity, as a step in the process of quantifying seismic hazard (Figure 2.2). The rule of thumb in natural seismicity is that $b = 1.0$ (Frohlich and Davis, 1993). However, El-Isa and Eaton (2014) suggest that statistically significant b-value variations occur in both space and time but this particular study uses regression of least-squares which introduces inherent bias (section 2.5.1.2). Nuannin et al. (2012) found that the b-value estimated during a period containing large events ($m_W \geqslant 7.0$) in the Sumatra-Andaman region were significantly lower than 1.0. On the other hand, prior research in New Zealand,

California and Venezuela showed a temporary increase in the b-value prior to a larger event and hence was considered as an earthquake precursor (Smith, 1981), as also observed in acoustic emissions from laboratory tests on undrained samples of fluid saturated porous rock (Sammonds et al., 1992). Additionally, the b-value may decrease before a major earthquake, followed by a subsequent increase (Main et al., 1989; Wiemer et al., 1998; Wyss et al., 2004). However, the very existence of reliable precursors of any type, including the b-value, has recently been called into question (Jordan et al., 2011). Nevertheless, pragmatic operational earthquake forecasts made in real time, as advocated by Jordan et al. (2011), may commonly be based on a best fit b-value that evolves in time, either due to sampling artefacts such as mixing of magnitude scales (Jackson and Kagan, 1999; Amorèse et al., 2010) or real fluctuations related to changes in the stress field, as seen in laboratory samples (Sammonds et al., 1992). Hence, it is important both to quantify the best fit and the uncertainty – systematic and random – in the estimate of the b-value, before assigning significance to b-value fluctuations in space and time.

2.6.2 Volcanic and volcano-tectonic seismicity

Roberts et al. (2015) challenged the widespread inference that b-values in volcanic seismicity are elevated (section 2.5.1.2) and found that these results are based on very different types and reliabilities of data and the methodology used. In particular, they showed that very small sample sizes (where $N < 200$) as well as small dynamic (magnitude) range, lead to significant biases in m_c calculations, which in return propagate into a bias in the b-value. However, the inherent link between sample size and dynamic range means that generally, small dynamic range will also have small sample size, leading to the possibility that the observed biases are in fact caused by the sample size and not the dynamic range. Even in cases where a 'like for like' comparison can be made using appropriate methods such as maximum likelihood inference, it is usually not possible to determine a statistically significant variability of b in space and time, as the evidence is often not sufficient. In certain circumstances, however,

notably using a stable network of seismometers, a consistent magnitude scale, and large datasets with small m_c and a large magnitude dynamic range, Roberts et al. (2016) showed that mode switching between states of tectonic ($b \approx 1$) and elevated ($b > 1$) b-values can occur, likely as intermittent activation and cessation of volcanic activity occurs in different parts of the volcanic system, rather than due to any systematically inferred underlying physical process.

2.6.3 Induced Seismicity

Induced seismicity is the term used to describe mainly low-magnitude seismic activity which is partially or completely caused by anthropogenic activities. There are many different forms of induced seismicity but the most common events are due to dam impoundment, oil and gas extraction, water injection in geothermal systems (Grigoli et al., 2018), subsurface waste fluid disposal, nuclear testing as well as carbon capture and storage (Keranen et al., 2014; Verdon and Stork, 2016; Foulger et al., 2018). There have also been an increased number of earthquakes associated with hydraulic fracturing (HF) (Clarke et al., 2014; Bao and Eaton, 2016). An increase in pore fluid pressure reduces effective stress and hence, from standard arguments in Coulomb friction lowers the effective normal stress and hence the shear stress required for failure (Fossen (2010), section 4.7 in the book). Similarly, the removal of any overlying mass reduces the normal stress, and leads to reduction in shear stress required for failure (Ellsworth, 2013). Fluid withdrawal can also induce seismicity, due to poro-elastic stresses generated in and around the reservoir by compaction (Segall et al., 1994; Candela et al., 2019). While most induced earthquakes are relatively small, larger induced earthquakes ($m > 4.0$) have occurred and have led to damage in recent years (Keranen et al., 2014; Hough and Page, 2015). The state of Oklahoma has been subject to increasingly large earthquakes due to water injection into deep disposal wells (McGarr et al., 2015; McGarr and Barbour, 2017) and will be discussed discussed later in this thesis (section 5.2.3.2).

Distinguishing induced from natural earthquakes remains a major challenge to this day. While it is easier in certain areas, especially where there is no prior record

of natural seismicity, and the induced events cluster, for example, around a reservoir, some locations allow both natural and induced seismicity to occur. In these areas it is extremely hard to discriminate and identify the origin of certain events (Verdon et al., 2019b). Furthermore, as a fairly recent form of seismicity, the literature on induced events is smaller, and the degree of understanding less established than for natural seismicity, making it difficult to identify exactly how the processes compare to those in tectonic events. Extensive monitoring and modelling of induced seismicity is used to develop time dependent hazard models that can be used to forecast the hazard and to mititgate the risks associated with induced seismicity (Bourne et al., 2018). One common mechanism for attempting to control the risk, including in the UK, is the 'traffic light system' described in the next section.

2.6.3.1 Traffic Light Systems (TLS)

To mitigate any adverse consequences of induced seismicity, it is necessary to estimate hazard and use this as a basis for risk assessment in real or near-real (several hours to one day) time. The traffic light system (TLS) was introduced as a risk management mechanism using continuous and real-time monitoring of induced seismicity for specific sites. This provides a clear basis for taking operational steps to manage and mitigate the risk of large induced events. TLS are underpinned by two key assumptions: that magnitudes increase relatively gradually during injection and that the mitigation action (i.e. halting injection) will reduce the earthquake magnitudes or entirely stop events from happening (Clarke et al., 2019). Generally, these TLS are used retroactively, i.e. actions are taken *after* an event of certain magnitude has occurred but Clarke et al. (2019) have shown that using the microseismicity with statistical models allows forecasts of such events and their magnitudes, which could prove to be an effective approach in future operations.

The generic setup of TLS are similar – thresholds are set at predetermined (normally) local earthquake magnitudes (m_L), set by a regulator, for example the Oil and Gas Authority (OGA) in the UK. If events reach the thresholds, further operation of the

sites are either modified (green or amber alert) or shut down entirely (red alert). These thresholds vary between countries and occasionally within countries too, as the risk assessment and tolerance for induced seismicity vary from region to region (Walters et al., 2015). However, Akkar et al. (2014) suggested that the peak ground velocity at which visible damage may occur, e.g. to infrastructure, translates to $m \approx 3.5$. Therefore, considering trailing events and possible magnitude jumps, the regulator's objective is to set this threshold about 2 orders of magnitude higher than the magnitude of events intended to be minimised (Verdon and Bommer, 2021).

The TLS thresholds are based on the following assumptions (Baisch et al., 2019):

1. The magnitude of induced earthquakes tends to increase steadily with the duration of operations
2. Operational measures for limiting the strength of induced earthquakes exist
3. The time delay between an alert and the mitigation measure (consequence) becoming effective is small enough to prevent magnitude escalation.

Table 2.1 is a selection of various TLS, with notably different thresholds due to a variety of factors such as politics, economics and understanding from the public.

Region	**Type**	**Green**	**Amber**	**Red**	**Mag. Scale**
California, USA	WD, HF	< 2.7	-	≥ 2.7	-
Illinois, USA	WD, HF	< 3.0	2.0 ≤ M < 4.0	≥ 5.0	-
Ohio, USA	WD, HF	< 1.0	-	≥ 1.0	-
Alberta, Canada	HF	< 2.0	2.0 ≤ M < 4.0	≥ 4.0	M_L
British Columbia, Canada	HF	< 4.0	-	≥ 4.0	M_L
UK	HF	< 0.0	0.0 ≤ M < 0.5	≥ 0.5	M_L

Table 2.1: TLS magnitude thresholds for selected induced seismicity regions. WD - wastewater disposal, HF - hydraulic fracturing. After Baisch et al. (2019); Kendall et al. (2019)

Using three different types of seismicity to test the models in section 2.4 will allow us to gain a better understanding of how these three types of seismicity are related in terms

of similar physical processes. The aim is to identify which assumptions and existing knowledge of scaling laws and frequency-magnitude distributions can or cannot also be applied to lower magnitude events, as often observed in induced and volcanic seismicity.

2.7 Summary

This chapter has introduced the basic principles and standard models used in earthquake science today and provided a context for the work described in this thesis. While many of these concepts provide a foundation for earthquake hazard assessment as we know it today, major uncertainties and points of discussion remain. Some of these have been discussed here, such as the seismic record required for accurate PSHA or to establish that modifications on the GR law are needed. However, the intrinsic variability of earthquake processes, including the finite sample sizes, mean that there will always be residual uncertainty. Further uncertainties due to the methods and data used in this work are investigated in more detail in the following chapter, with more focus on the specific problems to be addressed.

A focus of this thesis is the general assumption that earthquakes within a region follow the GR law. While this has been shown to be the case in various settings, it is not certain to be the case everywhere. In the case of induced seismicity, which has been studied much less than tectonic settings, it may be presumptuous and needs to be assessed on a case by case basis. Accordingly, I will test the MGR as a suitable alternate hypothesis for different types of earthquake catalogues, using the GR law as a suitable null hypothesis. Knowledge of the correct (or at least more likely) underlying frequency-magnitude distribution is crucial for future hazard assessment, even in the case of small to intermediate sized events characteristic of induced seismicity. This exercise requires accurate reporting of parameter values (sample size, magnitude dynamic range, m_c) and estimates of relevant uncertainties, without which it is often difficult to judge the significance of reported results.

Chapter 3

Methods and Data

This chapter describes the methods applied and the data used to test the hypotheses as stated in the Introduction (Chapter 1). Methods which have been touched upon in the Literature Review (Chapter 2) are explained here in more detail and why they are used in this work. Some discussion of methods has been kept to a minimum, where they appear in more detail in the published manuscript in Chapter 4. Additionally, I introduce and explain the generation of synthetic data which will be a major part of this thesis, followed by the real earthquake catalogues analysed and the motivation for choosing these.

3.1 Generation of synthetic magnitude distributions

The first step to generation of synthetic catalogues is the simulation of seismic moment or equivalent moment magnitude distributions. Therefore, I start by presenting the two methods used to generate synthetic catalogues sampled from the GR and MGR distributions.

3.1.1 Gutenberg-Richter (GR)

To create magnitudes following a GR distribution, I create a function in Python which takes N, the b-value and a minimum magnitude event (m_{min}) as input parameters.

This function then uses a built-in Python function which randomly draws samples from an exponential distribution given the rate parameter (rearranging equation 2.7 for magnitude) and the preferred output shape, i.e. the value of N. To ascertain that no magnitudes fall below the set value of m_{min}, this is added to the randomly drawn magnitudes. In the case of synthetic data here, m_{min} is equivalent to the magnitude of completeness m_c because these catalogues are created to be 100% complete. This magnitude is easily converted into moments using $\log(M_0) = A + Bm$.

3.1.2 Modified Gutenberg-Richter (MGR)

The generation of moments following the MGR distribution is performed according to Vere-Jones et al. (2001) and equations 41 - 43 within Kagan (2002a). This is done by generating two random variables (R1 and R2) which are then used to generate two sets of moments (M1 and M2) and through minimising, we obtain one set of MGR moments which are then used. The function I created to do this takes input parameters N, β (equivalent b-value), the corner moment M_θ and the moment of completeness M_c.

Now that I have created magnitudes (converted to moments) sampled from a GR and an MGR distribution, I can use these in the next section to estimate further parameters.

3.2 Parameter estimation

In a real case, the underlying parameters are not known and we have to estimate these from the data. To inform this inversion, I start by constructing and analysing a single synthetic catalogue where we know the underlying, true parameters. By then simulating a larger number of synthetic catalogues using Monte Carlo methods, and inverting for the frequency-magnitude distribution (FMD) parameters such as the b-value, I estimate the precision of the inversion (the random error), and by knowing the actual underlying parameters, I can estimate the accuracy (systematic error or bias). The specific parameters used for generating the random catalogues are described in section 3.5.1. In the following, I describe the methods used to estimate crucial parameters in

both the synthetic and real catalogues.

3.2.1 Estimation of m_c

The completeness magnitude, m_c, is the smallest magnitude at which we can be confident 100% of earthquakes are detected (Rydelek and Sacks, 1989). For a long time, estimation of this value has led to discussion and a lot of work has been undertaken to identify the most accurate method to determine m_c (section 2.5.1.1). A robust estimate of m_c is essential, as other parameters – such as b and a – depend on this estimate, and hence could be biased and/or more uncertain if m_c is not optimal. A value of m_c which is too high will lead to undersampling, as much of the useful and complete data is discarded, leading to large parameter uncertainties, whereas a value of m_c which is too low will also lead to biased estimates of other parameters due to the use of incomplete data (Mignan and Woessner, 2012).

Many methods exist to determine m_c for a dataset (Mignan and Woessner (2012); Herrmann and Marzocchi (2020); Marzocchi et al. (2020)) depending on the type of data but also in practice the preference of the researcher. The two main classes of approaches (catalog- and network-based methods) have been discussed in section 2.5.1.1. While there are many possible methods, for simplicity, and because a fully accurate estimate of m_c is not the main purpose of the work here, I have used two main, relatively straightforward methods based on easily accessible catalogue data – the Maximum Curvature (MaxC, Wyss et al. (1999); Wiemer and Wyss (2000)) and the b-value stability (BVS, Wiemer and Wyss (2000); Cao and Gao (2002)) methods.

Originally, I started by using the MaxC method because the intention was to identify a minimum possible value of m_c leading to a maximum sample size (N), to then increase m_c, resulting in a decrease in N, and estimating how this affects the evaluation of parameters of the dataset and the resulting bias. The non-parametric MaxC method is based on the evaluation of changes in the FMD, specifically calculating the highest value of the first derivative of the cumulative frequency curve. However, it generally understimates m_c (Mignan and Woessner, 2012; Roberts et al., 2015), even when there

is sufficient data, which in return will result in artificially low b-values. I realised this when analysing the real catalogues which I have chosen, leading to the inclusion of incomplete data and returning particularly biased b-values compared to published estimates of both m_c and b.

Therefore, I have decided to use the BVS as a preferred method of m_c estimation for all following chapters. The BVS method relies on the assumption that b-value estimates increase for cut-off magnitudes less than m_c and remain constant for cut-off magnitudes greater than or equal to m_c. It selects m_c as the magnitude above which the b-value stabilises within error of the b-value. Using synthetic catalogue analysis, Roberts (2016) previously suggested that the BVS method was the most reliable and accurate method for determination of m_c, proposing that at least 500 events are required if the data is incomplete and at least 200 events within a complete catalogue ($> m_c$) for a robust estimate of the b-value. In overall agreement, Woessner and Wiemer (2005) showed that the BVS method substantially underestimates m_c for $N \leq 250$ and that the uncertainty decreases with increasing sample size. For the reason that the focus of this thesis is on larger datasets ($N > 250$), I decided against the use of MaxC in favour of the BVS method. However, in three specific cases (see Chapter 5), the BVS method returned very high m_c values, leading to very small numbers of N and in these cases I have also used MaxC and/or the GFT method (Wiemer and Wyss, 2000) to compare with the BVS result. The GFT is a parametric method based on fitting the FMD with a linear function. It calculates m_c by comparing the observed FMD with a synthetic one and determining the residuals between the observed and synthetic data falling within a 95% confidence window. This also returns systematically low values for m_c, albeit not as low as the MaxC method. As part of the analysis conducted in Chapter 5, I also adapted the BVS method to assume the MGR MLE instead of the GR MLE and applied this to all real catalogues considered in that chapter, to see whether there are differences in the estimated value of m_c when both of these methods are applied.

It is crucial to realise that both MaxC and BVS (and most existing m_c methods) rely on the assumption of the underlying distribution being GR, immediately suggesting an

introduction of bias when performed on data for which the underlying distribution is unknown. However, this has shown that there is no 'one size fits all' option for choice of m_c method. Ideally, when looking at only one catalogue, the data should be assessed and visualised to determine the best option for estimating m_c. Roberts et al. (2015) showed that specific FMDs prefer different methods of m_c determination – that sharp-peaked incremental frequency data will return the most robust value when MaxC is used but that BVS is the preferred method for broad-peaked data, which is more common. Accordingly, a single method might not be best for another catalogue. For consistency, here the same method (BVS) is adopted for all case studies. In addition, it is assumed that the m_c of real catalogues is constant in both space and time, even though we know that this assumption does not generally hold (Mizrahi et al., 2021), for example due to limits of station coverage or event seismograms overlapping during aftershock sequences. Godano et al. (2014) agrees that m_c should generally not be estimated as one single value by showing that the b-value changes continuously with a cut-off magnitude threshold even at large cut-off values. In this work, the focus lies on increasing m_c to establish the resulting changes and bias in the b-value and hence not just one value of m_c is used.

Having outlined how m_c is estimated, the next two sections describe possible methods for estimating the model parameters b (or equivalent β) and m_θ (equivalent M_θ).

3.2.2 Least-squares regression

The method of using least-squares for the estimation of the b-value is a surprisingly commonly applied method, even though it is known to introduce substantial bias to the estimate (Aki, 1965) and is therefore not used in this thesis (section 2.5.1.2). I include a discussion here only for completeness. Naylor et al. (2010) explain very well, that conventional least-squares regression will return a biased b-value estimate because not all of the points in an FMD carry the same weight – often the points (or bins) in the tail, correspond to a single, large event, whereas the points at smaller magnitudes are likely

to contain tens, hundreds or even thousands of events. Furthermore, the variance is asymmetrically distributed at large magnitudes and therefore not Gaussian, which is a required assumption for this method. Nonetheless, many studies still use least-squares (see Table 1 within Roberts et al. (2015)) to estimate the b-value and it is crucial to be aware of this and assess outcomes accordingly. The following, preferred maximum likelihood method is the method used for parameter estimation in this thesis.

3.2.3 Maximum likelihood estimate (MLE)

The maximum likelihood estimate is a probabilistic method used to estimate the parameters of a distribution by maximising a likelihood function. The point which maximises the likelihood function is then called the maximum likelihood estimate (MLE) which is the point at which the observed data is the most probable. This is the preferred and standard method used in seismic hazard analysis (Mignan and Woessner, 2012). The assumptions are that the data is exponentially distributed as in equation 2.3 and that the maximum magnitude is at infinity. Using the MLE as in equation 2.7 to obtain the parameters of the GR distribution is common. The error estimate in b is given as $\sigma_b = \frac{b}{\sqrt{N}}$. Equations 4.4 and 4.5 show in detail how β and M_θ are determined.

3.3 Model selection

To select the best statistical model from a set of candidate models, formal model selection criteria are commonly applied. Choosing an appropriate statistical model is complex and challenging (Main et al., 2011). These are based on modern forms of Occam's razor which states that 'that which is accomplished with fewer is accomplished in vain with more'. In model selection this translates to selecting 'the simplest model consistent with the data'. A wide range of criteria exist, depending on what types of models are being used. For the specific purpose of this work, the aim is to identify which model (GR or MGR) is preferred for a specific dataset. In this section I introduce three commonly used information criteria (Likelihood Ratio, AIC and BIC defined below) devised to

address such questions and explain the preference for the BIC in this work.

3.3.1 Likelihood Ratio

The likelihood ratio is a goodness-of-fit test between two competing models, based on the ratio of their likelihoods. It is a commonly used method in statistical model selection and possible for both nested and non-nested models (Lewis et al., 2011). This method generally leads to a preference of a more complicated model and is prone to 'overfitting', i.e. starting to fit the noise. It cannot be used for models which have different number of parameters and is therefore not an option in this thesis.

3.3.2 Akaike Information Criterion (AIC)

The AIC (Akaike, 1974) is a penalised likelihood criteria comparing competing models. It selects the model which most adequately describes an unknown, high dimensional reality, meaning reality is never in the set of candidate models that are being considered.

The AIC (to base 10) is given as:

$$AIC = \frac{1}{\log(e)}(2k - 2\log(L)) \tag{3.1}$$

where k is the number of model parameters (which, including the variance, is two for the GR law and three for the MGR law) and L is the likelihood function. A lower AIC means a model is considered to be closer to the truth and seen as a higher level of empirical support than models with high AIC values. However, a downside is that the AIC is often too liberal and frequently prefers a more complex, wrong model over a simpler, true model.

3.3.3 Bayesian Information Criterion (BIC)

Just as the AIC, the BIC is another penalised likelihood criteria used to compare models (Schwarz, 1978) and tries to find the true model (the process that generated the data) among the set of candidate models. This is preferred over other commonly applied criteria such as the AIC or the simple Likelihood Ratio (which does not consider model

complexity at all) because the BIC penalises model complexity (i.e. the number of free parameters in the model) more heavily and hence the differences in models are highlighted more strongly. For this reason, the BIC is the criterion I will use throughout this thesis. Furthermore, synthetic catalogue examples have shown that the BIC is a more accurate discriminant for larger datasets, where $N > 46$ (Main et al., 1999). In this thesis, all of the catalogues analysed fulfil this criterion. I have used a modified version of the BIC (to base 10) according to Leonard and Hsu (1999):

$$BIC = \frac{1}{\log(e)}\left(-2\log(L) + k\log(N)\right) \tag{3.2}$$

where L is the likelihood function, N is the total number of events and k is the number of model parameters. Subtracting the BIC (ΔBIC_{MGR-GR}) of the two models returns the preferred model for each catalogue as shown in Figure 3.1. When the resulting value is positive, the GR distribution is preferred and when it is negative, the MGR distribution is preferred. This is an arbitrary example where I use 50 simulated catalogues all drawn from an MGR distribution with the same parameters, applying both GR and MGR MLEs and subtracting them to determine the preferred model. In a few cases (~ 7 catalogues), GR is preferred due to the random sampling nature, whereas all other catalogues suggest (the correct) MGR preference, especially at maximum number of events (1000). When catalogues are thinned to remove events, more catalogues prefer GR (i.e. at ~ 100 events).

The use of the BIC and the resulting ΔBIC plays a major role in discriminating preferred models and identifying truly high b-values later on in this thesis.

3.4 The assumption of a finite maximum magnitude

As I have described in section 3.2.3 and throughout the literature review in Chapter 2, the GR law on which the work in this thesis is based, assumes that the maximum magnitude possible, but possibly not yet observed, is infinite. This is a limitation in practice and in the application to real catalogues. Therefore, to test this assumption, I suggest two analytical methods to correct for this assumption by including a finite

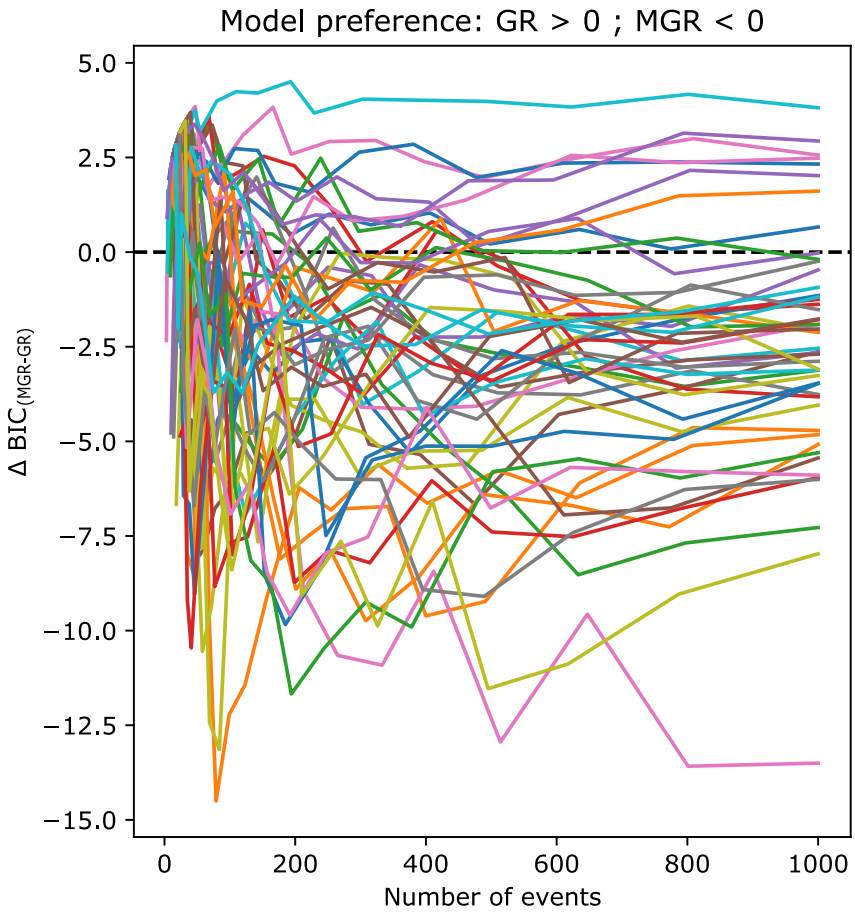

Figure 3.1: ΔBIC as a function of the number of events. GR is preferred > 0 and MGR is preferred < 0.

maximum magnitude in both the cases of the GR and MGR laws. Due to the extent of these derivations and analysis, they are shown separately in Chapter 6 rather than here.

3.5 Catalogue data

Synthetic earthquake catalogues are used primarily to test the methods applied and to identify potential biases and further issues that may arise when analysing real earthquake catalogues. The big advantage is that the underlying distribution and parameters are known, so we can assess accuracy as well as precision in the model selection and fitting. A full summary of the chosen real catalogues can be found in Table 3.1 at the end of this chapter and includes raw values obtained from the incomplete catalogues introduced in this section. Summary tables are also in Chapters 4 and 5 to then include the parameters estimated in those chapters, i.e. the b-value and m_θ.

3.5.1 Synthetic earthquake catalogues

Synthetic earthquake catalogue magnitudes are first created as described in section 3.1. This began with a fixed set of parameters (section 4.4.2) to replicate smaller magnitude earthquake catalogues as they may occur and be recorded in high-density networks for shallow volcanic or induced seismicity. The true, underlying b-value was initially set to 1.0, which is common of tectonic seismicity (Frohlich and Davis, 1993). All parameters (N, m_c, b, m_θ) are then varied to identify the extent of changes in biases, shown and discussed in Chapter 5. To simulate many randomly sampled catalogues with the same parameters and from that obtain uncertainties and variability between individual simulations, I use Monte-Carlo (MC) methods. More specific information on the generation of synthetic catalogues will be given in Chapter 4 and I now introduce the real catalogues used in this thesis, for examples selected where the data includes tectonic, volcanic and induced seismicity.

3.5.2 Tectonic earthquake catalogues

To test the GR and MGR hypotheses against real data, I have initially chosen three tectonic earthquake catalogues, individually introduced in the following subsections. Generally, these have all been selected due to their abundance of (accessible) data and dynamic range, as well as to showcase events from different geographical locations and seismicity settings. They will be further discussed in their respective analyses in sections 4.4.3 and 5.2.1.

3.5.2.1 Central Italy

The Central Apennines are a very tectonically active area, with ample seismicity from both smaller and larger magnitude ranges. This catalogue has been obtained from the Istituto Nazionale Geofisica e Vulcanologia (INGV, `http://terremoti.ingv.it/en`), beginning on January 1st 1990 until August 24th 2016, i.e. just before the $m_W = 6.2$ Central Italy earthquake on August 24th, 2016. This was the first catalogue I started working with and because the Central Italy earthquake sequence was still ongoing, I decided to not include the largest events to avoid potential bias recently introduced due to several large events at the end of the catalogue. A mixture of m_L, m_W and m_d (earthquake duration magnitude) values are provided. Magnitudes are used as given and no conversion between scales has been attempted as part of this thesis. Notably, the seismic network and sensitivity in this region improved substantially in 2005, as seen in Figure 3.2. Hence, this catalogue was split into two sub-catalogues which also resulted in very different values of m_c (section 4.4.3.1). The two sub-catalogues span the times 1990 - 2005 (hereafter labelled INGV 1990) and 2005 - 2016 (INGV 2005).

3.5.2.2 Reykjanes Peninsula, Iceland

The Reykjanes Peninsula is an area in the southwest of Iceland and the continuation of the submarine Reykjanes Ridge (as a part of the Mid-Atlantic Ridge), experiencing many tectonic earthquakes and hence providing us with much data for analysis. However, it

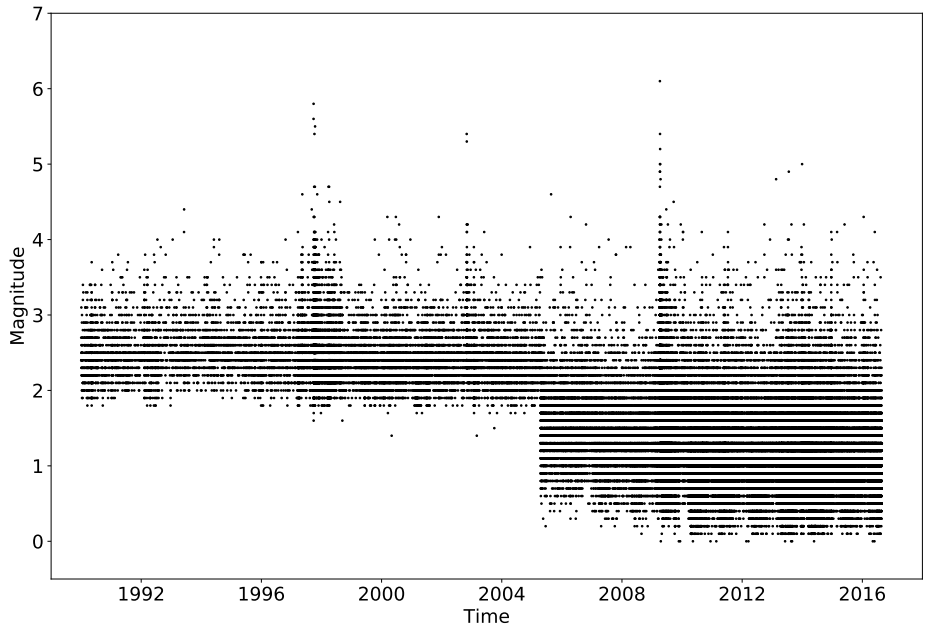

Figure 3.2: Earthquake magnitudes shown as a function of time for the full INGV catalogue. Note the clear drop in magnitude recordings during 2005.

is also marked by volcanic systems, which until recently (i.e. the Fagradalsfjall eruption between May and September 2021) have been fairly inactive. For this reason, I have put this catalogue in the 'tectonic' section. The data is obtained from the Icelandic Meteorological Office (IMO) under `http://hraun.vedur.is/ja/viku/` and I look at data obtained between January 1995 and January 2021. This catalogue offers two different magnitude estimates (more in section 5.2.1.1) both with ~ 50 000 events and a maximum observed magnitude of 5.61.

3.5.2.3 Southern California

Southern California is another tectonically very active region with a significant number of active faults, resulting in regular events of magnitudes > 5.0. Some of the larger recent events have included the 1992 $m_W = 7.3$ Landers earthquake and the 1999 $m_W = 7.1$ Hector Mine earthquake (Felzer et al., 2002) as well as the recent 2019 $m_W = 7.1$ Ridgecrest event (Chen et al., 2020). I chose Southern California as another tectonic catalogue due to its active seismic activity and the abundance of recorded events by

the Southern California Seismic Network (SCSN), archived by and obtained from the Southern California Earthquake Data Center (SCEDC, 2013, `https://service.scedc.caltech.edu/eq-catalogs/date_mag_loc.php`). This catalogue runs between January 1960 and May 2021 with over 750 000 events and is therefore the largest catalogue selected as part of this thesis. The maximum observed magnitude is 7.3.

3.5.3 Volcanic earthquake catalogues

Seismic events associated with volcanoes are often restricted to the smaller magnitudes, but still have the potential to be hazardous for populations. Therefore, it is crucial to assess the hazard and risk as best as possible but it is also important to note that b-values can also be used to make inferences about the physical behaviour driving such volcanic systems. For example, they could provide information on whether seismicity is driven by tectonic processes ($b \sim 1.0$) or by fluid movement ($b > 1.0$) (Roberts et al., 2016) and biased or inaccurate estimates of the b-value may also lead to problems for such physical studies. Therefore, to test the GR and MGR hypotheses on earthquakes associated with volcanic areas, I have chosen the four catalogues below. At this stage of the modelling, no efforts have been made at trying to discriminate between tectonic and volcanic seismicity in any of these catalogues.

3.5.3.1 Mount Etna, Italy

Mount Etna is an active stratovolcano in Sicily, Italy. It is one of the world's most active volcanoes, including many seismic events occurring both during and outwith eruptive phases. This catalogue was part of a published study in Chapter 4 where it is discussed in more detail.

3.5.3.2 El Hierro, Canary Islands

El Hierro is another catalogue chosen primarily for its volcanic seismicity within the volcanic island chain of the Canary Islands and also analysed in the following Chapter 4 in more detail.

3.5.3.3 Kilauea, Hawaii

Kilauea is an active shield volcano in Hawaii, most commonly known for its nearly continuous eruption between 1983 and 2018. Therefore, most of the earthquakes associated with this volcano are due to magma moving within the volcano and during eruptions, however, there are likely to be shallow tectonic earthquakes in the mix along faults in the crust and upper mantle (USGS). The catalogue used ranges from 1960 up to 2020 with nearly 60 000 events and a maximum magnitude of 4.7, which is the lowest of all volcanic catalogues used here. Both this catalogue and the following (Long Valley) have been obtained from the United States Geological Survey (`https://earthquake.usgs.gov/earthquakes/search/`).

3.5.3.4 Long Valley, California

The Long Valley Caldera is located in eastern California and formed as a result of a 'super eruption' about 760 000 years ago (USGS; Barton et al. (1999)). It is one of the planet's largest calderas. The caldera remains thermally active, leading to ongoing seismic events and deformation within the caldera and as a result, has a very high threat potential (Hill, 2006). Therefore, I assume that the majority of events recorded here are volcanic in origin, while the tectonic causes of volcanism and activity around Long Valley are still largely unknown, as it does not lie above a hotspot and is also not the result of subduction.

The catalogue used here has been obtained from the USGS and spans 1983 - 2020. Not many events were recorded prior to 1980 (see section 5.2.2.1 for details and results) but known earthquake swarms occurred at Long Valley between 1978 - 1983, 1990 - 1995, 1996 and 1997 - 1998. Even though there is available data from 1963, the assessed part of the catalogue only spans January 1st, 1983 onwards due to the lack of sensitivity and ability to record small events prior to this period. Including the period before this date will substantially affect the estimate of m_c and is another reason as to why it is excluded (Wiemer et al., 1998). The largest observed magnitude is 5.4 and this is the

volcanic catalogue with the most events (over 120 000) used in this thesis.

3.5.4 Induced earthquake catalogues

In this section, I introduce the induced seismicity catalogues which I have chosen to analysis in this thesis. It has proven much harder than expected to obtain large datasets of induced events, especially with maximum magnitudes significantly above m_c, because induced seismicity does not tend to include events as large as might be expected from the background tectonic seismicity. This reinforces the problem with analysing parameters, in particular the b-value, but highlights the importance of a critical analysis of the effect of 'narrow' dynamic range in evaluating the precision, accuracy and hence utility of model fits to such data.

3.5.4.1 The Geysers, California, USA

The Geysers geothermal complex in northern California is a rare dry steam field and the world's biggest single geothermal field with more than 350 wells spread across 78 km^2 (Leptokaropoulos et al., 2018a). It is located in the Mayacama Mountains, around 115km north of San Francisco and has been operating since 1960, with seismicity being monitored since 1972. The vast majority of earthquakes within this area are attributed to extraction of steam and the re-injection of condensate and wastewater into the reservoir (Henderson et al., 1999), although due to the nature of the setting, there may be some tectonic earthquakes within the recordings. I have included more information on this catalogue in the publication in Chapter 4.

3.5.4.2 Oklahoma, USA

In 2009, the frequency of earthquakes detected in Oklahoma increased rapidly. This was quickly found to be attributed to wastewater disposal (Keranen et al., 2013, 2014; Hough and Page, 2015; McClure et al., 2017). An event of $m_W = 5.8$ occurred on September 3rd, 2016 leading to major damage of buildings in Pawnee, 25km away from the epicenter, (Foulger et al., 2018) including several $m > 3.0$ aftershocks (Chen

et al., 2017). Due to Oklahoma's geological setting, which does not normally produce felt tectonic earthquakes on such timescales, the increase of events in Oklahoma over the last decade has received widespread coverage in the science community and is therefore a particularly interesting case to look at. For the work here, the catalogue used is directly obtained from the Oklahoma Geological Survey (OGS, `https://ogsweb.ou.edu/eq_catalog/`) and includes earthquakes between 1980 and 2017 (Walter et al., 2020) with the largest magnitude being the Pawnee earthquake in 2016. One would expect the threshold of m_c to vary (similar to INGV, section 3.5.2.1) due to improved network coverage and detection thresholds since the onset of increased seismicity in 2009. This is shown in Figure 3.3. However, in contrast to the INGV catalogue, I have decided to not split this catalogue because of its more recent improvement in detection and therefore more limited event data until the end of this catalogue in 2017.

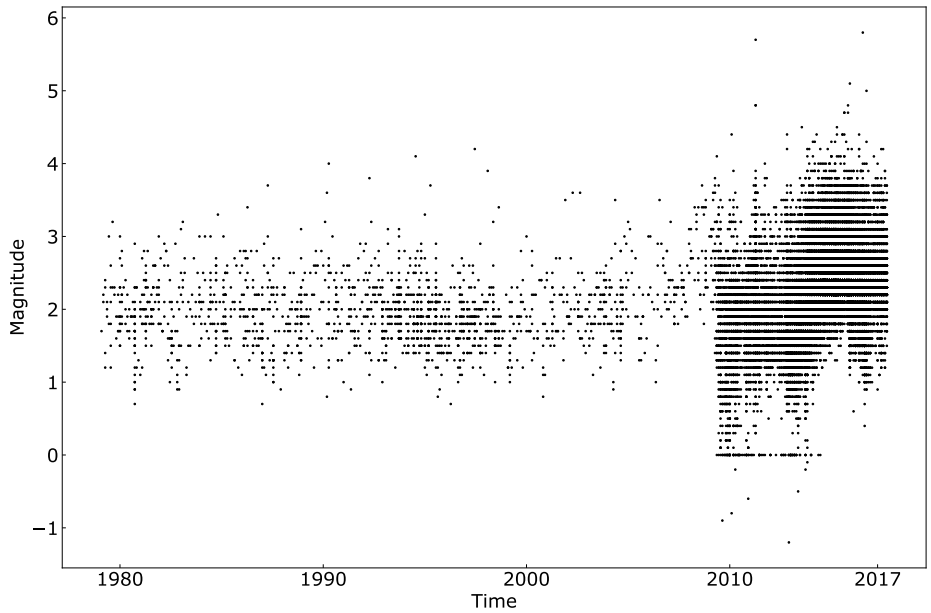

Figure 3.3: Earthquake magnitudes shown as a function of time for the full Oklahoma catalogue. Note the clear drop in magnitude recordings from 2009 onwards.

3.5.4.3 Preston New Road (PNR), UK

In October 2018, Cuadrilla Resources began hydraulic fracturing (HF) at Preston New Road (PNR) in Lancashire, which, throughout two operations in two wells (PNR-1z and PNR-2) was accompanied by microseismicity, the data for which were made publicly available in 2019 (`https://www.ogauthority.co.uk/exploration-production/onshore/onshore-reports-and-data/`). However due to equipment limitations, the magnitudes are problematic and a corrected magnitude catalogue provided by Dr. James Verdon is used instead (section 5.2.3.1). Hydraulic fracturing causes seismicity because the injection of high pressure fluid into rocks reduces the effective stress, and hence brings the stress state nearer the failure envelope in shear stress-effective normal stress space (section 4.7 within Fossen (2010)). This generates cracks which grow and produce microseismic events, often only recorded by specialist networks, including downhole monitoring to detect smaller events, deployed during these operations. This HF data at PNR undoubtedly provides a rare opportunity to analyse induced seismicity data in response to HF and has been used in several recent publications to improve our understanding of the processes involved in induced seismicity (Clarke et al., 2019; Mancini et al., 2019, 2020; Kettlety et al., 2021). It is unfortunate that microseismic data is so scarcely available (especially in real or near real time) because the studies cited above allow new and independent analysis of the data, and generate new understanding that may eventually lead to more effective risk reduction strategies in the long run.

3.6 Perspective

The methods and data detailed in this chapter have been used throughout the work presented in this thesis. The main focus of the work is on the model selection and quantification of likely accuracy and precision of the resulting optimal model parameters. Therefore, none of the catalogues introduced have been declustered and no attempts have been made in distinguishing mainshocks, aftershocks and foreshocks in any given seismic setting. Magnitude scales were not adjusted but used as given in the various

Table 3.1: This table indicates the important details for the real catalogues, including type of seismicity, the period during which the catalogue events were observed, the total number of events in this observed period (N), the number of events above m_c, the observed maximum magnitude within the catalogue and the value of the b-value stability m_c. WD - Wastewater Disposal, HF - Hydraulic Fracturing

Catalogue	**Seismicity Type**	**Observation Period**	**N**	**obs.** m_{max}
INGV 1990	Tectonic	01/01/1990 - 15/04/2005	17 166	5.8
INGV 2005	Tectonic	16/04/2005 - 24/08/2016	114 847	6.1
Reykjanes m	Tectonic	07/01/1995 - 31/01/2021	49 441	5.61
Reykjanes m_L	Tectonic	07/01/1995 - 31/01/2021	50 379	4.96
Southern California	Tectonic	01/01/1960 - 07/05/2021	767 380	7.3
Etna	Volcanic	19/08/1999 - 29/08/2016	15 688	5.7
El Hierro	Volcanic	30/11/2001 - 22/01/2017	22 494	5.1
Kilauea	Volcanic	01/01/1960 - 31/12/2020	59 128	4.7
Long Valley	Volcanic	01/01/1983 - 31/12/2020	122 748	5.4
The Geysers	Geothermal	18/03/1972 - 31/07/2018	281 572	5.01
Oklahoma	WD	01/01/1980 - 31/10/2017	28 219	5.8
Preston New Road PNR-1z	HF	15/10/2018 - 17/12/2018	38 452	1.9
Preston New Road PNR-2	HF	15/08/2019 - 23/08/2019	56 008	2.9

catalogues. Naturally, this will lead to limitations including uncertainties, problems with data quality and systematic biases which are addressed in more detail in the following results chapters.

Chapter 4

Biases in estimating b-values from small earthquake catalogues: How high are high b-values?

This chapter takes the form of a paper in *Geophysical Journal International*, which at the time of thesis submission has been published (Geffers et al., 2022). It has been re-formatted to follow the thesis style and no further changes have been made. Further (unpublished) analyses to this chapter have been included at the end in section 4.9. I first introduce the current science and understanding of seismic b-values in a literature review split into an introduction and a theory section. I then explain the methods applied to tackle the problems identified as research gaps in the literature review, as well as the rationale for including synthetic data and real earthquake catalogues. I then present the results obtained from the analysis, finishing with a discussion and conclusion. The results suggest that the dynamic range and the assumed distribution are major factors in data which shows high b-values. I add a reflection at the end of this chapter to place this work within the larger context of the thesis.

4.1 Abstract

The Gutenberg-Richter (GR) b-value describes the relative proportion of small to large earthquakes in a scale-free population, and is a critical parameter for probabilistic estimation of seismic hazard. At low magnitudes, the scale-free behaviour breaks down below the magnitude of completeness m_c due to censoring of the data, when the instrumentation used to construct the catalogue is incapable of completely recording all earthquakes in the study region above the background noise. At high magnitudes, it must also break down because natural tectonic and volcanic processes are incapable of an infinite release of energy. This breakdown at large magnitudes is commonly modelled as an exponential roll-off to either the incremental or cumulative GR distribution. This introduces an extra parameter and hence requires relatively more data to justify the additional model complexity. For tectonic seismicity, the estimated b-value is commonly close to unity. In contrast, studies of volcanic and induced seismicity often report significantly higher estimates of the b-value, albeit using relatively small datasets – both in sample size and dynamic (magnitude) range for data above m_c. Here, using synthetic data, we show that when we have low dynamic range, it is statistically challenging to test whether the sample is representative of the scale-free GR behaviour or whether it is controlled primarily by the finite size roll-off. We then explore the potential biases that arise when the data quality does not allow this distinction to be made and what the implications are for interpreting studies that have high estimated b-values. We find that systematically higher b-values than those used to generate the synthetic data are regularly obtained when assuming the wrong model and when having a too high m_c, resulting in too small catalogues. This is important because it changes our understanding of the accuracy of elevated or variable b-values in catalogues of different dynamic ranges, and quantifies the likely bias in the inferred b-value compared to the underlying true distribution and its associated uncertainty. Finally, we recommend steps to minimise this bias.

