Geophysical Journal International

Geophys. J. Int. (2024) **238,** 433–458
Advance Access publication 2024 May 03
GJI Seismology

https://doi.org/10.1093/gji/ggae159

The estimation of *b*-value of the frequency–magnitude distribution and of its 1σ intervals from binned magnitude data

S. Tinti [1] and P. Gasperini [1,2]

[1] *Dipartimento di Fisica e Astronomia 'Augusto Righi', Università di Bologna, Viale Berti Pichat, 8, I-40127 Bologna, Italy. E-mail: paolo.gasperini@unibo.it*
[2] *Istituto Nazionale di Geofisica e Vulcanologia, Sezione di Bologna, Viale Berti Pichat, 6/2, I-40127 Bologna, Italy*

Accepted 2024 April 27. Received 2024 March 25; in original form 2023 October 2

SUMMARY

The estimation of the slope (*b*-value) of the frequency–magnitude distribution of earthquakes is based on a formula derived by Aki decades ago, assuming a continuous exponential distribution. However, as the magnitude is usually provided with a limited resolution, its distribution is not continuous but discrete. In the literature, this problem was initially solved by an empirical correction (due to Utsu) to the minimum magnitude, and later by providing an exact formula such as that by Tinti and Mulargia, based on the geometric distribution theory. A recent paper by van der Elst showed that the *b*-value can be estimated also by considering the magnitude differences (which are proven to follow an exponential discrete Laplace distribution) and that in this case the estimator is more resilient to the incompleteness of the magnitude data set. In this work, we provide the complete theoretical formulation including (i) the derivation of the means and standard deviations of the discrete exponential and Laplace distributions; (ii) the estimators of the decay parameter of the discrete exponential and trimmed Laplace distributions and (iii) the corresponding formulas for the parameter *b*. We deduce (iv) the standard 1σ intervals for the estimated *b*. Moreover, we are able (v) to quantify the error associated with the Utsu minimum-magnitude correction. Furthermore, we have discussed the formulas to produce statistically independent magnitude differences. We tested extensively the *b*-value estimators on simulated synthetic data sets including complete catalogues as well as catalogues affected by a strong incompleteness degree such as aftershock sequences where the incompleteness is made to vary from one event to the next. We have also analysed the real aftershock sequence of the 30/10/2016 Norcia (central Italy) to integrate the finding of the simulations. To judge the performance of the various estimators we have introduced an index *p* that can be seen as a non-parametric extension of the Student's *t* index. The main outcomes of this paper are that (1) the *b*-value estimators devised for continuous magnitude data are not adequate for binned magnitudes, (2) for complete data sets, estimators based on magnitudes and on magnitude differences provide substantially equivalent results, (3) for incomplete magnitude data sets, estimators based on magnitude differences provide better results and (4) for incomplete aftershock sequences there is no evidence that methods based on positive magnitude differences are superior than other methods using differences. This conclusion is further confirmed by our analysis of the above-mentioned Norcia seismic sequence. This last finding contrasts with the van der Elst's claim that the so called b_+ method is the most adequate to treat real aftershock sequences.

Key words: Earthquake hazards; Earthquake interaction, forecasting and prediction; Statistical seismology.

INTRODUCTION

The *b*-value of the frequency–magnitude distribution (Gutenberg & Richter 1944)

$$\log_{10} N = a - bM \qquad (1)$$

is indicated by some researchers as a proxy of the level of differential stress within the Earth (Scholz 1968; Amitrano 2003, 2015) and thus as an index of the state of preparation of future strong earthquakes (Gulia & Wiemer 2010; Gulia *et al.* 2018, 2019, 2020). Some papers demonstrated that the *b*-value is negatively correlated with the rake of the focal mechanism (Schorlemmer *et al.* 2005;

Petruccelli *et al.* 2018, 2019a) and with the source depth (Spada *et al.* 2013; Petruccelli *et al.* 2019b). However, these results are controversial and others argued that b-value variations are statistically insignificant as they are due to artefacts of the methods used to determine the b-value (Kagan 1999, 2002, 2003; Bird & Kagan 2004).

One of the most critical aspects in b-value computations is the determination of the magnitude completeness threshold for the seismic data set used (e.g. Woessner & Wiemer 2005; Mignan & Woessner 2012) because an underestimation of the threshold might bias (lowering) the estimated b-value, whereas an overestimation might reduce the sample size so much that the resulting b-value uncertainty results intolerable.

Aki (1965), assuming a continuous exponential distribution of magnitudes, deduced the formulas for the estimation of the b-value and of its standard 1σ interval by the maximum likelihood method as:

$$b = \frac{1}{\ln(10)\left(\bar{M} - M_c\right)} \quad (2)$$

$$\sigma_b = \frac{b}{\sqrt{N}}, \quad (3)$$

where $\bar{M}$ is the average magnitude, M_c is the minimum (completeness) magnitude and N is the number of magnitude data in the sample. The eq. (2) was also derived by Utsu (1965) by the method of moments. Utsu (1966) evidenced that the value estimated by the eq. (2) is biased (higher) when magnitudes are binned (usually to one decimal digit) and proposed an approximate correction to the original formula:

$$b = \frac{1}{\ln(10)\left(\bar{M} - M_c + \delta\right)}, \quad (4)$$

where δ is one half of the binning size (e.g. 0.05).

Studying in detail the statistical distribution of b, Shi & Bolt (1982) suggested the following formula for the standard deviation σ_b of the continuous distribution:

$$\sigma_b = \ln(10)\, b^2 \sqrt{\frac{\sum_{i=1}^{N}\left(M_i - \bar{M}\right)^2}{N(N-1)}}. \quad (5)$$

Actually, if the magnitude data are binned, their distribution is not continuous anymore, but discrete and this implies changes in the estimators.

Bender (1983) analysed the problem of estimating the b-value from magnitude grouped data and found that the maximum likelihood estimate of b is the value for which:

$$\frac{q}{1-q} - \frac{nq^n}{1-q^n} = \sum_{i=1}^{n} \frac{(i-1)\,k_i}{N}, \quad (6)$$

where $q = \exp[-2\delta \ln(10) b]$, k_i is the number of earthquakes in the ith magnitude interval of width 2δ and n is the number of magnitude intervals from M_c to the maximum magnitude of the data set. An explicit expression for b was not derived by Bender (1983) and then Bender's method implies the numerical solution of the eq. (6).

Guttorp & Hopkins (1986) showed that the maximum likelihood estimate of b in case of magnitude data with limited accuracy 2δ is:

$$b = \frac{1}{2\delta \ln(10)} \ln\left[1 + \frac{2\delta}{\bar{M} - M_c}\right] = \frac{1}{2\delta \ln(10)} \ln\left[\frac{\bar{M} - M_c + 2\delta}{\bar{M} - M_c}\right] \quad (7)$$

Tinti & Mulargia (1987) derived the exact equation in the case of grouped magnitudes in a paper focused on the confidence intervals, providing the form:

$$b = -\frac{1}{2\delta \ln(10)} \ln\left[\frac{\frac{(\bar{M} - M_c + \delta)}{2\delta} - 0.5}{\frac{(\bar{M} - M_c + \delta)}{2\delta} + 0.5}\right] \quad (8)$$

which is perfectly equivalent to the eq. (7).

Marzocchi *et al.* (2020) suggested that, when data are binned, the b-value computed through the Utsu formula (eq. 4), say $b_{\rm Utsu}$, has to be corrected as follows:

$$b_{\rm corrected} = \frac{1}{2\delta \ln(10)} \ln\left[\frac{1 + b_{\rm Utsu}\delta \ln(10)}{1 - b_{\rm Utsu}\delta \ln(10)}\right] \quad (9)$$

Observe that, by substituting eq. (4) in the eq. (9), we have:

$$b_{\rm corrected} = \frac{1}{2\delta \ln(10)} \ln\left[\frac{1 + \frac{1}{\ln(10)\left(\bar{M} - M_c + \delta\right)} \ln(10)\,\delta}{1 - \frac{1}{\ln(10)\left(\bar{M} - M_c + \delta\right)} \ln(10)\,\delta}\right] = \frac{1}{2\delta \ln(10)} \ln\left[\frac{\frac{\bar{M} - M_c + \delta + \delta}{\left(\bar{M} - M_c + \delta\right)}}{\frac{\bar{M} - M_c + \delta - \delta}{\left(\bar{M} - M_c + \delta\right)}}\right] \quad (10)$$

that is exactly equivalent to the eq. (7).

Van der Elst (2021) showed that in case of discretized data, the exact formula for estimating b is:

$$b = \frac{1}{\delta \ln(10)} \coth^{-1}\left[\frac{1}{\delta}\left(\bar{M} - M_c + \delta\right)\right] \quad (11)$$

where $\coth^{-1}$ is the inverse of the hyperbolic cotangent function. Recalling the definition of $\coth^{-1}$, that is:

$$\coth^{-1}(x) = \frac{1}{2} \ln\left(\frac{x+1}{x-1}\right), \quad (12)$$

it is easy to show that even such an equation is equivalent to the eq. (7). Note that Zöller *et al.* (2010), treating magnitudes affected by uncertainties modelled as a box function of size 2δ, derived a formula close to the eq. (11) in their eq. (17) but did not derive an explicit expression for b o β and only provided the first-order approximation eq. (4) already suggested by Utsu (1966). Van der Elst (2021) also showed that the b-value can be consistently computed by using the absolute magnitude differences $|\Delta M|$ (since they follow the exponential discrete Laplace distribution) by means of the formula:

$$b = \frac{1}{2\delta \ln(10)} {\rm csch}^{-1}\left[\frac{1}{2\delta} \overline{|\Delta M|}\right] \quad (13)$$

where ${\rm csch}^{-1}$ is the inverse of the hyperbolic cosecant and $\overline{|\Delta M|}$ is the average of the absolute magnitude differences. Recalling the definition of ${\rm csch}^{-1}$, that is:

$${\rm csch}^{-1}(x) = \ln\left(\frac{1}{x} + \sqrt{\frac{1}{x^2} + 1}\right) \quad (14)$$

one finds that the eq. (13) can also be written in terms of natural logarithm as:

$$b = \frac{1}{2\delta \ln(10)} \ln\left[\frac{2\delta + \sqrt{4\delta^2 + \left(\overline{|\Delta M|}\right)^2}}{\overline{|\Delta M|}}\right] \quad (15)$$

As incompleteness also affects the Laplace distribution of magnitude differences, van der Elst (2021) suggested discarding all $\Delta M = 0$ and then only considering absolute differences not lower than the binning size $\Delta M'_c = 2\delta$. In this case, he showed that the b-value estimator becomes formally equivalent to that of binned magnitudes of the eq. (11) provided that $\overline{|\Delta M|}$ and $\Delta M'_{\mathrm{c}}$ replace $\bar{M}$ and M_{c}, respectively:

$$b = \frac{1}{\delta \ln(10)} \coth^{-1}\left[\frac{1}{\delta}\left(\overline{|\Delta M|} - \Delta M'_{\mathrm{c}} + \delta\right)\right] \quad (16)$$

It is obvious that the eq. (16) can be written in terms of natural logarithm as:

$$b = \frac{1}{2\delta \ln(10)} \ln\left(\frac{\overline{|\Delta M|} - \Delta M'_{\mathrm{c}} + 2\delta}{\overline{|\Delta M|} - \Delta M'_{\mathrm{c}}}\right) \quad (17)$$

Van der Elst did not derive any expressions for the 1σ intervals but suggested computing them by means of the bootstrap method (Hurvich & Tsai 1989). He also asserted that the estimation of the b-value is more stable and robust if only positive magnitude differences are used in the eq. (16). Indeed, van der Elst (2021) did not give too many details either on his formulation or on his preference for the positive differences and this might cause somebody to misapply the b-positive method and in particular to not correctly considering the problem of magnitude incompleteness. Hence, in this paper, we provide (see Appendices A–H): (i) the complete theoretical derivation of the first two moments of the discrete exponential distribution; (ii) the estimators of the decay parameter of the discrete exponential as well as of the discrete Laplace distributions, even in case of distribution trimming; (iii) the corresponding formulas for estimating the parameter b. Moreover, as explained in Appendix H we deduce (iv) the lower and upper sigma values, say σ_1 and σ_2, for the estimated b valid in the case of discrete exponential variables as:

$$\sigma_1 = \tilde{b} - b_1, \quad b_1 = \frac{1}{2\delta \ln(10)} \ln\left[\frac{c + \sqrt{\frac{c}{N}}}{1 + \sqrt{\frac{c}{N}}}\right] \quad (18)$$

$$\sigma_2 = b_2 - \tilde{b}, \quad b_2 = \frac{1}{2\delta \ln(10)} \ln\left[\frac{c - \sqrt{\frac{c}{N}}}{1 - \sqrt{\frac{c}{N}}}\right], \quad (19)$$

where

$$c = \exp\left(2\delta \ln(10)\,\tilde{b}\right) = 10^{2\delta\tilde{b}} \quad (20)$$

and $\tilde{b}$ is the estimate of b. The above values of b_1 and b_2 are the endpoints of the 1σ interval and the associated σ can be taken as the half-amplitude of such interval, that is as the mean of σ_1 and σ_2, which leads to the formula:

$$\sigma = \frac{\sigma_1 + \sigma_2}{2} = \frac{b_2 - b_1}{2} \quad (21)$$

This expression replaces the eq. (3) and the eq. (5) and applies to estimates made through the eq. (7) and the eq. (17). In addition, we derive v) the 1σ interval also when b is estimated through the formula (eq. 15), that is for the distributions of the absolute value of magnitude differences [see formulas (H17a), (H17b) and (H17c) in the Appendix H]. In the Appendix F, we demonstrate that (vi) the Utsu correction (eq. 4) coincides with the expansion of the exact formula (eq. 7) truncated to second order.

In this paper, we will evaluate the b-value and the corresponding 1σ intervals by using the formulas presented in the introduction on a large number of simulated data sets, with particular attention given to cases of incomplete magnitude samples. The details on how we produce complete and incomplete synthetic data sets and how we model aftershock sequences are given in Appendix I. Further, we will apply the methods also to a real earthquake sequence, taken from the Horus seismic catalogue (Lolli *et al.* 2020), consisting of the aftershocks of the $M_{\mathrm{w}} = 6.6$ Norcia (central Italy) earthquake occurred on 30/10/2016 (Chiaraluce *et al.* 2017; Improta *et al.* 2019).

COMPUTING MAGNITUDE DIFFERENCES

We point out that to compute magnitude differences in a N-size magnitude sequence taken in chronological order one can proceed essentially in two ways: in the first case, one computes the difference between magnitudes of the second and the first shocks and then between the third and the second one and so on up to the last one:

$$\Delta M_i = M_{i+1} - M_i, \quad i = 1, 2, \ldots, N - 1. \quad (22)$$

This maximizes the number of data (in all, $N - 1$), but the differences are not independent from one another since the magnitude M_i is used to compute two consecutive differences, and this might produce some statistical bias. In the second way, one computes the difference between magnitudes of the second and the first shocks and then between the fourth and the third one and so on up to the last one, that is:

$$\Delta M_i = M_{2i} - M_{2i-1}, \quad i = 1, 2, \ldots, N/2. \quad (23)$$

This grants that the differences are all independent from one another, but it halves the number of data.

This issue has not been treated adequately in the literature and will be given the deserved attention later in this paper.

PERFORMANCE INDEX

To evaluate the goodness of the various methods, we introduce an index p we have devised specifically to this purpose, that, given a random sample, is suitable to measure how close a given 'characteristic value' derived from the sample is to a given 'target value'. In our case, the sample is the set of the K estimators $\tilde{b}_i$, $(i = 1, 2, \ldots, K)$, derived through one of the estimation formulas given above, the 'characteristic value' we like to evaluate is the sample mean $\bar{b}_K$, while the 'target value' is the b-value used to generate the K random data sets. To compute the index p, let's count the number K_+ of $\tilde{b}_i$ that are larger than $\bar{b}_K$ and the number L_+ of $\tilde{b}_i$ that are larger than b. Further, let's count the number K_- of $\tilde{b}_i$ that are smaller than $\bar{b}_K$ and the number L_- of $\tilde{b}_i$ that are smaller than b. Usually, we expect that $K_+ + K_- = K$, and that $L_+ + L_- = K$, but it can happen that some of the $\tilde{b}_i$ accidentally equals $\bar{b}_K$ or b. Consequently, the sums $K_+ + K_-$ and $L_+ + L_-$ might be smaller than K, and it is more convenient to count all quantities separately. We define the performance index as:

$$\begin{aligned} p &= 1 \quad \text{if } \bar{b}_K = b \\ p &= \tfrac{L_+}{K_+} \quad \text{if } \bar{b}_K < b \\ p &= \tfrac{L_-}{K_-} \quad \text{if } \bar{b}_K > b. \end{aligned} \quad (24)$$

It's easy to see that the index p takes values between 0 and 1. Indeed, note that by definition, if $\bar{b}_K < b$, then L_+ cannot be larger than K_+ and therefore p cannot exceed 1. The same holds when $\bar{b}_K > b$. The way p is computed makes it a non-parametric index, very easy to obtain, and applicable to any kind of sample and to any pairs of variables, characteristic and target, one likes to compare. If the sample data number K is so large that one can approximate the sample frequency occurrences with a density probability curve, then the above criterion (eq. 24) can be expressed in terms of probability ratios as:

$$p = \frac{Prob(\tilde{b}>b)}{Prob(\tilde{b}>\bar{b}_K)} \quad \text{if } \bar{b}_K < b$$
$$p = \frac{Prob(\tilde{b}<b)}{Prob(\tilde{b}<\bar{b}_K)} \quad \text{if } \bar{b}_K > b. \qquad (25)$$

If variables $\tilde{b}_i$ happen to follow a symmetric distribution, then $\bar{b}_K$ coincides with the distribution median, implying that $Prob(\tilde{b} < \bar{b}_K) = Prob(\tilde{b} > \bar{b}_K) = 1/2$ and the criterion simplifies further to:

$$p = 2\,Prob\left(\tilde{b} > b\right) \; if \; \bar{b}_K < b; \quad p = 2\,Prob\left(\tilde{b} < b\right) \text{ if } \bar{b}_K > b. \qquad (26)$$

Further, if the distribution is Gaussian with standard deviation σ, the above formulas can be easily given in terms of the error function of the normalized variables and p identifies with the p-value of a two-sided Student's t test, that is $p = [1 - \text{erf}(|b - \bar{b}_K|/\sqrt{2}\sigma)]$.

