REVIEW REPORT FOR

Improving estimates of the b-value in regional catalogs by means of the the b-more positive method

published in *Seismica*

ROUND 1 REVIEWS

REVIEWER 1

See attachment

(See the pages following this compilation of text comments.)

REVIEWER 2

Review of "Improving estimates of the b-value in regional catalogs by means of the the b-more positive method" by Lipiello et al.
The paper compares the classical MLE b-value estimate with the b+ and B-more-positive estimators on both synthetic ETAS generated catalogs, made incomplete for both STAI and SNDI, as well as 5 regional/national catalogues. The paper concludes that the b++ estimators shows the best agreement with the synthetic prior value in the ETAS catalogs, where the other two estimators would underestimate the true prior value.

I find the paper interesting, however, do have some concerns:

1) the "true, prior" estimate seems to be determined by the traditional GR MLE estimate. It retrieves the imposed prior value, however, to what extend do the other estimators also retrieve the prior value for this catalog? Is there a possibility that the b-more-positive here would provide an overestimate? This should be investigated and discussed in the paper.

2) the authors provide no explanation for the observed underestimation of the MLE and b+ estimates. Especially considering that this behavior (the systhematic underestimation of the prior, imposed value) has not been seen in synthetic studies before, a more thourough assessment of this observation and the underlying causes would be in order. I think the firm conclusion requires such an assessment and a proper discussion of possible causes.

3) I think a throrough discussion of the findings of this paper for the regional b-values in comparison to previously published assessments is required. Have the other papers also found the lower b-values compared to the b++ estimate? Are there indications that spatial variability in b-values could cause discrepancies?

4) the cut-off values for dMth obtained seem quite large (>0.8) compared to those used in previous studies using b-positive and b-more-positive (0-0.3). Some discussion on the cause of this (maybe the spatial extend of the catalogues and inhomegeneous character of

Mc over these regions?) is required as well as a discusion on the implications of this for the application to other (regional) catalogues.

5) How sensitive is the b++ estimate to the imposed distance cut-off. To what extend does this, in combination with the imposed sigma of 0.4 for SNDI affect the estimates? Some discussion on this is necessary.

Small comments:

lines 165-168: to me the scheme is not completely clear, especially the summation in step b with respect to equation 10.

line 211: there are two questionmarks in this line. Please correct.

Answers to these reviews are inserted below, following this compilation of text comments.

ROUND 2 REVIEWS

REVIEWER 3

This manuscript benchmarks three b-value estimators, the classical MLE (β_0), the b-positive estimator (β_+; van der Elst, 2021), and the authors' b-more-positive estimator (β_{++}; Lippiello & Petrillo, 2024), using synthetic ETASI catalogs and five instrumental catalogs (Japan, Italy, Southern California, Northern California, New Zealand). The main conclusion is that β_0 is systematically biased low by STAI/SNDI incompleteness, while β_{++} recovers β_true accurately. The manuscript also claims statistically significant regional differences in b from β_{++}, and documents a non-monotonic dependence of β_{++} on δM_th that remains unexplained.
The topic is relevant and the idea is interesting, but in its current form the manuscript has several major scientific and statistical problems that need to be addressed before the conclusions can be supported.

Major issues

1) Internal inconsistency in the core derivation

Your central justification relies on Condition (7), Φ(m_j − M_c) = 1, which is used to obtain a pure exponential p(δm) with coefficient β. However, with your own Φ definition (Eq. 3), Φ never equals 1 for finite magnitudes: Φ(0)=0.5 and even Φ(2σ)≈0.98. The manuscript treats the condition as "satisfied" once δM_th ≳ 2σ, but you never quantify the residual bias introduced by Φ<1, nor propagate it into β_{++}.

What I need to see to accept the derivation: rewrite (7) as an inequality (Φ ≥ 1−ε), define ε explicitly, map δM_th and σ to ε numerically, and show that the induced bias in β_{++} is negligible for the δM_th values actually used.

2) Self-validation on a synthetic benchmark designed to match the estimator's assumptions

β_{++} comes from Lippiello & Petrillo (2024). The ETASI simulation framework is from the same group (Petrillo & Lippiello 2021, 2023), and it embeds exactly the incompleteness hypotheses (i) and (ii) that β_{++} is constructed to satisfy. Testing β_{++} on ETASI is therefore a consistency check, not independent validation. In addition, your baseline completeness estimators (MAXC, MBS, CV) are known to behave poorly under the kind of incompleteness you impose, which makes them a weak reference for "accuracy".

What I need to see: validation under at least one incompleteness model outside ETASI (e.g., log-linear time decay of M_c as in Helmstetter et al. 2006, or purely SNDI-driven incompleteness), and comparison against at least one robust alternative estimator not developed within the same framework.

3) Pair-selection rules are not shown to enforce hypotheses (i) and (ii) in real catalogs

The method claims to control completeness indirectly via pair selection (d_R and the "no larger intervening event" condition). In real catalogs, M_c is unobserved and varies in time and space. The current evidence that the rules avoid STAI/SNDI bias is circular: β_{++} appears stable, therefore the selection must have worked. That is not a scientific demonstration.

What I need to see:

- On synthetics: ablation tests separating STAI-only, SNDI-only, and combined incompleteness, showing which mechanism drives the bias in β_0/β_+ and why β_{++} avoids it.

- On real catalogs: diagnostics of the selected pairs—distributions of ($t_j - t_i$), d_{ij}, time since last larger event, and at least one proxy for completeness variation (e.g., station density / network evolution if available)—demonstrating that the selected j events occupy demonstrably higher-completeness conditions.

4) The δM_th selection rule (β_{++}*) is ad hoc and not statistically justified

The "best estimate" β_{++}* is defined using an MBS-style plateau criterion on β_{++}(δM_th). MBS was developed for a monotonic β_0(M_th) curve, while your β_{++}(δM_th) curves in real catalogs are explicitly non-monotonic. In that situation, the "plateau" can simply be the rising limb before the subsequent drop. Also, your rule depends on the averaging window and step size, but sensitivity is not explored. Finally, the uncertainty reported ($\delta\beta \approx \beta/\sqrt{N_{++}}$) does not include uncertainty from choosing δM_th.

What I need to see: sensitivity of β_{++}* to step size and window length, and uncertainty estimates that propagate the δM_th selection uncertainty. Preferably, replace the heuristic plateau rule with a likelihood-based or explicit bias–variance criterion tied to your $\Phi \geq 1-\varepsilon$ threshold.

5) "True β" language is not justified for instrumental catalogs

In synthetics, β_true is the simulation input; that is valid. In instrumental catalogs there is no external ground truth, yet the manuscript repeatedly states that β_{++} "recovers the true β" for Japan/Italy/etc. That is not a defensible statement given the current evidence.

Required change: in all instrumental-catalog discussion, replace "true β" language with a careful statement such as "β_{++} is less sensitive to incompleteness under the stated assumptions and yields systematically higher β than threshold-based estimators."

6) The hazard-direction statement is not supported as written

The conclusion states that biased-low b-values "may have led to systematic overestimation of seismic hazard." This is not established by the paper. The sign of hazard change is not determined by b alone; it depends on how the model is normalized, the magnitude range, M_max assumptions, and the return period of interest. As written, the statement is misleading and too strong.

Required change: remove the hazard-direction claim or support it with a demonstrator hazard/rate calculation that explicitly shows the sign under a stated PSHA parameterization. At minimum, restrict to "implications for seismicity rate extrapolation and parameter estimation."

7) The significance test for non-monotonic β_{++} is statistically invalid

The z-test compares β_1 (the maximum) and β_2 (a subsequent minimum) selected from the same $\beta_{++}(\delta M\_th)$ curve. This creates strong selection bias: the distribution of a max-minus-min statistic is not Gaussian with (σ_1,σ_2), and the curve points are not independent because thresholds are nested and share events. The reported confidence levels are therefore inflated.

What I need to see: a permutation/Monte Carlo-based null distribution for the same max–min statistic under a realistic null model (i.i.d. magnitudes + incompleteness), or else treat the non-monotonicity descriptively without formal significance claims.

