Are volcanic seismic b-values high, and if so when?

Citation for published version:
Roberts, NS, Bell, A & Main, I 2015, 'Are volcanic seismic b-values high, and if so when?' Journal of Volcanology and Geothermal Research. DOI: 10.1016/j.jvolgeores.2015.10.021

Digital Object Identifier (DOI):
10.1016/j.jvolgeores.2015.10.021

Published In:
Journal of Volcanology and Geothermal Research

Are volcanic seismic *b*-values high, and if so when?

Nick. S. Roberts[1], Andrew F. Bell[1], Ian G. Main[1]

[1]School of Geosciences, University of Edinburgh, James Hutton Road, Edinburgh, UK. EH9 3FE

Email: N.S.Roberts@sms.ed.ac.uk, a.bell@ed.ac.uk, ian.main@ed.ac.uk

1 Abstract

The Gutenberg-Richter exponent *b* is a measure of the relative proportion of large and small earthquakes. It is commonly used to infer material properties such as heterogeneity, or mechanical properties such as the state of stress from earthquake populations. It is 'well known' that the *b*-value tends to be high or very high for volcanic earthquake populations relative to *b*=1 for those of tectonic earthquakes, and that *b* varies significantly with time during periods of unrest. We first review the supporting evidence from of 34 case studies, and identify weaknesses in this argument due predominantly to small sample size, the narrow bandwidth of magnitude scales available, variability in the methods used to assess the minimum or cut-off magnitude *Mc*, and to infer *b*. Informed by this, we use synthetic realisations to quantify the effect of choice of the cut-off magnitude on maximum likelihood estimates of *b*, and suggest a new work flow for this choice. We present the first quantitative estimate of the error in *b* introduced by uncertainties in estimating *Mc*, as a function of the number of events and the *b*-value itself. This error can significantly exceed the commonly-quoted statistical error in the estimated *b*-value, especially for the case that the underlying *b*-value is high. We apply the new methods to data sets from recent periods of unrest in El Hierro and Mount Etna. For El Hierro we confirm significantly high *b*-values of 1.5-2.5 prior to the 10 October 2011 eruption. For Mount Etna the *b*-values are indistinguishable from *b*=1 within error, except during the flank eruptions at Mount Etna in 2001-2003, when 1.5<*b*<2.0. For the time period analysed, they are rarely lower than *b*=1. Our results confirm that these volcano-tectonic earthquake populations can have systematically high *b*-values, especially when associated with eruptions. At other times they can be indistinguishable from those of tectonic earthquakes within the total error. The results have significant implications for operational forecasting informed by *b*-value variability, in particular in assessing the significance of *b*-value variations identified by sample sizes with fewer than 200 events above the completeness threshold.

Keywords: *b*-value; volcano; seismology; completeness magnitude

2 Introduction

Volcanic earthquakes provide insight into physical processes acting at volcanoes, such as the mechanisms of deformation of the volcanic edifice and magma accumulation, and statistical analysis of earthquake catalogues are a key component of eruption forecasting methods (McNutt, 1996). Increased rates of earthquakes are a primary indicator of volcanic unrest, and changing locations of earthquake hypocentres can be used to map magma migration (Wiemer and Wyss, 2002). The frequency-magnitude distribution (FMD) of volcanic earthquakes can provide insight into the state of stress or material properties, and are a key component of most studies of volcanic seismicity.

Where the catalogue is completely reported, the FMD, commonly takes the form of a Gutenberg-Richter (GR) relation (Gutenberg and Richter, 1954):

$$\log(N) = a - bM, \qquad (1)$$

where *N* is the total number of earthquakes of magnitude equal to or greater than *M*, and *a* and *b* are real, positive constants characteristic of the specific catalogue. The parameter *a* is the logarithm of the number of earthquakes with $M \geq 0$, and is thus a measure of the seismicity rate of the region. The *b*-value represents the relative proportion of large and small events in the catalogue. It is best calculated or inferred using the maximum likelihood method (Aki, 1965), now used almost universally in earthquake seismology (Mignan and Woessner, 2012). Other methods such as a least squares fit of the data to equation 1 are known to produce a biased estimate (Naylor et al., 2010). In addition, if the bandwidth of data is narrow, or equivalently the sample is small, then it is easy to overestimate the underlying *b*-value (Main, 2000). Finally, the *b*-value may also be biased due to incorrect identification of the threshold for complete reporting, denoted *Mc* here (Mignan and Woessner, 2012). These and other sources of bias introduce an epistemic error to any inference from the data. In principle this should be accounted for in addition to the aleatory uncertainties inferred from the random error associated with measurement or statistical fluctuation in the data, but it is often neglected in studies of volcanic earthquake populations.

The Gutenberg-Richter form of the distribution holds, at least for small and intermediate events across a remarkable range of sizes and loading conditions, from laboratory experiments to volcanic and tectonic earthquakes (Main, 1996). In controlled laboratory tests, seismic *b*-values commonly change systematically with respect to a variety of controlling factors. These include the degree of material heterogeneity (Mogi, 1962), the level of applied stress (Scholz, 1968), the degree of stress concentration, i.e. the stress intensity normalised to the fracture toughness (Meredith and Atkinson, 1983), the chemical reactivity of the pore fluid (Meredith and Atkinson, 1983), and the pore fluid pressure (Sammonds et al., 1992). In nature other factors that affect the *b*-value systematically include the earthquake focal mechanism (Schorlemmer et al., 2005), the depth (Mori and Abercrombie, 1997), and the degree of coupling or strain partition between seismic and aseismic deformation at plate boundaries (Mazzotti et al., 2011).

The *b*-value for tectonic earthquakes, using best practice and large regional or global data sets, is commonly reported as taking values near unity (Frolich and Davis, 1993). In contrast the reported *b*-values from published studies of earthquake populations associated with volcanic unrest are commonly reported as being significantly higher than this, allowing for the random error expected for a *b*-value of unity (described in more detail below). The main question we address here is whether this difference is real or, at least to some extent, an artefact of the known sources of bias described above.

To examine this question we first use synthetic data to explore the effect of various factors on the estimated *b*-value, denoted $\tilde{b}$, and the underlying *b*-value, henceforth denoted *b*. Uncertainties in $\tilde{b}$ at one standard deviation, denoted $\sigma_{\tilde{b}}$, are estimated using the method of Shi & Bolt (1982), which correctly reflects the (approximately) Poisson 'counting errors' expected from sampling a whole number of events (Greenhough and Main, 2008). The advantage of using synthetic data is that we can distinguish between the random error $\sigma_{\tilde{b}}$, and the systematic error or bias $\tilde{b} - b$, or equivalently to errors of precision and accuracy respectively. We show how both depend intrinsically on the sample size. First we determine an optimum method of estimating the cut-off magnitude of complete reporting of events, *Mc,* for catalogues of different sizes, and then propose a formal workflow for the estimation of *Mc*. The proposed workflow is then applied to two volcanic seismic catalogues at Mount Etna and El Hierro as important examples of recently-active volcanic systems to address the questions: (a) are the *b*-values higher than 1? And (b) do they vary with time significantly outside the estimated margins of error? For these examples, *b* is remarkably stationary and similar to (~1) or only somewhat larger (1-1.5) than to those of tectonic earthquakes, except for specific transients where the *b*-value can be significantly greater than background at 95% confidence. The results presented here will provide greater confidence in identifying statistically-significant variations in *b*-value, and in identifying physical causes for this variability.

3 Review and synthesis of previous studies

In this section we extend the review of McNutt (2005), who summarised reported *b*-values and associated parameters such as source depth from 13 different volcanoes around the world. This review includes *b*-values as high as 3 in one case (McNutt, 2005). In Table 1 we extend this study to 21 volcanoes, and include a wider range of associated parameters, including: the number of events; the range of magnitudes used in the analysed catalogues; the methods used to calculate the completeness magnitude and fit the *b*-value; and the range of *b*-values reported in each study, including a typical value. Multiple studies use several methods for analysing *b*-value variations and thus the results are reported separately in Table 1, giving 34 separate results for comparison in this new synthesis. Information on all the different fields of data could not be found in all cases, e.g. how the threshold magnitude was estimated, resulting in some blank entries in Table 1.

The maximum reported *b*-values range between 1.4 and 3.5, with a peak at *b*=1.7 (Figure 1c). From Figures 1b there is no clear dependence on the magnitude and *b*-value. Bonnet et al. (2001) also found there was no direct dependence of the scaling exponent for fracture length on the scale of observation and that no significant trends could be determined in the type of faulting (Bonnet et al., 2001).

Figure 1 shows the distribution of *b*-values compared to the other variables in the study. There are no clear trends with depth (Figure 1a) or magnitude range or size (Figure 1b). However, there is a weak decreasing trend in the *b*-value as the number of events in the sample, *N*, increases (Figure 1c). The data only spans from 10 to 300 events covering just over one magnitude unit, with over half, (16 of 25) of the studies using catalogues with either 50 or 100 events. One further study (Ibanez et al., 2012) containing 7000 events reports a relatively high *b*-value of 1.57 that does not follow this trend. However, this study - and many others cited in Table 1 - use the Least Squares method to fit *b* or to check the results of the maximum likelihood estimation, introducing a known source of potential bias outlined in the introduction.

In summary this review has highlighted a significant variability in the reported values of *b*, and a significant variability in the methods of analysis used in the different studies. Typical *b*-values are usually in the range 1-1.2. They are never (for this list) less than one, and are occasionally very high (up to 3.5). The variability is much larger than any systematic trends, except that the *b*-value tends to decrease with increasing sample size. In this paper we use synthetically-generated data to address some of the most important origins of this variability, in particular the choice of threshold magnitude and the sample size.

4 Methods for analysis of Frequency-Magnitude Distributions

A variety of statistical methods have been used to model FMD's and to quantify whether those models are consistent with the observed data. Most methods involve modelling the proportion of the distribution above the completeness magnitude. Therefore there is a strong inter-dependence between estimates of the completeness magnitude and values of parameters of prospective FMD models. In this section we summarise the current methods used to address this problem.

4.1 Gutenberg-Richter parameters

There is a well-established literature that describes the merits of different statistical methodologies for FMD analysis. Methods involving regression on cumulative frequencies, or using least-squares regression, are known to give biased estimates of the *b*-value (Naylor et al., 2010) as they are known to give disproportionate weighting to higher magnitude events (Ghosh et al., 2008). The maximum likelihood technique has become standard in seismic hazard analysis (Mignan and Woessner, 2012). The data are assumed to be exponentially distributed (as in eq. 1) and the maximum possible magnitude is assumed to be at infinity (Aki, 1965). Physically, earthquakes must have a finite maximum size dependent on the size and strain limits within the Earth, but M_{max} is not well constrained by global data (Main et al., 2008; Holschneider et al., 2014). The maximum likelihood method weights each event equally and correctly allows for error structure of the data: in frequency data in the form of a Poisson distribution (Naylor et al., 2010). Formally, the maximum likelihood estimate of the *b*-value is:

$$\tilde{b} = \frac{\log_{10} e}{\overline{M} - (Mc - \Delta M/2)} \qquad (2)$$

where $\tilde{b}$ is the estimate of the *b*-value, $\overline{M}$ is the mean magnitude, M_c is the completeness magnitude, and ΔM is the magnitude bin size of the histogram (Aki, 1965). Aki also showed the uncertainty on this estimate at one standard deviation (67% confidence) can be approximated to:

$$\sigma_{\tilde{b}} = \frac{\tilde{b}}{\sqrt{N_c}} \qquad (3)$$

Where N_c is the number of events in the complete part of the catalogue, or 1.96 times this value at 95% confidence.

