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

Geophysical Journal International

Geophys. J. Int. (2022) **229,** 1840–1855
Advance Access publication 2022 January 27
GJI Seismology

https://doi.org/10.1093/gji/ggac028

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

G.-M. Geffers, I.G. Main and M. Naylor

School of Geosciences, University of Edinburgh, James Hutton Road, Edinburgh, EH9 3FE, UK. E-mail: G.Geffers@ed.ac.uk

Accepted 2022 0. Received 2021 December 13; in original form 2021 September 13

SUMMARY

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 data sets—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 minimize this bias.

Key words: Statistical methods; Induced seismicity; Statistical seismology; Volcano seismology.

1 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 eq. (1). This implies the frequency–size distribution of source rupture areas scales as a power law, indicating self-similarity on all scales (Kanamori & 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 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 $\sim$1.0 for shallow tectonic events (Frohlich & Davis 1993). However, Bell *et al.* (2013) used the global Harvard Centroid Moment Tensor (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 characterized by a 'corner moment' M_θ (Main & Burton 1984; Kagan 1991). Frohlich & Davis (1993) also investigated the b-value in teleseismic data for tectonic events and found that it may vary by 30 percent 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 (Wiemer *et al.* 1998; Jolly & McNutt 1999; Murru *et al.* 1999, 2005; Novelo-Casanova *et al.* 2006; Jacobs & McNutt 2010; Ibáñez *et al.* 2012). Specifically, Ibáñez *et al.* (2012) and Jolly & 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 1 summarizes 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 & 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 cut-off 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.

In the case of induced seismicity, Table 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 data sets 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 artificially 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.

2 THEORY

2.1 Frequency–magnitude distributions

The FMD describes either the incremental number of events within a magnitude window or the cumulative number of events above a series of magnitude thresholds (Fig. 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

Table 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 is the least-squares method and WLS is 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 & 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 & 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

Table 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 & Zoback (2011)	F	Fort Worth, TX, USA	4485	–	−2.5	LS	1.5–1.9
Henderson *et al.* (1999)	G	Geysers, CA, USA	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 to 4.4	0.5	MLE	1.06 ± 0.05
Langenbruch & Zoback (2016)	WD	Central OK, USA	–	3.0–?	–	MLE	1.41

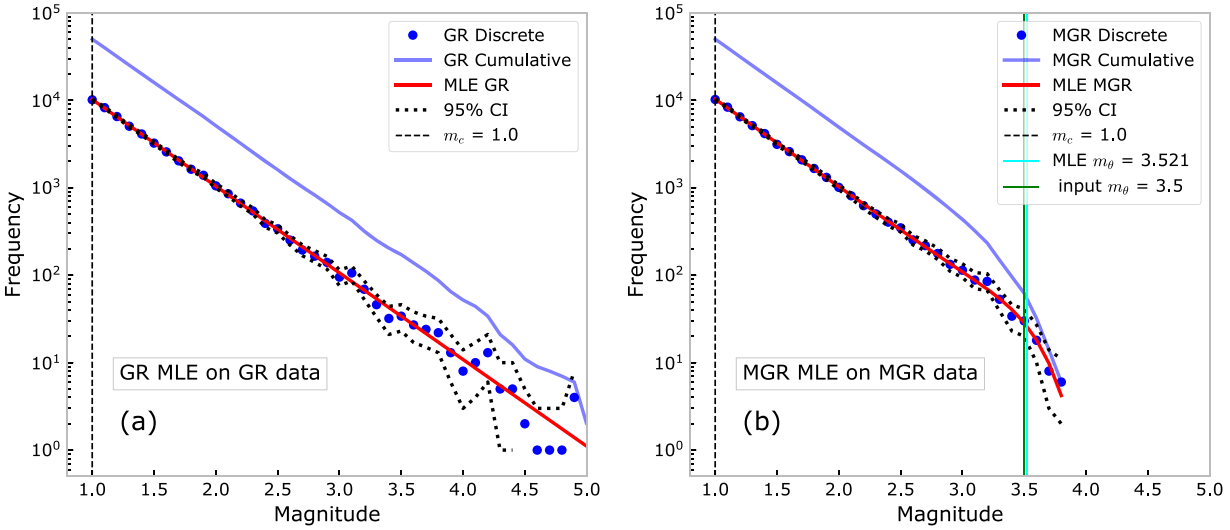

Figure 1. Synthetic incremental (blue circles) and cumulative (blue line) frequency–magnitude distributions for a catalogue of 50 000 events, assuming a complete catalogue (to the right of the leftmost vertical line) for both (a) the GR (eq. 2) and (b) the MGR (eq. 3) distributions, including 95 percent confidence intervals (CI).

& Richter 1944) in the form

$$\log_{10} F = a - bm, \tag{1}$$

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 & 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

$$F(M_0) \sim M_0^{-\beta}, \tag{2}$$

where $\beta = (2/3)b$ (Turcotte 1997; Kagan 2002); β being the equivalent of the *b*-value in the magnitude–frequency relation, where the factor 2/3 results from a log-linear model relating the moment-magnitude to the moment $\log(M_0) = A + Bm$. Hanks & Kanamori (1979) have taken this as a definition of moment magnitude with $A = 9.1$ and $B = 1.5$. The β-value in eq. (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 & Anderson 1975). The form of eq. (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 capitalized 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 (eqs 1 and 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 eq. (2) at large seismic moments, thus modifying the GR law as derived by Main & 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

$$F(M_0) \sim M_0^{-\beta} \exp(-M_0/M_\theta), \tag{3}$$

where M_θ is a 'corner moment' where the cumulative frequency $F(M)$ has dropped to a value $1/e$ 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 eqs (2) and (3) are shown in Fig. 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 2002) 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 Fig. 2, which shows that at $N = 200$, the b-value error is around 0.1. Fig. 2 uses synthetic data sampled randomly from an underlying GR 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.

Kagan & 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 $\pm$ 0.02 (equivalent b-value of $\sim$0.95 $\pm$ 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 & 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 & 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 & 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.

2.2 Estimating the parameters of the frequency–magnitude distribution

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

2.2.1 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 1 and 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.