8) Uncertainty quantification is inadequate for β_{++}

Pairs used by β_{++} are conditionally selected and can be dependent (triggering/clustering, shared events across thresholds). The variance formula $\delta\beta \approx \beta/\sqrt{N_{++}}$ assumes i.i.d.-like behavior and may substantially understate uncertainty.

What I need to see: uncertainty validated with an appropriate bootstrap (e.g., block bootstrap in time, or resampling independent sequences), plus reporting of how many distinct events actually contribute to β_{++}.

9) Declustering sensitivity is not tested

Hypothesis (i) is most likely to be violated in strongly clustered sequences (aftershock-rich catalogs), yet all instrumental catalogs are analyzed "as-is" with no sensitivity check.

What I need to see: repeat at least two catalogs with a standard declustering approach and report whether β_{++} and β_{++}^* materially change.

Minor issues

1. Reproducibility details are incomplete: magnitude type and homogenization, exact spatial polygons, treatment of blasts/induced events, and temporal changes in completeness/network evolution need clearer documentation.
2. The rounding correction (jitter) is shown for β_+ but not explicitly validated for β_{++}. Please demonstrate seed stability (e.g., 20–50 jitter realizations) and show that the additional variance is negligible compared with catalog uncertainty.
3. β vs b inconsistency: Table values are β (= b·ln10), but the text repeatedly says "b-value." This is confusing and will mislead many readers. Use one convention consistently or always state the conversion.
4. Several statements are too categorical (e.g., "traditional estimators systematically underestimate," "β_{++} yields highly accurate results"). These need to be qualified as "under the tested incompleteness model" / "in the studied catalogs" / "under stated assumptions."
5. Numerous typos and grammar issues (e.g., "intrumental," "New Zeland," "containse," "concerncs," "stnthetic," etc.) should be corrected before resubmission.
6. http://www.declustering.com/ seems to be slow and it is not loading (add dataset to zenodo)
7. Finally, please report b-values (base-10) rather than β throughout (or at least report both).
   At present the manuscript primarily reports β = b·ln(10) (e.g., Table 1), while the narrative repeatedly discusses "b-values." This is unnecessarily confusing for most readers and makes the reported ranges look "too large" unless the reader mentally converts (β≈2.3 corresponds to b≈1.0). I strongly recommend reporting b as the primary quantity in the main text, figures, and tables, and (if you want to retain β for derivations) include β only secondarily or in parentheses. At minimum, every table/figure should state explicitly whether it reports b or β, and provide the conversion used.

<u>Answers</u> to these reviews are inserted below, following this compilation of text comments.

ROUND 3 REVIEWS

<u>HANDLING EDITOR</u>

Based on reviews I have received, your manuscript is suitable for publication. I will not send it for further reviews, yes I have two requests to you in frame of a micro-revision.

After almost a year and several reviews, I find that the reviews+answers became a very important part of the discussion, even if there is no full agreement. Since in *Seismica* these reviewes+answers are published together with the paper, I would like to ask you to insert a sentence somewhere in the discussions or conclusions that further thoughts or elements of reasoning can be found in the reviews+answers document. This would be my first request.

The second request would be to include a generic sentence regarding the limitations of the methods (not only yours, but also some others in the literature). In essence, I get the impression that some methods (including b+ and b++) work only with large catalogue and perfect data, in perform less well in real applications. So if studies are conducted with methods developed for large catalogues, but on actually small datasets, it can result in some mis-interpretations. Adding a though on this regard to the manuscript, at a place convenient to you, would be appreciated.

Answers to these reviews are inserted below, following this compilation of text comments.

The paper is scarcely significant and not particularly well written.

In particular, the conclusions are rather poor and not particularly new: many papers (some of which are cited by authors) already demonstrated that traditional methods tend to underestimate the b-value. On the other hand, the only really new finding that the b++ tends to decrease at high δM_{th} is not particularly supported by statistically significant observations. Some other technical issues are noted below.

The authors completely omit the discussion on the effects of the binning to one decimal digit of most of the datasets used. As they well know, the binning produces an underestimation of the b-value, which can be approximately corrected by subtracting one half of the binning size (e.g. 0.05) from the minimum magnitude Mc or using other more appropriate formulations.

I cannot understand how in eq. (7) can be $\Phi\left(m_j - M_c\left(t_j, \vec{x}_j, \mathcal{H}_j \middle| m_i\right)\right)$ = 1 for $m_j < +\infty$. Based on the definition of eq. (3), $\Phi(0) = 1/2$. Such definition differs from that given in eq. (2) of Lippiello and Petrillo (2024) which gives $\Phi(0) = 1$.

This point makes the successive discussion almost completely meaningless. However, for the following comments I will assume that the correct definition of Φ is that given by Lippiello and Petrillo (2024).

At lines 166 to 168, what the algorithm does if the last condition (c) is not met?

To make the result reproducible, the authors must report a table with the values of parameters σ, Mc, and d_R for all the 15 analyzed dataset

I suggest reporting in table 1 even the result for the classical beta+ methods

Check the use of symbol “?” at line 211

Give some details (or reference) on the methods used to simulate catalogs

The variations (increase/decrease) described for various methods are mostly well within the error bounds and then they are substantially insignificant from the statistical point of view.

Answers point-to-point - 2025JH000863

September 9, 2025

1 Report of Reviewer #1

The paper is scarcely significant and not particularly well written. In particular, the conclusions are rather poor and not particularly new: many papers (some of which are cited by authors) already demonstrated that traditional methods tend to underestimate the b-value. On the other hand, the only really new finding that the b++ tends to decrease at high dMth is not particularly supported by statistically significant observations. Some other technical issues are noted below.

Answer: We are very sorry that Reviewer 1 judged our paper to be scarcely significant and not particularly well written. In the revised version of the manuscript, and following the suggestions of both reviewers, we have substantially improved the presentation of our results as well as the overall readability of the paper.
On the other hand, we must respectfully disagree with Reviewer 1's state-ment that the paper is "rather poor and not particularly new." We acknowledge that many previous studies have already shown that traditional methods tend to underestimate the b-value. However, the general consensus has been that by adopting a sufficiently large threshold magnitude M_{th}, the correct b- value can eventually be recovered.
In contrast, our manuscript demonstrates that all traditional methods - including the most conservative ones that employ large M_{th} values - systematically fail to provide the correct b-value. This is clearly highlighted both by numerical simulations using incomplete ETAS catalogs and by systematic analyses of five regional catalogs with the comparison with the b-more-positive estimator. From our perspective, this represents a truly novel result, showing that b-values used in forecasting studies prior to our work are significantly underestimated, thereby introducing a systematic bias into past seismic forecasting efforts.
A striking example is provided by the Italian catalog, which is typically associated with a b-value close to 1 (Lombardi, A.M. & Marzocchi, W. (2010). *The ETAS model for dailt forecasting of Italian seismicity in the CSEP experiment.* Annals of Geophysics, 53(3), 155-164). In contrast, the b-more-positive method yields a value that is approximately 15% larger (as shown in Fig. 2). We consider this a surprising and important finding, which significantly impact forecasting

in Italy.
Another aspect that was not sufficiently emphasized in the previous version of the manuscript, but which we now stress more clearly, concerns the variability of the b-value across different regions. Our new approach, and in particular the reduction of uncertainty in the estimation of the b-value, allows us to demonstrate that the common assumption of a universal b-value does not hold. Instead, the b-value varies significantly among regions, and this, again, has important implications for seismic hazard assessment.
We have clarified and reinforced the relevance of our findings and their implications for seismic forecasting in both the Abstract and the Conclusions of the revised manuscript.
Furthermore, regarding the last observation that β_{++} tends to decrease at high δM_{th}, we agree with Reviewer 1 that this trend was not strongly supported by statistically significant evidence in the previous version. In the revised ver-sion, we now perform more rigorous statistical tests, which demonstrate that for at least 3 out of the 5 catalogs the decrease is significant at the 95% confi-dence level. The details can be found in the specific point of our point-by-point response below.
We trust that the comprehensive responses provided here, together with the corresponding improvements made to the manuscript, will help convince Reviewer 1 of the significance and relevance of our study.

1) **The authors completely omit the discussion on the effects of the**

binning to one decimal digit of most of the datasets used. As they well know, the binning produces an underestimation of the b-value, which can be approximately corrected by subtracting one half of the binning size (e.g. 0.05) from the minimum magnitude Mc or using other more appropriate formulations.