4.2 Introduction

Earthquake hazard depends fundamentally on an estimate of the probability that an earthquake will occur within a given time window and geographic area and above a certain magnitude. Estimating this earthquake recurrence rate in space, time and magnitude, combined with attenuation rules translating the properties of the source population to strong ground motion remote from the source, allows assessment of seismic hazard for the specific region, expressed as maps of the likely levels of strong ground motion a building may be susceptible to (Reiter, 1991). Such maps are the basis for risk mitigation through informed development of regulations and actions in land use and infrastructure planning. While the seismic risk has been increasing due to increased exposure, the seismic hazard has remained fairly constant (averaged over long time periods) because of the relative stability of global dynamics (Main, 1996).

The maximum earthquake magnitude is an important factor in seismic hazard analysis because it constrains the properties of rare, extreme events (Main, 1995) and their effects and consequences on societies. Reiter (1991) defines the maximum possible earthquake magnitude as the upper bound determined by earthquake processes, however unlikely, associated with a specific earthquake source. Estimates of earthquake recurrence rates are generally obtained from data in the form of seismic catalogues and historical seismicity, often assuming a Gutenberg-Richter (GR) frequency-magnitude distribution (FMD) defined in equation 4.1. This implies the frequency-size distribution of source rupture areas scales as a power law, indicating self-similarity on all scales (Kanamori and Anderson, 1975). In a review of the then state of the art, Main (1996) found that the GR law applies only within a finite scale range both in model and natural seismicity due to natural fractal sets being constrained to lower and upper scale-lengths, challenging the assumption of the GR law operating at all scales. Currently, the form which the FMD takes at larger scales is still extremely uncertain (Kagan, 1999; Bell et al., 2013). Physically, an upper bound can be expected as this is limited by factors such as tectonic deformation rate (Main, 1996), regional stress regimes, seismic moment

finiteness (Kagan, 1997) and fault zone geometry.

Kagan (1999) showed that identifying all parameters within the GR distribution allows the assessment of key attributes of earthquake recurrence rates (e.g. return period for events at all sizes and the maximum expected magnitude). However, the finite deformation rate means it is not possible to extrapolate the GR law, without limits, to determine the recurrence rate of the rare, large events. This is a fundamental problem with the open-ended GR distribution in the common case of its slope, b, being less than $\frac{3}{2}$ (Main, 1995). Nonetheless, the GR law is the standard underlying assumption for the majority of methods used on earthquake data today. It provides crucial information such as the seismicity rate of a region at all event sizes, quantified by the exponent of the FMD, known as the b-value, which is generally ~ 1.0 for shallow tectonic events (Frohlich and Davis, 1993). However, Bell et al. (2013) used the global Harvard CMT catalogue to show that its FMD is inconsistent with an unbounded GR relation in time, further adding to the uncertainty in form of the GR relation for rare, large events. They have shown that between record-breaking events, the preferred model gradually converges to the modified GR (MGR) relation, which combines a power-law distribution of seismic moment for smaller earthquakes, with an exponential roll-off or taper at high magnitudes characterised by a 'corner moment' M_θ (Main and Burton, 1984; Kagan, 1991). Frohlich and Davis (1993) also investigated the b-value in teleseismic data for tectonic events and found that it may vary by 30% when calculated using different methods or catalogues and hence cannot always be constrained to a single optimum value. The MGR gives a much better fit to the global catalogue before the 2004 Sumatra event according to the Bayesian Information Criterion (BIC) but the occurrence of this large magnitude earthquake changed the balance of the BIC statistics, making the unbounded GR power law more suitable. This suggests that standard statistical tests may be unstable in the face of undersampled data. For example, Zöller (2013) argues that the findings of Bell et al. (2013) are a result from biased maximum likelihood estimates (MLE) of M_θ in strongly undersampled models, and that convergence to a specific distribution for the global earthquake catalogue is not possible with less than

at least 200 years of homogeneous recording of global seismicity.

Several studies on both volcanic and induced seismicity have shown that the b-value appears to be elevated above 1.0 in these cases (Jacobs and McNutt, 2010; Murru et al., 1999, 2005; Ibáñez et al., 2012; Jolly and McNutt, 1999; Wiemer et al., 1998; Novelo-Casanova et al., 2006). Specifically, Ibáñez et al. (2012) and Jolly and McNutt (1999) both used the least-squares method to fit the data, which is known to introduce significant bias (Aki, 1965; Naylor et al., 2010). Table 4.1 summarises some of the b-values obtained in previous studies, also showing that often the sample size and/or the magnitude dynamic range (here defined as the range between the magnitude of completeness, m_c (Rydelek and Sacks, 1989), and the maximum magnitude observed, thus strictly a logarithmic dynamic range) used are very small. Roberts et al. (2015) specifically investigated b-values obtained from volcanic seismicity. They show that – predominantly due to small sample size and lack of data, and the associated narrow dynamic range of magnitudes, as well as the major variability in the methods used to assess the cutoff magnitude for complete reporting – the b-value estimates can be both highly variable and biased. They found that volcanic earthquakes can have high b-values in some cases, but are also often indistinguishable from those of tectonic earthquakes, within error. Using random temporal sampling, Roberts et al. (2016) established that during eruptive phases at El Hierro, b-values can 'jump' between meta-stable values rather than evolving gradually with time. This was suggested to be due to switching between different parts of the volcanic system as opposed to spatially uniform sampling of a more gradually varying underlying process. Furthermore, Roberts et al. (2016) suggested that errors in b-values are often large, underestimated or biased due to the usage of a single value of m_c.

Table 4.1: Summary of selected studies on volcanic b-values. N is the number of events in a catalogue. dyn. r indicates the dynamic range of the catalogue and b_{typ} is the typical value (after Roberts et al. (2015) who selected these by eye). MLE is the maximum likelihood estimation, LS the least-squares method and WLS the weighted least-squares fit. (Adapted from Roberts et al. (2015))

Reference	**Volcano**	**Observation Period**	**N**	**dyn. r**	m_c	**Method** b	b_{typ}
Jacobs and McNutt (2010)	Augustine	2000 - 2006	100	-	0.1	MLE	1.4
Ibáñez et al. (2012)	El Hierro	19.07.11 - 16.09.11	7000+	1.3 - 2.7	1.3	LS	1.57
Murru et al. (1999)	Etna	1990 - 1997	50 (450)	2.5 -	~2.5	MLE/WLS	1.5 ± 0.3
Murru et al. (2005)	Etna	Jul - Aug 2001	50 (173)	1.5 - ?	2.6	MLE	1.5 ± 0.28
Jolly and McNutt (1999)	Mageik	Sep 1996 - Apr 1997	~300	-	0.7	WLS	1.56
Wiemer et al. (1998)	Long Valley	1989 - 1998	150	1.3 -	1.3	MLE	1.4
Novelo-Casanova et al. (2006)	Popocatepetl	Dec 2000 - Jan 2001	20 (225)	1.9 - 3.3	1.9	MLE	1.7

In the case of induced seismicity, Table 4.2 shows b-values obtained in previous studies, again showing that both the sample size and magnitude dynamic range used are often small, not stated or vague. Grünthal (2014) showed that inferred b-values regularly depended on the type of induced seismicity. Further studies have suggested that b-values in hydraulic fracturing are often $\gg 1.0$ and even closer to 2.0 (Maxwell et al., 2009; Eaton et al., 2014). However, several induced seismic catalogues have also shown b-values similar to those of tectonic seismicity (especially in natural gas exploitation and salt mining) as well as in some geothermal systems (Henderson et al., 1999; Bachmann et al., 2011; Grünthal, 2014; Schoenball et al., 2015).

Several of the studies described above have investigated b-values during or around volcanic eruptions and ongoing induced sequences, using limited, short-term data and assuming an underlying GR distribution only. Many b-value studies have been focused on global (tectonic) catalogues, but we know that these same, global rules may not apply to seismicity on smaller magnitude scales often found on local and regional scales. Here we estimate the reliability, accuracy and precision of the FMD parameters for tectonic, volcanic and induced seismicity. We use synthetic datasets with known underlying parameters to create a method to identify any systematically high b-values obtained due to the use of GR inference methods on underlying GR or MGR data, including the extent to which artificially high b-values can result from the exponential roll-off of the MGR, and when we can distinguish this artefact from truly high b-values. We then apply the understanding generated to real tectonic, volcanic and induced earthquake catalogues. We show that the incorrect assumption of the underlying distribution and lack of data can lead to artifically high b-values, but also in some cases to particularly low b-values, for example due to incompleteness of data. Finally, we demonstrate that a robust estimate of the optimal underlying b-value requires a thorough understanding of the sensitivity of the trade-off between catalogue sample size, magnitude dynamic range, the choice of the underlying distribution and the desired accuracy of the b-value.

Table 4.2: Summary of selected studies giving elevated b-values for different types of induced seismicity. EGS - Enhanced Geothermal Systems, F - Hydraulic fracturing, G - Geothermal, M - Mining, WD - Wastewater Disposal. Methods are maximum likelihood estimate (MLE) and least-squares (LS). (Adapted from Mousavi et al. (2017))

Reference	**Type**	**Location**	**N**	**dyn. r**	m_c	**Method** b	b_{typ}
Bachmann et al. (2012)	EGS	Basel, Switzerland	150	0.1 - 3.2	0.7 - 1.0	MLE	1.1 - 1.6
Vermylen and Zoback (2011)	F	Fort Worth, TX, USA	4485	-	-2.5	LS	1.5 - 1.9
Henderson et al. (1999)	G	Geysers, CA, USA	30 000 (200)	0.5 - ?	0.5	-	< 1.0
Grünthal (2014)	G	Tripoint GER-SUI-FRA	-	2.0 - 3.2	2.0	MLE	1.94 ± 0.21
Grünthal (2014)	M (Cu)	Legnica, Poland	-	- 4.5	-	MLE	2.13 ± 0.22
Mousavi et al. (2017)	WD	Guy, AR, USA	> 17309	-0.67 - 4.4	0.5	MLE	1.06 ± 0.05
Langenbruch and Zoback (2016)	WD	Central OK, USA	-	3.0 - ?	-	MLE	1.41

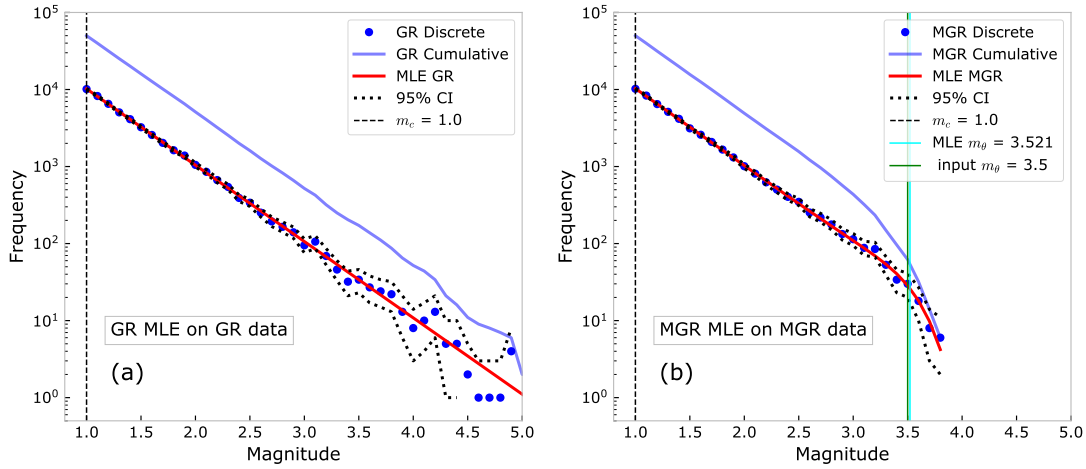

Figure 4.1: Synthetic incremental (blue circles) and cumulative (blue line) frequency-magnitude distributions for a catalogue of 50 000 events, assuming a complete catalogue (to the right of the left-most vertical line) for both (a) the GR (equation 4.2) and (b) the MGR (equation 4.3) distributions, including 95% confidence intervals (CI).

4.3 Theory

4.3.1 Frequency Magnitude Distributions

The frequency-magnitude distribution (FMD) describes either the incremental number of events within a magnitude window or the cumulative number of events above a series of magnitude thresholds (Figure 4.1). As such, real catalogues are commonly presented in a histogram. This distribution of frequency is most often represented as an exponential in magnitude (Gutenberg and Richter, 1944) in the form

$$\log F = a - bm \tag{4.1}$$

where F is the cumulative number of events at a magnitude greater than or equal to m in a given population, and a and b are constants describing the seismicity rate and the proportion of small to large earthquakes, respectively. The b-value is not just used in earthquake hazard analysis – it is often used to interpret material properties, such as heterogeneity, or to infer the state of stress from earthquake populations (Mogi, 1962; Scholz, 1968), or in non-destructive testing of the integrity of building and infrastructure

components (Colombo et al., 2003). Event magnitudes are often rounded to one decimal place, leading to issues with precision and if not accounted for, are a source of bias (Tinti and Mulargia, 1985). It is therefore simpler to work with the seismic moment, M_0, which is the product of source rupture area, average slip and the shear modulus of the surrounding medium. The magnitude of an event is related to the logarithm of the seismic moment. With this change of variables, the seismic moment-frequency relation has the form

$$F(M_0) \sim M_0^{-\beta} \tag{4.2}$$

where $\beta = \frac{2}{3}b$ (Turcotte, 1997; Kagan, 2002a); β being the equivalent of the b-value in the magnitude-frequency relation, where the factor $\frac{2}{3}$ results from a log-linear model relating the moment-magnitude to the moment $\log(M_0) = A + Bm$. Hanks and Kanamori (1979) have taken this as a definition of moment magnitude with A = 9.1 and B = 1.5. The β-value in equation 4.2 has a theoretical basis in the scale-free geometry of fault ruptures of different sizes and the filtering effect of the seismometer at different frequencies (Kanamori and Anderson, 1975). The form of equation 4.2 is a power law, and is diagnostic of scale-free behaviour because it has no characteristic size. In this work we will use the capitalised M to refer to values on the moment scale and m when they are quoted on the magnitude scale – they are equivalent and we can always convert between them using the log-linear relation. We will generally discuss parameters on the magnitude scale as the values are more intuitive, but all calculations on the synthetic data are done on the moment scale.

The scale-free form of the GR distribution (equations 4.1 and 4.2) is unbounded at the upper end which implies the potential for an unphysically infinite energy release. However, the finite size of the seismogenic lithosphere and/or the finite tectonic deformation rate necessitates a finite yet uncertain upper bound. Within models, this can be achieved by applying an exponential taper to equation 4.2 at large seismic moments, thus modifying the GR law as derived by Main and Burton (1984), from imposing a constraint of finite moment release rate, resulting in a modified gamma

distribution for the probability density function $p(M)$. In many applications the fit is to the probability density function for the tapered or modified Gutenberg-Richter (MGR) law, defined instead by a cumulative frequency distribution of the form

$$F(M_0) \sim M_0^{-\beta} \exp(-M_0/M_\theta) \tag{4.3}$$

where M_θ is a 'corner moment' where the cumulative frequency $F(M)$ has dropped to a value $\frac{1}{e}$ less than that expected by linear extrapolation from the power law term (Kagan, 1991). When we convert this 'corner moment' to the magnitude scale we will refer to it as m_θ. Examples of model fits to synthetic data for both incremental and cumulative frequency for equations 4.2 and 4.3 are shown in Figure 4.1. The best fit lines are shown for the incremental frequency because the data points are independent and provide a more realistic visual assessment of the scale of the confidence intervals shown. There are other common distributions (Kagan, 2002a) such as the 'characteristic distribution' (where there are more large events than expected by linear extrapolation than the GR law), the gamma distribution (where the exponential taper is applied to the probability density, not to the cumulative function as in MGR) or even a mixture of distributions, resulting in bilinear FMDs (Igonin et al., 2018; Staudenmaier et al., 2018) but for the purpose of the work here we focus on the GR and MGR distributions only.

In a previous study, Kagan (1991) stated that the dynamic range between m_c, and the largest magnitude within a catalogue needs to be 'widely separated' or else the standard estimate of the b-value in the GR relation will return a biased value. The amount of data and dynamic range needed to obtain a b-value estimate is linked to the desired accuracy and precision of the b-value. For example, Nava et al. (2017) show that when magnitudes are reported to a precision of 0.1 (as is often the case), an accurate estimate of b cannot be attained for a sample size less than 1000 events. Roberts et al. (2015) suggest from analysis of synthetic data that at least 200 events above m_c are required to statistically assess the b-value. This agrees with the results shown on the right hand side of Figure 4.2, which shows that at N = 200, the b-value error is around 0.1. Figure 4.2 uses synthetic data sampled randomly from an underlying GR

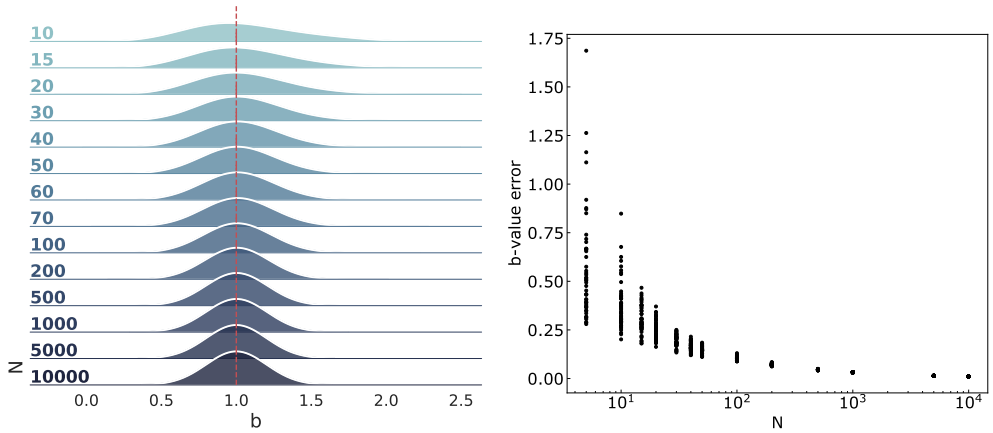

Figure 4.2: Left: Distribution of inferred b-values from randomly sampled data taken from a GR distribution, showing the convergence of the b-value with increased sample size. The true b-value of 1.0 is indicated by the vertical dashed red line. **Right:** The random error in b-value (using the maximum likelihood estimator described by Aki (1965), shown here to one standard deviation) when sampled from a GR distribution, with respect to increasing sample size.

distribution to show how substantial the spread in b-values can be when insufficient data is used for catalogue analysis, as well as the improvement in the accuracy (the peak is nearer the red dashed line) and precision (the spread on the left, or the error at one standard deviation on the right is lower) for larger data sets. The results suggest that N > 5000 for the b-value error to be less than 0.05.

Kagan and Jackson (1991) previously suggested that the reported regional variation of b-values could be an artefact of systematic errors in magnitude determination as opposed to non-universality of the earthquake scaling relationships. Using a consistent approach, Kagan (1997, 1999) further found the distribution of earthquake scaling to be universal, with β assuming a constant value of 0.63 ± 0.02 (equivalent b-value of ~ 0.95 ± 0.02 globally, with the exception of mid-ocean ridges). In some cases, frequency-magnitude data appear at least superficially to have a break in slope at a characteristic magnitude. For example, Kettlety et al. (2021) suggest a break in scaling between smaller events associated with hydraulic fracturing propagation (high b-value) and events linked to tectonic activation or re-activation of the fault (lower b-value more similar to tectonic events). Further, Herrmann and Marzocchi (2020) found

that in high-resolution catalogues, the exponential GR distribution does not hold in the low-magnitude ranges and should be used with caution for estimating any FMD properties such as m_c and b. Previously, Marzocchi and Sandri (2003) showed that Aki's (1965) MLE requires that, in practice, the GR law holds for a range of magnitudes $m_{max} - m_{min} \geq 3$. Eaton et al. (2014) have also shown that induced (micro)seismicity data do not always follow the GR distribution at large magnitudes, but instead exhibit a roll-off, which they attribute to an upper scaling limit arising from stratabound fracture networks where hydraulic fracturing takes place. We have already cited examples of elevated b-values in shallow earthquakes in induced and volcanic seismicity, but why are such reports so common? Godano and Pingue (2000) suggest that statistical inabilities due to the trade-off between having sufficient numbers of events and a reliable estimate of m_c can result in biased estimates of β.

4.3.2 Estimating the Parameters of the Frequency-Magnitude distribution

In this section, we summarise how to estimate the parameters in the FMDs introduced above, and also highlight some of the recognised pitfalls and problems of current practice in much of the published work in this area.

4.3.2.1 Estimating the b-value for a GR sample

One pitfall when estimating the frequency-magnitude parameters is to use least-squares regression to fit a straight line through raw frequency-magnitude data. The application of the least-squares method makes the b-value estimate overly sensitive to fluctuations in the tail of the distribution as not all of the data carry the same weight – it is much more likely for the tail of the distribution to contain only a few events in a single point whereas the points at lower magnitudes include hundreds of events. Furthermore, the residuals are asymmetrically distributed at high magnitudes and hence are not Gaussian-distributed (Naylor et al., 2010), as assumed in the least-squares method. This can lead to a systematic bias in the b-value estimation, which can be avoided by use of the maximum likelihood method (MLE, Aki (1965)). However, the misuse of

least-squares for b-value estimation is not uncommon (see Tables 4.1 and 4.2), as is the omittance of information regarding the method applied to calculate the b-value. In this paper, we use a maximum likelihood method and hence will not consider the biases derived from least-squares inference any further.

4.3.2.2 Estimating the parameters for the GR and MGR distribution

Here, we work with continuous moment data, which does not require corrections for magnitude binning (Tinti and Mulargia, 1985). Such continuous data is the basis for the maximum likelihood estimate for the b-value introduced by Aki (1965), but transferred here to the moment domain (we convert back to magnitudes for the discussion as these numbers are more physically intuitive). The implicit assumptions in using Aki's MLE method to estimate the GR b-value are a) that the data genuinely are power-law distributed in seismic moment, b) that the maximum moment (equivalent to maximum magnitude) is at infinity and c) that the m_c is well-defined. When data quality is poor, it is common practice to uncritically fall back on the parsimoniously simple GR model despite many of these assumptions breaking – the utility of b-values obtained under such conditions needs to be questioned.

Assuming a constant number of events N above m_c, the GR model (equation 4.2) only has one free parameter, β, which describes the slope of the frequency-moment distribution and can readily be converted to the b-value on the magnitude scale. The activity parameter, a is not free in this case as it is set by the total number of events (although it can be free if the number of events is expressed as an annual rate). For reasons described above, all synthetic calculations are performed in the moment domain. In the case of the GR distribution, the upper bound seismic moment is a 'hard' cut-off (Kagan, 2002a). To estimate the seismic moment distribution parameters (β for GR and β and M_θ for MGR), we apply a MLE by Kagan (2002a). For GR, the log-likelihood function for N observations of the seismic moment is:

$$l_{GR} = \log(L_{GR}) = N\beta \log(M_c) - \beta \sum_{i=1}^{N} \log M_i + \sum_{i=1}^{N} \log\left(\frac{\beta}{M_i}\right). \tag{4.4}$$

For the MGR distribution, the upper bound seismic moment is a more 'soft' taper than in the case of GR (Kagan and Jackson, 2000; Kagan, 2002a). In this case, where both parameters (β, M_θ) require estimation, the log-likelihood is given by:

$$l_{MGR} = \log(L_{MGR}) = N\beta \log(M_c) + \frac{1}{M_\theta}\left(NM_c - \sum_{i=1}^{N} M_i\right) - \beta \sum_{i=1}^{N} \log M_i + \sum_{i=1}^{N} \log\left(\frac{\beta}{M_i} + \frac{1}{M_\theta}\right) \tag{4.5}$$

where M_c is the completeness moment and M_i are the catalogue moments. Maximising these log-likelihoods returns estimates of the β (and M_θ) values.

To calculate the uncertainties of the estimated b-value based on seismic moment data, we use the standard estimate of error to one standard deviation without modifications (Aki, 1965; Kagan, 2002a), which is approximated to

$$\sigma_\beta = \frac{\beta}{\sqrt{N}} \tag{4.6}$$

where N is the number of events included to calculate $\tilde{b}$. For the MGR distribution, in particular when it incorporates the maximum observed moment (M_{max}) or the corner moment (M_θ), the standard error of β can be estimated by truncating the distribution at a moment M_x which is above the assumed or known value of M_{max} or M_θ (equation 25 in Kagan (2002a)):

$$\sigma_{\tilde{\beta}} = \frac{D}{\sqrt{n[D^2\tilde{\beta}^{-2} - (M_c/M_x)^\beta (\log M_x/M_c)^2]}} \tag{4.7}$$

where $D = 1 - (M_c/M_x)^\beta$.

4.4 Methods and Data

We first analyse the biases introduced when frequency-magnitude parameters are estimated for relatively small earthquake catalogues and when the form of the underlying distribution is known. This is done by drawing synthetic data from either a GR and MGR distribution whilst varying the dynamic (magnitude) range. These results are then generalised by using Monte Carlo (MC) simulations before applying the methods

to real catalogues from a variety of tectonic, volcanic and induced seismicity settings to understand the implications of our analysis.

4.4.1 Model Comparison using the ΔBIC

The criterion we use to distinguish between the GR and MGR models is a modified version of the Bayesian Information Criterion (BIC) by Leonard and Hsu (1999). This is preferred over other commonly used criteria such as the Akaike Information Criterion (AIC) or the Likelihood Ratio alone because the BIC penalises model complexity more heavily. Additionally, in synthetic trials the BIC proves to be more accurate a discriminant for the larger data samples used here (N > 46; Main et al. (1999)). It is defined by

$$BIC = \frac{1}{log(e)}\left(-2\log(L) + k\log(N)\right) \tag{4.8}$$

where L is the likelihood function, N is the total number of events and k is the number of model parameters (two for the GR law and three for the MGR, including variance). We compare the difference in BIC for the two models (ΔBIC_{MGR} - ΔBIC_{GR}) with a resultant Δ BIC > 0 preferring the GR distribution and Δ BIC < 0 preferring the MGR distribution (Bell et al., 2013).

4.4.2 Synthetic earthquake catalogues

We first create two synthetic event catalogues of 10 000 events, one randomly sampled from a parent GR distribution and one from a parent MGR distribution. The parameters used were b-value = 1.0, m_c = 1.0 and the MGR data used the extra parameter m_θ = 3.5. These parameters were chosen to mimic common magnitude ranges (and for commonly openly accessible data) found in volcanic and induced seismicity catalogues, which often have smaller magnitude events. We fit both the GR and MGR models to the continuous moment data. These two catalogues are then 'thinned' by increasing m_c in increments of 0.1. This creates a set of sub-catalogues of decreased sample size and dynamic range. Doing this enables the modelling of biases that may be introduced when

effectively sampling only the tail-end of the FMD, which is often the only accessible data in local volcanic and induced seismicity populations. Maximising the parameters in the likelihood functions for the GR and MGR models (equations 4.4 and 4.5) result in estimates of the b-value for each sub-catalogue for both GR and MGR distributions, as well as estimates for M_θ for the MGR distribution.

Additionally, we then use randomly generated Monte-Carlo (MC) sampling to simulate multiple catalogues with the same parameters, and look at the variability between individual realisations. Using the same parameters as above, for the single catalogues, we create 50 catalogues each for both GR and MGR and increase m_c in increments of 0.1, using the same process and analysis as for the single catalogue above. This allows us to explore the effect of random sampling on the precision and accuracy of the inferred b-values.

4.4.3 Real earthquake catalogues and determination of m_c

In the synthetic catalogues we can be sure we have complete data, but for the real catalogues we need to estimate m_c. If the estimated m_c is not accurate, bias is immediately introduced. Commonly used methods for the determination of m_c are: the maximum curvature method (MaxC; Wiemer and Wyss (2000)), the goodness-of-fit test (GFT; Wiemer and Wyss (2000)) and the b-value stability method (BVS; Cao and Gao (2002); Woessner and Wiemer (2005)). All of these methods assume self-similarity of the earthquake process, implying a power-law distribution (Woessner and Wiemer, 2005), but none of these methods is necessarily the 'correct' one. Mignan and Woessner (2012) give detailed explanations of how these methods work. The BVS method evaluates the estimated b-value for the data above a variable minimum magnitude. It selects m_c as the magnitude above which the b-value stabilises within error of the b-value. This method regularly tends to return the highest m_c of the three methods mentioned above, and is therefore also the most conservative (likely to be at or above the 'true' m_c). Due to its accuracy compared to the other methods (Roberts et al., 2015), it is the preferred method of m_c determination used here. We then again increase m_c in increments of

0.1 units, just as in the synthetic analysis, to determine the effect on the b-value and preferred model for a given catalogue.

For this study, we have chosen the following real catalogues as open-access examples covering a suitable range of possibilities – the Central Italy region (tectonic), Mount Etna (volcanic), El Hierro (volcanic) and The Geysers (induced by fluid injection). In line with the suggestion by Mizrahi et al. (2021) that observed decreases in the b-value could be an artefact of declustering algorithms on earthquake catalogues, no declustering has been performed on any of these catalogues. It is not clear in all catalogues which magnitude scales have been supplied, another potential source of magnitude uncertainties and bias, as extensively discussed in Herrmann and Marzocchi (2020). A summary of the sample size and magnitude ranges for each of these catalogues is given in Table 4.3. These catalogues have also been chosen due to their abundance in data (large numbers of events in the raw catalogue). However, Table 4.3 shows a large reduction in catalogue sample size and dynamic range for the complete data, especially for the Mount Etna, El Hierro and 1990 INGV catalogues.

4.4.3.1 Central Italy: Tectonic

The Central Italy region experiences a substantial number of earthquakes annually, including devastating ones such as the sequences in the summer of 2016. We use a tectonic catalogue managed by the Istituto Nazionale di Geofisica e Vulcanologia (INGV), starting in 1990 and ending just before the onset of the Central Italy Earthquake Sequence in August 2016. This project was started around this time and a detailed, relocated catalogue including the Central Italy sequences was under review at this point, so the work currently includes data only up to August 2016. However, in future work it would be a good idea to include data from the last five years now to see if, and if so how, the outcomes discussed here may have changed with more data. Sensitivity of seismometer equipment was improved in 2005, resulting in the detection of many more small earthquakes and in the overall enhancement of recording. Therefore, this catalogue was split into 2 catalogues with different start dates: 'INGV 1990' and 'INGV

2005′. The evaluation of m_c using BVS returned values of 2.7 and 1.7 and magnitude ranges above m_c of 3.1 and 4.4 respectively.

4.4.3.2 Mount Etna, Sicily: Volcanic

Mount Etna is an active stratovolcano in Sicily, Italy. It is one of the world's most active volcanoes, whose activity is accompanied by many seismic events both during and outwith eruptive phases. The catalogue chosen from this location contains mainly events recorded between 1999 and 2016. Most are associated with volcanic activity, although the possibility of some tectonic activity included in the catalogue remains. The BVS method indicated m_c = 2.5, resulting in a dynamic range of 3.2 above m_c. No efforts have been made at trying to discriminate between tectonic and volcanic seismicity here.

4.4.3.3 El Hierro, Canary Islands: Volcanic

This is another catalogue chosen primarily for its volcanic seismicity within the volcanic island chain of the Canary Islands. This catalogue was previously used by Roberts et al. (2016) to discover mode switching during the 2011 - 2013 El Hierro eruption. We have extended this catalogue in time (2001 – 2017) and to include several more events within the entirety of the island of El Hierro, however more than 95% of events are from the eruptive period and therefore the same as those used in Roberts et al. (2016). This extended catalogue returns m_c = 2.6 and a dynamic range of 2.5 above m_c.

Table 4.3: Summary of parameters for the real catalogues including type of seismicity, the period during which the catalogue events were observed, the total number of events in this observed period (N), the number of events above m_c, the observed maximum magnitude within the catalogue, the value of the b-value stability m_c, the magnitude dynamic range (dyn. r) as well as an estimated value of m_θ for when the MGR MLE is applied to all catalogues. These values are also indicated on the FMDs in Figure 4.7.

Catalogue	**Seismicity Type**	**Observation Period**	**N**	**N** $> m_c$	**obs.** m_{max}	m_c **(bvs)**	**dyn. r**	**est.** m_θ
Mount Etna	Volcanic	19/08/1999 - 29/08/2016	15 688	1928	5.7	2.5	3.2	~ 5.7
El Hierro	Volcanic	30/11/2001 - 22/01/2017	22 494	1782	5.1	2.6	2.5	~ 5.0
The Geysers	Geothermal	18/03/1972 - 31/07/2018	281 572	61 134	5.01	1.25	3.76	~ 4.6
INGV 1990	Tectonic	01/01/1990 - 15/04/2005	17 166	2917	5.8	2.7	3.1	~ 6.0
INGV 2005	Tectonic	16/04/2005 - 24/08/2016	114 847	16 629	6.1	1.7	4.4	~ 6.5

4.4.3.4 The Geysers, California: Geothermal

This naturally occurring steam field reservoir provides an energy source for the largest existing complex of geothermal power plants (Leptokaropoulos et al., 2018a). The vast majority of earthquakes within this region are attributed to the extraction of steam and the re-injection of condensate and wastewater into the reservoir (Henderson et al., 1999). The monitoring of seismicity began in 1972 and according to the USGS, seismicity has remained stable since the 1980s. The largest earthquake recorded in The Geysers had a moment magnitude of m_w = 5.01, on 14th December 2016. Seismicity in The Geysers is due to fluid re-injection and hence in principle, the rate of seismicity can be controlled by the rate of fluid injection and withdrawal (Majer et al., 2007). There are no large faults within the region and none of the mapped faults has been active in the last 10 000 years (Oppenheimer, 1986), making it unlikely for any of the events to be solely tectonic. From the Northern California Earthquake Data Center (NCEDC) we use a catalogue for The Geysers geothermal region from April 1972 up to July 2018 with a BVS m_c of 1.25 and a dynamic range of 3.76 above m_c. These choices minimise the possibility of including other events that may not be induced by geothermal activities.

4.5 Results

Here we present the results of the analysis on synthetic catalogue data drawn from a GR and MGR distribution and generalised using MC simulations (as described in section 4.4.2) before demonstrating the applications on the real data from the catalogues introduced above.

4.5.1 Synthetic catalogues

Figure 4.3 summarises the results obtained from analysis of the synthetic catalogues. In the first column, the FMDs are shown for two examples of GR and MGR catalogues – the original, full catalogue (N = 10 000 and m_c = 1.0, GR in red and MGR in green) as well

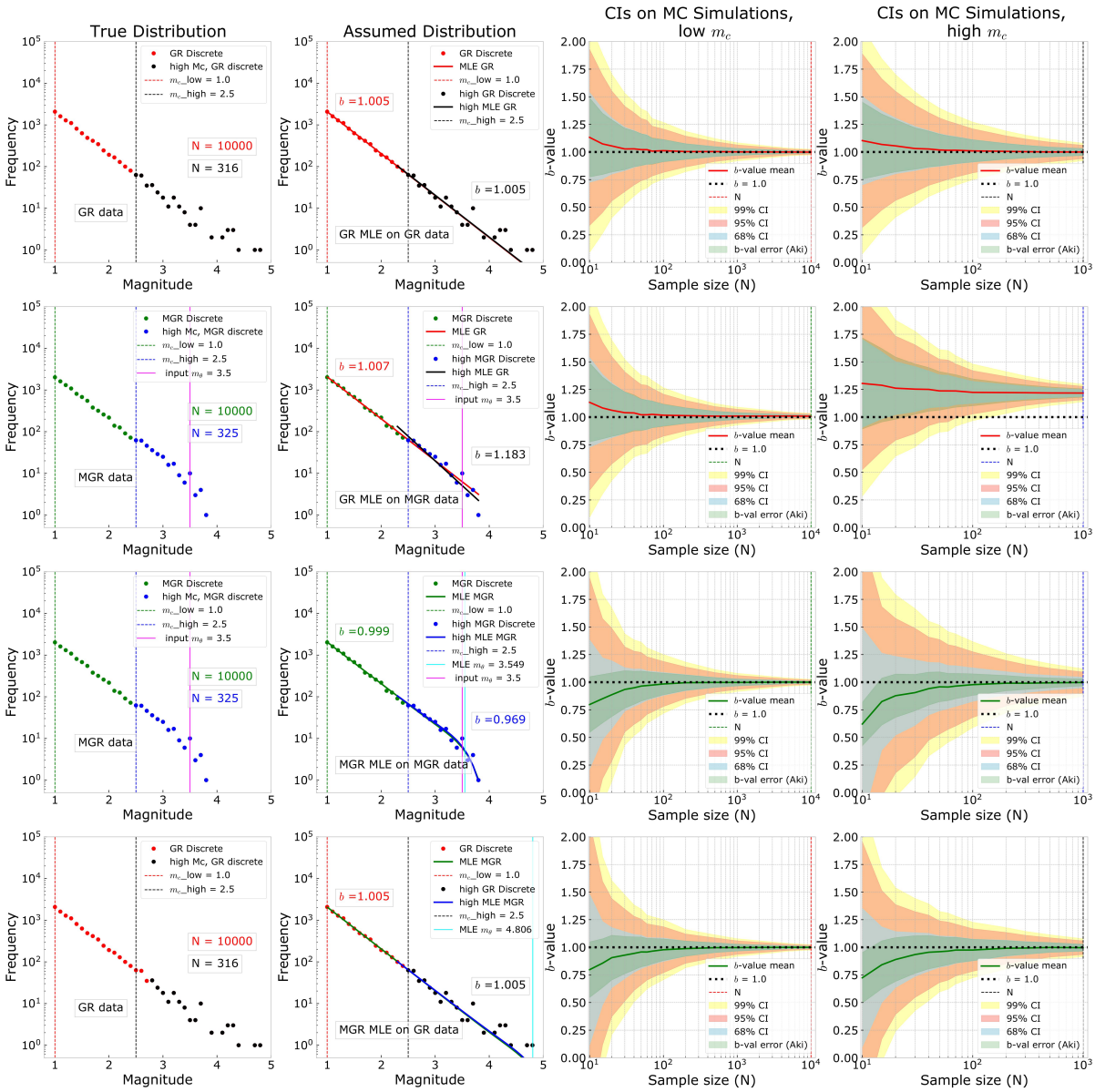

Figure 4.3: Summary of the synthetic results obtained using GR and MGR models. The first column represents the full and a smaller sub-catalogue of the true distribution – rows 1 and 4 show GR (red, sub-catalogue in black) and rows 2 and 3 show MGR (green, sub-catalogue in blue). The second column includes the assumed distribution – MLEs are indicated by solid lines fitting to the underlying data. The top row represents GR MLE on GR data, the second row shows GR MLE on MGR data, the third row represents MGR MLE on MGR data and the final row is MGR MLE on GR data. Respective b-values of the MLE are indicated on the panels for all catalogues. The last two columns show confidence intervals (68%, 95% and 99%) on simulated catalogues with two different values of N. N = 10 000 in column 3 and N = 1000 in column 4, which mimics the tail-end of a sub-catalogue. The solid red and green lines represent the b-value mean.

as one specific sub-catalogue (m_c = 2.5, GR in black, MGR in blue). The b-value is 1.0 for both and in the case of the MGR distribution, m_θ = 3.5. The smaller sub-catalogue is used as an arbitrary example of how parameters may change when sampling less data and concentrated on the tail-end of the distribution at larger magnitudes. The number of events in each of these catalogues is indicated on the respective FMD plots, in the respective colours. In the second column, we model the following (top to bottom): GR MLE on GR data, GR MLE on MGR data, MGR MLE on MGR data and MGR MLE on GR data. The b-values of these best fit models are indicated in the respective colour, on the plot. Both the models and the b-values show that the GR MLE fits GR data well and the MGR MLE fits MGR data well (rows 1 and 3), but the GR MLE does not fit MGR data well (row 2), when m_c = 2.5 (implying a lower dynamic range and sample size). The MGR MLE fits GR data (row 4) well, as the roll-off is not sampled (the m_θ estimate is given in cyan). In real catalogues, the underlying distribution is unknown, so the misidentification, for example based on an apparent good fit of the MGR law to GR data or vice versa, presents a possible source of bias. Columns 3 and 4 in Figure 4.3 show confidence intervals and the confidence intervals on the b-value from MC simulations performed on two different catalogue sizes: N = 10 000 (column 3) and N = 1000 (column 4). Column 4 has a lower N compared to the sub-catalogues in columns 1 and 2 because the sample sizes are very small. Columns 3 and 4 reveal the b-value bias and loss of precision that results when going from a catalogue of N = 10 000, down to a catalogue of N = 1000. The 67%, 95% and 99% confidence intervals are shown, as well as the mean b-value (solid red/green line) for all four scenarios. These mean b-values are always biased at small sample sizes, even when the correct model is applied to the data generated but the mean converges to the true b-value with increased sample size in all cases except for the GR distribution on MGR data (row 2) in the case N = 1000 (second row, fourth column), in which the mean b-value is always biased to high values (i.e. the red line is always above the black dashed line).

