How to be fooled searching for significant variations of the b-value

W. Marzocchi[1,⋆], I. Spassiani[2,⋆], A. Stallone[2,⋆], M. Taroni[2]

[1] *Dept. of Earth, Environmental and Resources Sciences, The University of Naples Federico II*

[2] *Istituto Nazionale di Geofisica e Vulcanologia, Rome, Italy*

⋆ *These authors contributed equally to this work*

SUMMARY

An unbiased estimation of the b-value and of its variability is essential to verify empirically its physical contribution to the earthquake generation process, and the capability to improve earthquake forecasting and seismic hazard. Notwithstanding the vast literature on the b-value estimation, we notice that some potential sources of bias that may lead to non-physical b-value variations are too often ignored in seismological common practice. The aim of this paper is to discuss some of them in detail, when the b-value is estimated through the popular Aki's formula. Specifically, we describe how a finite dataset can lead to biased evaluations of the b-value and its uncertainty, which are caused by the correlation between the b-value and the maximum magnitude of the dataset; we quantify analytically the bias on the b-value caused by the magnitude binning; we show how departures from the exponential distribution of the magnitude, caused by a truncated Gutenberg-Richter law and by catalogue incompleteness, can affect the b-value estimation and the search for statistically significant variations; we derive explicitly the statistical distribution of the magnitude affected by random symmetrical error, showing that the magnitude error does not induce any further significant bias, at least for reasonable amplitude of the measurement error. Finally, we provide some recipes to minimise the impact of these potential sources of bias.

Key words: b-value estimation, statistical seismology, statistical methods

1 INTRODUCTION

The magnitude-frequency distribution (MFD hereafter) of seismic events is typically described by the Gutenberg-Richter (G-R) exponential law [(Gutenberg & Richter(1944))]:

$$E[\text{Log}N(M)] = a - b(M - M_{min}), \tag{1}$$

where $E[\cdot]$ is the expectation operator, $N(M)$ is the number of events with magnitude $\geq M$, M_{min} is the minimum magnitude (i.e., all magnitudes $M \geq M_{min}$ are reported in the earthquake catalogue) and a, b are two positive parameters. Owing to the unavoidable binning of magnitudes, $N(M)$ may also indicate the number of events with magnitude M. The parameter b (hereafter b-value) controls the slope of the MFD in a semi-log plot: the smaller this parameter, the higher the proportion of larger shocks with respect to smaller shocks. Sometimes the GR law is written using the natural logarithm ln, i.e., the b-value is replaced by the parameter $\beta = b \ln 10$. Mathematically, the presence of the expectation operator on the left hand side of the GR law implies that the magnitude is a random variable (RV) with exponential distribution, i.e., $M \sim \mathcal{E}(\beta)$. We use the notation $X \sim D(\alpha, \dots)$ to indicate that the random variable X is distributed according to the probability distribution $D(\alpha, \dots)$ with parameters $\alpha, \dots$.

Though some authors [e.g. (Kagan(1999); Bird & Kagan(2004); Kamer & Hiemer(2015))] claim that the b-value is essentially constant and close to 1, many authors have reported spatial and temporal variations of this parameter, at which have been attributed a physical meaning and the capability to increase the skill in forecasting large earthquakes [e.g. (Scholz(1968); Smith(1980); Main et al.(1989); Imoto(1991); Schorlemmer et al.(2004); Schorlemmer et al.(2005); El-Isa & Eaton(2014)) and references therein]. Such a variability has been reported for limited space and/or time windows, while a constant in time b-value hypothesis holds well for wide areas [e.g. (Kagan et al.(2010))], as also testified by the pervasive use in the best performing earthquake forecasting models [(Schorlemmer et al.(2018))].

So far, the interpretation of the b-value variability in terms of physical processes has been mainly inferred from laboratory experiments, where significant changes in the b-value have been observed. Several processes have been proposed to explain these changes in the b-value: i) Changes in the stress conditions: the b-value is negatively correlated with differential stress. This has been observed for the b parameter of the frequency-magnitude relation of microfracturing events for rocks deformed during laboratory experiments [(Scholz(1968); Main et al.(1989); Main et al.(1992); Amitrano(2003))], as well as for the b-value of earthquakes [(Mori & Abercrombie(1997); Gerstenberger et al.(2001); Schorlemmer et al.(2005); Scholz(2015); Meng et al.(2018); Rodríguez-Pérez & Zuñiga(2018))]; ii) Changes in the rock heterogeneity: the b-value is positively correlated with the material heterogeneity. This result has been reported for both laboratory experiments [(Mogi(1962); Main et al.(1992))] and earthquakes analysis [(Mori & Abercrombie(1997); Gerstenberger et al.(2001))]; iii) Dilation and diffusion: the b-value is affected by fluid diffusion [(Henderson et al.(1994))]. Empirical evidence at the natural scale is less clear and more ambiguous at least in crustal tectonic regions [(Knopoff et al.(1982); Shi & Bolt(1982); Tinti & Mulargia(1985); Smith(1998); Frohlich & Davis(1993); Wiemer & Wyss(2002); Helmstetter & Sornette(2002); Parsons(2007); El-Isa & Eaton(2014); Raub et al.(2017))].

It goes without saying that an unbiased estimation of the b-value and of its variability is essential to test empirically its physical meaning in the earthquake generation process, and its capability to improve earthquake forecasting. Although this issue has been widely explored in the past [e.g. (Knopoff et al.(1982); Bender(1983); Wiemer & Wyss(2002); Helmstetter et al.(2003); Marzocchi & Sandri(2003); Woessner & Wiemer(2005); Parsons(2007))], and, at a first glance, there are no conceptual difficulties in estimating the parameter of the exponential distribution, we argue that some factors, which are very often neglected in seismology [(Michael & Wiemer(2010))], may induce bias in the b-value estimation and lead to apparent variations that do not have any real physical meaning. In this work, we present some new hints, which may assist scientists to reduce the possibility to be fooled by apparent b-value variations, when the b-value is estimated through the popular Aki's formulation ((Aki(1965))). Specifically, we analyse under which conditions the magnitude error can affect the b-value estimation; we investigate how the limited number of data can lead to biases and apparent statistically significant variations of the b-value; we quantify the bias introduced by the size of the magnitude binning when the most commonly used maximum likelihood estimation for the b-value is adopted, and show the effects on the expected seismicity rate for large magnitudes; we describe quantitatively the impact of departures from the GR law on the b-value estimation and its variability: the case of the truncated GR law, and of the catalogue incompleteness.

2 ESTIMATING THE b-VALUE AND ITS UNCERTAINTY

Although the b-value is still sometimes estimated through the least-square regression method [e.g. (Lopez-Pineda(2005); WGCEP(2003); Singh et al.(2011); El-Isa & Eaton(2014); Woessner et al.(2015))], this method has been proved to produce biased estimation of the b-value and of its uncertainty [(Castellaro et al.(2006); Sandri & Marzocchi(2007); Naylor et al.(2010))]. A more statistically grounded method is the maximum likelihood estimation (MLE) [(Aki(1965); Utsu(1965); Shi & Bolt(1982); Bender(1983); Marzocchi & Sandri(2003))], which is the most common approach for the b-value estimation [e.g. (Page(1968); Knopoff et al.(1982); Main et al.(1989); Wiemer et al.(1998); Kagan(1999); Wiemer & Wyss(2002); Schorlemmer et al.(2004); Schorlemmer et al.(2005); Ogata & Katsura(2006); Parsons(2007))]. When the magnitude can be considered a continuous RV (the binning problem will be discussed later) the MLE estimate of the parameter β is the reciprocal of the mean differential magnitude (i.e., the mean magnitude minus the minimum magnitude) [(Aki(1965); Utsu(1965))]. Specifically,

$$\widehat{\beta} = \frac{1}{(\overline{M} - M_{min})}, \tag{2}$$

where $\overline{M}$ is the sample mean magnitude of the events with $M \geq M_{min}$, and the hat sign indicates the estimator of either β or b-value. From equation 2, we can see that the mean magnitude $\overline{M}$ is a sufficient statistics for the calculation of the b-value. It is known that the MLE estimation given by 2 is biased and the unbiased estimation is (Ogata & Yamashina(1986))

$$\widehat{\beta}_{\text{corr}} = \widehat{\beta}\frac{N-1}{N}. \tag{3}$$

For most practical applications, the bias can be considered negligible with few tens of data. This is the reason for which equation 2 is the most commonly used formula to estimate the b-value.

As regards the b-value uncertainty, (Utsu(1966)) defines the confidence interval of the parameter β as:

$$\left(\frac{\chi^2_{1-\frac{\alpha}{2},2N}}{2N}\widehat{\beta},\, \frac{\chi^2_{\frac{\alpha}{2},2N}}{2N}\widehat{\beta} \right), \tag{4}$$

where $\chi^2_{\alpha,n}$ is the $100(\alpha)$ percentile of the chi-squared distribution with n degrees of freedom, that is, the value such that $\mathbb{P}(X \geq \chi^2_{\alpha,n}) = \alpha \quad (X \sim \chi^2_n)$. This expression can be derived explicitly from the $\widehat{\beta}$ distribution:

$$\widehat{\beta} \sim \frac{2N\beta}{\chi^2_{2N}} \tag{5}$$

For a sufficiently high number of events (tens of data), the normal approximation to the $\chi^2_{\alpha,n}$ distribution can be used [(Aki(1965))]. The Gaussian confidence interval is then defined as:

$$\left(\frac{z_{1-\frac{\alpha}{2}}}{\sqrt{N}}\widehat{\beta}, \frac{z_{\frac{\alpha}{2}}}{\sqrt{N}}\widehat{\beta}\right), \tag{6}$$

where z_α is the value such that $\mathbb{P}(Z \geq z_\alpha) = \alpha \quad (Z \sim \mathcal{N}(0,1)$ standard normal distribution). The b-value standard error is therefore $\widehat{b}/\sqrt{N}$.

In both (Aki(1965)) and (Utsu(1966)) formulations of the b-value uncertainty, the b-value is assumed to be the constant parameter of the exponential distribution; from a physical point of view is like to assume that the MFD is really exponential. This assumption may be relaxed accounting for a loosely defined "natural variability" of the b-value that may be due to several factors, such as, for example, the inhomogeneity of the seismic catalogue caused by different procedures to estimate the earthquake magnitude, and by non-detected variability of the catalog completeness in space and time. In these cases, (Shi & Bolt(1982)) proposed a formula for the standard error of $\widehat{b}$ that reads:

$$\sigma_{SB}(\widehat{b}) = 2.30\widehat{b}^2 \sqrt{\frac{\sum_{i=1}^{N}(M_i - \overline{M})^2}{N(N-1)}} \quad . \tag{7}$$

More recently, (Schorlemmer et al.(2003)) and (Amorese et al.(2010)) have proposed the bootstrap method as alternative and more reliable approach for quantifying the uncertainty (or natural variability) on the b-value estimate. By iteratively sampling (with replacement) from the empirical MFD, one can build the bootstrap distribution for statistical inference about the β parameter. This procedure returns more realistic error estimates by properly reproducing the MFD variability. Noteworthy, while (Aki(1965)), (Utsu(1966)) and (Shi & Bolt(1982)) formulas assume that the magnitude of completeness M_{min} has been correctly evaluated, the bootstrap error incorporates the uncertainties on the M_{min} estimate as well. The generality of the bootstrap techniques has been demonstrated by (Amorese et al.(2010)), which show that the bootstrap method returns error estimates that are consistent with the root mean square error, assumed to represent the true b-value variability; conversely, in both (Schorlemmer et al.(2003)) and (Amorese et al.(2010)), the standard error returned by the (Shi & Bolt(1982)) formula is proved to underestimate the true b-value variability.