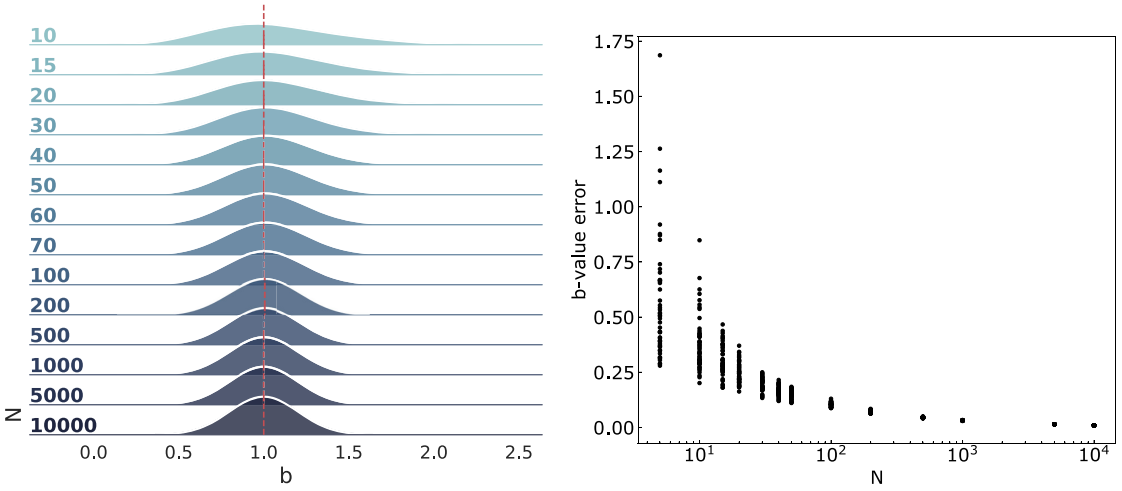

Figure 2. Left: distribution of inferred *b*-values from randomly sampled data taken from a GR distribution, showing the convergence of the *b*-value with increased sample size. The true *b*-value of 1.0 is indicated by the vertical dashed red line. Right: the random error in *b*-value [using the maximum likelihood estimator described by Aki (1965), shown here to one standard deviation] when sampled from a GR distribution, with respect to increasing sample size.

2.2.2 *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 & 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 (i) that the data genuinely are power-law distributed in seismic moment, (ii) that the maximum moment (equivalent to maximum magnitude) is at infinity and (iii) 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 (eq. 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 2002). To estimate the seismic moment distribution parameters (β for GR and β and M_θ for MGR), we apply an MLE by Kagan (2002). For GR, the log-likelihood function for N observations of the seismic moment is

$$l_{\mathrm{GR}} = \log(L_{\mathrm{GR}}) = N\beta\log(M_c) - \beta\sum_{i=1}^{N}\log M_i + \sum_{i=1}^{N}\log\left(\frac{\beta}{M_i}\right). \quad (4)$$

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

$$l_{\mathrm{MGR}} = \log(L_{\mathrm{MGR}}) = N\beta\log(M_c) + \frac{1}{M_\theta}\left(NM_c - \sum_{i=1}^{N}M_i\right) - \beta\sum_{i=1}^{N}\log M_i + \sum_{i=1}^{N}\log\left(\frac{\beta}{M_i} + \frac{1}{M_\theta}\right), \quad (5)$$

where M_c is the completeness moment and M_i are the catalogue moments on the real number ($\mathbb{R}$) line. Maximizing 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 2002), which is approximated to

$$\sigma_\beta = \frac{\beta}{\sqrt{N}}, \quad (6)$$

where N is the number of events included to calculate $\tilde{b}$. For the MGR distribution, we use eq. (25) in Kagan (2002) to estimate the standard error on β:

$$\sigma_{\tilde{\beta}} = \frac{D}{\sqrt{n[D^2\tilde{\beta}^{-2} - (M_c/M_x)^\beta(\log M_x/M_c)^2]}} \quad (7)$$

where $D = 1 - (M_c/M_x)^\beta$ and $M_x \gg \frac{M_\theta}{M_{\max}}$.

3 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 or MGR distribution whilst varying the dynamic (magnitude) range. These results are then generalized 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.

3.1 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 & 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 penalizes 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

$$\mathrm{BIC} = \frac{1}{\log_{10}(e)} \left(-2\log(L) + k\log(N)\right), \tag{8}$$

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 ($\Delta\mathrm{BIC}_{\mathrm{MGR}} - \Delta\mathrm{BIC}_{\mathrm{GR}}$) with a resultant $\Delta\mathrm{BIC} > 0$ preferring the GR distribution and $\Delta\mathrm{BIC} < 0$ preferring the MGR distribution (Bell *et al.* 2013).

3.2 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 $m_\theta = 3.5$. 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 subcatalogues 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. Maximizing the parameters in the likelihood functions for the GR and MGR models (eqs 4 and 5) result in estimates of the b-value for each subcatalogue for both GR and MGR distributions, as well as estimates for M_θ for the MGR distribution.

Additionally, we then use randomly generated MC sampling to simulate multiple catalogues with the same parameters, and look at the variability between individual realizations. 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.

3.3 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 & Wyss 2000), the goodness-of-fit test (GFT; Wiemer & Wyss 2000) and the b-value stability method (BVS; Cao & Gao 2002; Woessner & Wiemer 2005). All of these methods assume self-similarity of the earthquake process, implying a power-law distribution (Woessner & Wiemer 2005), but none of these methods is necessarily the 'correct' one. Mignan & 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 stabilizes 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 & Marzocchi (2020). A summary of the sample size and magnitude ranges for each of these catalogues is given in Table 3. These catalogues have also been chosen due to their abundance in data (large numbers of events in the raw catalogue). However, Table 3 shows a large reduction in catalogue sample size and dynamic range for the complete data above m_c, especially for the Mount Etna, El Hierro and 1990 INGV catalogues.

3.3.1 *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 two 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.

3.3.2 *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.

Table 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 Fig. 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	15688	1928	5.7	2.5	3.2	∼5.7
El Hierro	Volcanic	30/11/2001–22/01/2017	22494	1782	5.1	2.6	2.5	∼5.0
The Geysers	Geothermal	18/03/1972–31/07/2018	281572	61134	5.01	1.25	3.76	∼4.6
INGV 1990	Tectonic	01/01/1990–15/04/2005	17166	2917	5.8	2.7	3.1	∼6.0
INGV 2005	Tectonic	16/04/2005–24/08/2016	114847	16 629	6.1	1.7	4.4	∼6.5

3.3.3 *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 percent 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.

3.3.4 *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.* 2018). 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 2016 December 14. 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 have 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 minimize the possibility of including other events that may not be induced by geothermal activities.

4 RESULTS

Here we present the results of the analysis on synthetic catalogue data drawn from a GR and MGR distribution and generalized using MC simulations (as described in Section 3.2) before demonstrating the applications on the real data from the catalogues introduced above.