Answer: We completely agree with Reviewer 1 that the effect of bin-ning is a crucial point and, if not properly taken into account, can lead to a systematic bias in the b-value estimate. This problem does not occur in syn-thetic catalogs where magnitudes are not rounded. In instrumental catalogs we have already explicitly accounted for the effect of rounding by adding a small random perturbation, uniformly distributed in the interval $(-0.5\Delta, 0.5\Delta]$, to each magnitude in the catalog, where $\Delta = 0.1$.
We regret that this correction was not mentioned in the previous version of the manuscript and we thank Reviewer 1 for bringing this issue to our atten-tion. This approach had already been tested in our previous study (Lippiello & Petrillo, JGR 2024), but Reviewer 1's comment prompted us to dedicate an Appendix to this important aspect. In the revised manuscript, we show that our method provides an accurate estimate of the b-value, while avoiding some technical problems that can arise when applying the exact analytical expression proposed by Tinti and Gasperini, and previously by Van der Elst. These issues are specifically illustrated and discussed in the new Appendix A.

2) I cannot understand how in eq.(7) can be $\phi(m_j - M_c(t_j, x_j, \mathcal{H}_t|m_i)) = 1$ for $m_j < +\infty$. Based on the definition of eq. (3), $\phi(0) = 1/2$. Such definition differs from that given in eq. (2) of Lippiello and Petrillo (2024) which gives $\phi(0) = 1$. This point makes the successive discussion almost completely meaningless. However, for the following comments I will assume that the correct definition of ϕ is that given by Lippiello and Petrillo (2024)

Answer: Reviewer 1 is correct in pointing out that Eq. (7) holds exactly only in the limit $m_j \to \infty$. Nevertheless, due to the very rapid saturation of the error function, the approximation becomes extremely accurate as soon as $m_j - M_c > 2\sigma$. For example, the error function $\mathrm{erf}(x)$ increases from 0.9953 to 0.99998 when x varies from 2 to 3.
Accordingly, with very high accuracy one can neglect the dependence on δm in Eq. (6), obtaining

$$p(\delta m) \simeq \exp(-\beta\, \delta m),$$

from which it is possible to extract the correct value of β with a high accuracy. We emphasize that the fast saturation $\Phi(x) \simeq 1$ is the key feature that makes the b-positive method effective even in the presence of incomplete data.
In the revised version of the manuscript, we clarify that Eq. (7) is satisfied with very high accuracy, though not exactly.

3) At lines 166 to 168, what the algorithm does if the last condition (c) is not met?

Answer: We thank Reviewer 1 for this question. Condition (c) states: If $m_i < m_j < m_i + \delta M_{\mathrm{th}}$, then move to the subsequent i. Therefore, if condition (c) is not satisfied, the pair (i, j) is not considered in the β_{++} estimation, i.e., it is excluded from the sum $\sum(m_j - m_i)$. We have clarified this point in the revised manuscript (see lines 172-179 of the tracked manuscript).

4) To make the result reproducible, the authors must report a table with the values of parameters σ, M_c and d_R for all the 15 analyzed dataset

Answer: We use the same value of $\sigma = 0.4$ to generate all 10 synthetic ETASI catalogs. Conversely, as detailed in Sec.4.1, a unique M_c is not adopted for the synthetic catalogs. To implement realistic incompleteness, we use the spatially dependent M_c values obtained by Schorlemmer and Woessner (2008) for all synthetic catalogs. We also note that the time-dependence of M_c is implemented by considering a constant blind time, which is the same for all 10 synthetic catalogs.
Concerning d_R, this is a parameter of the b-more-positive algorithm, and we use $d_R = 40$ km for both the synthetic and the regional catalogs. This information was missing in the previous version of the manuscript and has been added to

the caption of Table 1 in the revised version. We have also verified that different choices of $d_R \in [15, 90]$ km leads to similar results (see Novel Fig. App.B)

5) I suggest reporting in table 1 even the result for the classical beta+ methods

Answer: We have followed the suggestion of Reviewer 1 and added a column with the β_+ values in Table 1.

6) Check the use of symbol "?" at line 211

Answer: We thank the reviewer for spotting this. The "?" symbol was due to a misformatted reference in the manuscript. It was intended to cite Tinti and Mulargia (1987) - Confidence intervals of b values for grouped magnitudes. Bulletin of the Seismological Society of America, 77(6), 2125–2134 and also Tinti and Gasperini (2024). These references has now been correctly inserted in the revised manuscript.

7) Give some details (or reference) on the methods used to simulate catalogs

Answer: In the revised manuscript, we have expanded the description of the synthetic catalog generation procedure. The simulations are based on an ETAS branching process, implemented using a generation tree algorithm. The process starts with background events (zero-th generation) occurring as a Poisson process. Each event (mother) generates a number of offspring events according to a Poisson distribution with mean productivity depending on the mother magnitude and occurrence time. Offspring magnitudes follow a Gutenberg-Richter law, and occurrence times are drawn from the Omori law probability density. Only events falling within the observational window are retained. This procedure is iterated recursively for successive generations until no further events fall within the simulation period. Changes at lines 187-197 of the tracked version of the manuscript.

8) The variations (increase/decrease) described for various methods are mostly well within the error bounds and then they are substantially insignificant from the statistical point of view.

Answer: We thank Reviewer 1 for this comment. In the revised version of the manuscript, we applied the z-test to assess whether the difference between the maximum value of β_{++} estimated at small δM_{th} and the subsequently lower values observed at larger δM_{th} is statistically significant. We find that, for three out of the five regions, this difference is statistically significant at the 95% confidence level. A new Table 2 has been added in the revised manuscript to present and discuss these results.

2 Report of Reviewer #2

The paper compares the classical MLE b-value estimate with the b+ and B-more-positive estimators on both synthetic ETAS generated catalogs, made incomplete for both STAI and SNDI, as well as 5 regional/national catalogues. The paper concludes that the b++ estimators shows the best agreement with the synthetic prior value in the ETAS catalogs, where the other two estimators would underestimate the true prior value. I find t he p aper i nteresting, h owever, do have some concerns:

Answer: We thank Reviewer 2 for the positive assessment and for summarizing the core scope of our work. We are pleased that the study is considered interesting and have revised the manuscript to improve clarity in the methodology and to better highlight the significance of our results. Below we address the reviewer specific concerns in detail.

1) the "true, prior" estimate seems to be determined by the traditional GR MLE estimate. It retrieves the imposed prior value, however, to what extend do the other estimators also retrieve the prior value for this catalog? Is there a possibility that the b-more-positive here would provide an overestimate? This should be investigated and discussed in the paper.

Answer: We believe that Reviewer 2 refers to the estimate of β in the synthetic ETASI catalog. In our synthetic tests, the "true prior" β is the exact value imposed in the ETAS model used to generate the catalogs, as reported in Table 1.
In the left panels of Fig. 1, we also plot (blue curve) the β value obtained using the traditional method on the ETAS catalog before introducing incompleteness. This value, aside from statistical fluctuations, coincides with the "true prior" β, as expected.
As shown in the left panel of Fig. 1, the estimate of the b -value via β_{++} converges to the "true prior" β as δM_{th} increases, consistently with theoretical expectations, and there is no indication of overestimation of the b-value. This conclusion is further supported by the systematic analysis over 10 synthetic ETASI catalogs illustrated in Table 1, which shows that the b-more-positive method (column 6) provides a β value in very good agreement with the "true prior" $\beta_{true} = 2.73$ in all catalogs.

2) the authors provide no explanation for the observed underestimation of the MLE and b+ estimates. Especially considering that this behavior (the systematic underestimation of the prior, imposed value) has not been seen in synthetic studies before, a more thourough assessment of this observation and the underlying causes would be in order. I think the firm conclusion requires such an assessment and a

proper discussion of possible causes.