We underline that the proper approach to estimate the b-value uncertainty critically relies on the assumption about the magnitude distribution. If we assume that the magnitude is exponentially distributed with a constant (unknown) b-value, equation 6 provide unbiased estimation of the b-value uncertainty. If we assume that there is a natural variability of the b-value, a bootstrap approach is more accurate to estimate the variability on the b-value; but in this case it would be important to specify what is the natural variability and what causes it. For instance, if this variability is caused by possible catalogue incompleteness, this may be cured or at least reduced. In section 3 we show that this variability cannot be caused by a reasonable magnitude error.

3 THE EFFECT OF THE MAGNITUDE ERROR ON THE b-VALUE ESTIMATION

The magnitude, as any kind of measurement, can be affected by errors which may modify the exponential distribution and its parameters. Here we consider the case in which the random error on the magnitude is distributed as the Gaussian noise of zero mean and standard deviation $\sigma_{\mathcal{N}}$, which seems a good representation of the real magnitude uncertainty [e.g. (Lolli & Gasperini (2012))]. The adjusted magnitude $M^\star$ is then obtained as the sum of exponential $M \sim \mathcal{E}(\beta)$ and Gaussian $\eta \sim \mathcal{N}(0, \sigma_{\mathcal{N}})$ random variables, that is, the adjusted magnitude $M^\star$ is an exponentially modified Gaussian (EMG) random variable, $M^\star = M + \eta$; for the sake of convenience we set $M^\star_{min} = 0$. From this definition we may infer that $E(M^\star) = E(M) + E(\eta)$; since $E(\eta) = 0$ by definition, and since the mean magnitude is a sufficient statistics for estimating the b-value, we can conclude that adding a random symmetrical error with zero average does not introduce any bias in the b-value estimation.

We now explore under which conditions the EMG distribution differs from the theoretical exponential distribution. The probability density function (PDF) of $M^\star$ is simply obtained by mathematical convolution and change of variable [(Marzocchi & Sandri(2003))]

$$p(M^\star) = \int_0^\infty \beta e^{-\beta x} \frac{1}{\sigma_{\mathcal{N}}\sqrt{2\pi}} e^{-\frac{(M^\star - x)^2}{2\sigma^2_{\mathcal{N}}}} dx \tag{8}$$

$$= \int_0^\infty \frac{\beta}{\sigma_{\mathcal{N}}\sqrt{2\pi}} e^{-\frac{(M^\star - \beta\sigma^2_{\mathcal{N}} - x)^2}{2\sigma^2_{\mathcal{N}}}} e^{\frac{(\beta\sigma_{\mathcal{N}})^2}{2} - M^\star\beta} dx \tag{9}$$

$$= \frac{\beta}{2} e^{\frac{\beta}{2}(\beta\sigma^2_{\mathcal{N}} - 2M^\star)} erfc\left(\frac{\beta\sigma^2_{\mathcal{N}} - M^\star}{\sigma_{\mathcal{N}}\sqrt{2}}\right), \tag{10}$$

where $erfc(x) = \frac{2}{\sqrt{\pi}} \int_x^\infty e^{-t^2} dt$ is the complementary error function. The EMG distribution can be considered as a shifted exponential function weighted by a function of the normal distribution. More precisely, the product $\beta\sigma_{\mathcal{N}}$ controls the EMG convergence to the normal or to the exponential function: if $\beta\sigma_{\mathcal{N}}$ is small, EMG and the exponential GR can be considered equivalent. Therefore, if the magnitude is not affected by a systematic error, and if their standard deviation is small, the MFD can be safely approximated by the exponential $\mathcal{E}(\beta)$ distribution. The similarity between EMG and GR can be verified quantitatively by analysing the ratio $R = \frac{p(M^\star)}{p_{GR}(M^\star)}$ between their relative PDFs, where $p_{GR}(M^\star) = \beta e^{-\beta M^\star}$. Since the addition of the random Gaussian noise may bring some magnitudes below the minimum magnitude (assumed to be zero here for convenience), but the opposite cannot happen, the dataset of the adjusted magnitudes will be unavoidably different from the initial one. To avoid this technical problem, we do not consider the magnitude range $[0, 3\sigma_{\mathcal{N}}]$ by opportunely normalising the PDFs in R, obtaining then

$$\begin{aligned} R &= \frac{e^{\frac{\beta^2\sigma_{\mathcal{N}}^2}{2}}}{2} erfc\left(\frac{\beta\sigma_{\mathcal{N}}^2 - M^\star}{\sigma_{\mathcal{N}}\sqrt{2}}\right) \frac{\int_{3\sigma_{\mathcal{N}}}^{\infty} \beta e^{-\beta M^\star} dM^\star}{\int_{3\sigma_{\mathcal{N}}}^{\infty} p(M^\star) dM^\star} \\ &= \frac{e^{-\beta\sigma_{\mathcal{N}}\left(3 - \frac{\beta\sigma_{\mathcal{N}}}{2}\right)}}{2 - \left[1 + erf\left(\frac{3}{\sqrt{2}}\right)\right] + e^{\beta\sigma_{\mathcal{N}}\left(\frac{\beta\sigma_{\mathcal{N}}}{2} - 3\right)}\left[1 + erf\left(\frac{3 - \beta\sigma_{\mathcal{N}}}{\sqrt{2}}\right)\right]} erfc\left(\frac{\beta\sigma_{\mathcal{N}}^2 - M^\star}{\sigma_{\mathcal{N}}\sqrt{2}}\right). \end{aligned} \quad (11)$$

The plot of this ratio is shown in Figure 1 for $\beta = 2.3$ and $\sigma_{\mathcal{N}} = 0.1, 0.3$: it is very close to 1, belonging to the range $1 \pm 0.003\, \forall M^\star$ if $\sigma_{\mathcal{N}} = 0.1$ and for $M^\star > 1$ if $\sigma_{\mathcal{N}} = 0.3$. This result clearly shows that magnitude error of reasonable size does not modify appreciably the exponential distribution that derives from the GR law. When excluding a wider magnitude range, e.g. $[0, 5\sigma_{\mathcal{N}}]$, R is still close to 1 for much larger values of $\sigma_{\mathcal{N}}$; in other words, for large $\sigma_{\mathcal{N}}$, the EMG approximates to the Gaussian distribution, but the extreme right tail ($M^\star > 5\sigma_{\mathcal{N}}$) is still exponentially distributed.

To complement these reasonings, we also provide results from a simple numerical simulation. Specifically, we simulate 1000 datasets, each with 100 magnitude values randomly drawn from a GR law. Then, we add Gaussian noise ($\sim N(0, \sigma_{\mathcal{N}})$) to the magnitude values. We test two values for the standard deviation of the noise: $\sigma_{\mathcal{N}} = 0.1$ and $\sigma_{\mathcal{N}} = 0.3$. Our goal is to inspect whether the hypothesis of exponentially-distributed magnitude is rejected at a higher rate when considering noisy magnitudes. We find that the goodness-of-fit Lilliefors test ($\alpha = 0.05$), which is a modification of the one-sample Kolmogorov-Smirnov test where the parameter is estimated by the data [(Lilliefors(1969))], is rejected 5.1% and 4.6% of times, for $\sigma_{\mathcal{N}} = 0.1$ and $\sigma_{\mathcal{N}} = 0.3$, respectively. Such rejection rates are in agreement with the chosen significance level. Therefore, our findings show that the magnitude error does not induce a bias in the b-value estimation for realistic values of $\sigma_{\mathcal{N}}$ [(Tinti & Mulargia(1985); Tinti & Mulargia(1987); Marzocchi & Sandri(2003))].

4 ESTIMATING THE b-VALUE AND ITS VARIABILITY WITH A LIMITED DATASET: THE CORRELATION BETWEEN LARGEST MAGNITUDE AND THE b-VALUE

An interesting and often overlooked feature of the b-value estimation is that the b-value estimated from data is correlated with the largest magnitude of the dataset. This may come as a surprise for some PSHA practitioners which are used to say that MLE estimation takes into account only of the small magnitudes and not of the largest [e.g., (Woessner et al.(2015))].

As a matter of fact, the mean of highly skewed distributions, such as the exponential distribution, is sensitive to outliers. Recalling that β is estimated by inverting the quantity $\overline{M} - M_{min}$, it follows that the b-value can be strongly influenced by outliers as well: in particular, the b-value decreases as the largest magnitude in the sample increases.

To investigate if this correlation exists and the effects on the apparent statistical significance of the b-value variability, here we analyse the b-value of the seismicity before a mainshock, which is often advocated as a possible indicator of stress changes before the mainshock. First, we simulate a synthetic earthquake catalogue according to the ETAS model [(Ogata(1988); Ogata(1998))] with a constant b-value $= 1$ for all earthquakes (see Appendix A). We choose the ETAS model because it represents the best forecasting model according to prospective tests [e.g. (Taroni et al.(2018))]. The synthetic catalogue includes 3,704,838 events and covers a period of 10,000 years. The magnitude ranges between 3 and 7.9. Second, we identify all the large earthquakes ($M \geq 6.5$) in the dataset, which sum up to 1,140 events. The threshold value of 6.5 represents an arbitrary choice often adopted in the literature for the mainshock definition [e.g. (Reasenberg(1999); Zöller & Hainzl(2002); Helmstetter & Sornette(2002); Ogata & Katsura(2014))]; for the sake of convenience, in the following we use the terms mainshocks for such events, and aftershocks (foreshocks) for the seismicity that follows (anticipates) the mainshocks, even though such a distinction is completely arbitrary for the ETAS model. Third, we estimate the b-value of the earthquakes that occurred in the time window which immediately precedes each mainshock. Since we are interested in investigating the foreshocks predictive skill, we consider only the case in which the events in the window have a magnitude lower than the mainshock, i.e., they correspond to the foreshocks of Type I described in (Helmstetter et al.(2003)) and (Helmstetter & Sornette(2002)). Additionally, the foreshocks are selected within a radius of 30 km from the mainshock, which is a common setting for the maximum accepted distance [e.g. (Schorlemmer et al.(2004)), (Parsons(2007)), (Papadopoulos et al.(2010)), (Wyss & Lee(1973))]. We adopt the most common choice for the window definition, namely a window of variable length, but fixed number of events. To inspect how the choice of the number of events per window can affect the results, we run our analysis for different values of the sample size, $N = 50, 200, 500, 2000$. The first two options ($N = 50, 200$) represent popular values in the literature [e.g., (Parsons(2007)), (Westerhaus et al.(2002)), (Schorlemmer et al.(2004)), (Shi & Bolt(1982)); (Nuannin(2006)), (Wyss & Lee(1973))].