The MC simulations are used to investigate how effective the BIC (see section 4.4.1) is at choosing the true model, with the key factor being the extra complexity (due to

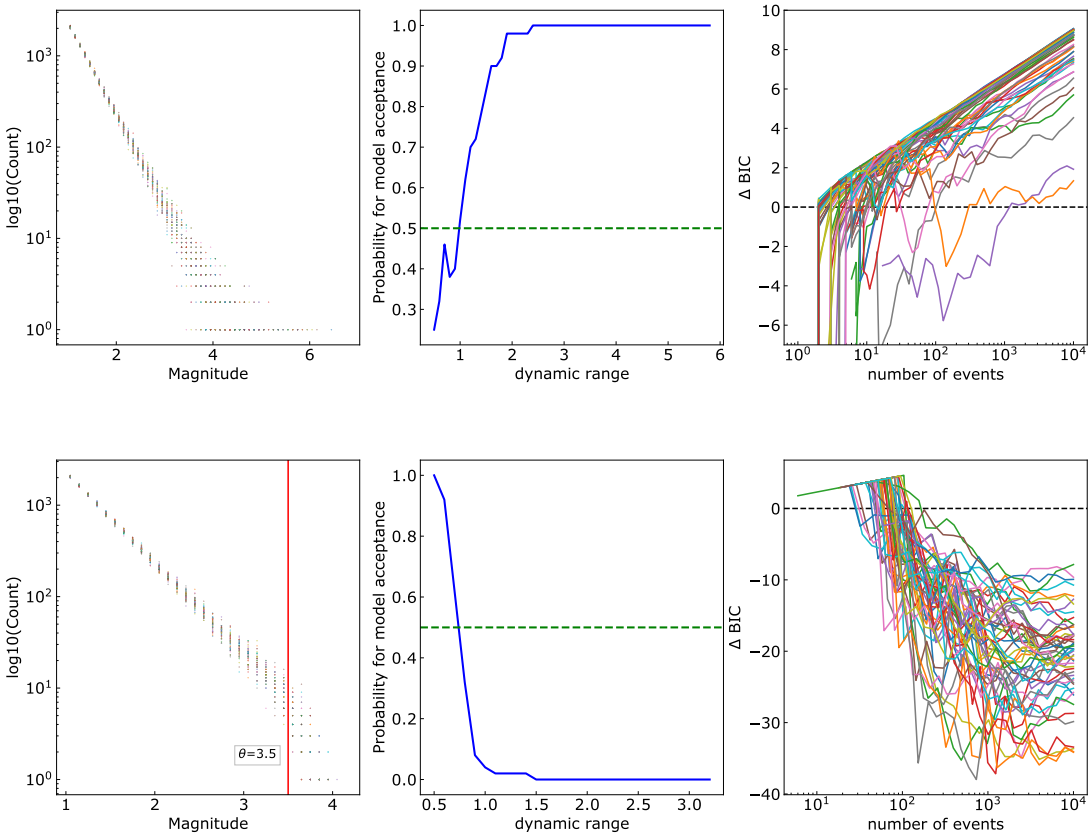

Figure 4.4: Left: FMDs for all GR (top) catalogue and MGR (bottom) catalogue simulations superposed. The solid red line in the MGR simulations indicates the input value of m_θ. **Center:** the probability that a model will be preferred dependent on the dynamic range, defined as the maximum sampled magnitude minus m_c, with 1 preferring GR and 0 preferring MGR in the case that the true distribution is GR (top) and MGR (bottom). **Right:** ΔBIC for all 50 GR (top) and MGR (bottom) simulations shown against number of events. GR is preferred when values are greater than 0, MGR is preferred when values are less than 0.

the additional model parameter M_θ) and the associated BIC penalty when applying the MGR model. Using m_θ = 3.5, Figure 4.4 shows the FMDs for 50 simulations in the case of an underlying GR (top) and MGR (bottom) distribution, along with the average probability that a certain model is preferred by ΔBIC and the evolution of ΔBIC with respect to m_c for the individual simulations. The GR distributed data results in a preference for the GR model at all times, with a strong preference (> 0.5) when the dynamic range is 2 or larger. The ΔBIC equally indicates that with increased number of events, the preferred model is more likely to be GR, i.e. the correct one. On the other hand, in the bottom panel (MGR distributed data), the FMD for all simulations shows that most of them sample the given roll-off of m_θ = 3.5, although not all (seen by some points being to the left of the vertical line indicating m_θ). The model acceptance in the center clearly shows that for the used parameters, at least one order of dynamic range is needed to identify the correct model (MGR in this case). Any dynamic range less than that, tends towards a preference for an incorrect GR model. The ΔBIC shows that the GR model is preferred for a small number of events (which coincides with small dynamic range) but the correct, MGR model is preferred for N > 100 in almost all the simulations. Corresponding to this, Figure 4.5 plots the evolution of ΔBIC as a function of the inferred b-value for both GR (top) and MGR (bottom) data, colour-coded by dynamic range. We show that in the case of a true GR distribution (top, with b = 1.0), small dynamic range (green triangles) leads to a larger spread in estimated b-values (we know that b = 1.0 in this case) as well as lack of confidence in the correct model selection (low values of ΔBIC). Several data points on Figure 4.5 have Δ BIC < 0, suggesting an incorrect preference of the MGR model, while estimated b-values spread between 0.5 and 1.8. With increased dynamic range (blue triangles), the bias of the b-value reduces (i.e. more points lie within the range b = 1.0 ± 0.2) and the preference for the GR model becomes stronger (greater ΔBIC). In the case of an underlying MGR distribution (Figure 4.5, bottom), a smaller dynamic range (green triangles) again results in increased misidentification of the correct model – many points have Δ BIC > 0, showing they incorrectly prefer a GR model. At the same time, the corresponding b-values mainly lie

between 2.0 and 3.5, which is abnormally high. The correct distribution is recovered on increasing the dynamic range (blue triangles), with estimated b-values nearer 1.0, close to that assumed in generating the synthetic data.

In Figure 4.6, we show the inferred GR b-value as a function of the dynamic range on underlying MGR data. Using six different values of m_θ (with 50 MC simulations for each), N ranging from 50 to 5000 and the true b-value being either 1.0 (left column) or 1.5 (right column). For a given value of m_θ, the dynamic range is varied by changing the value of m_c between 1.0 to 2.5 in increments of 0.1. This synthetic data illustrates how high b-values can arise when assuming an incorrect GR model on underlying MGR data. In the case of a true b-value of 1.0 and N = 50 (Figure 4.6, top left), b-value estimates range from $\sim$ 0.7 (at dynamic ranges > 1.5 and $m_\theta > 3.4$) to $\sim$ 3.5 (at a dynamic range of 0.5 and $m_\theta = 2.5$). Generally, the estimate of the b-value converges to the correct value of 1.0 as dynamic range is increased for a given value of m_θ, and for a given dynamic range as m_θ increases. The rest of column 1 shows how much smaller the b-value bias becomes with larger sample size. At N = 5000 (Figure 4.6, bottom left), estimated b-values range between 1.0 and 2.9 for all the catalogues compared to between 1.0 and 3.8 for N = 50, indicating a lower overall spread. However, the estimated b-value is also much closer to the true value of 1.0 for the larger m_θ cases at large dynamic ranges [highlighted in orange ($m_\theta = 3.7$) and purple ($m_\theta = 4.0$)]. Similar results can be seen in Figure 4.6, column 2, where the true b-value is 1.5, except that the spread in estimated b-values is even larger for all values of m_θ, especially in the cases of N = 50, 100 and 500. Lower dynamic ranges occur because of the steeper slope on the FMD compared to the case $b = 1.0$. Figure 4.6 proves the case that when a) the wrong model is assumed, b) the sample size is very small (i.e. N = 50) and c) the dynamic range is also very small, then the b-value is extremely biased to high values, and has a much bigger scatter, i.e. it is neither accurate nor precise, respectively. Accuracy and precision both improve with increased dynamic range and sample size. Nevertheless, Figure 4.6 suggests that an accurate estimate of the b-value can be obtained within a precision of approximately $\pm$ 0.2, even when the incorrect model is assumed, if a) the dynamic range is at least 1.5

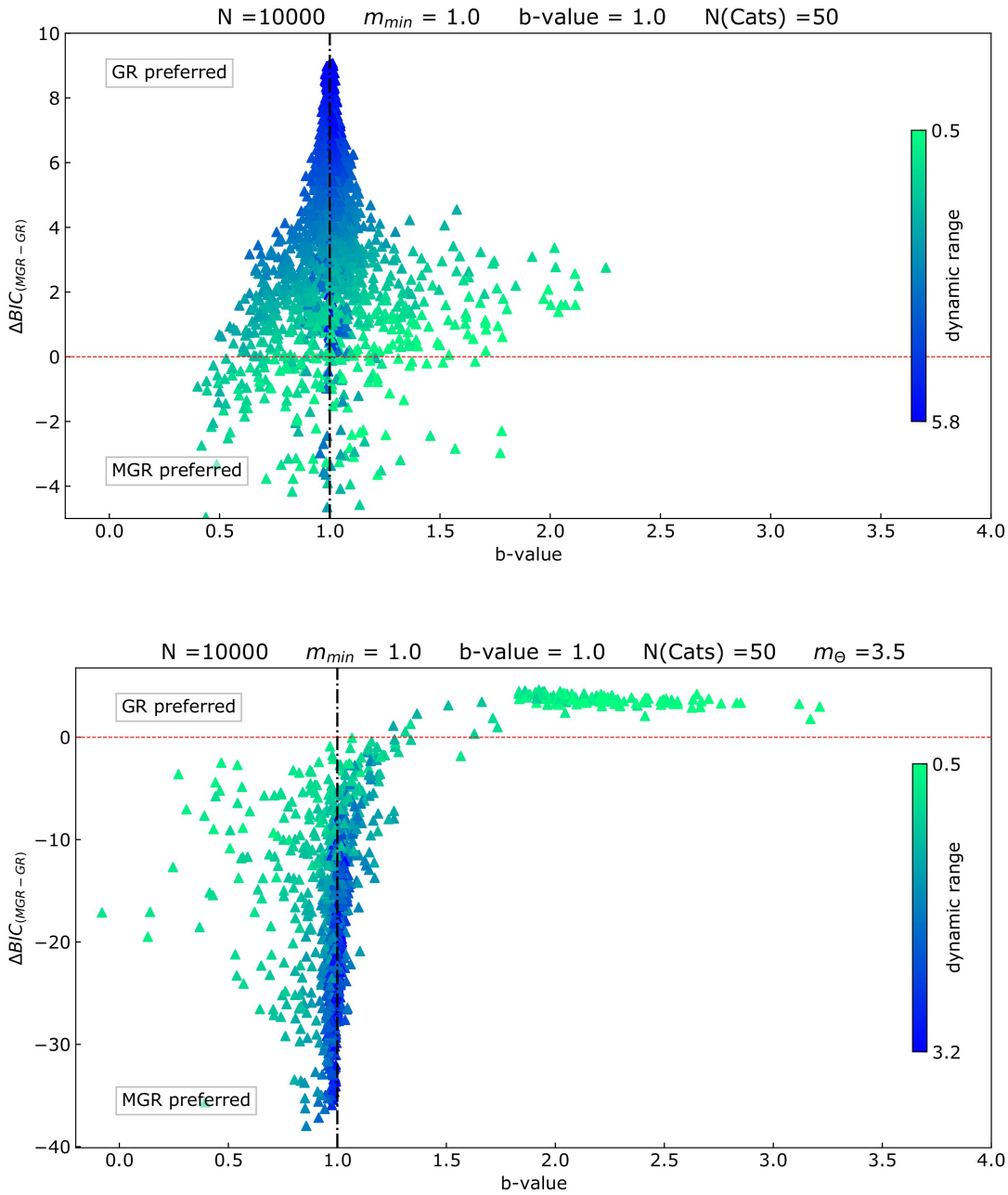

Figure 4.5: Analysis for GR data (top) and MGR data (bottom) showing: ΔBIC against the inferred b-value for all 50 catalogues with 10 000 events generated from a synthetic catalogue with $b = 1.0$, varying m_c for each. Positive ΔBIC implies a preference for the GR model, negative ΔBIC is a preference for the MGR model. Darker (blue) markers indicate larger dynamic ranges; lighter (green) markers indicate small dynamic ranges.

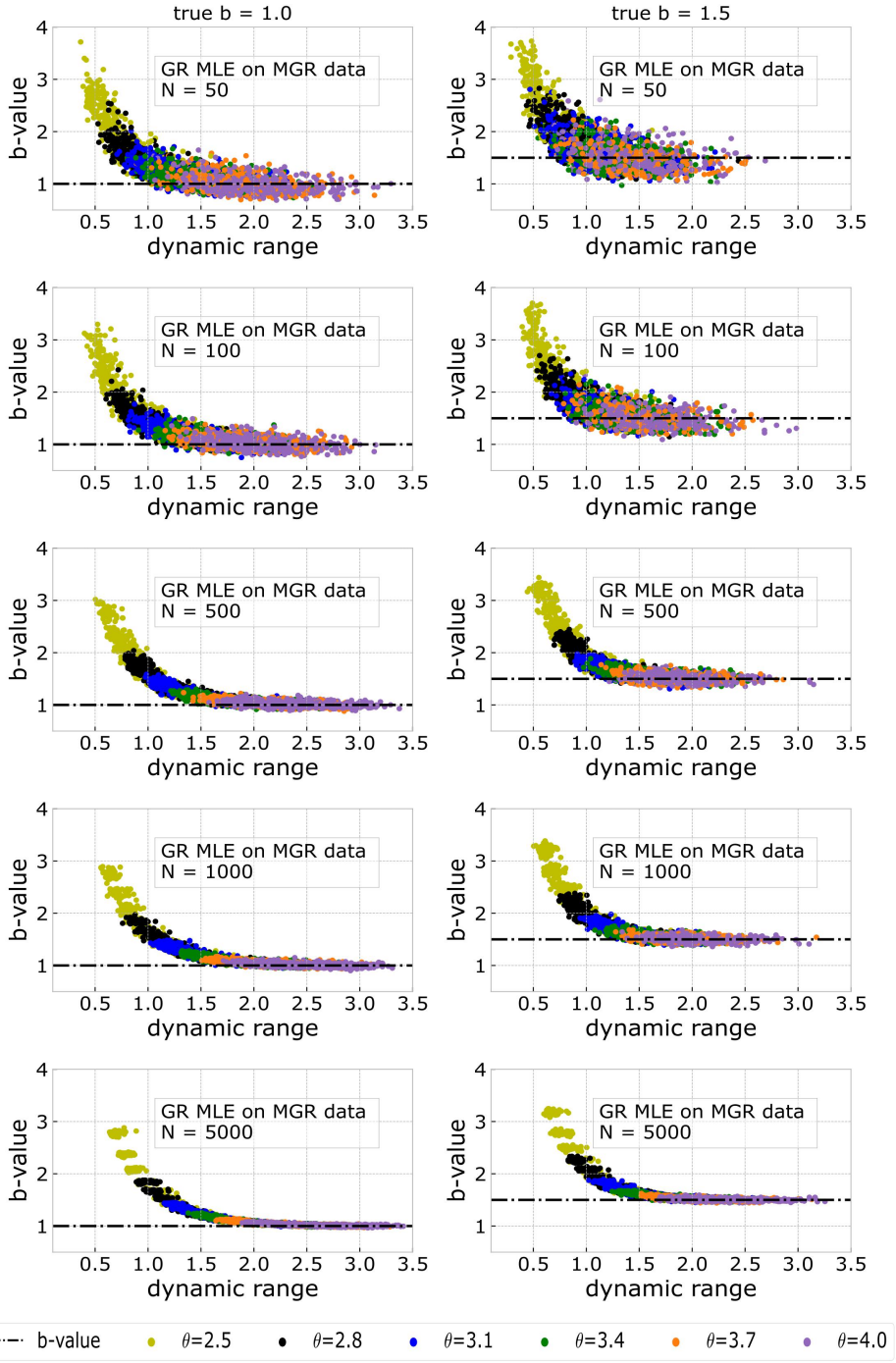

Figure 4.6: Estimated b-values assuming the GR model (applied on MGR data) as a function of dynamic range for **Left column:** a true b-value of 1.0; **right column:** a true b-value of 1.5. Colours represent different values of m_θ (see key immediately above).

and b) there are at least N = 1000 events in the sample.

4.5.2 Real catalogues

We now investigate the real catalogue data from the areas of tectonic, volcanic and induced seismicity introduced in section 4.4.3. Based on the results of our analysis of synthetic data, we now test the hypothesis developed there that many studies (including those cited in Tables 4.1 and 4.2) have found high b-values due to lack of data and/or a narrow dynamic range used in the analysis. The high b-values reported in the literature may be partly due to the smaller maximum magnitudes for volcanic or induced seismicity, but also because many b-value studies focus particularly on short-term data obtained during periods of volcanic unrest or during induced seismicity sequences. Hence we use more complete data available for the longer time scales outlined in section 4.4.3, and exploit the larger data sets in the five catalogues examined here. We begin by determining the m_c using the FMDs and BVS method (Figures 4.7 and 4.8), and then increasing the obtained m_c to thin the catalogue in the same way as we did for the synthetic data (section 4.4.2).

Figure 4.7 shows the FMDs for all catalogues, including the estimates of the BVS m_c (see Figure 4.8) in solid lines as well as the calculated value of m_θ as dotted lines. The Geysers data (black) has the lowest m_c of 1.25 and the INGV 1990 catalogue (green) the highest at 2.7. The higher-resolution INGV 2005 catalogue results in a significant improvement in complete reporting, with m_c lowered to 1.7 due to increased station density. The BVS curve is shown in more detail for the real catalogues in Figure 4.8, with errors shown at one standard deviation. An estimate of m_c is made where successive b-values for two trial values of m_c are the same within error. Estimates assuming an underlying GR model are shown in solid colours and estimates of the equivalent for an underlying MGR distribution are shown in lighter colours. With the exception of The Geysers case, the estimated b-values for the MGR distribution are systematically higher than those for the GR distribution, but the estimated values of m_c remain the same. The trend in the b-value shows a general increase in b and its error with respect

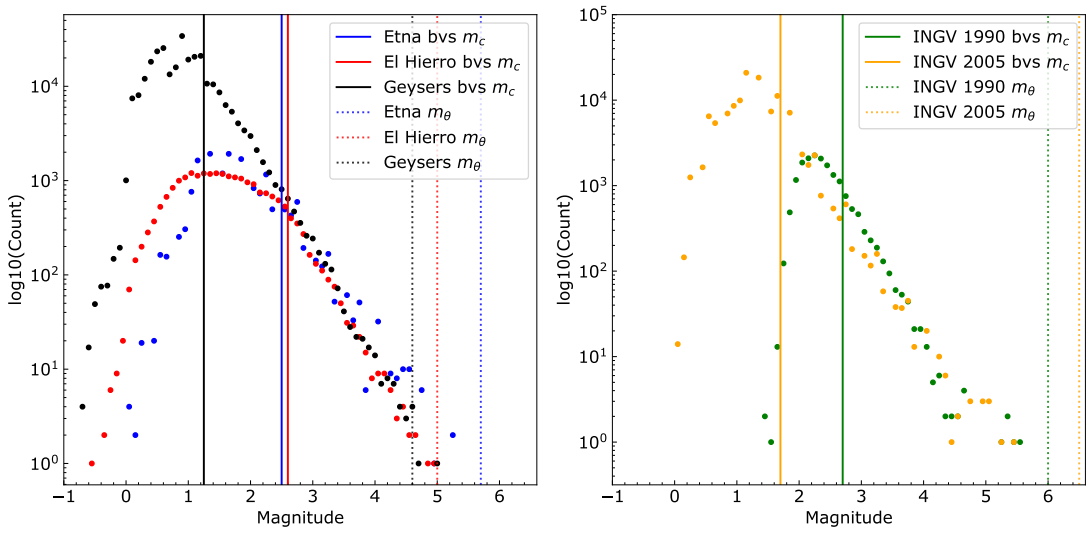

Figure 4.7: FMDs for all five catalogues. Solid vertical lines indicate the estimated m_c using the BVS method, dotted vertical lines show the calculated value of m_θ. The left panel shows results from volcanic and induced seismicity, and the right panel indicates a single tectonic setting, analysed for two catalogues of varying temporal and spatial resolution.

to m_c, with a more pronounced hiatus in the tectonic data due to its greater dynamic range. Conditional on the respective GR BVS m_c estimates, the estimated b-value is ~ 1.05 for The Geysers, ~ 1.0 for Etna and ~ 1.2 for El Hierro, remarkably close to $b = 1.0$ expected for tectonic seismicity. The two INGV catalogues show opposing trends. For the INGV 1990 catalogue, the estimated b-values rapidly increase with respect to m_c up to the best estimate of m_c (at which $b \sim 1.35$) and then sharply decreases between $3.7 \leq m_c \leq 4.8$ with b-values dropping as low as ~ 0.6 before another steep increase. Thus, the b-value estimate is not stable, and there may be systematic deviations from the GR law for this catalogue. The MGR and GR BVS b-value estimates tend to agree with each other during this 'descending' part between $3.7 \geq m_c \geq 4.8$, albeit with larger errors for the MGR model. When $m_c > 4.6$, b-value errors for both are very large and biased towards very high b-values. The INGV 2005 catalogue shows fairly steady b-value estimates between $1.7 \geq m_c \geq 4.0$, with an inferred b-value of ~ 1.15 at the GR BVS m_c estimate. Thereafter, the b-value tends to systematically increase with increasing m_c, confirming that data with small dynamic range may have inferred b-values that are

biased to higher values, irrespective of the assumed model.

In Figure 4.9, we show the estimated b-value as a function of the dynamic range for each of the real catalogues, obtained by varying m_c. The Geysers is the only catalogue which has a very large estimated b-value (> 2.5) at a very low dynamic range (~ 0.5) but then exhibits a continuous decreases until it reaches a b-value of ~ 1.0 at a dynamic range of ~ 3.7, again similar to that of tectonic seismicity. The two volcanic catalogues generally have a decreasing or stable b-value between dynamic ranges of 1.0 and 2.3 (for El Hierro) and 2.9 (for Etna), before they increase again just before their maximum dynamic range. This may be a statistical artefact of sampling for rare, large events. The INGV 1990 catalogue has a very low estimated b-value at low dynamic range ($b \approx 0.55$ at a dynamic range of 1.0). It continuously increases until it reaches a dynamic range of ~ 3.0. The INGV 2005 catalogue shows convergence to a b-value of 1.0 between dynamic ranges of 2.5 and 3.8, before it exhibits 'jumps' to larger b-values, again likely to be statistical artefacts. These systematic changes in estimated b-value with respect to dynamic range are also reflected in the trends for ΔBIC shown in Figure 4.10, where only The Geysers field shows a trend towards preferring an MGR model. Our results for the real data confirm the inference from the synthetic data shown in Figure 4.6 that increased dynamic range will result in a better chance of returning a consistent b-value, irrespective of the ultimate model preference. Logically, this implies that this consistent b-value is more likely to be nearer the true one for the scale-invariant part of the process, but in the case of real data this cannot be proven. In other respects, the real data do not show the relatively smooth variation in the sensitivity analysis with respect to dynamic range seen in the ideal case of Figure 4.6. Notably, the INGV catalogues behave different from the induced and volcanic seismicity catalogues, and are also distinct from one another.

Figure 4.10 shows how both ΔBIC and the inferred b-value change as a function of m_c (degree of shading) for the different catalogues (coloured symbols). For the INGV catalogues (green diamonds and orange crosses), all ΔBIC values remain above zero and hence within the GR preferred region. The Geysers data show that at full data

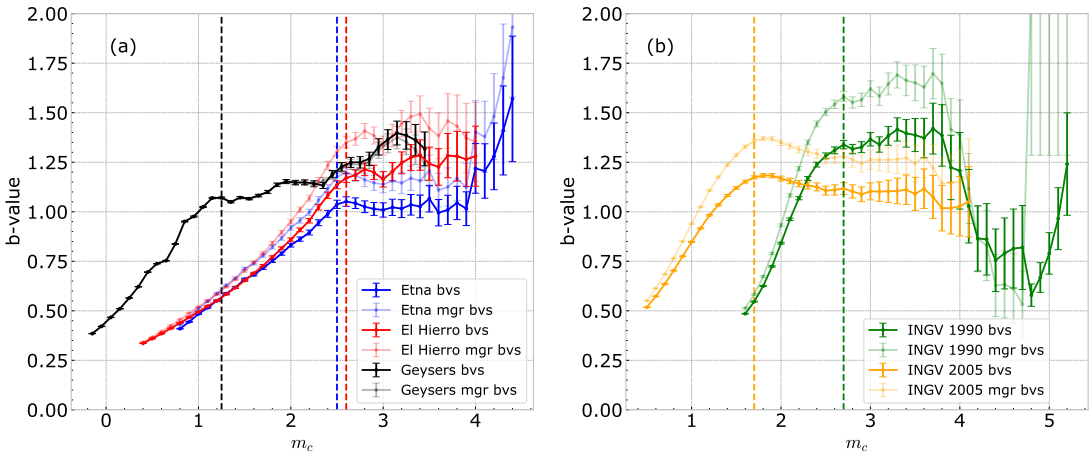

Figure 4.8: b-value stability plot using GR (solid colours) and MGR (lighter colours) for (a) the volcanic and induced catalogues and (b) the tectonic catalogues. Vertical dashed lines indicate the estimated m_c for each catalogue, based on successive b-values falling within error of the b-value. (Based on Roberts et al. (2015).)

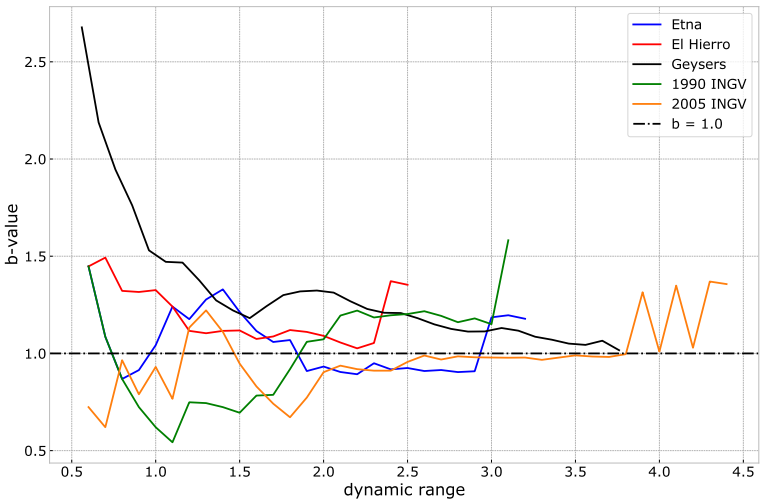

Figure 4.9: Equivalent of Figure 4.6 for the real catalogues, showing the b-value as a function of dynamic range for each of the real catalogues. The 'commonly' accepted b-value of 1.0 is shown as a dashed black line for reference.

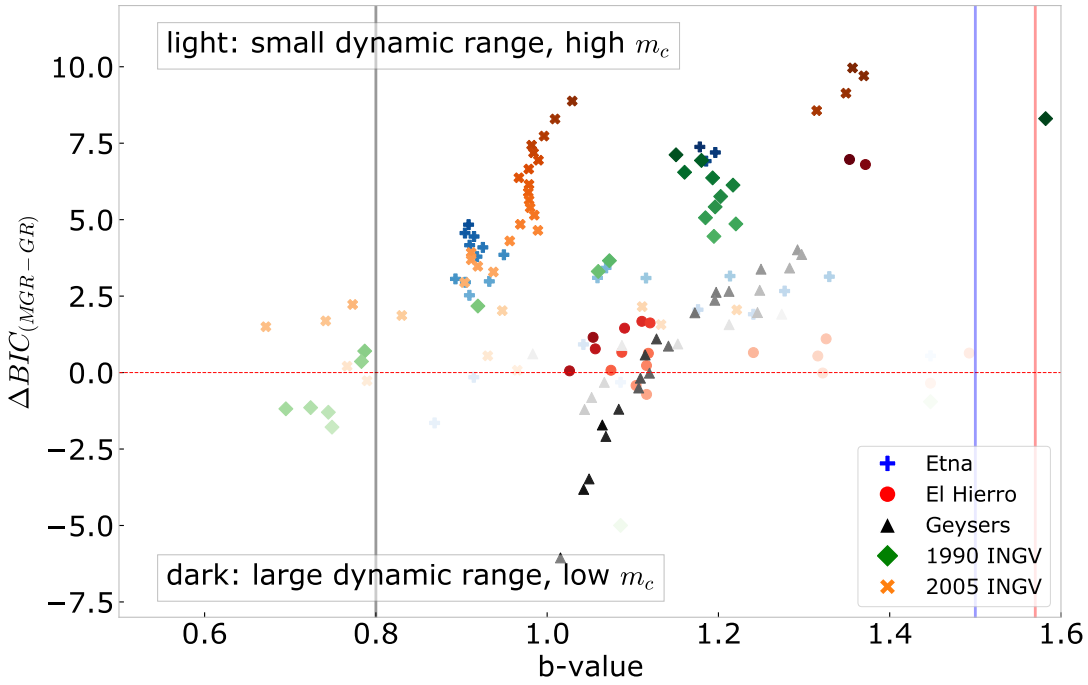

Figure 4.10: Δ BIC as a function of the inferred b-value for five different datasets - two volcanic (Etna and El Hierro), one induced (the Geysers) and one tectonic (Central Italy). Positive ΔBIC implies a preference for the GR model. b-values are calculated based on which model is preferred (MGR only for The Geysers). The minimum magnitude threshold is increased above the m_c determined by the b-value stability method - darker markers indicate lower magnitude thresholds and larger dynamic ranges; lighter markers indicated higher magnitude thresholds and lower dynamic ranges. The vertical lines indicate the values of b_{typ} (Tables 4.1 and 4.2) observed in previous studies of Etna (blue), El Hierro (red) and The Geysers (black). These are not shown for the INGV catalogues because the decision to split this original catalogue has not been previously investigated for b-values and there is no reference value.

availability, the optimal model chosen by the ΔBIC is MGR distributed, but when cut to a narrower dynamic range, the GR model is increasingly preferred and b-values are high. This is a typical example of how imposing a narrow dynamic range can result in a false preference for a GR model with an artifically high b-value. From the value of m_θ in Table 4.3 and Figure 4.7, The Geysers is the only catalogue here which samples a roll-off with sufficient data to prefer an MGR model (i.e. m_θ is within the FMD range and not outwith, as is the case with all other catalogues). The evolution of ΔBIC with respect to the dynamic range for The Geysers data shown in Figure 4.10 confirms this inference, i.e. the MGR model is preferred more strongly with more data. These results clearly illustrate the high potential for simultaneously a) misidentifying the underlying FMD distribution and b) returning too high a value for b in cases where the underlying distribution is more likely to have an MGR form, as demonstrated earlier on in our synthetic results. Both volcanic catalogues prefer the GR distribution at large dynamic range and their b-values cluster around ~ 0.9 to 1.2, i.e. not too different from that expected for tectonic earthquakes, but substantially smaller than the values of b_{typ} given in previous studies (see Table 4.3). For both volcanic and the INGV catalogues, GR is increasingly preferred with more data, which is the opposite to what we see for The Geysers. The vertical lines on Figure 4.10 indicate inferred b-values from previous studies (see Tables 4.1 and 4.2) for the two volcanic and induced catalogues. These are particularly high for the volcanic catalogues, and particularly low for The Geysers (where the Table 4.3 suggests a typical b-value of < 1.0), compared to our results.

4.6 Discussion

The synthetic results have shown that in the case of sampling from an underlying MGR distribution, the simpler, albeit incorrect model (GR) is more likely to be selected as the catalogue is thinned. This implies that, for low dynamic ranges, the GR model could be erroneously selected and the b-value estimate would then be biased to high values. When there is a lot of data, the GR model does a reasonable job of recovering the true

b-value irrespective of the underlying distribution, though some bias can result from the choice of m_c. There is no obvious difference in the general pattern of convergence between the $b = 1.0$ and $b = 1.5$ columns in Figure 4.6. That is, the bias in the b-value is largely controlled by the dynamic range of the catalogue magnitudes rather than the b-value itself.

The analysis of The Geysers data confirms the hypothesis developed from the analysis of synthetic data, that high b-values and a preference for the GR law are more likely to be seen in smaller catalogues with an underlying MGR distribution. This occurs because the method cannot distinguish between a steep (GR) slope and a (MGR) roll-off in the distribution for high m_c, in agreement with the results from Eaton et al. (2014), suggesting that an MGR distribution may lead to incorrectly high b-values, when this is not independently known to be the underlying distribution. Most sub-catalogues for volcanic seismicity prefer the GR model and have b-values near 1.0, though the b-value nearly always increases as the sample size becomes very small, as also seen in the synthetic data. However, the trends are not as monotonic in the real data. The most likely reason for the bias to high b-values in small catalogues following the GR law, is that the Aki (1965) formula assumes a maximum magnitude of infinity, whereas the data examined here have a finite upper bound to the sample maximum magnitude. For the tectonic catalogues, all sub-catalogues have a preference for the GR law, but b-values fluctuate within a surprisingly large range as the catalogues are thinned, associated with a strong variability at larger magnitudes (i.e. in the tail of the distribution). In turn, this may be due to random effects of statistical sampling, though we cannot rule out the possibility of complexity in the underlying distribution beyond that of the GR or MGR models.

In many volcanic and induced seismicity catalogues, the dynamic range and the number of data points in the literature is very limited, often $N \ll 1000$ above m_c (Tables 4.1 and 4.2). This renders it near impossible to test whether a true GR distribution is seen or whether the MGR roll-off is mistaken for a steep (high b) GR distribution. Both our synthetic and real catalogue results indicate that with too little data, b-values

can be biased by lack of knowledge of the underlying distribution. This is an example of an epistemic uncertainty in the b-value that cannot be captured by the statistical or aleatory uncertainty expressed by equation 4.6. Furthermore, many studies have focused exclusively on time spans around volcanic unrest or induced sequences. This introduces another sampling bias, because the time window results in a lower number for events, and hence a preference for a higher inferred b-value. Accordingly, we cannot assume such high b-values are representative in assessing long-term hazard, and are in fact likely to underestimate the hazard from larger events outside the range observed in the finite sample window.

Finally, we note that our estimates of m_c for the real data using the b-value stability method may be sub-optimal, although best practice at determining which method to determine m_c has been employed in this study. Recently, Marzocchi et al. (2020) and Herrmann and Marzocchi (2020) use a method that explicitly tests the suitability of an exponential FMD for the data above the candidate m_c values, resulting in a more sharply-defined optimal m_c. In future work, it would be interesting to examine the influence of using such alternate techniques and identify potential over- or undersampling issues, which easily lead to incorrect hazard estimates.

4.7 Conclusion

The main conclusions that can be drawn from the results presented here are as follows:

1. If you have too small a sample of magnitudes (typically < 1000) drawn from a GR distribution, then there is a substantial bias to higher b-values when assuming the wrong model. From Figures 4.4 and 4.5, the data suggests that two orders of dynamic range are required to correctly estimate a GR model and its b-values within $\pm$ 0.2 of $b = 1.0$ in the synthetic case, and one order of dynamic range is required for the MGR model. This compares to Marzocchi et al. (2020), who have shown that the bias on the b-value is $> 20\%$ for a dynamic range of 1.0 and $< 0.5\%$ for a dynamic range of ≥ 3.0, when using the GR model on MGR data. Many

published analyses of volcanic and time-dependent seismicity fail this initial test of sufficient dynamic range.

2. If the GR model is erroneously applied to MGR distributed data, then the b-values are erroneously high, especially when the number of events is limited (Figures 4.3, 4.5 and 4.6). From Figure 4.3, to obtain a b-value to a precision of ± 0.05, a minimum sample size of 10 000 is required when assuming an incorrect GR model on MGR data. This is not the case when assuming an MGR distribution on GR data.

3. If there is genuinely a high b-value, we would need at least N > 1000 events to make this conclusion with a reasonable degree of certainty, although this is highly dependent on the underlying distribution and the associated parameters. The interpretation of such high b-values must be taken with caution because we know this is the signal expected when we are sampling the finite size exponential tail in the GR and the roll-off in the MGR.

In summary, our results show that it is extremely important to be critical in assigning significance to high b-values from the analysis of small datasets. This highlights the need to obtain a greater dynamic range of event magnitudes, particularly for volcanic and induced seismicity catalogues, through station deployment including high-resolution networks, access to borehole data where possible, adopting methods such as template matching to identify a greater number of smaller events and employing machine learning techniques on whole waveform data, hence improving the statistics. However, as Herrmann and Marzocchi (2020) showed, this may not be easy, as high-resolution catalogues have their own pitfalls due to particular behaviour toward smaller magnitudes (e.g. over-/under-representation, scaling break of local magnitude, etc.). In the long run, current improvements in detecting greater numbers of events in a catalogue can only improve the quantification of hazard, the understanding of underlying processes of volcanic and induced seismicity, and better inform risk management and mitigation strategies. Therefore, we recommend that any b-value

studies should always report – as a minimum – details on sample size, magnitude dynamic range and the method and justification for m_c determination used for the assessment. Without these basic parameters, it is hard to judge the significance of the reported results.

We have shown that with decreasing dynamic range and catalogue sample size, it is inherently harder to resolve which model correctly fits the available data. We have demonstrated that in one case (The Geysers), assuming GR to be the preferred model, may not be correct. This may be due to having sufficient data (N $\gg$ 1000) only for this particular case (see Table 4.3 for N above m_c) of volcanic or induced seismicity. This epistemic error in assuming the wrong (or not proven) model, may then propagate into biased estimates of b-values which could result in significant underestimation of the hazard from events larger than the maximum recorded magnitude obtained in the past.

4.8 Acknowledgements

We would like to thank David Eaton and one anonymous reviewer for their time and thoughtful and constructive feedback on our manuscript. We would also like to acknowledge the support for this work by the NERC E3 Doctoral Training Partnership grant NE/L002558/1.

Data Availability

The Central Italy and Mount Etna catalogues were obtained from INGV (Istituto Nazionale di Geofisica e Vulcanologia, `http://terremoti.ingv.it/en` last accessed December 2016), the El Hierro catalogue was obtained from the Instituto Geográfico Nacional (`https://www.ign.es/web/ign/portal/sis-catalogo-terremotos`) and The Geysers catalogue was downloaded from the United States Geological Survey (USGS, `https://earthquake.usgs.gov/earthquakes/search/` last accessed December 2020).

4.9 Re-evaluation of published b-value estimates

As part of the further analysis to this paper, I have re-analysed three cases of volcanic and induced seismicity (Etna, El Hierro and The Geysers) which I have presented in Tables 4.1 and 4.2 as being particularly high or low b-values, including limited sample sizes and dynamic range.

Following on from Murru et al. (2005), I use the same observation period, resulting in 456 events with a BVS m_c of 2.6. Estimating b-values for the maximum dynamic range (1.3) results in a typical b-value of 1.2, which is substantially lower than the 1.5 suggested by (Murru et al., 2005) but still elevated. Increasing the value of m_c then also increases the b-value, resulting in values as high as 2.5 when the dynamic range is only 0.6.

In the case of El Hierro, I also use the observation period which Ibáñez et al. (2012) used, resulting in 7894 events, with a BVS m_c of 1.5 which is slightly higher than in their case. However, it is important to note that Ibáñez et al. (2012) use the least-squares method to estimate the b-value, while I continue to use the MLE, which returns a b-value of 1.15 at maximum dynamic range (3.1). The b-value here does not steadily increase with decreased dynamic range because the FMD observed for this (time-)limited dataset does not show GR behaviour, indeed, it looks more like a 'characteristic' distribution, with another 'bump' at larger magnitudes.

In the final case re-evaluated here, I chose The Geysers between 1989 and 1994, resulting in 23 000+ events with a BVS m_c of 1.4, which is substantially higher than the Henderson et al. (1999) estimate of 0.5, taken by observing the deviation from linearity of the FMD at low magnitudes. At maximum dynamic range (3.0), a b-value of 1.0 is obtained, whereas reducing this dynamic range to 0.5 results in b-values as high as 3.1. Therefore, the b-values obtained by Henderson et al. (1999) of < 1.0 cannot be replicated with this (USGS) data here used here, within the same time period.

These three cases generally strengthen the discussion and conclusions in the paper above, that b-values are an artefact of small dynamic range (and sample sizes) and

lower b-values are recovered when more dynamic range and data are available.

4.10 Reflection

The results from this paper have addressed a common issue encountered in identifying truly high b-values from artificially high b-values and included suggestions for minimum values of sample size and dynamic range required to reduce the bias in b-values. Furthermore, we have recommended that any b-value studies include detailed information on the catalogues used, so that the inferred b-values can be critically assessed. I will further investigate this in the following chapter, where I extend the analysis from this paper to parameter changes in synthetic data and six more real catalogues.

Chapter 5

Sensitivity Analysis

The aim of this chapter is to provide a more complete understanding of the sensitivity of model selection, and its accuracy and precision, to a range of input parameters such as the underlying b-value, the magnitude of completeness m_c and the corner magnitude m_θ, as a function of the dynamic range of the data. This extends the analysis of Chapter 4 to a wider range of possibilities. In particular, I extend the analysis to further datasets introduced in Chapter 3, and expand on the generation and analysis of synthetic data with a known underlying distribution, as a bridge to understanding the real catalogue data.

5.1 Advanced Synthetic Catalogue Analysis

As in Chapter 4, I use a 'standard' set of parameters as a benchmark, and then change only one parameter at a time to maintain a controlled test of the sensitivity. I also look at changing more than one parameter at a time as a proxy for changing between tectonic, volcanic and induced seismicity settings. The standard values used are N = 10 000, b-value = 1.0, m_c = 1.0 and m_θ = 3.5 for the generation of MGR data. Fifty random catalogues are created (see section 4.4.2) and I use them to investigate the effects of changing any of these parameters, in particular the sensitivity of model selection to the dynamic range of the synthetic data.

The true underlying distribution in all of these scenarios is MGR. I have not shown the changes in GR data in this section, because its main aim is to identify potential issues and biases of using GR methods on potentially underlying MGR data. This problem can be addressed best in the analysis of synthetic data where the underlying parameters are known. The ΔBIC criterion (section 3.3.3) is used to discriminate which model is preferred, conditional on the synthetic input data and model parameters.

5.1.1 Changing true b

Here, I employ the 'standard' parameters but change the true, underlying b-value, while the other parameters remain constant. I test the hypothesis that increasing the slope will steepen the frequency-magnitude distribution (FMD) and make the roll-off less obvious. Thus it may not be sampled adequately by all catalogue simulations and some of the catalogues with $b > 1.0$ may prefer a GR model as opposed to the correct, underlying MGR distribution. Figure 5.1 presents this analysis in the same way as Figure 4.4, except that the final column shows ΔBIC as a function of dynamic range as opposed to number of events. The 4 rows correspond to different b-values and show the resulting FMD (left), model acceptance as a function of the dynamic range (where 1 = GR and 0 = MGR; middle) as well as the ΔBIC as a function of the dynamic range where GR > 0 and MGR < 0 (right). The four different b-values used here are a) 0.8, b) 1.0 (the 'standard'), c) 1.2 and d) 1.5. The first is generally regarded as a 'low' estimate of b, whereas 1.5 is a 'high' estimate. The FMDs shown in the first column increase in slope for higher b-value with the red vertical line indicating the 'underlying' corner magnitude $m_\theta = 3.5$. It is clear from the first column of Figure 5.1 that while the roll-off is sampled in all simulations for both $b = 0.8$ and 1.0 (Figure 5.1a) and b), for the larger values of b, not all catalogue simulations sample the roll-off adequately; there is no obvious curvature in the simulated data at high magnitude. As a consequence, these specific catalogues return an incorrect preference for a GR model over an MGR model, at least in some of the individual realisations (best shown by the preference of ΔBIC in the right-hand column) and in the overall probability of returning

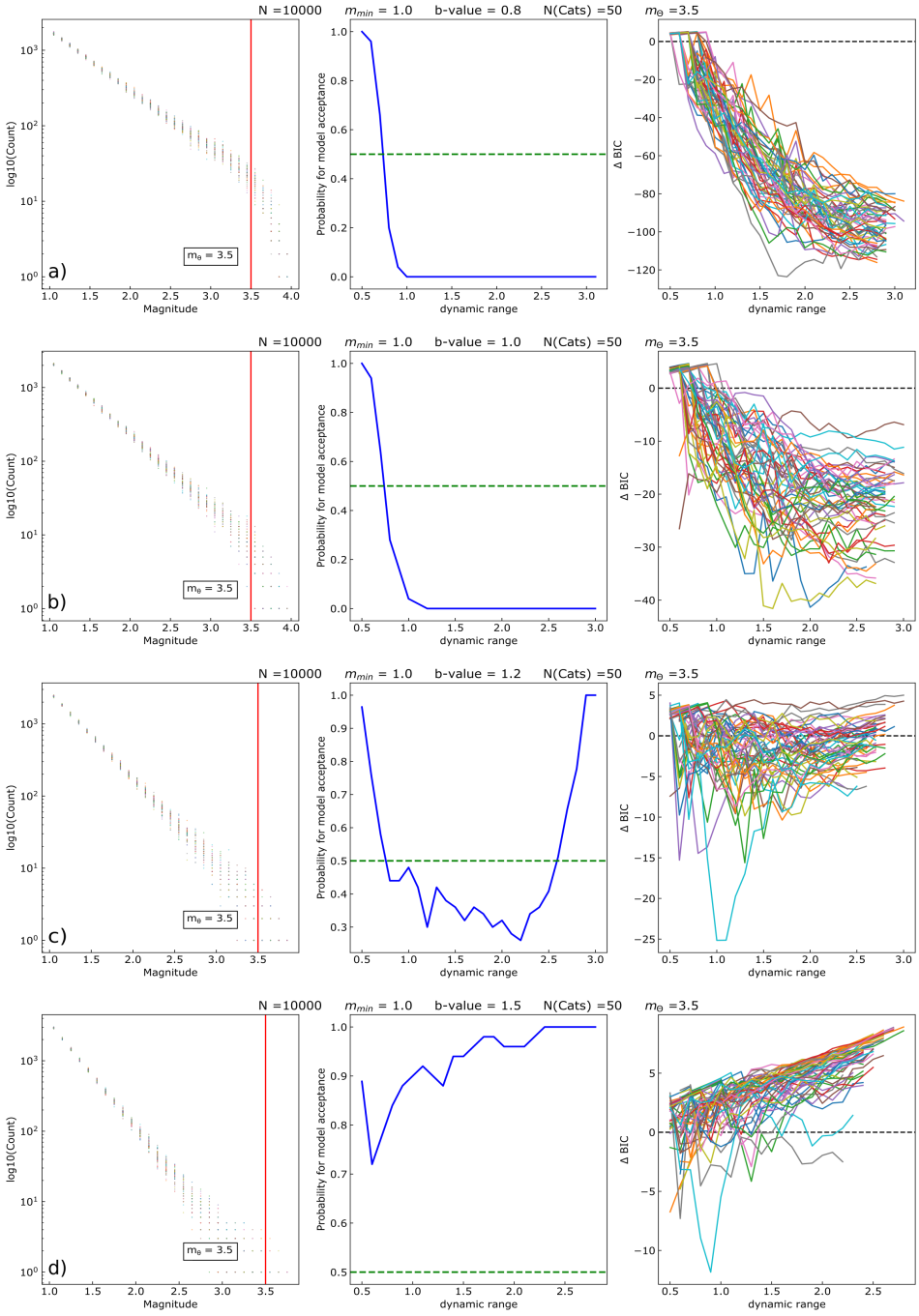

Figure 5.1: The FMD, model acceptance (1 = GR, 0 = MGR) and ΔBIC (GR > 0, MGR < 0) with respect to dynamic range are shown for each of four scenarios where a) $b = 0.8$, b) $b = 1.0$, c) $b = 1.2$ and d) $b = 1.5$. The vertical red line indicates the value of m_θ.

the incorrect model (middle column). This is particularly pronounced in the case of b = 1.5 (Figure 5.1d). This analysis implies underlying processes that produce an MGR distribution with higher b-values are more likely to be misidentified as GR. The model selection (middle column) always starts off preferring GR, because the dynamic range of < 1 provides too little data at all times and hence always prefers the simpler model, returning an incorrect result 100% of the time. Once the dynamic range is > 1.0, the model acceptance prefers MGR at high probability for the cases of b = 0.8 (Figure 5.1a) and 1.0 (b). For b = 1.2 in (c), increasing the dynamic range within the range shown may or may not result in the correct choice of the underlying MGR distribution. The GR is preferred again at large dynamic range, however, this is not part of a consistent pattern, and hence likely to be due to a sampling effect. In the case of b = 1.5 in d), visual inspection of the FMD realisations show that very few of the synthetic catalogues sample the roll-off adequately and ΔBIC generally prefers a GR model.