In this paper, we will rely on the index p to express the evaluation of the estimators. We will use it like the p-value of a null-hypothesis test, where the null hypothesis is that the estimator is acceptable. If $p < \alpha$ where α is the significance level, then we reject the null hypothesis and consider the estimator unacceptable. In this analysis, we take $\alpha = 0.05$. A further way we use p is assuming that the performance of the estimator method is an increasing function of p and that if the index of a method is significantly larger than the one of another, then the former shows a better performance.

RESULTS FOR SIMULATED COMPLETE DATA SETS

We first compare the various estimators of eqs (2), (4), (6), (7), (15) and (17) on complete binned data sets simulated by the procedure specified by the eqs (I1)–(I3) given in the Appendix I. In this analysis the parameter M_{min} used to generate the random samples and the parameter M_c appearing in the estimator formulas are assumed to be equal. In Table 1, we report the average values $\bar{b}_K$ and the corresponding standard deviations S_K computed on a set of $K = 10\,000$ simulated samples each including $N = 1000$ magnitudes, with binning size $2\delta = 0.1$ and with $b = 1$. When applying methods that use magnitudes (i.e. eqs 2, 4, 6 and 7), all samples have the same number of data ($N = 1000$). This is also true when untrimmed magnitude differences are used (see the eq. 15). In this case, however, the data number depends on the way such differences are computed, since differences made through the eq. (22) and the eq. (23) lead to $N = 999$ and $N = 500$, respectively. When null differences are trimmed away [as is required by the estimator of the eq. (17)], the number of remaining data changes from sample to sample, which gives the reason to introduce the mean number of data $\bar{N}$ in Table 1.

The results shown in Table 1 allow us to state that all methods reproduce the true b-value ($b = 1$) reasonably well with the exception of the simple Aki formula (2) for which the estimated b-value is significantly different from the true one. Since the corresponding p is less than $\alpha = 0.05$, we confirm statistically the well-known finding that the Aki estimator is not acceptable for binned data (e.g. Utsu 1966; Bender 1983). This result contradicts Marzocchi *et al.* (2020) who asserted instead that the bias of the Aki (1965) formula 'may be considered negligible in many applications when the binning is $\Delta M = 0.1$ or smaller'.

In Table 2, we report the mean $\bar{b}_K$ and the standard deviation S_K of the distribution of the K b-values computed with the aid of the eq. (7). For each data set we have also computed the standard deviations estimated by using Aki's and Shi-Bolt's formulas and also by eqs (18), (19) and (21) introduced in this paper. Their averages are also shown in Table 2. Note that the upper standard deviation $\bar{\sigma}_{2,K}$ is systematically larger than the lower one, $\bar{\sigma}_{1,K}$, which reflects the known feature that the b-values distribution is not symmetric around its midpoint. It is remarkable that for all cases, the 1σ intervals estimated via Aki and Shi-Bolt formulas as well as the ones based on eq. (21) of this paper correspond very well to the standard deviation S_K computed from the simulated data sets.

Table 3 reports results of simulation experiments where the magnitude differences are computed through the formulas (22) and (23). The b-values are computed for trimmed magnitude differences by the eq. (17) and the corresponding standard deviations are computed by means of eqs (18), (19) and (21), as specified earlier. As for the differences, we estimate b-values in three different ways, that is (i) by considering their absolute values, (ii) by considering only the subset of positive differences and (iii) by using only negative differences. These cases are, respectively, denoted in the second column of Table 3 as ABS, POS and NEG. The last column of Table 3 reports the ratio between the obtained standard deviations $\bar{\sigma}_K$ and the standard deviation S_K of the obtained b-value distribution. The analysis of the results leads us to some important observations. First, it is relevant to remark that when using absolute differences computed through the eq. (23), the ratio $\bar{\sigma}_K/S_K$ is very close to 1. On the contrary, when using non-independent differences computed via the eq. (22), the ratio is definitely less than 1 by about 22–23 per cent. This can be interpreted as the consequence of some sort of data correlation that reduces the number of 'effective' independent data, say N_e, in the difference data set, so that the calculated dispersion is less than the experimental one. Considering that the standard deviation scales inversely with the square root of the number of data, then the observed percentage discrepancy between $\bar{\sigma}_K$ and S_K corresponds to a decrease of about 40 per cent in the number of effective data, that is $N_e \sim 0.67N$. Then, we conclude that, when using absolute values, it is preferable to compute differences by means of the eq. (23). This will be our choice in all of the following computations shown in the paper and in the Supporting Information. Table 3 shows also that when using either only positive or only negative magnitude differences, the resulting ratio $\bar{\sigma}_K/S_K$ remains always very close to 1. This proves that the positive subsets and the negative subsets do not suffer of any significant level of correlation and that both formulas (22) and (23) can be used. Since using the latter implies halving the available data, and thus increasing the 1σ intervals, it follows that the eq. (22) is to be preferred.

Tables 3 and 4 allows us to make a further consideration. If we look at columns showing the resulting $\bar{b}_K$ and S_K, we can observe that all results are very good and that there are no significant discrepancies between values obtained by using various magnitude difference types: absolute, positive (or non-negative), negative (or non-positive) magnitude differences. Recalling that our finding regards data sets that are complete, we will address this issue later in the paper when incomplete magnitude sequences will be dealt with.

Even if, for most papers in the literature, the binning size is fixed to 0.1 as in Tables 1–3, larger bins can be assumed when the magnitude resolution is coarser, as it may occur for magnitudes derived from maximum macroseismic intensities. In Table 5 (which coincides with Table S14) we show the results for a binning size 2δ=0.5. We can note that in this case the eq. (4), that is the Aki (1965) estimator as corrected by Utsu (1966), significantly underestimates the simulated b-value, and that the corresponding performance index p does not pass the threshold. Therefore, it cannot be accepted as a valid estimator. This underestimation is observed even by varying the number of data N (100, 1000, 10 000) and the theoretical b-value (0.7, 1.0, 1.5) (see Tables S10–S18). This confirms that

Table 1. Estimates from complete simulated sets with $N = 1000$, $2\delta = 0.1$ and $b = 1$.

Estimator	Eq.	$\bar{b}_K$	S_K	$\bar{N}$	p
Aki (1965)	(2)	1.125 907	0.039 867	1000	0.000 587
Aki (1965), Utsu (1966)	(4)	0.996 582	0.031 225	1000	0.913 052
Bender (1983)	(6)	0.994 843	0.031 965	1000	0.869 062
This paper, magnitudes	(7)	1.001 003	0.031 644	1000	0.974 374
This paper, absolute differences by eq. (22)	(15)	1.001 331	0.040 499	999	0.974 819
This paper, absolute differences by eq. (23)	(15)	1.001 854	0.044 692	500	0.966 667
This paper, trimmed absolute differences by eq. (22)	(17)	1.001 663	0.043 709	885	0.970 893
This paper, trimmed absolute differences by eq. (23)	(17)	1.002 250	0.048 326	443	0.963 718

$\bar{b}_K$ and S_K are the average and the standard deviation of b-values computed for the $K = 10\,000$ simulated data sets, $\bar{N}$ is the average number of shocks in simulated data sets, p is the performance index computed by the eq. (24). The column 'Eq.' shows the equations used to estimate the b-values. Trimming with $\Delta M'_c = 0.1$.

Table 2. Estimated b-values and standard deviations for complete simulated sets with $2\delta = 0.1$.

b-Value	N	$\bar{b}_K$ Eq. (7)	S_K	$\bar{\sigma}_{b,K}$ Eq. (3)	$\bar{\sigma}_{b,K}$ Eq. (5)	$\bar{\sigma}_{1,K}$ Eq. (18)	$\bar{\sigma}_{2,K}$ Eq. (19)	$\bar{\sigma}_K$ Eq. (21)
0.7	1000	0.700 721	0.022 122	0.022 159	0.022 087	0.021 501	0.022 910	0.022 205
1	1000	1.001 003	0.031 644	0.031 654	0.031 516	0.030 746	0.032 768	0.031 757
1.5	1000	1.501 569	0.047 579	0.047 484	0.047 148	0.046 238	0.049 304	0.047 771

$\bar{b}_K$ and S_K are the average and the standard deviation of the distribution of the b-values estimated through the eq. (7) for the $K = 10\,000$ simulated data sets, each with N magnitudes; the standard deviations $\bar{\sigma}_K$ are the average of the sigma computed through the Aki formula (3), the Shi-Bolt formula (5) and the eqs (18), (19) and (21) proposed in this paper.

Table 3. Estimates of b-values and standard deviations for complete simulated sets with $2\delta = 0.1$ for trimmed absolute differences.

b-value	Diff. type	Eq.	$\bar{N}$	$\bar{b}_K$ Eq. (17)	S_K	$\bar{\sigma}_{1,K}$ Eq. (18)	$\bar{\sigma}_{2,K}$ Eq. (19)	$\bar{\sigma}_K$ Eq. (21)	$\frac{\bar{\sigma}_k}{S_K}$
0.7	ABS	(22)	919	0.701 100	0.029 949	0.022 415	0.023 950	0.023 182	0.774
		(23)	460	0.701 498	0.033 097	0.031 288	0.034 360	0.032 824	0.992
	POS	(22)	459	0.701 081	0.032 811	0.031 283	0.034 356	0.032 819	1.000
		(23)	230	0.703 577	0.046 774	0.043 609	0.049 800	0.046 704	0.999
	NEG	(22)	459	0.702 548	0.032 737	0.032 664	0.032 966	0.032 815	1.002
		(23)	226	0.702 548	0.047 092	0.043 538	0.049 718	0.046 628	0.990
1	ABS	(22)	885	1.001 663	0.043 709	0.032 652	0.034 940	0.033 796	0.773
		(23)	443	1.002 250	0.048 326	0.045 564	0.050 144	0.047 854	0.990
	POS	(22)	442	1.001 574	0.048 015	0.045 555	0.050 136	0.047 845	0.996
		(23)	221	1.005 455	0.068 193	0.063 497	0.072 730	0.068 113	0.999
	NEG	(22)	442	1.003 723	0.047 707	0.047 428	0.048 243	0.047 836	1.003
		(23)	219	1.003 723	0.068 949	0.063 383	0.072 600	0.067 992	0.986
1.5	ABS	(22)	828	1.502 808	0.068 133	0.050 696	0.054 403	0.052 549	0.771
		(23)	415	1.503 491	0.074 984	0.070 703	0.078 125	0.074 414	0.992
	POS	(22)	414	1.502 903	0.074 569	0.070 698	0.078 123	0.074 410	0.998
		(23)	207	1.508 829	0.106 490	0.098 496	0.113 473	0.105 984	0.995
	NEG	(22)	414	1.505 754	0.074 317	0.073 539	0.075 300	0.074 419	1.001
		(23)	202	1.505 754	0.106 555	0.098 296	0.113 242	0.105 769	0.993

$\bar{b}_K$ and S_K are the average and the standard deviation of b-values for the $K = 10\,000$ simulated data sets, $\bar{N}$ is the average number of shocks. The b-values are computed through the eq. (17). The standard deviations $\bar{\sigma}_{1,K}$, $\bar{\sigma}_{2,K}$ and $\bar{\sigma}_K$ are the average of the values computed through the eqs (18), (19) and (21) proposed in this paper. The ratio $\bar{\sigma}_K/S_K$ is used to judge the correlation of the data. The column 'Eq.' shows the equations used to compute the magnitude differences. Trimming with $\Delta M'_c = 0.1$.

Utsu (1966) correction to the Aki (1965) formula is approximate and is not adequate for large bin sizes. Table 5 also confirms that the Bender method, though adequate, has a performance slightly smaller than the eq. (7) for magnitudes and the formulas (15) and (17) for magnitude differences, but this might be due to a not fully accurate numerical minimization. It is relevant to note that the corresponding standard deviation S_K increases sensibly when passing from magnitudes to magnitude differences, since the number of data decreases, which means that the latter estimators can be considered less efficient.

RESULTS FOR SIMULATED INCOMPLETE DATA SETS

Experimental magnitude data sets are always affected by some degree of incompleteness. Therefore, evaluating how the estimators perform on incomplete data sets is of paramount importance. Van der Elst (2021), on analysing incomplete binned magnitude sequences, concluded that magnitude difference estimators are more robust if only the positive differences are used, since using also the negative differences might produce biased results. One of the purposes of this paper is to check this conclusion. Using a method

Table 4. Estimates of b-values and associated standard deviations for complete simulated sets with $2\delta = 0.1$ for untrimmed absolute differences and for untrimmed signed (non-negative, non-positive) differences.

b-value	Diff. type	Eq.	$\bar{N}$	$\bar{b}_K$	S_K	$\bar{\sigma}_{1,K}$	$\bar{\sigma}_{2,K}$	$\bar{\sigma}_K$	$\frac{\bar{\sigma}_k}{S_K}$
0.7	ABS	(22)	999	0.700 933	0.028 402	0.021 454	0.022 847	0.022 151	0.780
		(23)	500	0.701 315	0.031 301	0.029 964	0.032 751	0.031 358	1.002
	NONNEG	(22)	539	0.700 781	0.029 881	0.028 948	0.031 562	0.030 255	1.013
		(23)	270	0.702 983	0.042 582	0.040 385	0.045 644	0.043 015	1.010
	NONPOS	(22)	548	0.700 839	0.029 995	0.028 949	0.031 563	0.030 256	1.009
		(23)	270	0.702 077	0.043 129	0.040 327	0.045 577	0.042 952	0.996
1	ABS	(22)	999	1.001 331	0.040 499	0.030 588	0.032 560	0.031 574	0.780
		(23)	500	1.001 854	0.044 692	0.042 723	0.046 669	0.044 696	1.000
	NONNEG	(22)	552	1.001 041	0.042 217	0.040 779	0.044 414	0.042 597	1.009
		(23)	279	1.004 204	0.060 103	0.056 898	0.064 210	0.060 554	1.008
	NONPOS	(22)	555	1.001 211	0.042 236	0.040 782	0.044 417	0.042 600	1.009
		(23)	279	1.002 804	0.060 712	0.056 816	0.064 117	0.060 467	0.996
1.5	ABS	(22)	999	1.501 768	0.060 370	0.045 658	0.048 558	0.047 108	0.780
		(23)	500	1.502 550	0.066 694	0.063 783	0.069 583	0.066 683	1.000
	NONNEG	(22)	591	1.501 277	0.061 952	0.059 873	0.065 119	0.062 496	1.009
		(23)	293	1.505 961	0.088 278	0.083 555	0.094 107	0.088 831	1.006
	NONPOS	(22)	594	1.501 289	0.061 994	0.059 873	0.065 120	0.062 496	1.008
		(23)	293	1.503 813	0.089 153	0.083 435	0.093 971	0.088 703	0.995

$\bar{b}_K$ and S_K are the average and the standard deviation of b-values computed for the $K = 10\,000$ simulated data sets, $\bar{N}$ is the average number of shocks. For the absolute differences, the b-values are computed through the eq. (15) and for the other differences through the eq. (17). The standard deviations $\bar{\sigma}_{1,K}$, $\bar{\sigma}_{2,K}$ and $\bar{\sigma}_K$ are the average of the values computed for the absolute differences through the eqs (H17a)–(H17c) and for the other differences via the eqs (18), (19) and (21) proposed in this paper. The ratio $\bar{\sigma}_K/S_K$ is used to judge the correlation of the data. The column 'Eq.' shows the equations used to compute the magnitude differences. Trimming with $\Delta M'_c = 0.1$.

Table 5. Estimates from complete simulated sets with $N = 1000$, $2\delta = 0.5$ and $b = 1$.