A summary study by Marzocchi & Sandri, (2003), tested two further improvements on this estimation of *b* using binned magnitudes, equation (4) (Bender, 1983), and an improved uncertainty estimate (eq. 5) (Shi and Bolt, 1982; Marzocchi and Sandri, 2003):

$$\tilde{b} = \frac{1}{\ln 10[\hat{\mu} - (M_c - \Delta M)]} \qquad (4)$$

$$\sigma_{\tilde{b}} = 2.30\tilde{b}^2 \sqrt{\frac{\sum_{i=1}^{N}(M_i - \hat{\mu})^2}{N_c(N_c - 1)}}$$

where $\hat{\mu}$ is the average magnitude of the sample, and ΔM is the binned magnitude width. The *b*-value is relatively insensitive to the upper magnitude cut-off, so assuming an infinite cut-off in deriving equations (3) and (5) does not introduce a significant bias. However, in both cases the quoted error is formally conditional on the choice of *Mc*, which in practice must be estimated. This introduces an implicit source of bias that can be positive or negative. In this paper we will demonstrate that this additional source of uncertainty is comparable to or can greatly exceed the estimates from equations (3) or (5).

4.2 Calculating the completeness magnitude

Most studies apply a lower threshold or cut-off magnitude, *Mc*, above which the catalogue can be regarded as completely recorded (Wiemer and Wyss, 2000). *Mc* is the lowest magnitude at which 100 per cent of earthquakes in a space-time volume are detected (Rydelek and Sacks, 1989; Woessner and Wiemer, 2005; Mignan and Woessner, 2012). Earthquakes with smaller magnitudes are less likely to be completely reported when their amplitude becomes smaller than that of the ambient noise. This introduces a high-pass filter to the FMD, which could in principle be modelled and fitted to the data. However, this is rarely (if ever) done explicitly. In practice most studies assume *Mc* is the magnitude at which the log(cumulative frequency)-magnitude curve departs from a linear trend of eq. 1. There are three main techniques commonly used to estimate this magnitude, namely the Maximum Curvature (MaxC) method, the Goodness-of-Fit test (GFT) (Wiemer and Wyss, 2000) and *b*-value stability (BVS) method (Cao and Gao, 2002).

The MaxC method calculates the highest value of the first derivative of the cumulative frequency-magnitude curve. In practice this matches the frequency-magnitude bin with the highest number of events (Figure 2a). The main limitation of this method is that it will systematically underestimate *Mc* unless there is a sharp transition between the incomplete and complete portion of the catalogue, as illustrated in Figure 2a.

The GFT method calculates *Mc* by comparing the observed FMD with a synthetic one. The best-fit distribution is calculated for trial cut-off magnitudes using the maximum-likelihood estimates of *a*- and *b*-values of the observed dataset. The residuals between the data and the best fit distribution are then calculated as a function of cut-off magnitude (Figure 2b). The completeness threshold, *Mc,* is selected as being the first magnitude above which the residual between the synthetic straight line fit model and observed data falls within a 95% confidence window. If 95% confidence cannot be obtained then a 90% confidence window can be used as a compromise. This method tends to give systematically low values for *Mc* although not as low as the MaxC method (Wiemer and Wyss, 2000).

The BVS method simply evaluates the estimated *b*-value as a function of the cut-off magnitude. The assumption here is that $\tilde{b}$ will initially increase as the cut-off magnitude increases, until the cut-off magnitude equals *Mc* after which $\tilde{b}$ will stabilise. The inferred *b*-value is deemed to have stabilised once the average $\tilde{b}$ for the five successive cut-off magnitudes falls within error of the selected cut-off magnitude (Figure 2c). The BVS method tends to have high *Mc* values relative to other methods (Woessner and Wiemer, 2005) and consequently higher $\tilde{b}$ values.

5 Results for Synthetic catalogues

5.1 Generating synthetic catalogues

We now evaluate which of the three methods for calculating the *Mc* is the most accurate and reliable, by generating synthetic catalogues with known *Mc* and *b*-value, but different forms of the cut off below *Mc*. As a benchmark check we first generated synthetic data to determine $\tilde{b}$ and $\sigma_{\tilde{b}}$ for *b*=1 and *b*=2 as a function of the complete sample size N_c, conditioned on an exact value for *Mc*. This provided a good match to Fig. 1a,b of Marzocchi and Sandri (2003). However, in reality *Mc* is not known independently a priori. Ideally we would hope the incremental FMD would have a sharp and easily distinguishable peak at *Mc*, defining the lower limit of the complete catalogue (Figure 3a). In reality the peak of the distribution is often curved and much broader due to the complexity of the signal to noise ratio at the recording stations, and of locating and calculating magnitudes for small events, so defining *Mc* can be much more challenging (Figure 3b). This introduces an additional source of uncertainty that is the prime focus of the current paper.

To test each of the three methods, we use two end-member scenarios. The first has a sharp peak (Figure 3a) and the second a broader peak (Figure 3b). Both catalogues have *Mc* set to 1.0. The complete part of both catalogues was created by randomly generating individual events from an ideal parent Gutenberg-Richter law distribution with a *b*-value of 1.0. For the sharp-peaked distribution the incomplete part of the catalogue was generated using a filter with a linear slope of 3, for values below *Mc*=1.0 decaying to zero probability at M=0. For the broad-peaked distribution a GR distribution was used to generate events all the way down to M=0. The probability function shown in Figure 3c was then applied as a filter to remove events below the known threshold *Mc*=1.0, until the required number of events were left in the complete catalogue.

To examine the role of catalogue size, catalogues were generated with a complete size of 50, 100, 200, 500, 1000 and 5000 events. Finally the *b*-value was varied from a typical tectonic value of 1.0 to a significantly high *b*-value of 2.0, to test whether each method can reliably calculate *Mc* and inferred *b*-values for the case that the underlying *b*-value is high.

For each catalogue size, *b*-value, and distribution shape; 100 catalogue were iteratively generated, and the estimated *Mc* and *b*-values determined using the different methods described in section 4. A bin size *ΔM* of 0.1*M* is used throughout. Figure 3 shows both the average catalogue (solid line) and the spread of the outcomes associated with the finite sample size (dashed lines).

5.2 Synthetic Results

5.2.1. Sharp-peaked distribution

In this case the simulations of Figure 4 demonstrate that the MaxC method performs the best in terms of calculating *Mc*, closely followed by the BVS method. The GFT performs adequately for N_c=5000 but fails when N_c=50 as for over 90% of the catalogues *b* is not even calculated correctly within ±1.0 of the known value. When *b*=1 and N_c=5000, MaxC and BVS both correctly lead to a correct calculation of *b* with <0.01 error.

5.2.2. Broad-peaked distribution

Figure 5 shows histograms of the best estimates of *Mc* for the three methods, for different catalogue sizes and *b*-values, for the case of the broad-peaked distribution. When N_c=50 for both *b*=1 and *b*=2, MaxC and BVS both systematically underestimate *Mc*, because very few events have a greater magnitude than *Mc*=1.0 (Figure 6). Both MaxC and BVS methods give results with some scatter, centred on *b*=1, but several iterations had significantly higher *b*-values of 2 or above. Both methods perform poorly when *b*=2, as there too few events in the catalogue, with median values of $\tilde{b}$≈1.5. The GFT over-estimates *Mc* when *b*=1 but appears to give a reasonable estimate when *b*=2. However, the 95% confidence is only reached when *Mc* is very close to the maximum magnitude and thus the complete catalogue size is very small. This results in the inferred *b*-values being very high for both *b*=1 and *b*=2.

When N_c=5000 it becomes apparent that MaxC is not a good method for broad-peaked distributions. For *b*=1, *Mc* is heavily underestimated, with a median value of *Mc*=0.4, and resulting $\tilde{b}$-values all less than *b*=1. These underestimates are amplified when *b*=2 with median values of *Mc*=0.4 and $\tilde{b}$≈1.3. The GFT performs much better for both *b*=1 and *b*=2 however it gives a conservative estimate for both. The BVS method performs the best for a broad-peaked distribution, giving only a slightly conservative estimate of *Mc* with a median value of *Mc*=0.9 for *b*=1 and *b*=2. The BVS method returns the correct $\tilde{b}$=1.0 in over 80 iterations. The median value for *b*=2 is also approximately correct, however there is a very broad range of results with a slight skew towards values higher than *b*=2.0. This is a very large catalogue and the BVS method is clearly the best when *b*=2. Our results show that it is intrinsically more difficult to calculate high *b*-values, however it is possible to find an estimate with a correct median value with the BVS method, albeit with a large spread in $\tilde{b}$.

5.2.3. Comparison of method performance

For a sharp-peaked distribution the MaxC method correctly calculates *Mc* the highest proportion of times for both high and low *b*-values. This outcome is not surprising as the MaxC method finds the magnitude bin with the highest number of events that, trivially, is the *Mc* set by the parent distribution.

The BVS method performs almost as well as the MaxC method for low *b*-values, but with higher *b*-values the method returns too high estimates of *Mc*. However, as long as for larger catalogue sizes the BVS method continues to return good estimates of the *b*-value. The GFT method does not work with small catalogues as the 95% confidence threshold is only reached when the *Mc* is very close to the maximum magnitude event, therefore there are a minimal number of earthquakes left in the catalogue, and thus the uncertainty is very large. For larger catalogues GFT performs much better. However for both *b*=1 and *b*=2, using the GFT-calculated value of *Mc* results in fewer correct calculations of $\tilde{b}$ than the MaxC and BVS methods. Therefore we consider it to be the least-well performing method. For *b*=2 the steeper slope of the complete catalogue leads to a larger spread of calculated $\tilde{b}$-values for all three methods than for *b*=1. This is due to the random scattering of data due to sampling which has a greater influence on the FMD at high *b* compared to low *b*-values, and is not inherently linked to any of the methodologies.

Figure 7 and Figure 8 compare the performance of the different methods for the case of a broad-peaked distribution, using the mean and standard deviations of $\tilde{b}$ calculated from the data in Figure 6. For both *b*-values the GFT method does not reliably calculate *Mc,* resulting in a biased estimate of the *b*-value. For N_c≤500 the correct *b*-value is calculated within the statistical error, but the distribution is heavily skewed towards high *b*-values, meaning that this method performs sub-optimally for these small catalogue sizes. However for larger catalogues (N_c=1000 & 5000) the GFT method does calculate accurate *b*-value estimates for both *b*=1 and *b*=2. The MaxC method returns a systematically-low estimate of *Mc* for all catalogue sizes, resulting in under-estimates of the *b*-value for both *b*=1 and *b*=2. We conclude that it is not an appropriate method for calculating *Mc* for a broad-peaked distribution.

The estimates of *Mc* returned by the BVS method increase in accuracy with catalogue size. For N_c≥200 the BVS method correctly calculates *Mc* within the 95% confidence limits for both *b*=1 (Figure 7) and *b*=2 (Figure 8). When *b*=1 and the catalogue size is N_c≥200, the 95% confidence spread around the true *b*-value is very small, ±0.25. Using the BVS method with smaller catalogue sizes can result in *b*-value estimates as high as 2 even with *b*=1 (Figure 7). This observation suggests that care must be taken to not over-interpret high *b*-values calculated for small catalogues sizes. For *b*=2, the standard deviation of results is independent of catalogue size at about ±0.75. However, the median and mean of the $\tilde{b}$-value estimates tend towards the parent *b*=2 as catalogue size increases. Again for N_c≥200 for *b*=2 the BVS method estimates $\tilde{b}$ to within 95% confidence.

In terms of defining a threshold minimum complete catalogue size, when N_c≥500 our results show both *b*=1 and *b*=2 can be estimated accurately and precisely (Figure 7). For N_c=100 the statistical error in estimating *b*=1 is large, indicating a lack of precision, and for *b*=2 the average and median values are significantly below 2, indicating a residual bias. However, a threshold of 500 for completely-reported events is a relatively large number for many practical applications. From the results in Figure 7, a pragmatic choice of N_c=200 is an acceptable threshold for a trade–off between accuracy, precision, and realistic catalogue size.