In summary, the results prove that the correct model (MGR in this case) is chosen when $b \leq 1.0$, and with absolute certainty above a dynamic range of approximately 1.2. In contrast, the incorrect model is chosen at high probability when b = 1.5, even when the dynamic range is large. In the case of b = 1.2, there is approximately a 30% chance of picking the incorrect model in the dynamic range of 1 and 2.5.

5.1.2 Changing true m_{min}

In this section, I examine the sensitivity of model selection to m_{min}, for a constant b-value of 1.0 and keep all other parameters at the 'standard' values (section 5.1). The motivation is to reflect the variability of m_{min} in different catalogue types with magnitude thresholds for complete reporting. This can range anywhere from what is common in microseismicity (sometimes less than magnitude zero with access to nearby borehole instruments), up to reasonable values typical of a regional network.

The values of m_{min} here are a) -1.0, b) 0.5, c) 2.0 and d) 3.0. The 'standard' value of 1.0 is shown in Figure 5.1b) and has therefore not been reproduced here. At m_{min} = -1.0 (Figure 5.2a), left diagram), visual inspection of the FMD shows that for all catalogue

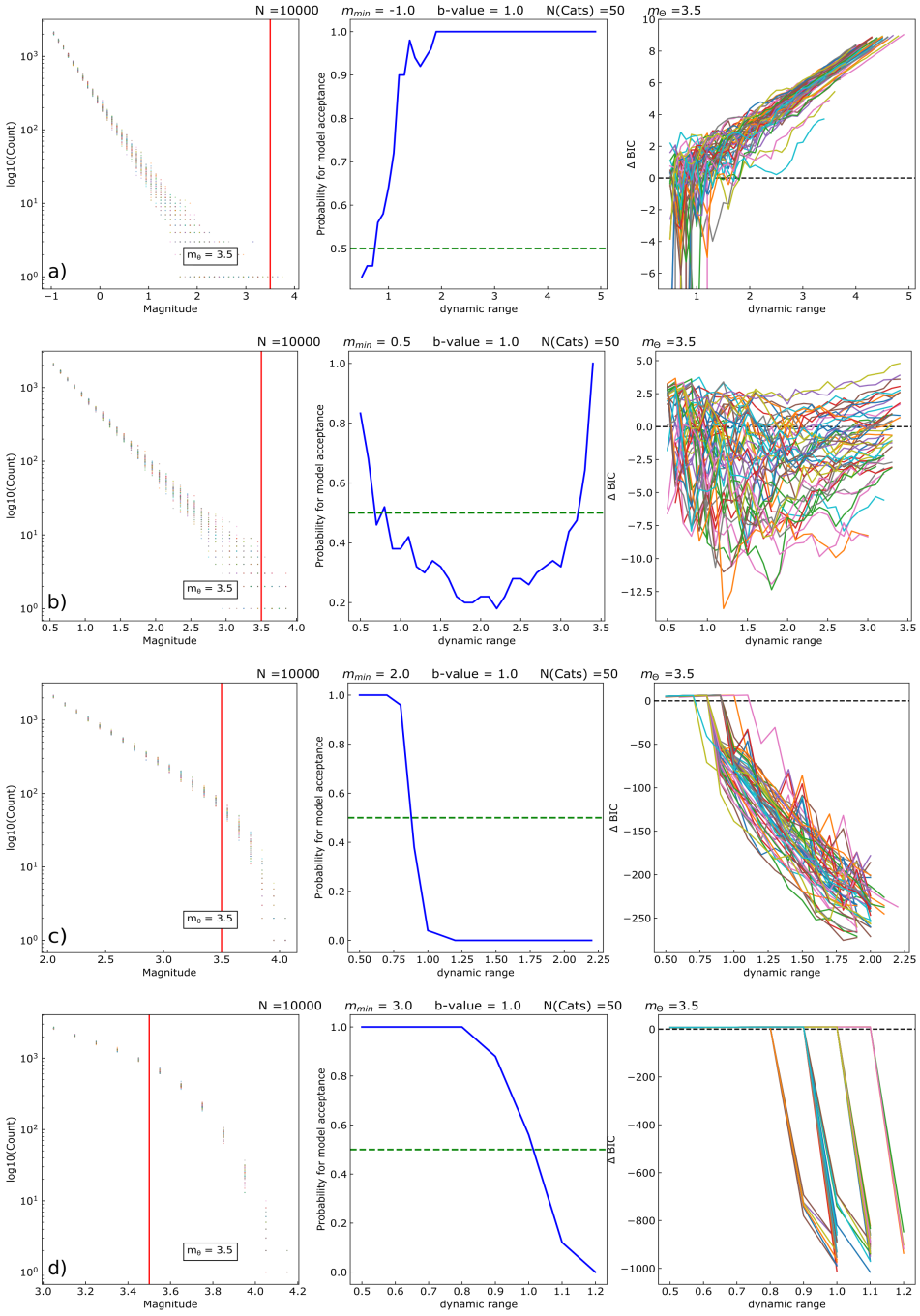

Figure 5.2: The FMD, model acceptance (1 = GR, 0 = MGR) and ΔBIC (GR > 0, MGR < 0) with respect to dynamic range are shown for each of four scenarios where a) m_{min} = -1.0, b) m_{min} = 0.5, c) m_{min} = 2.0 and d) m_{min} = 3.0.

simulations, the roll-off is not sampled adequately by most of the synthetic catalogues, resulting in the incorrect preference of the GR model at a very high probability (Figure 5.2a), middle diagram) due to lack of data. If the number of events in these catalogues were increased (e.g. to 100 000), then more of the catalogues would sample the roll-off sufficiently and return a correct preference for the MGR model. With increasing m_{min}, more and more of the simulated catalogues in Figure 5.2b), c) and d), first column, sample the roll-off adequately, allowing the correct model to be preferred, as confirmed in the model acceptance (middle column) and ΔBIC (right column) panels. In Figure 5.2b), GR is preferred above a dynamic range of ~ 2.5. In Figure 5.2c) and d), the dynamic range is significantly reduced, due to large m_{min} (in comparison to m_θ). Nevertheless, the correct model is returned at high probability despite lack of data. However, in the case of m_{min} = 3.0, the dynamic range is so small that almost only the roll-off is sampled, not returning a very significant result (also indicated by the magnitude of ΔBIC, Figure 5.2d), right column).

Summarising these results, they show that the correct model is returned 100% of the time when m_{min} is at least 1.0 (from Figure 5.1b) or larger (Figure 5.2c) and d) and has a minimum dynamic range of 1.2. The correct model is returned approximately 70% of the time when m_{min} = 0.5 (b), even though the incorrect model (GR) is preferred again at larger dynamic range due to a sampling effect. At m_{min} = -1.0 the value of m_θ is relatively too large for the FMD to sample it with the number of magnitude data points available, returning only the incorrect GR model, when the dynamic range is > 1.

5.1.3 Changing true m_θ

Here, I again return b and m_{min} to the 'standard' parameters (section 5.1) and examine the sensitivity of model choice with respect to changing m_θ. The true, underlying distribution remains MGR.

The values of m_θ I have chosen for the sensitivity analysis in this case are a) 2.0, b) 3.0, c) 4.0 and d) 5.0. The results for the 'standard' value of 3.5 is shown in Figure 5.1b) and has therefore not been reproduced here. As with the previous two parameter

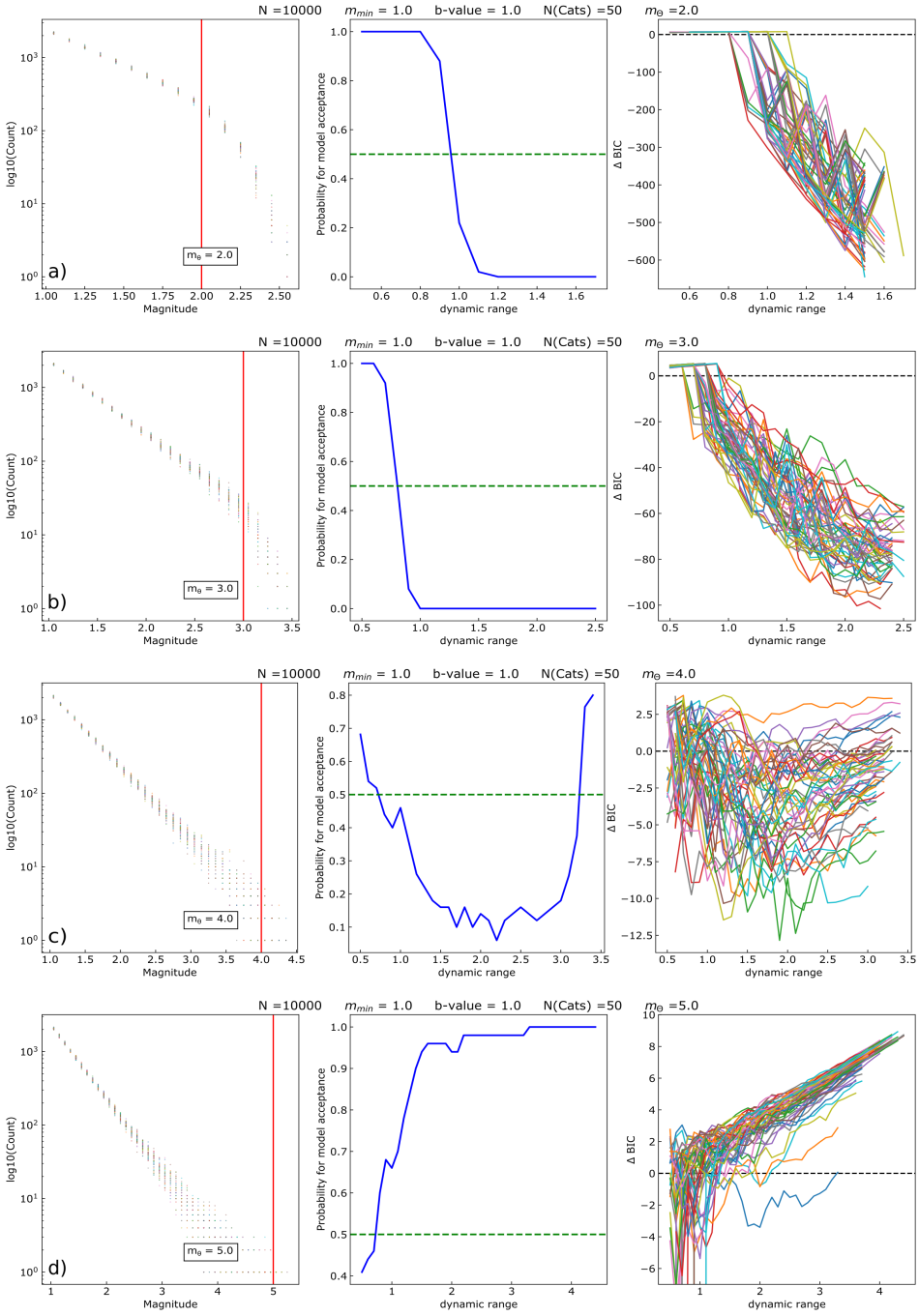

Figure 5.3: The FMD, model acceptance (1 = GR, 0 = MGR) and ΔBIC (GR > 0, MGR < 0) with respect to dynamic range are shown for each of four scenarios where a) m_θ = 2.0, b) m_θ = 3.0, c) m_θ = 4.0 and d) m_θ = 5.0.

changes, there is a clear trend in model preference in Figure 5.3. The FMDs suggest that the roll-off is sampled by all catalogue simulations in the first two scenarios where m_θ = 2.0 and 3.0 (Figure 5.3a) and b), returning the correct preference of underlying distribution – MGR in the middle column. Figure 5.3 shows a mixture of preferences for GR and MGR, with catalogues tending towards GR with increasing dynamic range, and entirely exhibiting this behaviour for a dynamic range above ~ 1.5 in d). This preference for the incorrect underlying distribution occurs because the value of m_θ is overestimated for the data simulated in the best fit MGR model.

The results here prove that when the range between m_{min} and m_θ is < 3, the correct model is returned when the dynamic range is at least 1.0. When m_θ is 4.0, about 90% of the simulations return the correct model between dynamic ranges of 1.0 and 3.0. After this point, the sampling effect returns the preference back to the incorrect model. When m_θ = 5.0, it is too large to be sampled by most catalogues and the incorrect model is returned at all times.

5.1.4 Changing multiple parameters

As I have shown in all three sections above, changing just one variable is a good test to isolate the sensitivity of the results to the changes in individual parameters. However, changing just one parameter is not necessarily representative of how real data might change in earthquake catalogues from different settings. Therefore, in this section, I have changed multiple of the 'standard' parameters simultaneously (see Table 5.1), to simulate the three types of seismicity (tectonic, volcanic and induced) which I examine later using the real data (section 5.2).

Figure 5.4 shows two different cases of tectonic seismicity with parameters of a) m_{min} = 2.5, b = 0.95 and m_θ = 6.0 and b) m_{min} = 4.5, b = 1.0 and m_θ = 9.0. The underlying magnitudes are sampled from a parent MGR distribution but it is clear that the roll-off used is not adequately sampled in either case, even for the relatively large data sample shown. This results in an increasing preference for the GR model with increased dynamic range, as the straight line portion of the MGR distribution is extended.

Mimic Catalogue	m_{min}	b**-value**	m_θ
Tectonic	2.5	0.95	6.0
Tectonic	4.5	1.0	9.0
Volcanic	1.5	1.0	3.5
Volcanic	2.0	1.2	5.5
Induced	-1.0	1.0	1.5
Induced	-1.0	1.5	1.5
Induced	-1.0	1.0	3.5

Table 5.1: Summary of the parameters changed and represented in Figures 5.4 (tectonic), 5.5 (volcanic) and 5.6 (induced).

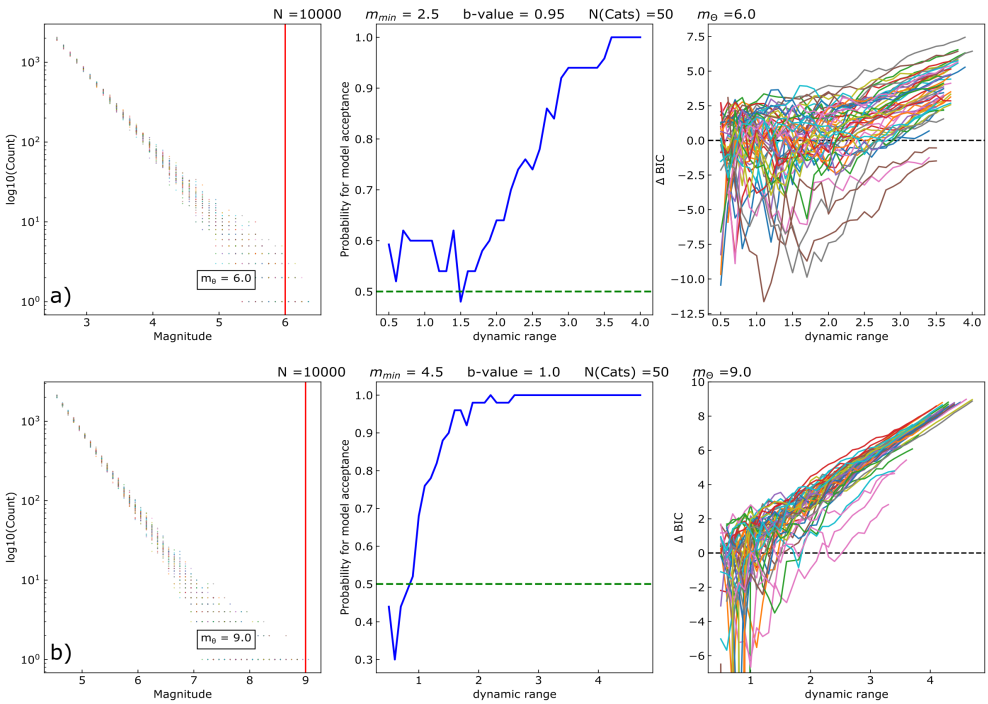

Figure 5.4: The FMD, model acceptance (1 = GR, 0 = MGR) and ΔBIC (GR > 0, MGR < 0) with respect to dynamic range are shown for a) where m_{min} = 2.5, b = 0.95 and m_θ = 6.0 and b) where m_{min} = 4.5, b = 1.0 and m_θ = 9.0 to mimic tectonic seismicity. The vertical red line is the value of m_θ.

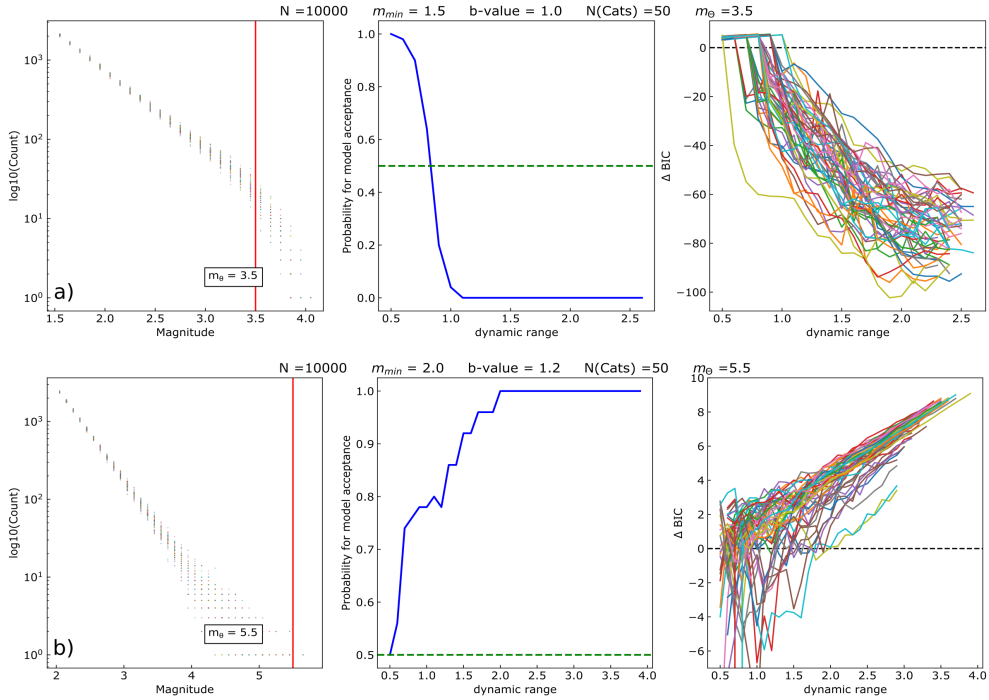

Figure 5.5: The FMD, model acceptance (1 = GR, 0 = MGR) and ΔBIC (GR > 0, MGR < 0) with respect to dynamic range are shown for a) where m_{min} = 1.5, b = 1.0 and m_θ = 3.5 and b) where m_{min} = 2.0, b = 1.2 and m_θ = 5.5 to mimic volcanic seismicity. The vertical red line is the value of m_θ.

Figure 5.5 shows a similar analysis for a) m_{min} = 1.5, b = 1.0, m_θ = 3.5 and b) m_{min} = 2.0, b = 1.2, m_θ = 5.5, to reflect possible volcanic seismicity settings. Figure 5.5 shows clearly that a) prefers the correct, MGR model whereas in b) the incorrect GR model is returned. The reason, again, is that a) has a good sample of the roll-off and b) does not (see first column of Figure 5.5). In both cases, the model preference becomes increasingly certain with respect to increasing dynamic range, as the straight-line portion of the curve becomes better sampled.

Figure 5.6 repeats the analysis for parameter values that could be observed in induced microseismicity. The parameters used are a) where m_{min} = -1.0, b = 1.0, m_θ = 1.5, b) where m_{min} = -1.0, b = 1.5, m_θ = 1.5 and c) where m_{min} = -1.0, b = 1.0, m_θ = 3.5. Figure 5.6a) uses parameters which sample the roll-off well, cases b) and c) do not (left column). As a consequence, a) shows increasing preference for the correct model as the dynamic range increases, and b) and c) show the opposite. This implies that

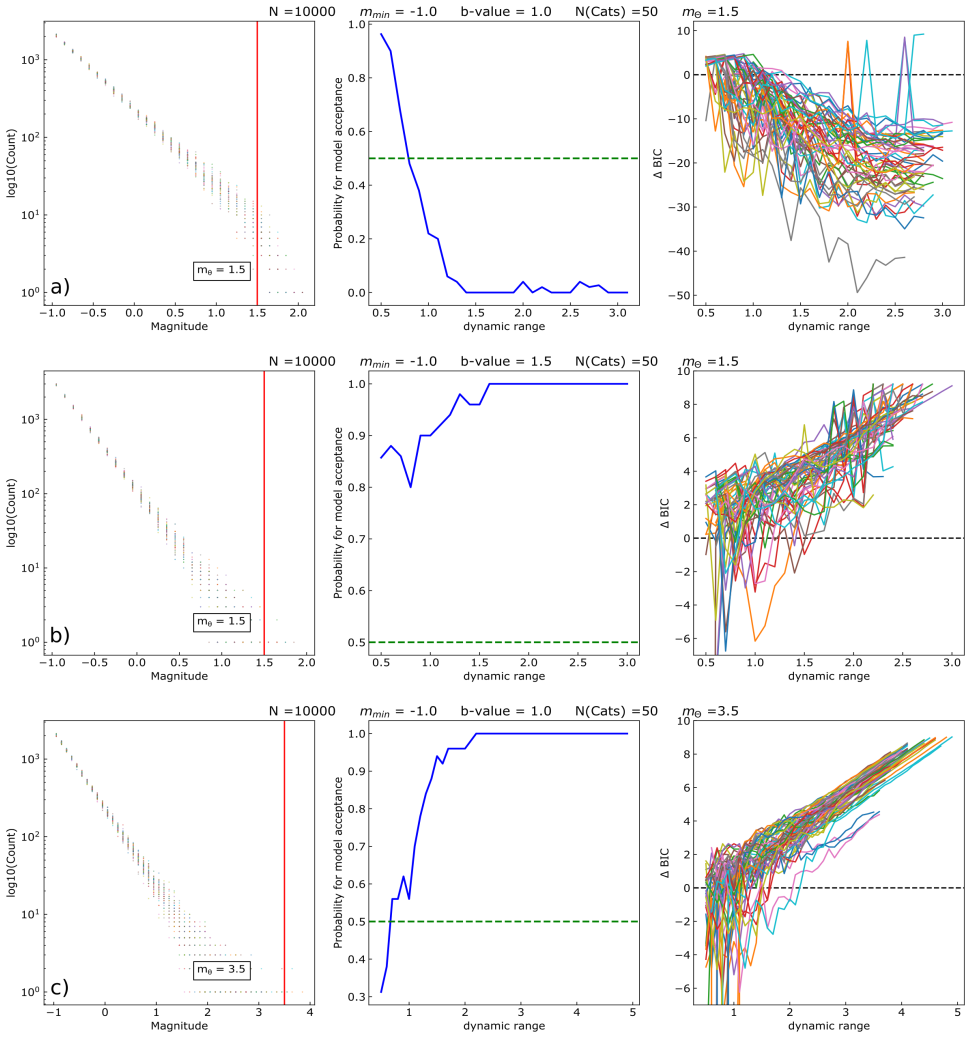

Figure 5.6: The FMD, model acceptance (1 = GR, 0 = MGR) and ΔBIC (GR > 0, MGR < 0) with respect to dynamic range are shown for a) where m_{min} = -1.0, b = 1.0 and m_θ = 1.5, b) where m_{min} = -1.0, b = 1.5 and m_θ = 1.5 and c) where m_{min} = -1.0, b = 1.0 and m_θ = 3.5 to mimic microseismicity. The vertical red line is the value of m_θ.

microseismic data with similar underlying parameters is almost certain to return the correct model for a dynamic range of 1.5 and above, with the sample sizes used. For the parameters used in b) and c), this is less obviously the case, as the roll-off is only sampled by few of the 50 catalogue simulations. In reality, the choices of m_θ used here are very low for microseismicity (especially in Figures 5.6a) and b) because it is not necessarily common to observe MGR in these cases. Here, the value of m_θ is small enough for events to fall to the right side of this value and the roll-off is observed in the synthetics. These results show that the underlying m_θ is crucial because this controls whether the correct underlying distribution can be identified, for a given number of events and dynamic range.

These results show that when mimicking tectonic seismicity (on underlying MGR data, Figure 5.4), the GR model prevails in both parameter changes shown. However, for Figure 5.5, in a) with the parameters chosen (m_{min} = 1.5, b = 1.0 and m_θ = 3.5), the correct model (MGR) is always returned for dynamic ranges greater than 1.1. This is no longer the case when increasing all parameter values (in b), where the preferred model is GR with certainty (model preference = 1) in all cases with dynamic ranges greater than 2.0. When mimicking induced seismicity in Figure 5.6, similar observations can be made – only in a) is the preferred model certainly MGR (model preference = 0) at a dynamic range greater than 1.5. In cases b) and c), when both b and m_θ are varied, the tendency is towards GR with increasing dynamic range (> 1.5).

5.1.5 Biases in b-value estimation

Here, I examine the potential biases in b-values inferred from Aki's (1965) method (based on a maximum likelihood fit to magnitude data) and those obtained from the maximum log-likelihood method applied to seismic moment data, then converting back to magnitudes. In Figure 5.7, I show two synthetic examples, 5.7a) is based on underlying GR data and 5.7b) is based on underlying MGR data. Both of these use the 'standard' parameters of m_{min} = 1.0, b = 1.0 and m_θ = 3.5 for the MGR events. I show the relationship between the Aki (GR) b-value (calculated using magnitudes and equation

2.7 reproduced from (Aki, 1965)) and the estimated GR (red circles) and MGR (blue circles) b-values obtained from the maximum log-likelihood (see equations 4.4 and 4.5). Both plots show that the Aki b-value and the b-value inferred from the maximum likelihood method on seismic moment data assuming a GR distribution (red circles) are basically equal (the difference between the two are of order 10^{-5}). This confirms there is no bias in this case between the Aki method and the maximum likelihood method applied to seismic moment data, irrespective of the underlying distribution.

I also show the pattern obtained when using the maximum log-likelihood method on seismic moment data to estimate MGR b-values (blue circles) on both the GR (top) and MGR (bottom) data in Figure 5.7. The most noticeable observation is the difference in the spread of data (blue circles). For the GR data and for the MGR data with approximately $b > 1.7$, the GR b estimate again is in perfect or very good agreement with the Aki result (red and blue circles respectively). In contrast, the MGR b-values below magnitude 1.7 are commonly biased to high values (cloud of blue circles) when applying Aki's method. This is because the curvature in the MGR law is interpreted as a steeper slope in the data.

For the MGR data, the cluster of MGR b-values is much more pronounced around $b \sim 1.0$ in Figure 5.7b), than it is in a). In this cluster, the corresponding Aki (GR) b-values are exclusively larger than the underlying value of 1.0, with a strong focus on $1.0 < b < 1.3$, and some returning values of b close to 2.0. This confirms that using Aki's (GR) formula to estimate b in the case of an underlying MGR distribution, returns high, and incorrect, b-values. In contrast, large underlying MGR b-values (> 1.7) in b) are well captured by the Aki GR estimates, even though the data is MGR distributed. In comparison, MGR b-value estimates in Figure 5.7a) do not return values (excluding outliers) larger than ~ 1.5.

The results presented here suggest Aki b-values are likely to be biased to high values, especially in cases where the underlying b-value < 1.5 and the underlying distribution is MGR.

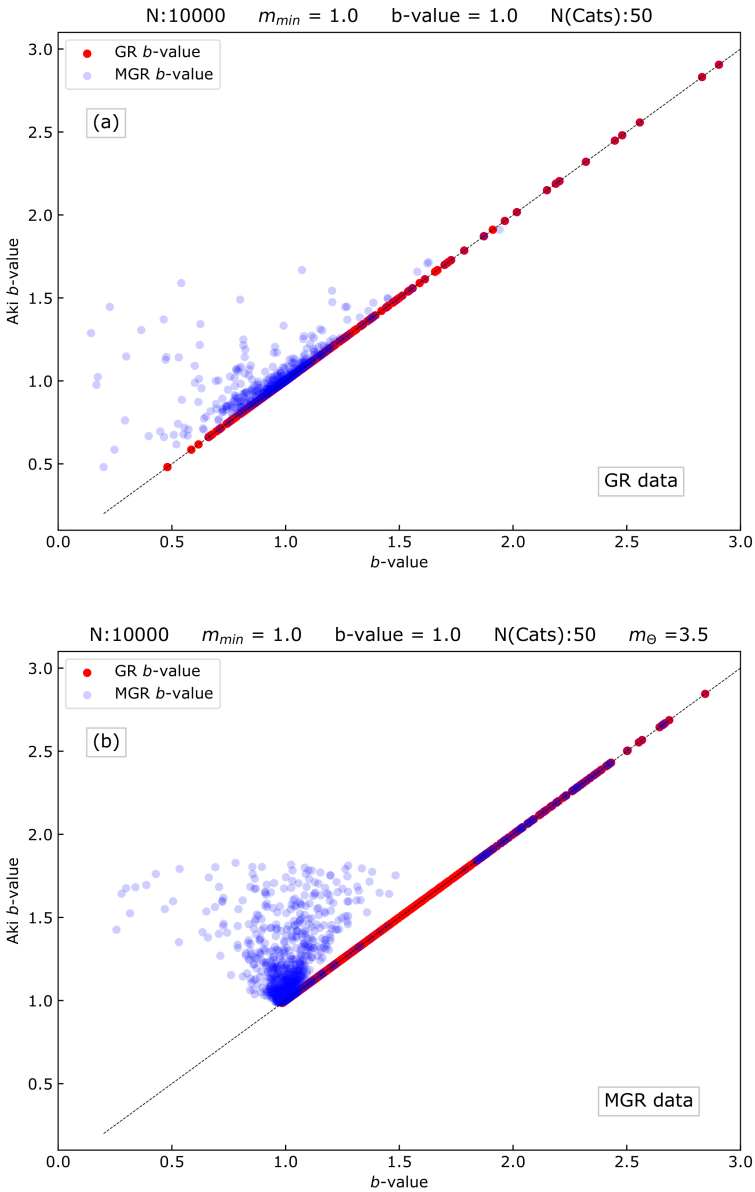

Figure 5.7: Aki b-value as a function of b-value estimated through maximising the log-likelihood function on seismic moment data. a) has underlying GR data and b) has underlying MGR data. Red circles are GR b-values, blue circles are MGR b-values.

5.1.6 Changes with dynamic range

One of the main motivations for investigating potential biases here is the relatively low dynamic range involved in many published studies. Therefore, I now examine the effect of increasing the dynamic range on the b-value by increasing the maximum magnitude in any given synthetic catalogue and varying the minimum magnitude of completeness to achieve different dynamic ranges for a given maximum magnitude.

Figure 5.8 shows the sensitivity to maximum magnitude and dynamic range with 50 catalogues randomly simulated for each set of synthetic data. The red circles are for underlying GR data and the blue circles for underlying MGR data. The number of events (N) in each catalogue for a given dynamic range is not the same. After the catalogues have been simulated, they are thinned by increasing m_c and once a specific dynamic range (i.e. 1 to 6) has been reached, the corresponding maximum magnitude and b-value are shown in Figure 5.8. In the case of MGR (last row), there is not enough data to exceed a dynamic range of 3 due to the relatively small (in global terms) m_θ. The headings of the subplots indicate the size of the dynamic range that is in any given subplot; for GR these range from 1 to 6, for MGR these range from 1 to 3. Each subplot shows b-value estimates and maximum magnitudes for the different synthetic catalogues, with the true b-value indicated by a horizontal, dashed black line. Due to the random nature of catalogue sampling, there are fewer data points for larger dynamic ranges, as fewer synthetic catalogues sample these larger dynamic ranges. For a GR dynamic range of 1 (Figure 5.8, top left diagram), the b-value estimates range from 0.4 to 1.6 (with a true b of 1.0). This implies that b-values are not necessarily biased to 'high' values in datasets of limited dynamic range but instead they simply become much more uncertain under these conditions. With increased dynamic range, this range of b-value estimates becomes continuously smaller. With a dynamic range of 3 (Figure 5.8, top right diagram), b-values still range between 0.7 and 1.2, in this set of synthetics. However, the estimated b-value and its range is converging to the true value of $b \sim 1.0$ as the dynamic range increases. A b-value of 1.2 is still regarded as 'high' in

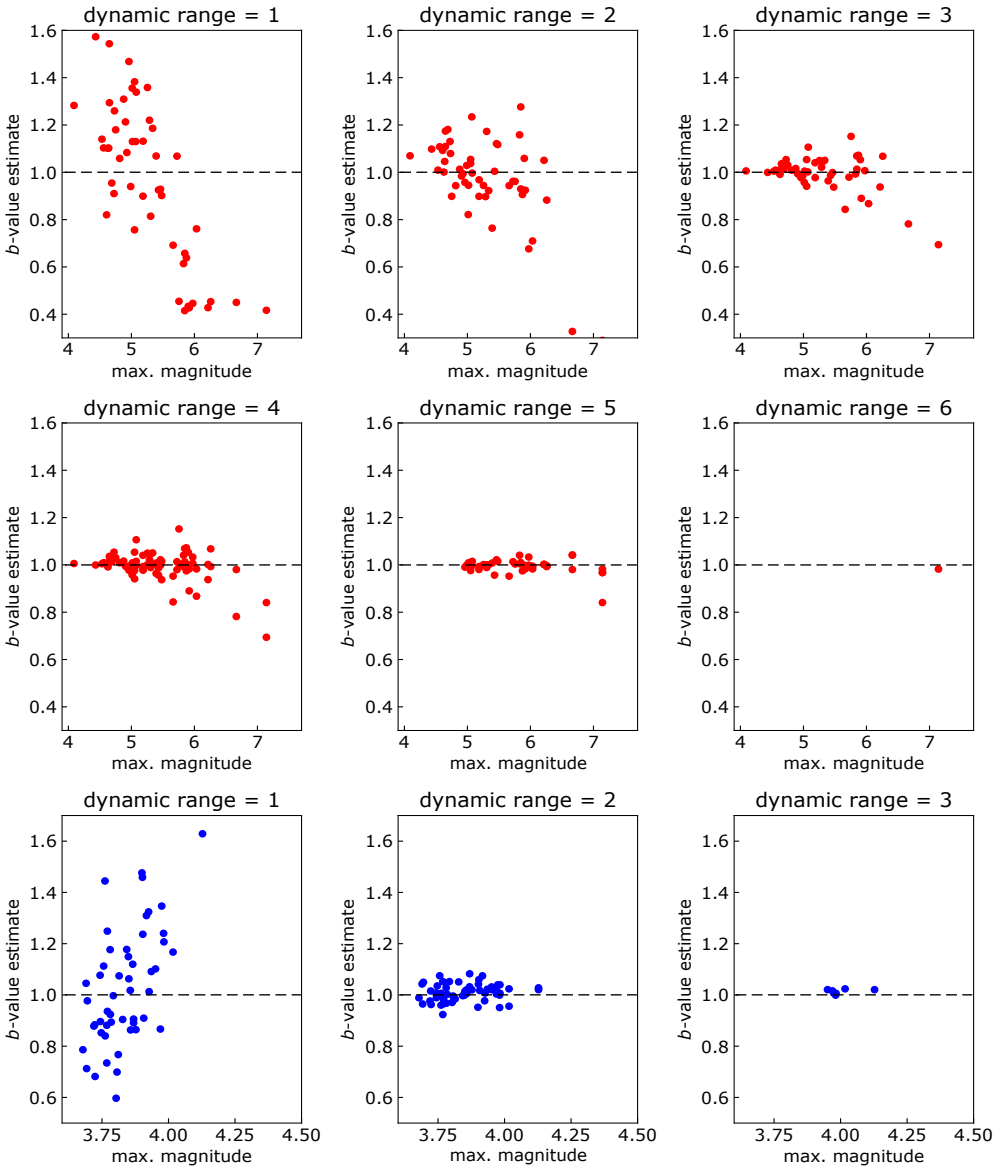

Figure 5.8: *b*-value estimates shown as a function of the maximum magnitude for different dynamic ranges (given in the subplot headers). The first 6 plots show GR data (red circles) and the bottom three plots show MGR data (blue circles). The true *b*-value of 1.0 is shown as a horizontal, dashed line.

most literature. Once a dynamic range of 5 is reached (Figure 5.8, middle diagram), the estimated b-value is close to the true value, with one apparent outlier at high m_{max} with $b < 1.0$. No b-value estimates reach above 1.05. A similar analysis of MGR synthetic data in 50 catalogues (Figure 5.8, lower row), also shows a large scatter around the true b-value for a dynamic range of only one unit, but much quicker convergence to the true b-value for dynamic ranges of 2-3. This suggests the covariance between b and m_θ is responsible for the bias and scatter in the top row, i.e. b is much less scattered and nearer the true value for an MGR model. For example, at a dynamic range of 2, all b-value estimates are already between 0.9 and 1.1. In summary, the spread in b-value estimates substantially decreases with increased dynamic range and more quickly for the MGR simulations.

5.1.7 Importance of m_θ within an FMD

With the focus of changing single and multiple parameters in the sections above, comes one main realisation – that the controlling factor as to whether a catalogue exhibits true and unbiased MGR data or not (as shown by the probability of model acceptance in the middle column and the value of ΔBIC of Figures 5.1 through 5.6), is only dependent on the numbers of events above and below the value of m_θ. Throughout this chapter, I have shown in how many ways this number of events can be altered and how that changes the model preference of the simulated catalogues. However, in sections 5.1.1 through 5.1.4, when selecting the cases where the model acceptance and ΔBIC tends to the correct MGR distribution, the corresponding FMDs show that at least ~ 50 events are required for the roll-off to be observed and the correct model to be chosen, agreeing with Main et al. (1999). I have chosen to plot model acceptance and ΔBIC in all cases against dynamic range rather than number of events; even though they are inherently linked, I have emphasised the importance of having enough dynamic range (not just enough sample size) for accurate determination of the correct model. This is further investigated in sections 5.1.8 and 5.3, particularly with regards to the b-value. If m_θ is large, no simulated catalogues sample the roll-off and if m_θ is small, all simulated

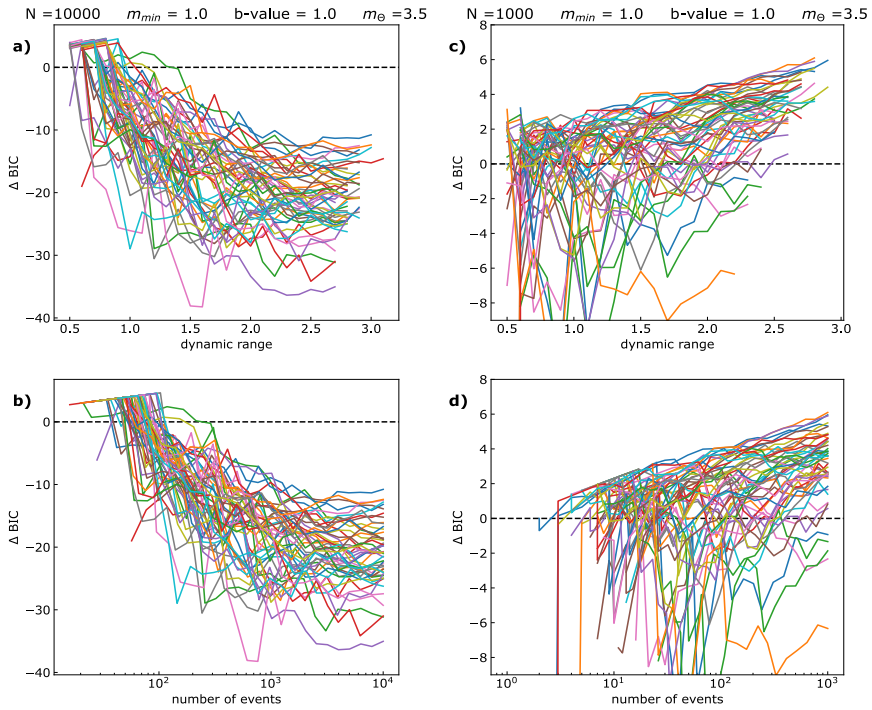

Figure 5.9: ΔBIC for synthetic catalogues as a function of dynamic range in a) and c) and ΔBIC as a function of number of events in b) and d) for parameters shown at the top.

catalogues will sample the roll-off and hence the correct MGR distribution. Therefore, careful and critical assessment of m_θ is required to determine whether it has practical or only theoretical applicability.

5.1.8 Relation between dynamic range and N

While I have targeted the importance of dynamic range, as many previous studies have focused solely on the importance of N (see Chapter 2), I acknowledge that there is an inherent relationship between these two factors; generally, small dynamic range will also have low N. Therefore, the observed biases in the b-value would likely be caused by low N and not lack of dynamic range. I present some further analysis on synthetic data to clarify the importance of dynamic range. To do this, I have chosen to only focus on the ΔBIC, both as a function of dynamic range and N for the same parameters. These are shown in Figures 5.9 and 5.10. The number of events (bottom row) in all cases has been shown on the log scale for visual ease.