Estimator	Eq.	$\bar{b}_K$	S_K	$\bar{N}$	p
Aki (1965)	(2)	1.884 281	0.106 413	1000	0.000 000
Aki (1965), Utsu (1966)	(4)	0.903 155	0.024 395	1000	0.000 000
Bender (1983)	(6)	0.996 548	0.033 689	1000	0.916 258
This paper, magnitudes	(7)	1.001 296	0.033 480	1000	0.966 446
This paper, absolute differences	(15)	1.001 698	0.041 874	500	0.960 875
This paper, trimmed absolute differences	(17)	1.004 231	0.069 217	240	0.954 589

$\bar{b}_K$ and S_K are the average and the standard deviation of b-values computed for the $K = 10\,000$ simulated data sets, $\bar{N}$ is the average number of shocks and p is the performance index computed by the eq. (24). The column 'Eq.' shows the equations used to estimate the b-values. Trimming with $\Delta M'_c = 0.1$.

that is detailed in Appendix I we build $K = 10\,000$ incomplete data sets by starting from 11 000 exponentially distributed magnitudes with $M_{min} = 0.4$ and decay parameter $b = 1$, and by applying a thinning model where the incompleteness function is assumed as a cumulative Gaussian with mean $\mu = 1$ and standard deviation $\lambda = 0.2$. After the reduction, the remaining magnitudes are less than 10 per cent. More precisely the mean number of earthquakes per data set is $\bar{N}$=1093. The histogram of one of the simulated data sets is portrayed in Fig. 1 and is manifestly incomplete. Seismologists usually refrain from estimating the b-value using all data of such incomplete data sets, and in common practice one discards the smaller magnitude classes that are visibly incomplete. Nonetheless, it is interesting to see the behaviour of the estimators in the extreme case where M_c is taken to be equal to M_{min}, although it is clear from Fig. 1 that such kind of samples are far from fitting an exponential distribution. It is worth stressing that M_c enters in all formulas based on magnitudes, namely eqs (2), (4), (6) and (7), but not in the ones based on magnitude differences, that is eqs (15) and (17). The results are shown in Table 6.

As expected, all methods provide estimates quite far from the true value, and correspondingly the index p lies below the significance threshold $\alpha = 0.05$. But it is relevant to observe that on using differences one obtains better results.

In Table 7, we show the results considering exactly the same data sets as before, but assuming that M_c is equal to 1.1, corresponding to the maximum curvature magnitude (say M_{mxc}) of the frequency magnitude distribution, as suggested by Wiemer & Wyss (2000). It is clear that increasing the value of the completeness magnitude improves the results for all methods. One common feature is that all estimates lie below the true value, with one exception. Indeed, the Aki (1965) estimator is seemingly better than the estimator corrected by Utsu (1966) and also than the exact formula using magnitudes. However, this is an artefact since the typical underestimation due to incompleteness is overcompensated for the Aki method by the effect of magnitude binning. Most importantly, observe that methods based on magnitude differences and either on the positive differences or on the negative differences produce the best results in terms of closeness to the true b and in terms of the

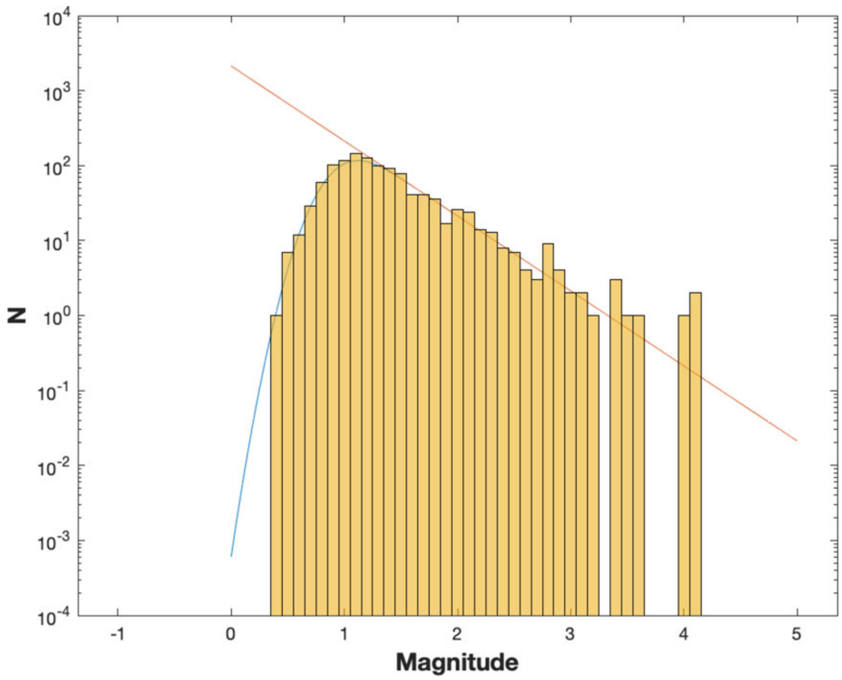

Figure 1. Histogram of one of the K incomplete simulated sets obtained by thinning with parameters: $\mu = 1$, $\lambda = 0.2$; $M_{min} = 0.4$, $2\delta = 0.1$, $b = 1$.

Table 6. Estimates for incomplete simulated data sets with $\mu = 1$, $\lambda = 0.2$, $2\delta = 0.1$, $b = 1$, $M_c = 0.4$.

Estimator	Eq.	$\bar{b}_K$	S_K	$\bar{N}$	p
Aki (1965)	(2)	0.460 944	0.006 947	1093	0.000 000
Aki (1965), Utsu (1966)	(4)	0.437 711	0.006 264	1093	0.000 000
Bender (1983)	(6)	0.387 450	0.024 053	1093	0.000 000
This paper, magnitudes	(7)	0.438 082	0.006 280	1093	0.000 000
This paper, absolute differences—eq. (23)	(15)	0.862 855	0.032 991	546	0.000 000
This paper, trimmed absolute differences—eq. (23)	(17)	0.890 224	0.036 483	506	0.005 059
This paper, trimmed positive differences—eq. (22)	(17)	0.890 039	0.036 662	506	0.005 283
This paper, trimmed negative differences—eq. (22)	(17)	0.891 447	0.036 558	506	0.002 304

$\bar{b}_K$ and S_K are the average and the standard deviation of b-values computed for the $K = 10\,000$ simulated data sets, $\bar{N}$ is the average number of shocks and p is the performance index computed by means of the eq. (24). The column 'Eq.' shows the equations used to estimate the b-values. Trimming with $\Delta M'_c = 0.1$.

Table 7. Estimates for incomplete simulated data sets with $\mu = 1$, $\lambda = 0.2$, $2\delta = 0.1$, $b = 1$, $M_c = 1.1$.

Estimator	**Eq.**	$\bar{b}_K$	S_K	$\bar{N}$	p
Aki (1965)	(2)	1.026 523	0.037 518	786	0.478 636
Aki (1965), Utsu (1966)	(4)	0.917 912	0.029 991	786	0.008 349
Bender (1983)	(6)	0.911 953	0.031 097	786	0.006 099
This paper, magnitudes	(7)	0.921 364	0.030 332	786	0.012 220
This paper, absolute differences—eq. (23)	(15)	0.973 845	0.047 540	393	0.582 759
This paper, trimmed absolute differences—eq. (23)	(17)	0.986 348	0.051 871	353	0.788 916
This paper, trimmed positive differences—eq. (22)	(17)	0.986 018	0.051 915	353	0.788 513
This paper, trimmed negative differences—eq. (22)	(17)	0.988 299	0.051 836	353	0.815 571

$\bar{b}_K$ and S_K are the average and the standard deviation of b-values computed for the $K = 10\,000$ simulated data sets, $\bar{N}$ is the average number of shocks in simulated data sets and p is the performance index computed by means of the eq. (24). The column 'Eq.' shows the equations used to estimate the b-values. Trimming with $\Delta M'_c = 0.1$.

performance index p. Such evidence seems to confirm the strategy suggested by van der Elst (2021) that using magnitude differences is preferable than using magnitudes. However, his claim that positive differences lead to better estimates than using negative differences is not evident from this Table.

In Table 8, we show the results for the same parameters when the minimum magnitude M_c is increased up to 1.3 corresponding to the magnitude of maximum curvature plus 0.2, that is the way M_c is commonly set in the literature (Wiemer & Wyss 2000; Mignan & Woessner 2012).

Table 8. Incomplete simulated sets with $\mu = 1$, $\lambda = 0.2$, $2\delta = 0.1$, $b = 1$, $M_c = 1.3$.

Estimator	Eq.	$\bar{b}_K$	S_K	$\bar{N}$	p
Aki (1965)	(2)	1.107 743	0.052 196	541	0.027 810
Aki (1965), Utsu (1966)	(4)	0.982 229	0.041 025	541	0.670 191
Bender (1983)	(6)	0.976 523	0.042 257	541	0.579 654
This paper, magnitudes	(7)	0.986 471	0.041 560	541	0.744 560
This paper, absolute differences—Eq. (23)	(15)	0.998 481	0.060 113	270	0.979 558
This paper, trimmed absolute differences—Eq. (23)	(17)	1.001 747	0.064 811	240	0.975 911
This paper, trimmed positive differences—Eq. (22)	(17)	1.001 059	0.064 781	240	0.986 776
This paper, trimmed negative differences—Eq. (22)	(17)	1.006 768	0.064 375	240	0.917 414

$\bar{b}_K$ and S_K are the average and the standard deviation of b-values computed for the $K = 10\,000$ simulated data sets, $\bar{N}$ is the average number of shocks in simulated data sets and p is the performance index computed by means of the eq. (24). The column 'Eq.' shows the equations used to estimate the b-values. Trimming with $\Delta M'_c = 0.1$.

Notice that increasing the value of the completeness magnitude M_c has the effect of reducing the size of the samples and, consequently, of increasing the simulated standard deviation S_K. Further, it lowers the degree of incompleteness, and therefore leads to better estimates. As for the rest, most considerations made for Table 7 apply also to Table 8. Most precisely, we can see that now the Aki (1965) formula (2) clearly overestimates the b-value and is thus inadequate, whereas all other estimators using magnitudes as well as the estimators based on magnitude differences give acceptable results. In this case, the estimator using the differences (the last 4 rows), either trimmed or not, irrespective of the difference type, give equivalent results.

Fig. 2 displays the histogram of the absolute differences of the same data set plotted in Fig. 1. It includes also the null differences, which however are neglected in cases that the difference are trimmed. It is quite evident from this graph, that the histogram exhibits an exponential behaviour much more than the corresponding graph of Fig. 1, and it is much more so if the first column on the left-hand side is discarded. This induces the expectation that methods based on differences, rather than on magnitudes, can provide better estimates.

The resilience of trimmed differences estimators to magnitude incompleteness can be further proven through the following experiment. We take exactly the same data sets analysed before and with estimates given in Table 6: they have an average size $\bar{N} = 1093$ and have $M_c = 0.4$. They are strongly incomplete (see the exemplary sample in Fig. 1). Let us estimate the b-value through the eq. (17) and change the value of $\Delta M'_c$. The last three rows of Table 6 show the results for trimmed differences estimators (absolute, positive and negative) when we apply the basic trimming, which means that we set $\Delta M'_c = 2\delta$, so discarding the differences equal to 0.

In Table 9, we show the results obtained by increasing $\Delta M'_c$ up to $\Delta M'_c = 10\delta$ (i.e. 5 times the bin size). In the Table we also include the outputs for $\Delta M'_c = 2\delta$ for the sake of comparison. Very remarkably, it can be seen that the deviations of the estimated b-values from the theoretical one progressively decrease when incrementing the amount of trimming, and at the end, they become almost negligible for all of the three estimators. It is also worth stressing that increasing the trimming threshold $\Delta M'_c$ obviously reduces the number of available data and leads to larger standard deviations, but less dramatically than when a similar increment is applied to the magnitude threshold M_c. Even in this experiment, for a given trimming one cannot find any significant discrepancy among the methods based on magnitude differences, in particular the van der Elst's claim (2021) that positive differences should be preferred to absolute or negative differences is not supported by these simulations.

RESULTS FOR SIMULATED DATA SETS WITH INCOMPLETENESS CHANGING WITH TIME

It is known that incompleteness affects seismic catalogues systematically after a strong main shock, owing to the superposition of the waveforms in the recorded seismograms that prevents the correct location and size determination of many small shocks in the hours or days after the main shocks. It is also known that under these circumstances, incompleteness is time-dependent since it tends to decrease in time, which means that the completeness magnitude can be modelled as a decreasing function of time.

We generate the synthetic data sets following the procedure described in Appendix I and that can be found in the literature (see e.g. Lolli *et al.* 2009), We set the magnitude of the main shock to $m = 5.6$ in the eq. (I4), the Omori-Utsu law parameters to $p_O = 1$ and $c_O = 0.01$ in the eq. (I5) and the sequence ending time to $T_E = 5$ days in eqs (I8)–(I13). We produce $K = 10000$ aftershock sequences of $N = 40000$ earthquakes that are made incomplete through the thinning mechanism, that is a probabilistic and time-dependent process. The magnitude mean of the thinning law $\mu(t)$ is given in the eq. (I4) and λ is set to 0.2. Further, we compute the magnitude M_c using the criterion adopted for the analysis shown in Table 8, that is M_c is equal to the magnitude M_{mxc} corresponding to the maximum curvature of the sample frequency magnitude curve plus twice the bin size. Then, we eliminate all magnitudes smaller than M_c. The mean number of surviving data is $\bar{N} = 1041$. In Table 10, we present the results of the estimation when M_c is equal to 1.3. Our finding is that none of the estimators is performing very well. However, we observe that estimators based on magnitudes tend to severely underestimate the theoretical b, and the corresponding values of the performance index p is much smaller than the 5 per cent. Conversely, the estimators based on differences give results that are superior and can be considered satisfactory. A further remark is that among methods using differences, one cannot see any relevant superiority for methods based on positive differences with respect to the others.

To examine the performance of the estimators on aftershock sequences we have varied the number of data and the theoretical b-value (see Tables from S19 to S27). Note that Table 10 coincides with Table S23. We have found confirmation that methods based on differences provide better results than methods based on magnitudes

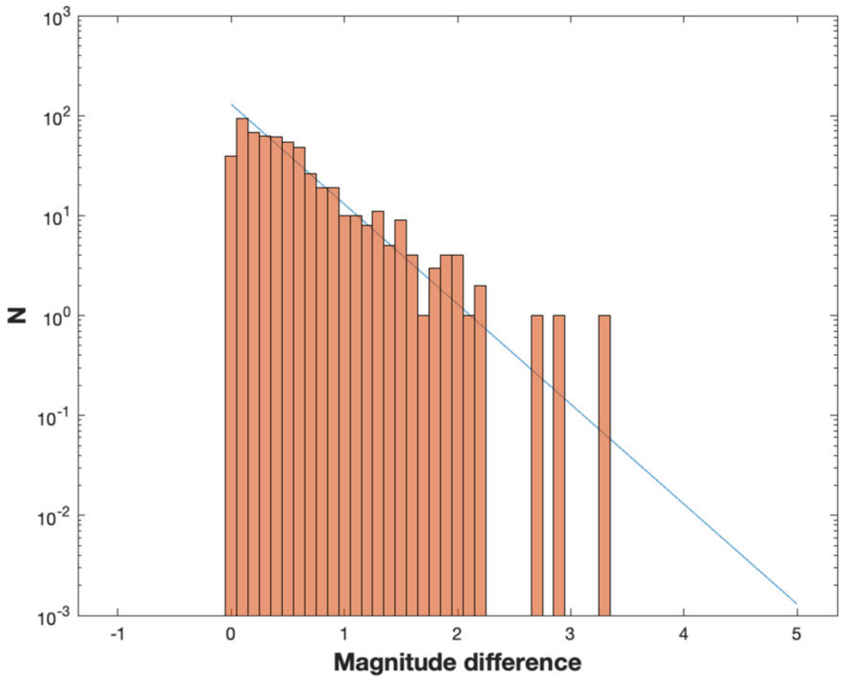

Figure 2. Histogram of the absolute differences computed for the data set shown in Fig. 1.

Table 9. Estimates for incomplete simulated sets generated with parameters $\mu = 1$, $\lambda = 0.2$, $2\delta = 0.1$, $b = 1$, $M_c = 0.4$.