5.3 A proposed workflow for the calculation of *Mc*

Informed by this analysis, we propose a workflow for analysing the FMD of volcanic earthquake catalogues (Figure 9). As discussed above, we considered that the minimum catalogue size for reliable estimation of the *b*-value is N_c=200.

First, *Mc* is estimated using each of the MaxC, GFT and BVS methods. If all three *Mc* estimates agree within ±0.1, the FMD can be modelled by a sharp-peaked distribution, and so the MaxC estimate of *Mc* should be used. If the *b*-value calculated using this *Mc* has an error of ≤±0.25 it should be considered to be reliable. An error of >±0.25 makes it difficult to interpret the *b*-value and may indicate an unreliable estimate of *Mc*.

If the three estimates of *Mc* vary by ≥0.1, or the *b*-value calculated from the MaxC estimate of Mc is ≥0.25, we recommend that the BVS method should be used. If the resulting *b*-value has an error of ≤0.25 it should be considered to be reliable. If this is not the case, the GFT analysis should be used. If a *b*-value with an error of ≤0.25 cannot be obtained using any of the 3 methods, we argue that the catalogue is too small for reliable FMD analysis. If the complete catalogue has over 5000 events and the *b*-value uncertainty is still too high, it is likely that the FMD is not consistent with an underlying Gutenberg-Richter distribution.

For the analysis of variations in FMDs, a large volcanic earthquake catalogue can be split on the basis of spatial or temporal windows, and this workflow applied to each sub-catalogue in turn. However, the same minimum complete catalogue size and reliability criteria rules apply to sub catalogues too.

5.4 Error introduced from the completeness magnitude

We now use the workflow of Figure 9 to consider the relative effect of *Mc* estimation for catalogues of different size on the accuracy and precision of the estimate of $\tilde{b}$ for the case of a broad-peaked distribution. Figure 10 shows a histogram of the $\tilde{b}$ for 100 catalogue realizations with *b*=2, along with examples of its standard deviation $\sigma_{\tilde{b}}$ estimated from equation 5. $\tilde{b}$ is beyond 1 standard deviation of *b* in more than 1/3 of the cases, indicating a significant epistemic error in the estimation. We show in this section that this is due to the bias $\tilde{b} - b$ in the finite-sized sample. The error due to calculating *Mc* for individual realisations is illustrated as a blue bar at one standard deviation in Figure 9. The median $\tilde{b}$ is close to the true value (the central blue dot is near the vertical dashed line), so the residual bias due to estimating *Mc* is near zero for a large population of trials. However, the standard deviation in the error due to *Mc* is much larger than the average statistical error for similar *b*-values (the black error bars).

To quantify this error in the general case, we ran many simulations for different values of *b* and N_c, with the results shown in Figure 11. Figure 11a shows the average statistical error from equation (5), Figure 11b the average error in $\tilde{b}$ due to propagating uncertainties in estimating *Mc* as illustrated by the blue horizontal error bar in Figure 10, and Figure 11c the ratio of the two. The ratio was calculated 5 times for each of 15 catalogue sizes between 50-5,000 events and *b*-values of 0.5, 1.0, 1.5, 2.0 & 3.0, with the average value indicated by the colour scheme in Figure 11. The ratio varies between 1.2 and a factor 14 or so for the range studied, implying that the sample bias error is always greater than, and often much greater than the estimated statistical uncertainty in $\tilde{b}$ from equation (5). This finding means that the statistical error commonly used on its own to quantify the $\tilde{b}$-value uncertainty is not an adequate description of the total error, though it approaches the total error for large numbers of events and low underlying *b*-values. In Figure 11c the ratio can reach an order of magnitude for *b*>2 and event numbers above 1000. This is because the statistical error $\sigma_{\tilde{b}}$ is very small when N_c is large. However the sample bias also increases with N_c for high *b*. This somewhat counter-intuitive result is because the magnitude range over which *Mc* can be calculated is much smaller at low N_c than at high N_c, so the uncertainty is bounded to a greater degree at low N_c, and hence reduces at low N_c. The template of Figure 11c can be used empirically to determine a more appropriate error for *b*-value estimation.

5.5 Application to volcanic catalogues

We apply our proposed workflow to earthquake catalogues for Mount Etna volcano, Sicily (Murru et al., 1999; Murru et al., 2005; Murru et al., 2007) and El Hierro volcano, Canary Islands (Ibanez et al., 2012; López et al., 2012; Becerril et al., 2013; Marti et al., 2013; García et al., 2014) to test the reliability of any previously reported variations in *b*-values. This is simply to compare results from the proposed workflow to previous volcanic *b*-value's and not to make any interpretation about the behaviour of the volcanos.

We analyse the Instituto Geográfico Nacional (IGN) earthquake catalogue for El Hierro between July 2011 and December 2013, a period associated with significant seismic activity associated with magma emplacement, and including a submarine eruption that began on 10th October 2011 (Ibanez et al., 2012; López et al., 2012). The catalogue contains over 20,000 events, and so it is possible to subdivide it into several phases to analyse *b*-value variations. Figure 12 shows how each phase is defined by changes in event rate, with the first three phases following the scheme of Ibanez et al. (2012). The start of each phase is defined as midnight at the start of the selected day, however, if necessary the resolution of the boundaries can be increased as most catalogues give event time to the nearest second. All phases have over 200 events at or above *Mc*, thus the catalogues should be large enough to calculate reliable $\tilde{b}$-values following the synthetic analysis. At this stage the catalogue is simply divided temporally, so earthquakes may originate from different portions of the volcanic edifice. Should this occur, the *b* estimate may represent an average between sub-catalogues representative of different processes or stress conditions.

The results of applying our proposed workflow to the El Hierro catalogue are shown in Figure 12. These show a very high *b*-value of $\tilde{b}$=2.39±0.10 before the onset of the eruption, followed by a fluctuating $\tilde{b}$-value between 1-1.5 for the remainder of the catalogue. $\tilde{b}$-value uncertainties are determined using equation 5. The $\tilde{b}$-value is always above 1 within these statistical errors. These results are similar to those of Ibanez et al. (2012), who reported a *b*-value before the eruption of 2.25±0.05 followed by values of *b*=1.34±0.04 and *b*=1.12±0.01 for the second and third phases respectively (Ibanez et al., 2012). However, the Ibanez study used the 90% Goodness-of-fit method to estimate *Mc,* and least-squares regression to estimate *b*. The *Mc* values they report are significant under-estimates, and this means that the biased least-squares *b*-value estimates are, coincidently, close to the values reported here.

We also analyse the Istituto Nazionale di Geofisica e Volcanologia (INGV) earthquake catalogue for Mt Etna between January 1999 and December 2014. This catalogue spans several eruptive episodes, including the 2001 and 2002-03 flank eruptions and more recent paroxysmal activity at the new South East Crater. The catalogue contains 8000 events, with an event rate that is more stable through time than the El Hierro catalogue (Figure 12 and Figure 13). We divide the catalogue into 10 sub-phases on the basis of changes in earthquake rate, with each phase ideally containing between 200-5000 events.

Figure 13 shows the $\tilde{b}$-values calculated for Mt Etna using our proposed workflow. During the 2001 and 2002-03 flank eruptions the $\tilde{b}$-value is 1.5 or greater. However from the end of the 2002-03 flank eruption, the $\tilde{b}$-value appears to have stabilised at 1.0±0.2. Murru et al. (2007) analysed the spatial distribution of the *b*-value at Mt Etna between 1999 and 2005 and found an average of approximately 1.5, with an increase in average *b*-value with depth from *b*=1.2 to *b*=1.9.

Although the $\tilde{b}$-values for Mt Etna from 2004 onwards are close to 1.0 and there is no systematic trend in values, the $\tilde{b}$-values do not encompass *b*=1 within error for over half of the sub-phases in Figure 13. As the Shi & Bolt $\tilde{b}$-value uncertainty (eq. 5) defines one standard deviation error in the $\tilde{b}$-value we would expect 68% of the calculated *b*-values to capture *b*=1 within error if the underlying *b*-value is stationary. We might then conclude that the hypothesis that *b*=1 can be rejected at this confidence level. However, we have shown that the total error, including sample bias, can be significantly underestimated in Figure 11.

Accordingly we now apply the contour plot for the error multiplication values in Figure 11c to estimate a more realistic total error for our calculated *b*-value. For the 2011-13 El Hierro catalogue (Figure 14a) the high *b*-values at the start of the catalogue now have dramatically increased errors, and 3 of the 6 following *b*-values that sat between $1>\tilde{b}>1.5$ now lie within 1 standard deviation error around *b*=1.0. Using the Shi & Bolt uncertainty for the 2004-2014 Etna catalogue, the estimated $\tilde{b}$-values for only 2 of 10 phases (20%) lie within one standard deviation of *b*=1.0. However, once the modified error is applied to the catalogue (Figure 14b), the estimated $\tilde{b}$-value for 6 of the 10 phases (60%) lie within 1 standard deviation of *b*=1.0. The high *b*-values associated with the 2001 and 2002-03 flank eruptions also increase in error and could be consistent with *b*-value of no more than 1.5. The *b*-values for 3 of the 10 phases do not lie within 2 standard deviations of *b*=1 using the modified error. Therefore it would be hard to reject the hypothesis that *b* is a constant near unity for these phases, except at marginal significance.

6 Conclusions

The almost axiomatic inference that *b*-values are systematically higher for volcanic earthquakes is based on data and methodology that are often insufficient to address the question, notably the very small sizes of the samples used, the methods of parameter estimation and the different methods used to infer the completeness magnitude *Mc*. The Maximum Curvature method is simple, and can be used when a catalogue has a sharp peak in the discrete data. Otherwise the *b*-value stability method is the most favourable. If that does not generate a *b*-value with a standard error ≤0.25 the Goodness-of-Fit method can be used as a third option. If a stable value of *b* cannot be obtained then the sample size must be increased in space and/or time. Our results imply a pragmatic minimum of 200 events above *Mc* is generally needed. From further simulations, we also recommend a minimum of 500 events when dealing with raw incomplete catalogues before this workflow can be applied. This logic is captured in a new workflow for estimating *Mc*. Even when this best practice is followed, there can be a significant residual error from calculating *Mc* in a single sample. This is comparable to or much greater than the statistical error, particularly for higher values of *b*. Nevertheless, when this is accounted for we confirm *b*-values for the El Hierro catalogue are generally higher than 1 at a confidence level of 95%, and may be significantly higher during eruptive phases. For Mount Etna the hypothesis *b*=1 can be rejected for only two time intervals, one associated with a flank eruption. We conclude seismic *b*-values can be high for volcanic earthquake populations, especially when associated with eruptive phases. Otherwise they appear to be very close to those obtained for tectonic earthquakes at the 95% confidence level.

7 Acknowledgements

Nick Roberts is a NERC funded PhD student at the University of Edinburgh. We thank the Instituto Geográfico Nacional and INGV Sezione di Catania for making the seismic catalogues for El Hierro and Mount Etna respectively available, Mark Naylor for providing feedback on earlier drafts of the paper, and three anonymous reviewers, and Jackie Caplan-Auerbach for improving the paper with their critical feedback.

8 References

Aki, K., 1965. Maximum Likelihood Estimate of *b* in the Formula log *N = a - bM* and its confidence limits. Bulletin of the Earthquake Research Institute, 43: 237-239.

Becerril, L., Cappello, A., Galindo, I., Neri, M. and Del Negro, C., 2013. Spatial probability distribution of future volcanic eruptions at El Hierro Island (Canary Islands, Spain). Journal of Volcanology and Geothermal Research, 257: 21-30.

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

Bonnet, E., Bour, O., Odling, N.E., Davy, P., Main, I., Cowie, P. and Berkowitz, B., 2001. Scaling of fracture systems in geological media. Review of Geophysics, 29(3): 347-383.