Answer: We thank Reviewer 2 for this comment. The underestimation of the true prior b-value by the MLE and β_+ estimators in our synthetic tests is a direct consequence of the incompleteness deliberately imposed on the catalogs via the STAI and SNDI models.
This systematic bias has not been highlighted before, because to our knowledge no previous study has focused on synthetic regional catalogs implementing both STAI and SNDI.
The mechanism responsible for the underestimation is twofold. First, due to STAI, the completeness magnitude is very large, since even very large earthquakes can remain undetected in the aftermath of even larger events. At the same time, because of SNDI, a spatially dependent static M_c exists. When M_{th} is set larger than the maximum static M_c, the slope of the β value versus M_{th} (see inset in the left panel of Fig. 1) appears to decrease, approaching a plateau. Common estimators, such as Maximum Curvature or MBS, interpret this decrease as a plateau and associate to it the corresponding β value.
Nevertheless, as noted above, because of STAI this is not yet the correct value; only larger M_{th} values can provide the true b-value.
We have expanded the discussion in the revised manuscript to clarify this mechanism and to highlight its consistency with the underlying theory and with prior results obtained for incomplete datasets. Change at lines 237-242 of the tracked version of the manuscript.

3) I think a throrough discussion of the findings of this paper for the regional b-values in comparison to previously published assessments is required. Have the other papers also found the lower b-values compared to the b++ estimate? Are there indications that spatial variability in b-values could cause discrepancies?

Answer: We are not aware of previous studies applying the b-more-positive or b-positive estimators to regional catalogs. In the presence of a spatially varying b-value, the measurement of a single b becomes less meaningful, as there is no unique value to estimate.
In many forecasting studies, a single b-value is typically adopted for the entire region, but we agree with Reviewer 2 that spatial variations in b can be expected. A more rigorous analysis would restrict the study to smaller regions with a more homogeneous tectonic setting; however, this reduces the dataset and may increase statistical uncertainty.
Nevertheless, we consider this a valuable suggestion for future studies. We comment on this point in the Conclusions of the revised version of the manuscript.

4) the cut-off values for d Mth obtained seem quite large (> 0.8) compared to those used in previous studies using b-positive and b-more-positive (0-0.3). Some discussion on the cause of this (maybe the spatial extend of the catalogues and inhomegeneous character of

Mc over these regions?) is required as well as a discusion on the implications of this for the application to other (regional) catalogues.

Answer: Reviewer 2 is perfectly right, and we thank her for this observation, which was indeed not addressed in the previous version of the manuscript. Previous studies usually focus on the estimation of the b-value within single aftershock sequences. In this case, as extensively illus-trated in Lippiello & Petrillo (2024), it is sufficient to have $dM_{\mathrm{th}} > \sigma$ to recover the correct b-value, where σ is the quantity appearing in Eq. (3). The mech-anisms responsible for σ are typically diurnal and seasonal variations, staffing changes, or more general noise, which generally leads to $\sigma \sim 0.3$.
In our study, however, we consider regional catalogs where M_c also varies spatially. The constraint on d_R reduces the impact of spatial variations in M_c, but does not eliminate it. Combining the two effects, a larger $d\ M_{\mathrm{th}}$ is necessary. This is evident in the left panel of Fig. 1: even with $\sigma = 0.4$, the saturation of β_{++} is achieved only for $\delta M_{\mathrm{th}} = 0.8$, as correctly observed by Reviewer 2. In contrast, in previous studies (Lippiello & Petrillo, 2024), when $\sigma = 0.4$ but spatial heterogeneity of M_c was not included, $\delta M_{\mathrm{th}} = 0.4$ was sufficient.
We comment on this point in the revised version of the manuscript.

5) How sensitive is the b++ estimate to the imposed distance cut-off. To what extend does this, in combination with the imposed sigma of 0.4 for SNDI affect the estimates? Some discussion on this is necessary.

Answer: In Appendix B of the revised manuscript, we present results obtained from regional catalogs by considering different values of $d_R \in [15, 500)$ km. We find that the results depend only weakly on d_R for $d_R < 100$ km, whereas significant differences are observed for larger values. This finding is consistent with our theoretical expectations, as it is reasonable that for dis-tances smaller than 100 km the completeness magnitude remains similar, while at larger distances this is no longer the case. We have explicitly added this discussion in the Appendix B of the revised manuscript.
Concerning the effect of σ on the b-value estimate with the b-positive estima-tors, this aspect was already discussed in the previous manuscript (Lippiello & Petrillo, 2024). As expected, the larger the value of σ, the larger the δM_{th} required to recover the correct b-value. In the revised manuscript, we now comment more clearly on this point.

lines 165-168: to me the scheme is not completely clear, especially the summation in step b with respect to equation 10.

Answer: We thank the referee for pointing out that the original description of the algorithm (lines 165–168) was not fully clear, especially in step (b) and its relation to Eq.(10). We have revised the text to provide a clearer and more explicit description of the procedure.

line 211: there are two question marks in this line. Please correct.

Answer: We thank the reviewer for pointing this out. The two ? symbols in line 211 were caused by misformatted citations. They should have referred to Tinti and Mulargia (1987), Confidence intervals of b values for grouped magnitudes, Bulletin of the Seismological Society of America, 77(6), 2125–2134, and to Tinti and Gasperini (2024). These citations have now been correctly formatted and appear properly in the revised manuscript.

Dear Editor,

Thank you very much for careful handling of our manuscript, “Improving estimates of the b-value in regional catalogs by means of the b-more-positive method.” We are grateful for the time and effort invested by you and the reviewers in evaluating our work.

Please find enclosed the revised version of the manuscript, together with:

- a clean version of the revised manuscript,
- a marked-up version highlighting all changes,
- and a detailed point-by-point response to the reviewers.

We have carefully revised the manuscript in light of the new review and, more generally, we took this opportunity to improve the paper substantially. In particular, the revised version includes:

- a clearer treatment of the role of the detection function and of the finite-\sigma bias, including a quantitative discussion.
- a more careful and qualified interpretation of the results obtained on instrumental catalogs, avoiding any reference to a “true” value in real data;
- a revised discussion of the implications for hazard, now formulated in a more cautious and accurate way;
- an expanded presentation of the data, methods, and reproducibility details, including the treatment of catalog rounding, magnitude conventions, and data/code availability;
- additional clarification of the behavior of \beta_{++} as a function of \delta M_{th}, including a revised statistical discussion in the Appendix;
- correction of typographical errors, notation inconsistencies, and language issues throughout the manuscript.

We hope that the revised version satisfactorily addresses the reviewer’s concerns and that the manuscript may now be considered suitable for publication in *Seismica*. Of course, we remain fully available should further clarification or revision be needed.

Thank you again for your time and consideration.

With kind regards,

Cataldo Godano
on behalf of all authors

Answers point-to-point - R2

March 19, 2026

- Reviewer questions are reported in **black**.
- Author answers are reported in red.

1 Report of Reviewer #1

The modifications made in response to my criticisms to the first version of the manuscript only marginally improve its clarity, while other questions remain unanswered or inadequately answered.

Answer: We regret that the Reviewer did not fully appreciate the novelty and the significance of the work presented in this manuscript. In the following sections, we provide a detailed, point-by-point response to the additional criticisms raised, clarifying the original contributions of our study.

1a. To justify the novelty of their work, the authors assert that in previous studies "the general consensus has been that by adopting a sufficiently large threshold magnitude Mth, the correct b-value can eventually be recovered". The authors should specify which studies explicitly report this claim and concerning with which methods.