4.1 Synthetic catalogues

Fig. 3 summarizes 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 as one specific subcatalogue ($m_c = 2.5$, GR in black, MGR in blue). The b-value is 1.0 for both and in the case of the MGR distribution, $m_\theta = 3.5$. The smaller subcatalogue 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 Fig. 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 subcatalogues 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 percent 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 3.1) is at choosing the true model, with the key factor being the extra complexity (due to the additional model parameter M_θ) and the associated BIC penalty when applying the MGR model. Using $m_\theta = 3.5$, Fig. 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, that is, 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

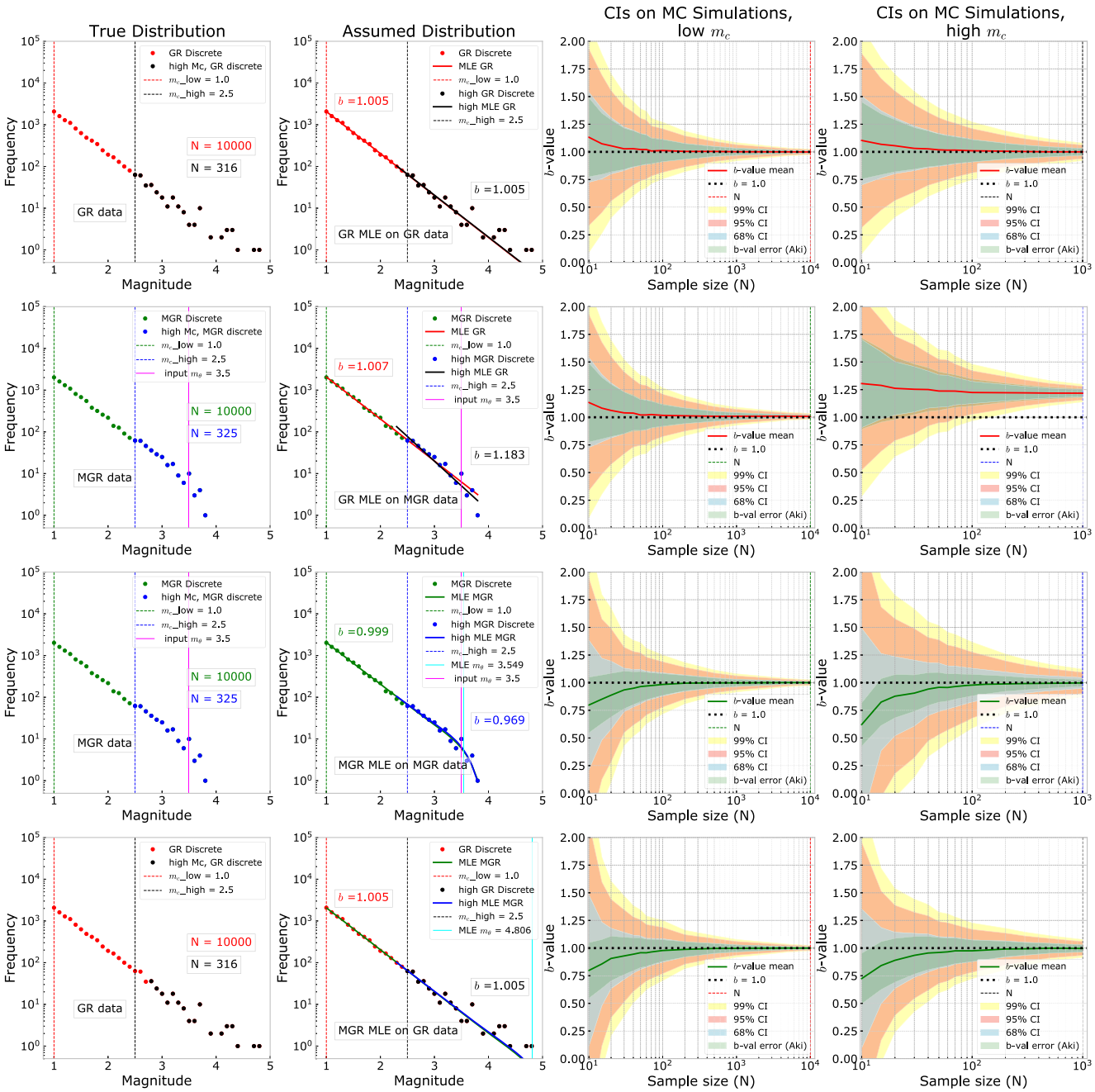

Figure 3. Summary of the synthetic results obtained using GR and MGR models. The first column represents the full and a smaller subcatalogue of the true distribution—rows 1 and 4 show GR (red, subcatalogue in black) and rows 2 and 3 show MGR (green, subcatalogue in blue). The second column includes the 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 percent) 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 subcatalogue. The solid red and green lines represent the *b*-value mean.

$= 3.5$, although not all (seen by some points being to the left of the vertical line indicating m_θ). The model acceptance in the centre 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, Fig. 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 Fig. 5 have ΔBIC < 0,