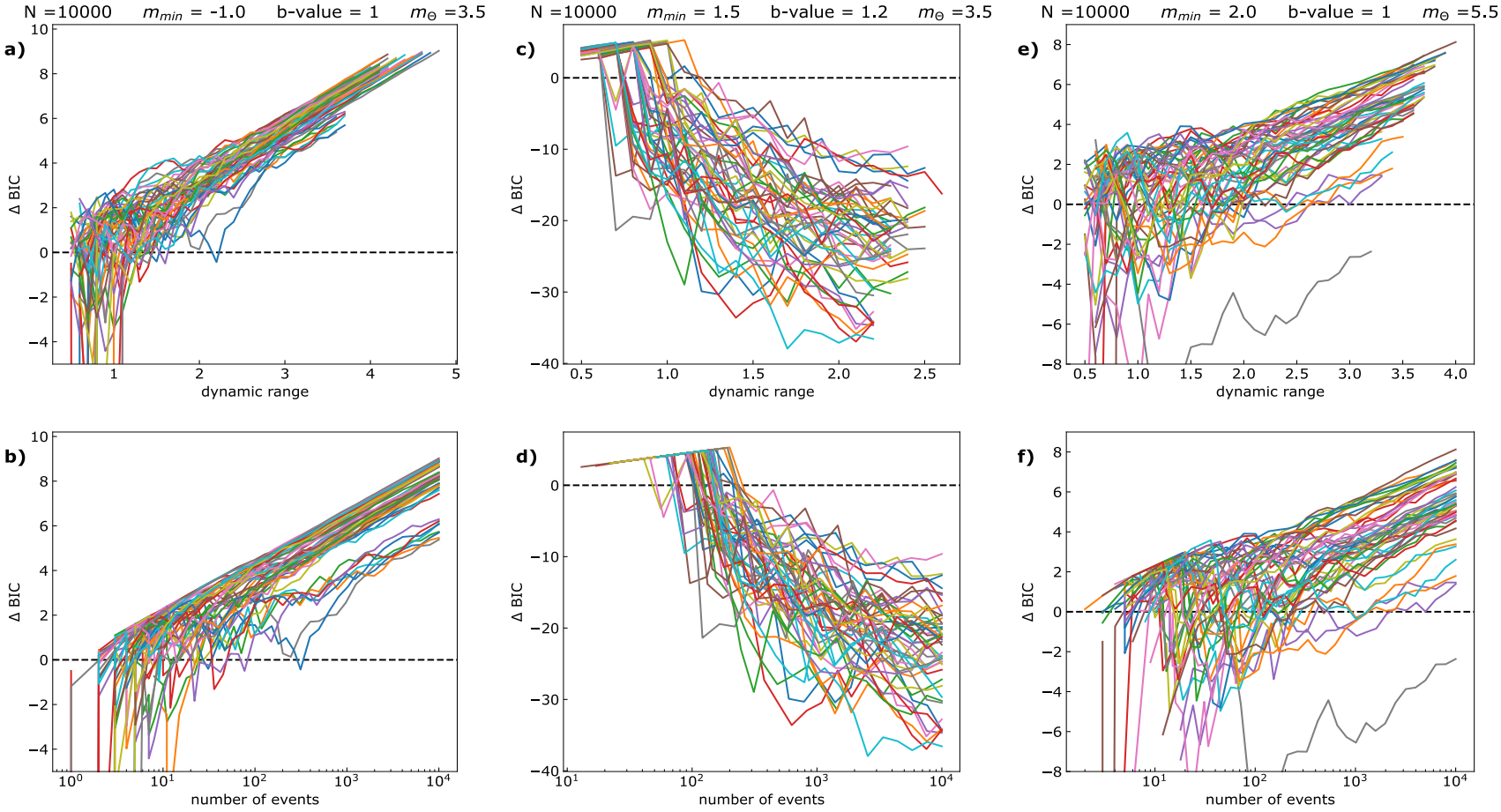

Figure 5.10: ΔBIC for further synthetic catalogues as a function of dynamic range in a), c) and e) and ΔBIC as a function of number of events in b), d) and f).

In Figure 5.9, I show the same 'standard' parameters (m_{min} = 1.0, b = 1.0, m_θ = 3.5) in both the left column and the right column, with the only difference being the number of events – 10 000 on the left and 1000 on the right. The true underlying distribution is MGR. This is returned by all simulated catalogues in Figure 5.9a) above a dynamic range of 1.0 and with more than ~ 100 events (Figure 5.9b). Using the same parameters but reducing N to 1000 now only returns the correct MGR distribution in very few catalogues and with more dynamic range increasingly prefers the incorrect GR distribution in both Figures 5.9c) and d). Interesting is the comparison between Figures 5.9a) and c). Even though a) has many more events, the average dynamic range observed in the catalogues is around 2.8. In the case of c), the average dynamic range is approximately 2.6, which is only minimally smaller than in the case with more events. Not only does this show the importance of dynamic range but it also shows that N and dynamic range can be separated to some extent. In Figure 5.10, I show further examples, where in all cases N = 10 000 but the observed dynamic range varies depending on the parameters of m_{min}, b and m_θ used. Even though there is ample dynamic range in the left column (Figure 5.10a) and b), a limitation of these parameter choices is that basically no events occur to the right hand side of m_θ (section 5.1.7), shown by the fact that all catalogues prefer the incorrect GR distribution at dynamic ranges above 1 and when N > 100. The middle column shows the smallest dynamic range, averaging 2.4, despite the same value of N being used. Again, this is due to the fact that the parameters are favourable towards events above m_θ, leading to the roll-off being observed (unlike in Figure 5.10a) and b). The parameters used to obtain Figures 5.10e) and f) have only few events occurring above m_θ, shown by the preference for the GR distribution.

This section has exemplified the issue discussed in section 5.1.7, in that the controlling factor for observing the MGR distribution is the events above and below the value of m_θ. I have shown here that this varies with any change of parameters but also that N and dynamic range are not necessarily proportional. These findings are also directly applicable to real datasets, where we have no immediate control over N or the dynamic range available.

5.2 Advanced Real Catalogue Analysis

This section presents a more advanced analysis of the real data analysed in Chapter 4, informed by the sensitivity analysis presented in section 5.1. All the additional catalogues looked at in this section have been briefly introduced in section 3.5. To ease comparison in the analysis, the results and figures are presented in the same format as those in Chapter 4.

5.2.1 Tectonic Seismicity Catalogues

Further to the Italy catalogue(s) discussed in Chapter 4, I use two other tectonic catalogues to assess potential b-value biases stemming from potentially poor assessment of the underlying FMD. The two additional catalogues are from the Reykjanes Peninsula, Iceland and Southern California in the US, chosen for their abundance of accessible data (Table 3.1) prior to estimating m_c, and the potential to compare my results with significant published literature on these much studied tectonic regions.

5.2.1.1 Reykjanes Peninsula

Iceland has the highest seismic hazard in Northern Europe, comparable to that of Southern Europe (Jónasson et al., 2021). This is due to its unique geographical location on the Mid-Atlantic Ridge, a very volcanically and seismically active area due to its setting at the plate boundary between the North American and the Eurasian plates. The Icelandic Meteorological Office (IMO) is the country's national weather service but is also in charge of volcanic and seismic monitoring. I have briefly introduced the Reykjanes catalogue in section 3.5.2.2 explaining the importance of the abundance of events and the very active seismological setting, with shallow strike-slip tectonic earthquakes and volcanic earthquakes. This catalogue was obtained from the IMO website (`http://hraun.vedur.is/ja/viku/`) and provides the following data: Date/Time, latitude, longitude, depth and two magnitudes - one of which is designated as m – a local moment magnitude (Panzera et al., 2017) and the other being m_L, a local

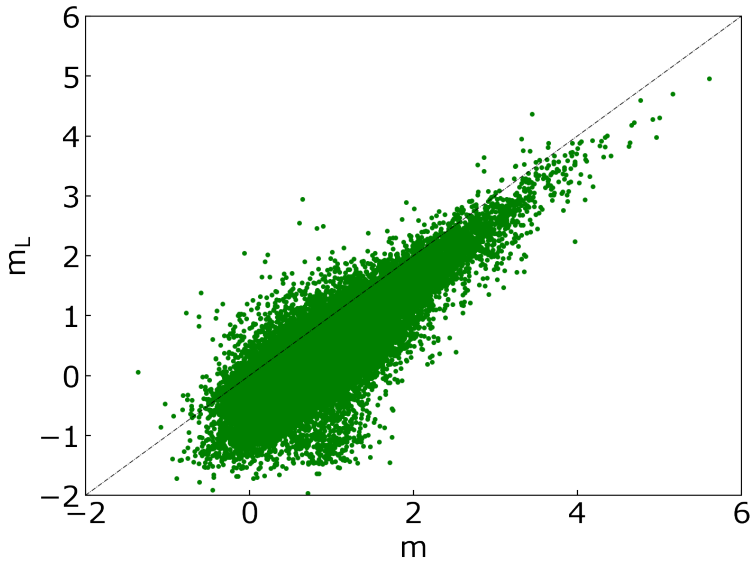

Figure 5.11: Comparison of m_L as a function of m, with most event magnitudes falling below the dashed black line representing $m = m_L$.

magnitude. The comparison of these magnitudes is shown in Figure 5.11. Most points fall beneath the dashed line indicating $m = m_L$, suggesting that in most cases where both magnitudes exist for a specific event, m is larger than the equivalent m_L.

The incomplete catalogue contains 50 453 events, which is reduced first by removing the events without existing magnitudes. While most events have a reported m and m_L, some events only have a reported m or an m_L value, rendering two catalogues of different sizes – m with 49 441 events, m_L with 50 379 events.

In Figure 5.12, I show the FMD of the full, incomplete, catalogue and the BVS curve. Both magnitude types are shown, with m in blue and m_L in red. It is clear that the m_L values are generally smaller than the local moment magnitude m values, but also give a relatively more complete FMD at the very lowest magnitudes (Figure 5.12a). In Figure 5.12b), I show the b-value estimate as a function of m_c to determine the optimal value of m_c within errors to one standard deviation. Using both MaxC or GFT methods (section 3.2.1), returns smaller values for m_c, keeping more events in the nominally 'complete' catalogue. However, I adopt a conservative approach and choose BVS to ensure completeness and minimise any associated systematic bias in the

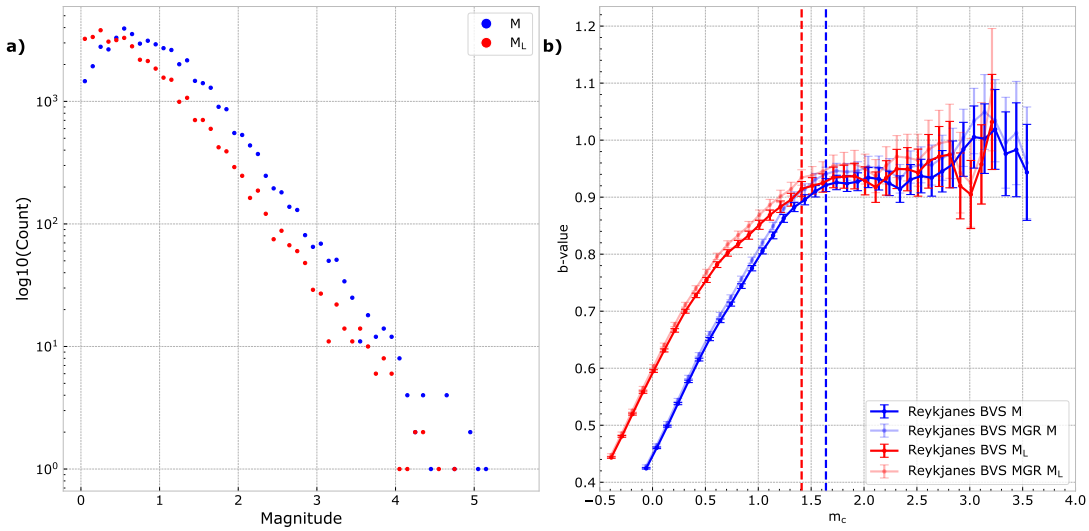

Figure 5.12: a) FMDs for both magnitude types (m in blue, m_L in red), including incompletely reported data at small magnitudes. b) The BVS curve using GR (solid colours) and MGR (lighter colours) for both magnitude types. Vertical dashed lines indicate the returned BVS m_c estimates, based on successive b-values falling within error (one standard deviation) of the b-value.

b-value estimate to low values associated with analysing potentially incomplete low magnitude data. I have also modified the BVS method (which assumes an underlying GR model) to assume an underlying MGR model. To my knowledge, this is the first time this has been adapted for such purpose. The results of both of these analyses are shown – the BVS GR in solid colours and the BVS MGR in lighter colours in Figure 5.12. In this specific case, the GR and MGR return the same values of m_c, but this is not always the case (section 5.2.3.1). The errors between GR and MGR for a given trial m_c are similar, suggesting a potential lack of sensitivity of the BVS to the assumed models in this tectonic region. For m, the BVS m_c is 1.64 and for m_L, the BVS m_c estimate is 1.41 (see catalogue details in Table 5.2).

Figure 5.13a) demonstrates that both FMDs (for m and m_L) converge to values of approximately $b = 0.92$. There is a local peak observed in the b-value for the m_L catalogue between a dynamic range of 1.0 and 1.8, which is absent for the same plot for the m catalogue. This is also reflected in Figure 5.13b) – with both magnitude types having a different range of b-value estimates. While m_L exhibits values of $b \sim 0.9$

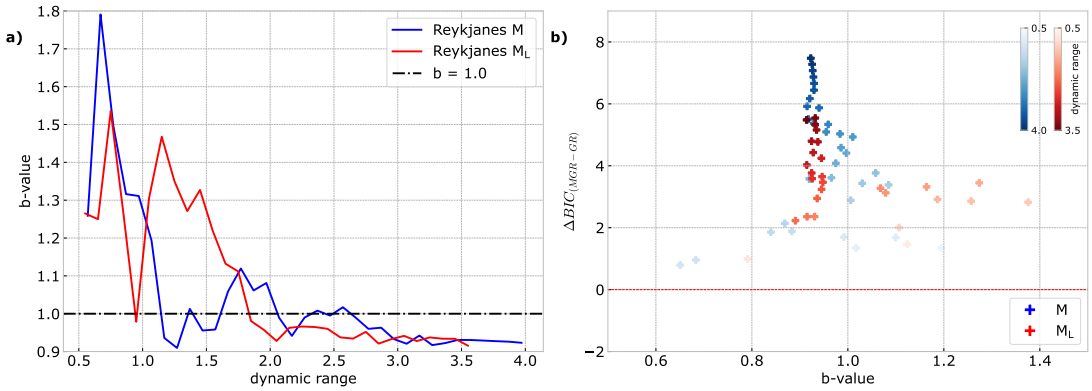

Figure 5.13: a) Change in b-value as a function of dynamic range. b) ΔBIC as a function of the inferred b-value. A positive ΔBIC implies a preference for the GR model. Darker markers indicate lower magnitude thresholds and larger dynamic ranges; lighter markers indicate higher magnitude thresholds and smaller dynamic ranges.

or larger (up to 1.4, with an outlier around 0.8), m returns more consistent b-values between 0.8 and 1.2 (with outliers near 0.6). The GR ($\Delta BIC > 0$) is always preferred, albeit stronger for the larger dynamic range catalogues of the m catalogue (ΔBIC lies between 4 and 8 for the majority of catalogues and between 2 and 6 for the m_L estimates). Based on the MLE (sections 5.1.5 and 3.2.3), estimates of m_θ for m are between 5.4 and 5.8, whereas the estimate of m_θ for m_L is much smaller between 4.4 and 5.1.

The results shown here suggest that b-value estimates for both magnitude types are near unity (from Figure 5.13a) at dynamic ranges larger than 2.0 (b-values are very variable with a lower dynamic range) and the preferred distribution is GR in both cases (Figure 5.13b). Both of these observations are consistent with our understanding of b-values and model selection in tectonic regions (section 2.5.1.2). In contrast, the behaviour of the Reykjanes data is very different to that obtained previously with the INGV catalogue (section 4.5.2), where for example, the BVS estimates and b-value errors vary greatly between the BVS GR and MGR methods (see Figure 4.8, right side). The next subsection looks at another tectonic catalogue – Southern California, where in a historical context the idea and theory of the GR distribution was first developed soon after the local magnitude was invented.

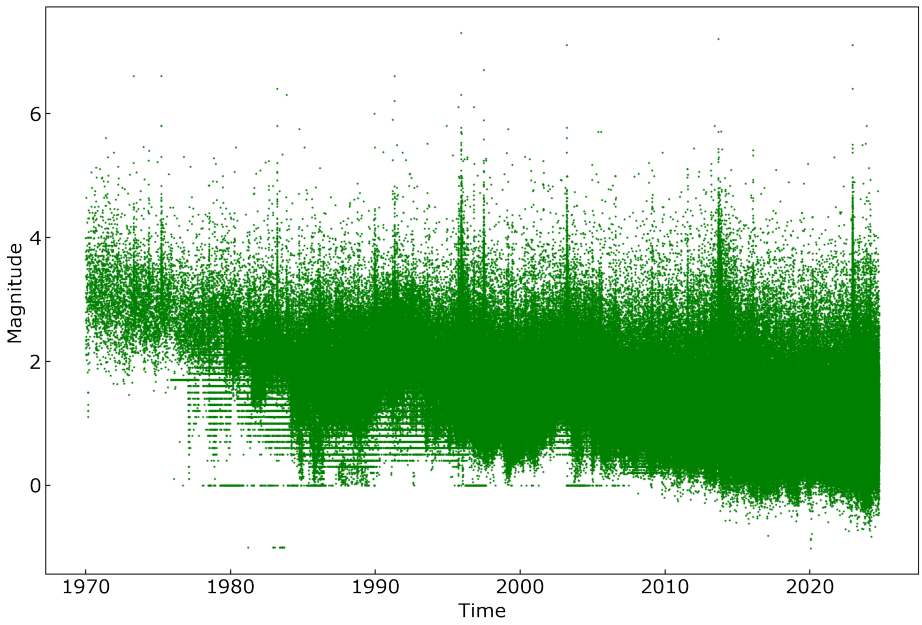

Figure 5.14: Plot of catalogue magnitudes with time to indicate the gradual improvement in network sensitivity and recording.

5.2.1.2 Southern California (SoCal)

This catalogue is the largest catalogue I have used and also contains much larger magnitudes than any of the other catalogues I have used in this thesis. It was obtained from the archive of the Southern California Earthquake Data Center (SCEDC), recorded by the Southern California Seismic Network (SCSN, Hutton et al. (2010)).

Hutton et al. (2010) give the m_c estimate for their catalogue between 1932 and 2008 to be ~ 3.25 in early years, down to ~ 1.8 at present or even lower in the most densely instrumented areas. The caveat here is the decision of where and with what parts of the catalogue to estimate this m_c. Here, I have chosen to use the entire available catalogue, as the sensitivity of event recording (shown in Figure 5.14) improves more gradually than, for example, in the case of the INGV data (Figure 3.2), and is therefore more difficult to cut at specific times.

Visualising the FMD of this catalogue, Figure 5.15a) shows a strong curve indicating a very long incomplete part (compared to the Reykjanes data, for example; see Figure 5.12a) of the catalogue at low magnitude. From Figure 5.15b), the b-value estimate,

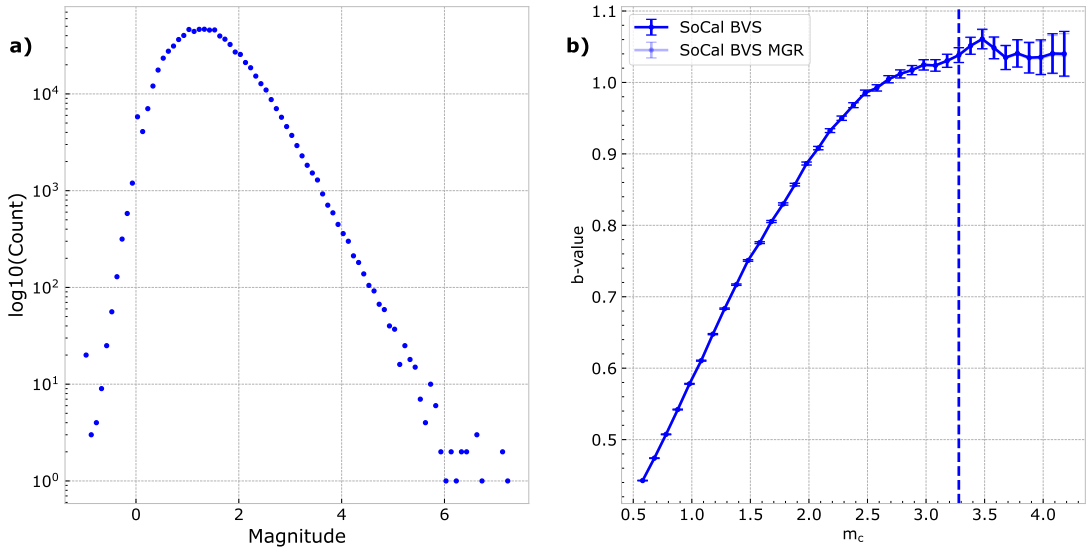

Figure 5.15: a) FMD for the SoCal data, including incompletely reported data at small magnitudes. b) The BVS curve using GR (solid colours) and MGR (lighter colours). The vertical dashed line indicates the returned BVS m_c estimate, based on successive b-values falling within error (one standard deviation) of the b-value.

here assuming an underlying GR distribution, is particularly low ($0.3 < b < 0.9$) up until an m_c of approximately 2.0, after which it remains stable within the one standard deviation error bars shown. Similarly to what we saw in the Reykjanes data (section 5.2.1.1), using the BVS m_c estimate for the SoCal catalogue (Figure 5.15b), the data is substantially diminished, returning 9023 events in the complete catalogue – just over 1% of those in the original (incomplete) catalogue. In this case, there is no visible difference between the BVS GR and BVS MGR estimates.

In agreement with the expectation from the analysis of synthetic data, Figure 5.16a) exhibits convergence of the estimated b-value to a value around 1.05 for a dynamic range larger than 2.5. In agreement with the Reykjanes data in Figure 5.13a), inferred b-values for a dynamic range of less than 2.0 returns very variable and questionable b-values. The estimate of the roll-off magnitude is $7.0 < m_\theta < 7.5$. Similarly, Figure 5.16b) suggests a strong preference towards the GR model. However, with a lower dynamic range and sample size (lighter markers), estimated b-values are < 1.0, even as low as 0.5 and the choice of model increasingly prefers MGR.

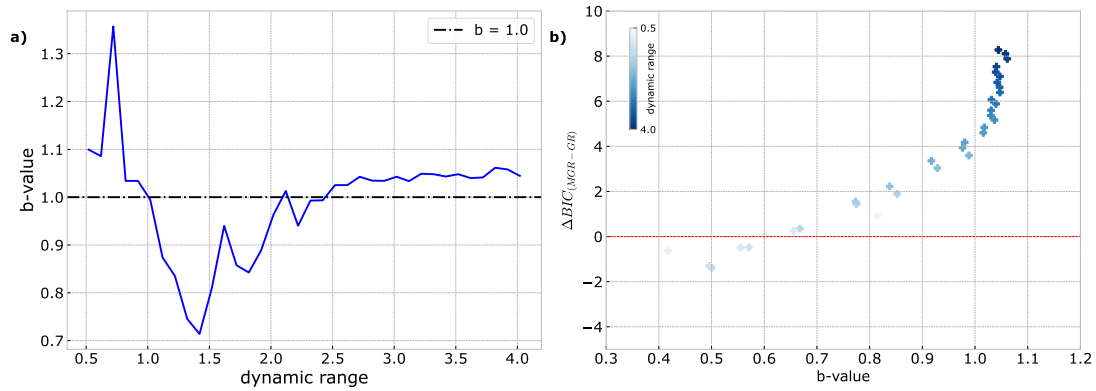

Figure 5.16: a) Change in b-value as a function of dynamic range. b) ΔBIC as a function of the inferred b-value. A positive ΔBIC implies a preference for the GR model. Darker markers indicate lower magnitude thresholds and larger dynamic ranges; lighter markers indicate higher magnitude thresholds and smaller dynamic ranges.

Using this catalogue data, I have shown a stable BVS curve (Figure 5.15b) as well as b-value convergence to a value of around 1.05 for dynamic ranges greater than 2.5 (Figure 5.16a), as expected for tectonic seismicity. It is more surprising that lack of dynamic range and sample size in Figure 5.16b) returns small b-values and a slight preference for the MGR distribution. This is in contrast to the results of section 4.5.1, where I have shown that such lack of data generally leads to a b-value bias towards larger values.

While Southern California is a good, comprehensive 'reference' catalogue to use, this may not be representative of many of the smaller (magnitude) catalogues that are often associated with volcanic and induced seismicity, which are now investigated in the following subsection.

5.2.2 Volcanic Seismicity Catalogues

In addition to the two volcanic catalogues discussed in Chapter 4, I have also chosen to investigate the Long Valley and Kilauea data, both of which are volcanoes located in the US. I considered using Mount St. Helens earthquake data, but it is known that Mount St. Helens follows neither a GR nor MGR distribution, so was not included in

the assessment of this thesis. Instead, it follows a 'characteristic' distribution (Main, 1987) which I briefly introduced in section 2.4.3. However, looking at an extended catalogue with increased seismic monitoring and recording of Mount St. Helens (from 1972 until 2021 obtained from the Pacific Northwest Seismic Network (PNSN), not just the 1980 eruption), there is a trend to a less obvious characteristic distribution as, for example, just the 1980 event and could be an interesting lead to follow in future work.

5.2.2.1 Long Valley

The Long Valley Caldera, introduced in section 3.5.3.4, is an additional volcanic catalogue I have analysed in this thesis. It is one of the larger catalogues assessed here, with 122 748 events in the incomplete catalogue. The ongoing thermal activity at the caldera results in much seismic activity being monitored by the USGS through a dense seismic network. The largest event recorded at Long Valley was a magnitude 6.1 earthquake in May 1980. However, due to poor sensitivity and ability to record particularly small events prior to 1983, the more homogeneous catalogue used here starts in January 1983 and therefore omits this largest recorded event.

The incomplete catalogue's FMD is shown in Figure 5.17a), with the BVS curve shown in b). The BVS curve stabilises around $b = 1.1$ when m_c is estimated to be 1.56. This results in 16 204 events in the complete part of the catalogue (Table 5.2). With increased trial m_c, the inferred b-values increase and spike at $b \approx 1.35$ when $m_c = 2.8$. From this, the b-value is in general slightly elevated above 1.0 for this data. The errors to one standard deviation increase with large m_c, which is expected due to lack of data and discarding substantial portions of the complete part of the catalogue. The difference between GR and MGR BVS estimates is again barely visible and the best inferred value of m_θ lies between 5.1 and 5.7, with a maximum observed magnitude of 5.4.

Figure 5.18 confirms the inference of elevated b-values, where in a), above between a dynamic range of 1.5 and 3.0, the b-value lies between 1.2 and 1.3 before it converges to around 1.1 for a dynamic range > 3.0. The preferred model is the GR distribution for

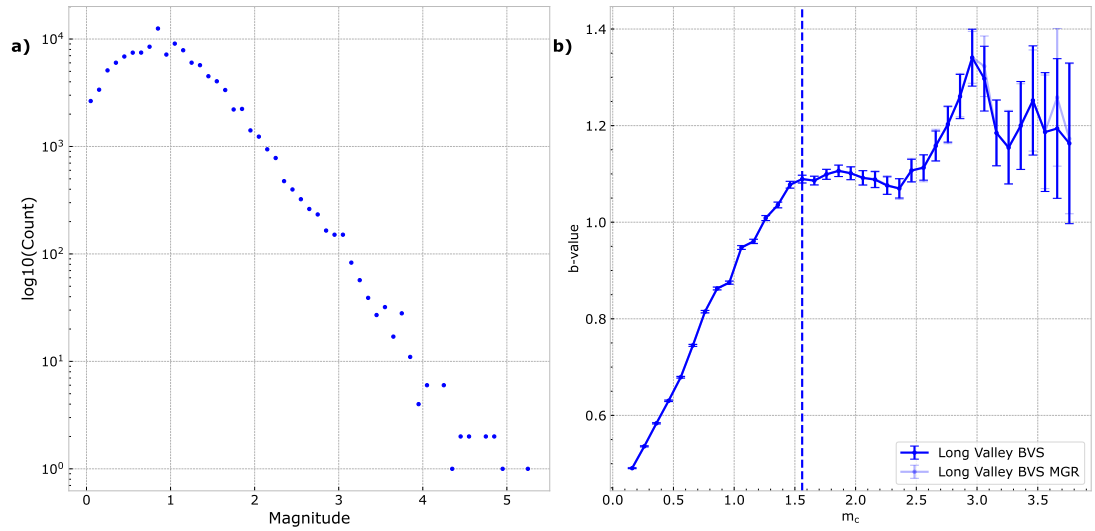

Figure 5.17: a) FMD for the Long Valley data, including incompletely reported data at small magnitudes. b) The BVS curve using GR (solid colours) and MGR (lighter colours). The vertical dashed line indicates the returned BVS m_c estimate, based on successive b-values falling within error (one standard deviation) of the b-value.

most cases (Figure 5.18b). Only when sub-catalogues are small, with high m_c, does the preference for GR become less and inferred b-values decrease to < 1.0, similar to the results seen in the Southern California catalogue (Figure 5.16). This again suggests the influence of sampling artefacts due to clear lack of data.

The b-values at Long Valley are slightly elevated and consistent with findings in previous studies (Wiemer et al., 1998; Barton et al., 1999), suggesting that this may be linked to a correlation to locations of magma chambers. The b-values estimated here are somewhat elevated at enough dynamic range (> 3.0) and sample size, however, when this is minimised, b-values become very low and the preference for the GR model also becomes less strong as seen in Figure 5.18b). In the next section, I compare these results to another volcanic catalogue.

5.2.2.2 Kilauea

Kilauea is the most active of the Hawaiian volcanoes. Continuous, detailed monitoring of the volcano has been a big factor in assessing the hazard and risk of the island and

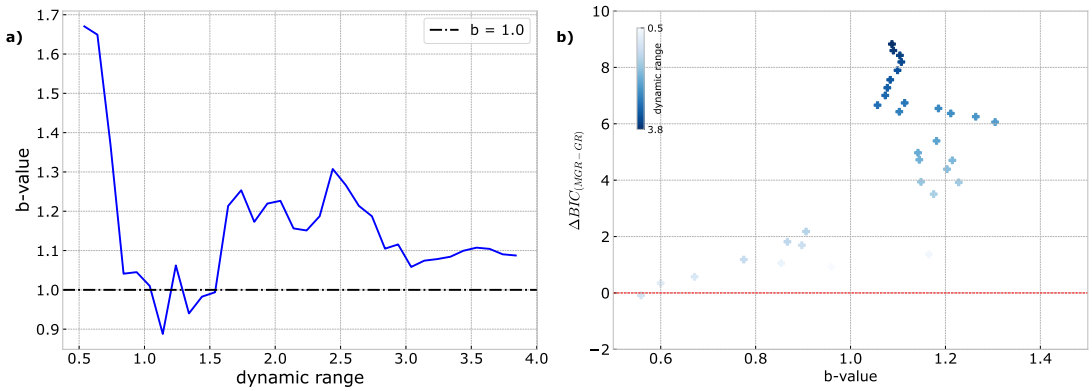

Figure 5.18: a) Change in b-value as a function of dynamic range. b) ΔBIC as a function of the inferred b-value. A positive ΔBIC implies a preference for the GR model. Darker markers indicate lower magnitude thresholds and larger dynamic ranges; lighter markers indicate higher magnitude thresholds and smaller dynamic ranges.

allows access to abundant data to investigate the volcano's statistics and properties of the volcanic seismicity, such as the b-value. Wyss et al. (2001) previously showed that the seismicity originating in the South Flank of the volcano has anomalously high b-values, assumed consequences of active magma reservoirs at depths down to 8km. However, in other areas of the volcano, they found b-values to be exceptionally low (b = 0.5), possibly as a result of faulting in a more homogeneous medium. Llenos et al. (2021) suggest that the Kilauea caldera collapse events are a source of seismic hazard and should therefore always be included in a complete seismic hazard assessment. They further suggest that the event occurrence model is non-Poissonian but can be approximated to Poisson-like at low ground motion probabilities - an approach which may be generalisable for other seismogenic processes. This research has also shown that the recent 2018 eruption and collapse is not GR distributed. To build on this, Figure 5.19a) shows the FMD. It is quite symmetric, especially compared to any of the other catalogues I have looked at. In part b), the BVS curve returns large m_c estimates of 3.2 for both the GR and MGR estimates but the b-value estimates and errors are again slightly larger for the MGR BVS curve (in lighter colours). At BVS m_c, the b-value is near 2.0, which is very high.

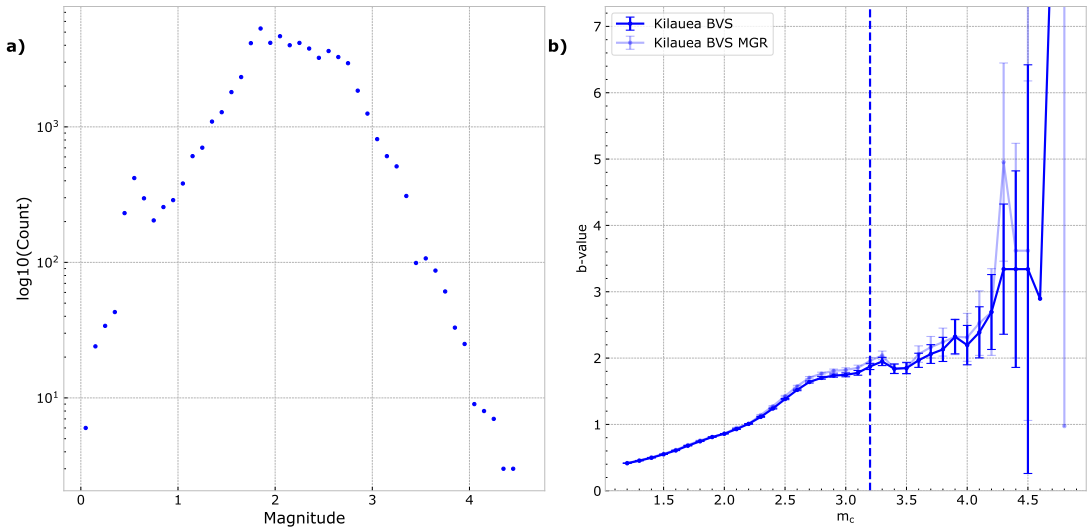

Figure 5.19: a) FMD for the Kilauea data, including incompletely reported data at small magnitudes. b) The BVS curve using GR (solid colours) and MGR (lighter colours). The vertical dashed line indicates the returned BVS m_c estimate, based on successive b-values falling within error (one standard deviation) of the b-value.

In Figure 5.20a), I show the estimated b-value with dynamic range. The two most obvious points are i) the b-values decreases with increased dynamic range and ii) the resulting dynamic range when using a BVS m_c estimate is extremely small (only 1.4) at its maximum. For this reason, I have chosen to also plot the cases of using both the GFT (red) and MaxC (green) methods to estimate m_c, which return m_c values of 2.7 and 1.9 respectively (section 3.2.1). These are substantially smaller than the BVS m_c estimate and therefore include more data in the 'complete' part of the catalogue. This results in 1263 events above BVS m_c, 8736 events above GFT m_c and 39 656 events above MaxC m_c, resulting in larger dynamic ranges as shown in Figure 5.20a) and continuing the trend of decreasing b-values with increased dynamic range, albeit minimally. For BVS, the largest dynamic range is 1.5, at which a b-value of 2.0 is returned. For GFT the largest dynamic range is 2.0, at which a b-value of 1.75 is returned and for MaxC the largest dynamic range is 2.8, at which a b-value of < 1.0 is obtained, clearly due to inclusion of incompletely reported data. For Figure 5.20b), I show the ΔBIC for all three m_c methods. This demonstrates that the majority of data points with large

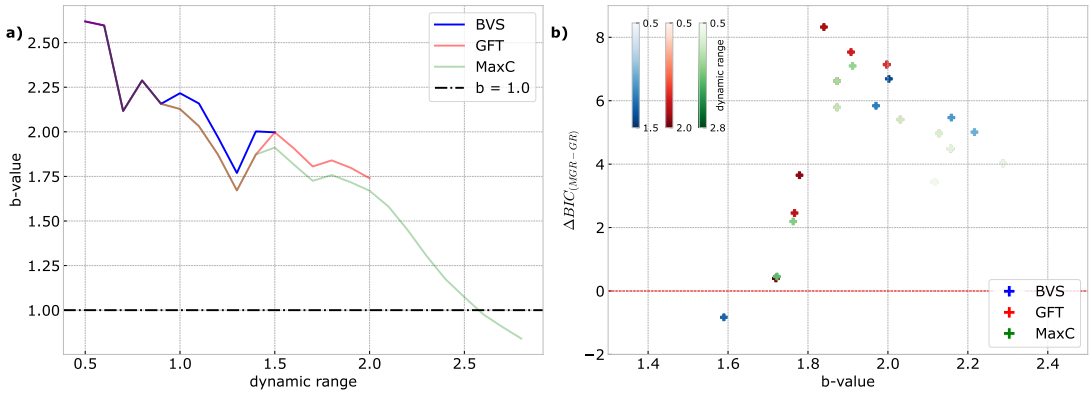

Figure 5.20: a) Change in b-value as a function of dynamic range shown for BVS (blue), GFT (red) and MaxC (green) m_c estimates. b) ΔBIC as a function of the inferred b-value. A positive ΔBIC implies a preference for the GR model. Darker markers indicate lower magnitude thresholds and larger dynamic ranges; lighter markers indicate higher magnitude thresholds and smaller dynamic ranges.

dynamic range (and associated large numbers of events) for MaxC have $b \sim 1.2$ and large, negative values of ΔBIC (up to -3000). These large amplitude estimates are not shown because they would skew the view of the rest of the plot. These anomalously low values of ΔBIC are consistent with the MaxC m_c estimate being too low, resulting in incomplete data at low magnitude, a strong preference for the incorrect model, and very low b-values (compared to the other methods). The estimated value of m_θ is 4.3 in this case. This is similar to the maximum observed magnitude in this catalogue (Figure 5.19a) and implies the roll-off is not sampled and hence the GR distribution is the preferred model, in agreement with the ΔBIC results shown in Figure 5.20b).

Summarising these results, I have shown that due to the availability and range of magnitudes in this catalogue, this is a complicated catalogue to analyse – there is very limited dynamic range above m_c in both BVS and GFT cases, whereas the MaxC estimate is too radical and returns dynamic ranges and b-values which need to be assessed equally critically. Therefore, it is important to note that some flexibility in decision-making on a case-by-case basis is needed on which method of m_c estimation to use. Nonetheless, in all cases, the b-value is strongly elevated (between 1.7 and 2.1) in agreement with the results of Wyss et al. (2001).

5.2.3 Induced Seismicity Catalogues

In this last subsection, I look at a further two induced seismicity catalogues – from i) Preston New Road (due to fracking) and ii) Oklahoma (due to wastewater injection).

5.2.3.1 Preston New Road (PNR)

The PNR data are associated with the fracking operations undertaken near Preston, England in 2018 and 2019, as introduced in section 3.5.4.3. During both the first and second operations (resulting in two separate published catalogues labelled PNR-1z and PNR-2), microseismicity exceeded the UK Traffic Light Scheme (section 2.6.3.1) red threshold of $m_L = 0.5$ on multiple occasions, leading in each case to a mandatory, immediate pause of injection for a minimum of 18 hours. The largest event to occur in the first operation had a magnitude of $m_L = 1.9$ and the maximum magnitude event in the second operation had a magnitude $m_L = 2.9$. Smaller events not recorded by the surface catalogue seem to be irretrievably lost during data transfer and storage. Baptie et al. (2020) show that the downhole catalogues for both the PNR-1z and PNR-2 catalogues show a roll-off in the FMD at larger magnitudes, possibly due to limited dynamic range and frequency of the instrumentation, or the parameterisation and method used to determine the moment magnitudes.The magnitudes used in the surface catalogues are more reliable, but the number of events are much more limited than those in the downhole catalogues with a lower detection threshold. Due to these biases from instrumental issues in the downhole catalogue, a catalogue which corrects these magnitudes to m_W is used (provided by Dr. James Verdon) as opposed to the raw catalogue.

The seismicity at PNR is suggested to have occurred on a previously unmapped strike-slip fault (Verdon et al., 2019a). Using the raw data, Mancini et al. (2019) found b-values of ~ 1.5 for the first set of operations at PNR-1z including potentially large uncertainties in magnitude scale conversion between the two (surface and downhole) networks. They further suggest that the downhole catalogue is complete to a magnitude

of -1.0 using BVS but the dynamic range above this is relatively small. Nonetheless, they found the PNR-1z catalogue to obey the GR law. The authors recommend further work to test the robustness of the conclusions made. Fracturing in the PNR-2 well began in August 2019. Following the induced seismicity from these operations, the UK government announced a moratorium on high-volume HF in November 2019.

The FMDs for both operations in Figure 5.21a) show that the event magnitudes in the first (PNR-1z) operation are overall larger than in the second (PNR-2), although the largest magnitude events all occurred in PNR-2. The PNR-2 FMD for the raw data shows a very sharp-peaked distribution. In this case, Roberts et al. (2015) concluded that MaxC is a suitable method to estimate m_c. However, the BVS method should also return robust values of m_c for sharp-peaked distributions and therefore, both should return good values of m_c when applied to (specifically) the PNR-2 catalogue. Therefore, I will use and compare the outcomes for both BVS and MaxC methods in this specific case.

In the BVS curve (Figure 5.21b), the first operation returns an m_c estimate of -0.16 for both the GR (solid colours) and MGR (lighter colours) BVS estimates, leaving 1335 events in the complete catalogue. The MGR BVS curve is only slightly higher than the GR BVS curve. In the PNR-2 operation, both GR and MGR BVS m_c estimates return -0.88 (4858 events in the complete catalogue) and both curves are nearly identical. (It is worth noting that when using the raw, uncorrected magnitudes that the BVS m_c values are not the same when using GR and MGR.) Using the MaxC method returns a lower m_c = -1.26 for PNR-1z (25 959 events above m_c) and -1.78 (45 151 events above m_c) for PNR-2. Specifically for PNR-2, this value is not as close to the GR BVS estimate as expected, from the discussion in the previous paragraph.

Figure 5.22a) shows that b-values are elevated (PNR-1z using MaxC/BVS) or fairly low (PNR-2 using Maxc/BVS) at small dynamic ranges (0.5 to 1.0) but even at large dynamic ranges (see colorbars on Figure 5.22b), b-values are elevated. Using BVS m_c estimates, at resulting maximum dynamic range (2.0 for PNR-1z and 3.7 for PNR-2), inferred b-values are $\sim$ 1.6 and $\sim$ 1.3, respectively. Using MaxC to increase the number

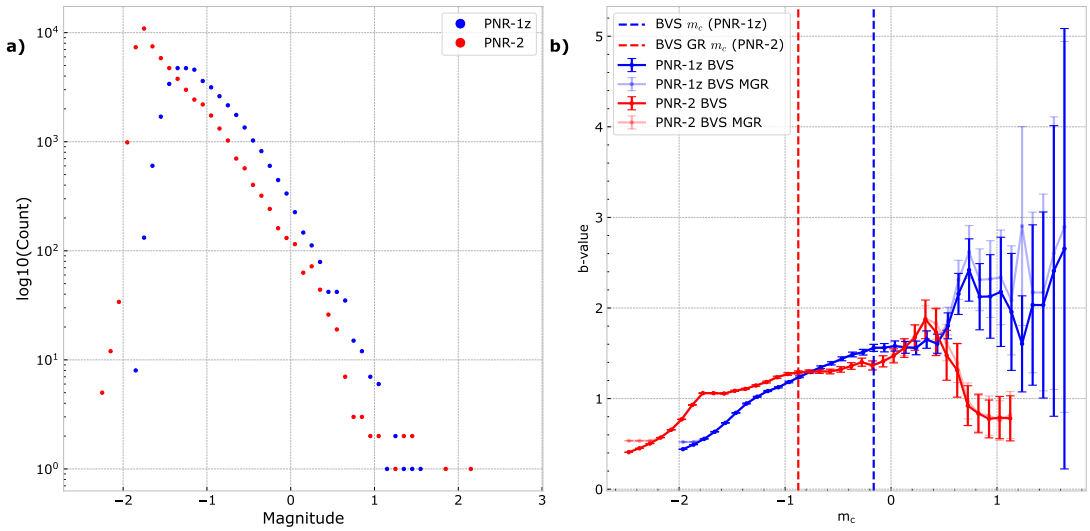

Figure 5.21: a) FMDs for both operations (PNR-1z in blue, PNR-2 in red), including incompletely reported data at small magnitudes. b) The BVS curve using GR (solid colours) and MGR (lighter colours) for both operations. Vertical dashed lines indicate the returned BVS m_c estimates, based on successive b-values falling within error (one standard deviation) of the b-value.

of events in the 'complete' catalogue and increase dynamic range further, b-values decrease continuously for PNR-1z (green triangles) where at maximum dynamic range (3.1), the b-value is $\sim$ 1.2 and similarly for PNR-2 (purple triangles), at largest dynamic range (4.7), the smallest b-value or around 1.05 is obtained.However, the same problem as in the Kilauea data (Figure 5.20a) is observed – the values of m_c obtained from the MaxC method are very small, and could lead to the inclusion of incomplete parts of the catalogue, leading to artifically low b-values. Therefore, the above b-values of 1.2 and 1.05 for Maxc in PNR-1z and PNR-2, respectively, may be true and therefore not truly elevated (i.e. not considerably above 1.0) when all data is available and used, or may be biased and artificially low due to incomplete catalogue data. If these trends had been extended to larger magnitudes outside the data range, the hazard resulting from these potential extreme events might have been underestimated or overestimated respectively. Figure 5.22b) confirms the fairly elevated b-values observed in the PNR fracking operations (especially when determining m_c using BVS), as well as the preferred model being GR (ΔBIC $>$ 0), in nearly all cases, even in the smallest sub-catalogues.