Answer: We thank the Reviewer for this comment, which allows us to clarify the theoretical context of our statement. The assumption that a correct b-value can be recovered by choosing a sufficiently large M_{th} is not merely an isolated claim, but the fundamental premise underlying almost all b-value estimation studies since the introduction of the Maximum Likelihood Estimator (Aki, 1965; Bender, 1983). In standard practice, researchers identify M_c (the completeness magnitude) and assume that for any $M_{th} \geq M_c$, the frequency-magnitude distribution perfectly follows an exponential law. This implies that the b-value estimate should remain stable (a 'plateau') for increasing M_{th}. This approach is explicitly reported and utilized in all the historical works that measure the b-value, included the papers (Petruccelli et al., 2018, https://doi.org/10.1785/0120170328, Spada et al., 2013, https://doi.org/10.1029/2012GL054198) suggested by the reviewer. Here we demonstrate that in the presence of combined STAI and SNDI, this 'plateau' is illusory: the b-value continues to increase with M_{th} without ever stabilizing at the true value. To the best of our knowledge this kind of analysis as never been performed before on synthetic catalogs and not having a benchmark represented by the novel methodology of the b-more-positive, was not possible to be enlightened before in instrumental data sets. Nevertheless, we would like to emphasize that the primary novelty of our study does not merely lie in identifying the saturation issues of the traditional b-value estimation method. Rather, it resides in the introduction and validation of the b-more-positive method as a significantly more accurate and effective tool for handling the simultaneous presence of STAI and SNDI. This is the first time that the b-more-positive approach has been applied to estimate b-values in regional seismic catalogs. This application represents a fundamental shift in methodology: while traditional approaches struggle with completeness fluctuations, our method demonstrates superior resilience and precision in these challenging contexts. We have updated the Introduction to clarify that the idea that the correct b-value is recovered at sufficiently large Mth is the fundamental premise underlying almost all b-value estimation based on Eq.(2). We have also cited Petruccelli et al, 2018 and Spada et al., 2013 in the Introduction of the revised version of the manuscript.

2-3. In particular, this claim does not apply to most recent studies using magnitude differences (e.g., van der Elst, 2021; Tinti and Gasperini, 2024). In my opinion, the authors should only focus on comparing their method b++ with the b+ by van der Elst (2021) since the clear superiority of the latter over classic methods has already been proved by recent literature. The demonstration that b-value may vary significantly across different regions and depth is also not particularly new (e.g., Petruccelli et al., 2018, https://doi.org/10.1785/0120170328, Spada et al., 2013, https://doi.org/10.1029/2012GL054198)

Answer: We wish to observe that the practice of using a fixed $M_{th} \geq M_c$ threshold is very active even in the very recent literature (e.g., studies published in September and November 2025: Piegari (2025), Li 2025 (see ref below)) which continues to rely on this traditional approach, assuming that staying above M_c guarantees an unbiased estimate. We believe that maintaining a direct comparison with the traditional method is essential: it provides a familiar baseline for the seismological community and effectively highlights the specific advantages and superior performance of the b-more-positive approach. Furthermore, we respectfully address the Reviewer's point regarding the superiority of the b-positive method compared to the traditional one. While it is clearly true in presence of STAI, its performance on regional catalogs with significant spatial M_c variability (such as in Southern California) is not inherently superior. In these cases, without a spatial constraint (like our d_r parameter), the standard b-positive method can be as biased as—or only slightly better than—the traditional MLE. This is also evident from ETASI simulations (Fig.1a) where β_+

behaves similarly to the traditional estimator. This specific limitation is exactly what motivates our study: it justifies the inclusion of the traditional method for context and, more importantly, demonstrates why the *b*-more-positive method is a necessary evolution for accurate regional stress estimation.

- Piegari, E., Corrado, P., Herrmann, M., & Marzocchi, W. (2025). Structural heterogeneities and spatial variations of seismicity drive temporal b-value changes. Geophysical Research Letters, 52, e2025GL116118. https://doi.org/10.1029/2025GL116118
- Li, C.; Li, Z.; Duan, M.; Zhou, L. Spatiotemporal Distribution of the Magnitude of Completeness and b-Values in Mainland China Based on a Fused Multi-Source Earthquake Catalog. Entropy 2025, 27, 1137. https://doi.org/10.3390/e27111137

4. Regarding the question of the rounding to the first decimal of real catalog, I can object that the issues against the use of the formulas of van der Elst (2021) and Tinti and Gasperini (2024) discussed in Appendix A are rather weak as there is no particular reason to assume a non-integer k in the expression of dMth (and apparently they do not) and to perform calculations using single-precision floating-point arithmetic rather than double-precision. On the other hand, the methods of adding a uniform perturbation in the interval from -0.05 to 0.05 arbitrarily modifies the magnitudes compared to the observed ones and it is not entirely correct because, due to the Gutenberg-Richter law, there should be a slight overshoot of negative differences compared to positive ones. I understand that such approach does not greatly affect the results, but I do not understand why it should be preferred over the exact formula for binned data.

Answer: First, we would like to clarify that we fully recognize the validity and rigor of the analytical solutions proposed by Van der Elst (2021) and Tinti & Gasperini (2024) regarding the binning correction. Here we propose an alternative numerical approach that, while less formal, offers the advantage of being straightforward to implement while yielding equivalent results. In fact, we have verified that all the curves presented in this study remain unchanged regardless of whether the analytical formula or our numerical correction is applied. Therefore, the comparison between these two methods is a secondary aspect of our work, which we have included in the Appendix as supplementary information. We believe this discussion is valuable for the community because it highlights practical implementation challenges. For instance, we noted that failing to use double precision in these calculations can lead to significant biases. By sharing this observation in the Appendix, we aim to provide a practical 'best practice' to help other researchers avoid systematic errors in *b*-value estimation.

5a. Regarding the discussion about eq. (7), the changes made do not solve the problem. First, in eq. (7) the equality sign should be

replaced by $\simeq$. Furthermore, the so-called "fast saturation" of the error function is not fast enough if x for example is < 2. Since s is assumed to be 0.4, for $x > 2$ one would need dMth> 0.8 magnitude units, which is rather higher than that used by b+ (usually 0.1). This point should be clearly highlighted in the manuscript because such a threshold strongly reduces the size of the dataset and thus increases the error of the b-value.

Answer: In the revised version of the manuscript, we address this point in detail, specifically regarding the necessary adjustments when $\sigma \neq 0$. We have numerically evaluated the correction term ϵ that accounts for a finite $\delta M_{th}/\sigma$ ratio. Consequently, the revised manuscript now features a generalized expression for β_{++} that incorporates this correction, supported by a comprehensive discussion in the new Appendix A.

5b. Furthermore, it is not clear which dMth the authors use for method b+. In any case they should also show the results of the b+ method using dMth=0.8. Tinti and Gasperini (2024) demonstrated, by a statistical test, that using such a large dMth, the b+ method well reproduces the theoretical b-value even for strongly incomplete datasets with both STAI and SNDI.

Answer: We are somewhat surprised by this objection, as our entire methodology is applied across a wide range of threshold magnitudes, $\delta M_{th} \in [0, 2]$, with increments of 0.1. In Figures 1, 2, and 3, we present the results as a function of N_{th} (the number of events), which is itself a direct function of M_{th} for β evaluation and of δM_{th} for β_+ and β_{++} evaluation. This allows us to express different quantities such as β and β_+ as function of the same variable N_{th}. To make this relationship explicit, the inset in Figure 1 specifically shows the β curve as a function of M_{th}, demonstrating that varying N_{th} or M_{th} yields the same physical insights. In the revised version we better specify that plotting vs N_{th} or vs δM_{th} is a fully equivalent representation. We remark that the specific value of $\delta M_{th} = 0.8$, mentioned by the Reviewer, is well within the range we considered. However, as clearly seen in the synthetic catalog (Figure 1A), at $\delta M_{th} = 0.8$ the estimated β with the b-positive method β_+ is still significantly far from the true value. We would like to take this opportunity to emphasize the key point of our manuscript: the primary limitation of the standard b-positive method in regional contexts is not the specific choice of the magnitude threshold (δM_{th}), but rather the absence of a spatial constraint (d_r). From our assessment, the numerical study by Tinti and Gasperini (2024) does not account for spatial coordinates. The SNDI in that study is introduced assuming a unique, uniform detection threshold, rather than a threshold that varies in space, as is more realistically implemented in our numerical study. It is evident that in such a context, the spatial constraint d_r becomes meaningless, rendering the b-positive and b-more-positive approaches practically equivalent. The agreement between β_+ and the theoretical value observed by Tinti and

Gasperini, setting $\delta M_{th} = 0.8$, is expected given their use of a uniform M_c. Had they introduced an M_c that varies in space, this agreement would likely not have been recovered. Furthermore, as clarified in our revised version, the critical factor is not the absolute value of δM_{th}, but the ratio $\delta M_{th}/\sigma$. In their analysis, Tinti and Gasperini use $\sigma = 0.2$ (λ in their notation) and find good agreement with $\delta M_{th} = 0.8$. However, if a larger value of σ were used, the same level of agreement would only be reached for larger values of δM_{th}. In the discussions of the revised version of the manuscript we have further explained the limit of the b-positive method in presence of a threshold that does change in space.