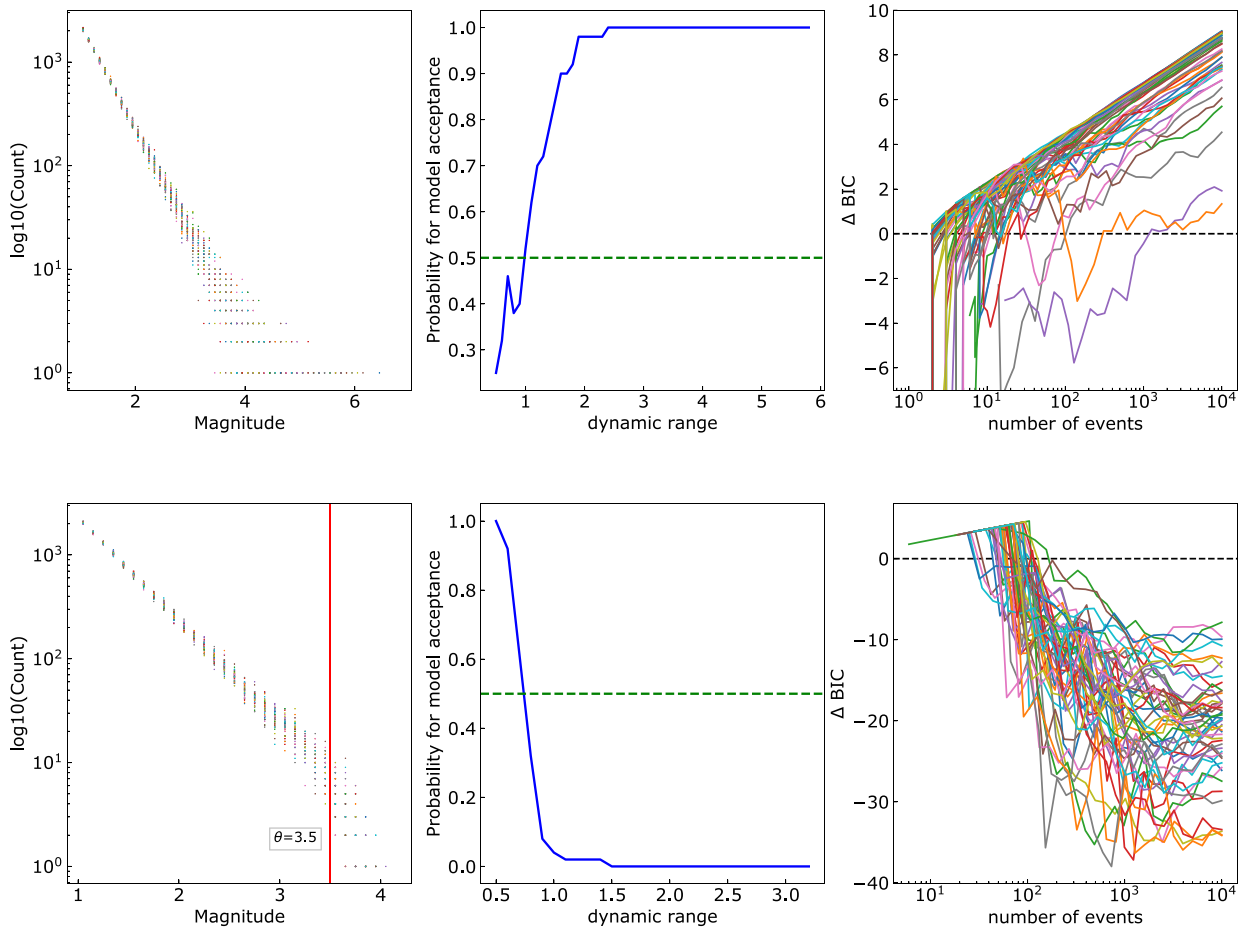

Figure 4. Left: FMDs for all GR (top) catalogue and MGR (bottom) catalogue simulations superposed. The solid red line in the MGR simulations indicates the input value of m_θ. Centre: the probability that a model will be preferred dependent on the dynamic range, defined as the maximum sampled magnitude 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.

suggesting an incorrect preference of the MGR model, while estimated b-values range 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 $b = 1.0 \pm 0.2$) and the preference for the GR model becomes stronger (greater ΔBIC). In the case of an underlying MGR distribution (Fig. 5, bottom), a smaller dynamic range (green triangles) again results in increased misidentification of the correct model—many points have ΔBIC > 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 Fig. 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 and 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 $N = 50$ (Fig. 6, top left), b-value estimates range from $\sim$ 0.7 (at dynamic ranges > 1.5 and $m_\theta >$ 3.4) to ∼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 $N = 5000$ (Fig. 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 Fig. 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 $b = 1.0$. Fig. 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, that is it is neither accurate nor precise, respectively. Accuracy and precision both improve with increased dynamic range and sample size. Nevertheless, Fig. 6 suggests that an accurate estimate of the b-value can be obtained within a precision of approximately ±0.2, even when the incorrect model is assumed, if (i) the dynamic range is at least 1.5 and (ii) there are at least $N = 1000$ events in the sample.

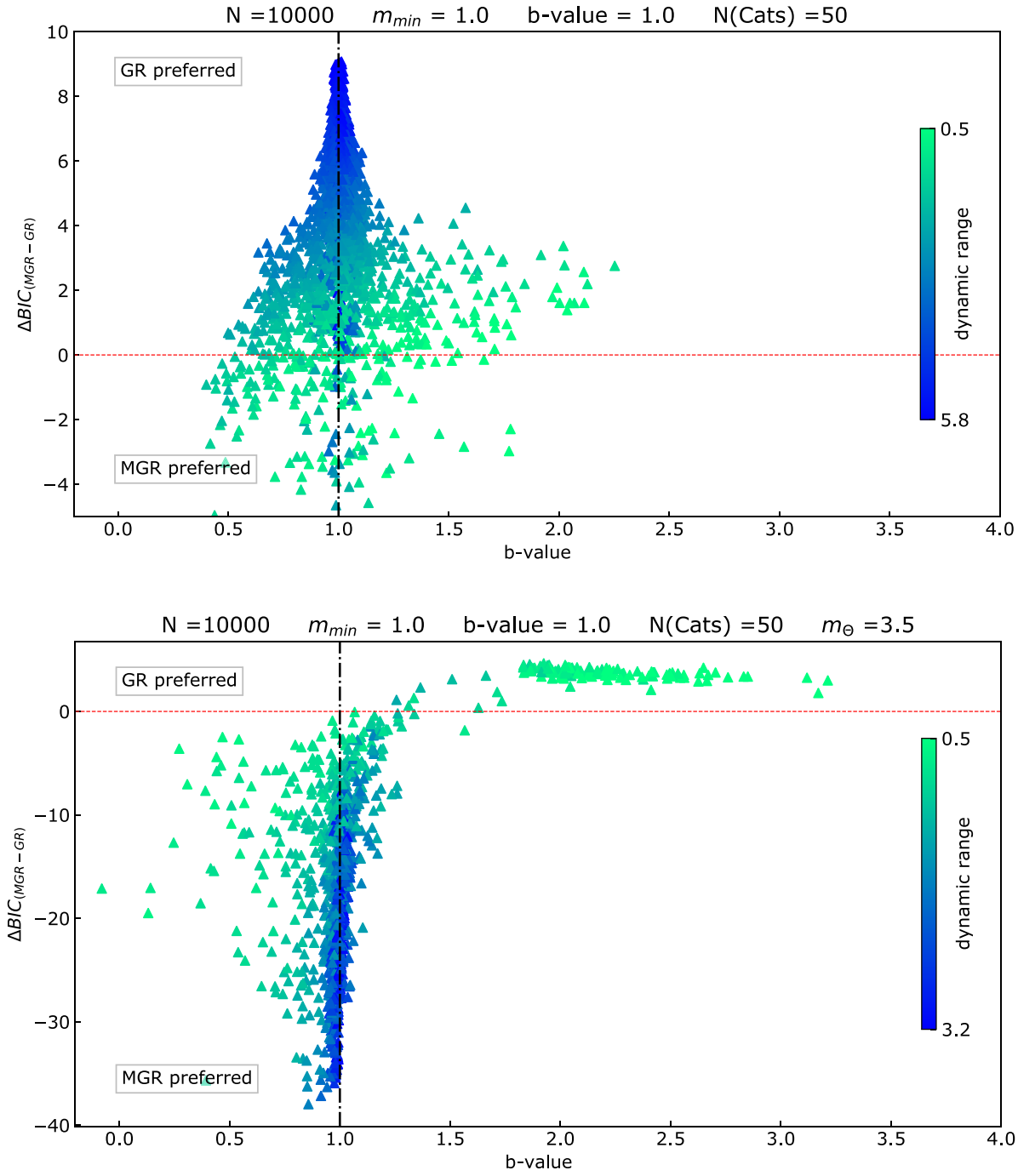

Figure 5. Analysis for GR data (top) and MGR data (bottom) showing: ΔBIC against the inferred *b*-value for all 50 catalogues with 10 000 events generated from a synthetic catalogue with $b = 1.0$, varying m_c for each. Positive ΔBIC implies a preference for the GR model, negative ΔBIC is a preference for the MGR model. Darker (blue) markers indicate larger dynamic ranges; lighter (green) markers indicate small dynamic ranges.