Bridges, D.L. and Gao, S.S., 2006. Spatial variation of seismic b-values beneath Makushin Volcano, Unalaska Island, Alaska. Earth and Planetary Science Letters, 245: 408-415.

Cao, A. and Gao, S.S., 2002. Temporal variation of seismic b-values beneath northeastern Japan island arc. Geophysical Research Letters, 29(9): 1-3.

Centamore, C., Patane, G. and Tuve, T., 1999. Maximum entropy estimation of b values at Mt. Etna: comparison with conventional least squares and maximum likelihood results and correlation with volcanic activity. Annali Di Geofisica, 42(3): 515-528.

Farrell, J., Husen, S. and Smith, R.B., 2009. Earthquake swarm and b-value characterization of the Yellowstone volcano-tectonic system. Journal of Volcanology and Geothermal Research, 188: 260-276.

Frolich, C. and Davis, S.D., 1993. Telesiesmic b values; Or, Much Ado About 1.0. Journal of Geophysical Research, 98(No. B1): 631-644.

García, A., Fernandez-Ros, A., Berrocoso, M., Marrero, J.M., Prates, G., De la Cruz-Reyna, S. and Ortiz, R., 2014. Magma displacements under insular volcanic fields, applications to eruption forecasting: El Hierro, Canary Islands, 2011–2013. Geophys. J. Int.

Ghosh, A., Newman, A.V., Thomas, A.M. and Farmer, G.T., 2008. Interface locking along the subduction megathrust from b-value mapping near Nicoya Peninsula, Costa Rica. Geophysical Research Letters, 35(L01301).

Greenhough, J. and Main, I., 2008. A Poisson model for earthquake frequency uncertainties in seismic hazard analysis. Geophysical Research Letters, 35.

Gutenberg, B. and Richter, C.F., 1954. Seismicity of the Earth, 2nd ed., 310.

Holschneider, M., Zoller, G., Clements, R. and Schorlemmer, D., 2014. Can we test for the maximum possible earthquake magnitude? J. Geophys. Res. Solid Earth, 199: 2019-2028.

Ibanez, J.M., De Angelis, S., Diaz-Moreno, A., Hernandez, P., Alguacil, G., Posdas, A. and Perez, N., 2012. Insights into the 2011-12 submarine eruption off the coast of El Hierro (Canary Islands, Spain) from statistical analyses of earthquake activity. Geophys. J. Int., 191: 659-670.

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.

López, C., Blanco, M.J., Abella, R., Brenes, B., Rodríguez, V.M.C., Casas, B., Cerdeña, I.D., Felpeto, A., Villalta, M.F., Del Fresno, C., García, O., García-Arias, M.J., García-Cañada, L., Moreno, A.G., Gonzalez-Alonso, E., Pérez, J.G., Iribarren, I., López-Díaz, R., Luengo-Oroz, N., Meletlidis, S., Moreno, M., Moure, D., Pablo, J.P., Rodero, C., Romero, E., Sainz-Maza, S., Domingo, M.A.S., Torres, P.A., Trigo, P., Villasante-Marcos, V., de Villalta, M.F. and de Pablo, J.P., 2012. Monitoring the volcanic unrest of El Hierro (Canary Islands) before the onset of the 2011–2012 submarine eruption. Geophysical Research Letters, 39(13).

Main, I., 1996. Statistical Physics, Seismogenesis, and Seismic Hazard. Review of Geophysics, 34(4): 433-462.

Main, I., 2000. Apparent Breaks in Scaling in the Earthquake Cumulative Frequency-Magnitude Distribution: Fact or Artifact? Bulletin of the Seismological Soceity of America, 90(1): 86-97.

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., Li. L, McCloskey, J. and Naylor, M., 2008. Effect of the Sumatran mega-earthquake on the global magnitude cut-off and event rate. Nature Geoscience, 1: 142.

Marti, J., Pinel, V., Lopez, C., Geyer, A., Abella, R., Tarraga, M., Blanco, M.J., Castro, A. and Rodriguez, C., 2013. Causes and mechanisms of El Hierro submarine eruption (2011-2012) (Canary Islands). JGR: 1-47.

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.

Mazzotti, S., Leonard, L.J., Cassidy, J.F., Rogers, G.C. and Halchuk, S., 2011. Seismic hazard in western Canada from GPS strain rates versus earthquake catalog. Journal of Geophysical Research, 116(B12310): 1-17.

McNutt, S.R., 1996. Seismic Monitoring and Eruption Forecasting of Volcanoes: A Review of the State-of-the-Art and Case Histories. 9-146.

McNutt, S.R., 2005. Volcanic Seismology. Annual Review of Earth and Planetary Sciences, 33(1): 461-491.

Meredith, P.G. and Atkinson, B.K., 1983. Stress corrosion and acoustic emission during tensile crack propagation in Whin Sill dolerite and other basic rocks. Geophys. J. R. astr. Soc., 75(1): 1-21.

Mignan, A. and Woessner, J., 2012. Estimating the magnitude of completeness for earthquake catalogs. Community Online Resource for Statistical Seismicity Analysis: 1-45.

Mogi, K., 1962. Magnitude frequency relations for elastic shocks accompanying fractures of various materials and some related problems in earthquakes. Bull. Earthquake Res. Inst. Univ. Tokyo, 40: 831-853.

Mori, J. and Abercrombie, R.E., 1997. Depth dependence of earthquake frequency-magnitude distributions in California: Implications for rupture initiation. Journal of Geophysical Research, 102: 15081-15090.

Murru, M., Console, R., Falcone, G., Montuori, C. and Sgroi, T., 2007. Spatial mapping of the *b* value at Mount Etna, Italy, using earthquake data recorded from 1999 to 2005. J. Geophys. Res., 112(B12303).

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).

Murru, M., Wyss, M. and Privitera, E., 1999. The locations of magma chambers at Mt. Etna, italy, mapped by b-values. Geophysical Research Letters, 26(16): 2553-1556.

Naylor, M., Orfanogiannaki, M. and Hart, D., 2010. Exploratory data analysis: magnitude, space, and time. Community Online Resource for Statistical Seismicity: 1-42.

Novelo-Casanova, D.A., Martinez-Bringas, A. and Valdes-Gonzalez, C., 2006. Temporal variations of Qc-1 and b-values associated to the December 2000–January 2001 volcanic activity at the Popocatepetl volcano, Mexico. Journal of Volcanology and Geothermal Research, 251: 347-358.

Patane, D., Caltabiano, T., Del Pezzo, E. and Gresta, S., 1992. Time variation of B and Qc at Mt. Etna (1981-87). Physics of the Earth and Planetary Interiors, 71: 137-140.

Power, J.A., Wyss, M. and Latchman, J.L., 1998. Spatial variations in the frequency-magnitude distribution of earthquakes at Soufriere Hills Volcano, Montserrat, West Indies Geophysical Research Letters, 29(19): 3652-3656.

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

Sammonds, P.R., G., M.P. and Main, I.G., 1992. Role of pore fluids in the generation of seismic precursors to shear fracture. Nature, 359: 228-230.

Sanchez, J.J., Gomez, J.A., Torres, P.A., Calvache, M.L., Ortega, A., Ponce, A.P., Acevedo, A.P., Gil-Cruz, F., Londono, J.M., Rodriguez, S.P., Patino, J.D.J. and Bohoroquez, O.P., 2005. Spatial mapping of the b-value at Galeras volcano, Columbia, using earthquakes recorded from 1995 to 2002. Earth Sci. Res. J., 9(1): 30-36.

Sanchez, J.J., McNutt, S.R., Power, J.A. and Wyss, M., 2004. Spatial Variations in the Frequency-Magnitude Distribution of Earthquakes at Mount Pinatubo Volcano. Bullertin of Seismological Society of America, 94(2): 430-438.

Scholz, C.H., 1968. The frequency-magnitude relation of microfracturing in rock and its relation to earthquakes. BSSA, 58(1): 399-415.

Schorlemmer, D., Wiemer, S. and Wyss, M., 2005. Variations in earthquake-size distribution across different stress regimes. Nature, 437(7058): 539-542.

Shi, Y. and Bolt, B.A., 1982. The standard error of the magnitude-frequency b value. Bulletin of the Seismological Soceity of America, 72(5): 1677-1687.

Wiemer, S. and McNutt, S.R., 1997. Variations in the frequency-magnitude distribution with depth in two volcanic areas: Mount St. Helens, Washington, and Mt. Spurr, Alaska. Geophysical Research Letters, 24(2): 189-192.

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

Wiemer, S. and Wyss, M., 2002. Mapping spatial variability of the frequency-magnitude distribution of earthquakes. Advances in Geophysics, 45: 259-302.

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.

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., Shimazaki, K. and Wiemer, S., 1997. Mapping active magma chambers by b values beneath the off-Ito volcano, Japan. Journal of Geophysical Research, 102(B9): 20413-20422.

9 Tables

Table 1 - Compilation of *b*-values and range of magnitudes for volcanic seismic catalogues

Reference	Volcano	Dates	Depth, km	*N*	Method *Mc*	Mag. range	Method *b*	b_{min}	b_{typ}	b_{max}
(Jacobs and McNutt, 2010)	Augustine	2000 - 2006	-2-0	100	ZMAP	-	MLE	0.8	1.4	2.1
(Jacobs and McNutt, 2010)	Augustine	17/11/05 - 10/12/05	-2-0	~250	ZMAP	-0.1-0.7	MLE	-	-	1.85
M. Wyss (written comm.)	Coso		0.8-3					-	-	1.7
(Ibanez et al., 2012)	El Hierro	19/7/11 - 16/9/11	8-16	7000+	90GFT	1.3-2.7	LS	1.12	1.57	2.25
(Ibanez et al., 2012)	El Hierro	19/07/2011	8-16	200	90GFT	-	LS	0.75	1.25	2.55
(Marti et al., 2013)	El Hierro	14/8/11 - 18/8/11	8-16	-	-	-	MLE	0.8	1.1	2.3
(Ibanez et al., 2012)	El Hierro	19/7/11 - 28/7/11	8-16	-	90GFT	1.5-2.6	LS	0.81	1.2	3.01
(Patane et al., 1992)	Etna	1984	-	200	-	2.8-	MLE	0.8	1.1	1.7
(Patane et al., 1992)	Etna	29/3/1983 - 6/8/1983	-	-	-	2.5-	MLE	0.7	1.0	2.1
(Murru et al., 1999)	Etna	-	9-15	50	MaxC	2.5-	MLE	1.4	1.5	3.5
(Centamore et al., 1999)	Etna	1/1/1990 - 31/12/92	-	100	-	2.3-5.1	LS	0.5	1.2	1.9
(Centamore et al., 1999)	Etna	1/1/1990- 31/12/92	-	100	-	2.3-5.1	MLE	0.9	1.1	1.7
(Murru et al., 2007).	Etna	July - Aug 2001	0-2	50	GFT	2.6-3.5	MLE	0.7	1	2.6
(Murru et al., 2005)	Etna	July - Aug 2001	0-12	50	90GFT	2.6	MLE	0.8	1.5	2.50
(Murru et al., 2007)	Etna	Aug 1999 - Dec 2005	1-3	100	90GFT	2.5	MLE	0.7	1.0	1.86
(Sanchez et al., 2005)	Galeras	Sep 1995 - Jun 2002	0-2	300	-	1.2-2.8	MLE	0.65	1.0	1.4
(Jolly and McNutt, 1999)	Katmai	-	6-8	-	-	-	-	1.0	1.3	1.6
(Wyss et al., 2001)	Kilauea	-	4-7,20	-	-	-	-	-	-	1.9
(Wyss et al., 2001)	Kilauea	1979 - 1997	4-7	50	-	1.8-2.6	MLE & LS	0.6	1.0	1.73
(Wiemer et al., 1998)	Long Valley	1989 - 1998	1-11	150	MaxC	1.3-	MLE	1.1	1.4	2.0
(Jolly and McNutt, 1999)	Mageik	Sep 1996 - April 1997	0-5	-	-	-	WLS	1.0	1.5	2.0
(Bridges and Gao, 2006)	Makushin	July 1996 - April 05	0-8	50	74GFT	0.9-3.9	MLE	0.73	1.21	2.03
(Wiemer et al., 1998)	Mammoth Mtn.	1989 - 1990.5	3-4,7-9	150	MaxC	1.3-	MLE	0.95	1.2	1.6
(Jolly and McNutt, 1999)	Martin/Mageik	Sep 1996 - April 1997	-2-10	-	-	0.7-4.5	WLS	-	-	1.56
(Wiemer and McNutt, 1997)	Mount Spurr	1991 - 1995	2.3-4.5	100	Inspection	0.1-2.2	MLE & LS	0.6	1.1	1.8
(Main, 1987)	Mount St Helens	20 Mar - 18May 1980	na	~300	Inspection	3.5-5	MLE	0.5	1.0	1.5
(Wiemer and McNutt, 1997)	Mount St. Helens	1988 - Jan 1996	2.7-3.8	100	Inspection	0.4-2.8	MLE & LS	0.8	1.2	1.6
(Wyss et al., 1997)	Off-Ito	1982 - 1996	7-15	100	MaxC	1.6-2.5	MLE	0.44	1.0	1.54
M. Wyss (written comm.)	Oshima		4					-	-	1.5
(Sanchez et al., 2004)	Pinatubo	29 June - 19 Aug 1999	0-4,8-13	100	ZMAP	0.73-	MLE	1.0	1.3	1.7
(Novelo-Casanova et al., 2006)	Popocatepetl	Dec 2000 - Jan 2001	2-7	20	Inspection	1.9-3.3	MLE	1.0	1.7	2.70
S. Wiemer (written. comm.)	Redoubt		3-4,6-8					-	-	1.7
(Power et al., 1998)	Soufriere Hills	Aug 1995 - Mar 1996	2.0-2.5	100	-	1.7-2.4	MLE	0.9	1	3.07
(Farrell et al., 2009)	Yellowstone	1984 - 2006	4-18	>10	EMR	1.5-	MLE	0.5	1.0	1.5