Outliers supporting a preference for MGR only occur due to the inclusion of catalogue data that is likely to be incomplete when using MaxC to estimate m_c. The values of m_θ lie between 0.8 and 2.2 for PNR-1z and between 2.7 and 4.0 for PNR-2 (when using the MaxC m_c).

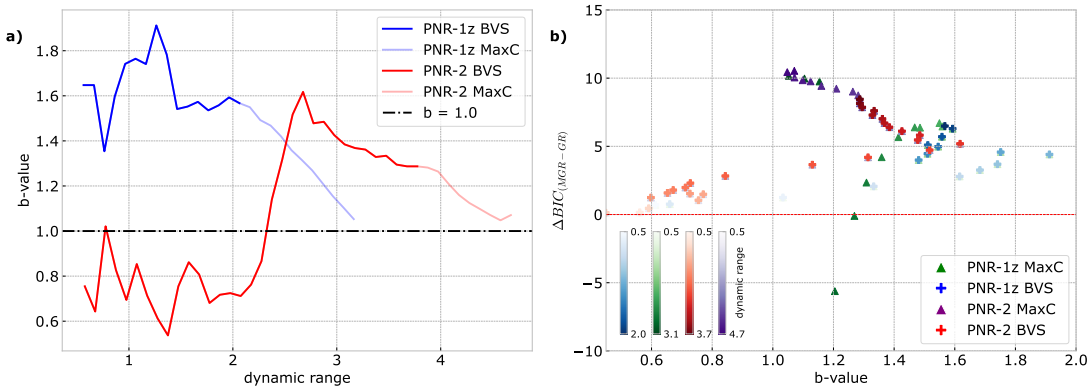

Figure 5.22: a) Change in b-value as a function of dynamic range shown for BVS (blue), GFT (red) and MaxC (green) m_c estimates. b) ΔBIC as a function of the inferred b-value. A positive ΔBIC implies a preference for the GR model. Darker markers indicate lower magnitude thresholds and larger dynamic ranges; lighter markers indicate higher magnitude thresholds and smaller dynamic ranges.

In summary, in this case study, it is difficult to decide on a reliable estimate on which the b-value would ultimately converge. Increasing the dynamic range by decreasing m_c (using MaxC) returns a larger dynamic range but is likely to bias the results, particularly in model preference and returns a b-value closer to 1.0, likely due to inclusion of incomplete data, as the b-values at PNR are known to be elevated above 1.2 (Verdon et al., 2019a; Kettlety et al., 2021). As a consequence, the b-value may well be artificially high (BVS) or low (MaxC), due to the sampling and model selection issues highlighted in the analysis of synthetic data in section 5.1. This can lead to a systematic underestimate or overestimate of the probability of occurrence of events larger than those in the sample while attempting to manage regulatory compliance and the risk associated with induced seismicity. Kendall et al. (2019) previously stated that it is not uncommon for $b > 2$ to be observed in hydraulic fracture settings, most likely because fluids reduce the shear strength required to rupture the rock. This is consistent with the analysis

shown here from the PNR operations, where b is significantly elevated at medium to low dynamic range. The analysis also highlights the issues of reliable characterisation of induced seismicity and instrumentation biases during ongoing operations which use only surface networks and hence suffer from a narrow dynamic range. Nonetheless, the largest dynamic ranges obtained here for BVS estimates (at minimum 2 orders of magnitude in PNR-1z) are an improvement to other induced seismicity settings as well as the raw PNR catalogue where this was only just above one order of dynamic range. This suggests that the b-values obtained at maximum dynamic range for BVS (2 for PNR-1z and 3.7 for PNR-2) are likely to be representative of the true b-values at PNR, i.e. somewhat elevated between 1.3 and 1.5.

5.2.3.2 Oklahoma

This induced seismicity catalogue has been chosen due to the importance of understanding events that have been ongoing in Oklahoma since 2009 due to wastewater injection (Keranen et al., 2014). The seismicity induced through this process has much larger magnitudes that the microseismicity induced from fracking shown in the previous section (5.2.3.1), up to magnitude 5.8 (see section 3.5.4.2 for more details). The events in Oklahoma are often large enough to be felt and some have caused considerable damage to buildings and towns (Langenbruch and Zoback, 2016; Foulger et al., 2018). In cases such as these, it is particularly important to have a good understanding of the processes to be able to assess the hazard and to manage the associated risk to the communities.

The FMD for the raw Oklahoma data is almost an ideal Gaussian one (Figure 5.23a). The effect of this anomalous raw FMD is also seen in Figure 5.23b), where the BVS curve reaches a maximum b-value of 1.7 (2.3 for MGR) for a trial m_c of 3.7, before the inferred b-value decreases. The estimated GR BVS m_c is 3.8, while the MGR BVS m_c = 4.0 – again, the values are different as was the case for PNR-2 (Figure 5.21b). The MGR BVS curve also has substantially higher b-value estimates than the GR BVS curve does. This is likely due to the higher value of m_c, otherwise it would be logical for the trade off between having a roll-off and not to result in lower b-values for the MGR fits.

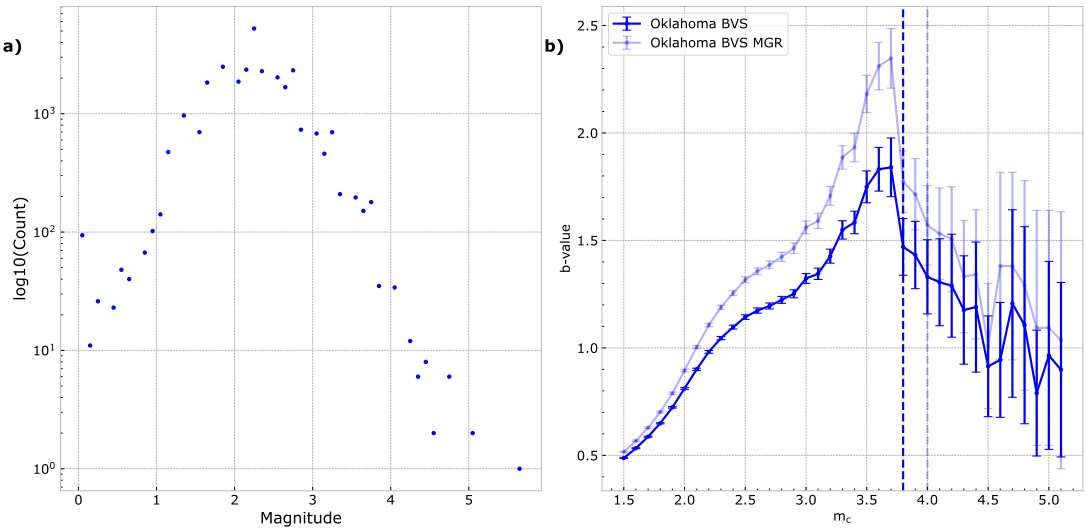

Figure 5.23: a) FMD for the Oklahoma data, including incompletely reported data at small magnitudes. b) The BVS curve using GR (solid colours) and MGR (lighter colours). The vertical dashed line indicates the returned BVS m_c estimate, based on successive b-values falling within error (one standard deviation) of the b-value.

The sensitivity of b-values in this catalogue to dynamic range is shown in Figure 5.24a). The results suggest that b-values at dynamic ranges less than 1.5 are lower than 1.0 and only with dynamic ranges larger than this, the b-value is elevated (up to nearly 1.5 when the dynamic range is ~ 2.2). With a dynamic range larger than 2.2, the b-value decreases again until it is near 1.0 for a dynamic range of 3.4. The GFT m_c was also applied here for comparison, resulting in dynamic ranges larger than 2.0, which is the maximum dynamic range seen when using the BVS method. The MaxC method returned a similar m_c estimate to the GFT. Figure 5.24b) indicates that the GR is the preferred model for all sub-catalogues, albeit being a stronger preference for the darker markers which indicate low m_c and a larger dynamic range. For the largest catalogue (obtained through a GFT m_c estimate), the inferred b-value is very close to 1.0, before it increases in the following smaller sub-catalogues. The MLE of m_θ ranges between 5.7 and 6.5, which is larger than the maximum magnitude observed in this catalogue (5.8, Table 5.2). As a consequence, any roll-off is not sampled by the data and therefore this catalogue is unlikely to prefer the MGR distribution.

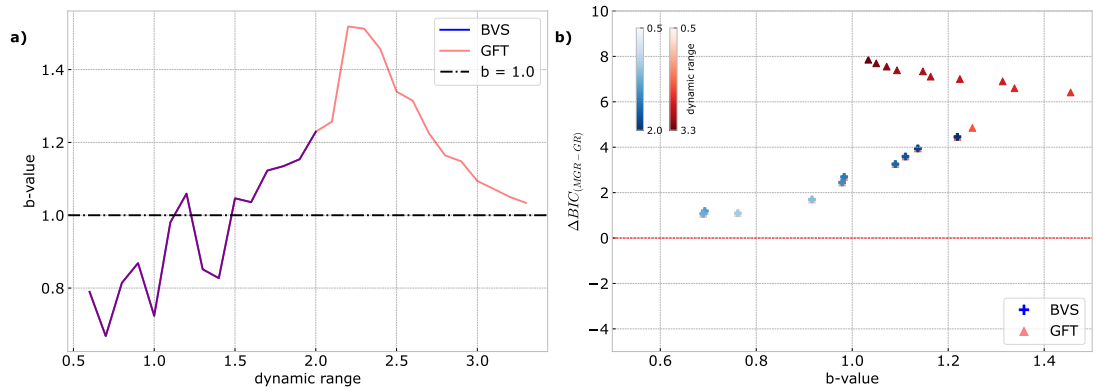

Figure 5.24: a) Change in b-value as a function of dynamic range shown for BVS (blue) and GFT (red) m_c estimates. b) ΔBIC as a function of the inferred b-value. A positive ΔBIC implies a preference for the GR model. Darker markers indicate lower magnitude thresholds and larger dynamic ranges; lighter markers indicate higher magnitude thresholds and smaller dynamic ranges.

To summarise, I have shown again that the dynamic range resulting from a BVS m_c estimate is too narrow to estimate a reliable b-value – inferred b-values range from 0.7 to 1.2 between dynamic ranges of 0.5 and 2.0 (from Figure 5.24a) and are unusually low up to a dynamic range of 1.5 (considering it is likely that the BVS m_c estimate overestimates the m_c, in all previous cases resulting in particularly high b-values). Additionally, in this case it was difficult to obtain a dynamic range larger than 2.0 with the same method (BVS) as in all previous catalogues – prompting the use of a less conservative method (GFT used here) than BVS to estimate m_c. Nonetheless, from Figure 5.24b), this catalogue has somewhat elevated estimated b-values. This is likely to be a true estimate of the underlying b-value because the b-value remains slightly high even with larger dynamic range (but still much lower than the estimate of the b-value for the PNR catalogue, section 5.2.3.1). In any case, the preferred model is the GR distribution, as for the induced PNR data (section 5.2.3.1). Due to the nature of the (micro)seismicity, it is reasonable that similar issues to the Kilauea data (section 5.2.2.2, such as lack of data) are encountered, shown by the need to choose multiple methods to estimate m_c, as the BVS is overestimates m_c and returns very small dynamic ranges and sample size.

Table 5.2: This table indicates the important details for the real catalogues, including type of seismicity, the period during which the catalogue events were observed, the total number of events in this observed period (N), the number of events above m_c, the observed maximum magnitude within the catalogue and the value of the b-value stability m_c. WD - Wastewater Disposal, HF - Hydraulic Fracturing

Catalogue	Seismicity Type	Observation Period	N	N > m_c	obs. m_{max}	m_c (bvs)	dyn. r	est. m_θ
Reykjanes M1	Tectonic	07/01/1995 - 31/01/2021	49 441	5683	5.61	1.64	3.97	~ 5.4 - 5.8
Reykjanes M2	Tectonic	07/01/1995 - 31/01/2021	50 379	4255	4.96	1.41	3.55	~ 4.4 - 5.1
Southern California	Tectonic	01/01/1960 - 07/05/2021	767 380	9023	7.3	3.28	4.02	~ 7.0 - 7.5
Kilauea	Volcanic	01/01/1960 - 31/12/2020	59 128	1263	4.7	3.2	1.5	~ 4.3
Long Valley	Volcanic	01/01/1983 - 31/12/2020	122 748	16 204	5.4	1.56	3.84	~ 5.1 - 5.7
Oklahoma	WD	01/01/1980 - 31/10/2017	28 219	107	5.8	3.8	2.0	~ 5.7 - 6.5
Preston New Road PNR-1z	HF	15/10/2018 - 17/12/2018	38 452	1335	1.9	-0.16	2.06	~ 0.8 - 2.2
Preston New Road PNR-2	HF	15/08/2019 - 23/08/2019	56 008	4858	2.9	-0.88	3.78	~ 2.7 - 4.0

5.3 Alternative thinning of earthquake catalogues

Throughout this thesis, I have thinned synthetic and real catalogues by increasing the value of m_c to consistently decrease both dynamic range and N. An alternative method of creating sub-catalogues is to randomly sample a subset of events from a large overall population. If earthquakes are random processes, as we assume them to be, this is what happens when b-values are assessed in time and space. I have chosen one synthetic catalogue and one real catalogue to create sub-catalogues in this way and discuss the results compared to the method used throughout the thesis.

For the synthetic catalogue, I return to the 'standard' (MGR) parameters, where N = 10 000, m_{min} = 1.0, b = 1.0 and m_θ = 3.5. From this, I randomly sample 3 sub-catalogues with N = 5000, 1000 and 700 as shown in Figures 5.25a), b) and c) respectively, where the ΔBIC is shown as a function of the b -value, with the dynamic range indicated by the shading of triangles.

In the sub-catalogues shown in Figure 5.25, at largest dynamic range (blue triangles), it is the GR distribution which is preferred. This is surprising, especially as the preference also improves (i.e. some simulations prefer the correct MGR distribution) with lower N (Figure 5.25c). This suggests that sampling specific N from a larger overall population also leads to biases in model preference in the opposing direction. The relation between dynamic range and N here is proportional to some extent, as N decreases, the maximum observed dynamic range also decreases. However, this is not the case in the Southern California data in Figure 5.26 and again a result of the randomness with which the events are picked. In this case, the largest N (5000, blue crosses) has a very small observed dynamic range, comparable to that of the sub-catalogue with only N = 700. This result backs the discussion in section 5.1.7, here showing that small dynamic range (1.5) can be obtained even when N is large. Furthermore, the largest sub-catalogue (blue crosses) in N prefers an MGR distribution at all times. This is due to its narrow dynamic range, as I have shown in section 5.2.1.2 that Southern California does not exhibit any MGR behaviour at any point, when

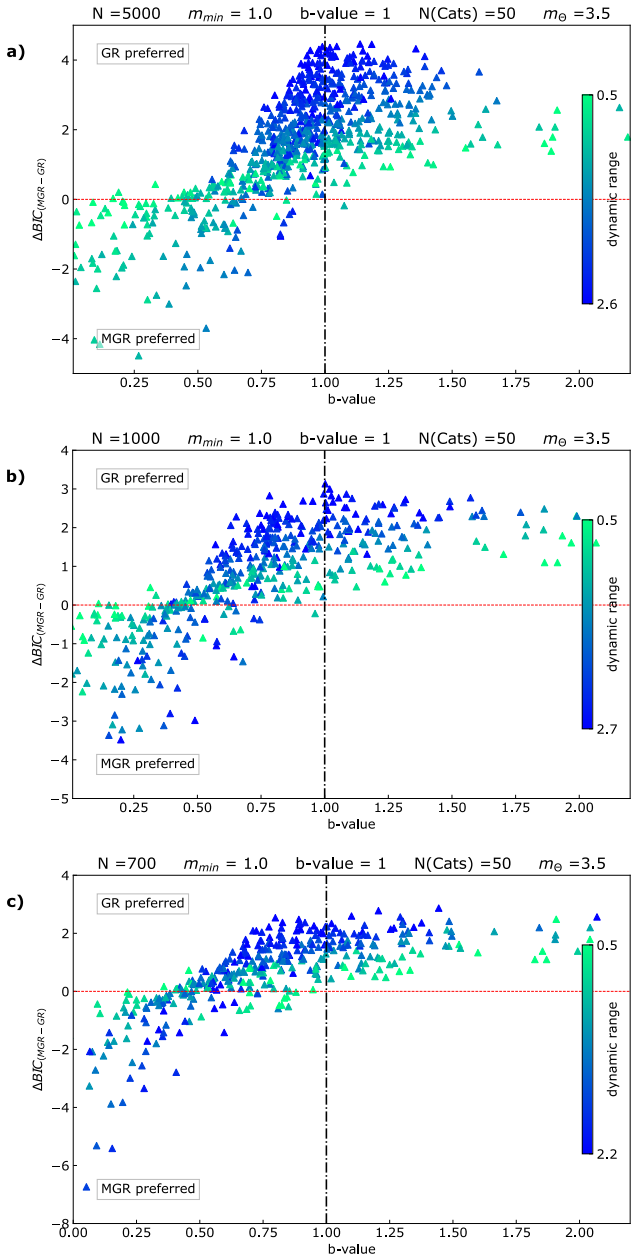

Figure 5.25: ΔBIC as a function of the b-value for a) $N = 5000$, b) $N = 1000$ and c) $N = 700$ sampled from the same overall population. Dynamic range is indicated by shading of triangles.

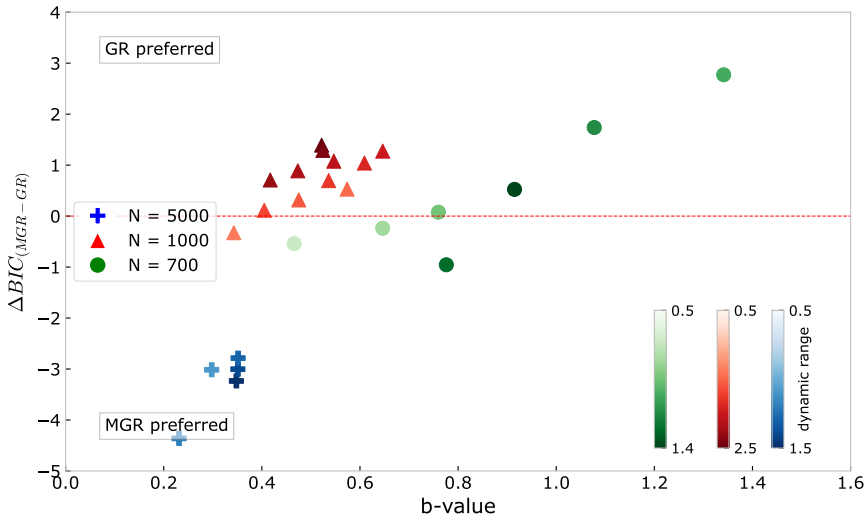

Figure 5.26: ΔBIC as a function of the b-value for Southern California data when $N = 5000$ (blue crosses), $N = 1000$ (red triangles) and $N = 700$ (green circles). Dynamic ranges are indicated by degree of shading.

thinned by increasing m_c.

The observed b-values both in Figures 5.25 and 5.26 show strong b-value biases to low values (also at larger dynamic ranges) and is different to what I have shown so far, when thinning by increasing m_c. Summarising the work in this section, I have shown that whether catalogues are thinned by consistently increasing m_c or by randomly sampling a smaller value of N from a larger, overall population, both the model preference and the resulting b-values are likely to be biased which will be consequential to any hazard analysis.

5.4 Summary

The analysis of synthetic data presented in this chapter was aimed at enhancing the understanding of changes in model preference and inferred model parameters such as the b-value, as a function of the parameters used to generate the synthetic data, notably its dynamic range. The main goal was to establish this sensitivity before applying this reasoning to the analysis of real earthquake catalogues from a variety of settings. I have shown in particular, that Aki b-values are likely to be biased to high b-values in cases

where the underlying true b-value < 1.5 and the distribution is MGR (section 5.1.5) and that the accuracy and precision of b-values increases with increased dynamic range (section 5.1.6).

The FMDs of all real catalogues show that with the raw, incomplete, catalogue, b-values are unrealistically small, reminding us why an accurate (and not too low) estimate of m_c is crucial. Therefore, the more conservative BVS method was the main choice of method to estimate m_c. I developed and applied a new BVS method, based on the original one, but which assumes an underlying MGR distribution, on all real catalogues assessed in this chapter. In most cases, the values of m_c obtained by both methods are indistinguishable, except for the two induced seismicity catalogues (sections 5.2.3.1, 5.2.3.2) where visible discrepancies in the resulting m_c estimates are seen. A summary for the estimates for the BVS m_c and m_θ are shown in Table 5.2.

I have shown in sections 5.2.2.2, 5.2.3.1 and 5.2.3.2 that the use of the BVS method to estimate m_c resulted in relatively large m_c estimates, returning very small dynamic ranges and making it hard to assess whether the inferred b-values are accurate or precise. Therefore, it is sometimes necessary to choose a different method of m_c estimation, tailored to individual situations, and to examine the resulting trade-off between accuracy and precision in using a less conservative method. This led to using different m_c methods for both induced and one volcanic catalogue and comparing the outcomes. In a real application it may be sensible to apply both methods and incorporate the two outcomes as two branches of a logic tree for the assessment of epistemic uncertainty in any hazard analysis. For the additional methods of m_c used, the MaxC (in sections 5.2.2.2, 5.2.3.1) returns particularly low b-values at large dynamic ranges, but this is likely to be due to the inclusion of incomplete event data (as shown in Figures 5.20a) and 5.22b). While this is a general rule, low b-values are also obtained at high m_c in the BVS analysis (Figure 5.24a). More often, high b-values are obtained when m_c is high and the dynamic range is small – and this is exacerbated if m_c is overestimated. These issues, ultimately due to lack of data and dynamic range, were not observed in any of the tectonic catalogues assessed here or in Chapter 4 and reinforces

the requirement for large (in sample size and dynamic range) volcanic and induced seismicity catalogues, to definitively assess the FMDs' b-value and the resulting hazard.

Both tectonic catalogues shown here (sections 5.2.1.1, 5.2.1.2) return stable b-values near unity for dynamic ranges larger than 2.0, but also elevated b-values for smaller dynamic ranges and a model preference for the GR model with enough data – in line with the general expectation for tectonic seismicity (section 2.5.1.2). In the Long Valley catalogue (section 5.2.2.1), inferred b-values are slightly elevated at large dynamic ranges, but less 'high' than those suggested by Wiemer et al. (1998) (Table 4.1), who used a sample size of only 150 events. For Kilauea, the b-values are the highest seen in all of the catalogues here, and are in better agreement with the results of Wyss et al. (2001). Elevated b-values are also seen in both induced seismicity catalogues – substantially higher for the PNR catalogues (section 5.2.3.1) than for the Oklahoma data (section 5.2.3.2). From the analysis here, only the Kilauea and PNR catalogues exhibit truly high (> 1.5) b-values.

In conclusion, none of the real catalogues in this chapter have confirmed the hypothesis that high b-values may be due to the incorrect use of GR methods on an underlying MGR distribution, but the sensitivity analysis for the synthetic data has shown that this remains a possibility in the more general case, for example as shown for The Geysers data in section 4.5.2.

Chapter 6

Analytical solutions for bias in the Aki b-value applied to synthetic and real data

In an attempt to explain the results observed in sections 4.5 and 5.2, this chapter begins by introducing two analytical solutions for the relationship between the estimated b-value using Aki's (1965) method (i.e. assuming an unbounded GR distribution) on i) an underlying bounded GR distribution and ii) an underlying bounded MGR distribution, both for the case of an infinite sample with a finite maximum sampled magnitude in the distribution, ω. A new theory is then presented for the convergence of the mean magnitude $\bar{m}$ in the case of an exponential FMD, for a finite sample with a given dynamic range of observations up to the maximum sampled magnitude ω and the consequent convergence of the Aki-estimated b-value, denoted b_{Aki} here. In order to simplify the mathematical treatment somewhat, the term MGR in this chapter refers to a gamma distribution in the probability density function (pdf, section 4.3.1), rather than the cumulative probability form introduced in section 2.2 of Kagan and Jackson (2000) and adopted elsewhere in this thesis. The first version of the theory presented here was provided by Ian Main, unless otherwise indicated. This included

the original derivations for sections 6.1.1 and 6.1.2, which then underwent corrections by myself after my first attempt to test them on synthetic data highlighted some issues. These analytical results then further propagated into theories derived and discussed in sections 6.1.3 and 6.1.4.

6.1 Analytical theory for convergence of the b-value due to truncation and finite sampling

In the following two subsections, I show the derivations for corrections of the maximum likelihood (MLE) solutions for the estimated b-value for the underlying GR (section 6.1.1) and MGR (section 6.1.2) distributions in the case of a finite sampled maximum magnitude, assuming a perfectly sampled distribution. These derivations are based on expectation values for the mean magnitude in a finite sample, therefore assuming no uncertainty in primary measurement or finite sampling effects. This assumption will be challenged later in the chapter. Accordingly, the answers are therefore subject to additional uncertainties and potentially some bias, when compared to synthetic or real data. Section 6.1.3 will then show the derivation for the convergence of the mean magnitude and 6.1.4 use the previously derived solutions to establish a relationship between the b-value and the mean magnitude.

6.1.1 Correction of the MLE for the GR law in the case of a finite sampled maximum magnitude

The maximum likelihood solution of Aki (1965), outlined in section 3.2.3, assumes an infinite maximum magnitude in the underlying distribution, or an unbounded GR law. This simplifies the mathematics, but comes at the expense of having an infinite seismic moment release rate (section 2.4.1). The derivation is reproduced here to act as a reference for the case of a truncated distribution. Aki's (1965) formula for the expectation value of the magnitude as a function of the GR parameter $\lambda = b\ln(10)$

(where λ is the equivalent exponent for the base e), takes the form:

$$\langle m \rangle_\infty = \frac{\int_{m_c}^{\infty} m e^{-\lambda m}\, dm}{\int_{m_c}^{\infty} e^{-\lambda m}\, dm} \tag{6.1}$$

where the subscript ∞ acknowledges this equation holds when the sampled maximum magnitude is infinite. After integration, the denominator is:

$$\left| \frac{e^{-\lambda m}}{-\lambda} \right|_{m_c}^{\infty} = \frac{e^{-\lambda m_c}}{\lambda}$$

Then, to solve for the numerator let $u = m$, $du = dm$, $v = -e^{-\lambda m}$, $dv = \lambda e^{-\lambda m} dm$ and integrate by parts as follows:

$$\begin{aligned}
\int u\,dv = uv - \int v\,du &= \lambda \int m e^{-\lambda m} dm \\
&= \left| -m e^{-\lambda m} \right|_{m_c}^{\infty} - \int_{m_c}^{\infty} -e^{-\lambda m}\, dm \\
&= m_c e^{-\lambda m_c} + \left| \frac{e^{-\lambda m}}{-\lambda} \right|_{m_c}^{\infty} \\
&= m_c e^{-\lambda m_c} + \frac{e^{-\lambda m_c}}{\lambda}.
\end{aligned}$$

We can then combine the numerator and denominator to get:

$$\langle m \rangle_\infty = \frac{\frac{\int_{m_c}^{\infty} u\,dv}{\lambda}}{\frac{e^{-\lambda m_c}}{\lambda}}$$

$$\langle m \rangle_\infty = m_c + \frac{1}{\lambda} \tag{6.2}$$

Rearranging equation 6.2 for λ_∞ will give

$$\lambda_\infty = \frac{1}{\langle m \rangle_\infty - m_c} \tag{6.3}$$

Equation 6.3 is identical to that derived by Aki (1965), i.e. $\lambda_{Aki} = \lambda_\infty$. The same result can be obtained independently from maximum likelihood estimation (MLE, e.g. see Appendix A of Kijko and Smit (2012)), so Aki's solution is often called the MLE of λ.

To solve this equation for real data, Aki assumed that the mean magnitude in a finite sample was a good approximation for the expectation value, i.e.

$$\langle m \rangle = \bar{m}. \tag{6.4}$$

This is a common and often accurate assumption in statistical analysis, for example in statistical physics, in cases where where we are dealing with large sample numbers, for example Avogadro's number in a mole fraction of a gas. However, here we are often dealing with relatively small samples, so the mean magnitude is likely to be biased, and biased to low values compared to $\langle m \rangle$ because of the relative absence of rarer large events in a small sample. This hypothesis is tested later in section 6.1.3. Transforming equation 6.3 to base 10 using $e^{-\lambda m} = 10^{-bm}$ and including equation 6.4, gives finally

$$b_{Aki} = \frac{1}{(\bar{m} - m_c)\ln(10)}. \tag{6.5}$$

This is the form of Aki's equation that is usually quoted in the literature (section 2.5.1.2).

Derivation - modification for a finite sampled maximum magnitude

Now, we consider the case of a finite maximum observed magnitude ω:

$$\langle m \rangle_\omega = \frac{\int_{m_c}^{\omega} m e^{-\lambda m}\, dm}{\int_{m_c}^{\omega} e^{-\lambda m}\, dm} \tag{6.6}$$

After integration, the denominator is:

$$\left| \frac{e^{-\lambda m}}{-\lambda} \right|_{m_c}^{\omega} = \frac{e^{-\lambda m_c} - e^{-\lambda \omega}}{\lambda}$$

and the numerator is:

$$\frac{\left[\left| -m e^{-\lambda m} \right|_{m_c}^{\omega} + \left| \frac{e^{-\lambda m}}{-\lambda} \right|_{m_c}^{\omega} \right]}{\lambda}$$
$$= \frac{m_c e^{-\lambda m_c} - \omega e^{-\lambda \omega}}{\lambda} + \frac{e^{-\lambda m_c} - e^{-\lambda \omega}}{\lambda^2}$$

Combining the numerator and denominator and simplifying gives us the final analytical solution:

$$\langle m \rangle_\omega = \frac{\frac{\lambda[m_c e^{-\lambda m_c} - \omega e^{-\lambda \omega}]}{\lambda}}{e^{-\lambda m_c} - e^{-\lambda \omega}} + \frac{1}{\lambda} \tag{6.7}$$

In general, $\langle m \rangle_\infty$ (equation 6.2) and $\langle m \rangle_\omega$ (equation 6.7) will be different in an infinite sample for the same value of λ because of the missing events above magnitude ω in the latter. However, in the case of a finite sample as in any analysis of real data, our estimates of $\langle m \rangle_\infty$ or $\langle m \rangle_\omega$ will be the mean magnitude observed in the sample (equation 6.4). If we substitute the expectation values in equations 6.2 and 6.7 with the same mean observed magnitude, we obtain

$$\bar{m} = \frac{m_c e^{-\lambda m_c} - \omega e^{-\lambda \omega}}{e^{-\lambda m_c} - e^{-\lambda \omega}} + \frac{1}{\lambda} = m_c + \frac{1}{\lambda_\infty}$$

Rearranging this and solving for $\frac{1}{\lambda_\infty} = \frac{1}{\lambda_{Aki}}$ gives

$$\boxed{\frac{1}{\lambda_{Aki}} = \frac{1}{\lambda} + \frac{m_c e^{-\lambda m_c} - \omega e^{-\lambda \omega}}{e^{-\lambda m_c} - e^{-\lambda \omega}} - m_c.} \tag{6.8}$$

The final analytical solution (equation 6.8) shows that when the maximum observed magnitude, ω, approaches infinity, $\lambda_{Aki} \rightarrow \lambda$ and given $m_c < \omega$, then $\frac{1}{\lambda_{Aki}} < \frac{1}{\lambda}$, and hence we would expect $\lambda_{Aki} > \lambda$. Both imply the λ_{Aki} is likely to be based to high values in small samples, and to converge to the true λ from above. The bias is then equal to the difference between λ_{Aki} and the true λ.

6.1.2 Correction of the MLE for the MGR law in the case of a finite sampled maximum magnitude

This section presents a similar correction of the MLE for the MGR law in the case of a finite sampled maximum magnitude.

Derivation - modification for a finite sampled maximum magnitude

The moment magnitude equation is given as:

$$M(m) = 10^{cm+d} \tag{6.9}$$

where c = 1.5, d = 9.1 in SI units. For the MGR distribution in M, in a finite sample of magnitudes between m_c and ω,

$$p(m) \sim M(m)^{-\beta} e^{-\frac{M(m)}{M_\theta}} \sim e^{-\lambda m} e^{-\left(\frac{10^{cm+d}}{10^{cm_\theta+d}}\right)} = e^{-\lambda m} e^{-10c(m-m_\theta)}$$

where

$$\beta = \frac{b}{c} \text{ and } \lambda = b\ln(10)$$

The probability density function for the unlimited Gutenberg-Richter law is recovered, $p(m) \sim e^{-\lambda m}$ in the case m_θ, $M_\theta \to \infty$. In the general case of finite M_θ, m_θ, the expectation value for the magnitude can be expressed in the form

$$\langle m \rangle - m_c = \frac{\int_{m_c}^{\omega} m e^{-\lambda m} e^{-10^{c(m-m_\theta)}} dm}{\int_{m_c}^{\omega} e^{-\lambda m} e^{-10^{c(m-m_\theta)}} dm} - m_c \tag{6.10}$$

where the denominator in the first term on the right-hand side (RHS) ensures the integral of the probability density function over all values is unity. From equation 6.4, the left hand side (LHS) of this equation can be substituted by $\frac{1}{\lambda_{Aki}}$. Thus, for the MGR distribution,

$$\boxed{\frac{1}{\lambda_{Aki}} = \frac{\int_{m_c}^{\omega} m e^{-\lambda m} e^{-10^{c(m-m_\theta)}} dm}{\int_{m_c}^{\omega} e^{-\lambda m} e^{-10^{c(m-m_\theta)}} dm} - m_c.} \tag{6.11}$$

Unfortunately, it is not possible to reduce this equation further analytically, so it is solved by numerical integration.

6.1.3 Derivation of the convergence of the mean magnitude for an exponential FMD

This section explores the veracity of the assumption that the mean magnitude in a finite sample is a good approximation for the expectation value of the magnitude in an infinite sample (equation 6.4) and shows how one converges to the other as more data are added in a random sample of a finite number of events.

The probability density function for an exponential distribution with a scale factor $\frac{1}{\lambda}$ for $x \geqslant 0$ is

$$p(x) = \lambda e^{-\lambda x}. \tag{6.12}$$

The probability density function for the mean $\bar{x}$ in a sample of size n taken from an exponential distribution in a variable takes the form

$$p(\bar{x}) = \frac{\lambda^n n^n}{\Gamma(n)} x^{n-1} e^{-n\lambda x} \tag{6.13}$$

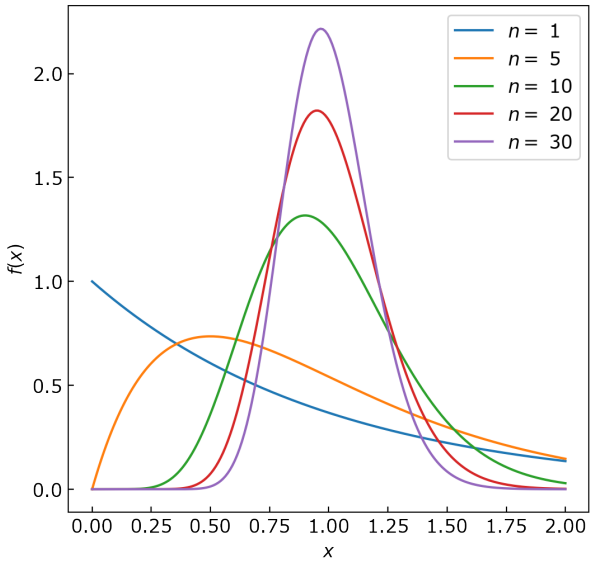

Figure 6.1: The gamma function(n, 1/n) for several values of n.

where $\Gamma(n) = (n-1)!$ is the gamma function (see `http://webhome.auburn.edu/~carpedm/courses/stat3610b/documents/Exponential_sampling.pdf` for this specific form, noting the difference in nomenclature). Therefore, when $n = 1$, this is the standard exponential probability density function (Figure 6.1), where the most likely outcome (the mode) occurs at $\bar{x} = 0$. When n is large, $p(\bar{x})$ approaches a Gaussian with a mode centred on the mean of the distribution $p(\bar{x})$. Prior to this the mode is less than the mean magnitude. The maximum likelihood for the mean value of $\bar{x}$ is defined by the criterion

$$\frac{dp(\bar{x})}{dx} = 0$$

This expresses the mean magnitude that is *most likely to be sampled* in a given population of trials, for example in the analysis of synthetic catalogues produced by random sampling of an underlying GR distribution (section 4.4.2). Substituting equation 6.13 using the product rule for differentiation then gives

$$\frac{\lambda^n n^n}{\Gamma(n)} \frac{d(\bar{x}^{n-1} e^{-n\lambda\bar{x}})}{dx} = \frac{\lambda^n n^n}{\Gamma(n)} \left[(n-1)\bar{x}^{n-2} e^{-n\lambda\bar{x}} - \bar{x}^{n-1} \lambda n e^{-n\lambda\bar{x}} \right] = 0.$$

The term outside the brackets is a constant, so the term in brackets is zero. After

taking out a common factor $x^{n-2}e^{-n\lambda x}$, we have

$$\bar{x}^{n-2}e^{-n\lambda\bar{x}}\Big[(n-1)-\bar{x}\lambda n\Big]=0.$$

This implies a single maximum in $p(\bar{x})$ when

$$\bar{x}=\frac{(n-1)}{\lambda n}. \tag{6.14}$$

We recover $\bar{x}=0$ at $n=1$ and $\bar{x}=\frac{1}{\lambda}=\bar{x}_\infty$ for an infinite sample. The convergence for $\bar{x}$ then takes the form

$$\bar{x}=\bar{x}_\infty\frac{(n-1)}{n} \tag{6.15}$$

This confirms that we expect the most likely sample mean to start at zero in a sample of $n=1$ and trend in a non-linear fashion asymptotically to $\bar{x}=\bar{x}_\infty$ from below.

This equation specifies the convergence of mean magnitude in a finite sample of events above m_c as a function of the number of events n. However, the focus in this thesis is the dynamic range $r=\omega-m_c$ where ω is the largest magnitude in a sample. We account for this in the next section.

Accounting for a finite minimum in x

If we have a finite threshold such that $x\geqslant x_{min}$, then the previous equation 6.15 becomes

$$\bar{x}-x_{min}=(\bar{x}_\infty-x_{min})\frac{(n-1)}{n}. \tag{6.16}$$

In this section, we calculate a relationship between the total number of events in a sample and the dynamic range, assuming the total number of events in the largest sampling bin $n(\omega-dm,\omega)=1$. We do this by calculating the total number of events in the sample for different threshold magnitudes 0, m_c and $\omega-dm$. The integrals of the probability density function for the incremental version above these thresholds are then

$$I_0=\int_0^\infty\lambda e^{-\lambda x}=e^{-0}-e^{-\infty}=1.$$

This equation proves the functional form of equation 6.12 is correct, i.e. unit total probability is achieved when the pre-exponential factor equals the exponent λ, so the

pdf is specified by a single variable. In the case of a finite magnitude threshold m_c and finite sampled maximum magnitude ω, the cumulative probability is

$$I_{m_c} = \int_{m_c}^{\omega} \lambda e^{-\lambda x} = e^{-\lambda m_c} - e^{-\lambda \omega}$$

and for the largest bin

$$I_{\omega - dm} = \int_{\omega - dm}^{\omega} \lambda e^{-\lambda x} = e^{-\lambda(\omega - dm)} - e^{-\lambda \omega}$$

The ratio of the two cumulative probabilites I_{m_c} and $I_{\omega - dm}$ must be equal to the ratio of the numbers of events in the two samples n_{m_c} and $n_{\omega - dm}$. Given we know $n_{\omega - dm} = 1$ and $\frac{n_\omega - m_c}{n_\omega - dm} = \frac{I_{mc}}{I_\omega - dm}$ by proportion, we have

$$n_{m_c} = \frac{e^{-\lambda m_c} - e^{-\lambda \omega}}{e^{-\lambda(\omega - dm)} - e^{-\lambda \omega}} = \frac{e^{\lambda(\omega - m_c)} - 1}{e^{\lambda dm} - 1}.$$

Substituting this solution for $n(\lambda, m_c, \omega, \delta m)$ into equation 6.16, the mean magnitude then converges (changing variables from x to magnitude m) according to

$$\boxed{\bar{m} - m_c = (\bar{m}_\infty - m_c)\left[1 - \left(\frac{e^{\lambda dm} - 1}{e^{\lambda(\omega - m_c)} - 1}\right)\right].} \tag{6.17}$$

6.1.4 Relationship between λ_{Aki} and the convergence of the mean magnitude

I combined equations 6.3 and 6.17 to derive the following relationship between λ_{Aki}, λ, m_c, ω and dm

$$\boxed{\frac{1}{\lambda_{Aki}} = \frac{1}{\lambda}\left[1 - \left(\frac{e^{\lambda dm} - 1}{e^{\lambda(\omega - m_c)} - 1}\right)\right].} \tag{6.18}$$

In equation 6.4, I stated the often implicit assumption that the mean magnitude $\bar{m}$ is a good approximation for the expectation value, $\langle m \rangle$. However, the derivation in section 6.1.3 shows that in the range where convergence has not yet been reached, this assumption is not correct and the sampled $\bar{m}$ is likely to actually be an underestimate of the true expectation value $\langle m \rangle$. Therefore, equation 6.4 becomes an inequality, and consequently equation 6.3 does too:

$$\frac{1}{\lambda_{Aki}} > \bar{m} - m_c. \tag{6.19}$$

This analytical theory section has derived four solutions to estimate the GR (equation 6.8) and MGR (equation 6.11) MLEs with a finite sampled maximum magnitude, the mean magnitude convergence for an exponential FMD (equation 6.17) and the comparison of λ_{Aki} compared to equation 6.17 (equation 6.18). These four equations are all stated in boxes in the derivation above for ease of cross-reference. In the following two sections, I will test the validity of these four equations in comparison with both synthetic data (section 6.3) and real data (section 6.4).

6.2 Solving analytical derivations numerically

I solved equations 6.8 and 6.11 numerically by calculating their component parts, equating the difference between the LHS and RHS sides to zero, and looking for a minimum in the square of the function LHS-RHS = 0 in each case.

In the case of equation 6.17, I use an existing scipy function in Python to estimate the non-linear least squares fit. This returns optimal parameters values after iteration, so that the sum of the squared residuals is minimised. The parameters returned by this optimisation were then fed into equation 6.18 and solved for λ_{Aki} (and equivalent b_{Aki}).