4.2 Real catalogues

We now investigate the real catalogue data from the areas of tectonic, volcanic and induced seismicity introduced in Section 3.3. 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 1 and 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 3.3, 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 (Figs 7 and 8), and then increasing the obtained m_c to thin

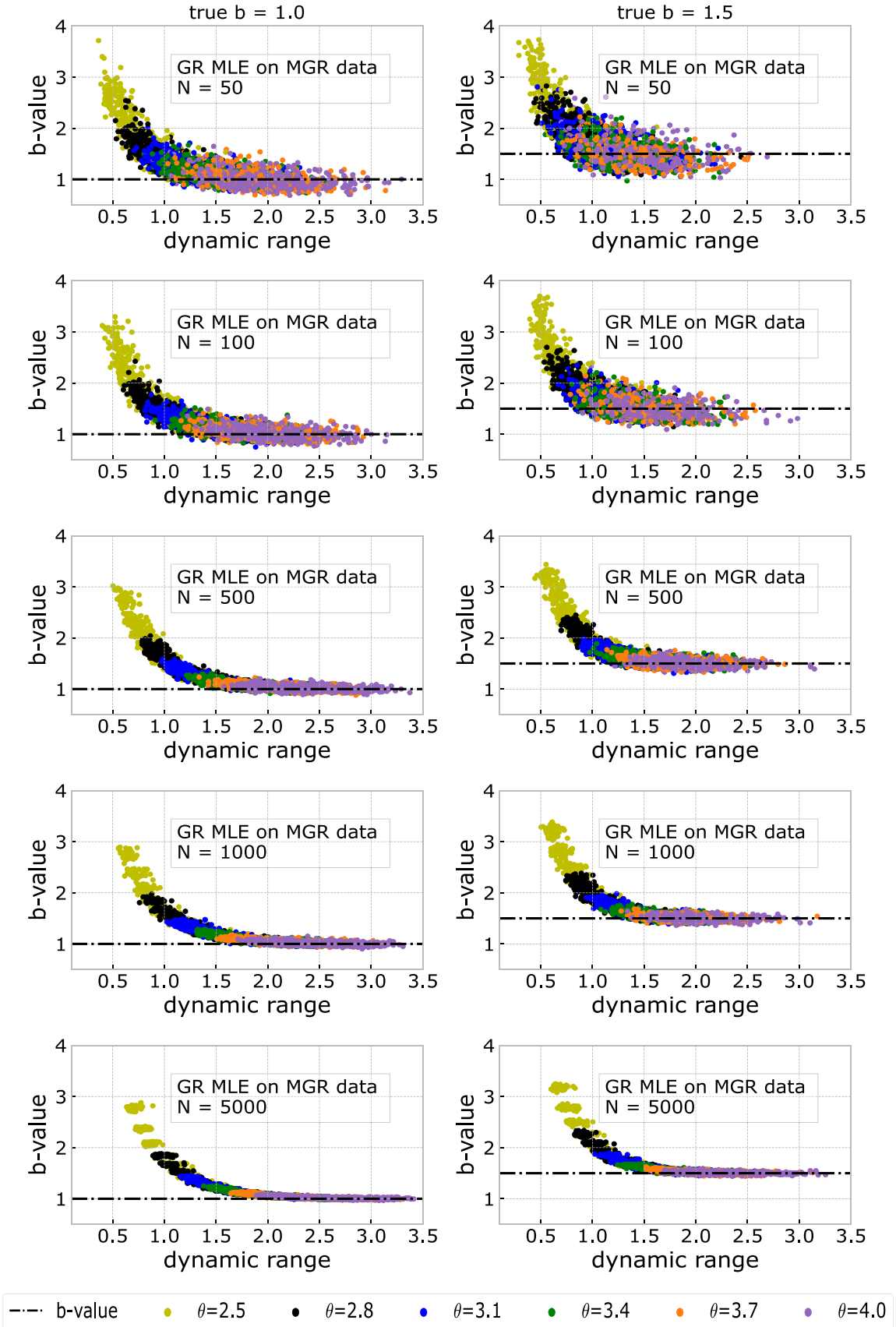

Figure 6. Estimated *b*-values assuming the GR model (applied on MGR data) as a function of dynamic range for Left column: a true *b*-value of 1.0; right column: a true *b*-value of 1.5. Colours represent different values of m_θ (see key immediately above).

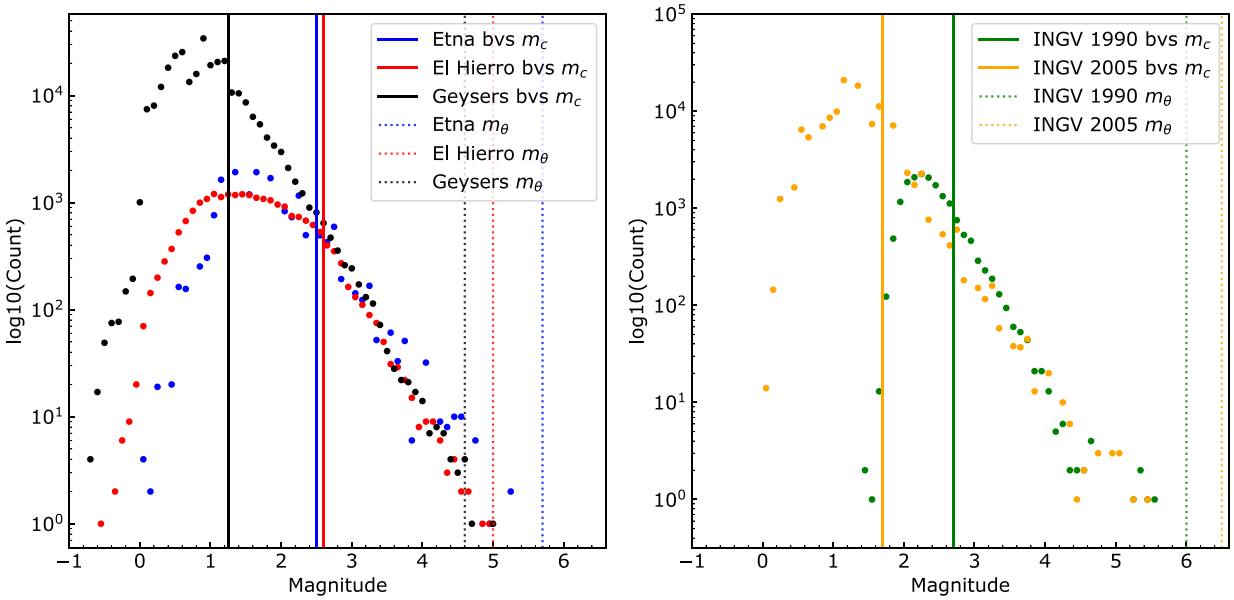

Figure 7. FMDs for all five catalogues. Solid vertical lines indicate the estimated m_c using the BVS method, dotted vertical lines show the calculated value of m_θ. The top row shows results from volcanic and induced seismicity, and the bottom row indicates a single tectonic setting, analysed for two catalogues of varying temporal and spatial resolution.