For each mainshock, we estimate the b-value before its occurrence by implementing equation 2. To verify whether the b-value is significantly correlated to the maximum foreshock magnitude, we use the Kendall-tau test [(Kendall(1948))], a non-parametric test which returns a measure of rank correlation between two variables. In Figure 2, for each sample size, we plot the estimated b-values against the magnitude of the largest foreshock occurred in the sequence they have been evaluated for. Results from the Kendall-tau test are also shown. The correlation between the b-value and the maximum foreshock magnitude is significantly negative for all the explored sample sizes. It goes without saying that the same correlation holds for samples randomly chosen from an exponential distribution; the reason is that for an ETAS model with b-value $= 1$, the magnitudes observed in any time interval are a random sample from the same exponential distribution.

Our findings suggest that the number of events typically selected for the b-value estimation ($N = 50, \ldots, 500$) exposes the b-value estimation to a significant dependence on the magnitude of the largest foreshock; this dependence is strongly reduced only for more than 1000 events (be aware though that the correlation is still not zero for $N = 2000$). In practice this means that even though the b-value is the same for all earthquakes, we expect to observe a lower b-value when at least one strong foreshock is observed, in particular for dataset of few hundreds earthquakes or less.

Now we investigate the effects of this correlation on the search for significant variations of the b-value before a mainshock. We stack all the foreshocks sequences with at least 100 earthquakes at a distance smaller than 30 km from the mainshock in a fixed 1-year time window before each mainshock. The length of the time window is not relevant here, but we chose a sufficiently long time window to have enough data to calculate the b-value. We use the Wilcoxon signed rank test [(Wilcoxon(1945))] to assess whether the median of the distribution of the estimated b-value for all stacked seismic sequences is significantly smaller than 1. Our analysis returns a clear systematic deviation of the median b-value from the theoretical value at the significance level of 0.01. Since the synthetic catalogue is built with b-value $= 1$, this departure stems from the specific nature of the ETAS model for which the stronger the foreshock the larger the likelihood of a mainshock in a specific time window. Figure 3 shows the histogram of the estimated b-values, with highlighted the mean, the median and the theoretical value of 1.

To complement the information returned by the rank test, we analyse the same synthetic catalogue and count the fraction of times for which we estimate the b-value significantly lower than 1. To do that, we implement the one-sample left-tailed t-test for each group of data [(Boslaugh(2012))]. The t-statistic is:

$$t_{stat} = \frac{\widehat{b} - b}{\widehat{\sigma}}, \tag{12}$$

where $\widehat{b}$ is the b-value estimated for a given catalogue at a given completeness magnitude M_{min} ($M \geq M_{min}$); b is the true b-value used by the model (i.e., $b = 1$); $\widehat{\sigma}$ is the b-value error estimated from the sample according to equation 6. The degrees of freedom are $N - 1$. When adopting a significance level $\alpha = 0.05$, we find that the b-value is significantly smaller than 1 in 23% of the number of windows, instead of 5%. In other words, we find a b-value significantly lower than

1 about one out of four times. The main finding here is that when considering the distribution of the b-value (calculated with a few hundreds of data) of a set of foreshock sequences, an apparent statistically significant small b-value can be found more frequently than expected by chance.

5 THE BIAS INTRODUCED BY THE BINNING OF MAGNITUDES

In all seismological applications the errors in the magnitude estimation require the use of binned values, e.g., the magnitudes are grouped in selected ΔM intervals, with usually $\Delta M = 0.1$ (more recent instrumental catalogues may have $\Delta M = 0.01$). In other words, magnitude is not anymore a continuous RV, but it has discrete values. (Bender(1983)) and (Tinti & Mulargia(1987)) provide a detailed analysis about how to estimate the b-value accounting for the binning. Nonetheless, equation 2 is still the most common procedure to estimate the b-value, likely due to its simplicity; the only modification is that $M_{min} = M_c - \Delta M/2$ and M_c is the completeness binned magnitude of the catalogue [(Marzocchi & Sandri(2003))], namely the smallest magnitude above which only a negligible number of earthquakes are lost. The still large use of equation 2 motivates us to explore more in depth the bias induced by the binning. The presence of this bias is not a novelty. (Utsu(1966)) shows that the magnitude binning induces a significant bias when $b\ \Delta M > 0.25$; (Bender(1983)) in her figure 5g-h shows graphically that the equation 2 is affected by a significant negative bias when $\Delta M = 0.6$. Here we contribute adding an analytical formulation to estimate the unbiased b-value starting from estimation obtained by equation 2.

Since the mean magnitude is a sufficient statistic for estimating the b-value (see equation 2), we only need to inspect the extent to which the magnitude binning could influence the mean magnitude. It is easy to demonstrate that the binning does not have any influence on the b-value estimation if the central value of the bin coincides with the mean of the magnitudes inside the bin. This is always true for symmetrical distributions, but not for highly skewed distributions such as the exponential one, which is actually the distribution adopted for earthquake magnitudes. The difference between the (exponential) mean and the central value (uniform mean) of the bin needs therefore to be inspected, as it may cause a bias due to the binning procedure.

To compute this difference, we group the continuous magnitudes generated by an exponential distribution in $[M_i - \Delta M/2; M_i + \Delta M/2)$ bins of size ΔM. The centre of each bin is M_i by construction; the exponential mean ($\overline{M}_i^{(\mathrm{exp})}$) is instead derived as:

$$\overline{M}_i^{(\text{exp})} = \frac{1}{F(M_i, \Delta M)} \int_{M_i - \Delta M/2}^{M_i + \Delta M/2} m\phi(m) dm, \tag{13}$$

where $\phi(m) = \beta e^{-\beta(m - M_{min})}$ is the PDF of the exponential GR distribution, and $F(M_i, \Delta M) = \Phi(M_i + \Delta M/2) - \Phi(M_i - \Delta M/2)$ is the normalising function given by the difference between the CDF calculated in the integral extremes. Easy computations give that

$$\overline{M}_i^{(\text{exp})} = M_i + \frac{1}{\beta} - \frac{\Delta M}{2} \cdot \frac{1 + e^{-\beta \Delta M}}{1 - e^{-\beta \Delta M}}, \tag{14}$$

which leads to the difference between the exponential and uniform mean points in each bin of size ΔM equals to:

$$\delta = M_i - \overline{M}_i^{(\text{exp})} = -\frac{1}{\beta} + \frac{\Delta M}{2} \cdot \frac{1 + e^{-\beta \Delta M}}{1 - e^{-\beta \Delta M}}. \tag{15}$$

This equation has been obtained also by (Bender(1983)) using a different procedure (see her equation A5).

Notice that the bias tends to zero for $\Delta M \to 0$, and it is the same for all magnitude bins, no matter what the central value is; this is a direct consequence of the memoryless property of the exponential distribution. Combining equations 2 and 15 we get the difference between the β-value estimated using the correct exponential mean and $\widehat{\beta}$ obtained by equation 2:

$$\frac{1}{\widehat{\beta}} - \frac{1}{\beta} = \delta. \tag{16}$$

Then, combining equations 16 and 15 we can obtain the correct estimation of the b-value from the $\widehat{b}$-value estimated by equation 2 and ΔM:

$$b = \ln \left[\frac{1 + \widehat{b}\, \ln(10) \Delta M/2}{1 - \widehat{b}\, \ln(10) \Delta M/2} \right] \frac{1}{\ln(10) \Delta M}. \tag{17}$$

In Figure 4 we show the plot of the ratio $b/\widehat{b}$ (17 and 2) as a function of ΔM for different b-values. The solutions of this equation are comparable to the numerical solutions provided by (Utsu(1966)) for a few selected combinations of $b\Delta M$.

For the sake of example, if $\Delta M = 0.6$ and the true $\beta = 2.3$, $\widehat{\beta}$ estimated by equation 2 is about 13% less. We investigate this bias numerically, plotting in Figure 5 the frequencies of the b-values estimated for 1000 simulated datasets of 100 magnitudes each, for $\Delta M = 0.1$ and $\Delta M = 0.6$. The histogram shows that a high value of ΔM induces a lower estimation of the b-value. These results agree with the one obtained in (Bender(1983)), where the binning procedure is stated to be relevant for seismic hazard when $\Delta M = 0.6$. In particular, in our last example, if we do not correct the $\widehat{b}$-value and use the same seismicity rate for $M \geq 4$, the seismicity rate for $M \geq 6$ and $M \geq 7$ is about 1.8 and 2.5 times larger than the correct seismicity rate. It goes without saying that this has usually a significant impact, increasing artificially the seismic hazard.

In conclusion, our findings show that for instrumental catalogues with $\Delta M = 0.1$ or smaller the use of equation 2 does not lead to any appreciable bias. Otherwise, if a larger ΔM is used, the correction given by equation 17 has to be applied.

6 BIAS INTRODUCED BY DEPARTURES FROM THE EXPONENTIAL DISTRIBUTION: THE CASE OF THE TRUNCATED GUTENBERG-RICHTER LAW, AND THE CATALOGUE INCOMPLETENESS

6.1 Are the data consistent with the Gutenberg-Richter law?

The b-value estimation and its uncertainty through equations 2 and 6 makes sense only if $M \sim \mathcal{E}(b)$, i.e., the magnitude is exponentially distributed; otherwise, the scientific relevance of estimating the b-value becomes questionable. Although this claim may sound trivial, we notice that classical goodness-of-fit tests are rarely done in seismological literature to verify if this assumption is supported by the data [(Clauset et al.(2009)); (Kamer & Hiemer(2015))]. As a general principle, any b-value estimation has to be made only when the data do not show clear evidence against the exponential distribution; a popular test of this hypothesis is given by the Lilliefors test, which has been described before. Departures from the exponential distribution can be due to different technical or physical causes. Here we consider two common reasons: the magnitude is distributed according to a different statistical distribution, such as a truncated GR law; the magnitude does not follow the exponential distribution because the earthquake catalogue is incomplete.

6.2 The bias introduced by truncation

The truncated GR law consists in a modification of the classical GR in which an upper hard magnitude cutoff M_{up} is introduced to account for the limited capability to accumulate seismic energy in one specific region or fault. The PDF of the truncated GR distribution is

$$\phi_{\text{trunc}}(M) \quad = \quad \frac{\beta e^{-\beta(M-M_{min})}}{1-e^{-\beta(M_{up}-M_{min})}}, \quad M_{min} \leq M \leq M_{up}. \tag{18}$$

For a given range of $M_{up} - M_{min}$, the Trunc-GR law becomes approximately equal to the GR relationship:

$$\frac{\phi_{\text{trunc}}(M)}{\phi_{\text{GR}}(M)} \to 1 \qquad \Leftrightarrow \qquad \left| \frac{1}{1 - e^{-\beta(M_{up} - M_{min})}} - 1 \right| < \varepsilon \quad \Leftrightarrow \quad M_{up} > \frac{1}{\beta} \ln\left(\frac{1+\varepsilon}{\varepsilon}\right) + M_{min}, \tag{19}$$

for an arbitrary small constant ε. For example, if we set $\varepsilon = 0.001$, we get that the two laws are approximately the same for $M_{up} > 3 + M_{min}$, as already mentioned in (Marzocchi & Sandri(2003)).