Estimator	$\Delta M'_c$	$\bar{b}_K$	S_K	$\bar{N}$	p
This paper, trimmed absolute differences	2δ	0.890 224	0.036 483	506	0.005 059
This paper, trimmed positive differences	2δ	0.890 039	0.036 662	506	0.005 283
This paper, trimmed negative differences	2δ	0.891 447	0.036 558	506	0.002 304
This paper, trimmed absolute differences	4δ	0.927 973	0.042 749	428	0.101 695
This paper, trimmed positive differences	4δ	0.927 803	0.043 113	428	0.106 245
This paper, trimmed negative differences	4δ	0.929 565	0.042 655	428	0.110 231
This paper, trimmed absolute differences	6δ	0.957 032	0.049 715	355	0.394 683
This paper, trimmed positive differences	6δ	0.956 623	0.049 740	355	0.385 656
This paper, trimmed negative differences	6δ	0.959 462	0.049 503	355	0.404 973
This paper, trimmed absolute differences	8δ	0.977 009	0.057 063	290	0.683 715
This paper, trimmed positive differences	8δ	0.976 942	0.056 726	290	0.685 273
This paper, trimmed negative differences	8δ	0.980 056	0.056 776	290	0.709 012
This paper, trimmed absolute differences	10δ	0.990 306	0.064 465	235	0.879 491
This paper, trimmed positive differences	10δ	0.989 968	0.064 246	234	0.877 323
This paper, trimmed negative differences	10δ	0.994 486	0.064 548	234	0.925 205

$\Delta M'_c$ is the trimming difference, $\bar{b}_K$ and S_K are the average and the standard deviation of b-values computed for the K = 10 000 simulated data sets, $\bar{N}$ is the average number of shocks in simulated data sets, and p is the performance index computed by the eq. (24). Absolute differences are computed through the eq. (23), all other differences through the eq. (22). The b-values are computed through the eq. (17).

Table 10. Aftershock sequence with time-dependent incompleteness; parameters: $\lambda = 0.2$, $m = 5.6$, $p_O = 1$, $c_O = 0.01$, $T_E = 5$ days, N = 40 000 (before thinning), $2\delta = 0.1$, $b = 1$, $M_c = 1.3$.

Estimator	Eq.	$\bar{b}_K$	S_K	$\bar{N}$	p
Aki (1965)	(2)	0.835 400	0.025 265	1041	0.000 000
Aki (1965), Utsu (1966)	(4)	0.762 046	0.021 019	1041	0.000 000
Bender (1983)	(6)	0.752 022	0.022 715	1041	0.000 000
This paper, magnitudes	(7)	0.764 015	0.021 183	1041	0.000 000
This paper, absolute differences—eq. (23)	(15)	0.952 553	0.040 146	520	0.244 594
This paper, trimmed absolute differences—eq. (23)	(17)	0.965 537	0.043 553	469	0.430 854
This paper, trimmed positive differences—eq. (22)	(17)	0.966 745	0.043 654	468	0.442 129
This paper, trimmed negative differences—eq. (22)	(17)	0.967 363	0.043 399	470	0.446 190

$\bar{b}_K$ and S_K are the average and the standard deviation of b-values computed for the K = 10 000 simulated data sets, $\bar{N}$ is the average number of shocks in simulated data sets and p is the performance index computed by means of the eq. (24). The column 'Eq.' shows the equations used to estimate the b-values. Trimming with $\Delta M'_c = 0.1$.

and that methods based on trimmed differences tend to be slightly better than the others and practically equivalent to one another.

Interestingly, Tables S24 and S27 show that all estimates exhibit a performance index p below the significance level. Considering also Table S21, and comparing results with the ones of the other Tables, one can note that estimates made with the largest sample size (i.e. $N >\sim 4000$ in our experiments) are worse than estimates performed with samples of smaller size, which is exactly the opposite of what one expects, and is an apparent paradox, that does not manifest with complete data sets. Indeed, this effect disappears when increasing trimming: that is we found that using a trimming larger than $\Delta M'_c = 0.1$ produces better and acceptable results for larger samples, according to the expectations.

RESULTS FOR A REAL AFTERSHOCK SEQUENCE

All data sets analysed so far, either complete or incomplete, were produced through numerical simulations, and though the methods to produce random variables following exponential distributions are quite standard and the methods to produce incompleteness are commonly applied in the literature, nonetheless they might not reproduce well real sequences of recorded magnitudes. Therefore, we have considered to complement our analysis by taking into account a real aftershock sequence to test our finding. We have decided to use a single sequence as an example. We have selected a well recorded subset of the aftershocks of the Norcia M_w 6.6 earthquake, that is the last and the largest of the 3 main shocks characterizing the 2016 Amatrice-Visso-Norcia seismic sequence (central Italy). We have made use of the Horus catalogue, including the instrumental Italian earthquakes from 1960 (Lolli *et al.* 2020) that is maintained and updated as a permanent activity by the INGV Bologna section. We have considered a set consisting of the main shock and the aftershocks occurred within a radius of 50 km in the first 17 hr from the main shock up to a total of $N = 1000$ events. Figs 3 and 4 show the histograms of the magnitude data and of their absolute differences with bin amplitude $2\delta = 0.1$. The former figure exhibits that magnitude data are characterized by a remarkable level of incompleteness below magnitude $M = 3$, while the latter one shows that magnitude differences are somewhat closer to an exponential distribution.

Table 11 summarizes the estimates we made of the b-value. We focus our attention only on methods using magnitude differences. Obviously, there is no true b-value to compare our estimates to, and therefore, we are not able to evaluate in absolute the performance of individual methods. However, we are able to make some relevant observations, also in virtue of what we learned from our previous simulations. Indeed, we saw that methods based on absolute differences computed through the eq. (23) are unreliable since lead to standard deviations that are strongly underestimated (1st and 7th row). On the other hand, methods based on signed differences computed through the eq. (22) (see 4th, 10th and 12th row) reduce too much the sample size, resulting in unnecessarily larger standard deviations. Thus, we restrict our attention to difference methods using almost half of the data, that is about 500 in the present case. Further, simulations have shown that methods based on trimmed differences give better results than the others. Taking all this into account, we make the choice that the preferable estimates of the sequence b-value are: $b = 1.04 \pm 0.05$ (8th row), 1.03 ± 0.05 (9th row), 1.01 ± 0.05 (11th row), with truncation made to the second decimal figure. The obtained values are equivalent to one another and show that there is no significant change in using absolute, positive or negative trimmed differences.

DISCUSSION OF THE RESULTS

Computing the b-value of a magnitude catalogue or a magnitude sequence is a classical activity in standard seismicity analyses and is often used in advanced studies. For complete data sets, when the magnitudes are taken as continuous exponentially distributed variables, the problem of estimating b and the related confidence intervals was given a definite solution, respectively, by Aki (1965) and by Shi & Bolt (1982) who made recourse to the chi-square distribution. When magnitudes are grouped in bins of equal size, the problem was also given a final solution by Guttorp & Hopkins (1986) for estimating b and by Tinti & Mulargia (1986, 1987) to compute the corresponding confidence intervals. Catalogue incompleteness affects remarkably all the estimations, but this problem was long overlooked, since there was the belief that it regards only the smaller magnitudes range and that it is easily possible to find a threshold magnitude M_c above which the data set is complete and is sufficiently large to allow accurate estimates. More sophisticated views emerged progressively as well as criteria to find M_c (see Mignan & Woessner 2012), but it was only recently that the problem was tackled from a very different point of view. It was van der Elst's (2021) who, analysing aftershock sequences where the completeness is known to change quite rapidly with time, proposed to consider data sets of differences of magnitudes in place of magnitudes and to base on them all the statistical inferences. This was a remarkable turning point in the discipline. He also suggested that using data sets of positive differences was the only correct way because the alternative choice of using absolute differences of magnitudes or negative differences leads to inaccurate estimates.

In this paper, we have explored the potential of the new approach and we have considered synthetic data sets since this is the best way to evaluate the performance of inferential estimators. Indeed, one can generate easily a very high number (10 000 in this paper) of pseudorandom samples of magnitudes of any reasonable size. In addition, we have also applied the new approach to a real aftershock sequence.

First of all, we have addressed the problem of how to compute the magnitude differences in such a way that the resulting sample can be considered formed by independent random variables, that is an issue overlooked by previous studies. Making consecutive differences (see eq. 22) does not alter significantly the set size (it passes from N to $N - 1$), but it introduces an undesirable correlation in the data. The alternative way to compute differences (see eq. 23), by definition produces independent random variables, but has the cost of halving the size of the sample. It was shown (see Tables 3 and 4) that if one considers the set of the absolute magnitude differences obtained by the eq. (22), the effect of such correlation is that the standard deviation of the b-value estimates, S_K, is larger than the theoretical one, $\bar{\sigma}_k$, by an amount of about 22 per cent. This implies that the variance increases by a factor of 1.49 or, correspondingly, that the equivalent number of independent data is less than $N - 1$. Therefore, when treating absolute differences, in this paper we have opted to work only with data sets formed by independent variables obtained through eq. (23). We have further shown that if one uses signed differences, practically the effect of the correlation implicitly introduced by eq. (23) results to vanish, since the values of S_K and $\bar{\sigma}_k$ tend to overlap. Consequently, when using signed differences, the adoption of the eq. (23) is admissible and indeed even preferable

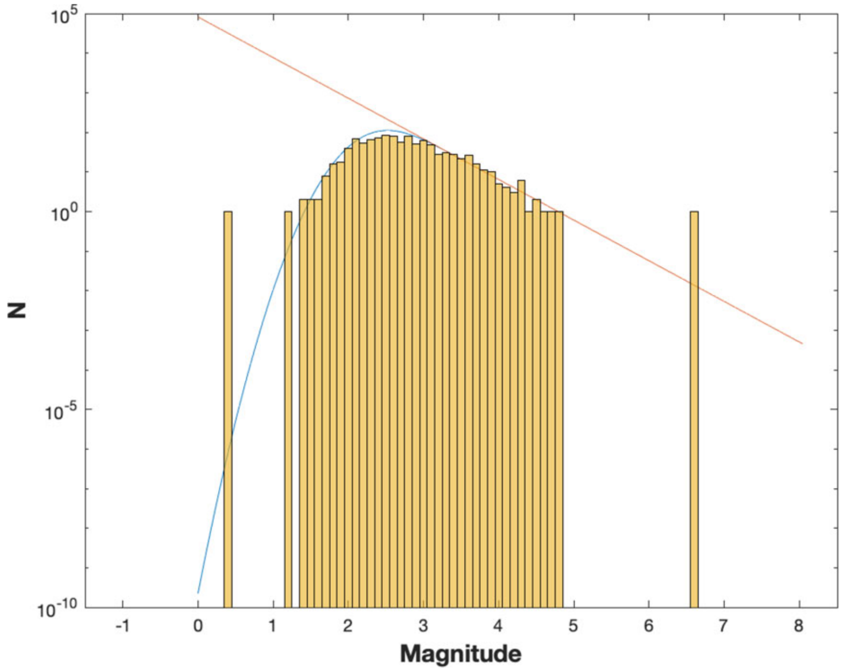

Figure 3. Histogram of magnitudes of the first 1000 shocks from the 30/10/2016 Norcia (Italy) earthquake ($M_w = 6.6$), including also the main shock. Data are grouped in bins with size $2\delta = 0.1$. Data from the Horus catalogue (Lolli *et al.* 2020).

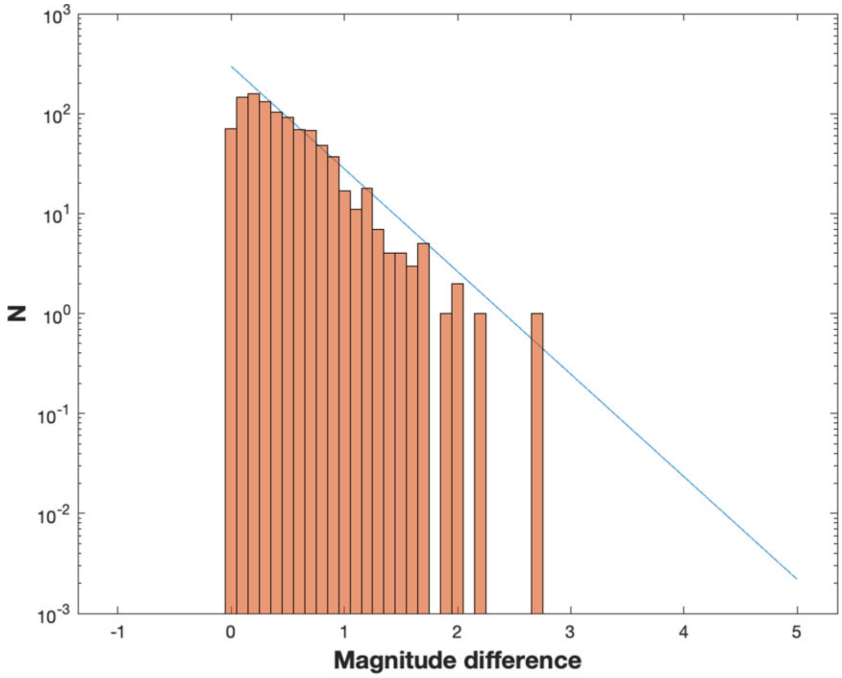

Figure 4. Histogram of magnitude differences of the real data set of the first 1000 shocks from the 30/10/2016 Norcia (Italy) earthquake ($M_w = 6.6$). Data from the Horus catalogue (Lolli *et al.* 2020).

since using eq. (22) would reduce the data to about one quarter of the original size. With the above choices, if magnitude data sets have size N, then the magnitude differences data sets, either absolute differences obtained through the eq. (23) or the signed differences obtained by means of the eq. (22), have about $N/2$ data.

To judge the goodness of the various estimators based on magnitudes and magnitude differences we have devised a new performance index denoted by p (eq. 24), that can be interpreted as a non-parametric variant of the two-side Student's t. Depending on the way it is defined, that is a sort of normalization, the index p, like the Student's t, tends to tolerate larger discrepancies for estimators with larger standard deviations, and does not prize the estimator efficiency. Bearing this in mind, we consider p a suitable index to evaluate estimators that are known a priori to operate on different data set sizes like the ones working on magnitudes and on magnitude differences.

Table 11. Estimates of *b*-value and related uncertainties for the 30/10/2016 Norcia (Italy) aftershock sequence. Data from the Horus catalogue (Lolli *et al.* 2020).

Estimator	Eq1	Eq2	N	b	σ_1	σ_2	σ
This paper, absolute differences	(15)	(22)	999	0.972 094	0.029 702	0.031 618	0.030 660
	(15)	(23)	500	0.995 501	0.042 455	0.046 377	0.044 416
This paper, non-negative differences	(17)	(22)	530	0.941 596	0.039 265	0.042 853	0.041 059
	(17)	(23)	277	0.945 544	0.053 681	0.060 587	0.057 134
This paper, non-positive differences	(17)	(22)	514	0.898 246	0.038 006	0.041 533	0.039 769
	(17)	(23)	459	0.932 026	0.056 058	0.063 757	0.059 908
This paper, trimmed absolute differences	(17)	(22)	922	1.016 543	0.032 478	0.034 706	0.033 592
	(17)	(23)	459	1.039 070	0.046 433	0.051 014	0.048 724
This paper, trimmed positive differences	(17)	(22)	460	1.026 253	0.045 810	0.050 324	0.048 067
	(17)	(23)	239	1.025 553	0.062 427	0.071 126	0.066 776
This paper, trimmed negative differences	(17)	(22)	462	1.007 057	0.044 857	0.049 266	0.047 061
	(17)	(23)	220	1.054 166	0.066 717	0.076 440	0.071 578

b is the estimated b-value, N is the number of shocks; σ_1 and σ_2 are the lower and upper sigma values and σ is their mean, that are computed through the eqs (H17a)–(H17c) for the two first rows, and through the eqs (18), (19) and (21) for the other rows. The columns denoted Eq1 and Eq2 show the equations used to estimate the b-values and to compute the magnitude differences, respectively. Trimming with $\Delta M'_c = 0.1$.

When treating complete binned data sets, our analysis confirms that the classical estimator of the eq. (7) works quite well and better than estimators for continuous magnitudes either in the original form (eq. 2) or in the one corrected to account for binning (see the eq. 4). It is relevant to point out that also the methods based on magnitude differences and given in the eqs (15) and (17) provide equivalently good results.

The most important finding we obtain is that magnitude samples showing a very severe incompleteness transform to samples with an almost exponential decay when magnitudes are replaced by magnitude differences (compare Figs 1 to 2, and also Figs 3 to 4). This seems to be a strong support for the strategy of using absolute differences or signed differences to evaluate b. Results reported in Tables 6–8 show that increasing the magnitude value M_c in the traditional estimator (eq. 7) from the very low value of patent incompleteness to the magnitude of maximum curvature of the frequency magnitude distribution and even further, improves its performance very much. Nonetheless, from all those Tables it emerges that better evaluations are attained by the estimator (15) that applies to absolute differences and by the estimator (eq. 17) that applies to trimmed differences, either absolute or one-sign. Table 9 explores the performance of the estimator (17) when the magnitude data set is very incomplete (M_c is assumed to be very low) and the trimming is made progressively more substantial. It is found that increasing the trimming also improves the results. In virtue of these outcomes, one could play with M_c and with the amount of trimming $\Delta M'_c$ to optimize the estimates. However, refining this strategy is not the scope of this paper.