6. Finally, assuming a detection function modeled as an Erf function, is only a reasonable choice, fist proposed by Ogata and Katsura (1993) (please add the citation), but it does not always correspond to the real incompleteness of the data, see for example the Italian catalogs CSTI (https://doi.org/10.13127/csti.1.1), CSI (https://doi.org/10.13127/CSI.1.1) and Horus (https://doi.org/10.13127/HORUS) before 2003.

Answer: We thank the referee for signaling the missing reference to Ogata amd Katsura, that we added in the revised version. We agree with the reviewer, that the functional form of the detection function describes only quantitatively the real detection function. We wish also to remark, that however the efficiency of the method does not depend explicitly on the functional form of the detection function. We added a brief comment to this point in the revised manuscript.

7. Please indicate the space, time and magnitude ranges (and not just the numbers of data) of real catalogs analyzed.

Answer: We thank the referee for this suggestion. In the revised manuscript is explicitly state that only earthquakes with magnitude $m > 0$ and depth smaller than 50 km are considered. In addition, we have included a new figure showing the spatial distribution of seismicity in the five study regions (Japan, Italy, Southern California, Northern California, and New Zealand). This figure allows the reader to directly visualize the spatial extent of the analyzed catalogs and the geographic domains considered in the study.

8. Please provide some words (or citations) about the method (thinning?) adopted to introduce STAI and SNDI into the complete simulated catalogdev.

Answer: We thank the reviewer for this suggestion. We now clarify in the revised manuscript that the introduction of STAI and SNDI into the complete ETAS catalog is implemented via a history- and space-dependent probabilistic thinning procedure. Specifically, events are retained or removed according to a retention probability defined through the detection function, which depends on preceding magnitudes (STAI) or spatially varying completeness levels (SNDI). We have added appropriate references (e.g.; Ogata, 1981) to situate this procedure within the standard thinning framework for point processes.

2 Report of Reviewer #2

This manuscript benchmarks three b-value estimators, the classical MLE (β_0), the b-positive estimator (β_+; van der Elst, 2021), and the authors' b-more-positive estimator (β_{++}; Lippiello & Petrillo, 2024), using synthetic ETASI catalogs and five instrumental catalogs (Japan, Italy, Southern California, Northern California, New Zealand). The main conclusion is that β_0 is systematically biased low by STAI/SNDI incompleteness, while β_{++} recovers β_{true} accurately. The manuscript also claims statistically significant regional differences in b from β_{++}, and documents a non-monotonic dependence of β_{++} on δM_{th} that remains unexplained. The topic is relevant and the idea is interesting, but in its current form the manuscript has several major scientific and statistical problems that need to be addressed before the conclusions can be supported.

Answer: We would like to express our sincere gratitude to the Reviewer for their detailed and insightful analysis of our manuscript. We particularly appreciated the timely nature of the review and the constructive comments provided, which have been instrumental in improving the quality of our work. We are also encouraged by the Reviewer's assessment "The topic is relevant and the idea is interesting". In the following we address in detail each specific remark.

MAJOR ISSUES

1) Internal inconsistency in the core derivation. Your central justification relies on Condition (7), $\phi(m_j - M_c) = 1$, which is used to obtain a pure exponential $p(\delta m)$ with coefficient β. However, with your own ϕ definition (Eq. 3), ϕ never equals 1 for finite magnitudes: $\phi(0) = 0.5$ and even $\phi(2\sigma) = 0.98$. The manuscript treats the condition as "satisfied" once $\delta M_{th} \geq 2\sigma$, but you never quantify the residual bias introduced by $\phi < 1$, nor propagate it into β_{++}. What I need to see to accept the derivation: rewrite (7) as an inequality $\phi \geq 1 - \epsilon$, define ϵ explicitly, map δM_{th} and σ to ϵ numerically, and show that the induced bias in β_{++} is negligible for the δM_{th} values actually used..

Answer: In the revised version of the manuscript, we address this point in detail, specifically regarding the correction required when the cumulative distribution does not saturate exactly at one. We present a comprehensive analysis of the expected error arising when $\sigma \neq 0$. We have numerically evaluated the correction term ϵ that accounts for a finite $\delta M_{th}/\sigma$ ratio. Consequently, the revised manuscript now features a generalized expression for β_{++} that incorporates this correction, supported by a comprehensive discussion in the new Appendix A. Our results demonstrate that for a b-value (in $\ln_{10}$ terms) where $\beta \approx \ln(10)$, and even for a threshold $\Delta M_{th} = 2\sigma$, the expected correction remains below 2%. This is significantly smaller than the standard uncertainties typically encountered in b-value estimation.

2) Self-validation on a synthetic benchmark designed to match the estimator's assumptions β_{++} comes from Lippiello & Petrillo (2024). The ETASI simulation framework is from the same group (Petrillo & Lippiello 2021, 2023), and it embeds exactly the incompleteness hypotheses (i) and (ii) that β_{++} is constructed to satisfy. Testing β_{++} on ETASI is therefore a consistency check, not independent validation. In addition, your baseline completeness estimators (MAXC, MBS, CV) are known to behave poorly under the kind of incompleteness you impose, which makes them a weak reference for "accuracy". What I need to see: validation under at least one incompleteness model outside ETASI (e.g., log-linear time decay of M_c as in Helmstetter et al. 2006, or purely SNDI-driven incompleteness), and comparison against at least one robust alternative estimator not developed within the same framework..

Answer 2:

Regarding the Reviewer's specific suggestion, we would like to point out that a detection threshold M_c that decays logarithmically over time—as proposed by Helmstetter et al. 2006 to model catalog incompleteness—is actually a direct consequence of the constant blind time approach. This relationship is explicitly derived in Hainzl (2016) (see his Equation 7) and is further illustrated in Figure 3 of that same study. Therefore, we argue that the 'blind time' approach is more general and fundamentally incorporates the mechanism suggested by the Reviewer, providing a robust physical basis for the observed incompleteness. Moreover, we wish to point to the reviewer attention that in the study by de Arcangelis et al. (2019), several masking mechanisms were compared; it was found that a constant blind time, despite its relative simplicity, yields results very similar to more complex formulations. More generally, regardless of the specific modeling of STAI-induced shadowing, our theoretical analysis demonstrates that the b-positive approach effectively filters out inconsistencies related to STAI, independently of its explicit functional implementation. Consequently, we believe that repeating the entire set of simulations with a different STAI implementation would not yield further insights. However, should the Referee consider this a critical requirement, we are prepared to perform these additional tests.

Regarding the use of a more robust estimator, our study demonstrates that the combined effect of STAI and SNDI causes the estimated b-value to monotonically increase with M_{th} (as shown in Fig. 1a). Consequently, any chosen completeness estimator—no matter how large is the output value of Mc—results in a systematic underestimation of the true b-value. This directly challenges the common assumption that an accurate b-value can always be recovered by simply choosing a sufficiently large M_{th}.Our results (see Fig. 1 and Fig. 3) suggest that the issue is not the specific M_c estimator used, but rather an intrinsic impossibility of obtaining an accurate b-estimate under these conditions. We have employed the most widely used indicators in the literature to identify the completeness magnitude; however, we are open to testing any other indicators the Reviewer might suggest. That said, based on our analysis, we anticipate that any other metric would similarly yield underestimated b-values. In the revised version we have added a sentence at pag.10 where we specify that the underestimation of the b-value with the traditional estimator is independent of the value of M_{th} adopted.

3) Pair-selection rules are not shown to enforce hypotheses (i) and (ii) in real catalogs. The method claims to control completeness indirectly via pair selection (d_R and the "no larger intervening event" condition). In real catalogs, M_c is unobserved and varies in time and space. The current evidence that the rules avoid STAI/SNDI bias is circular: β_{++} appears stable, therefore the selection must have worked. That is not a scientific demonstration. What I need to see: - On synthetics: ablation tests separating STAI-only, SNDI-only, and combined incompleteness, showing which mechanism drives the bias in β_0/β_+ and why β_{++} avoids it. - On real catalogs: diagnostics of the selected pairs-distributions of $(t_j - t_i)$, d_{ij}, time since last larger event, and at least one proxy for completeness variation (e.g., station density / network evolution if available)-demonstrating that the selected j events occupy demonstrably higher-completeness conditions.