As regards the bias introduced in the b-value estimation, (Bender(1983)) notices that a truncation of the GR law may induce a bias in her procedure to estimate the b-value. Here we explore the same kind of bias when estimating the b-value through equation 2. We first perform a MLE analysis to find the correct β estimator for the Trunc-GR law. The log-likelihood for the parameter β in 18 is then given by $N \ln \beta - N\beta(\overline{M} - M_{min}) - N \ln\left(1 - e^{-\beta(M_{up} - M_{min})}\right)$, where N is the number of events. Deriving this function, we obtain that the estimated β is the solution of the equation

$$\frac{1}{\beta} - \overline{M} + M_{up} - \frac{M_{up} - M_{min}}{1 - e^{-\beta(M_{up} - M_{min})}} = 0. \tag{20}$$

Although the latter formulation differs from equation 2, very often in practical applications the parameter β of Trunc-GR magnitudes is estimated by using 2 that has been derived from the classical GR case. To investigate the bias introduced by such a procedure, we perform 100 simulations of $N = 100, 200, 300, 400$ magnitudes and we compute the mean difference $\text{bias}_{\text{Trunc-GR}} = \mathbb{E}[\beta - \widehat{\beta}\,]$ between the MLE solution β of equation 20 and $\widehat{\beta}$ of equation 2. Since equation 20 is of transcendental type, we use a numerical approach (notice that the remaining small bias is due to the binning problem discussed before). Figure 6 shows that the $\text{bias}_{\text{Trunc-GR}}$ is not significantly different from zero when $M_{up} - M_{min} \geq 3$; conversely, the estimation of β (b-value) using equation 2 may be significantly affected when $M_{up} - M_{min} = 2$ or less. This means that, for example, equation 2 can overestimate significantly β (b-value) when considering normal earthquakes in the global catalogue, because the completeness magnitude is about 5.5 and the maximum or corner magnitude is about 7.5 (Bird & Kagan(2004)). Notice that this bias is opposite to the one obtained by (Bender(1983)) for her procedure to estimate the b-value. In Figure 7 we show the bias percentage as a function of $M_{up} - M_{min}$. In the same figure we show the same quantity for the tapered Gutenberg-Richter distribution; the results have the same trend of the truncated distribution, but with a slightly smaller effect.

Finally, since the truncated GR is actually the GR law with only an upper bound for the magnitudes, we expect that the bias introduced by binning Trunc-GR distributed random magnitudes is the same as for the classical GR case. This is mathematically confirmed by explicitly computing the integral $\overline{M}^{(\text{exp})} = \frac{1}{F(M_i, \Delta M)} \int_{M_i - \Delta M/2}^{M_i + \Delta M/2} m \phi_{\text{Trunc}}(m) dm$, where

$F(M_i, \Delta M) = \Phi_{\text{trunc}}(M_i - \Delta M/2) - \Phi_{\text{trunc}}(M_i + \Delta M/2)$, which reduces exactly to that of the classical (non-truncated) GR.

6.3 Bias on the of the b-value introduced by incomplete earthquake catalogues

The magnitude may deviate from the exponential GR law at lower magnitudes, due to several factors (network detection capability, seismic stations density, low signal-to-noise ratio, post-processing, incomplete recording after larger events etc.). The usual approach is to determine the completeness magnitude M_c, using different procedures [see (Woessner & Wiemer(2005)) and references therein, and also (Schorlemmer & Woessner(2008))]. Although some methods do not require explicitly the estimation of M_c [e.g. (Ogata & Katsura(1993); Kijko & Smit(2017))], it is still possible to identify a magnitude M_c [see (Woessner & Wiemer(2005))]. No matter how one estimates such M_c, it is always possible that some earthquakes with $M \geq M_c$ will be missing; in practice, the term 'negligible' in the above definition means that these missing earthquakes do not induce any significant bias in the estimation of the GR parameters. Since missing earthquakes are not a random sample from the GR law (having preferentially small magnitudes), the most obvious procedure to check if the earthquake catalogue is practically complete is through testing whether the magnitude $M \geq M_c$ follows the exponential distribution ((Clauset et al.(2009))); if so, it is implicitly assumed that the missing earthquakes do not affect the analyses.

The importance of a correct assessment of M_c in the b-value estimation is far to be a novelty and has been addressed in the past by several authors which show that underestimating M_c implies an underestimation of the b-value [e.g., (Knopoff et al.(1982); Woessner & Wiemer(2005); Mignan & Woessner(2012)) and references therein, (Godano et al.(2014))]. Here we want to contribute with new insights by showing how subtle this bias could be, showing that passing a goodness-of-fit test for $M \geq M_c$ is a necessary, but not sufficient condition to retrieve an unbiased estimation of the b-value.

To explore in detail the effects of incompleteness, we carry out two simple simulations in which we demonstrate how earthquake catalogues with magnitudes apparently exponentially distributed may contain significant bias in the b-value estimation. In the first case, we simulate 1000 synthetic earthquake catalogues, each one with exponentially distributed magnitudes, b-value=1, and different sample size, $N = 50, 200, 500, 2000$. Then, for each simulated dataset, we add an arbitrary and known incompleteness at the lower magnitudes, adopting the procedure suggested by (Ogata & Katsura(1993); Ogata & Katsura(2006)); in particular, we filter the data using the cumulative normal distribution

$$F(M|\mu, \sigma) = \int_{-\infty}^{M} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx, \tag{21}$$

where μ and σ are constants: the first is the magnitude with a missing rate $\nu(M_c) = 0.5$ (i.e., $\nu(M)$ is the probability that an earthquake with magnitude M will be missing); the latter is the standard deviation of the normal distribution. $F(M|\mu, \sigma)$ is the probability of detection at magnitude M, i.e., $\nu(M) = 1 - F(M|\mu, \sigma)$. For our simulations, we set $\mu = 2$ and $\sigma = 0.2$ (the exact values of these parameters does not affect the final results, since we are dealing with a synthetic dataset, for which we are simulating an arbitrary incompleteness).

In Figure 8, for each considered sample size, we group the estimated b-values as a function of M_c (and the associated ν). For all N values, there is clear evidence of the bias affecting the b-value estimation for M_c between μ and $\mu + 2 \cdot \sigma$ ($0.02 < \nu(M) < 0.5$). As expected, the b-value distribution exhibits a lower dispersion as the sample size increases. In agreement with previous studies [e.g., (Woessner & Wiemer(2005); Mignan & Woessner(2012)) and references therein, (Godano et al.(2014))], the medians are positively correlated with M_c. More interesting and novel insights come from Table 1, which shows the results of both the Lilliefors goodness-of-fit test and the t-test of the null hypothesis b-value=1 for all the sample sizes. The most striking result is that the t-test of the null hypothesis $b = 1$ has a rejection rate that is much higher than the rejection rate of the Lilliefors test, which tests the exponential distribution null hypothesis. This means that often, when the earthquake catalogue appears to be complete because the magnitude appears to follow an exponential distribution, the missing events may still lead to significant underestimation of the b-value. Notably, this occurs also for large dataset ($N = 2000$); in particular, when $\nu(M) = 0.07$, the Lilliefors test seems unable to find significant discrepancies from the exponential distribution, while the rejection rate of the t-test is 3 times higher than expected. This effect is likely due to the fact that the nonparametric goodness-of-fit tests have less power that the parametric t-test, which is more sensible to any departure from the exponential distribution. In other words, the incompleteness of the earthquake catalogue may be undetected by goodness-of-fit tests, but it may induce significantly lower b-values. Although the choice of the minimum reasonable M_c is always the best strategy to use the largest number of earthquakes in the analysis, we recommend to test the stability of the outcomes for larger M_c.

The problem of the incompleteness emerges also in other kinds of applications that are mimicked by this second simulation. In several application, the check of completeness is carried out for the whole earthquake catalogue, but the b-value is calculated in some space and/or time partitions of the catalogue (subsets); for instance, this happens often when the b-value is calculated in spatial grids [e.g. (Wiemer et al.(1998); Wyss & Wiemer(2000); Westerhaus et al.(2002); Wiemer & Wyss(2002); Schorlemmer et al.(2004); Wiemer et al.(2005); Nanjo & Yoshida(2017))]. To investigate on this issue we simulate the case where the chosen M_c holds perfectly for all subsets of the earthquake catalogue but one, and we show

how this can affect the search for significant b-value variations. Our simple case consists of an iterative application (1000 times) of the following algorithm:

(i) Generate a catalogue of 1000 magnitudes following the GR distribution, for which we set a minimum magnitude of 2 and a bin width of 0.1;

(ii) Divide the catalogue in 5 subsets, each with 200 data;

(iii) Model incompleteness for one subcatalogue only (the fifth), following the same procedure explained in Section 6.3;

(iv) Estimate the b-value assuming $M_c = 2$ for the whole aggregated catalogue (composed by 4 complete subsets and one incomplete) and for each subset.

The outputs from the iterative simulation can be then summarised by finding the proportions of times when: i) the Lilliefors test is rejected in each subset and in the aggregated catalogue; ii) the t-test is rejected in each subset and in the aggregated catalogue. Results are shown in Table 2. Figure 9 shows the distributions of the b-value estimated for the subcatalogues and the aggregated catalogue in the 1000 simulations. From Table 2 and Figure 9 we can see that most of the times $M_c = 2$ is a reasonable completeness magnitude for the whole aggregated catalogue since the Lilliefors test is not rejected and the b-value is not significantly different from 1. However, this is not true for the incomplete subcatalogue, for which we find a b-value significantly lower than 1 about the 70% of times; in essence the apparently significant b-value variability among the subsets is just derived from different subcatalogues completeness. This bias can be limited by carrying out specific completeness analysis in each subset, avoiding to extrapolate to any subset the M_c estimated for the whole catalogue. Although this comes to the price of having less data to estimate M_c and the b-value, exposing more strongly the analyses to the biases discussed in this paper, we think this procedure has to become a standard for this kind of analysis. To our knowledge, only a limited number of studies implement this best practice [e.g. (Cao & Gao(2002); Wiemer et al.(2005); Papadopoulos et al.(2010); Chan et al.(2012); Kijko & Smit(2012); Kamer & Hiemer(2015); García Hernández et al.(2018))].

7 CONCLUSIONS AND RECIPES

In this work, we have contributed to the discussion of how we may be fooled in estimating the b-value and its variability. Below, we summarise the most important points, which could serve as a guide for a proper evaluation of significant b-value variations, in particular when the b-value is estimated through the popular equation 2:

(i) The correlation between the b-value and the maximum magnitude of the earthquake catalogue may induce significant unrealistic b-value variations, even when using hundreds of data. In particular, we show that stacking the b-value of hundreds of sequences before mainshocks for a clustered process could lead to the artificial identification of statistically significant low b-values. This problem can be minimised using at least one thousand of data, or replacing the classical statistical test for checking anomalous b-values (i.e., the t-test) with a comparison between the observed b-value variability with the same variability of a reference model which accounts for earthquake clustering, but having a constant b-value.

(ii) A proper estimation of the variability of the b-value depends on the process we are assuming, not on (reasonable) magnitude errors. If we assume that the sample comes from an exponential distribution with a constant parameter, equation 4 and its normal approximation equation 6 hold. Bootstrap and other methods can be justified when we assume that the b-value has its own natural variability violating the exponential distribution of the magnitude. This assumption can be tested with goodness-of-fit tests. In case the magnitude does not conform to an exponential distribution, it is necessary to explain what the estimated b-value means.

(iii) Magnitude binning leads to systematic underestimation of the b-value. This bias may be considered negligible in many applications when the binning is $\Delta M = 0.1$ or smaller. Conversely, if the b-value is calculated for $\Delta M \geq 0.2$ (e.g. using historical catalogues), the b-value could be severely underestimated. If applied to seismic hazard analysis, this may lead to important overestimation. Here we provide an analytical equation to correct the bias caused by the binning to the b-value estimated through equation 2.