With this clue in mind, we have considered aftershock sequences. Their main feature is not only that they are strongly incomplete data sets, but also that the completeness magnitude changes quickly during the process, and formally it changes from one earthquake to the next [according to the eq. (I4) in Appendix I]. Nonetheless, each sequence gives rise to a frequency magnitude histogram that can be examined as in the previous analysis, that is by establishing the maximum curvature magnitude M_{mxc} and by determining M_c through the formula given above. The results of Tables S19–S27 in the supplementary material confirm the superiority of the formulas based on magnitude differences, but they also add that the estimator (eq. 17) including trimming is generally better than formula (eq. 15).

The second most important result of our analysis is that the claim by van der Elst (2021) that in aftershock sequences one necessarily finds that positive magnitude differences are distributed exponentially much better than the negative ones and therefore the most successful estimator is the estimator (eq. 17) applied to the subsets of positive differences is not supported by enough evidence. As a matter of fact, we found that the estimates are very close to one another and that in some cases positive differences provide better values while in other cases the reverse is true. One could suspect that this finding is due to the difference between the methods used in producing simulated aftershock sequences (that is between the one adopted here and described in Appendix I and the one adopted by van der Elst (2021), that is an ETAS model obtained as a variant of the Hardebeck *et al.* (2008) code. However, we stress that this is the outcome not only of the analysis performed on the simulated aftershock sequences, but also of the estimates made for the real seismic sequence of the 30/10/2016 Norcia earthquake (see Table 11), which is a convincing evidence that this result does not depend on the way the aftershock sequences are simulated.

A final remark regards the 1σ intervals. Van der Elst's contribution brought into the seismological arena the two new estimators eqs (15) and (17), without providing however the related 1σ intervals. The formula (17) is practically the same as eq. (7) but applied to magnitude differences since the supporting distribution (discrete exponential) is the same. This means that the corresponding theoretical confidence intervals based on the geometric distribution are already known for any given number N of sample data (Tinti & Mulargia 1986, 1987). In this paper, we have provided explicit simple formulas (18), (19) and (21), to compute the 1σ intervals of the b-value when the sample size is large enough (N larger than about 30–40) that the distribution of the mean (either $\bar{M}$ or $\overline{|\Delta M|}$) approximates a Gaussian, which is normally the case in seismological practice. On the other hand, to the authors' knowledge, the confidence intervals of the distribution underlying the estimator eq. (15) have not yet been given a general theoretical solution in the seismological literature. In this context, and under the same assumption of Gaussian distribution of either $\bar{M}$ or $\overline{|\Delta M|}$, we propose the formulas (H17a and H17b) derived in Appendix H to compute the endpoints of the 1σ intervals of the b-value and, in addition, the formula (H17c) to compute its amplitude.

CONCLUSIONS

This paper consisted of two parts and was inspired by the new approach due to van der Elst (2021) to deal with a strongly incomplete data set of magnitudes to make estimates of the b-value. The theoretical part, mostly developed in the Appendixes, recalls in a plane way the main properties of the binned magnitude distributions, more precisely the discrete exponential and the discrete Laplace distribution, the latter being analysed also in its variants of absolute differences (either including or not null differences) and one-sign differences. It is a systematic analysis reproducing known results, but also providing clarifications and leading to new outcomes, such as the expressions to compute the 1σ limits of the estimators. The second part, chiefly illustrated in the main text and complemented by the supplementary material, is an attempt to evaluate the classical estimators of the b-value compared to the ones based on magnitude differences. For estimators based on the distribution of magnitudes, exact formulations (eq. 7) are always preferable with respect to the approximate formula by Aki (1965) with Utsu (1966) correction (eq. 4), in particular when the bin size is larger than 0.1. The uncorrected formula by Aki (1965) (eq. 2) usually overestimates the theoretical b-value, but sometimes may deceptively appear to work well when, by chance, the overestimation due to the binning almost exactly compensates the underestimation due to incompleteness.

In the case of substantially incomplete catalogues, it was shown that the distributions of magnitude differences happen to be closer to an exponential decay. Therefore, estimators using magnitude differences (eqs 15 and 17) are more robust with respect to magnitude incompleteness than those using magnitudes (eq. 7) and give correct b-values when the magnitude cutting threshold M_c is not lower than the magnitude of maximum curvature of the frequency magnitude distribution. Conversely, estimators using magnitudes (eq. 7) give correct results only for M_c not lower than the magnitude of maximum curvature plus 0.2. The latter finding confirms the goodness of a common choice, made in current literature (Mignan & Woessner 2012), to establish the magnitude completeness threshold.

For aftershock sequences, where completeness is time-dependent, we have found that estimates based on magnitude absolute magnitude differences, obtained through formula (23), or on one-signed magnitude differences, calculated by formula (22), are comparable, which contradicts the van der Elst (2021) assert that positive magnitude differences have to be preferred in seismological practice.

Finally, we want to stress that all the above considerations only hold if the frequency-magnitude distribution is really exponential, otherwise, even the concept of b-value becomes meaningless as it was evidenced by Herrmann and Marzocchi (2020) analysing high-resolution catalogues of Southern California and Central Italy.

ACKNOWLEDGMENTS

We thank Peng Han and an anonymous referee for thoughtful comment and suggestions that helped much to improve the paper. This study was partially supported by the Real-time earthquake rIsk reduction for a reSilient Europe (RISE) project, funded by the European Union's Horizon 2020 research and innovation program under Grant Agreement Number 821115.

AUTHOR CONTRIBUTIONS

S. Tinti (Conceptualization [equal], Investigation [equal], Methodology [equal], Supervision [equal], Validation [equal], Writing – original draft [equal], Writing – review and editing [equal]), and P. Gasperini (Conceptualization [equal], Funding acquisition [equal], Investigation [equal], Methodology [equal], Software [equal], Supervision [equal], Validation [equal], Writing – original draft [equal], Writing – review and editing [equal])

SUPPORTING INFORMATION

Supplementary data are available at *GJI* online.

Table S1. Estimates from complete simulated sets with $N = 100$, $2\delta = 0.1$ and $b = 0.7$.

Table S2. Estimates from complete simulated sets with $N = 1000$, $2\delta = 0.1$ and $b = 0.7$.

Table S3. Estimates from complete simulated sets with $N = 10\,000$, $2\delta = 0.1$ and $b = 0.7$.

Table S4. Estimates from complete simulated sets with $N = 100$, $2\delta = 0.1$ and $b = 1.0$.

Table S5. Estimates from complete simulated sets with $N = 1000$, $2\delta = 0.1$ and $b = 1.0$.

Table S6. Estimates from complete simulated sets with $N = 10\,000$, $2\delta = 0.1$ and $b = 1.0$.

Table S7. Estimates from complete simulated sets with $N = 100$, $2\delta = 0.1$ and $b = 1.5$.

Table S8. Estimates from complete simulated sets with $N = 1000$, $2\delta = 0.1$ and $b = 1.5$.

Table S9. Estimates from complete simulated sets with $N = 10\,000$, $2\delta = 0.1$ and $b = 1.5$.

Table S10. Estimates from complete simulated sets with $N = 100$, $2\delta = 0.5$ and $b = 0.7$.

Table S11. Estimates from complete simulated sets with $N = 1000$, $2\delta = 0.5$ and $b = 0.7$.

Table S12. Estimates from complete simulated sets with $N = 10\,000$, $2\delta = 0.5$ and $b = 0.7$.

Table S13. Estimates from complete simulated sets with $N = 100$, $2\delta = 0.5$ and $b = 1.0$.

Table S14. Estimates from complete simulated sets with $N = 1000$, $2\delta = 0.5$ and $b = 1.0$.

Table S15. Estimates from complete simulated sets with $N = 10\,000$, $2\delta = 0.5$ and $b = 1.0$.

Table S16. Estimates from complete simulated sets with $N = 100$, $2\delta = 0.5$ and $b = 1.5$.

Table S17. Estimates from complete simulated sets with $N = 1000$, $2\delta = 0.5$ and $b = 1.5$.

Table S18. Estimates from complete simulated sets with $N = 10\,000$, $2\delta = 0.5$ and $b = 1.5$.

Table S19. Aftershock sequence with time-dependent incompleteness; parameters: $\lambda = 0.2$, $m = 5.6$, $p_O = 1$, $c_O = 0.01$, $T_E = 5$ d, $N = 1600$ (before thinning), $2\delta = 0.1$, $b = 0.7$, $M_c = M_{mxc} + 4\delta = 1.4$.

Table S20. Aftershock sequence with time-dependent incompleteness; parameters: $\lambda = 0.2$, $m = 5.6$, $p_O = 1$, $c_O = 0.01$, $T_E = 5$ d, $N = 16\,000$ (before thinning), $2\delta = 0.1$, $b = 0.7$, $M_c = M_{mxc} + 4\delta = 1.4$.

Table S21. Aftershock sequence with time-dependent incompleteness; parameters: $\lambda = 0.2$, $m = 5.6$, $p_O = 1$, $c_O = 0.01$, $T_E = 5$ d, $N = 160\,000$ (before thinning), $2\delta = 0.1$, $b = 0.7$, $M_c = M_{mxc} + 4\delta = 1.4$.

Table S22. Aftershock sequence with time-dependent incompleteness; parameters: $\lambda = 0.2$, $m = 5.6$, $p_O = 1$, $c_O = 0.01$, $T_E = 5$ d, $N = 4000$ (before thinning), $2\delta = 0.1$, $b = 1.0$, $M_c = M_{mxc} + 4\delta = 1.3$.

Table S23. Aftershock sequence with time-dependent incompleteness; parameters: $\lambda = 0.2$, $m = 5.6$, $p_O = 1$, $c_O = 0.01$, $T_E = 5$ d, $N = 40\,000$, (before thinning), $2\delta = 0.1$, $b = 1.0$, $M_c = M_{mxc} + 4\delta = 1.3$.

Table S24. Aftershock sequence with time-dependent incompleteness; parameters: $\lambda = 0.2$, $m = 5.6$, $p_O = 1$, $c_O = 0.01$, $T_E = 5$ d, $N = 400\,000$, (before thinning), $2\delta = 0.1$, $b = 1.0$, $M_c = M_{mxc} + 4\delta = 1.3$.

Table S25. Aftershock sequence with time-dependent incompleteness; parameters: $\lambda = 0.2$, $m = 5.6$, $p_O = 1$, $c_O = 0.01$, $T_E = 5$ d, $N = 16\,000$, (before thinning), $2\delta = 0.1$, $b = 1.5$, $M_c = M_{mxc} + 4\delta = 1.2$.

Table S26. Aftershock sequence with time-dependent incompleteness; parameters: $\lambda = 0.2$, $m = 5.6$, $p_O = 1$, $c_O = 0.01$, $T_E = 5$ d, $N = 160\,000$, (before thinning), $2\delta = 0.1$, $b = 1.5$, $M_c = M_{mxc} + 4\delta = 1.2$.

Table S27. Aftershock sequence with time-dependent incompleteness; parameters: $\lambda = 0.2$, $m = 5.6$, $p_O = 1$, $c_O = 0.01$, $T_E = 5$ d, $N = 1600\,000$, (before thinning), $2\delta = 0.1$, $b = 1.5$, $M_c = M_{mxc} + 4\delta = 1.2$.

DATA AND SOFTWARE AVAILABILITY STATEMENT

Matlab codes written by authors for simulated and real data analyses as well as the data of the first 1000 shocks from the 30/10/2016 M_w 6.6 earthquake taken from the Horus catalogue at horus.bo.ingv.it, are provided freely at https://github.com/pgaspy/b-value-testing.

REFERENCES

Aki, M., 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**(B1), doi:10.1029/2001JB000680.

Bender, B., 1983. Maximum likelihood estimation of *b* values from magnitude grouped data. *Bull. seism. Soc. Am.,* **73,** 831–851.

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. seism. Soc. Am.,* **94,** 2380–2399.

Chiaraluce, L. *et al.*, 2017. The 2016 Central Italy seismic sequence: a first look at the mainshocks, aftershocks, and source models. *Seismol. Res. Lett.,* **88,** 757–771.

Gulia, L., Rinaldi, A.P., Tormann, T., Vannucci, G., Enescu, B. & Wiemer, S., 2018. The effect of a mainshock on the size distribution of the aftershocks. *Geophys. Res. Lett.,* **45**(24), 13 277–13 287.

Gulia, L. & Wiemer, S., 2010. The influence of tectonic regimes on the earthquake size distribution: a case study for Italy. *Geophys. Res. Lett.,* **37**(10), doi:10.1029/2010GL043066.

Gulia, L. & Wiemer, S., 2019. Real-time discrimination of earthquake foreshocks and aftershocks. *Nature,* **574,** 193–199.

Gulia, L., Wiemer, S. & Vannucci, G., 2020. Pseudoprospective evaluation of the Foreshock Traffic-Light system in Ridgecrest and implications for aftershock hazard assessment. *Seismol. Res. Lett.,* **91,** 2828–2842.

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

Guttorp, P. & Hopkins, D., 1986. On estimating varying *b* values. *Bull. seism. Soc. Am.,* **76,** 889–895.

Hardebeck, J.L., Felzer, K.R. & Michael, A.J., 2008. Improved tests reveal that the accelerating moment release hypothesis is statistically insignificant. *J. geophys. Res.,* **113**(B8), doi:10.1029/2007JB005410.

Helmstetter, A., Kagan, Y.Y. & Jackson, D.D., 2006. Comparison of short-term and time-independent earthquake forecast models for southern California. *Bull. seism. Soc. Am.,* **96**(1), 90–106.

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

Hurvich, C.M. & Tsai, C.L., 1989. Regression and time series model selection in small samples. *Biometrika,* **76,** 297–307.

Improta, L., Latorre, D., Margheriti, L., Nardi, A., Marchetti, A. & Lombardi, A.M.& The Bollettino Sismico Italiano Working Group., 2019. Multi-segment rupture of the 2016 Amatrice-Visso-Norcia seismic sequence (central Italy) constrained by the first high-quality catalog of early aftershocks. *Sci. Rep.,* **9,** doi:10.1038/s41598-019-43393-2.

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

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

Kagan, Y.Y., 2003. Accuracy of modern global earthquake catalogs. *Phys. Earth planet. Int.,* **135,** 173–209.

Lolli, B., Boschi, E. & Gasperini, P., 2009. A comparative analysis of different models of aftershock rate decay by maximum likelihood estimation of simulated sequences. *J. geophys. Res.,* **114**(B1), doi:10.1029/2008JB005614.

Lolli, B., Randazzo, D., Vannucci, G. & Gasperini, P., 2020. The Homogenized instrumental Seismic Catalog (HORUS) of Italy from 1960 to present. *Seismol. Res. Lett.,* **91,** 3208–3222.

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

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

Ogata, Y., 1981. On Lewis' simulation method for point processes. *IEEE Trans. Inform. Theory,* **IT-27,** 23–31.

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

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

Petruccelli, A., Gasperini, P., Tormann, T., Schorlemmer, D., Rinaldi, A.P., Vannucci, G. & Wiemer, S., 2019b. Simultaneous dependence of the earthquake-size distribution on faulting style and depth. *Geophys. Res. Lett.,* **4,** 11 044–11 053.

Petruccelli, A., Schorlemmer, D., Tormann, T., Rinaldi, A.P., Wiemer, S., Gasperini, P. & Vannucci, G., 2019a. The influence of faulting style on the size-distribution of global earthquakes. *Earth planet. Sci. Lett.,* **527,** doi:10.1016/j.epsl.2019.115791.

Petruccelli, A., Vannucci, G., Lolli, B. & Gasperini, P., 2018. Harmonic fluctuation of the slope of the frequency–magnitude distribution (b-value) as a function of the angle of rake. *Bull. seism. Soc. Am.,* **108**(4), 1864–1876.

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

Scholz, C.H., 2015. On the stress dependence of the earthquake b value. *Geophys. Res. Lett.,* **42,** 1399–1402.

Schorlemmer, D., Wiemer, S. & Wyss, M., 2005. Variations in earthquake-size distribution across different stress regimes. *Nature,* **437,** 539–542.

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

Spada, M., Tormann, T., Wiemer, S. & Enescu, B., 2013. Generic dependence of the frequency-size distribution of earthquakes on depth and its relation to the strength profile of the crust. *Geophys. Res. Lett.,* **40,** 709–714.

Tinti, S. & Mulargia, F., 1986. On the estimation of the geometric distribution parameter in seismology. *Statistica,* **46,** 163–188.
Tinti, S. & Mulargia, F., 1987. Confidence intervals of b values for grouped magnitudes. *Bull. seism. Soc. Am.,* **77,** 2125–2134.
Utsu, T., 1961. A statistical study of the occurrence of aftershocks. *Geophys. Mag.,* **30,** 521–605.
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, (in Japanese with English summary).
Utsu, T., 1966. A statistical significance test of the difference in b-value between two earthquake groups. *J. Phys. Earth,* **14,** 34–37.
van der Elst, N.J., 2021. B-positive: a robust estimator of aftershock magnitude distribution in transiently incomplete catalogs. *J. geophys. Res.,* **126,** e2020JB021027.
Wiemer, S. & Wyss, M., 2000. Minimum magnitude of completeness in earthquake catalogs: examples from Alaska, the western United States, and Japan. *Bull. seism. Soc. Am.,* **90,** 859–869.
Woessner, J. & Wiemer, S., 2005. Assessing the quality of earthquake catalogues: estimating the magnitude of completeness and its uncertainty. *Bull. seism. Soc. Am.,* **95,** 684–698.
Zöller, G., Hainzl, S. & Holschneider, M., 2010. Recurrence of large earthquakes: bayesian inference from catalogs in the presence of magnitude uncertainties. *Pure appl. Geophys.,* **167,** 845–853.