These algorithms are applied to the analysis of synthetic data in section 6.3 followed by real data in section 6.4.

6.3 Application to synthetic catalogues

In this section, I show the results of comparing all four analytical solutions derived in section 6.1 above with finite data samples taken from underlying GR and MGR distributions.

Firstly, to compute the analytical solution in equation 6.8, I chose a fixed value of λ = $\ln 10$ (so that b = 1.0 and representative of an active tectonic zone, but not the most active) and set the value of m_c to 1.0, similar to trial values used in the previous two chapters. The value of the maximum possible observed magnitude, ω is fixed at 7.0 and in the case of 6.11 is governed by m_θ which equals 3.5 (in line with the 'standard'

parameters for volcanic and induced seismicity employed throughout this thesis, see sections 3.5.1 and 4.4.2). The solutions are solved numerically to return values for λ_{Aki} and the equivalent b_{Aki}. To obtain a comprehensive picture of the full solution, I then increase the value of m_c in increments of 0.1 to 'thin' the catalogues and run the process again.

The figures in this section (as well as the next section 6.4) follow the same visual representation – I plot catalogue data (mean magnitudes on the top plot and b-values on the bottom plot, both as a function of dynamic range) as blue or green circles with a solid black line indicating the sampled mode. Green circles (in synthetic data only) indicate mean magnitudes and b-values computed from 2 or fewer data points in a sample. Most of these plot on a straight line (shown in brown in the case of the synthetic data), representing a hard edge to the range of possibilities – a clear artefact of such small samples. These outliers resulted in a significant bias in the best fit to equations 6.17 and 6.18 and hence were discarded prior to obtaining the best fit and the confidence intervals. The area in light blue represents the 95% confidence intervals of the outcomes determined by the scatter in the data indicated in blue circles for the synthetic data, and the error estimates (given in the next paragraph) of $\bar{m}$ and b for the real data. Equation 6.8 is shown as a solid red curve, equation 6.11 as a dashed red curve and equation 6.18 as a solid bright green curve on the bottom plot. Equation 6.17 is shown as a solid bright green curve on the upper plot.

Having described the generic properties of the figures, I now discuss the metrics that can be extracted from them to allow a quantitative comparison of the rate of convergence in our estimates of the b-value and the most likely, or modal value of, the mean magnitude $\bar{m}$. The aleatory or statistical uncertainty in the estimated b-value at 95% confidence for a single catalogue is the product of a factor 1.96 and the standard deviation in b expected from the standard deviation in $\bar{m}$ (Aki, 1965):

$$\delta b = \pm \frac{1.96b}{\sqrt{n}} \tag{6.20}$$

which is labelled as 'δb' on the figures in b). Aki (1965) derived this formula by

propagating the equivalent uncertainty in $\bar{m}$ using equation 6.3. The primary sampling error in $\bar{m}$ at 95% confidence is:

$$\delta\bar{m} = \pm\frac{1.96(\bar{m} - m_c)}{\sqrt{n}} \tag{6.21}$$

which is labelled as '$\delta\bar{m}$' on the figures in a). These represent the irreducible random errors in our estimates of b and $\bar{m}$. We consider an estimate of the mode in $\hat{b}$ to be accurate when the systematic error or bias $\hat{b} - b \leqslant 0.1$, and to be precise when $\delta b \leqslant 0.1$. In the case of the real data, the true values of m_∞ and b are not known, so we cannot define an accuracy. The equivalent criteria for the estimates of mean magnitude are a bias $\hat{\bar{m}} - \bar{m}_\infty \leqslant 0.05$ and a precision $\delta m \leqslant \frac{0.1}{\ln(10)} \leqslant 0.05$. Strictly exact equivalence occurs when $\delta m \leqslant \frac{0.1}{\ln(10)}$ but pragmatically we choose the approximation $\frac{0.1}{2}$ here. In the case of the synthetic data, there are multiple catalogues, and the resulting data scatter provides a more complete range of uncertainties. In this case, I estimate the confidence intervals from the data scatter. These reflect better the total uncertainty, including epistemic uncertainty resulting from lack of complete knowledge in a single sample. Table 6.1 shows a summary of the threshold dynamic ranges $r_a(b, \bar{m})$ and $r_p(b, \bar{m})$ for the figures both below and in the following section (6.4), where these thresholds are met for accuracy and precision respectively for estimates of b and $\bar{m}$. Therefore, in the following, I will be focusing on how the equations stated in section 6.1 behave compared to this data.

In Figure 6.2, the data is randomly sampled from a GR distribution. In Figure 6.2a), $\bar{m}$ converges from below as a function of dynamic range and there is good agreement between the mode and that predicted by equation 6.17. From a dynamic range of ~ 2.1 onwards, the mode is within ± 0.05 units of $\bar{m}_\infty$. The analytical upper bound (brown line) for $\bar{m}$ is shown for reference. Figure 6.2b) shows the corresponding Aki b-values as a function of dynamic range. The analytical solution for the truncated GR law (equation 6.8) falls within the scatter in the synthetic data but overestimates the observed mode at dynamic ranges less than 2.0, which is where it converges to the true $b = 1.0 \pm 0.1$ units. In the case of the truncated MGR distribution (equation 6.11), a limited dynamic

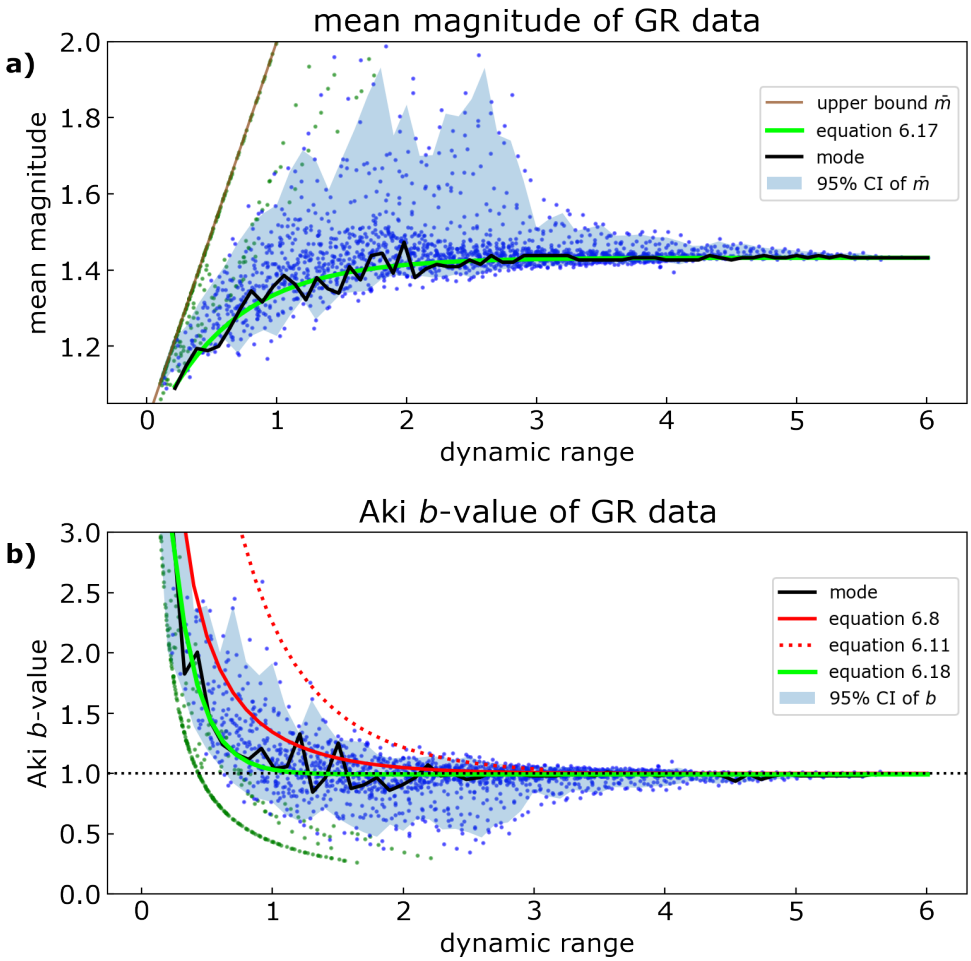

Figure 6.2: a) Mean magnitude as a function of dynamic range for GR data. The straight brown line indicates an upper bound to the mean magnitude in a sample of 1. The solid black line shows the mode for the plotted synthetic data (blue circles) and the bright green curve represents equation 6.17 fitted to this data. b) b-values as a function of dynamic range for GR data with green circles being discarded data, the solid black line the mode and the bright green curve equation 6.18 fitted to the mode. The bounded GR solution (equation 6.8) is shown as a solid red curve and the bounded MGR solution (equation 6.11) as a dashed red curve.

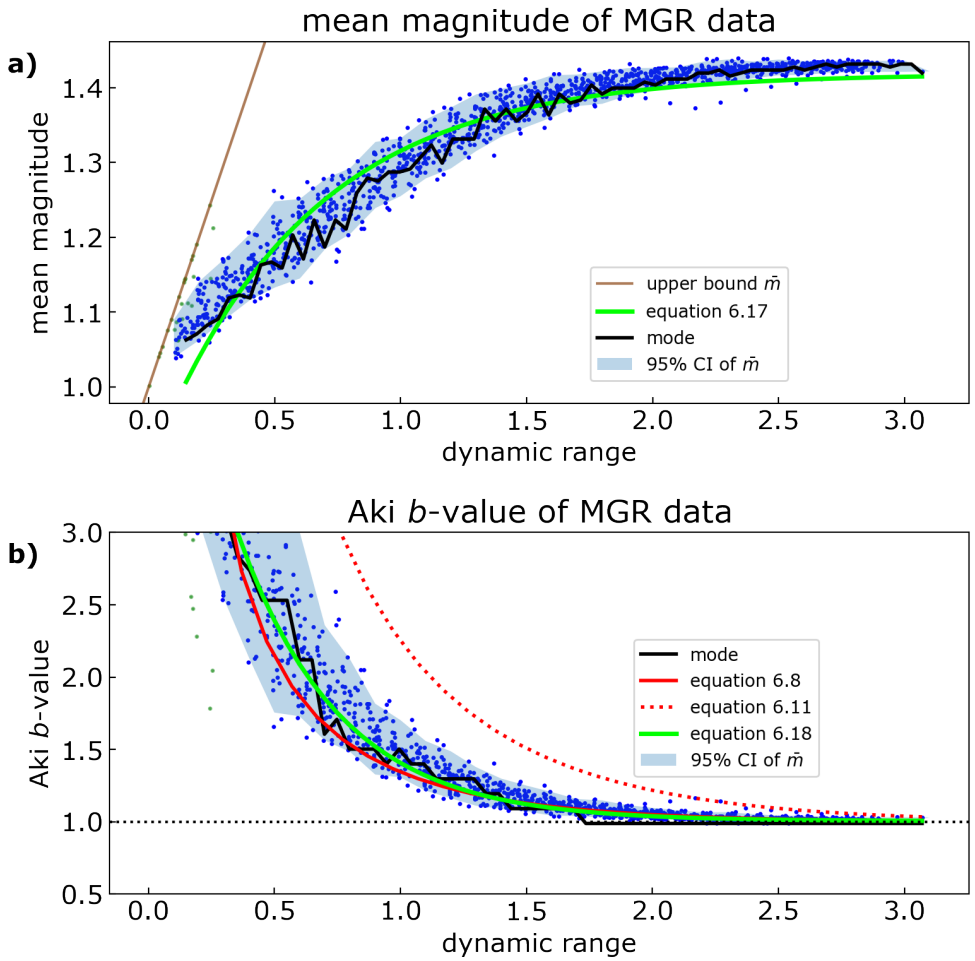

Figure 6.3: a) Mean magnitude as a function of dynamic range for MGR data. b) b-values as a function of dynamic range for MGR data, presented as explained in the caption to Figure 6.2.

range is available due to the characteristic magnitude term $m_\theta = 3.5$, which governs the resulting value of ω. Equation 6.11 grossly overestimates the b-values, i.e. at a dynamic range of 1, equation 6.11 returns $b = 2.2$, whereas equation 6.8 returns 1.3 while the best fit 6.18 suggests b has already converged to true $b = 1.0$ within $\pm$ 0.1 and therefore much quicker. This is not surprising because, in this case, we know MGR is the incorrect distribution.

I now repeat this analysis for synthetic data sampled from an MGR distribution in Figure 6.3. The dynamic range observed in the synthetic data is considerably smaller than for GR data due to the value of m_θ and the associated roll-off in frequency for

larger events. In Figure 6.3a), the convergence in equation 6.17 is not as good a fit to the mode as is the case in Figure 6.2a) and struggles to fit the synthetic data even above dynamic ranges of 2.0. On the other hand, both the bounded GR solution (equation 6.8) and the modal b-value estimate (equation 6.18) in Figure 6.3b) are in very good agreement with the observed mode of the sampled b-values. Perhaps surprisingly, the bounded MGR solution (equation 6.11) does not fit the underlying, MGR distributed, data better than the bounded GR solution.

In summary, equation 6.8 is a moderately good first order estimate to both the synthetic GR (Figure 6.2a) and MGR (Figure 6.3a) data shown here and confirms that we would expect the estimated b_{Aki} to be biased to high values for small samples, and converge to the correct value from above as ω increases. In the case of equation 6.11 however, the bounded MGR solution overestimates the underlying b-values as it cannot sample the true roll-off. The inequality correction in equation 6.19 is particularly exacerbated in the case of the bounded MGR model in Figure 6.3b), possibly explaining the substantial overestimate in b-values at small dynamic range (section 2.4.2).

In the next sub-section I will show the same set up of figures as in this section but for two real earthquake catalogues analysed in chapters 4 and 5.

6.4 Application to real catalogues

As demonstrated in section 6.3, the generation of many catalogues with known true b-value allows us to identify the bias and the precision from the mode and the spread in the estimated values of $\bar{m}$ and b for many finite samples of an underlying FMD. However, in the real data, we only have one sample, i.e. a single catalogue and we do not know the true b-value. For this part of the analysis, I have chosen one catalogue which I have previously suggested to be MGR distributed – The Geysers (section 4.5.2). I have also chosen to revisit a tectonic catalogue 5.2.1.2 (Southern California) because it strongly prefers the GR distribution (see Figure 5.16b) in nearly all cases of thinned sub-catalogues. In both cases, I have set a strong prior of $b = 1.0$ because I have shown

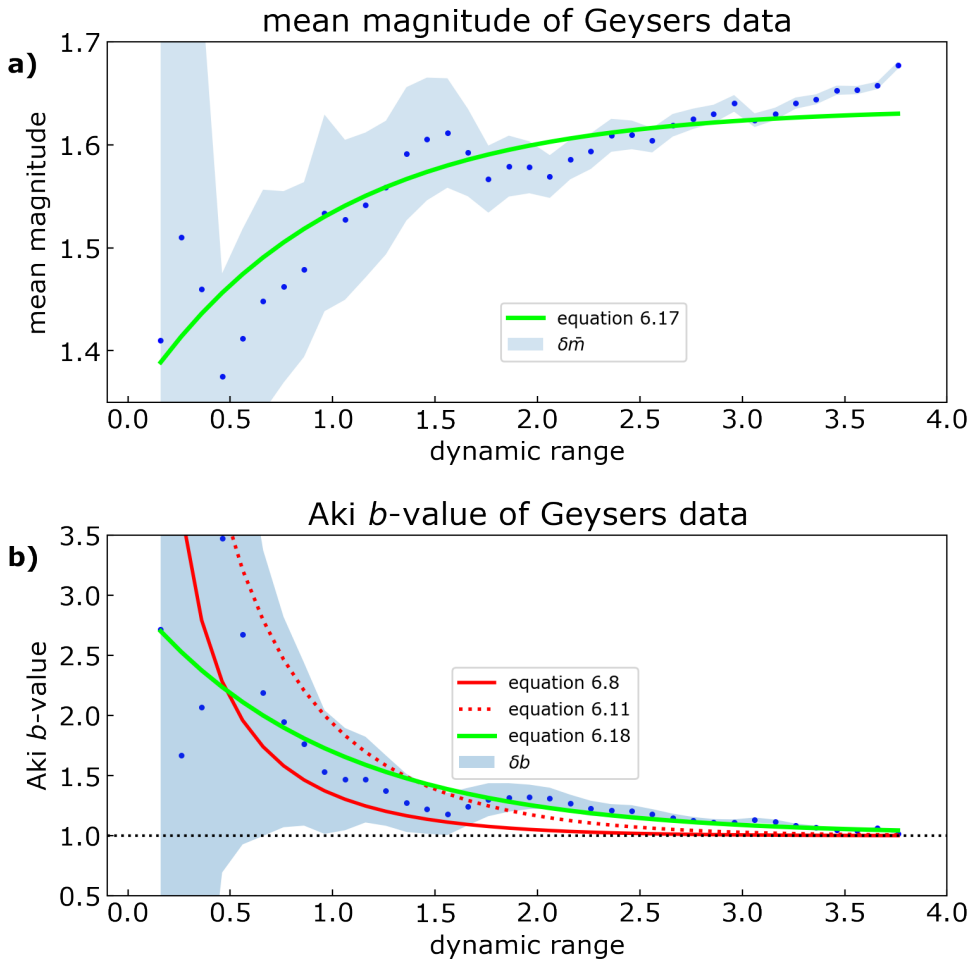

Figure 6.4: a) Mean magnitude as a function of dynamic range for The Geysers data. b) b-values as a function of dynamic range for The Geysers data, presented as explained in the caption to Figure 6.2.

in chapters 4 and 5 that with sufficient data and dynamic range, b-values are not likely to be elevated in these catalogues. The mean magnitude in Figure 6.4a) is very variable, mirrored in equation 6.17 which does not fit the data as well as in the case for the synthetic data. An accuracy here has not been defined for $\bar{m}$ and b (see Table 6.1) because the true, underlying values are unknown. Nonetheless, equation 6.17 does not seem to reach convergence to a flat asymptote within the dynamic range observed and nor does the data. In the case of Figure 6.4b), fitting both bounded corrections, the b-values lie roughly in the middle of these solutions up to a dynamic range of ~ 1.5. This suggests that at dynamic ranges lower than this, The Geysers catalogue exhibits

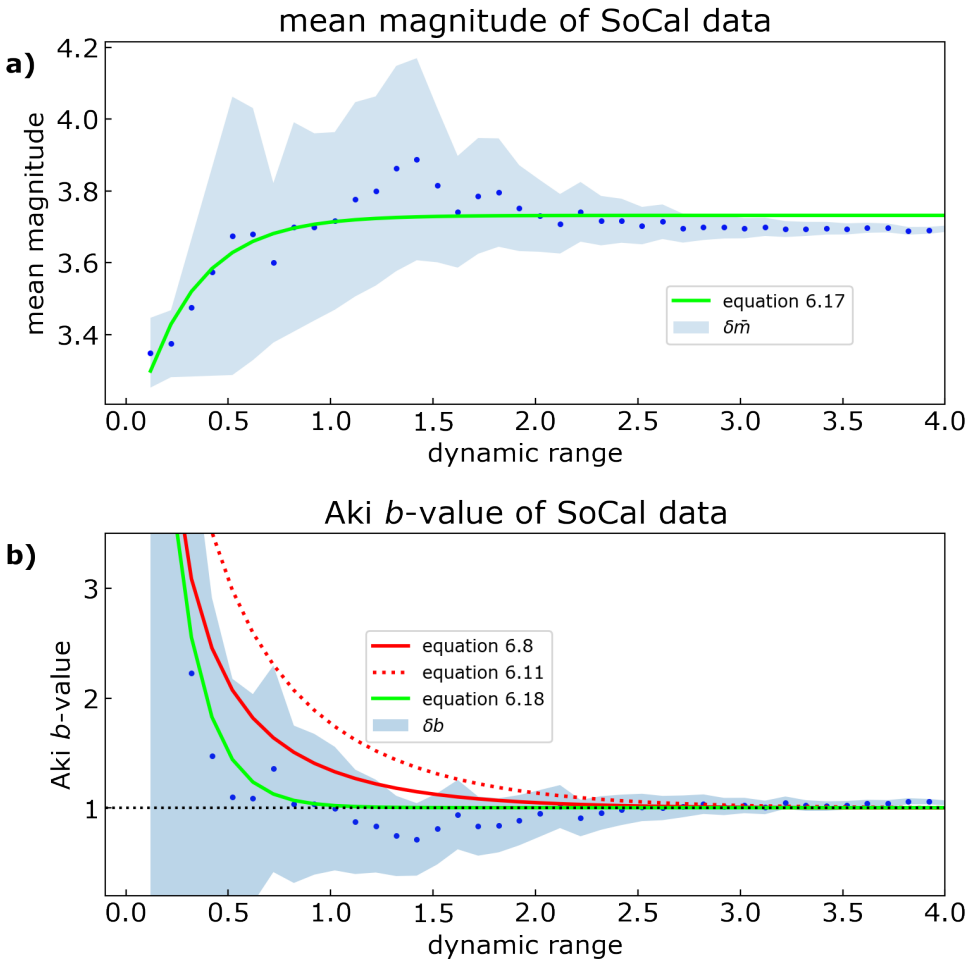

Figure 6.5: a) Mean magnitude as a function of dynamic range for SoCal data. b) b-values as a function of dynamic range for SoCal data, presented as explained in the caption to Figure 6.2.

neither (bounded) GR or MGR behaviour – one solution underestimates the b-values while the other one overestimates them before convergence is reached to ± 0.1 unit at a dynamic range of 1.8 for equation 6.8 and 2.5 for equation 6.11. The mean magnitude for the SoCal data in Figure 6.5a) does not exhibit as much variability as in the case of The Geysers (Figure 6.4a). The catalogue data becomes stable at a dynamic range of around 2.5 ± 0.05 units. The best fit (equation 6.18) falls within the 95% confidence intervals up to a dynamic range of 3.5 and converges to true $b = 1.0$ within ± 0.1 units at a dynamic range of 0.9. The bounded GR and MGR solutions overestimate the observed b-values until convergence (within ± 0.1 units) at 1.9 for the bounded GR and 2.5 for

Figure	Data	dyn.$r_a(\bar{m})$	dyn.$r_p(\bar{m})$	dyn.$r_a(b)$	dyn.$r_p(b)$	b-value	est.$\bar{m}_\infty$	δm
6.2	synthetic GR	3.5	4.1	3.5	4.0	1.0	1.43	0.05
6.3	synthetic MGR	1.2	1.8	2.0	2.3	1.0	1.42	0.05
6.4	The Geysers	-	1.7 - 3.1	-	2.3	1.0	1.64	0.05
6.5	Southern California	-	2.2	-	2.0	1.0	3.73	0.05

Table 6.1: Threshold dynamic ranges for accuracy r_a and precision r_p at 95% confidence when bias or precision respectively are less than 0.1 units in b and less than 0.05 units in $\bar{m}$. The accuracy cannot be estimated for the real data because the underlying values are not known.

the bounded MGR.

In this section, I have shown the application of the analytical theory in section 6.1 on two real catalogues chosen due to their preferences for the MGR and GR distributions. The results show the finite sampling theory (sections 6.1.3 and 6.1.4) provides a better fit to the mode than the infinite sample theories (sections 6.1.1 and 6.1.2), and that convergence in the real data is both more erratic and less likely to be established within the finite ranges observed, than for the ideal synthetic data. I now discuss the implications of the analyses presented in sections 6.4 and 6.3.

6.5 Discussion

The analysis in sections 6.3 and 6.4 and the precision of the b-values, explicitly stated in Table 6.2, has shown that the ratio of 95% confidence limits, $\frac{\delta b_s}{\delta b_R}$ for synthetic GR data compared to Southern California data, increases for increased dynamic range – from a factor of 1.0 at dynamic range of 1.0 up to a factor of 3.0 at a dynamic range of 4.0. In the case of synthetic MGR data, compared to The Geysers, the precision of the δb-value also increases, albeit much slower. At small dynamic ranges, the factor is only $\frac{1}{2}$, up to a factor of 1.0 at a dynamic range of 3.0. This means the epistemic uncertainty (Aki, 1965) in a single catalogue is less than the aleatory uncertainty obtained from the scatter of the synthetic data and especially in the case of MGR. This highlights the importance of considering both epistemic and aleatory uncertainty in probabilistic seismic hazard analysis (2.3).

Figure	Data	dyn. r = 1.0	dyn. r = 2.0	dyn. r = 3.0	dyn. r = 4.0
6.2	synthetic GR	1.1	0.65	0.3	0.15
6.5	Southern California	1.1	0.45	0.15	0.05
	ratio $\frac{\delta b_s}{\delta b_R}$	**1.0**	**1.4**	**2.0**	**3.0**
6.3	synthetic MGR	0.45	0.1	0.05	-
6.4	The Geysers	0.95	0.2	0.05	-
	ratio $\frac{\delta b_s}{\delta b_R}$	**0.5**	**0.5**	**1.0**	**-**

Table 6.2: Convergence of the precision in the estimated b-value (δb) as a function of dynamic range for GR and MGR sampled data compared to Southern California and The Geysers data, respectively. For the synthetic data, δb is estimated from the scatter in outcomes at 95% confidence, denote δb_s and for the real data from equation 6.20, also representing 95% confidence, denoted δb_R. The rows in bold state the ratio of $\frac{\delta b_s}{\delta b_R}$.

In the case of the synthetic data, the thresholds for convergence in dynamic range for accuracy $r_p(b)$ and precision $r_p(b)$ in the estimated b-value define the following ranges in all cases:

$r < r_a$ estimates are neither accurate nor precise

$r_a \leqslant r < r_p$ estimates are accurate but not precise

$r > r_p$ estimates are accurate and precise.

For the real data, the accuracy cannot be estimated, and convincing convergence to a flat asymptote is often not reached, even if the precision can be very good. Sometimes the precision gives a false impression of accuracy defined in this way – it can be very small when convergence has not been established. This observation highlights two issues with the primary catalogue – i.e. it contains events that are not independent e.g. aftershocks, and is much more susceptible to catalogue heterogeneity. The theory for the sampled synthetic data assume events are independent and identically-distributed in the case of a perfect homogeneous catalogue of magnitudes. In a real catalogue, the magnitudes are measured indirectly from an evolving network with a finite number of stations with a finite sample of the radiation pattern. This leads to much larger variability in the convergence trend, and often a lack of convergence to a flat asymptote.

This provides an additional, aleatory source of error in the case of real catalogues.

The convergence of mode in the mean magnitude expected in a finite sample (equation 6.17) in the case of an exponential FMD has shown good correlation to the observed mode in the mean magnitude in the cases of synthetic data (section 6.3). For The Geysers case, the trend in the mean magnitude distribution with respect to dynamic range is very variable, not allowing a conclusive estimate of the value of $\bar{m}$ at convergence. At the same time, the precision can be very small, and smaller than the variability in the mode. This suggests that in such cases the estimated Aki b-value is precise but not accurate. On the other hand, the improvement in accuracy of the estimated b-value in synthetic cases as a function of dynamic range (Table 6.1) can be attributed to the improvement in the estimated mean magnitude as sample sizes increase. This effect, captured in equation 6.18, provides a better explanation for the trend than the analytical corrections of equations 6.8 and 6.11, in particular for the resultant high b-values at small dynamic range. Additionally, I have assumed a strong prior of $b = 1.0$ in all cases here, in line with the results stated in sections 4.5.2 and 5.4. However, using the estimated $\bar{m}_\infty$ of equation 6.17 to estimate the b-value may suggest a more robust value to which b converges and should be examined in future work.

The bounded MGR solution (equation 6.11) overestimates b-values in all cases and does not fit well to synthetic or real data. The closest example is The Geysers (Figure 6.4) where it lies within confidence intervals at least for dynamic ranges less than 2.0. The general overestimation of the analytical MGR solution compared to the underlying b-value can be attributed to finite Poisson sampling (for the scatter in data and its reduction with increased dynamic range), empty bins, and using the incorrect model as well as the double exponential in the numerator of equation 6.11.

I have also shown that if, in a finite sample (i.e. equation 6.8), the value of $\langle \mathrm{m} \rangle$ is underestimated, then through the relationship between m and λ in $\langle \mathrm{m} \rangle = m_c + \frac{1}{\lambda}$, the value of b_{Aki} is an overestimate. This observation is an additional explanation for why b-values are biased to high values at low dynamic ranges, both seen in synthetic (section 6.3) and real data in this chapter, and previously also observed in chapters 4 and 5.

Finally, the real catalogues (particularly The Geysers catalogue) contain very few (but > 2 and hence not discarded) events at dynamic ranges < 0.5. Hence, removing these is likely to lead to a better fit to equations 6.17 and 6.18. This could be investigated in future work. Additionally, looking at more real catalogues would be beneficial in further quantifying the accuracy and precision of the solutions presented here. Similarly, the epistemic uncertainty arising from a finite sample could be explored by Monte-Carlo sampling of the best estimate of the underlying distribution parameters, to see if the net uncertainty from this is greater than that estimated from equation 6.3.

6.6 Summary

To summarise the analysis in this chapter, I have suggested four analytical models to explain why inferred b-values increase as m_c increases and dynamic range decreases, and proven that the accuracy of the b-value depends on the accuracy of the assumption that the mean magnitude in a finite sample is representative of the expectation value of the magnitude in an infinite sample. This confirms that given an underlying GR model, small dynamic range results in systematically high b-values, both analytically (section 6.1) and due to the convergence of the mode in the sampled mean event magnitude (section 6.1.3). Together this analysis indicated that many published studies in the literature use dynamic ranges where we would expect significant bias in the b-value, resulting in a systematic error in the likelihood of large events in a larger sample. The equations presented here provide the possibility for a correction applied to b-values estimated from small dynamic ranges and sample size and in future work, should be tested in practice to determine whether it provides a useful improvement to biased b-values.

Chapter 7

Discussion

Throughout this thesis, I have examined how the b-value varies in synthetic and real catalogues in different seismicity settings, primarily as a function of limited dynamic range in finite data samples. Additionally, I have considered the impact of the misinterpretation of the underlying distribution and incorrect applications of GR methods on such data, leading to further sources of bias in estimated b-values.

In this chapter I discuss the significance and implications of these results in a wider context and how they can be applied in future work, particularly in varied settings of seismicity and how this can feed into seismic hazard and PSHA calculations.

7.1 Bias of the b-value

I have used synthetic data in sections 4.4.2 and 5.1 to show that b-values are likely to be overestimated when there is little data, little dynamic range, or when the assumed distribution is not in agreement with that of the underlying data. This is expected from previous studies that have investigated the minimum number of events required to estimate the b-value (Zöller, 2013; Roberts et al., 2015; Nava et al., 2017). However, most of these focus exclusively on the number of events, without explicit consideration of the dynamic range. A combination of two or more of these cases (low number of events, low dynamic range and disagreement between the assumed distribution and

underlying data), will result in substantially larger bias in the estimated b-value.

When sampling data from an underlying MGR distribution, the simpler, GR model is preferred by the model selection criterion (BIC) when the catalogue is reduced in data and dynamic range (section 4.5), even when the underlying data is sampled from an MGR distribution. The analysis of model preference and estimated b-values in both sections 4.5.2 and 5.2 on real catalogue data has provided further confidence in estimated b-values being biased to larger values than they should be. The catalogues used here suggest that this is less due to misjudgment of the underlying distribution (GR or MGR) and use of incorrect methods if this were the case, but rather suggests that the majority of biased, high b-values are due to lack of data and dynamic range. Only in a single case identified here – The Geysers (induced seismicity) – was the reasoning for high b-values (and model preference) a result of misjudgement of the underlying distribution. At large dynamic range and data, The Geysers data returns b-values $\sim$ 1.0 (compared to < 1.0 in Henderson et al. (1999)), with a model preference of MGR. These b-values at The Geysers are increasingly biased to high values, and the model preference increasingly changes to GR, with reduction of The Geysers catalogue in data and dynamic range (section 4.5.2). This is a case where all three of the above mentioned limitations – lack of data, dynamic range and misinterpretation of the underlying distribution – are combined, returning results that are inaccurate and imprecise when small sub-catalogues are used. This is in agreement with Eaton et al. (2014) where the method applied cannot distinguish between a GR distribution with a steep slope and the more gradual roll-off to a steeper slope displayed in the MGR when m_c is high in a finite data sample. This is a strong example of issues that may occur in estimating b-values in real catalogues, particularly in volcanic and induced seismicity settings, where the catalogues are often small and lacking data, as extensively discussed in section 4.2. Furthermore, the synthetic analyses in sections 4.5.1 and 5.1 demonstrate that the bias in the b-value is largely controlled by the dynamic range of the catalogue magnitudes rather than the true b-value itself.

One of the obvious limitations leading to biased b-values is the required accurate

estimate of m_c. While this is known in the synthetic cases – and can therefore be said to not be a factor in biased, high b-values for synthetic data – this value is unknown in real catalogues. Many methods have been developed, past and present, to determine the most accurate m_c possible (Cao and Gao, 2002; Herrmann and Marzocchi, 2020; Wiemer and Wyss, 2000). In the real catalogues analysed in sections 4.5.2 and 5.2, this estimation of m_c (using BVS) introduces bias to the resulting b-values, especially as most of the BVS m_c estimates obtained overestimate the possibly 'correct' value of m_c (Roberts et al., 2015). However, the fundamental aim of the work performed here was to show the change in b-value bias (and model preference) when that value of m_c is increased, regardless of what its 'correct' value is. I have shown the effect of under- and overestimation of m_c on the estimated b-value, particularly in the case studies of Kilauea (Figure 5.20a), PNR (Figure 5.22a) and Oklahoma (Figure 5.24a). As a consequence, the b-value may be artifically high (BVS) or low (MaxC), due to the sampling and model selection issues highlighted in the analysis of synthetic data in section 5.1, leading respectively to a systematic underestimate or overestimate (Mignan and Woessner, 2012) of the probability of occurrence of events larger than those in the sample. This is important because such extrapolations are usually needed to define probabilities of occurrence for magnitudes that have not yet occurred in a sample, but are possible and to manage regulatory compliance and the risk associated with large events that may not yet have occurred, both in PSHA and in operational earthquake forecasting, including induced seismicity. My results suggest that finding a robust (BVS) m_c is harder for volcanic and induced catalogues than it is for tectonic catalogues, because the event magnitudes are often smaller and obtaining accurate magnitude, and hence m_c, estimates is much harder in small earthquakes, where errors can easily be larger than an order of magnitude in scale. This underpins the requirement of applying one 'best' method to a specific catalogue after investigation of its FMD, rather than applying one general method to all datasets.

Except for The Geysers, no other catalogues used here exhibit MGR preference but in the case of all volcanic catalogues used here, the b-value increases when m_c is

increased (and the data and dynamic range is reduced). Occasionally, at very high m_c, b-values are then extremely low (i.e. in the case of Long Valley, Figure 5.18b) where the dynamic range is too small, resulting in sampling artefacts. In the case of tectonic catalogues, b-values (from lack of data and dynamic range) are either biased to low values (both INGV catalogues in Figure 4.10; Southern California in Figure 5.16b) or to high values (Reykjanes in Figure 5.13b), confirming issues with bias exist in either direction on real data when there is obvious lack of primary data. Strongly elevated b-values are seen in both the Kilauea and PNR catalogues, both of which exhibit truly high (> 1.5) b-values. In order to expand the analysis to a broader set of underlying distributions, I developed and applied an adapted BVS method, based on the original, GR-assumed (Aki, 1965) one, but which assumes an underlying MGR distribution, on all real catalogues assessed in this chapter. In most cases, the values of m_c obtained by both methods are indistinguishable, except for the two induced seismicity catalogues (sections 5.2.3.1 and 5.2.3.2) where visible discrepancies in the resulting m_c estimates are seen.

Throughout Chapter 4, I argued that the most likely reason for the common bias to high b-values is because Aki's (1965) formulation assumes an infinite maximum magnitude but the real data analysed in this thesis (and even all real data) have a finite upper bound. An attempt at this correction has been made in Chapter 6 with equation 6.8 introducing a finite maximum magnitude for the MLE of the GR law and equation 6.11 introducing a finite maximum magnitude for the MLE of the MGR law. For the correction for the GR law, I have shown that the analytical correction presents a fairly good first-order correlation (Figures 6.2b) and 6.3b) to synthetic data previously presented in sections 4.5.1 and 5.1. It also presents a decent first-order estimate to the two real catalogues shown in Figures 6.4b) and 6.5. However, the dominant factor controlling the bias of high b-values is the convergence of the mean magnitude (section 6.1.3) with respect to dynamic range, showing that the mode of the data matches the maximum likelihood expected in a random sample. This also suggests that in samples with a small dynamic range, the expectation value $\langle m \rangle$ is systematically and

significantly underestimated, hence leading to an overestimate of the b-value and an increasing probability to biased, high b-values as the dynamic range reduces.

7.2 How common is MGR?

In reality, and in particular in datasets of volcanic and induced seismicity, it is not common for MGR distributions to be observed within the sampling range (van der Elst et al., 2016), which is also seen throughout this thesis. Of the case studies I have analysed, only once (The Geysers in Chapter 4) is data MGR distributed, adding significance to the results. This case highlights the potential for bias when incorrectly applying a GR distribution to MGR data. The event magnitudes in The Geysers data are fairly large (up to 5.01) for induced seismicity, but the estimated value of M_θ is considered small, which is why the roll-off is observed in this dataset. This implies that there is a fundamental, physical factor which controls this roll-off, limiting the event magnitudes within this setting. These will be limited for example by the finite fault length determined by the volume stimulated by human activities (Shapiro et al., 2013) and conservation of energy. In all other case studies in this thesis, the value of M_θ is high and lies outside the observed range of magnitudes. This was previously shown by van der Elst et al. (2016) who reasoned that the largest magnitudes observed at fluid injection sites are are consistent with the sampling statistics of a GR distribution, assuming an infinite maximum magnitude. Their results suggest that the frequency of earthquakes induced is related to the injection but the magnitude is controlled by the tectonics and therefore finds that MGR is not a commonly observed distribution in (injection) induced seismicity and the seismic hazard associated with these earthquakes should be treated in the same way in which tectonic hazard is approached.

On the other hand, the MGR distribution can be artifically observed in induced seismicity for example due to limits of equipment (for example in the raw PNR data, see section 5.2.3.1) or physical restrictions such as finite fault length. Nonetheless, in a case study of coal mining in the UK, Verdon et al. (2018) found that earthquakes were

directly induced by mining rather than stress transfer on pre-existing faults and that these did not follow a GR but an MGR distribution. The rupture dimensions of the largest magnitude events corresponded to the sizes of the coal seam panels which were being mined and explain the MGR distribution observed. Therefore, although there is much theoretical relevance to the work performed here regarding GR methods on MGR data, I summarise that the MGR distribution is more likely to be seen in global (or at least tectonic) seismicity as opposed to volcanic and induced seismicity. Even so, observing MGR in induced seismicity would ultimately be a positive thing as it implies a fundamental limit to event magnitudes, which would also be smaller than the maximum magnitudes following a GR distribution.

Additionally, the hazard estimate would be somewhat more complex than originally suggested in Chapter 4 where I have stated that if the b-value is high, the large events may not be sampled and the hazard is likely to be underestimated. However, in the rare case of the true, underlying distribution actually being MGR (which I have shown to be difficult to determine), then any seismic hazard assessment which assumes GR distributed data and leading to the artifical b-value overestimate, would actually overestimate the hazard, which would also be concerning.

7.3 Uncertainties, accuracy and precision

Many published studies regarding volcanic and induced data (see Tables 4.1 and 4.2 and references within; Table 1 within Roberts et al. (2015)) have focused exclusively on time spans limited to episodes of volcanic unrest or short-lived induced seismicity sequences. This introduces substantial sampling bias as time windows result in lower numbers of events, for reasons given in section 7.1, in particular a preference for higher inferred b-values, and hence an underestimate of seismic hazard from rare large events which may not yet have occured in the sample.

In the synthetic results of Figure 4.3, I have shown how important it is to estimate uncertainties of the b-value in identifying whether b-values are truly high or only

apparently high as a result of sample bias. This has been supported further by the analysis of the BVS stability curve errors in section 5.2, and the 95% confidence intervals shown in Figures 6.2 through to 6.5. It is clear that increased dynamic range in real catalogues (low m_c, but not too low!) will substantially reduce the bias of the b-value in all situations. From the results preferring a GR distribution, I suggest an absolute minimum dynamic range of 2 is needed to constrain b-values to an accuracy of ± 0.1 when $b \approx 1.0$ and a minimum dynamic range of 3 to confirm that b-values are truly high to the same accuracy, i.e. > 1.4. In the case of an underlying MGR distribution, there will always be less dynamic range available than for the equivalent GR case for the same number of events, due to the roll-off initiated at m_θ. In the one real case examined here which prefers MGR (The Geysers), a dynamic range of 3.5 is required to return the correct model preference and an unbiased b-value of ~ 1.0. The synthetic results suggest that a lower dynamic range (~ 2 from Figure 5.8 for $b = 1.0 \pm 0.1$) may be sufficient in other cases with an underlying MGR distribution. Section 5.1 has shown that in the case of an MGR distribution, the required dynamic range can change substantially depending on the underlying b-value and m_{min} (m_c in real data) as well as the estimate of m_θ. Nonetheless, in all cases, I support the need for at least N > 1000 events (Nava et al. (2017), see also Figure 4.3) and have shown that both the accuracy and precision will improve significantly with even more data. In the specific case that GR methods are applied to underlying MGR distributed data (Figure 4.3, second row), this minimum requirement extends to N > 10 000. More data is likely to return a larger dynamic range and vice versa and this should always be checked before analysis is begun. Decisions made on the exact data required for a specific catalogue needs to be amended on a case by case basis. Results obtained for b-values from estimations with less data or dynamic range than suggested here require extremely cautious discussion of significance or lack thereof. This applies both to any inferences made for the accuracy and precision of the inferred b-value and likely known-on effects to the calculation of seismic hazard through quantification of occurrence rates.

The epistemic uncertainty in the b-value arising from the incorrect assumption of

the underlying distribution cannot be captured solely by the statistical uncertainty expressed in equations 4.6 (Aki, 1965; Kagan, 2002a) or 4.7 (Kagan, 2002a). Additionally, as catalogues are thinned by increasing m_c, there is strong variability in the magnitudes in the tail of the distribution, likely to be due to random effects of the statistical sampling. However, these could also be due to increased complexity in the underlying distribution in cases when this is more complex than a GR or MGR distribution. Many FMDs follow a mixture of distributions or other distributions (section 2.4), but there has not been time to examine these cases here. The MLE correction for the MGR law (equation 6.11) has shown such a problem in its increased complexity which cannot easily be dealt with with the datasets we commonly have available (Zöller, 2013). In Table 6.1 I summarised the necessary dynamic ranges required for data (synthetic and real, where for real data we can only estimate the precision) to be both accurate and precise within the 95% confidence intervals portrayed. These suggest that in the case of the b-value for GR data, dynamic ranges of 3 or larger are required to obtain accurate estimates of the b-value (and even more to obtain accuracy *and* precision) and a similar dynamic range is required to obtain convergence of the mean magnitude. In the case of MGR this is less due to limited existing dynamic range.

7.4 Further limitations

In addition to the limitations that come with estimating m_c, b and also m_θ, there are further factors that may affect the results shown in this thesis. One of these is the lack of declustering of catalogues. No declustering was performed as the aim was to use as much earthquake data as available, disregarding whether these events were categorised as fore-, main- or aftershocks and in which inter-event times they occurred. However, as seen in section 4.4.3, bias of b-values to high values in declustered catalogues may also be an artefact of declustering algorithms (Mizrahi et al., 2021), in addition to the random sampling issues analysed here. Main (1995) further suggests that it would be dangerous to decluster catalogues as the after- and foreshocks are the strongest

evidence for criticality as the underlying statistical physical process. Furthermore, there has not been time to analyse b-values spatio-temporally within the real catalogues used.

From the discussion in sections 7.1 and 7.3, I also note that high b-values (unless obtained with the prerequisites stated in section 7.3 above) are in general not representative of those which ideally should be used in assessing long-term hazard in a given region. Instead, they are likely to underestimate the hazard from larger events (which would return smaller b-values) which lie outside the range observed in the finite sample window.