(iv) Applying equation 2 to truncated Gutenberg-Richter law may induce significant overestimation of the b-value when the completeness magnitude is 2 units below the maximum magnitude. This bias sharply reduces when the difference between the maximum and the completeness magnitude is 3 or more. We argue that the same bias may hold when using any reasonable tapered version of the Gutenberg-Richter law, instead of the truncated one.

(v) Testing the completeness of the catalogue as a whole is not enough if we analyse the b-value in subsets of the catalogue itself. It is important to check the reliability of the exponential distribution in each subset of data in which we calculate the b-value.

(vi) The b-value and its uncertainty can be meaningfully estimated from the data when the magnitude has an exponential distribution, or distributions that do not show significant departures from it. This can be checked statistically through classical goodness-of-fit tests, such as the Lilliefors test. However, some important departures from an exponential distribution due to catalogue incompleteness that may induce significant low b-values are not always identified by goodness-of-fit tests;

although this comes to the cost of reducing the dataset, we suggest verifying the outcomes of the analysis increasing the completeness magnitude inside a reasonable range.

The general message of this paper is that it is very easy to be fooled in searching for significant variations of the b-value through retrospective analyses, and the recipes listed above can likely reduce, not removing completely, the problem. As far as the search for b-value variations are aimed to improve earthquake forecasting, the best strategy to test this scientific hypothesis is through prospective tests ((Zechar et al. (2010); Schorlemmer et al.(2018); Marzocchi (2018))).

ACKNOWLEDGMENTS

TBD

REFERENCES

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

Amitrano, D., 2003. Brittle-ductile transition and associated seismicity: Experimental and numerical studies and relationship with the b-value, *J. Geophys. Res.*, **108**, 2044.

Amorese, D., Grasso, J.-R., & Rydelek, P. A., 2010. On varying b-values with depth: results from computer-intensive tests for Southern California, *Geophys. J. Int.*, **180**(1), 347–360.

Bender, B., 1983. Maximum likelihood estimation of b-values for magnitude grouped data, *Bull. Seismol. Soc. Am.*, **73**.

Bird, P. & Kagan, Y. Y., 2004. Plate-tectonic analysis of shallow seismicity: Apparent boundary width, beta, corner magnitude, coupled lithosphere thickness, and coupling in seven tectonic settings, *Bull. Seismol. Soc. Am.*, **94**, 2380–2399.

Boslaugh, S., 2012. *Statistics in a Nutshell*, O'Reilly Media, Incorporated.

Cao, A. & Gao, S., 2002. Temporal variation of seismic b-values beneath Northeastern Japan island arc, *Geophys. Res. Lett.*, **29**(9), 481–483.

Castellaro, S., Mulargia, F., & Kagan, Y. Y., 2006. Regression problems for magnitudes, *Geophys. J. Int.*, **165**(3), 913–930.

Chan, C. H., Wu, Y. M., Tseng, T. L., L., T.-L., & Chen, C. C., 2012. Spatial and temporal evolution of b-values before large earthquakes in Taiwan, *Tectonophysics*, **532-535**, 215–222.

Clauset, A., Shalizi, C. R., & Newman, M. E. J., 2009. Power-law distributions in empirical data, *J. Soc. Ind. App. Math. review*, **51**(4), 661–703.

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

Felzer, K. R. & Brodsky, E., 2006. Decay of aftershock density with distance indicates triggering by dynamic stress, *Nature*, **441**, 735–8.

Felzer, K. R., Becker, T. W., Abercrombie, R. E., Ekström, G., & Rice, J. R., 2002. Triggering of the 1999 MW 7.1 Hector Mine earthquake by aftershocks of the 1992 MW 7.3 Landers earthquake, *J. Geophys. Res. Solid Earth*, **107**(B9), ESE 6–1–ESE 6–13.

Frohlich, C. & Davis, S., 1993. Teleseismic b values - or, much ado about 1.0, *J. Geophys. Res.*, **98**, 631–644.

García Hernández, R., D'Auria, L., Barrancos, J., & Padilla, G. D., 2018. On the functional expression of frequency–magnitude distributions: A comprehensive statistical examination, *Bull. Seismol. Soc. Am.*, **109**, 482–486.

Gerstenberger, M., Wiemer, S., & Giardini, D., 2001. A systematic test of the hypothesis that the b value varies with depth in California, *Geophys. Res. Lett.*, **28**.

Godano, C., Lippiello, E., & de Arcangelis, L., 2014. Variability of the b value in the Gutenberg–Richter distribution, *Geophys. J. Int.*, **199**(3), 1765–1771.

Gutenberg, B. & Richter, C. F., 1944. Frequency of earthquakes in California, *Bull. Seismol. Soc. Am.*, **34**(8), 185–188.

Helmstetter, A. & Sornette, D., 2002. Foreshocks explained by cascades of triggered seismicity, *J. Geophys. Res.*, **108**.

Helmstetter, A., Sornette, D., & Grasso, J.-R., 2003. Mainshocks are aftershocks of conditional foreshocks: How do foreshock statistical properties emerge from aftershock laws, *J. Geophys. Res. Solid Earth*, **108**(B1), 2046.

Henderson, J., Main, I., Pearce, R., & Takeya, M., 1994. Seismicity in North-Eastern Brazil: fractal clustering and the evolution of the b value, *Geophys. J. Int.*, **116**(1), 217–226.

Imoto, M., 1991. Changes in the magnitude frequency b-value prior to large (m >= 6.0) earthquakes in Japan, *Tectonophysics*, **193**, 311–325.

Kagan, Y. Y., 1999. Universality of the seismic moment-frequency relation, *Pure Appl. Geophys.*, **155**(2), 537–573.

Kagan, Y. Y., Bird, P., % Jackson, D. D., 2010. Earthquake Patterns in Diverse Tectonic Zones of the Globe, *Pure Appl. Geophys.*, **167**, 721–741.

Kamer, Y. & Hiemer, S., 2015. Data-driven spatial b value estimation with applications to California seismicity: To b or not to b, *J. Geophys. Res. Solid Earth*, **120**, 5191–5214.

Kendall, M. G., 1948. *Rank correlation methods*, C. Griffin.

Kijko, A. & Smit, A., 2012. Extension of the Aki-Utsu b-value estimator for incomplete catalogs, *Bull. Seismol. Soc. Am.*, **102**, 1283–1287.

Kijko, A. & Smit, A., 2017. Estimation of the frequency-magnitude Gutenberg-Richter b-value without making assumptions on levels of completeness, *Seismol. Res. Lett.*, **88**(2A), 311–318.

Knopoff, L., Y. Kagan, Y. Y., & Knopoff, R., 1982. b-values for foreshocks and aftershocks in real and simulated earthquake sequences, *Bull. Seismol. Soc. Am.*, **72**(5), 1663–1676.

Lilliefors, H., 1969. On the Kolmogorov-Smirnov test for the Exponential Distribution with mean unknown, *J. Amer. Statis. Assoc.*, **64**, 387–389.

Lolli, B., & Gasperini, P., 2012. A comparison among general orthogonal regression methods applied to earthquake magnitude conversions, *Geophys. J. Int.*, **190**, 1135–1151.

Lopez-Pineda, L., 2005. Source characteristics of the mw 6.2 Loreto earthquake of 12 march 2003 that occurred in a transform fault in the middle of the gulf of California, Mexico, *Bull. Seismol. Soc. Am.*, **95**, 419–430.

Main, I. G., Meredith, P. G., & Jones, C., 1989. A reinterpretation of the precursory seismic b-value anomaly from fracture mechanics, *Geophys. J. Int.*, **96**(1), 131–138.

Main, I. G., Meredith, P. G., & Sammonds, P. R., 1992. Temporal variations in seismic event rate and b-values from stress corrosion constitutive laws, *Tectonophysics*, **211**(1), 233–246.

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

Marzocchi, W., 2018. Predictive seismology, *Seismol. Res. Lett.*, **89**, 1998–2000.

Meng, X., Yang, H., & Peng, Z., 2018. Foreshocks, b value map, and aftershock triggering for the 2011 Mw 5.7 Virginia Earthquake, *J. Geophys. Res. Solid Earth*, **123**(6), 5082–5098.

Michael, A. & Wiemer, S., 2010. Corssa: the community online resource for statistical seismicity analysis, *Community Online Resource for Statistical Seismicity Analysis*, **1**, 1–13.

Mignan, A. & Woessner, J., 2012. Theme IV - understanding seismicity catalogs and their problems, Tech. rep., Technical Report doi: https://doi. org/10.5078/corssa-00180805.

Mogi, K., 1962. Study of elastic shocks caused by the fracture of heterogeneous materials and its relations to earthquake phenomena, *Bull. Earthq. Res. Inst. Tokyo Univ.*, **40**.

Mori, J. & Abercrombie, R., 1997. Depth dependence of earthquake frequency-magnitude distributions in California: Implication for rupture initiation, *J. Geophys. Res.*, **1021**, 15081–15090.

Nanjo, K. Z. & Yoshida, A., 2017. Anomalous decrease in relatively large shocks and increase in the p and b values preceding the April 16, 2016, M7.3 earthquake in Kumamoto, Japan, *Earth, Planets and Space*, **69**(1), 13.

Naylor, M., Orfanogiannaki, K., & Harte, D., 2010. Exploratory data analysis: Magnitude, space, and time, *Community Online Resource for Statistical Seismicity Analysis*, **3**, 1–42.

Nuannin, P., 2006. The potential of b-value variations as earthquake precursors for small and large events, *(Doctoral dissertation, Acta Universitatis Upsaliensis)*.

Ogata, Y., 1988. Statistical models for earthquake occurrences and residual analysis for point processes, *J. Amer. Statis. Assoc.*, **83**, 9–27.

Ogata, Y., 1998. Space-time point-process models for earthquake occurrences, *Ann. Inst. Stat. Math.*, **50**, 379–402.

Ogata, Y. & Katsura, K., 1993. Analysis of temporal and spatial heterogeneity of magnitude frequency distribution inferred from earthquake catalogues, *Geophys. J. Int.*, **113**(3), 727-738.

Ogata, Y. & Katsura, K., 2006. Immediate and updated forecasting of aftershock hazard, *Geophys. Res. Lett.*, **33**(10).

Ogata, Y. & Katsura, K., 2014. Comparing foreshock characteristics and foreshock forecasting in observed and simulated earthquake catalogs, *J. Geophys. Res. Solid Earth*, **119**(11), 8457–8477.

Ogata, Y. & Yamashina, K., 1986. Unbiased estimate for b-value of magnitude frequency, *J. Phys. Earth*, **34**, 187–194.

Page, R., 1968. Aftershocks and microaftershocks of the great Alaska earthquake of 1964, *Bull. Seismol. Soc. Am.*, **58**(3), 1131–1168.

Papadopoulos, G., Charalampakis, M., Fokaefs, A., & Minadakis, G., 2010. Strong foreshock signal preceding the L'Aquila (Italy) earthquake (mw 6.3) of 6 April 2009, *Nat. Hazards Earth Sys.*, **10**, 19–24.

Parsons, T., 2007. Forecast experiment: Do temporal and spatial b value variations along the Calaveras fault portend M >= 4.0 earthquakes?, *J. Geophys. Res. Solid Earth*, **112**(B3).