APPENDIX A: THE CONTINUOUS AND DISCRETE EXPONENTIAL DISTRIBUTIONS

Earthquake magnitudes, when taken as random variables, are supposed to follow an exponential distribution at least beyond a certain (completeness) magnitude threshold M_c. Generally, they are provided up to a few decimal digits (usually one) and therefore they can be naturally binned in classes of equal size, say 2δ. In common practice, they are treated either as a discrete set of variables or as a continuous set. In the former case, if M_0 is the magnitude of the first class, the magnitude of the ith class is given by:

$$M_i = M_0 + 2\delta i. \tag{A1}$$

The integer i identifying the class is a discrete random variable obeying the probability distribution:

$$P_i = A(\alpha)\, e^{-\alpha i} \quad i = 0, 1, 2, \ldots \tag{A2}$$

where α is assumed to be a positive decay parameter. Because P_i represents a probability for the random variable i, its distribution must satisfy the normalization condition, i.e. the sum of all probabilities must be equal to 1. By imposing it, we obtain:

$$\sum_{i=0}^{\infty} A(\alpha)\, e^{-\alpha i} = A(\alpha) \sum_{i=0}^{\infty} e^{-\alpha i} = \frac{A(\alpha)}{1 - e^{-\alpha}} = 1. \tag{A3}$$

It follows that ($A2$) can be rewritten as:

$$P_i = \left(1 - e^{-\alpha}\right)\, e^{-\alpha i} \quad i = 0, 1, 2, \ldots \tag{A4}$$

On the other hand, when treating the magnitude M as a continuous variable, its probability density function is given by:

$$P(M) = \beta e^{-\beta(M - M_c)} \quad M - M_c \geq 0 \tag{A5a}$$

or

$$P(M) = \beta e^{-\beta(M - M_0 + \delta)} \quad M - (M_0 - \delta) \geq 0 \tag{A5b}$$

depending on the decay factor β. The formula ($A5b$) is justified since, usually, the first value of the discrete set of magnitudes M_0 is taken as the midpoint of the first magnitude class, that is the one with the lower endpoint in M_c. This means that:

$$M_c = M_0 - \delta \tag{A5}$$

It can be shown that the decay factors α and β of the discrete and continuous distributions are linked by the relation:

$$\alpha = 2\delta\beta \tag{A7}$$

Indeed, if we consider the scaled variable:

$$y = \frac{M - M_c}{2\delta} \tag{A8}$$

then its probability density has the form:

$$P(y) = P(M)\frac{\mathrm{d}M}{\mathrm{d}y} = 2\delta\beta\, e^{-2\delta\beta y} = \alpha e^{-\alpha y} \quad y \geq 0. \tag{A9}$$

In the following, the random variables we will consider are the continuous variable y defined in ($A8$) and the discrete variable i defined in ($A1$). We will see that all statistical formulas we will derive for the discrete variable i will tend to the corresponding formulas of the continuous variable y as the bin size 2δ becomes increasingly small. More specifically, if we approximate $e^{-\alpha}$ with 1 and $(1 - e^{-\alpha})$ with $2\delta\beta$, then the discrete-case expressions transform into the continuous-case ones.

Mean, variance and standard deviation

The formulas for the mean and variance of the continuous exponential distribution ($A9$) are well known and will be given here for the sake of completeness. They are:

$$\mu_{CE} = \frac{1}{\alpha} \tag{A10}$$

$$var_{CE} = \frac{1}{\alpha^2};\quad \sigma_{CE} = \frac{1}{\alpha} \tag{A11}$$

where the subscript CE denotes a continuous exponential random variable. As regards the discrete distribution ($A4$), we start with computing its mean μ_{DE} that is defined as:

$$\mu_{DE} = \sum_{i=0}^{\infty} i\ P_i = \left(1 - e^{-\alpha}\right) \sum_{i=1}^{\infty} i\ e^{-\alpha i} \tag{A12}$$

Consider that the sum, say S_1, in the last term can be easily evaluated as:

$$S_1 = \frac{e^{-\alpha}}{(1 - e^{-\alpha})^2} \tag{A13}$$

On substituting this expression in the definition ($A12$), we eventually get:

$$\mu_{DE} = \left(1 - e^{-\alpha}\right) S_1 = \frac{e^{-\alpha}}{1 - e^{-\alpha}} \tag{A14}$$

The second moment, say $M_{2,DE}$, of the discrete exponential distribution, is by definition:

$$M_{2,DE} = \sum_{i=0}^{\infty} i^2 P_i = \left(1 - e^{-\alpha}\right) \sum_{i=1}^{\infty} i^2 e^{-\alpha i} \tag{A15}$$

To evaluate the sum of the series, one can consider that:

$$S_2 = \sum_{i=1}^{\infty} i^2 e^{-\alpha i} = e^{-\alpha} \sum_{i=0}^{\infty} (i+1)^2 e^{-\alpha i} = e^{-\alpha} \left(S_2 + 2 S_1 + \frac{1}{1 - e^{-\alpha}} \right) \tag{A16}$$

Solving the eq. (A16), one gets:

$$S_2 = \frac{e^{-\alpha}\left(1 + e^{-\alpha}\right)}{(1 - e^{-\alpha})^3} \tag{A17}$$

From the definition ($A15$), one eventually obtains:

$$M_{2,DE} = \left(1 - e^{-\alpha}\right) S_2 = \frac{e^{-\alpha}\left(1 + e^{-\alpha}\right)}{(1 - e^{-\alpha})^2} \tag{A18}$$

As is well known, the variance of a distribution can be computed from the mean and the second moment, which leads us to the formula:

$$var_{DE} = \frac{e^{-\alpha}\left(1 + e^{-\alpha}\right)}{(1 - e^{-\alpha})^2} - \frac{e^{-2\alpha}}{(1 - e^{-\alpha})^2} = \frac{e^{-\alpha}}{(1 - e^{-\alpha})^2} = \frac{1}{4}\left(\mathrm{csch}\frac{\alpha}{2}\right)^2 \tag{A19}$$

Consequently, the standard deviation σ_{DE} takes the form:

$$\sigma_{DE} = \frac{e^{-\frac{\alpha}{2}}}{1 - e^{-\alpha}} = \frac{1}{2}\mathrm{csch}\frac{\alpha}{2} \tag{A20}$$

We point out here that the formula ($A14$) for the mean μ_{DE} is already known in the seismological literature and forms the basis for the estimator of b (eq. 7) given in the main text (Guttorp & Hopkins 1986; Tinti & Mulargia 1987; van der Elst 2021). On the contrary, the expression ($A20$) for the standard deviation of the discrete distribution is original and first derived in this paper.

APPENDIX B: THE CONTINUOUS AND DISCRETE DISTRIBUTIONS OF THE DIFFERENCES OF EXPONENTIAL VARIABLES (LAPLACE DISTRIBUTIONS)

If we consider the scaled random variables y and z, both following the exponential distribution ($A9$), then the random variable $w = z - y$ can be proven to obey the continuous Laplace distribution with density function defined as:

$$P(w) = \frac{\alpha}{2} e^{-\alpha|w|} \quad -\infty < w < +\infty \tag{B1}$$

It is a continuous density function with two symmetric exponential tails, and the same decay parameter α as the original distributions and it is known as Laplace distribution.

Let us now consider the differences of integer random variables i and j, both following the discrete exponential distribution ($A4$) with the same parameter α. If they are independent variables, the joint probability distribution P_{ij} for the pair (i, j) is given by the product:

$$P_{i,j} = P_i P_j = \left(1 - e^{-\alpha}\right)^2 e^{-\alpha(i+j)} \tag{B2}$$

We introduce the random variable $d = j - i, d \in Z$, and compute its distribution P_d. First, we assume that $j \geq i$, and therefore that $d \geq 0$. Given d, all pairs (i, j) having differences equal to d, are of the type $(i, i + d)$ with $i \in N$. It follows that:

$$P_d = \sum_{i=0}^{\infty} P_i P_{i+d} = \left(1 - e^{-\alpha}\right)^2 e^{-\alpha d} \sum_{i=0}^{\infty} e^{-2\alpha i} \tag{B3}$$

Considering that the terms to be summed can be seen as the elements of a geometric series with a constant ratio $e^{-2\alpha}$, we obtain the expression:

$$P_d = \frac{\left(1 - e^{-\alpha}\right)^2}{1 - e^{-2\alpha}} e^{-\alpha d} = \frac{1 - e^{-\alpha}}{1 + e^{-\alpha}} e^{-\alpha d}, d \geq 0 \tag{B4}$$

When $j \leq i$, following an analogous procedure, we can get a similar expression. Indeed, we should sum up all probabilities of the pairs $(j + |d|, j)$ getting the result:

$$P_d = \frac{1 - e^{-\alpha}}{1 + e^{-\alpha}} e^{\alpha d}, \quad d < 0 \tag{B5}$$

Both expressions $(B4)$ and $(B5)$ can be synthesized in the form:

$$P_d = \frac{1 - e^{-\alpha}}{1 + e^{-\alpha}} e^{-\alpha|d|} = \tanh \frac{\alpha}{2} e^{-\alpha|d|} \quad -\infty < d < +\infty \tag{B6}$$

In the following the distribution $(B6)$ will be referenced as a discrete Laplace distribution.

Mean, variance and standard deviation

The computation of the mean of the continuous Laplace distribution μ_{CL} is straightforward, since $P(w) = P(-w)$, which implies that:

$$\mu_{CL} = 0 \tag{B7}$$

Owing to the vanishing of μ_{CL}, the second moment of the Laplace distribution $(B1)$ coincides with its variance:

$$var_{CL} = \int_{-\infty}^{+\infty} w^2 P(w)\, dw = \alpha \int_{0}^{+\infty} w^2 e^{-\alpha w} dw = \frac{2}{\alpha^2} \tag{B8}$$

Hence, the corresponding standard deviation is:

$$\sigma_{CL} = \frac{\sqrt{2}}{\alpha} \tag{B9}$$

Also, the mean μ_{DL} of the discrete distribution $(B6)$ is zero due to its symmetry around the origin (i.e. $P_{-d} = P_d$) and, as a consequence, its second moment and variance are coincident:

$$var_{DL} = \sum_{d=-\infty}^{\infty} d^2\ P_d = 2\sum_{d=1}^{\infty} d^2\ P_d = 2\frac{1 - e^{-\alpha}}{1 + e^{-\alpha}} S_2 = \frac{2e^{-\alpha}}{(1 - e^{-\alpha})^2} = \frac{1}{2}\left(\mathrm{csch}\frac{\alpha}{2}\right)^2 \tag{B10}$$

The corresponding standard deviation results to be:

$$\sigma_{DL} = \frac{\sqrt{2}e^{-\frac{\alpha}{2}}}{1 - e^{-\alpha}} = \frac{1}{\sqrt{2}}\mathrm{csch}\frac{\alpha}{2} \tag{B11}$$

On comparing expressions $(A11)$ with $(B8)$ and $(A19)$ with $(B10)$, it is worth noting that the variances of the Laplace distributions are twice larger than the corresponding variances of the exponential distributions, that is:

$$var_{CL} = 2var_{CE}\ var_{DL} = 2var_{DE} \tag{B12}$$

Indeed, the results reached in this section could be anticipated simply by remembering that the mean and the variance of the difference of two independent random variables are, respectively, the difference of their mean and the sum of their variances, which entails that the resulting mean is zero and the resulting variance is twice as large.

APPENDIX C: THE CONTINUOUS AND DISCRETE ONE-SIGN DIFFERENCES DISTRIBUTIONS

If we restrict the attention only to one-sign differences, it is trivial to see that their distribution is exponential. Indeed, for the continuous case, the distribution $(B1)$ becomes:

$$P(w) = \alpha e^{-\alpha|w|} \quad -\infty < w \leq 0 \tag{C1a}$$

$$P(w) = \alpha e^{-\alpha w} \qquad 0 < w < +\infty \tag{C1b}$$

Likewise, for the discrete case, the distribution $(B6)$ splits into:

$$P_d = \left(1 - e^{-\alpha}\right) e^{-\alpha|d|} \quad -\infty < d \leq 0 \tag{C2a}$$

$$P_d = \left(1 - e^{-\alpha}\right) e^{-\alpha d} \qquad 0 \le d < +\infty \tag{C2b}$$

It follows that the corresponding means, variances, and standard deviation have the expressions ($A10$) and ($A11$) given in Appendix A for the continuous case, and (A15), (A19) and (A20) for the discrete case. It is worth stressing that the result regarding the means was derived first by van der Elst (2021).

APPENDIX D. THE CONTINUOUS AND DISCRETE ABSOLUTE DIFFERENCES DISTRIBUTIONS

Let us consider the absolute values of the differences, which are $|w|$ and $|d|$, respectively. It is worth outlining that for the continuous case, the distribution is exponential, while for the discrete variables, this is not true. In the former case, we can write:

$$P\left(|w|\right) = \alpha e^{-\alpha|w|} \; 0 \le |w| \le +\infty \tag{D1}$$

On the other hand, for the discrete variable $|d|$, we should distinguish the case of null differences from the others, and their probability distributions results to be:

$$P_0 = \frac{1 - e^{-\alpha}}{1 + e^{-\alpha}} \tag{D2a}$$

$$P_{|d|} = 2\frac{1 - e^{-\alpha}}{1 + e^{-\alpha}} e^{-\alpha|d|} \quad 1 \le |d| < +\infty \tag{D2b}$$

Mean, variance and standard deviation

The absolute values of the continuous differences are exponential variables and their statistical moments that are relevant in our context can be taken from the expressions displayed in Appendix A. We can write them explicitly here below:

$$\mu_{CA} = \frac{1}{\alpha}, \; var_{CA} = \frac{1}{\alpha^2}, \; \sigma_{CA} = \frac{1}{\alpha} \tag{D3}$$

In the adopted notation the subscript CA stands for continuous absolute differences. The mean of the absolute values of the discrete differences is by definition given by:

$$\mu_{DA} = \sum_{|d|=1}^{\infty} |d| \, P_{|d|} = 2\frac{1 - e^{-\alpha}}{1 + e^{-\alpha}} \sum_{i=1}^{\infty} i e^{-\alpha i} = \frac{2e^{-\alpha}}{(1 + e^{-\alpha})(1 - e^{-\alpha})} \tag{D4}$$

Likewise, the second moment $M_{2,DA}$ is computed as:

$$M_{2,DA} = \sum_{|d|=1}^{\infty} |d|^2 P_{|d|} = 2\frac{1 - e^{-\alpha}}{1 + e^{-\alpha}} \sum_{i=1}^{\infty} i^2 e^{-\alpha i} = \frac{2e^{-\alpha}}{(1 - e^{-\alpha})^2} \tag{D5}$$

It follows that the variance is:

$$var_{DA} = \frac{2e^{-\alpha}\left(1 + e^{-2\alpha}\right)}{(1 + e^{-\alpha})^2 (1 - e^{-\alpha})^2} \tag{D6}$$

The related standard deviation is therefore given by:

$$\sigma_{DA} = \frac{\sqrt{2e^{-\alpha}\left(1 + e^{-2\alpha}\right)}}{(1 + e^{-\alpha})(1 - e^{-\alpha})} \tag{D7}$$

It is relevant to observe that the variance of the absolute differences ($D6$) is smaller than the variance of the discrete Laplace distribution ($B10$), i.e.:

$$var_{DA} = var_{DL} \frac{1 + e^{-2\alpha}}{(1 + e^{-\alpha})^2} < var_{DL} \tag{D8a}$$

since the adjusting factor is smaller than 1. Similarly, we can conclude that:

$$\sigma_{DA} < \sigma_{DL} \tag{D8b}$$

APPENDIX E: THE EFFECT OF TRIMMING

For trimming we mean here the removal of all values below a predefined limit. Therefore, for the continuous variable y, we will consider only values $y \ge y' > 0$, and, likewise, for the continuous difference w we will take into account only values $w \ge w' > 0$ or $w \le -w' < 0$. It is very easy to see that the distribution of y follows the exponential distribution with mean:

$$\mu_{T,CE} = \mu_{CE} + y' = \frac{1}{\alpha} + y' \tag{E1a}$$

while variance and standard deviation remain unchanged, i.e.:

$$var_{T,CE} = var_{CE}\frac{1}{\alpha^2}\ \sigma_{T,CE} = \sigma_{CE} = \frac{1}{\alpha} \tag{E2}$$

Here the additional subscript T denotes the trimmed distribution. Further, it is immediate to observe that also the one-sign differences $w - w'$ and the absolute differences $|w - w'|$ follow an exponential distribution, that is:

$$P(w) = \alpha e^{-\alpha(w-w')} \quad w \geq w' > 0 \tag{E3a}$$

$$P(w) = \alpha e^{-\alpha|w-w'|} \quad w \leq -w' < 0 \tag{E3b}$$

$$P(|w|) = \alpha e^{-\alpha|w-w'|} \quad |w| \geq w' > 0 \tag{E3c}$$

Therefore, even for these distributions, the mean results to be shifted by an amount equal to w', whereas variance and standard deviation do not change.