Values for *N* are the number of events analysed in each catalogue. These figures are either given or estimated from figures. The methods for calculating the completeness magnitude, *Mc*, are; using ZMAP software; the Goodness-of-Fit method (GFT) with given percentage threshold (e.g. 90GFt is 90% fit); the Maximum Curvature method (MaxC); Inspection is choosing a *Mc* by eye; and using the Entire Magnitude Range method (EMR). The methods for approximating the *b*-value are the Maximum Likelihood Estimation (MLE) and the Least Squares and Weighted Least Squares fit (LS & WLS). The *b*-value ranges in each study are described by the minimum (b_{min}) and maximum (b_{max}) quoted values in the study, with a typical value (b_{typ}) being estimated by eye.

10 Figures

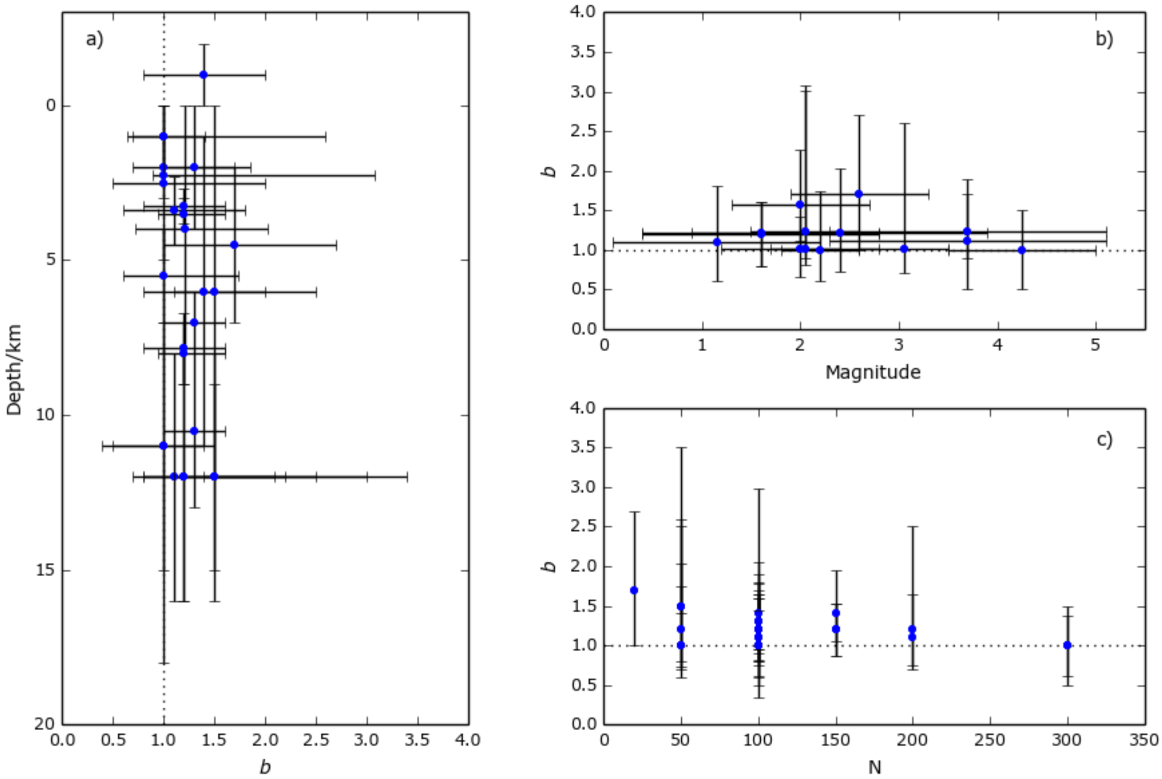

Figure 1 – Synthesis of *b*-value distributions compared to a) depth, b) Magnitude, and c) the number of events in each catalogue, *N*. The errors bars show the minimum and maximum values of *b* from Table 1, and the range of depth/magnitude over which the catalogue was comprised. The blue dots show the typical *b*-values. Dotted line marks *b*=1.

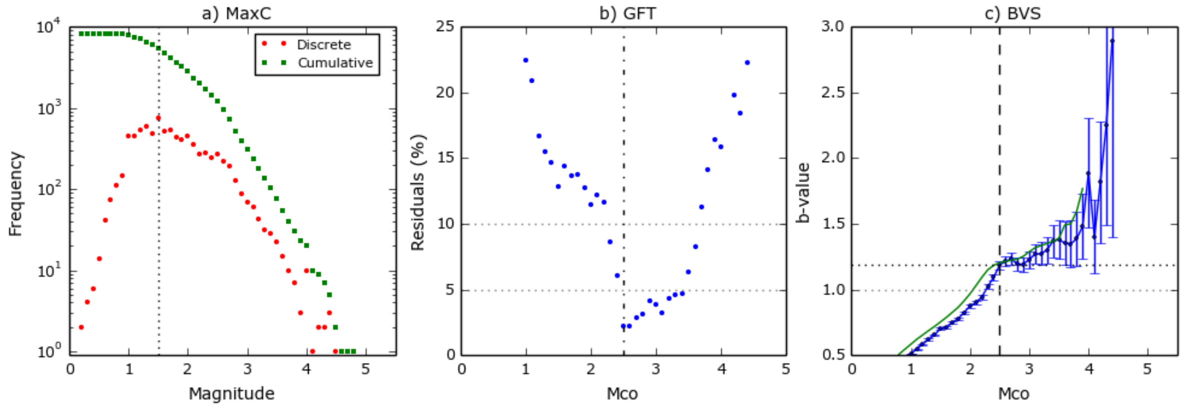

Figure 2 – a) Discrete and cumulative frequency-magnitude distributions, demonstrating the Maximum Curvature Method (MaxC). The vertical dotted line represents the estimate of *Mc* at the highest discrete magnitude bin at (*Mc*=1.5). b) Residuals of the Goodness-of-Fit method (GFT) as a function of trial cut-off. Once the residual falls beneath 5% the completeness magnitude is selected, in this case *Mc*=2.5. c) *b*-value stability curve showing the *b*-values for each cut-off magnitude. The vertical dashed line indicates when successive *b*-values (green line) fall within error of the *b*-value. Here *Mc*=2.5.

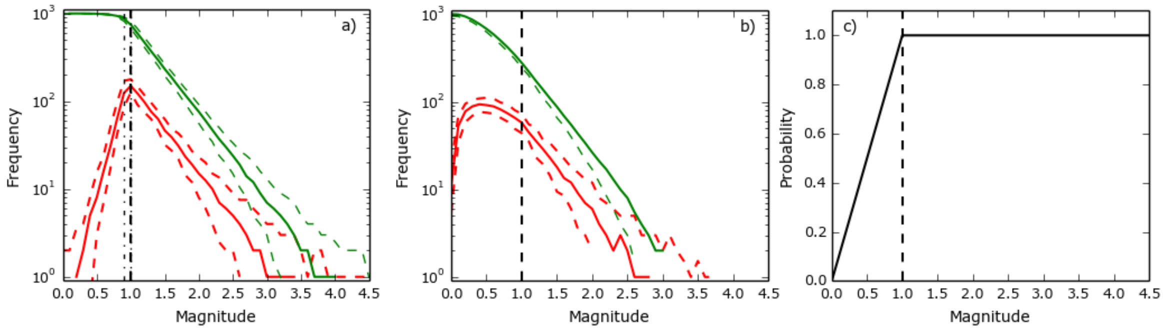

Figure 3 – a) Example of a sharp-peaked frequency-magnitude distribution. b) Example of a broad-peaked frequency-magnitude distribution. Both catalogues have an *Mc* of 1.0 and a *b*-value of 1.0. Discrete distributions are in reds, cumulative distributions are in green. The dashed lines show the 95% confidence intervals representing the scatter in the synthetic data c) The probability filter applied to b). Above *Mc*=1.0 all generated events are kept in the catalogue. Beneath *Mc*=1.0 there is a constantly decreasing probability that that will remain in the catalogue, creating the broad peak in the filtered discrete FMD.