Answer: We agree with the Reviewer that an exact, point-wise estimation of M_c is missing; indeed, this lack of precision is the core difficulty in handling incomplete catalogs. However, our method relies on two properties of M_c that are widely considered reasonable in the literature. First, the masking effect causes M_c to decrease over time (the basis of our Hypothesis 1). Second, if two events occur close to each other in space, the network's ability to detect them is nearly identical (Hypothesis 2). While defining the exact spatial threshold for this similarity is challenging, as it depends on local network geometry, our choice of $d_r \approx 20$ km is supported by regional M_c maps (e.g., see Shoerlemmeer & Woessner (2008) for Southern California). These maps show that within such a radius, M_c can be effectively considered constant. We remark that in our analysis, we found that results remain robust for d_r values ranging from 15 to 90 km, confirming that d_r is a stable model parameter. We maintain that our theoretical framework provides a robust foundation for the claims presented in this study. The tests performed on incomplete ETAS catalogs—which currently represent the most accurate statistical description of seismicity—further validate our theoretical approach. While this does not definitively prove that the measured β_{++} in instrumental catalogs is the "true" value, it demonstrates that, given current constraints, it represents the most statistically sound estimation available. Regarding the Referee's suggestion to validate our scheme on real-world data, if we have understood the request correctly, such a test would require a precise estimation of local completeness. However, we contend that reconstructing such local completeness levels is not feasible with the information currently available in instrumental catalogs. Indeed, it would necessitate knowing the exact network configuration and triggering thresholds for every single event pair at their specific occurrence times. The estimation of the completeness magnitude suggested by the reviewer would remain qualitative at best and, ultimately, less informative than the more rigorous assertion that if earthquake i has been detected, its magnitude m_i is realistically above M_c. This last assertion constitutes the core of the b-positive estimator. Our 'distance-link' approach is not only computationally efficient but also avoids making ad hoc assumptions about the explicit functional form of M_c. Finally, concerning the request to separate STAI and SNDI effects in numerical catalogs, as stated above, we note that this analysis was already performed in our previous work, where the 'b-more-positive' method was shown to bypass completeness issues in both cases. (see Lippiello & Petrillo, 2024). The specific objective of the current study is to investigate their combined effect. In the revised manuscript, we have expanded the discussion on the role of d_r to better clarify its physical and modeling significance and specified that the separate analysis of STAI and SNDI was already performed.

4) The δM_{th} selection rule (β^*_{++}) is ad hoc and not statistically justified The "best estimate" β^*_{++} is defined using an MBS-style plateau criterion on $\beta_{++}(\delta M_{th})$. MBS was developed for a monotonic $\beta_0(M_{th})$ curve, while your $\beta_{++}(\delta M_{th})$ curves in real catalogs are explicitly non-monotonic. In that situation, the "plateau" can simply be the rising limb before the subsequent drop. Also, your rule depends on the averaging window and step size, but sensitivity is not explored. Finally, the uncertainty reported $\delta\beta \approx \beta/\sqrt{N_{++}}$ does not include uncertainty from choosing $\delta_{M_{th}}$. What I need to see: sensitivity of β^*_{++} to step size and window length, and uncertainty estimates that propagate the $\delta_{M_{th}}$ selection uncertainty. Preferably, replace the heuristic plateau rule with a likelihood-based or explicit bias–variance criterion tied to your $\phi \geq 1 - \epsilon$ threshold.

Answer: We demonstrate that the 'plateau criterion' is highly effective for ETASI catalogs, where β_{++} indeed becomes constant beyond a certain δM_{th} threshold, in full agreement with theoretical predictions. We agree with the Reviewer that when the b-value does not exhibit monotonic behavior—as observed with experimental catalogs—our initial definition of the 'best b-value' may be less applicable. Consequently, in the revised version of the manuscript, we have introduced two distinct definitions for the b-value: $\beta_{++}max$ which represents the value expected based on theoretical considerations, assuming that a single b-value describes the magnitude distribution. The second one $\overline{\beta_{++}}$ (as suggested by the Reviewer) is obtained by performing a Maximum Likelihood Estimation within the interval between the M_{th} that yields β_{++}^{max} and the M_{th} that produces the subsequent β_{++}^{min}. By providing both metrics, we offer a more flexible approach that accounts for the complexities and fluctuations inherent in real-world seismic data. Regarding the Referee's request to evaluate the effect of the "averaging window and step size," we must clarify that it is not entirely clear to us how these parameters should be implemented within the specific context of our methodology. Since the β_{++} framework does not inherently rely on a sliding-window temporal or spatial averaging, we are currently unable to perform the requested test. We wish to specify that the MBS criterion identifies the optimal β_{++} value as the one that remains consistent across different choices of δM_{th}. Specifically, the criterion is satisfied when the variations in β_{++} are significantly smaller than the associated statistical uncertainty. In our implementation, we increase δM_{th} in steps of 0.1. While the uncertainty related to the choice of δM_{th} could, in principle, be associated with the difference between two subsequent estimates of β_{++}, by definition, this difference is much smaller than the overall statistical uncertainty of the estimator. If the Referee could provide further clarification on how these parameters relate to our analysis, we would be happy to investigate this aspect.

5) "True β" language is not justified for instrumental catalogs. In synthetics, β_{true} is the simulation input; that is valid. In instrumental catalogs there is no external ground truth, yet the manuscript repeatedly states that β_{++} "recovers the true β" for Japan/Italy/etc. That is not a defensible statement given the current evidence. Required change: in all instrumental-catalog discussion, replace "true β" language with a careful statement such as "β_{++} is less sensitive to incompleteness under the stated assumptions and yields systematically higher β than threshold-based estimators.".

Answer: The referee is right, we made an improper use of the term "true". We have corrected this point in the revised version of the manuscript.

6) The hazard-direction statement is not supported as written The conclusion states that biased-low b-values "may have led to systematic overestimation of seismic hazard." This is not established by the paper. The sign of hazard change is not determined by b alone; it depends on how the model is normalized, the magnitude range, M_{max} assumptions, and the return period of interest. As written, the statement is misleading and too strong. Required change: remove the hazard-direction claim or support it with a demonstrator hazard/rate calculation that explicitly shows the sign under a stated PSHA parameterization. At minimum, restrict to "implications for seismicity rate extrapolation and parameter estimation.".

Answer: We thank the referee for the suggestion. We have reformulated this statement to avoid misunderstanding.

7) The significance test for non-monotonic β_{++} is statistically invalid. The z-test compares β_1 (the maximum) and β_2 (a subsequent minimum) selected from the same $\beta_{++}(\delta M_{th})$ curve. This creates strong selection bias: the distribution of a max-minus-min statistic is not Gaussian with (σ_1, σ_2), and the curve points are not independent because thresholds are nested and share events. The reported confidence levels are therefore inflated. What I need to see: a permutation/Monte Carlo-based null distribution for the same max–min statistic under a realistic null model (i.i.d. magnitudes + incompleteness), or else treat the non-monotonicity descriptively without formal significance claims..

Answer: The reviewer is perfectly right β_1 and β_1 are strongly correlated. We thank the referee for enlightening our mistake. In the new version of the manuscript we performed a comparison via the generalized t-test:

$$t = \frac{\beta_1 - \beta_2}{\sqrt{\sigma_1^2 + \sigma_2^2 - 2r\sigma_1\sigma_2}} \tag{1}$$

where σ_1 and σ_2 are the standard deviations of β_1 and β_2 respectively and r is the correlation between β_1 and β_2 that can be estimated as

$$r \approx \sqrt{\frac{n}{n+m}} \tag{2}$$

We discuss the result of this test in the Appendix D of the revised version of the manuscript.

8) Uncertainty quantification is inadequate for β_{++}. Pairs used by β_{++} are conditionally selected and can be dependent (triggering/clustering, shared events across thresholds). The variance formula $\delta\beta \approx \beta/\sqrt{N_{++}}$ assumes i.i.d.-like behavior and may substantially understate uncertainty. What I need to see: uncertainty validated with an appropriate bootstrap (e.g., block bootstrap in time, or resampling independent sequences), plus reporting of how many distinct events actually contribute to β_{++}.