When considering the continuous Laplace distribution, appropriate for the differences, trimming is realized by considering the variables with absolute values larger than the threshold. The related density function is split into:

$$P(w) = \frac{\alpha}{2}e^{-\alpha(w-w')} \quad w \geq w' > 0 \tag{E4a}$$

$$P(w) = \frac{\alpha}{2}e^{-\alpha|w-w'|} \quad w \leq -w' < 0 \tag{E4b}$$

It is symmetric, centred in zero, and therefore, if we denote its mean by $\mu_{T,CL}$, we can write:

$$\mu_{T,CL} = 0 \tag{E5}$$

As for the variance, it identifies with the second moment and can be written as:

$$var_{T,CL} = 2\int_{w'}^{\infty} w^2 P(w)\,dw = \alpha e^{\alpha w'}\int_{w'}^{\infty} w^2 e^{-\alpha w}dw = \frac{1}{\alpha^2}\left[\left(1+\alpha w'\right)^2 + 1\right] \tag{E6}$$

and correspondingly the standard deviation:

$$\sigma_{T,CL} = \frac{1}{\alpha}\sqrt{(1+\alpha w')^2 + 1} \tag{E7}$$

Both expressions tend to the respective values ($B8$) and ($B9$) of the untrimmed distributions as w' tends to zero, that is:

$$var_{T,CL} \to var_{CL} \text{ and } \sigma_{T,CL} \to \sigma_{CL} \text{ as } w' \to 0 \tag{E8}$$

Notice further that both are increasing functions of w'.

As regards the discrete distributions, trimming is realized by considering only variables beyond specified integer thresholds, say i' and d'. Even in this case, the trimmed exponential distributions and the one-sign differences are exponential, that is:

$$P_i = (1-\alpha)\,e^{-\alpha(i-i')} \quad i \geq i' > 0 \qquad \text{positive differences} \tag{E9}$$

$$P_d = (1-\alpha)\,e^{-\alpha|d-d'|} \quad d \leq -d' < 0 \qquad \text{negative differences} \tag{E10a}$$

$$P_d = (1-\alpha)\,e^{-\alpha(d-d')} \quad d \geq d' > 0 \qquad \text{absolute differences} \tag{E10b}$$

So the means are affected by trimming, whereas variances and standard deviations are not.

We observe that trimming affects substantially the distribution of the absolute differences. Indeed, since trimming discards the value $d = 0$, the resulting distribution becomes exponential. It is worth to write it down explicitly:

$$P_{|d|} = (1-\alpha)\,e^{-\alpha(|d|-d')} \quad |d| \geq d' > 0 \tag{E11}$$

Its relevant statistical indices are quite different from the ones of the untrimmed distribution (see expressions ($D4$), ($D6$) and ($D7$)). They are:

$$\mu_{T,DA} = \frac{e^{-\alpha}}{1-e^{-\alpha}} + d';\ var_{T,DA} = \frac{e^{-\alpha}}{(1-e^{-\alpha})^2};\ \sigma_{T,DA} = \frac{e^{-\frac{\alpha}{2}}}{1-e^{-\alpha}} \tag{E12}$$

For the differences distributed according to the discrete Laplace distribution, trimming leads to the following expression for the probabilities:

$$P_d = \frac{1-e^{-\alpha}}{1+e^{-\alpha}}e^{-\alpha(|d|-d')} = \tanh\frac{\alpha}{2}e^{-\alpha(|d|-d')} \quad d \leq -d' < 0 \text{ or } d \geq d' > 0 \tag{E13}$$

It is a symmetric distribution with mean equal to zero, that is:

$$\mu_{T,DL} = 0 \tag{E14}$$

Its variance can be computed as:

$$var_{T,DL} = 2\sum_{d=d'}^{\infty} d^2 P_d = 2\frac{1-e^{-\alpha}}{1+e^{-\alpha}}e^{\alpha d'}\sum_{j=0'}^{\infty}\left(j+d'\right)^2 e^{-\alpha j} \tag{E15}$$

The summation in the RHS of the last equation can be further elaborated:

$$\sum_{j=0'}^{\infty}\left(j+d'\right)^2 e^{-\alpha j} = S_2 + 2d' S_1 + \frac{d'^2}{1-e^{-\alpha}} \tag{E16}$$

Combining $(E15)$ and $(E16)$ after some calculations we eventually get:

$$var_{T,DL} = \frac{2}{\left(1+e^{-\alpha}\right)\left(1-e^{-\alpha}\right)^2}\left\{e^{-\alpha} + \left[e^{-\alpha} + d'\left(1-e^{-\alpha}\right)\right]^2\right\} \tag{E17a}$$

$$\sigma_{T,DL} = \frac{1}{1-e^{-\alpha}}\sqrt{\frac{2}{\left(1+e^{-\alpha}\right)}\left\{e^{-\alpha} + \left[e^{-\alpha} + d'\left(1-e^{-\alpha}\right)\right]^2\right\}} \tag{E17b}$$

It is worth noting that when d' is set equal to zero, both the above expressions transform into the corresponding untrimmed variables indices, that is var_{DL} and σ_{DL}.

APPENDIX F: ESTIMATING THE DECAY PARAMETERS BY MEANS OF THE MEAN METHOD

For magnitudes obeying the Gutenberg–Richter formula (1) the decay parameter is b. If we opt for the canonical exponential expressions $(A5)$, the decay parameter is β. If we consider binned magnitudes, the decay parameter is α. Since these three parameters are linked by constant factors, we can estimate any one of them and very easily deduce the others. In this paper, the main attention goes to sequences of binned magnitudes and therefore here we concentrate on methods suitable to estimate α and only on discrete distributions. In this Appendix, we will consider methods based on the mean value of the distributions. If we denote the generic mean by μ, and if it happens to depend on α, that is if $\mu = f(\alpha)$, then we can obtain α by means of the expression $\alpha = f^{-1}(\mu)$ where f^{-1} is the inverse function of f, provided that the inverse function exists. On the other hand, the mean of any distribution can be estimated from experimental data, and approximated by the sample mean value, the approximation being better and better as the data number N increases. The goodness of the estimate of μ reflects directly on how good the estimate of α is. With this strategy in mind, we will consider separately the distributions treated so far, pointing out, however, that the method cannot be applied to the discrete Laplace distributions, either trimmed or untrimmed, because their mean μ_{DL} and $\mu_{T,DL}$ are identically zero, and thus not depending on α.

Estimates based on exponential distributions

The exponential distribution applies to binned trimmed or untrimmed magnitudes, as well as to binned trimmed or untrimmed one-sign magnitude differences, and also to binned trimmed absolute differences. In all these cases the formula for the mean can be written as (see Appendix E):

$$\mu = \frac{e^{-\alpha}}{1-e^{-\alpha}} + k \tag{F1}$$

where k is the trimming threshold and is equal to zero for untrimmed distributions.

The expression $(F1)$ can be inverted easily and leads to:

$$\alpha = \ln\left(\frac{\mu-k+1}{\mu-k}\right) \tag{F2}$$

Interestingly, we can observe that the ratio in the formula $(F2)$ can be written also as:

$$\frac{\mu-k+1}{\mu-k} = \frac{x+1}{x-1} \tag{F3a}$$

where we have posed:

$$x = 2\left(\mu-k+\frac{1}{2}\right) \tag{F3b}$$

Taking advantage of the identity:

$$\coth^{-1}(x) = \frac{1}{2}\ln\left(\frac{x+1}{x-1}\right) \tag{F4}$$

the variable α in $(F2)$ can be alternatively given also as:

$$\alpha = 2\coth^{-1}\left(2\left(\mu-k+\frac{1}{2}\right)\right) \tag{F5}$$

We stress that in the above formulas, α is the true value of the decay parameter. Therefore, we can interpret formulas $(F2)$ and $(F5)$ as unbiased estimators of α, say $\tilde{\alpha}$, if we replace μ with its sample mean, since the sample mean tends to μ when the amount of data in the sample increases.

In terms of binned magnitudes M_i given by $(A1)$ the above formulas $(F2)$ and $(F5)$ for the estimator $\tilde{\alpha}$ take the form:

$$\tilde{\alpha} = \ln\left(\frac{\bar{M}-M_k+2\delta}{\bar{M}-M_k}\right) = 2\coth^{-1}\left(\frac{1}{\delta}\left(\bar{M}-M_k+\delta\right)\right) \tag{F6}$$

where $\bar{M}$ is the sample mean magnitude and:

$$M_k = M_0 + 2\delta k \quad k \geq 0 \tag{F7}$$

is defined as the trimming threshold magnitude which coincides with the magnitude of the lowest bin if no trimming is applied.

Since $\alpha = 2\delta b \ln(10)$, the corresponding estimator of the decay parameter b is:

$$\tilde{b} = \frac{1}{2\delta \ln(10)} \ln\left(\frac{\bar{M} - M_k + 2\delta}{\bar{M} - M_k}\right) = \frac{1}{\delta \ln(10)} \coth^{-1}\left(\frac{1}{\delta}\left(\bar{M} - M_k + \delta\right)\right) \tag{F8}$$

Observe that the expressions $(F8)$ coincide with the estimators (8) and (12) in the main text, where however we used a different notation and called M_k as M_c.

The logarithmic version of the above formula can be rewritten as:

$$\tilde{b} = \frac{1}{2\delta \ln(10)} \ln\left(1 - \frac{2\delta}{\bar{M} - M_k}\right) \tag{F9}$$

This is a version expandable in series. If we truncate the expansion to the second order, we obtain:

$$\tilde{b} = \frac{1}{2\delta \ln(10)}\left[\frac{2\delta}{\bar{M} - M_k} - \frac{1}{2}\frac{4\delta^2}{\left(\bar{M} - M_k\right)^2}\right] = \frac{1}{\ln(10)\left(\bar{M} - M_k\right)}\left(1 - \frac{\delta}{\bar{M} - M_k}\right) \tag{F10}$$

It is interesting to observe that the formula (F10) when k=0, coincides with the first terms of the expansion of the expression (4) in the main text. Indeed:

$$\tilde{b} = \frac{1}{\ln(10)\left(\bar{M} - M_0 + \delta\right)} = \frac{1}{\ln(10)\left(\bar{M} - M_0\right)}\frac{1}{1 + \frac{\delta}{\bar{M} - M_0}} \approx \frac{1}{\ln(10)\left(\bar{M} - M_0\right)}\left(1 - \frac{\delta}{\bar{M} - M_0}\right)$$

Therefore, we can state that the estimator (4) is an approximation of the estimator for binned exponential magnitudes corrected at the second order in the variable $\delta/(\bar{M} - M_0)$.

When considering the binned magnitude differences, we come to analogous expressions for the estimator. If we denote by ΔM the magnitude differences, by $\overline{\Delta M}$ the related sample mean value, and by ΔM_k the trimming threshold, then for trimmed positive differences we obtain:

$$\tilde{b} = \frac{1}{2\delta \ln(10)} \ln\left(\frac{\overline{\Delta M} - \Delta M_k + 2\delta}{\overline{\Delta M} - \Delta M_k}\right) \quad \Delta M \geq \Delta M_k = 2k\delta \quad k \geq 0 \tag{F11}$$

For trimmed negative differences the formula is:

$$\tilde{b} = \frac{1}{2\delta \ln(10)} \ln\left(\frac{\overline{|\Delta M|} - \Delta M_k + 2\delta}{\overline{\Delta M} - \Delta M_k}\right) \quad \Delta M \leq \Delta M_k = -2k\delta \quad k \geq 0 \tag{F12}$$

Eventually, for the trimmed absolute differences, we get:

$$\tilde{b} = \frac{1}{2\delta \ln(10)} \ln\left(\frac{\overline{|\Delta M|} - \Delta M_k + 2\delta}{\overline{|\Delta M|} - \Delta M_k}\right) \quad |\Delta M| \geq \Delta M_k = 2k\delta \quad k \geq 1 \tag{F13}$$

All the above expressions $(F11) - (F13)$ can be also given in terms of the inverse hyperbolic cotangent, like in $(F8)$. They coincide with the formula given in the main text as eq. (17), provided that we change notation replacing ΔM_k with $\Delta M'_c$.

Estimates based on the untrimmed absolute differences distribution

The mean μ_{DA} of the distribution of the untrimmed absolute differences is given by:

$$\mu_{DA} = \frac{2e^{-\alpha}}{1 - e^{-2\alpha}} \tag{F14}$$

It is an invertible function of α. Indeed, the expression $(F14)$ can be transformed into:

$$\mu_{DA}e^{-2\alpha} + 2e^{-\alpha} - \mu_{DA} = 0 \tag{F15a}$$

that can be interpreted as a quadratic equation in the unknown $e^{-\alpha}$, with roots:

$$e^{-\alpha} = \frac{-1 \pm \sqrt{1 + \mu_{DA}{}^2}}{\mu_{DA}} = -\frac{1}{\mu_{DA}} \pm \sqrt{\frac{1}{\mu_{DA}{}^2} + 1} \tag{F15b}$$

Of the two roots, only the positive one is an admissible solution, since the exponential in the LHS must be positive. Thus we can write:

$$\alpha = \ln\left(\frac{-1 + \sqrt{1 + \mu_{DA}{}^2}}{\mu_{DA}}\right)^{-1} = \ln\left(\frac{1 + \sqrt{1 + \mu_{DA}{}^2}}{\mu_{DA}}\right) = \ln\left(\frac{1}{\mu_{DA}} + \sqrt{\frac{1}{\mu_{DA}{}^2} + 1}\right) = \operatorname{csch}^{-1}\left(\mu_{DA}\right) \tag{F16}$$

The last equality has been introduced by virtue of the following identity involving the natural logarithm and the inverse of the hyperbolic cosecant:

$$\mathrm{csch}^{-1}(x) = \ln\left(\frac{1}{x} + \sqrt{\frac{1}{x^2} + 1}\right) \tag{F17}$$

Substituting μ_{DA} with the sample mean, we obtain an unbiased estimator $\tilde{\alpha}$ and, in terms of the sample mean $\overline{|\Delta M|}$ of the absolute magnitude differences, we obtain an expression for $\tilde{\beta}$ that coincides with the formula (14) of the main text, that is:

$$\tilde{b} = \frac{1}{2\delta\ln(10)} \ln\left[\frac{2\delta + \sqrt{4\delta^2 + \left(\overline{|\Delta M|}\right)^2}}{\overline{|\Delta M|}}\right] = \frac{1}{2\delta\ln(10)} \mathrm{csch}^{-1}\left(\frac{\overline{|\Delta M|}}{2\delta}\right) \quad |\Delta M| \geq 0 \tag{F18}$$

APPENDIX G: ESTIMATING THE DECAY PARAMETERS THROUGH THE MAXIMUM LIKELIHOOD METHOD.

The decay parameter can be estimated also by means of the Maximum Likelihood (ML) approach. As a general observation, the main conceptual difference between the ML method and the method of the mean is that the former applies to empirical samples, while the latter uses relations proper of the theoretical distribution. However, provided that we replace the expected value of the distributions with the related sample means, the two methods are expected to lead to the same result. This is exactly what we will prove here, but we outline that there is an important caveat that we need to express for the estimator of the binned differences.

Using samples of the discrete exponential distribution

For the sake of simplicity, we will consider here only the untrimmed exponential distribution of binned magnitudes. Making recourse to the ML method, we introduce the Likelihood Function $L_N(\alpha)$ for a sample of N data $(i_1, i_2, .., i_N)$, that we write:

$$L_N(\alpha) = \left(1 - e^{-\alpha}\right)^N \prod_{s=1}^{n} e^{-\alpha i_s} \left(1 - e^{-\alpha}\right)^N e^{-\alpha N \bar{i}} \quad i_s \geq 0 \tag{G1}$$

where $\bar{i}$ is the arithmetic mean of the sample.

The ML estimate of the parameter α is that value, say $\tilde{\alpha}$, that maximizes $L_N(\alpha)$ and can be found by solving the equation;

$$\frac{d}{d\alpha}\left(1 - e^{-\tilde{\alpha}}\right) - \bar{i}\left(1 - e^{-\tilde{\alpha}}\right) = 0 \tag{G2}$$

Its solution is:

$$\tilde{\alpha} = \ln\left(\frac{1 + \bar{i}}{\bar{i}}\right) \tag{G3}$$

This corresponds to the expression $(F2)$, once we pose $k = 0$ and substitute the theoretical mean μ with the sample mean $\bar{i}$.