Raub, C., Martínez-Garzón, P., Kwiatek, G., Bohnhoff, M., & Dresen, G., 2017. Variations of seismic b-value at different stages of the seismic cycle along the North Anatolian fault zone in Northwestern Turkey, *Tectonophysics*, **712-713**, 232 – 248.

Reasenberg, P., 1999. Foreshock occurrence before large earthquakes, *J. Geophys. Res. Solid Earth*, **104**(B3), 4755–4768.

Rodríguez-Pérez, Q. & Zuñiga, F. R., 2018. Imaging b-value depth variations within the Cocos and Rivera plates at the Mexican subduction zone, *Tectonophysics*, **734-735**, 33–43.

Sandri., L., & Marzocchi, W., 2007. A technical note on the bias in the estimation of the b-value and its uncertainty through the least squares technique, *Ann. Geophys.*, **50**.

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

Scholz, C., 2015. On the stress dependence of the earthquake b-value, *Geophys. Res. Lett.*, **42**.

Schorlemmer, D., Neri, G., Wiemer, S., & Mostaccio, A., 2003. Stability and significance tests for b-value anomalies: Example from the Tyrrhenian Sea, *Geophys. Res. Lett.*, **30**(16).

Schorlemmer, D., Wiemer, S., & Wyss, M., 2004. Earthquake statistics at Parkfield: 1. stationarity of b values, *J. Geophys. Res. Solid Earth*, **109**(B12).

Schorlemmer, D., Wiemer, S., & Wyss, M., 2005. Variations in earthquake-size distribution across different stress regimes, *Nature*, **437**, 539–42.

Schorlemmer, D., & Woessner, J., 2008. Probability of Detecting an Earthquake, *Bull. Seismol. Soc. Am.*, **98**, 2103–2117.

Schorlemmer, D., Werner, M. J., Marzocchi, W., Jordan, T. H., Ogata, Y., Jackson, D. D., Mak, S., Rhoades, D. A., Gerstenberger, M. C., Hirata, N., Liukis, M., Maechling, P., Strader, A., Taroni, M., Wiemer, S., Zechar, J. D., & Zhuang, J., 2018. The collaboratory for the study of earthquake predictability: achievements and priorities, *Seismol. Res. Lett.*, **89**, 1305–1313.

Shi, Y. & Bolt, B., 1982. The standard error of the magnitude-frequency b value, *Bull. Seismol. Soc. Am.*, **72**, 1677–1687.

Singh, A., Mishra, O. P., Rastogi, B., & Kumar, D., 2011. 3-D seismic structure of the Kachchh, Gujarat, and its implications for the earthquake hazard mitigation, *Nat. Hazards*, **57**, 83–105.

Smith, W., 1980. The b-value as an earthquake precursor, *Nature*, **289**, 136–139.

Smith, W. D., 1998. Resolution and significance assessment of precursory changes in mean earthquake magnitudes, *Geophys. J. Int.*, **135**(2), 515–522.

Taroni, M., Marzocchi, W., Schorlemmer, D., Werner, M., Wiemer, S., Zechar, J., Heiniger, L., & Euchner, F., 2018. Prospective CSEP evaluation of 1-day, 3-month, and 5-yr earthquake forecasts for Italy, *Seismol. Res. Lett.*, **89**(4), 1251–1261.

Tinti, S. & Mulargia, F., 1985. Effectes of magnitude uncertainties on estimating the parameters in the Gutenberg-Richter frequency-magnitude law, *Bull. Seismol. Soc. Am.*, **75**, 1681–1697.

Tinti, S. & Mulargia, F., 1987. Confidence intervals of b values for grouped magnitudes, *Bull. Seismol. Soc. Am.*, **77**(6), 2125–2134.

Utsu, T., 1965. A method for determining the value of "b" in a formula log n= a-bm showing the magnitude-frequency relation for earthquakes, *Geophys. Bull. Hokkaido Univ.*, **13**, 99–103.

Utsu, T., 1966. A statistical significance test of the difference in b-value between two earthquake groups, *J. Phys. Earth*, **14**(2), 37–40.

Westerhaus, M., Wyss, M., Yilmaz, R., & Zschau, J., 2002. Correlating variations of b values and crustal deformations during the 1990s may have pinpointed the rupture initiation of the Mw= 7.4 Izmit earthquake of 1999 August 17, *Geophys. J. Int.*, **148**(1), 139–152.

WGCEP, W. G. o. C. E. P., 2003. Earthquake probabilities in the San Francisco Bay region: 2002–2031, Tech. rep., U.S. Geological Survey Open-File Report 2003-214.

Wiemer, S. & Wyss, M., 2002. Mapping spatial variability of the frequency-magnitude distribution of earthquakes, in *Adv. geophys.*, vol. 45, pp. 259–V, Elsevier.

Wiemer, S., McNutt, S., & Wyss, M., 1998. Temporal and three-dimensional spatial analyses of the frequency–magnitude distribution near Long Valley Caldera, California, *Geophys. J. Int.*, **134**(2), 409–421.

Wiemer, S., Yoshida, A., Hosono, K., Noguchi, S., & Takayama, H., 2005. Correlating seismicity parameters and subsidence in the Tokai region, Central Japan, *J. Geophys. Res. Solid Earth*, **110**(B10).

Wilcoxon, F., 1945. Individual comparisons by ranking methods, *Biometrics bulletin*, **1**(6), 80–83.

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

Woessner, J., Laurentiu, D., Giardini, D., Crowley, H., Cotton, F., Grünthal, G., Valensise, G., Arvidsson, R., Basili, R., & Demircioglu, M. B. o., 2015. The 2013 European seismic hazard model: key components and results, *Bull. Earth. Eng.*, **13**(12), 3553–3596.

Wyss, M. & Lee, W. H. K., 1973. Time variations of the average earthquake magnitude in Central California, in *Proceedings of the conference on tectonic problems of the San Andreas fault system*, pp. 24–42, Stanford University Geol. Sci.

Wyss, M. & Wiemer, S., 2000. Change in the probability for earthquakes in Southern California due to the Landers magnitude 7.3 earthquake, *Science*, **290**(5495), 1334–1338.

Zechar, J. D., Schorlemmer, D., Liukis, M., Yu, J., Euchner, F., Maechling, P.J., & Jordan, T. H., 2010. The Collaboratory for the Study of Earthquake Predictability perspective on computational earthquake science, *Concurrency Comput. Pract. Ex.* , **22**, 1836–1847.

Zöller, G. & Hainzl, S., 2002. A systematic spatiotemporal test of the critical point hypothesis for large earthquakes, *Geophys. Res. Lett.*, **29**(11), 53–1.

Table 1. Percentage of synthetic catalogues for which the Lilliefors (first row) and the *t*-test (second row) is rejected, as a function of the missing rate $\nu(M)$ and the sample size.

	$\nu(M) = 0.5$	$\nu(M) = 0.31$	$\nu(M) = 0.16$	$\nu(M) = 0.07$	$\nu(M) = 0.02$	$\nu(M) = 0.01$
			$N = 50$			
Lilliefors Test	15.1%	7.6%	5.9%	5.9%	5%	5.1%
t-test	29%	14.7%	8%	6.5%	5.6%	6.1%
			$N = 200$			
Lilliefors Test	41%	12.6%	5.7%	5.1%	4.3%	4.8%
t-test	70.8%	28.8%	12.2%	7.1%	6.2%	6.2%
			$N = 500$			
Lilliefors Test	78.1%	24.9%	9.6%	6.1%	5.2%	4.3%
t-test	95.8%	51.7%	17.6%	9.7%	7.7%	6.8%
			$N = 2000$			
Lilliefors Test	100%	77.5%	18.1%	6.8%	4.8%	4.8%
t-test	100%	96.9%	43.6%	15.4%	8.5%	7.7%

Table 2. Proportion of subcatalogues/aggregated catalogues for which the Lilliefors (first row) and the *t*-test (second row) is rejected.

	Subcat 1	Subcat 2	Subcat 3	Subcat 4	Subcat 5	Whole catalogue
Lilliefors Test	4.6%	4.3%	4.4%	5.6%	40.1%	10.6%
t-test	6.1%	6.2%	5.9%	6.7%	69.1%	28.5%

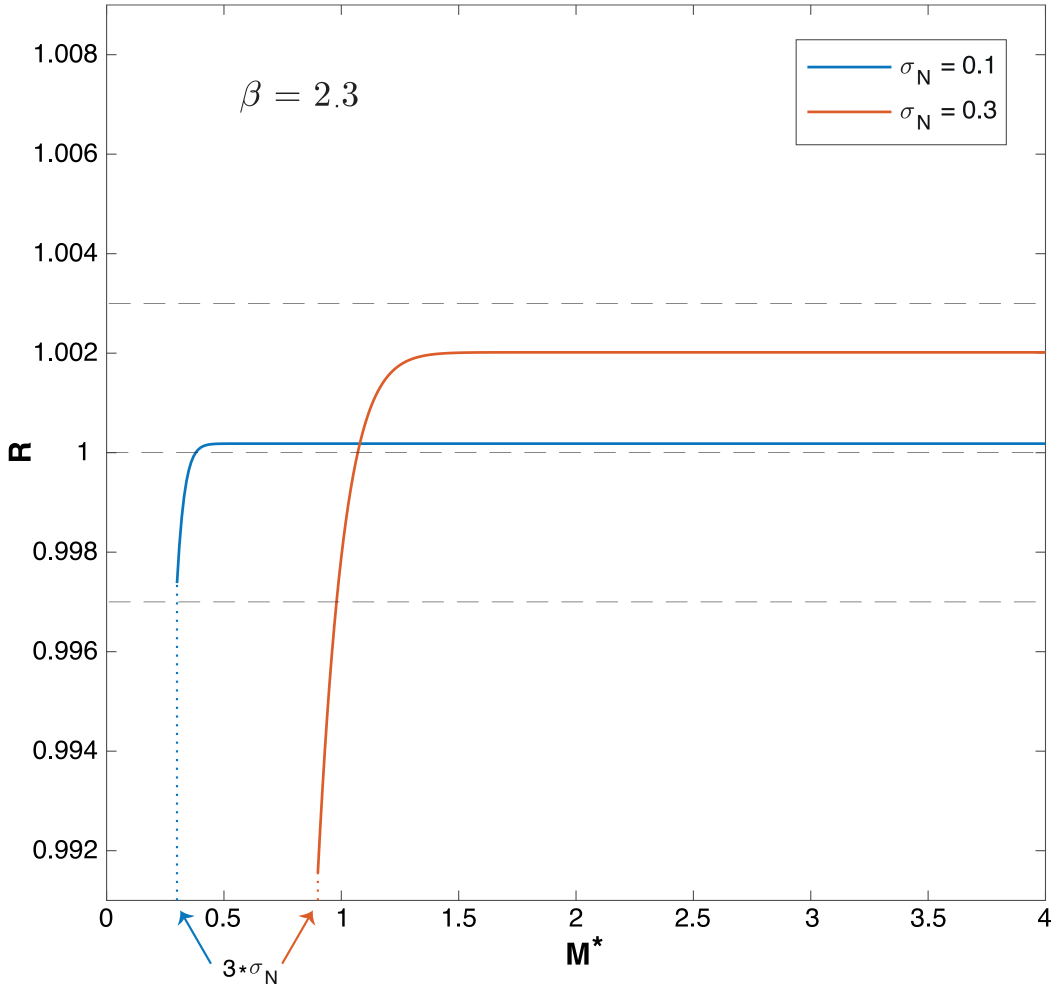

Figure 1. Plot of the ratio R (equation 11) between the magnitude PDFs with and without the random Gaussian error, obtained for $\beta = 2.3$ and $\sigma_{\mathcal{N}} = 0.1, 0.3$. As a reference, dashed horizontal lines identify the range 1 ± 0.003, in which the ratio could be considered negligible. The vertical dotted lines show the initial points $3\sigma_{\mathcal{N}}$ of the magnitude ranges to be considered to avoid loss and/or addition of data caused by the Gaussian error (see text for more details).