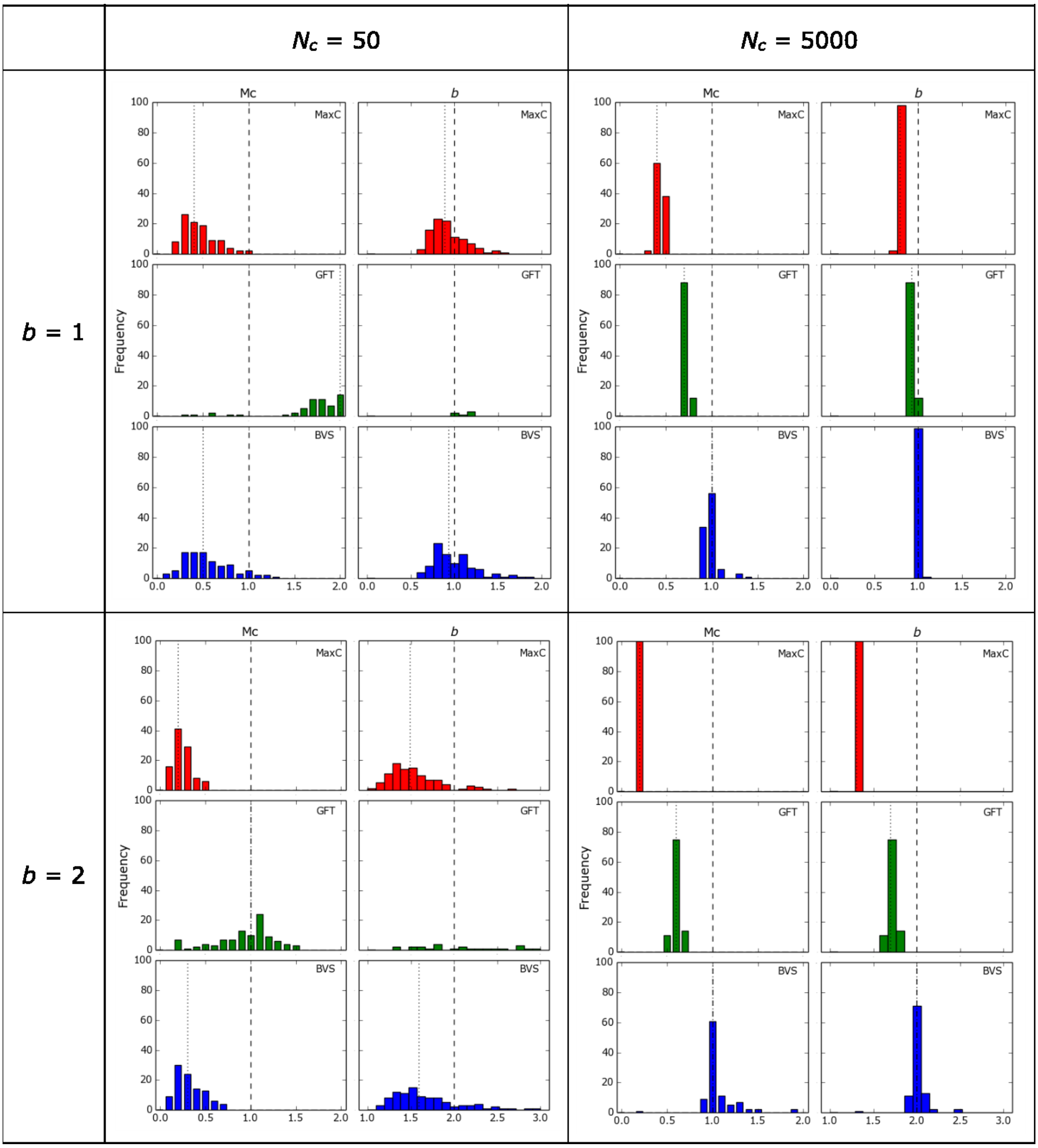

Figure 4 – Histograms for the estimated *Mc* and *b*-value for the MaxC (red), GFT (green), and BVS (blue) methods for different catalogue sizes (columns) and *b*-values (rows) for the sharp-peaked distribution. The known values of *Mc*=1.0 and *b*=1.0 are marked with vertical bold dashed lines. The median value calculated by each method is shown by the vertical dotted line.

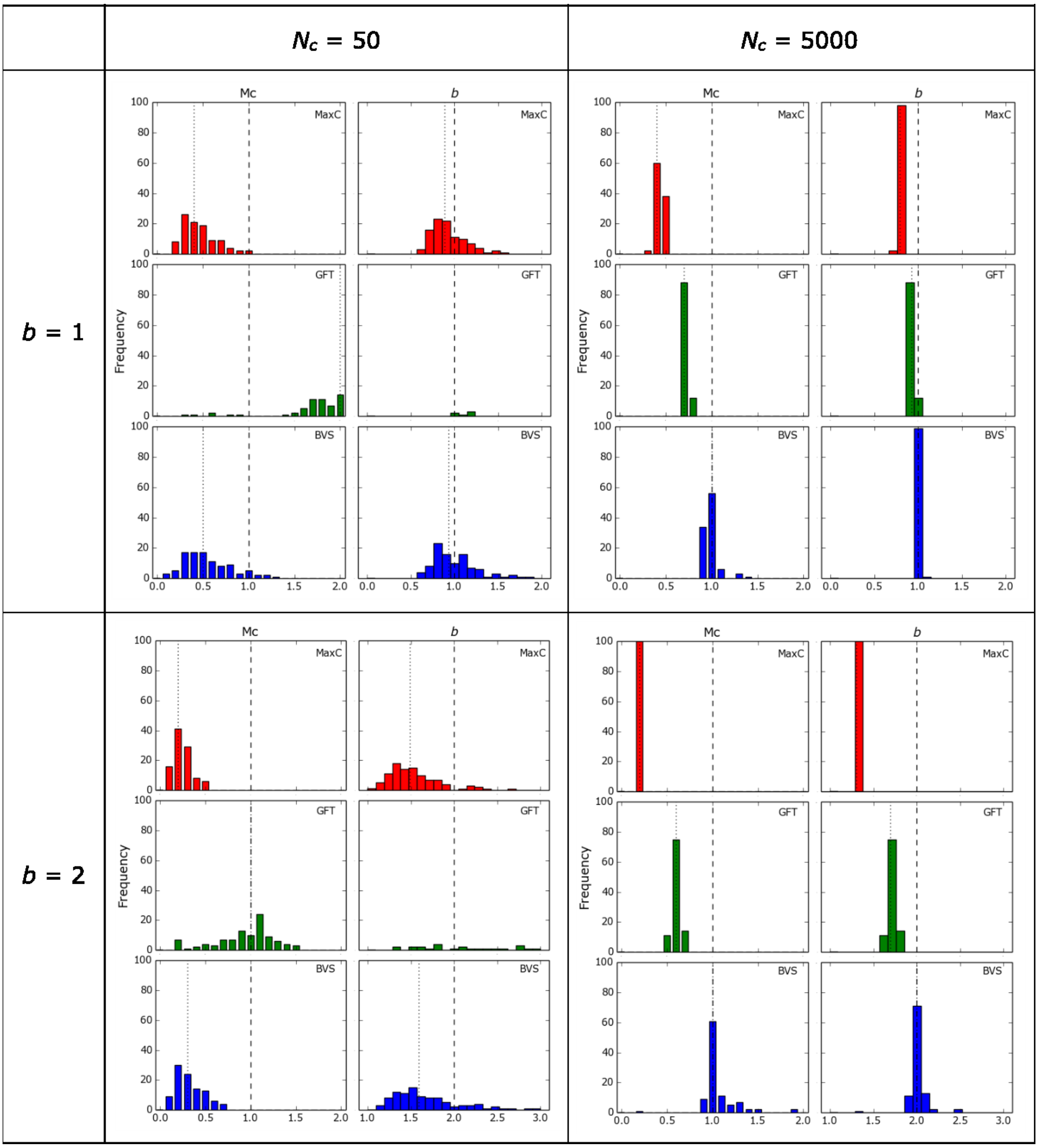

Figure 5 – Histograms as in Figure 4 except for a broad-peaked distribution.

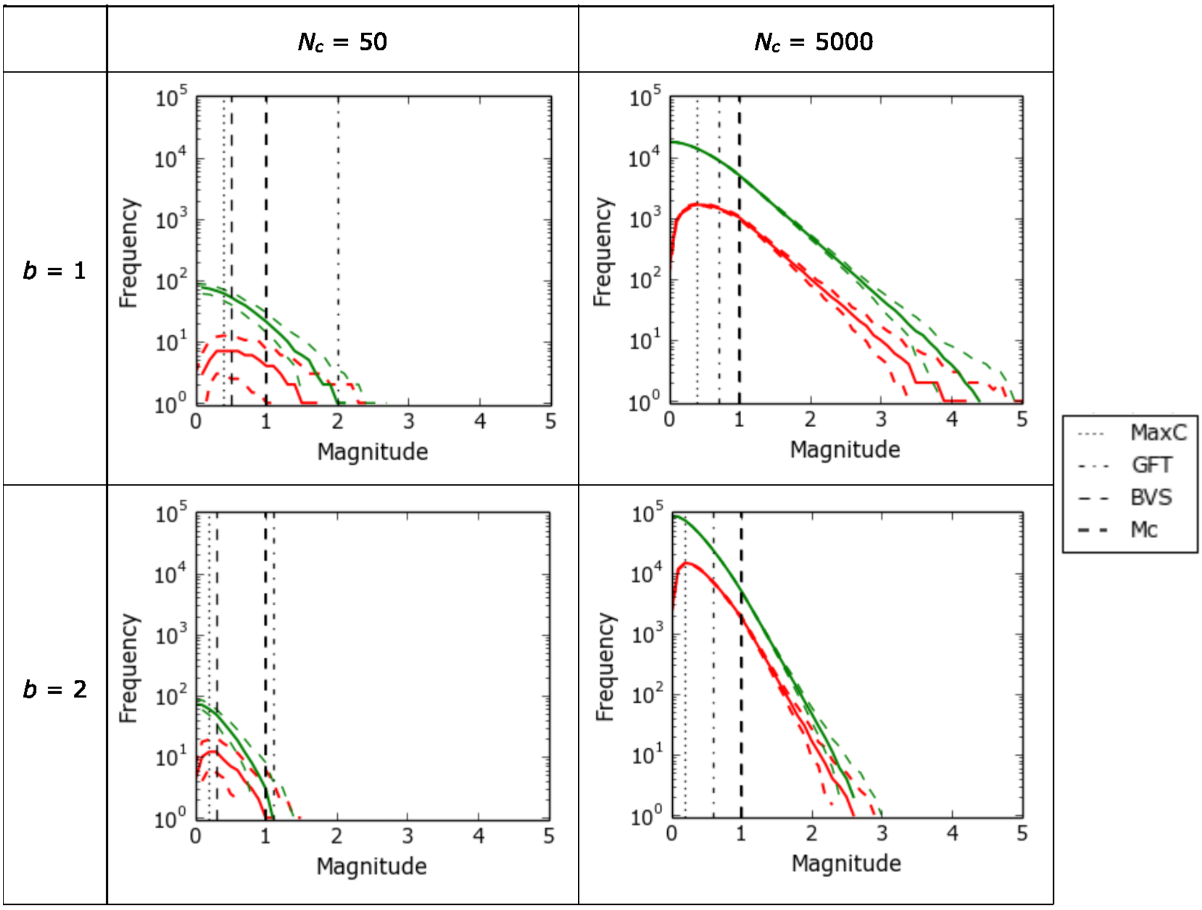

Figure 6 – Frequency-magnitude distributions for b=1 & 2, and N_c=50 & 5000 in the case of a broad-peaked distribution. Red shows discrete frequency and green cumulative frequency. The solid red and green lines show the average values of the 100 catalogues. The dashed lines represent a 95% confidence window. The vertical dashed black lines show the known Mc of the catalogue, Mc=1.0, and the Mc's calculated by each method.