Using samples of the discrete Laplace distribution

We have remarked that the Laplace distribution of the magnitude differences is unsuitable to the application of the mean method since its mean is identically zero. However, we can apply the ML method. Let us consider the Likelihood Function $L_N(\alpha)$ as the product of three functions $L_{N_+}(\alpha)$, $L_{N_-}(\alpha)$ and $L_{N_0}(\alpha)$, where $N = N_+ + N_- + N_0$ and where the subscripts denote the absolute frequencies of differences, respectively, greater than, smaller than, and equal to zero. If we pose (see $(B6)$):

$$B(\alpha) = \tanh\frac{\alpha}{2} \tag{G4}$$

then we can write for positive differences:

$$L_{N_+}(\alpha) = (B(\alpha))^{N_+} e^{-\alpha \sum_{k=1}^{k=N_+} d_k} \quad d_k > 0 \tag{G5a}$$

Analogously, for negative differences, we have:

$$L_{N_-}(\alpha) = (B(\alpha))^{N_-} e^{-\alpha \sum_{k=1}^{k=N_-} |d_k|} \quad d_k < 0 \tag{G5b}$$

And for differences equal to zero:

$$L_{N_0}(\alpha) = (B(\alpha))^{N_0} \tag{G5c}$$

As a consequence, the Likelihood Function $L_N(\alpha)$ can be given the expression:

$$L_N(\alpha) = L_{N_+}(\alpha)\ L_{N_-}(\alpha)\ L_{N_0}(\alpha) = (B(\alpha))^N e^{-\alpha N \overline{|d|}} \qquad -\infty < d < \infty \tag{G6}$$

By imposing that the first derivative of $L_N(\alpha)$ with respect to α is equal to zero, we get the ML solving equation, that is:

$$\frac{dB(\tilde{\alpha})}{d\alpha} - \overline{|d|} B(\tilde{\alpha}) = 0 \qquad -\infty < d < \infty \tag{G7}$$

After some calculations, we get the solution:

$$\sinh\tilde{\alpha} = \frac{1}{|d|} \qquad -\infty < d < \infty \tag{G8}$$

which leads to the final expression for the estimator:

$$\tilde{\alpha} = \operatorname{csch}^{-1}\left(|d|\right) \qquad -\infty < d < \infty \tag{G9}$$

It is important to stress that the formula ($G9$) identifies with the formula ($F16$) that resulted from the application of the mean method to the binned untrimmed absolute differences.

APPENDIX H: 1σ INTERVALS

The decay parameters α and b derived in the previous Appendix F are functions of the mean μ of a distribution of a discrete variable i with probability P_i and standard deviation σ. Let us say that:

$$p = g(\mu) \tag{H1}$$

where p denotes the parameter and g the function. The corresponding estimator $\tilde{p}$ has been computed through the same function g as:

$$\tilde{p} = g(\bar{\mu}_N) \tag{H2}$$

where $\bar{\mu}_N$ is the mean of an empirical sample of N data. The sample mean, being a linear combination of random variables, is in turn a random variable with an expected value equal to μ and with standard deviation

$$\sigma_N = \frac{1}{\sqrt{N}}\sigma \tag{H3}$$

According to this view, the function g maps the random variable $\bar{\mu}_N$ into the random variable $\tilde{p}$.

If we call P_I the probability that μ belongs to a given interval $I_{\mu_N} = [\mu_1, \mu_2]$, then, in virtue of the mapping, it results that the parameter p has the same probability to belong to the interval $I_{p_N} = [p_{1,N}, p_{2,N}]$, where $p_{1,N}$ and $p_{2,N}$ are, respectively, the smaller and the larger of the images of the endpoints μ_1 and μ_2. Formally it can be written that:

$$P_I = P\left(\mu \in I_{\mu_N}\right) = P\left(p \in I_{p_N}\right) \tag{H4}$$

Since in general the function g is not linear, $\tilde{p}$ is not the midpoint of the interval. It is common practice to provide $\tilde{p}$ as the estimator of p. If one takes as $\mu_1 = \bar{\mu}_N - \sigma_N$ and $\mu_2 = \bar{\mu}_N + \sigma_N$ then the extremes $p_{1,N}$ and $p_{2,N}$ of the image interval I_{p_N} can be considered endpoints of the 1σ interval. We stress that instead of $\tilde{p}$ as given in ($H2$) it would be more correct to provide the midpoint of the image interval and its half-length as the result of the estimation process, that is:

$$\tilde{p}_N = \frac{1}{2}\left(p_{1,N} + p_{2,N}\right) \tag{H5}$$

$$\Delta\tilde{p}_N = \frac{1}{2}\left(p_{2,N} - p_{1,N}\right) \tag{H6}$$

Note that in the above formulas, we assume to know σ that, through ($H3$), would allow us to know σ_N. In practice, however, σ is not known. It could be estimated from the empirical standard deviation. Here we make the choice to estimate it as a function of the estimator $\tilde{p}$ given by ($H2$). More specifically, the procedure we propose to compute the 1σ interval is:

1. Compute $\bar{\mu}_N$ from the N-sample data.
2. Calculate the estimator $\tilde{p}$.
3. Obtain σ through a proper function of $\tilde{p}$, say $\sigma = \sigma(\tilde{p})$.
4. Compute σ_N via ($H3$).
5. calculate the endpoints $p_{1,N}$ and $p_{2,N}$.

1σ intervals for exponential distributions

As an illustrative example of the exponential distributions addressed in this paper, let us consider the discrete untrimmed exponential distribution and write the function g as:

$$\tilde{\alpha} = g(\bar{\mu}_N) = \ln\left(\frac{\bar{\mu}_N + 1}{\bar{\mu}_N}\right) \tag{H7}$$

Then we compute the standard deviation of an N-size sample in terms of the computed $\tilde{\alpha}$:

$$\sigma_N = \frac{1}{\sqrt{N}}\frac{e^{-\frac{\tilde{\alpha}}{2}}}{1 - e^{-\tilde{\alpha}}} \tag{H8}$$

The further step is to compute the endpoints of the interval I_{μ_N} :

$$\bar{\mu}_N \pm \sigma_N = \frac{e^{-\tilde{\alpha}}}{1 - e^{-\tilde{\alpha}}} \pm \frac{1}{\sqrt{N}}\frac{e^{-\frac{\tilde{\alpha}}{2}}}{1 - e^{-\tilde{\alpha}}} = \frac{1}{c - 1}\left(1 \pm \sqrt{\frac{c}{N}}\right) \qquad c = e^{\tilde{\alpha}} \tag{H9}$$

The lower end of the interval I_{p_N} is:

$$p_{1,N} = \ln\left(\frac{\bar{\mu}_N + \sigma_N + 1}{\bar{\mu}_N + \sigma_N}\right) = \ln\left(\frac{\frac{1}{c-1}\left(1+\sqrt{\frac{c}{N}}\right)+1}{\frac{1}{c-1}\left(1+\sqrt{\frac{c}{N}}\right)}\right) = \ln\left(\frac{c+\sqrt{\frac{c}{N}}}{1+\sqrt{\frac{c}{N}}}\right) \qquad c = e^{\tilde{\alpha}} \tag{H10}$$

Likewise, the upper-end results to be:

$$p_{2,N} = \ln\left(\frac{\bar{\mu}_N - \sigma_N + 1}{\bar{\mu}_N - \sigma_N}\right) = \ln\left(\frac{c-\sqrt{\frac{c}{N}}}{1-\sqrt{\frac{c}{N}}}\right) \qquad c = e^{\tilde{\alpha}} \tag{H11}$$

These calculations allow us to compute the 1σ interval for the decay parameter b:

$$b_{1,N} = \frac{p_{1,N}}{2\delta\ \ln(10)} = \frac{1}{2\delta\ \ln(10)}\ln\left(\frac{c+\sqrt{\frac{c}{N}}}{1+\sqrt{\frac{c}{N}}}\right) \tag{H12a}$$

$$b_{2,N} = \frac{p_{2,N}}{2\delta\ \ln(10)} = \frac{1}{2\delta\ \ln(10)}\ln\left(\frac{c-\sqrt{\frac{c}{N}}}{1-\sqrt{\frac{c}{N}}}\right) \tag{H12b}$$

The above formulas are the ones proposed in the main text as eqs (20) and (21).

1σ intervals for the absolute difference distribution

The decay parameter α of the distribution of the absolute values of the differences is linked to the distribution mean through the formula $(F16)$ that therefore provides us with the function g:

$$\tilde{\alpha} = g(\bar{\mu}_N) = \ln\left(\frac{1+\sqrt{1+\bar{\mu}_N^2}}{\bar{\mu}_N}\right) = \operatorname{csch}^{-1}(\bar{\mu}_N) \tag{H13}$$

This formula was derived also by applying the ML method to the distribution of the differences, as noted before, but the standard deviation to use here is the one of the absolute differences shown in $(D7)$, while the formula $(B11)$ is unsuitable and would lead to incorrect evaluations. By using it, we can compute the sample standard deviation as:

$$\sigma_N = \frac{1}{\sqrt{N}}\frac{\sqrt{2e^{-\tilde{\alpha}}\left(1+e^{-2\tilde{\alpha}}\right)}}{\left(1+e^{-\tilde{\alpha}}\right)\left(1-e^{-\tilde{\alpha}}\right)} \tag{H14}$$

The endpoints of the interval I_{μ_N} are:

$$\bar{\mu}_N \pm \sigma_N = \frac{2e^{-\tilde{\alpha}}}{\left(1+e^{-\tilde{\alpha}}\right)\left(1-e^{-\tilde{\alpha}}\right)} \pm \frac{1}{\sqrt{N}}\frac{\sqrt{2e^{-\tilde{\alpha}}\left(1+e^{-2\tilde{\alpha}}\right)}}{\left(1+e^{-\tilde{\alpha}}\right)\left(1-e^{-\tilde{\alpha}}\right)} \tag{H15}$$

After some manipulations, this formula can be given the following version:

$$\bar{\mu}_N \pm \sigma_N = \left(1 \pm \sqrt{\frac{\cosh\tilde{\alpha}}{N}}\right)\operatorname{csch}\tilde{\alpha} \tag{H16}$$

In terms of the absolute difference magnitudes, the endpoints of the 1σ uncertainty interval are:

$$b_{1,N} = \frac{1}{2\delta\ \ln(10)}\ln\left[\frac{2\delta + \sqrt{4\delta^2 + (\operatorname{csch}\tilde{\alpha})^2\left(1+\sqrt{\frac{\cosh\tilde{\alpha}}{N}}\right)^2}}{\left(1+\sqrt{\frac{\cosh\tilde{\alpha}}{N}}\right)\operatorname{csch}\tilde{\alpha}}\right] = \frac{1}{2\delta\ \ln(10)}\operatorname{csch}^{-1}\left(\left(1+\sqrt{\frac{\cosh\tilde{\alpha}}{N}}\right)\operatorname{csch}\tilde{\alpha}\right) \tag{H17a}$$

$$b_{2,N} = \frac{1}{2\delta\ \ln(10)}\ln\left[\frac{2\delta + \sqrt{4\delta^2 + (\operatorname{csch}\tilde{\alpha})^2\left(1-\sqrt{\frac{\cosh\tilde{\alpha}}{N}}\right)^2}}{\left(1-\sqrt{\frac{\cosh\tilde{\alpha}}{N}}\right)\operatorname{csch}\tilde{\alpha}}\right] = \frac{1}{2\delta\ \ln(10)}\operatorname{csch}^{-1}\left(\left(1-\sqrt{\frac{\cosh\tilde{\alpha}}{N}}\right)\operatorname{csch}\tilde{\alpha}\right) \tag{H17b}$$

and the estimated σ is given by:

$$\sigma = \frac{b_{2,N} - b_{1,N}}{2} \tag{H17c}$$

The above formulas are not provided in the main text.

APPENDIX I: SIMULATION OF COMPLETE AND INCOMPLETE MAGNITUDE DATA SETS

To generate a complete random data set of magnitudes $M \geq M_{min}$ with exponential distribution, we use the inverse exponential transformation:

$$M = -\frac{\ln\{\text{rand}\,]0:1[\}}{b\ln(10)} + M_{min} \tag{I1}$$

where rand]0 : 1[is a pseudo-random number with uniform distribution in the interval]0 : 1[.

The binning of magnitudes is obtained by:

$$M_{binned} = \text{round}\left(\frac{M}{2\delta}\right)2\delta \tag{I2}$$

where round(x) indicates the closest integer to the argument value x and 2δ is the binning size. In such a case, in order for the simulated data set to be complete, the latter must include magnitudes down to $M_{min} - \delta$:

$$M = -\frac{\ln\{\text{rand}\,]0:1[\}}{b\ln(10)} + M_{min} - \delta \tag{I3}$$

Therefore, the eq. (I3) is the one adopted to generate all the complete magnitude data sets in the paper. As suggested by Ogata & Katsura (1993), magnitude data incompleteness can be mimicked by a cumulative Gaussian probability distribution with mean μ and standard deviation λ:

$$P(m \leq M|\mu,\lambda) = \frac{1}{\lambda\sqrt{2\pi}}\int_{-\infty}^{M} e^{-\frac{(m-\mu)^2}{2\lambda^2}}\,dm \tag{I4}$$

In this formulation, the mean μ corresponds to the threshold magnitude at which $P = 0.5$, which means that below it, the 50 per cent of earthquakes cannot be correctly evaluated and are lost for the frequency magnitude analyses.

It can be introduced in the simulated data set using the thinning method (Ogata 1981), which consists of discarding the magnitudes for which an extracted random number in the interval]0 : 1[is larger than the cumulative Gaussian probability (I4).

Van der Elst (2021) simulated data sets with time-varying incompleteness as it may be found in the first hours or days after a strong main shock. For modelling such a decaying incompleteness threshold, Helmstetter *et al.* (2006) proposed the empirical equation:

$$m_c(t) = m - 4.5 - 0.75\log_{10}t \tag{I5}$$

where m_c is the time-dependent magnitude threshold of completeness, m is the magnitude of the main shock and t is the time elapsed since the main shock in days. The law (I5) is a decreasing function of time and implies that after a single day the threshold lowers down to $m - 4.5$. Van der Elst suggested using equation (I5) to set the time-varying mean $\mu(t)$ in the eq. (I4), which entails the assumption that m_c is the magnitude below which half of the earthquakes are lost.

In order to simulate the time t of each shock after a main shock, we assumed a simple Omori-Utsu decay law (Utsu 1961) with the equation:

$$r(t) = \frac{K_O}{(t + c_O)^{p_O}} \tag{I6}$$

where $r(t)$ is the time-varying rate (in shocks per day) of a non-homogeneous Poisson process, p_O and c_O are empirical parameters and K_O is a normalization factor depending on the number of shocks and the considered time interval. Usually, $p_O \approx 1$ and c_O is of the order of some tens of minutes (about 0.01 d). The time integration:

$$\tau = \int_0^t r(s)\,ds = F(t) \tag{I7}$$

produces a set of transformed times that follow a stationary Poisson process with intensity 1 (Ogata 1988).

Conversely, given a set of times τ_i generated according to a stationary Poisson process with intensity 1, the inverse integral transformation:

$$t = F^{-1}(\tau) \tag{I8}$$

corresponds to a non-homogeneous Poisson process with rate $r(t)$.

Moreover, it is often useful to generate sequences of exactly N events over a given time interval $[0, T_E]$, i.e.:

$$\int_0^{T_e} r(s)\,ds = N \tag{I9}$$

This implies that:

$$K_O = \begin{Bmatrix} \frac{N(1-p_O)}{(T_E+c_O)^{1-p_O} - c_O{}^{1-p_O}} & p_O \neq 1 \\ \frac{N}{\ln(T_E+c_O)-\ln(c_O)} & p_O = 1 \end{Bmatrix} \tag{I10}$$

Then, the direct timescale transform is:

$$\tau = F(t) = \int_0^t r(s)\,ds = \begin{Bmatrix} N\frac{(t+c_O)^{1-p_O} - c_O{}^{1-p_O}}{(T_E+c_O)^{1-p_O} - c_O{}^{1-p_O}} & p_O \neq 1 \\ \mathrm{N}\frac{\ln(t+c_O)-\ln(c_O)}{\ln(T_E+c_O)-\ln(c_O)} & p_O = 1 \end{Bmatrix} \tag{I11}$$

and the inverse timescale transform is:

$$t = F^{-1}(\tau) = \left\{ \begin{array}{ll} \left[\tau \frac{(T_E+c_O)^{1-p_O} - {c_O}^{1-p_O}}{N} + {c_O}^{1-p_O}\right]^{1/(1-1-p_O)} - c_O & p_O \neq 1 \\ \exp\left[\tau \frac{\ln(T_E+c_O)-\ln(c_O)}{N} + \ln(c_O)\right] - c_O & p_O = 1 \end{array} \right\} \quad \text{(I12)}$$

The set of stationary Poisson times with intensity 1 can be generated by cumulating exponentially distributed interevent times (starting from $\tau_1 = -\ln\{1 - \text{rand}]0:1[\ \}$):

$$\tau_i = \tau_{i-1} - \ln\{1 - \text{rand}\,]0:1[\ \},\quad i = 2,\ N \quad \text{(I13)}$$