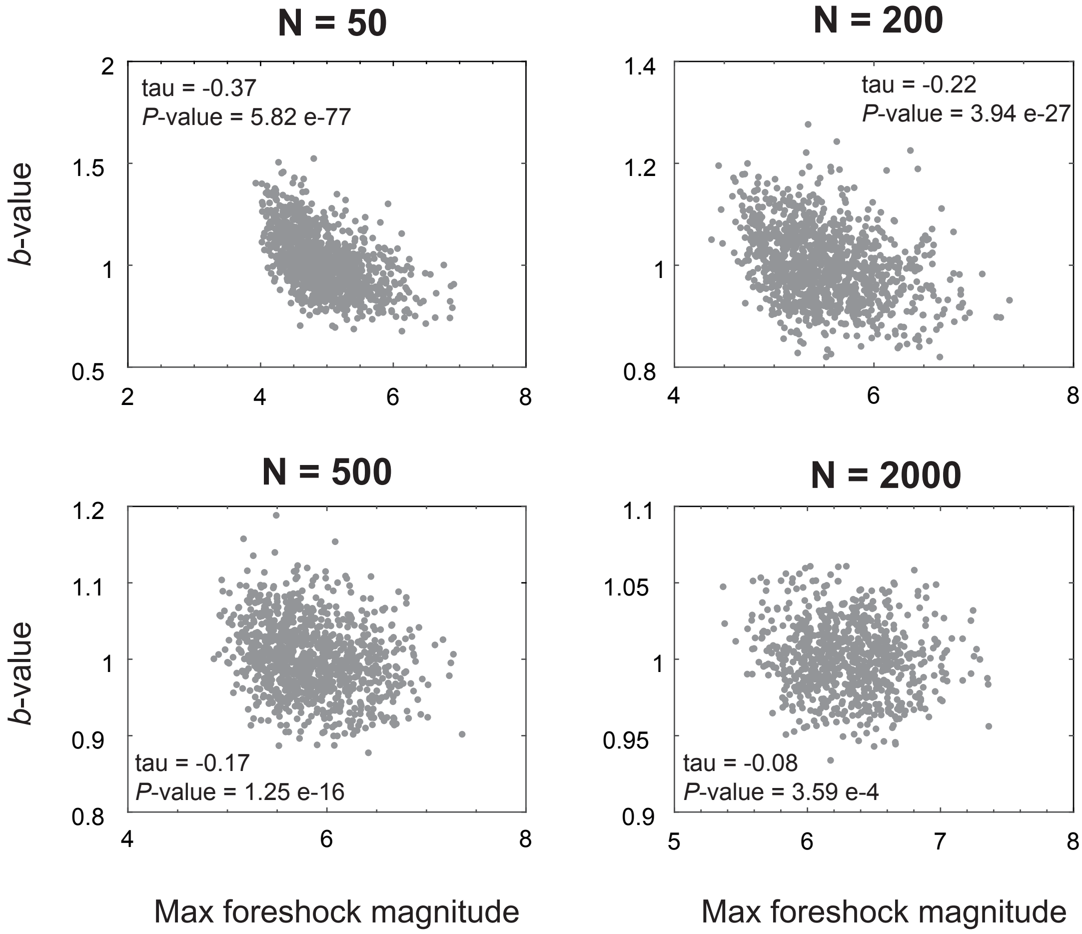

Figure 2. Plot of the b-value obtained by equation 2 (with the modification $M_{min} = M_c - \Delta M/2$ to account for binning) as a function of the maximum magnitude foreshock in synthetic ETAS catalogues. The four panels are relative to a different number of data used to calculate the b-value. In each plot we report the P-value obtained by the Kendall-tau test for the null hypothesis of no correlation.

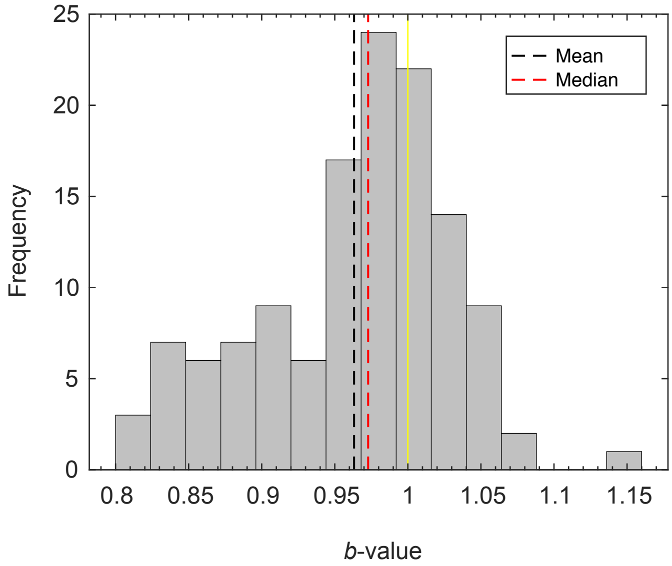

Figure 3. Histogram of the b-value estimated through equation 2 (with the modification $M_{min} = M_c - \Delta M/2$ to account for binning) for all the stacked foreshock sequences in the ETAS synthetic catalogue (see text for more information). The vertical solid yellow line shows the theoretical value used to generate the synthetic ETAS catalogue, and the dashed red and black lines show, respectively, the median and the mean of the distribution of the b-value.

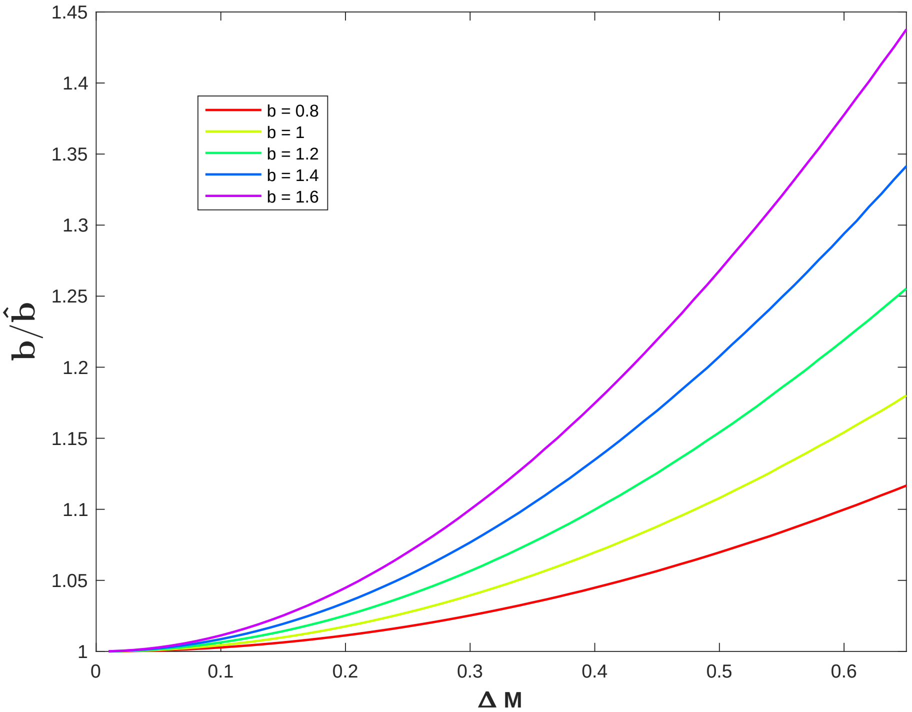

Figure 4. Plot of the ratio between the corrected b-value and the estimation given by equation 2 as a function of the amplitude of the magnitude binning. Colors refer to different b-values.

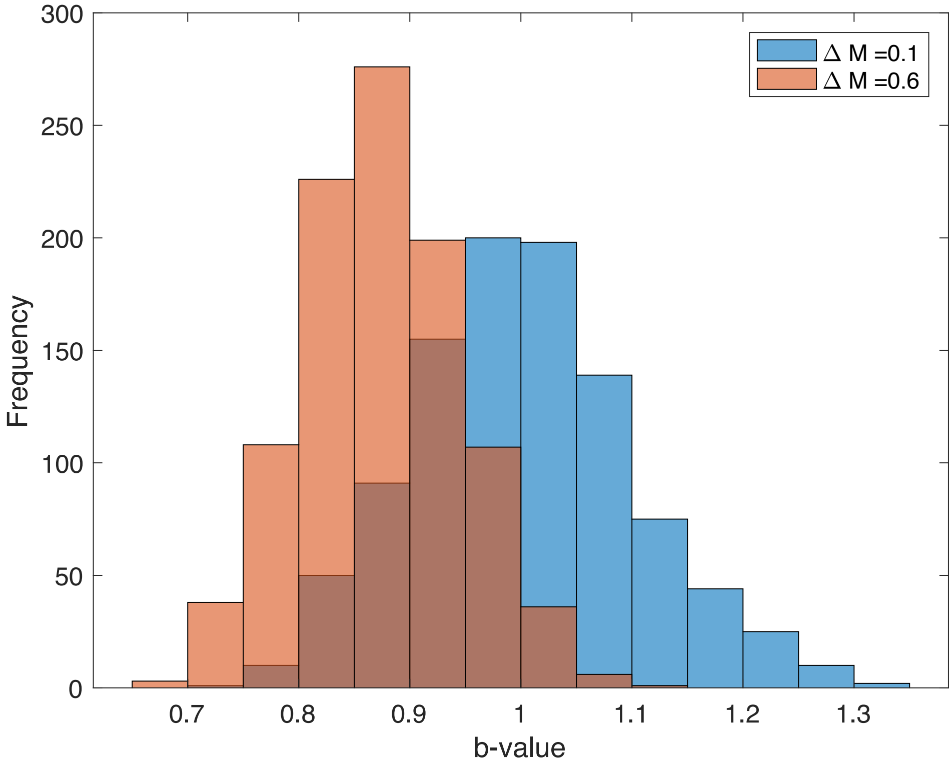

Figure 5. Histogram of the b-values estimated through equation 2 for 1000 simulations of 100 magnitudes each, with $\Delta M = 0.1$ and 0.6.