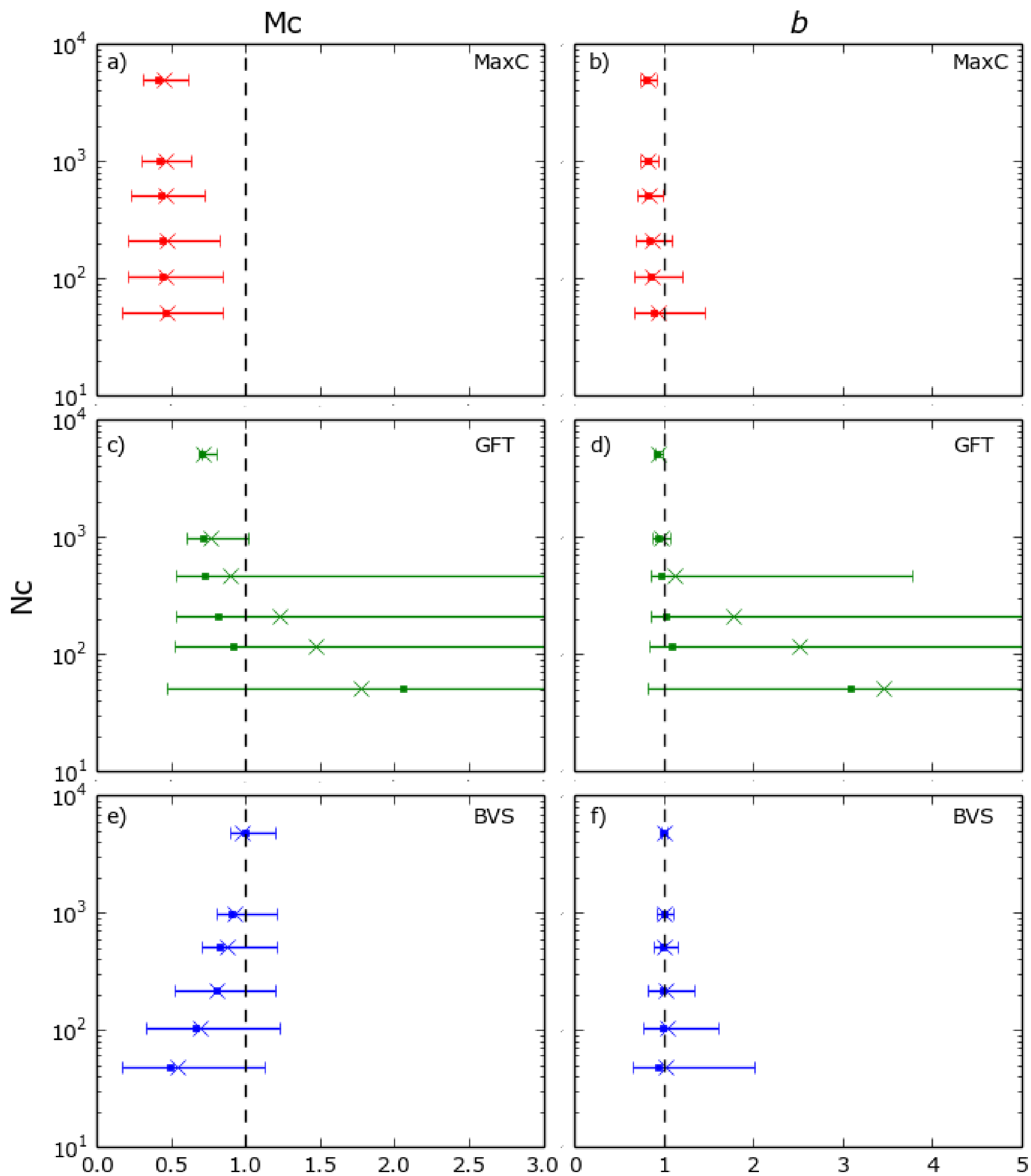

Figure 7 - Summary of histograms for broad-peaked distributions in Figure 5 for *b*=1. They show the spread of *Mc*'s and *b*-value's against catalogue size, *N*, for each of the three methods. Error bars represent a 95% spread of the data, with dots representing the median value and x's the average. The known *Mc*=1.0 and *b*=1.0 are marked with a vertical dashed line.

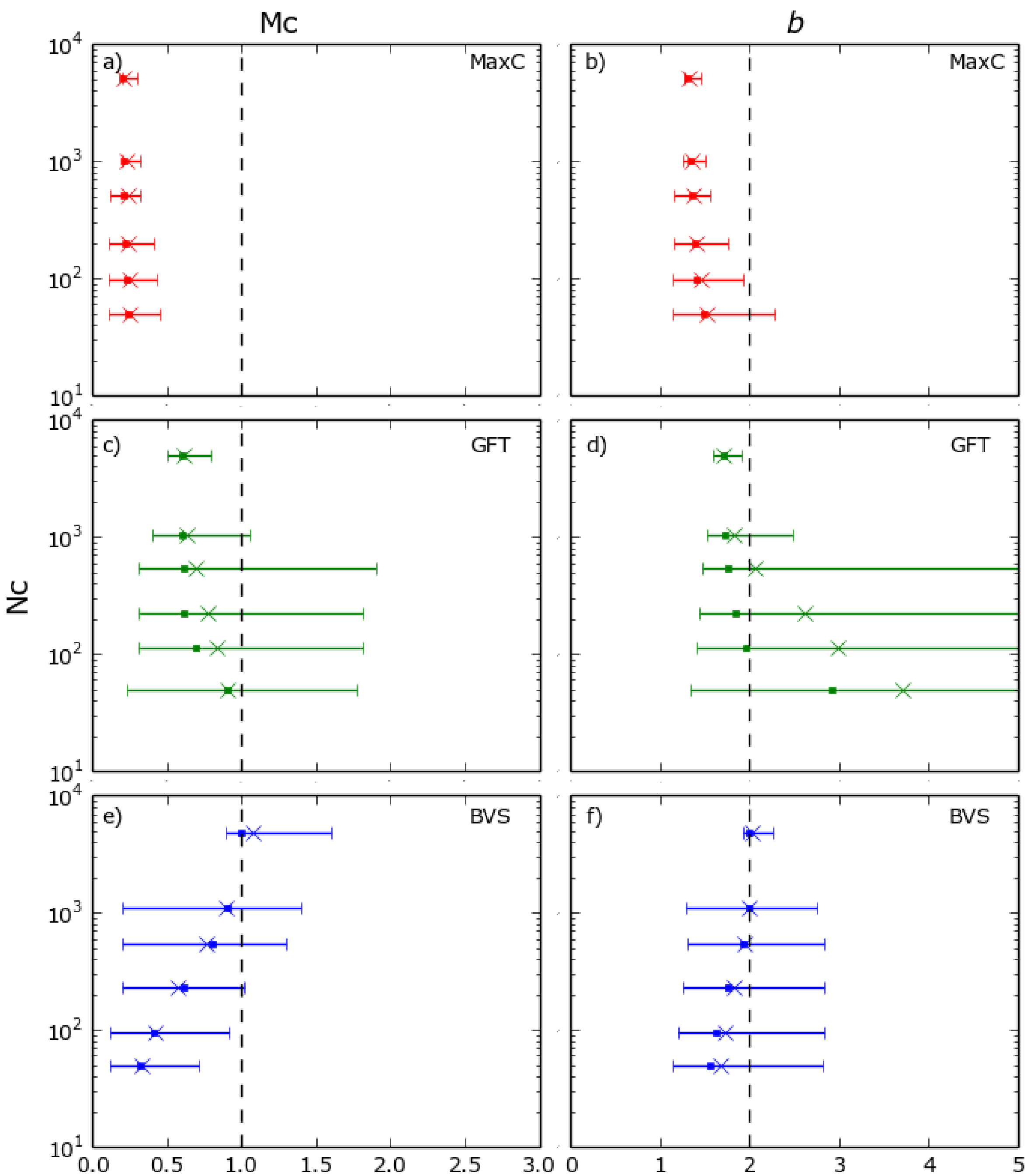

Figure 8 - Summary graphs as in Figure 7 but for *b*=2.

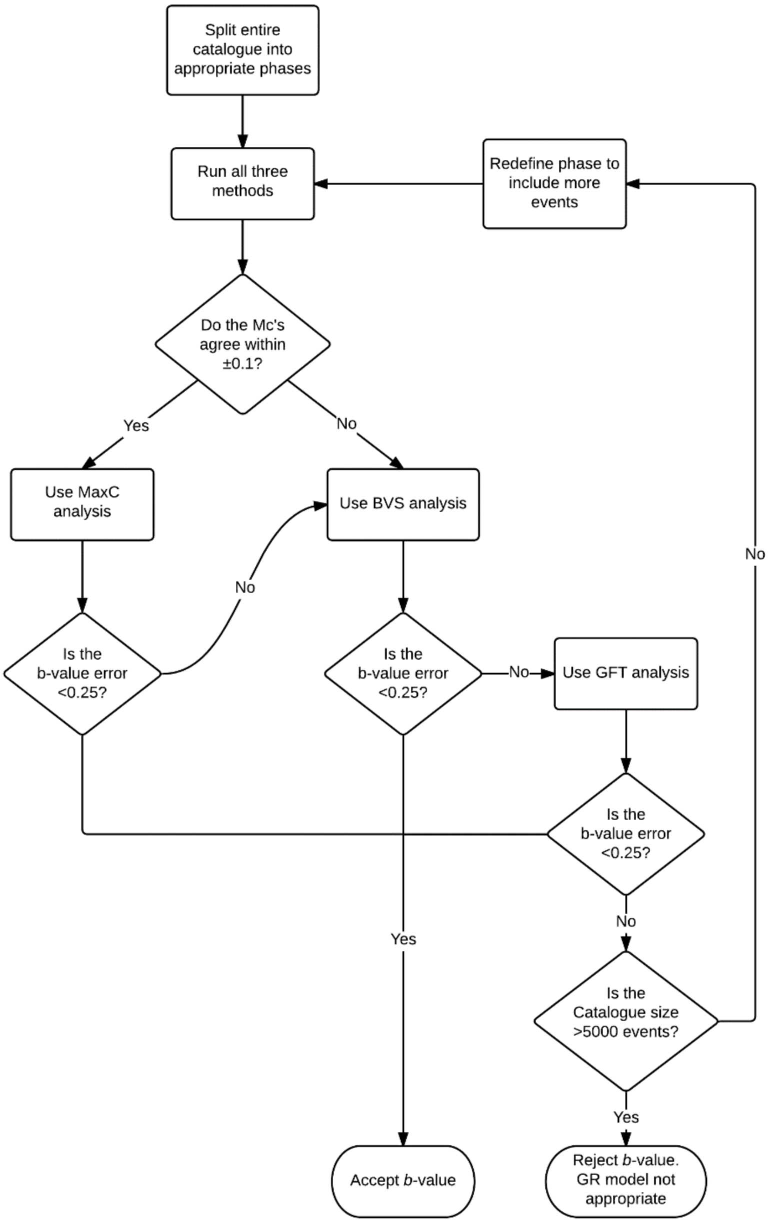

Figure 9 - Proposed workflow for best practice based on synthetic analysis.

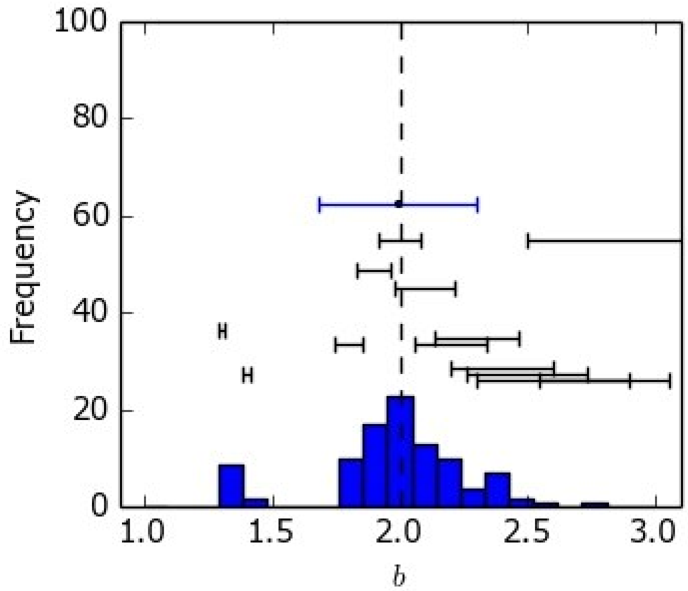

Figure 10 - *b*-value frequency plot for 100 synthetic catalogues when N_c=1000 and *b*=2. The blue (epistemic) error bar represents one standard deviation error in the data centred on the median *b*-value. The black error bars show the average aleatoric (Shi & Bolt *b*-value uncertainty) error for each bin.