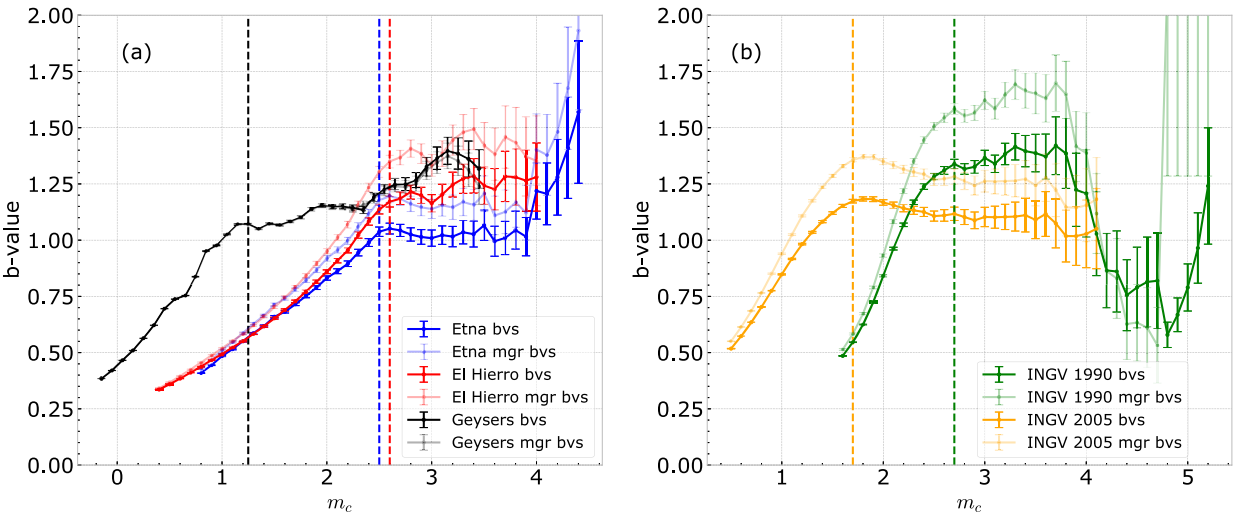

Figure 8. *b*-value stability plot using GR (solid colours) and MGR (lighter colours) for (a) the volcanic and induced catalogues and (b) the tectonic catalogues. Vertical dashed lines indicate the estimated m_c for each catalogue, based on successive *b*-values falling within error of the *b*-value. (Based on Roberts *et al.* (2015)).

the catalogue in the same way as we did for the synthetic data (Section 3.2).

Fig. 7 shows the FMDs for all catalogues, including the estimates of the BVS m_c (see Fig. 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 Fig. 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 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 $\sim$1.05 for The Geysers, $\sim$1.0 for Etna and $\sim$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 $3.7 \geq m_c \geq 4.8$ with *b*-values dropping as low as $\sim$ 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 Fig. 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 ($\sim$0.5) but then exhibits a continuous decreases until it reaches a *b*-value of $\sim$1.0 at a dynamic range of $\sim$3.7, again similar to that of tectonic seismicity. The two volcanic catalogues generally have a decreasing or stable *b*-value between dynamic ranges of 1.0 and 2.3 (for El Hierro) and 2.9 (for Etna), before they increase again just before their maximum dynamic range. This may be a statistical artefact of sampling for rare, large events. The INGV 1990 catalogue has a very low estimated *b*-value at low dynamic range ($b \approx 0.55$ at a dynamic range of 1.0). It continuously increases until it reaches a dynamic range of $\sim$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 Fig. 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 Fig. 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 Fig. 6. Notably, the INGV catalogues behave different from the induced and volcanic seismicity catalogues, and are also distinct from one another.

Fig. 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 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 artificially high *b*-value. From the value of m_θ in Table 3 and Fig. 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 Fig. 10 confirms this inference, that is 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 $\sim$0.9 to 1.2, that is not too different from that expected for tectonic earthquakes, but substantially smaller than the values of b_{typ} given in previous studies (see Table 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 Fig. 10 indicate inferred *b*-values from previous studies (see Tables 1 and 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 3 suggests a typical *b*-value of < 1.0), compared to our results.

5 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 Fig. 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 subcatalogues 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 subcatalogues 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 1 and 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 eq. (6). Furthermore, many

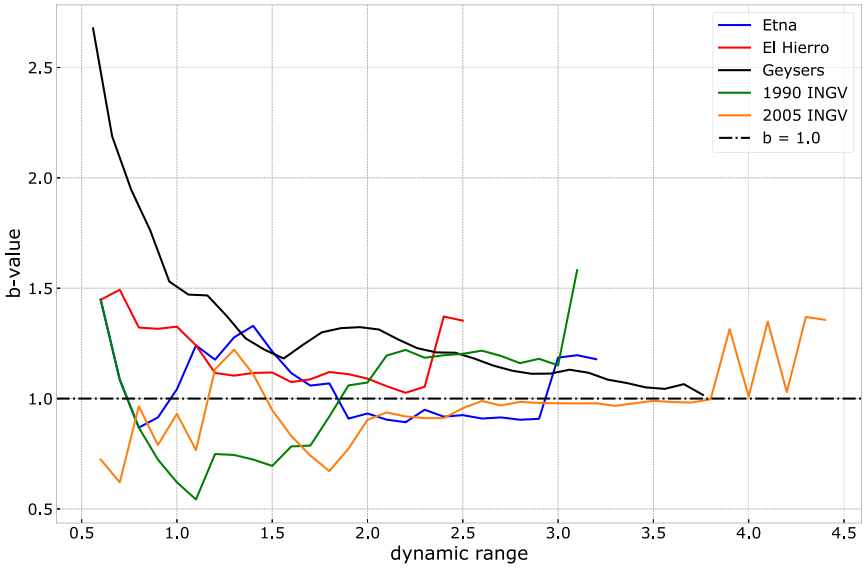

Figure 9. Equivalent of Fig. 6 for the real catalogues, showing the *b*-value as a function of dynamic range for each of the real catalogues. The 'commonly' accepted *b*-value of 1.0 is shown as a dashed black line for reference.