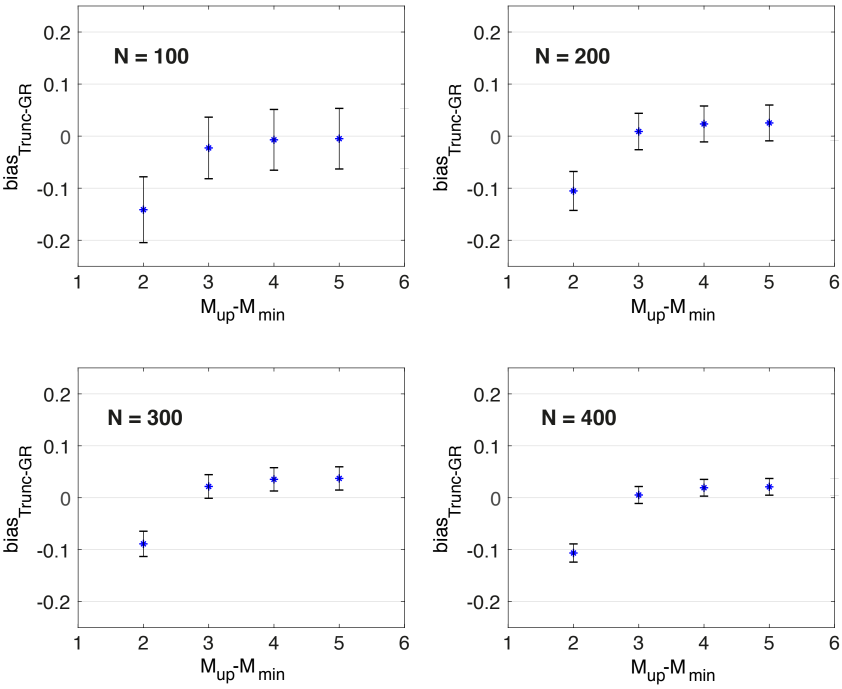

Figure 6. Mean difference between the β-values estimated for 100 simulated datasets drawn from both N Trunc-GR (formula 20) and N classical GR (formula 2) distributions, for $N = 100, 200, 300, 400$; one-σ error bars are also shown.

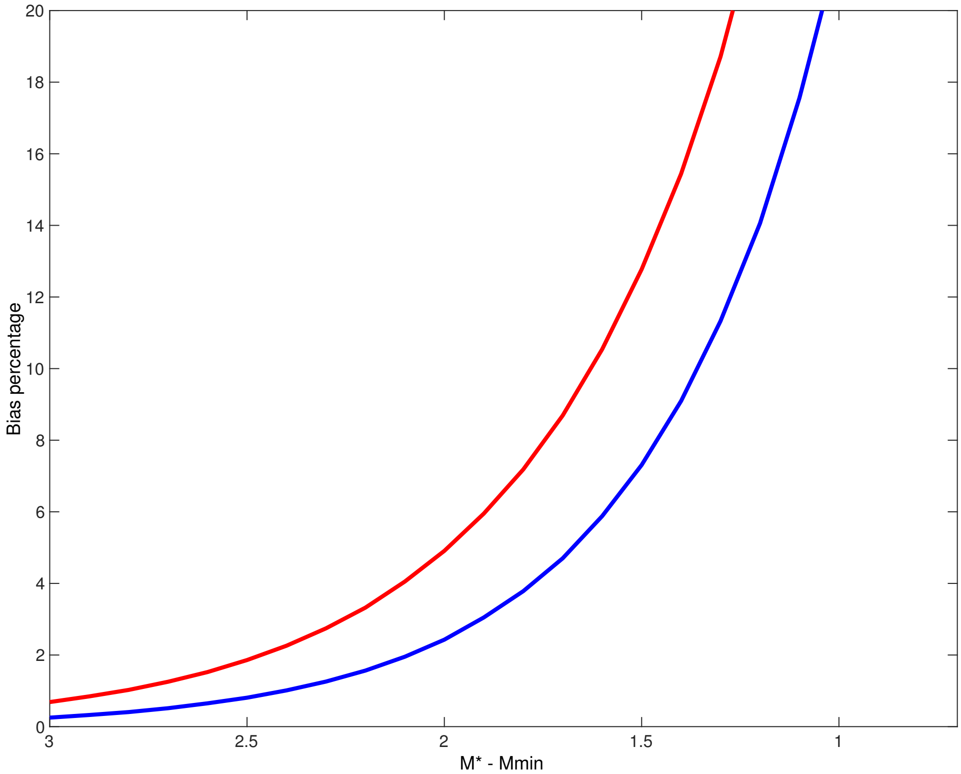

Figure 7. Bias percentage in the calculation of the b-value through equation 2 as a function of the difference between upper and lower limit of the magnitude. In x-axis the term $M^{\star}$ is the maximum magnitude of the truncated GR (red curve), and the corner magnitude of the tapered GR (blue curve).

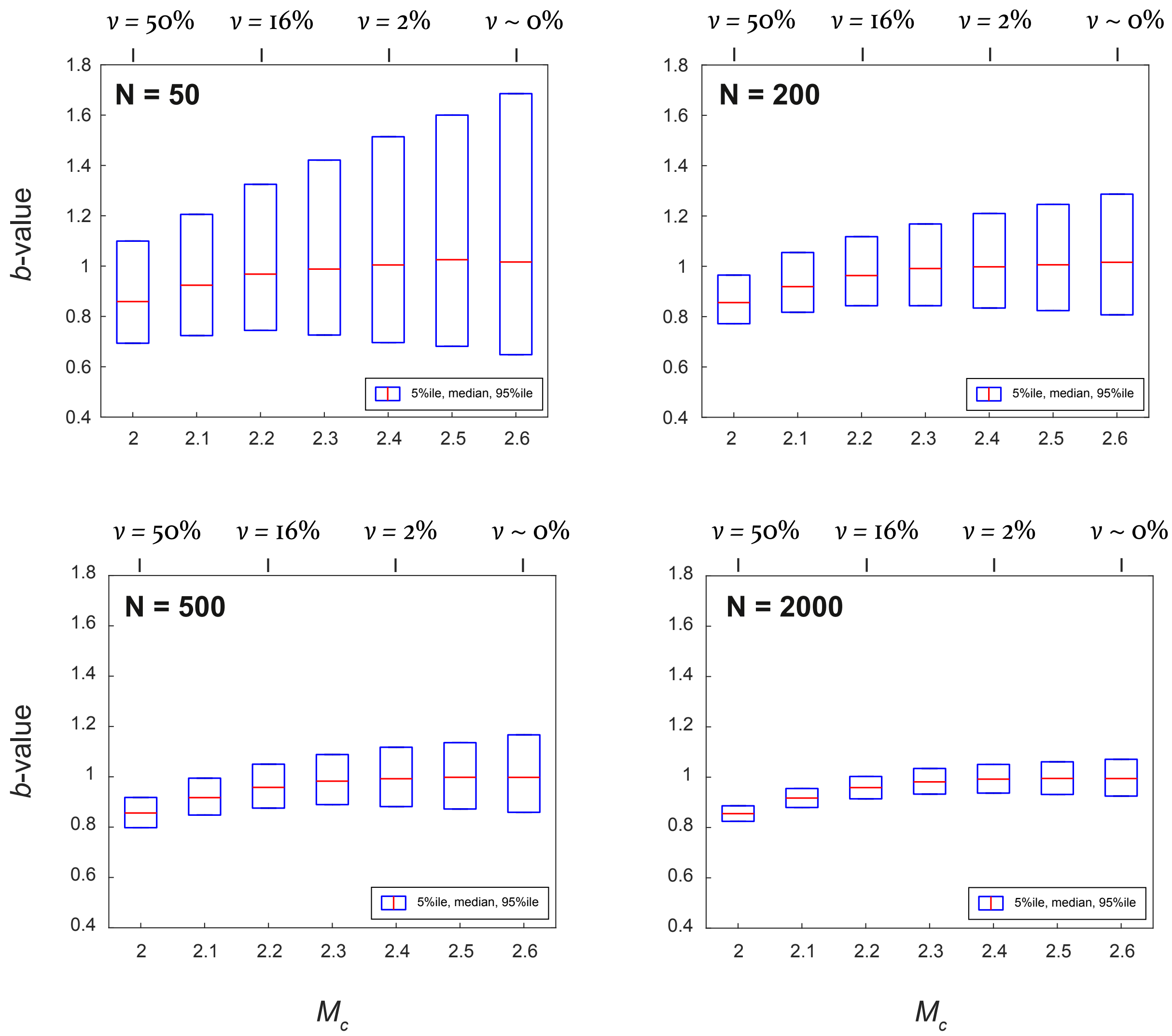

Figure 8. For each sample size ($N = 50, 200, 500, 2000$), we group the estimated b-values by the threshold magnitude M_c, with $\mu \leq M_c \leq \mu + 3 \cdot \sigma$ (and the associated ν). The obtained distributions are summarized by boxes, whose upper (lower) limit corresponds to the 95^{th} (5^{th}) percentile. The middle red line represents the median of each distribution.

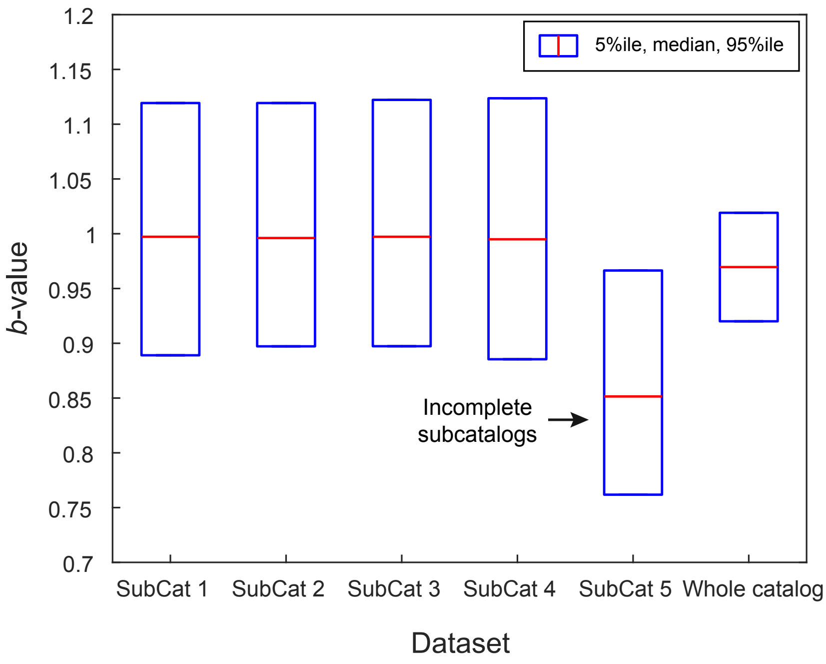

Figure 9. The distributions of the 1000 b-values estimated for the five subsets and the whole catalogue are shown. They are summarized by boxes, whose upper (lower) limit corresponds to the 95^{th} (5^{th}) percentile. The middle red line represents the median of each distribution. The box referring to Subcatalogue 5 represents the b-values estimated for the incomplete subsets.

Table A1. Input parameters

Gutenberg-Richter Law	Omori's Law			Aftershocks decay with distance
b-value	c	p	a	n
1	0.095	1.34	0.008	1.37

APPENDIX A: ETAS SIMULATION

We use the stochastic program from (Felzer et al.(2002)) to simulate a synthetic catalogue based on the ETAS model [(Ogata(1988))]. The code models the Gutenberg-Richter law for earthquake magnitudes, the Omori's law for earthquake times of occurrence and the (Felzer & Brodsky(2006)) law of aftershock decay with distance. Inverse method is used to draw samples from the empirical distributions. We model all the events as point sources and we leave unchanged the parameters as provided in the code (see Table A1), except the range of possible distances, in km, between triggered and triggering earthquakes, which we set equals to [0.001 100].