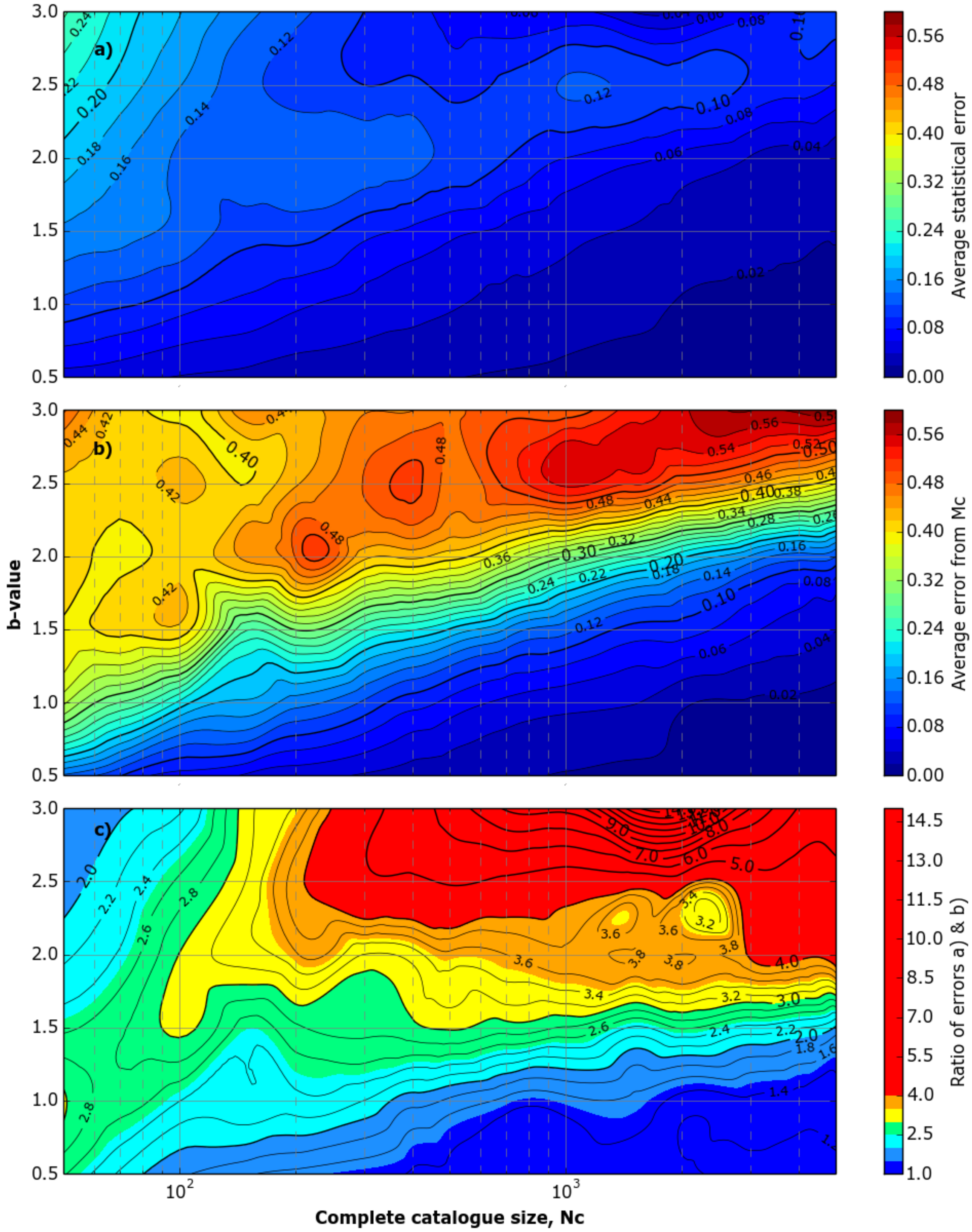

Figure 11 – Contour plots showing a) the statistical error in *b*-value estimated from eq. (5) as a function of varying complete catalogue size, N_c, and *b*-value. b) The error in *b*-value associated with the uncertainty in calculating *Mc*, estimated as in the example given in Fig 10 as a blue horizontal error bar c) The ratio of the error in (*b*) to the statistical error in (a).

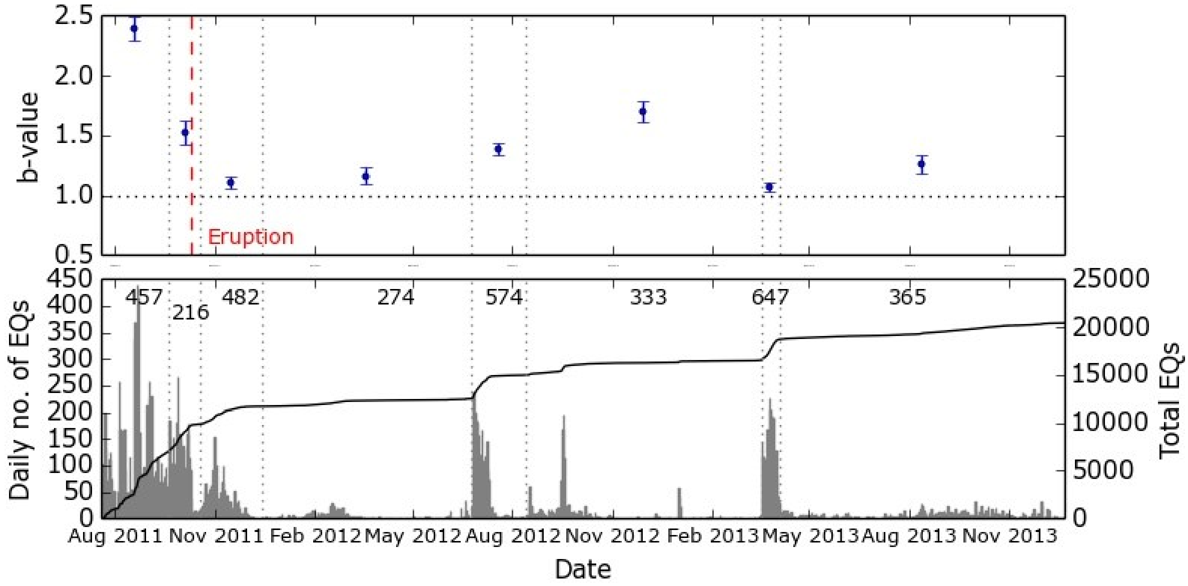

Figure 12 – Top: *b*-value variation through time for the July 2011 to December 2013 El Hierro seismic catalogue using the proposed workflow. The eruption date is marked by the red dashed line. Bottom: Daily number of events (grey bars) and cumulative number of events (black line). The phase divisions are marked by vertical grey dotted lines with the number of events in the complete catalogue of each phase noted at the top of the plot.

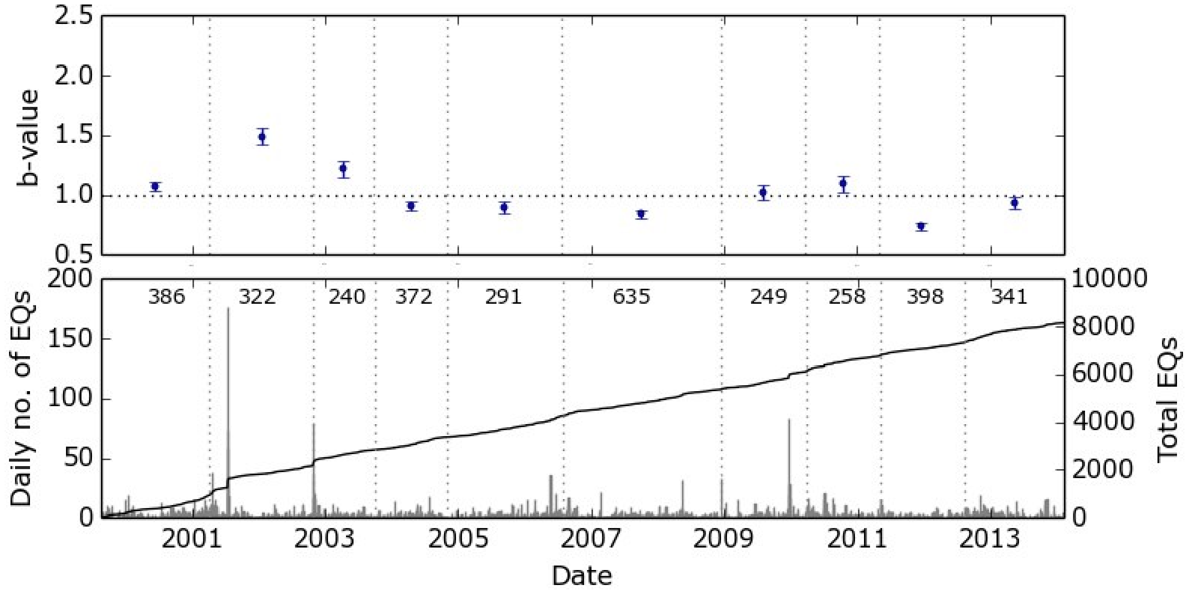

Figure 13 – Plots as in Figure 12 but for the 1999 - 2014 Mount Etna seismic catalogue.

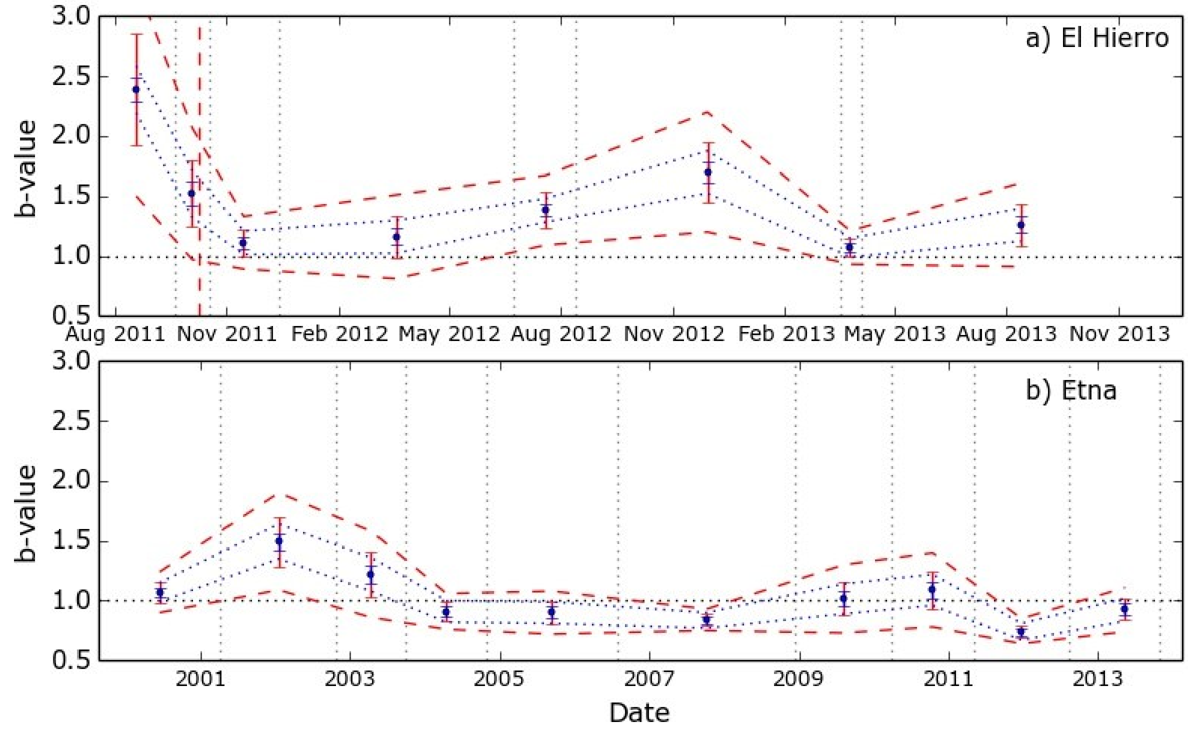

Figure 14 - *b*-value variation through time for a) the 2011-13 El Hierro catalogue, and b) the 1999 - 2014 Mount Etna seismic catalogue. Sample bias errors in are blue and estimated epistemic error are in grey. One standard deviation error is represented by the error bars and the grey dashed and blue dotted line respectively represent the 2 standard deviation error envelope.