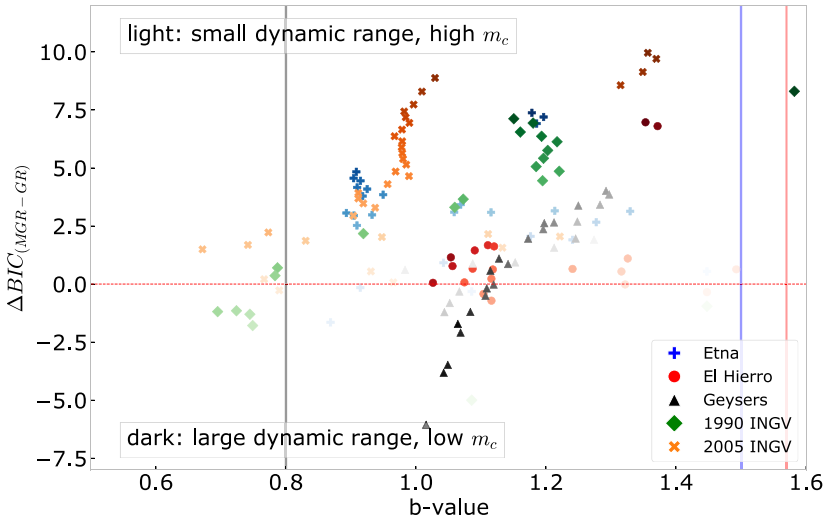

Figure 10. ΔBIC as a function of the inferred *b*-value for five different data sets: two volcanic (Etna and El Hierro), one induced (the Geysers) and one tectonic (Central Italy). Positive ΔBIC implies a preference for the GR model. *b*-values are calculated based on which model is preferred (MGR only 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 1 and 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.

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 suboptimal. Recently, Marzocchi *et al.* (2020) and Herrmann & 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.

6 CONCLUSION

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

(i) 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 Figs 4 and 5, the data suggest that two orders of dynamic range are required to correctly estimate a GR model and its *b*-values within ±

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 percent for a dynamic range of 1.0 and < 0.5 percent 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.

(ii) 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 (Figs 3, 5 and 6). From Fig. 3, to obtain a b-value to a precision of $\pm$ 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.

(iii) 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 data sets. 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 & 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 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.

ACKNOWLEDGMENTS

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 Natural Environment Research Council (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).

REFERENCES

Aki, K., 1965. Maximum likelihood estimate of b in the Formula log $N=a-bM$ and its confidence limits, *Bull. Earthq. Res. Inst.,* **43,** 237–239.

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

Bachmann, C.E., Wiemer, S., Goertz-Allmann, B.P. & Woessner, J., 2012. Influence of pore-pressure on the event-size distribution of induced earthquakes, *Geophys. Res. Lett.,* **39**(L09302), doi:10.1029/2012GL051480.

Bell, A.F., Naylor, M. & Main, I.G., 2013. Convergence of the frequency–size distribution of global earthquakes, *Geophys. Res. Lett.,* **40,** 2585–2589.

Cao, A. & Gao, S.S., 2002. Temporal variation of seismic b-values beneath northeastern Japan island arc, *Geophys. Res. Lett.,* **29**(9), 48–1–48-3.

Colombo, I., Main, I. & Forde, M., 2003. Assessing damage of reinforced concrete beam using "b-value" analysis of acoustic emission signals, *J. Mater. Civil Eng.,* **15**(3), 280–286.

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

Frohlich, C. & Davis, S.D., 1993. Teleseismic b values; Or, much ado about 1.0, *J. geophys. Res.,* **98**(B1), 631–644.

Godano, C. & Pingue, F., 2000. Is the seismic moment–frequency relation universal? *Geophys. J. Int.,* **142,** 193–198.

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. & Richter, C.F., 1944. Frequency of earthquakes in California, *Bull. seism. Soc. Am.,* **34**(4), 185–188.

Hanks, T.C. & Kanamori, H., 1979. A moment magnitude scale, *J. geophys. Res.,* **84**(B5), 2348–2350.

Henderson, J.R., Barton, D.J. & Foulger, G.R., 1999. Fractal clustering of induced seismicity in The Geysers geothermal area, California, *Geophys. J. Int.,* **139,** 317–324.

Herrmann, M. & Marzocchi, W., 2020. Inconsistencies and lurking pitfalls in the magnitude–frequency distribution of high-resolution earthquake catalogs, *Seismol. Res. Lett.,* **92**(2A), 909–922.

Ibáñez, J.M., De Angelis, S., Díaz-Moreno, A., Hernández, P., Alguacil, G., Posadas, A. & 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, *Geophys. J. Int.,* **191,** 659–670.

Igonin, N., Zecevic, M. & Eaton, D.W., 2018. Bilinear magnitude–frequency distributions and characteristic earthquakes during hydraulic fracturing, *Geophys. Res. Lett.,* **45**(23), 866–12.

Jacobs, K.M. & 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, chap. 3, pp. 59–83.

Jolly, A.D. & McNutt, S.R., 1999. Seismicity at the volcanoes of Katmai National Park, Alaska; July 1995–December 1997, *J. Volc. Geotherm. Res.,* **93,** 173–190.

Kagan, Y.Y., 1991. Seismic moment distribution, *Geophys. J. Int.,* **106,** 123–134.

Kagan, Y.Y., 1997. Seismic moment–frequency relation for shallow earthquakes: regional comparison, *J. geophys. Res.,* **102**(B2), 2835–2852.

Kagan, Y.Y., 1999. Universality of the seismic moment–frequency relation, *Pure appl. Geophys.,* **155,** 537–573.

Kagan, Y.Y., 2002. Seismic moment distribution revisited: I. Statistical results, *Geophys. J. Int.,* **148,** 520–541.

Kagan, Y.Y. & Jackson, D.D., 1991. Seismic gap hypothesis: ten years after, *J. geophys. Res.,* **96**(B13), 419–421.

Kagan, Y.Y. & Jackson, D.D., 2000. Probabilistic forecasting of earthquakes, *Geophys. J. Int.,* **143,** 438–453.

Kanamori, H. & Anderson, D.L., 1975. Theoretical basis of some empirical relations in seismology, *Bull. seism. Soc. Am.,* **65**(5), 1073–1095.

Kettley, T., Verdon, J.P., Butcher, A., Hampson, M. & Craddock, L., 2021. High-resolution imaging of the M_L 2.9 August 2019 earthquake in Lancashire, United Kingdom, induced by hydraulic fracturing during Preston New Road PNR-2 operations, *Seismol. Res. Lett.,* **92**(1), 151–169.