Access to (sufficient) dynamic range in real data has been a major limitation in the work performed in this thesis and is likely to continue in future work. Therefore, the results produced in this thesis advocate the improvement for event detection and even more importantly, the accessibility of data, regardless of type of seismicity or location. This is particularly important in areas of high hazard such as in volcanic settings or in areas of induced seismicity which are located near populations, requiring hazard and risk assessments which in turn require a good, holistic understanding of the hazard within that specific region. This is difficult to obtain with lack of available data. Nonetheless, with improved recording, including new-generation high-resolution catalogues where events are detected and located from whole waveform analysis (Spallarossa et al., 2021), there remains a trade-off between how much data can be used and how much data is required for best results. For large catalogues, it is difficult to choose one value of m_c when recording sensitivity will have improved over time – where do we make this cut and split the catalogue (i.e. section 4.4.3.1) or do we opt for an overall higher value of m_c to include the older, less sensitive period together with the more recent one and aim for a more holistic image of overall seismicity, trading in accuracy? These issues have not been resolved here, and are likely to continue to require formal estimates of their effects in the estimation of epistemic uncertainties in PSHA.

A further limitation lies within the methods used throughout this thesis. Not only do FMDs from real datasets have many inconsistencies, they often deviate from

exponentiality (the GR distribution at all times and the MGR distritbutions up until the roll-off is introduced at a certain corner magnitude m_θ) which is one of the main assumptions on which this work is based. The assumption of an infinite maximum magnitude in the methods applied to the data, is another limitation which will introduce bias to parameter estimation. I have shown that this can be mitigated to some extent in Chapter 6, by using corrections based on the analytical solutions for a finite maximum magnitude in an infinite sample, and the associated bias in the estimates of the mean magnitude associated with small samples.

7.5 Suggestions for future work

From the discussions above, I now suggest multiple areas of future work which would enhance the results shown throughout this thesis.

Firstly, widening the Central Italy catalogue used in section 4.4.3.1 to include the Central Italy earthquake sequence from 2016 and event data from the last 5 years would allow a direct comparison on how the data and results have changed compared to the catalogue used in this thesis, which ends in August 2016. None of the catalogues analysed in this thesis, except for The Geysers, have been shown to be MGR distributed. However, many more catalogues could be investigated to test this hypothesis and I would suggest looking at more induced seismicity catalogues, especially geothermal ones. There may be a connection between The Geysers and other geothermal catalogues exhibiting MGR distributed event data and it would be interesting to follow up on this. Even then, this may be due to it having substantial data available above the (minimum) BVS m_c – over 60 000 events, whereas most other catalogues have far fewer than 10 000 events (Table 3.1), suggesting that the common assumption (Nava et al., 2017) of requiring N = 1000 to calculate a robust b-value, may not always be good guidance, as many more events are needed in some cases, particularly when investigating non-tectonic catalogues.

To test the model preference for a given catalogue, it would be worth applying

different model selection criteria (i.e. AIC, Bayesian evidence and the associated Bayes factor) to see how this affects the results discussed in this thesis and whether this would change the model preference for some of the catalogues used here, which are close to the border between GR and MGR. Also, declustering the catalogues would provide an interesting comparison to the results here, although this may mean obvious lack of data (N) for many of the catalogues used here and therefore introduce its own problems of data sampling. It would further be possible to extend the methods applied here to other distributions commonly observed in earthquake data, such as the characteristic or gamma distributions or even a mixture of distributions.

Personally, I would be interested in investigating why the modified method to return a BVS estimate assuming an underlying MGR distribution (3.2.1 and first applied in section 5.2.1.1) returns the same values of BVS m_c as the GR assuming one, except in the case of Oklahoma in Figure 5.23b) and whether this is a real effect of the adaptation of the method or a further artefact due to (lack of) data. As additional work for Chapter 6, I would suggest looking at further real catalogues as well as creating synthetic data with parameters of the real catalogues to compare these with each other. Additionally, testing the equations provided as corrections for biased b-values in practice would be beneficial in determining their robustness and possibility for use in future b-value studies.

All of the above suggestions could complement the data analysed in this thesis and strengthen the results, or even identify further issues which I may have not yet come across while performing this work. Either way, they are likely to strengthen our general understanding of b-value estimates in the context of hazard analysis and could help suggest improved steps for such, particularly in areas of volcanic and induced seismicity.

Chapter 8

Conclusions

The process of estimating parameters in the frequency-magnitude distribution for earthquakes requires robust methods to return accurate, and precise, values. This is particularly important for the b-value which quantifies the proportion of large to small events in a given catalogue in the Gutenberg-Richter (GR) law or in its modified form (MGR). In this thesis, I have focused on estimating this b-value with regards to (lack of) sufficient data and a possible misinterpretation of the underlying distribution for a given catalogue as sources of epistemic uncertainty.

The results obtained from the work in this thesis can be summarised as follows:

- the elevated b-values reported in many studies of volcanic and induced seismicity are likely to be biased to high b-values and hence not truly as high as claimed. This bias is introduced due to the narrow dynamic range of data often available for such analysis
- the source of the bias can best be explained by an analytical theory for the most likely mean magnitude to be sampled in a finite data set of independent elements, drawn from an identically underlying distribution, resulting in systematic convergence of mean magnitude to an asymptotic limit from below in random samples
- the convergence of the mean magnitude propagates into convergence of the

b-value from above to an asymptotic minimum. The associated convergence in b-value is generally achieved only with a relatively large dynamic range (typically > 3.5) and hence substantially more data is required to distinguish a truly high b-value from a biased high b-value

- all b-value studies should consistently give clear details on sample sizes used, the dynamic range above the magnitude of completeness m_c and the method and justification for m_c. Additionally, assumptions regarding the underlying distribution and parameter estimation methods (i.e. maximum likelihood estimation) should also be clearly stated

- as a consequence of incorrect m_c estimation, artificially high or low b-values can easily be returned due to sampling and model selection issues

- existing methods should be adapted for a particular distribution, i.e. using GR methods (without adapting) on MGR distributed data will introduce bias which can be easily avoided by modifying the methods applied to assume the correct distribution (if the distribution is known). When the distribution is not known, either a) model selection criteria should be used or b) both models should be considered, and the results propagated into the likelihood of future earthquakes as a source of epistemic uncertainty

- the number of events in a catalogue (above m_c) is very important but I have shown that the dynamic range is also critical

- any bias in the b-value to high values will likely to result in underestimation of the hazard from rare, large events, particularly for sample sizes with a narrow dynamic range typical of volcanic or induced seismicity settings

- while the systematic error in b is high at low magnitudes, the random error in b to 95% confidence provides a reasonable estimate on the precision (albeit with the wrong b-value), but at large dynamic ranges the random error in b significantly underestimates the precision obtained from multiple random samples

- analysis of the random samples show that the b-value is neither accurate nor precise at small dynamic range, accurate but not precise at intermediate dynamic range and accurate and precise only at large dynamic range.

In summary, I have demonstrated that it is prudent to adopt a cautious interpretation of b-values and their importance and significance in seismic hazard analysis. There is no 'one size fits all' answer to how many events or dynamic range is required in specific catalogues for an accurate or precise estimate of b. Nonetheless, the b-value estimates become more reliable and precise as the sample size and its dynamic range increases. Finally, I note that these conclusions are not necessarily restricted to the applications in earthquake hazard. A range of natural hazards exhibit power-law frequency-size distributions (avalanches or volcanic eruptions by volume, forest fires by area, floods by area, height or volume, hurricanes, by area or severity etc), resulting in an exponential frequency-magnitude (i.e. some logarithmic measure of size) relation. Hence, the same issues of convergence, accuracy and precision in finite samples will apply.


"Finishing a PhD is like finishing
a group project where your partner
made a ton of mistakes
at the beginning of
the assignment.
Except your partner is just
you 4 years ago."
John M. Mola
I don't get it..

Bibliography

Akaike, H. (1974). A New Look at the Statistical Model Identification. *IEEE Transactions on Automatic Control*, 19(6):716–723.

Aki, K. (1965). Maximum Likelihood Estimate of b in the Formula logN=a-bM and its Confidence Limits. *Bulletin of the Earthquake Research Institute*, 43:237–239.

Akkar, S., Sandıkkaya, M. A., and Bommer, J. J. (2014). Empirical ground-motion models for point- and extended-source crustal earthquake scenarios in Europe and the Middle East. *Bulletin of Earthquake Engineering*, 12(1):359–387.

Amorèse, D., Grasso, J. R., and Rydelek, P. A. (2010). On varying b-values with depth: Results from computer-intensive tests for Southern California. *Geophysical Journal International*, 180:347–360.

Bachmann, C. E., Wiemer, S., Goertz-Allmann, B. P., and Woessner, J. (2012). Influence of pore-pressure on the event-size distribution of induced earthquakes. *Geophysical Research Letters*, 39(L09302):1–7.

Bachmann, C. E., Wiemer, S., Woessner, J., and Hainzl, S. (2011). Statistical analysis of the induced Basel 2006 earthquake sequence: introducing a probability-based monitoring approach for Enhanced Geothermal Systems. *Geophysical Journal International*, 186(2):793–807.

Baisch, S., Koch, C., and Muntendam-Bos, A. (2019). Traffic light systems: To what extent can induced seismicity be controlled? *Seismological Research Letters*, 90(3):1145–1154.

Baker, J. W. (2013). Probabilistic Seismic Hazard Analysis. Technical report.

Bao, X. and Eaton, D. W. (2016). Fault activation by hydraulic fracturing in western Canada. *Science*, 354(6318):1406–1409.

Baptie, B., Luckett, R., Butcher, A., and Werner, M. J. (2020). Robust relationships for magnitude conversion of PNR seismicity catalogues. Technical report, Multihazards and Risk Programme Open Report OR/20/042.

Barton, D. J., Foulger, G. R., Henderson, J. R., and Julian, B. R. (1999). Frequency-magnitude statistics and spatial correlation dimensions of earthquakes at Long Valley caldera, California. *Geophysical Journal International*, 138(2):563–570.

Bell, A. F., Naylor, M., and Main, I. G. (2013). Convergence of the frequency-size distribution of global earthquakes. *Geophysical Research Letters*, 40:2585–2589.

Bird, P. and Kagan, Y. Y. (2004). Plate-tectonic analysis of shallow seismicity: Apparent boundary width, beta, corner magnitude, coupled lithosphere thickness, and coupling in seven tectonic settings. *Bulletin of the Seismological Society of America*, 94(6):2380–2399.

Boettcher, M. S., McGarr, A., and Johnston, M. (2009). Extension of Gutenberg-Richter distribution to Mw -1.3, no lower limit in sight. *Geophysical Research Letters*, 36(L10307):1–5.

Bourne, S. J., Oates, S. J., and van Elk, J. (2018). The exponential rise of induced seismicity with increasing stress levels in the Groningen gas field and its implications for controlling seismic risk. *Geophysical Journal International*, 213(3):1693–1700.

Burridge, R. and Knopoff, L. (1967). Model and theoretical seismicity. *Bulletin of the Seismological Society of America*, 57(3):341–371.

Candela, T., Osinga, S., Ampuero, J., Wassing, B., Pluymaekers, M., Fokker, P. A., Wees, J., Waal, H. A., and Muntendam-Bos, A. G. (2019). Depletion-Induced Seismicity at the Groningen Gas Field: Coulomb Rate-and-State Models Including Differential Compaction Effect. *Journal of Geophysical Research: Solid Earth*, 124(7):7081–7104.

Cao, A. and Gao, S. S. (2002). Temporal variation of seismic b -values beneath northeastern Japan island arc. *Geophysical Research Letters*, 29(9):48–1.

Chen, K., Avouac, J. P., Aati, S., Milliner, C., Zheng, F., and Shi, C. (2020). Cascading and pulse-like ruptures during the 2019 Ridgecrest earthquakes in the Eastern California Shear Zone. *Nature Communications*, 11(22):1–8.

Chen, X., Nakata, N., Pennington, C., Haffener, J., Chang, J. C., He, X., Zhan, Z., Ni, S., and Walter, J. I. (2017). The Pawnee earthquake as a result of the interplay among injection, faults and foreshocks. *Scientific Reports*, 7(4945):1–18.

Clarke, H., Eisner, L., Styles, P., and Turner, P. (2014). Felt seismicity associated with shale gas hydraulic fracturing: The first documented example in Europe. *Geophysical Research Letters*, 41(23):8308–8314.

Clarke, H., Verdon, J. P., Kettlety, T., Baird, A. F., and Kendall, J. (2019). Real-Time Imaging, Forecasting, and Management of Human-Induced Seismicity at Preston New Road, Lancashire, England. *Seismological Research Letters*, 90(5):1902–1915.

Clauset, A., Shalizi, C. R., and Newman, M. E. (2009). Power-law distributions in empirical data. *Society for Industrial and Applied Mathematics Review*, 51(4):661–703.

Colombo, I., Main, I., and Forde, M. (2003). Assessing damage of reinforced concrete beam using "b-value" analysis of acoustic emission signals. *Journal of Materials in Civil Engineering*, 15(3):280–286.

Cornell, C. A. (1968). Engineering Seismic Risk Analysis. *Bulletin of the Seismological Society of America*, 58(5):1583–1606.

Corral, A. (2004). Long-Term Clustering, Scaling, and Universality in the Temporal Occurrence of Earthquakes. *Physical Review Letters*, 92(10):1–4.

Eaton, D. W., Davidsen, J., Pedersen, P. K., and Boroumand, N. (2014). Breakdown of the Gutenberg-Richter relation for microearthquakes induced by hydraulic fracturing: Influence of stratabound fractures. *Geophysical Prospecting*, 62(4):806–818.

Ekström, G., Nettles, M., and Dziewoński, A. M. (2012). The global CMT project 2004-2010: Centroid-moment tensors for 13,017 earthquakes. *Physics of the Earth and Planetary Interiors*, 200-201:1–9.

El-Isa, Z. H. and Eaton, D. W. (2014). Spatiotemporal variations in the b-value of earthquake magnitude-frequency distributions: Classification and causes. *Tectonophysics*, 615-616:1–11.

Ellsworth, W. L. (2013). Injection-Induced Earthquakes. *Science*, 341(6142):1–7.

Ellsworth, W. L., Giardini, D., Townend, J., Ge, S., and Shimamoto, T. (2019). Triggering of the Pohang, Korea, Earthquake (Mw 5.5) by Enhanced Geothermal System Stimulation. *Seismological Research Letters*, 90(5):1844–1858.

England, P. and Jackson, J. (2011). Uncharted seismic risk. *Nature Geoscience*, 4(6):348–349.

Felzer, K. R., Becker, T. W., Abercrombie, R. E., Ekström, G., and Rice, J. R. (2002). Triggering of the 1999 M W 7.1 Hector Mine earthquake by aftershocks of the 1992 M W 7.3 Landers earthquake . *Journal of Geophysical Research: Solid Earth*, 107(B9):6–1.

Field, E. H., Arrowsmith, R. J., Biasi, G. P., Bird, P., Dawson, T. E., Felzer, K. R., Jackson, D. D., Johnson, K. M., Jordan, T. H., Madden, C., Michael, A. J., Milner, K. R., Page, M. T., Parsons, T., Powers, P. M., Shaw, B. E., Thatcher, W. R., Weldon, R. J., and Zeng, Y. (2014). Uniform California Earthquake Rupture Forecast, version 3 (UCERF3) -The time-independent model. *Bulletin of the Seismological Society of America*, 104(3):1122–1180.

Field, E. H., Jordan, T. H., Jones, L. M., Michael, A. J., Blanpied, M. L., and Participants, O. W. (2016). The Potential Uses of Operational Earthquake Forecasting. *Seismological Research Letters*, 87(2A):313–322.

Fossen, H. (2010). *Structural Geology*. Cambridge University Press, Cambridge.

Foulger, G. R., Wilson, M. P., Gluyas, J. G., Julian, B. R., and Davies, R. J. (2018). Global review of human-induced earthquakes. *Earth-Science Reviews*, 178:438–514.

Frohlich, C. and Davis, S. D. (1993). Teleseismic b Values; Or, Much Ado About 1.0. *Journal of Geophysical Research*, 98(B1):631–644.

García-Hernández, R., D'Auria, L., Barrancos, J., Padilla, G. D., and Pérez, N. M. (2021). Multiscale Temporal and Spatial Estimation of the b-Value. *Seismological Research Letters*, 92(6):3712–3724.

Gasperini, P., Lolli, B., and Vannucci, G. (2013). Empirical calibration of local magnitude data sets versus moment magnitude in Italy. *Bulletin of the Seismological Society of America*, 103(4):2227–2246.

Geffers, G.-M., Main, I. G., and Naylor, M. (2022). Biases in estimating b-values from small earthquake catalogues: How high are high b-values? *Geophysical Journal International*, 229(3):1840–1855.

Godano, C., Lippiello, E., and De Arcangelis, L. (2014). Variability of the b value in the Gutenberg-Richter distribution. *Geophysical Journal International*, 199:1765–1771.

Godano, C. and Pingue, F. (2000). Is the seismic moment-frequency relation universal? *Geophysical Journal International*, 142:193–198.

Goertz-Allmann, B. P., Edwards, B., Bethmann, F., Deichmann, N., Clinton, J., Fäh, D., and Giardini, D. (2011). A new empirical magnitude scaling relation for Switzerland. *Bulletin of the Seismological Society of America*, 101(6):3088–3095.

González, (2017). The Spanish National Earthquake Catalogue: Evolution, precision and completeness. *Journal of Seismology*, 21(3):435–471.

Greenhough, J. and Main, I. G. (2008). A poisson model for earthquake frequency uncertainties in seismic hazard analysis. *Geophysical Research Letters*, 35(L19313):1–4.

Grigoli, F., Cesca, S., Rinaldi, A. P., Manconi, A., López-Comino, J. A., Clinton, J. F., Westaway, R., Cauzzi, C., Dahm, T., and Wiemer, S. (2018). The November 2017 Mw5.5 Pohang earthquake: A possible case of induced seismicity in South Korea. *Science*, 360(6392).

Grünthal, G. (2014). Induced seismicity related to geothermal projects versus natural tectonic earthquakes and other types of induced seismic events in Central Europe. *Geothermics*, 52:22–35.

Gutenberg, B. and Richter, C. F. (1944). Frequency of Earthquakes in California. *Bulletin of the Seismological Society of America*, 34(4):185–188.

Gutenberg, B. and Richter, C. F. (1956). Magnitude and Energy of Earthquakes. *Annali di Geofisica*, 9:1–15.

Hanks, T. C. and Kanamori, H. (1979). A Moment Magnitude Scale. *Journal of Geophysical Research*, 84(B5):2348–2350.

Heaton, T. H., Tajima, F., and Mori, A. W. (1986). Estimating Ground Motions using Recorded Accelerograms. *Surveys in Geophysics*, 8:25–83.

Henderson, J. R., Barton, D. J., and Foulger, G. R. (1999). Fractal clustering of induced seismicity in The Geysers geothermal area, California. *Geophysical Journal International*, 139:317–324.

Herrmann, M. and Marzocchi, W. (2020). Inconsistencies and Lurking Pitfalls in the Magnitude–Frequency Distribution of High-Resolution Earthquake Catalogs. *Seismological Research Letters*, 92(2A):909–922.

Hill, D. P. (2006). Unrest in Long Valley Caldera, California, 1978-2004. In Troise, C., De Natale, G., and Kilburn, C. R. J., editors, *Mechanisms of Activity and Unrest at Large Calderas*, volume 269, pages 1–24. Geological Society, London, special publications edition.

Holschneider, M., Zöller, G., Clements, R., and Schorlemmer, D. (2014). Can we test for the maximum possible earthquake magnitude? *Journal of Geophysical Research: Solid Earth*, 119:2019–2028.

Hough, S. E. and Page, M. (2015). A century of induced earthquakes in Oklahoma? *Bulletin of the Seismological Society of America*, 105(6):2863–2870.

Howell, B. F. (1981). On the Saturation of Earthquake Magnitudes. *Bulletin of the Seismological Society of America*, 71(5):1401–1422.

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.

Ibáñez, J. M., De Angelis, S., Díaz-Moreno, A., Hernández, P., Alguacil, G., Posadas, A., and Pérez, N. (2012). Insights into the 2011-2012 submarine eruption off the coast of El Hierro (Canary Islands, Spain) from statistical analyses of earthquake activity. *Geophysical Journal International*, 191:659–670.

Igonin, N., Zecevic, M., and Eaton, D. W. (2018). Bilinear Magnitude-Frequency Distributions and Characteristic Earthquakes During Hydraulic Fracturing. *Geophysical Research Letters*, 45(23):866–12.

Jackson, D. D. and Kagan, Y. Y. (1999). Testable Earthquake Forecasts for 1999. *Seismological Research Letters*, 70(4):393–403.

Jacobs, K. M. and McNutt, S. R. (2010). Using Seismic b-Values to Interpret Seismicity Rates and Physical Processes During the Preeruptive Earthquake Swarm at Augustine Volcano 2005–2006. *USGS Professional Paper*, 1769(3):59–83.

Jolly, A. D. and McNutt, S. R. (1999). Seismicity at the volcanoes of Katmai National Park, Alaska; July 1995-December 1997. *Journal of Volcanology and Geothermal Research*, 93:173–190.

Jónasson, K., Bessason, B., Helgadóttir, , Einarsson, P., Guðmundsson, G. B., Brandsdóttir, B., Vogfjörd, K. S., and Jónsdóttir, K. (2021). A Harmonised Instrumental Earthquake Catalogue for Iceland and the Northern Mid-Atlantic Ridge. In *Natural Hazards and Earth System Sciences Discussions*, pages 1–26.

Jordan, T. H., Chen, Y. T., Gasparini, P., Madariaga, R., Main, I., Marzocchi, W., Papadopoulos, G., Sobolev, G., Yamaoka, K., and Zschau, J. (2011). Operational

earthquake forecasting: State of knowledge and guidelines for utilization. *Annals of Geophysics*, 54(4):316–391.

Kagan, Y. Y. (1991). Seismic moment distribution. *Geophysical Journal International*, 106:123–134.

Kagan, Y. Y. (1997). Seismic moment-frequency relation for shallow earthquakes: Regional comparison. *Journal of Geophysical Research*, 102(B2):2835–2852.

Kagan, Y. Y. (1999). Universality of the Seismic Moment-frequency Relation. *Pure Applied Geophysics*, 155:537–573.

Kagan, Y. Y. (2002a). Seismic moment distribution revisited: I. Statistical results. *Geophysical Journal International*, 148:520–541.

Kagan, Y. Y. (2002b). Seismic moment distribution revisited: II. Moment conservation principle. *Geophysical Journal International*, 149:731–754.

Kagan, Y. Y. and Jackson, D. D. (1991). Seismic gap hypothesis: Ten years after. *Journal of Geophysical Research*, 96(B13):419–421.

Kagan, Y. Y. and Jackson, D. D. (2000). Probabilistic forecasting of earthquakes. *Geophysical Journal International*, 143:438–453.

Kanamori, H. and Anderson, D. L. (1975). Theoretical basis of some empirical relations in seismology. *Bulletin of the Seismological Society of America*, 65(5):1073–1095.

Kendall, J.-M., Butcher, A., Stork, A. L., Verdon, J. P., Luckett, R., and Baptie, B. J. (2019). How big is a small earthquake? Challenges in determining microseismic magnitudes. *First Break*, 37:51–56.

Keranen, K. M., Savage, H. M., Abers, G. A., and Cochran, E. S. (2013). Potentially induced earthquakes in Oklahoma, USA: Links between wastewater injection and the 2011 Mw 5.7 earthquake sequence. *Geology*, 41(6):699–702.

Keranen, K. M., Weingarten, M., Abers, G. A., Bekins, B. A., and Ge, S. (2014). Sharp increase in central Oklahoma seismicity since 2008 induced by massive wastewater injection. *Science*, 345(6195):448–451.

Kettlety, T., Verdon, J. P., Butcher, A., Hampson, M., and Craddock, L. (2021). High-Resolution Imaging of the ML 2.9 August 2019 Earthquake in Lancashire, United Kingdom, Induced by Hydraulic Fracturing during Preston New Road PNR-2 Operations. *Seismological Research Letters*, 92(1):151–169.

Kijko, A. (2004). Estimation of the maximum earthquake magnitude, mmax. *Pure and Applied Geophysics*, 161:1655–1681.

Kijko, A. and Smit, A. (2012). Extension of the Aki-Utsu b-Value estimator for incomplete catalogs. *Bulletin of the Seismological Society of America*, 102(3):1283–1287.

Koshimura, S., Hayashi, S., and Gokon, H. (2014). The impact of the 2011 Tohoku earthquake tsunami disaster and implications to the reconstruction. *Soils and Foundations*, 54(4):560–572.

Langenbruch, C. and Zoback, M. D. (2016). How will induced seismicity in Oklahoma respond to decreased saltwater injection rates? *Science Advances*, 2(e1601542):1–9.

Leonard, T. and Hsu, J. S. J. (1999). *Bayesian Methods*. Cambridge University Press, Cambride, UK.

Leptokaropoulos, K., Staszek, M., Lasocki, S., Martínez-Garzón, P., and Kwiatek, G. (2018a). Evolution of seismicity in relation to fluid injection in the North-Western part of The Geysers geothermal field. *Geophysical Journal International*, 212(2):1157–1166.

Leptokaropoulos, K. M., Adamaki, A. K., Roberts, R. G., Gkarlaouni, C. G., and Paradisopoulou, P. M. (2018b). Impact of magnitude uncertainties on seismic catalogue properties. *Geophysical Journal International*, 213(2):940–951.

Lewis, F., Butler, A., and Gilbert, L. (2011). A unified approach to model selection using the likelihood ratio test. *Methods in Ecology and Evolution*, 2(2):155–162.

Llenos, A. L., Michael, A. J., Moschetti, M. P., Powers, P. M., McNamara, D. E., and Petersen, M. D. (2021). Assessing the Hazard of Highly Clustered M˜5 Caldera Collapse Earthquakes in Kīlauea Volcano, Hawai'i. In *Poster*. Seismological Society of America.

Main, I. (1995). Earthquakes as critical phenomena: implications for probabilistic seismic hazard analysis. *Bulletin of the Seismological Society of America*, 85(5):1299–1308.

Main, I. (1996). Statistical physics, seismogenesis, and seismic hazard. *Reviews of Geophysics*, 34(4):433.

Main, I. (2000). Apparent breaks in scaling in the earthquake cumulative frequency-magnitude distribution: Fact or artifact? *Bulletin of the Seismological Society of America*, 90(1):86–97.

Main, I. and Burton, P. (1984). Information theory and the earthquake frequency-magnitude distribution. *Bulletin of the Seismological Society of America*, 74(4):1409–1426.

Main, I. G. (1987). A characteristic earthquake model of the seismicity preceding the eruption of Mount St. Helens on 18 May 1980. *Physics of the Earth and Planetary Interiors*, 49:283–293.

Main, I. G., Leonard, T., Papasouliotis, O., Hatton, C. G., and Meredith, P. G. (1999). One slope or two? Detecting statistically significant breaks of slope in geophysical data, with application to fracture scaling relationships. *Geophysical Research Letters*, 26(18):2801–2804.

Main, I. G., Meredith, P. G., and Jones, C. (1989). A reinterpretation of the precursory seismic b-value anomaly from fracture mechanics. *Geophysical Journal International*, 96(1):131–138.

Main, I. G., Naylor, M., Greenhough, J., Touati, S., Bell, A. F., and McCloskey, J. (2011). Model selection and uncertainty in earthquake hazard analysis. In *Applications of Statistics and Probability in Civil Engineering*, pages 735–743.

Majer, E. L., Baria, R., Stark, M., Oates, S., Bommer, J., Smith, B., and Asanuma, H. (2007). Induced seismicity associated with Enhanced Geothermal Systems. *Geothermics*, 36(3):185–222.

Mancini, S., Segou, M., Werner, M. J., and Baptie, B. J. (2019). Statistical Modelling of the Preston New Road Seismicity: Towards Probabilistic Forecasting Tools. Technical report, British Geological Survey Commissioned Report, CR/19/068.

Mancini, S., Werner, M., Baptie, B., and Segou, M. (2020). Statistical Modelling and Forecasting of the Preston New Road Seismicity. Technical report, British Geological Survey Commissioned Report, CR/20/032.

Marzocchi, W. and Sandri, L. (2003). A review and new insights on the estimation of the b-value and its uncertainty. *Annals of Geophysics*, 46(6):1271–1282.

Marzocchi, W., Spassiani, I., Stallone, A., and Taroni, M. (2020). How to be fooled searching for significant variations of the b-value. *Geophysical Journal International*, 220(3):1845–1856.

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.

Maxwell, S., Jones, M., Parker, R., Miong, S., Leaney, S., Dorval, D., Logel, J., Anderson, E., and Hammermaster, K. (2009). Fault Activation During Hydraulic Fracturing. In *SEG Technical Program Expanded Abstracts*, pages 1552–1556. Society of Exploration Geophysicists.

McClure, M., Gibson, R., Chiu, K. K., and Ranganath, R. (2017). Identifying potentially induced seismicity and assessing statistical significance in Oklahoma and California. *Journal of Geophysical Research: Solid Earth*, 122(3):2153–2172.

McGarr, A. and Barbour, A. J. (2017). Wastewater Disposal and the Earthquake Sequences During 2016 Near Fairview, Pawnee, and Cushing, Oklahoma. *Geophysical Research Letters*, 44(18):9330–9336.

McGarr, A., Bekins, B., Burkardt, N., Dewey, J., Earle, P., Ellsworth, W., Ge, S., Hickman, S., Holland, A., Majer, E., Rubinstein, J., and Sheehan, A. (2015). Coping with earthquakes induced by fluid injection. *Science*, 347(6224):830–831.

Meng, F., Wong, L. N. Y., and Zhou, H. (2019). Power law relations in earthquakes from microscopic to macroscopic scales. *Scientific Reports*, 9(10705):1–11.

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 Woessner, J. (2012). Estimating the magnitude of completeness for earthquake catalogs. *Community Online Resource for Statistical Seismicity Analysis*, pages 1–45.

Mizrahi, L., Nandan, S., and Wiemer, S. (2021). The Effect of Declustering on the Size Distribution of Mainshocks. *Seismological Research Letters*, 92(4):2333–2342.

Mogi, K. (1962). Study of the elastic shocks caused by the fracture of heterogeneous materials and its relation to earthquake phenomena. *Bulletin of the Earthquake Research Institute*, 40:125–173.

Mousavi, S. M., Ogwari, P. O., Horton, S. P., and Langston, C. A. (2017). Spatio-temporal evolution of frequency-magnitude distribution and seismogenic index during initiation of induced seismicity at Guy-Greenbrier, Arkansas. *Physics of the Earth and Planetary Interiors*, 267:53–66.

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.

Murru, M., Montuori, C., Console, R., and Lisi, A. (2005). Mapping of the b value anomalies beneath Mt. Etna, Italy, during July-August 2001 lateral eruption. *Geophysical Research Letters*, 32(L05309):1–4.

Murru, M., Montuori, C., Wyss, M., and Privitera, E. (1999). The locations of magma chambers at Mt. Etna, Italy, mapped by b-values. *Geophysical Research Letters*, 26(16):2553–2556.

Musson, R. M. (2012). The effect of magnitude uncertainty on earthquake activity rates. *Bulletin of the Seismological Society of America*, 102(6):2771–2775.

Nava, F. A., Márquez-Ramírez, V. H., Zúñiga, F. R., Ávila-Barrientos, L., and Quinteros, C. B. (2017). Gutenberg-Richter b-value maximum likelihood estimation and sample size. *Journal of Seismology*, 21:127–135.

Naylor, M., Orfanogiannaki, K., and Harte, D. (2010). Exploratory Data Analysis: Magnitude, Space, and Time. *Community Online Resource for Statistical Seismicity Analysis*, pages 1–42.

Novelo-Casanova, D. A., Martínez-Bringas, A., and Valdés-González, C. (2006). Temporal variations of Qc- 1and b-values associated to the December 2000-January 2001 volcanic activity at the Popocatepetl volcano, Mexico. *Journal of Volcanology and Geothermal Research*, 152:347–358.

Nuannin, P., Kulhánek, O., and Persson, L. (2012). Variations of b-values preceding large earthquakes in the Andaman-Sumatra subduction zone. *Journal of Asian Earth Sciences*, 61:237–242.

Ogata, Y., Matsu'ura, R. S., and Katsura, K. (1993). Fast likelihood computation of epidemic type aftershock-sequence model. *Geophysical Research Letters*, 20(19):2143–2146.

Ogata, Y. and Yamashina, K. (1986). Unbiased estimate for b-value of magnitude frequency. *Journal of Physics of the Earth*, 34:187–194.

Oppenheimer, D. H. (1986). Extensional Tectonics at The Geysers Geothermal Area, California. *Journal of Geophysical Research*, 91(B11):463–474.

Panzera, F., Mignan, A., and Vogfjörð, K. S. (2017). Spatiotemporal evolution of the completeness magnitude of the Icelandic earthquake catalogue from 1991 to 2013. *Journal of Seismology*, 21(4):615–630.

Parsons, T. (2008). Monte Carlo method for determining earthquake recurrence parameters from short paleoseismic catalogs: Example calculations for California. *Journal of Geophysical Research: Solid Earth*, 113(B03302):1–14.

Petersen, M. D., Shumway, A. M., Powers, P. M., Mueller, C. S., Moschetti, M. P., Frankel, A. D., Rezaeian, S., McNamara, D. E., Luco, N., Boyd, O. S., Rukstales, K. S., Jaiswal, K. S., Thompson, E. M., Hoover, S. M., Clayton, B. S., Field, E. H., and Zeng, Y. (2020). The 2018 update of the US National Seismic Hazard Model: Overview of model and implications. *Earthquake Spectra*, 36(1):5–41.

Pisarenko, V. F. and Sornette, D. (2003). Characterization of the frequency of extreme earthquake events by the Generalized Pareto Distribution. *Pure and Applied Geophysics*, 160(12):2343–2364.

Reiter, L. (1991). *Earthquake Hazard Analysis*. Columbia University Press, New York.

Richter, C. F. (1935). An instrumental earthquake magnitude scale. *Bulletin of the Seismological Society of America*, 25(1):1–20.

Roberts, N. S. (2016). *Earthquake Distributions at Volcanoes: Models and Field Observations*. PhD thesis, The University of Edinburgh.

Roberts, N. S., Bell, A. F., and Main, I. G. (2015). Are volcanic seismic b-values high, and if so when? *Journal of Volcanology and Geothermal Research*, 308:127–141.

Roberts, N. S., Bell, A. F., and Main, I. G. (2016). Mode switching in volcanic seismicity: El Hierro 2011–2013. *Geophysical Research Letters*, 43:4288–4296.

Rydelek, P. A. and Sacks, I. S. (1989). Testing the completeness of earthquake catalogues and the hypothesis of self-similarity. *Nature*, 337:251 – 253.

Sammonds, P. R., Meredith, P. G., and Maint, I. G. (1992). Role of pore fluids in the generation of seismic precursors to shear fracture. *Nature*, 359:228–230.

Schoenball, M., Davatzes, N. C., and Glen, J. M. G. (2015). Differentiating induced and natural seismicity using space-time-magnitude statistics applied to the Coso Geothermal field. *Geophysical Research Letters*, 42(15):6221–6228.

Scholz, C. H. (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.

Schorlemmer, D., Wiemer, S., and Wyss, M. (2004). Earthquake statistics at Parkfield: 1. Stationarity of b values. *Journal of Geophysical Research B: Solid Earth*, 109(12):1–17.

Schorlemmer, D. and Woessner, J. (2008). Probability of detecting an earthquake. *Bulletin of the Seismological Society of America*, 98(5):2103–2117.

Schwartz, D. P. and Coppersmith, K. J. (1984). Fault behavior and characteristic earthquakes - examples from the Wasatch and San Andreas fault zones. *Journal of Geophysical Research*, 89(B7):5681–5698.

Schwarz, G. (1978). Estimating the Dimension of a Model. *The Annals of Statistics*, 6(2):461–464.

Segall, P., Grasso, J.-R., and Mossop, A. (1994). Poroelastic stressing and induced seismicity near the Lacq gas field, southwestern France. *Journal of Geophysical Research*, 99(B8):15423–15438.

Shapiro, S. A., Krüger, O. S., and Dinske, C. (2013). Probability of inducing given-magnitude earthquakes by perturbing finite volumes of rocks. *Journal of Geophysical Research: Solid Earth*, 118(7):3557–3575.

Shcherbakov, R., Zhuang, J., Zöller, G., and Ogata, Y. (2019). Forecasting the magnitude of the largest expected earthquake. *Nature Communications*, 10(4051):1–11.

Smith, W. D. (1981). The b-value as an earthquake precursor. *Nature*, 289:136–139.

Sornette, D. and Sornette, A. (1999). General theory of the modified Gutenberg-Richter law for large seismic moments. *Bulletin of the Seismological Society of America*, 89(4):1121–1130.

Spallarossa, D., Cattaneo, M., Scafidi, D., Michele, M., Chiaraluce, L., Segou, M., and Main, I. G. (2021). An automatically generated high-resolution earthquake catalogue for the 2016-2017 Central Italy seismic sequence, including P and S phase arrival times. *Geophysical Journal International*, 225(1):555–571.

Staudenmaier, N., Tormann, T., Edwards, B., Deichmann, N., and Wiemer, S. (2018). Bilinearity in the Gutenberg-Richter Relation Based on M L for Magnitudes Above and Below 2, From Systematic Magnitude Assessments in Parkfield (California). *Geophysical Research Letters*, 45(14):6887–6897.

Stein, S., Geller, R., and Liu, M. (2011). Bad Assumptions or Bad Luck: Why Earthquake Hazard Maps Need Objective Testing. *Seismological Research Letters*, 82(5):623–626.

Stein, S., Geller, R. J., and Liu, M. (2012). Why earthquake hazard maps often fail and what to do about it. *Tectonophysics*, 562-563:1–25.

Stein, S. and Wysession, M. (2003). *An Introduction to Seismology, Earthquakes and Earth Structure*. Blackwell Publishing, Oxford, 14 edition.

Stirling, M. W. (2014). The Continued Utility of Probabilistic Seismic-Hazard Assessment. In *Earthquake Hazard, Risk and Disasters*, pages 359–376. Elsevier Inc.

Suckale, J. (2010). Moderate-to-large seismicity induced by hydrocarbon production. *The Leading Edge*, 29(3):310–319.

Telford, J., Cosgrave, J., and Houghton, R. (2006). Joint evaluation of the international response to the Indian Ocean tsunami: Synthesis Report. Technical report, Tsunami Evaluation Coalition, London.

Tinti, S. and Mulargia, F. (1985). Effects of Magnitude Uncertainties on Estimating

the Parameters in the Gutenberg-Richter Frequency-Magnitude Law. *Bulletin of the Seismological Society of America*, 75(6):1681–1697.

Tinti, S., Rimondi, R., and Mulargia, F. (1987). On Estimating Frequency-Magnitude Relations From Heterogeneous Catalogs. *Pageoph*, 125(1-18).

Touati, S., Naylor, M., and Main, I. G. (2009). Origin and Nonuniversality of the Earthquake Interevent Time Distribution. *Physical Review Letters*, 102(168501):1–4.

Turcotte, D. (1997). *Fractals and Chaos in Geology and Geophysics*. Cambridge University Press, Cambridge, UK.

van der Elst, N. J., Page, M. T., Weiser, D. A., Goebel, T. H., and Hosseini, S. M. (2016). Induced earthquake magnitudes are as large as (statistically) expected. *Journal of Geophysical Research: Solid Earth*, 121(6):4575–4590.

van Stiphout, T., Zhuang, J., and Marsan, D. (2012). Seismicity Declustering. *Community Online Resource for Statistical Seismicity Analysis*, pages 1–25.

Vanneste, K., Stein, S., Camelbeeck, T., and Vleminckx, B. (2018). Insights into earthquake hazard map performance from shaking history simulations. *Scientific Reports*, 8(1855):1–10.

Verdon, J., Kettlety, T., and Kendall, M. (2019a). Geomechanical Interpretation of Microseismicity at the Preston New Road PNR-1z Well, Lancashire, England. Technical report, Outer Limits, Report Commissioned by the Oil and Gas Authority, London.

Verdon, J. P., Baptie, B. J., and Bommer, J. J. (2019b). An improved framework for discriminating seismicity induced by industrial activities from natural earthquakes. *Seismological Research Letters*, 90(4):1592–1611.

Verdon, J. P. and Bommer, J. J. (2021). Green, yellow, red, or out of the blue? An assessment of Traffic Light Schemes to mitigate the impact of hydraulic fracturing-induced seismicity. *Journal of Seismology*, 25(1):301–326.

Verdon, J. P., Kendall, J. M., Butcher, A., Luckett, R., and Baptie, B. J. (2018). Seismicity induced by longwall coal mining at the Thoresby Colliery, Nottinghamshire, U.K. *Geophysical Journal International*, 212(2):942–954.

Verdon, J. P. and Stork, A. L. (2016). Carbon capture and storage, geomechanics and induced seismic activity. *Journal of Rock Mechanics and Geotechnical Engineering*, 8:928–935.

Vere-Jones, D., Robinson, R., and Yang, W. (2001). Remarks on the accelerated moment release model: problems of model formulation, simulation and estimation. *Geophysical Journal International*, 144:517–531.

Vermylen, J. P. and Zoback, M. D. (2011). Hydraulic Fracturing, Microseismic Magnitudes, and Stress Evolution in the Barnett Shale, Texas, USA. Technical report.

Walter, J. I., Ogwari, P., Thiel, A., Ferrer, F., Woelfel, I., Chan, J. C., Darold, A. P., and Holland, A. A. (2020). The Oklahoma geological survey statewide seismic network. *Seismological Research Letters*, 91(2 A):611–621.

Walters, R. J., Zoback, M. D., Baker, J. W., and Beroza, G. C. (2015). Characterizing and responding to seismic risk associated with earthquakes potentially triggered by fluid disposal and hydraulic fracturing. *Seismological Research Letters*, 86(4):1110–1118.

Wesson, R. L., Boyd, O. S., Mueller, C. S., Bufe, C. G., Frankel, A. D., and Petersen, M. D. (2007). Revision of Time-Independent Probabilistic Seismic Hazard Maps for Alaska Open-File Report 2007-1043. Technical report.

Wiemer, S., 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*, 134:409–421.

Wiemer, S. and Wyss, M. (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.

Woessner, J. and Wiemer, S. (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.

World Bank (2012). The great east Japan earthquake: Learning from megadisasters. Technical report, The World Bank, Washington D.C.

Wyss, M., Hasegawa, A., Wiemer, S., and Umino, N. (1999). Quantitative mapping of precursory seismic quiescence before the 1989, M7.1 off-Sanriku earthquake, Japan. *Annali di Geofisica*, 42(5):851–869.

Wyss, M., Klein, F., Nagamine, K., and Wiemer, S. (2001). Anomalously high b-values in the South Flank of Kilauea volcano, Hawaii: evidence for the distribution of magma below Kilauea's East rift zone. *Journal of Volcanology and Geothermal Research*, 106:23–37.

Wyss, M., Sammis, C. G., Nadeau, R. M., and Wiemer, S. (2004). Fractal dimension and b-value on creeping and locked patches of the San Andreas fault near Parkfield, California. *Bulletin of the Seismological Society of America*, 94(2):410–421.

Yaghmaei-Sabegh, S. and Ostadi-Asl, G. (2021). Bayesian estimation of b-value in Gutenberg–Richter relationship: a sample size reduction approach. *Natural Hazards*, 110:1783–1797.

Zechar, J. D., Marzocchi, W., and Wiemer, S. (2016). Operational earthquake forecasting in Europe: progress, despite challenges. *Bulletin of Earthquake Engineering*, 14:2459–2469.

Zöller, G. (2013). Convergence of the frequency-magnitude distribution of global earthquakes: Maybe in 200 years. *Geophysical Research Letters*, 40:3873–3877.