Langenbruch, C. & Zoback, M.D., 2016. How will induced seismicity in Oklahoma respond to decreased saltwater injection rates? *Sci. Adv.,* **2**(11), e1601542.

Leonard, T. & Hsu, J.S.J., 1999. *Bayesian Methods,* Cambridge Univ. Press.

Leptokaropoulos, K., Staszek, M., Lasocki, S., Martínez-Garzón, P. & Kwiatek, G., 2018. Evolution of seismicity in relation to fluid injection in the North-Western part of The Geysers geothermal field, *Geophys. J. Int.,* **212**(2), 1157–1166.

Main, I., 1995. Earthquakes as critical phenomena: implications for probabilistic seismic hazard analysis, *Bull. seism. Soc. Am.,* **85**(5), 1299–1308.

Main, I., 1996. Statistical physics, seismogenesis, and seismic hazard, *Rev. Geophys.,* **34**(4), 433–462.

Main, I. & Burton, P., 1984. Information theory and the earthquake frequency–magnitude distribution, *Bull. seism. Soc. Am.,* **74**(4), 1409–1426.

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

Majer, E.L., Baria, R., Stark, M., Oates, S., Bommer, J., Smith, B. & Asanuma, H., 2007. Induced seismicity associated with enhanced geothermal systems, *Geothermics,* **36**(3), 185–222.

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

Marzocchi, W., Spassiani, I., Stallone, A. & Taroni, M., 2020. How to be fooled searching for significant variations of the *b*-value, *Geophys. J. Int.,* **220**(3), 1845–1856.

Maxwell, S. *et al.*, 2009. Fault Activation During Hydraulic Fracturing, in *SEG Technical Program Expanded Abstracts,* pp. 1552–1556, Society of Exploration Geophysicists.

Mignan, A. & Woessner, J., 2012. Estimating the magnitude of completeness for earthquake catalogs, Community Online Resource for Statistical Seismicity Analysis, doi:10.5078/corssa-00180805.

Mizrahi, L., Nandan, S. & Wiemer, S., 2021. The effect of declustering on the size distribution of mainshocks, *Seismol. Res. Lett.,* **92,** 2333–2342.

Mogi, K., 1962. Study of the elastic shocks caused by the fracture of heterogeneous materials and its relation to earthquake phenomena, *Bull. Earthq. Res. Inst.,* **40,** 125–173.

Mousavi, S.M., Ogwari, P.O., Horton, S.P. & Langston, C.A., 2017. Spatio-temporal evolution of frequency–magnitude distribution and seismogenic index during initiation of induced seismicity at Guy-Greenbrier, Arkansas, *Phys. Earth Planet. Inter.,* **267,** 53–66.

Murru, M., Montuori, C., Wyss, M. & Privitera, E., 1999. The locations of magma chambers at Mt. Etna, Italy, mapped by *b*-values, *Geophys. Res. Lett.,* **26**(16), 2553–2556.

Murru, M., Montuori, C., Console, R. & Lisi, A., 2005. Mapping of the *b* value anomalies beneath Mt. Etna, Italy, during July–August 2001 lateral eruption, *Geophys. Res. Lett.,* **32**(L05309), doi:10.1029/2004GL021545.

Nava, F.A., Márquez-Ramírez, V.H., Zúñiga, F.R., Ávila-Barrientos, L. & Quinteros, C.B., 2017. Gutenberg–Richter *b*-value maximum likelihood estimation and sample size, *J. Seismol.,* **21,** 127–135.

Naylor, M., Orfanogiannaki, K. & Harte, D., 2010. Theme III–statistical foundations exploratory data analysis: magnitude, space, and time, Community Online Resource for Statistical Seismicity Analysis, doi:10.5078/corssa-92330203.

Novelo-Casanova, D.A., Martínez-Bringas, A. & Valdés-González, C., 2006. Temporal variations of Q_c^{-1} and *b*-values associated to the December 2000–January 2001 volcanic activity at the Popocatepetl volcano, Mexico, *J. Volc. Geotherm. Res.,* **152,** 347–358.

Oppenheimer, D.H., 1986. Extensional tectonics at the Geysers Geothermal Area, California, *J. geophys. Res.,* **91**(B11), 463–474.

Reiter, L., 1991. *Earthquake Hazard Analysis,* Columbia University Press.

Roberts, N.S., Bell, A.F. & Main, I.G., 2015. Are volcanic seismic *b*-values high, and if so when? *J. Volc. Geotherm. Res.,* **308,** 127–141.

Roberts, N.S., Bell, A.F. & Main, I.G., 2016. Mode switching in volcanic seismicity: El Hierro 2011–2013, *Geophys. Res. Lett.,* **43,** 4288–4296.

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

Schoenball, M., Davatzes, N.C. & Glen, J. M.G., 2015. Differentiating induced and natural seismicity using space–time–magnitude statistics applied to the Coso Geothermal field, *Geophys. Res. Lett.,* **42**(15), doi:10.1002/2015GL064772.

Scholz, C.H., 1968. The frequency–magnitude relation of microfracturing in rock and its relation to earthquakes, *Bull. seism. Soc. Am.,* **58**(1), 399–415.

Staudenmaier, N., Tormann, T., Edwards, B., Deichmann, N. & 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), *Geophys. Res. Lett.,* **45**(14), 6887–6897.

Tinti, S. & Mulargia, F., 1985. Effects of magnitude uncertainties on estimating the parameters in the Gutenberg–Richter frequency–magnitude law, *Bull. seism. Soc. Am.,* **75**(6), 1681–1697.

Turcotte, D., 1997. *Fractals and Chaos in Geology and Geophysics,* Cambridge Univ. Press.

Vermylen, J.P. & Zoback, M.D., 2011. Hydraulic fracturing, microseismic magnitudes, and stress evolution in the Barnett Shale, Texas, USA, Tech. Rep.

Wiemer, S. & Wyss, M., 2000. Minimum magnitude of completeness in earthquake catalogs: examples from Alaska, the Western United States, and Japan, *Bull. seism. Soc. Am.,* **90**(4), 859–869.

Wiemer, S., McNutt, S.R. & Wyss, M., 1998. Temporal and three-dimensional spatial analyses of the frequency–magnitude distribution near Long Valley Caldera, California, *Geophys. J. Int.,* **134,** 409–421.

Woessner, J. & Wiemer, S., 2005. Assessing the quality of earthquake catalogues: estimating the magnitude of completeness and its uncertainty, *Bull. seism. Soc. Am.,* **95**(2), 684–698.

Zöller, G., 2013. Convergence of the frequency–magnitude distribution of global earthquakes: maybe in 200 years, *Geophys. Res. Lett.,* **40,** 3873–3877.