Answer: We wish to draw the referees' attention to the fact that, despite the presence of clustering in both space and time, the standard working hypothesis in seismology assumes that earthquake magnitudes are Independent and Identically Distributed (IID) variables. Strictly speaking, we have previously demonstrated, specifically by analyzing the subtle effects of incompleteness on magnitude differences, that some correlations between subsequent magnitudes may be present in instrumental datasets (Lippiello et al., Physical Review Letters 2024). Nevertheless these correlations are sufficiently weak as to have a negligible impact on uncertainty estimates. Furthermore, we remark that an analytical expression for the uncertainty of b-positive estimators has been formally derived by Tinti & Gasperini (2024), which we have adopted in the present work. Conversely, in our previous study (Lippiello & Petrillo, 2024), we employed a bootstrap method to estimate the error. Here we have verified that both approaches yield comparable results. In the revised version of the manuscript, we now explicitly state that the hypothesis of magnitude independence is widely accepted and provides a robust basis for error estimation.

9) Declustering sensitivity is not tested Hypothesis (i) is most likely to be violated in strongly clustered sequences (aftershock-rich catalogs), yet all instrumental catalogs are analyzed "as-is" with no sensitivity check. What I need to see: repeat at least two catalogs with a standard declustering approach and report whether β_{++} and β^*_{++} materially change.

Answer: We respectfully disagree that hypothesis (i) should be violated in the presence of strongly clustered sequences. This hypothesis is grounded in the standard observation that the completeness magnitude M_c increases sharply, such that $M_c \approx m_i$, immediately following an event of magnitude m_i, and subsequently decays over time. While the occurrence of a new event may cause M_c to rise again, it can only exceed the initial magnitude m_i if an event larger than m_i occurs. By definition, the β_{++} method considers the first subsequent event at time $t_j > t_i$ such that $m_j > m_i$. Consequently, it is physically impossible for M_c at time t_j to be larger than m_i, regardless of the density of the aftershock cluster. High clustering may cause t_j to occur closer to t_i in time, but this does not affect the validity of the estimator. Regarding the referee's request to estimate β_{++} after declustering, we remark that such a procedure would destroy the fundamental structure of b-positive methods. Declustering routines, by their nature, remove events from the catalog; this directly contradicts the essential requirement that the pair $(m_i < m_j)$ must be "consecutive", meaning no event with $m_k > m_i$ exists between them within the distance d_r. This condition cannot be satisfied if one considers only a declustered subset of the catalog. In the revised manuscript, we now specify that the procedure is incompatible with declustered catalogs or, more generally, with any catalog where a pre-filtering procedure has been applied.

MINOR ISSUES

Reproducibility details are incomplete: magnitude type and homogenization, exact spatial polygons, treatment of blasts/induced events, and temporal changes in completeness/network evolution need clearer documentation..

Answer: We thank the referee for this important comment. In particular, we now state the type of magnitude for each catalog and specify that the magnitude threshold adopted in the analysis is $m > 0$, together with the depth cutoff (< 50 km). We have also added a new figure (Fig.1) showing the spatial distribution of seismicity in the five study regions, so that the geographic extent of the analyzed catalogs is immediately visible. We have not implemented any specific procedure to remove quarry blasts or induced seismicity events. We agree that this is an important detail for the characterization of the dataset; therefore, we have explicitly clarified this in the revised version of the manuscript. While these events can locally affect b-value estimates, we expect their impact on the global robustness of the β_{++} framework to be marginal in the context of our study. Concerning temporal changes in completeness/network evolution, the use of the b-more-positive estimator strongly mitigates these effects. This estimator is specifically designed to be robust with respect to missing small events, which allows us to include earthquakes with $m > 0$ without introducing significant bias in the estimation.

The rounding correction (jitter) is shown for β_+ but no explicitly validated for β_{++}. Please demonstrate seed stability (e.g., 20–50 jitter realizations) and show that the additional variance is negligible compared with catalog uncertainty..

Answer: The effect of magnitude rounding is identical for both β_+ and β_{++}; therefore, duplicating the analysis shown in Figure 5 for β_{++} would be redundant. Nevertheless, we have verified for each curve that the results remain consistent when incorporating the analytical expression for rounding proposed by Tinti & Gasperini (2024). We have clarified this point in the revised manuscript to reassure the reader of the robustness of our estimates against rounding effects.

β vs b inconsistency: Table values are $\beta (= b \ln 10)$, but the text repeatedly says "b-value." This is confusing and will mislead many readers. Use one convention consistently or always state the conversion..

Answer: We thank the reviewer for this suggestion. In the revised manuscript we have clarified the notation throughout. We now explicitly state in the Introduction that all reported numerical results are given in terms of $\beta = b \ln(10)$, and that the corresponding base-10 b-values are obtained as $b = \beta / \ln(10)$. In addition, all tables and figure captions now clearly indicate whether the reported quantities correspond to β, and the conversion is explicitly provided. This removes any possible ambiguity.

Several statements are too categorical (e.g., "traditional estimators systematically underestimate," "β_{++} yields highly accurate results"). These need to be qualified as "under the tested incompleteness model" / "in the studied catalogs" / "under stated assump-

tions.".

Answer: We thank the reviewer for this important comment. We agree that some statements in the original version of the manuscript were formulated in an overly categorical manner. In the revised manuscript, we have carefully qualified these statements throughout the Abstract, Results, and Discussion sections. In particular: (i) claims regarding the systematic underestimation by traditional estimators have been rephrased to explicitly refer to the tested ETASI incompleteness model and to the instrumental catalogs analyzed here; (ii) statements describing the performance of the β_{++} estimator have been reformulated to clarify that the reported accuracy holds under the stated assumptions and within the investigated framework; (iii) in the analysis of regional catalogs, we now avoid referring to a "true" value and instead describe relative differences among estimators.

Numerous typos and grammar issues (e.g., "intrumental," "New Zeland," "containse," "concerncs," "stnthetic," etc.) should be corrected before resubmission..

Answer: Done.

http://www.declustering.com/ seems to be slow and it is not loading (add dataset to zenodo) .

Answer: We thank the reviewer for pointing this out. The website www.declustering.com is currently temporarily unavailable due to technical maintenance issues. To ensure long-term accessibility and reproducibility, we have now deposited the code for the generation of incompleteness in a dedicated Zenodo repository https://doi.org/10.5281/zenodo.18831071. The Data Availability section has been updated accordingly, and the Zenodo DOI is now provided in the revised manuscript.

Finally, please report b-values (base 10) rather than β throughout (or at least report both). At present the manuscript primarily reports $\beta = b\ln(10)$ (e.g., Table 1), while the narrative repeatedly discusses "b-values." This is unnecessarily confusing for most readers and makes the reported ranges look "too large" unless the reader mentally converts ($\beta \approx 2.3$ corresponds to $b \approx 1.0$). I strongly recommend reporting b as the primary quantity in the main text, figures, and tables, and (if you want to retain β for derivations) include β only secondarily or in parentheses. At minimum, every table/figure should state explicitly whether it reports b or β, and provide the conversion used..

Answer: We thank the reviewer for this important suggestion. We agree that the distinction between b and $\beta = b\ln(10)$ should be made fully explicit to avoid confusion. In the revised manuscript, we have clarified at the beginning of the Introduction that all reported numerical results are given in terms of β, while the corresponding b-values can be obtained through the relation $b = \beta/\ln(10)$. We have also explicitly stated in all table captions and figure captions whether the reported values correspond to β, and we provide the conversion formula. We retain β as the primary reported quantity for consistency with the analytical derivations, but the relationship to the base-10 b-value is now clearly indicated throughout the manuscript.

Dear Editor,
we would like to sincerely thank you for your positive decision to proceed with the publication of our manuscript.
Following your requests, we have introduced two additional sentences in the Conclusions section. The first one points the reader to further analyses provided in the Appendices and to the review–response document available online, where additional elements of reasoning can be found. The second one addresses the general limitations of the methods, including a brief discussion on their applicability to real datasets.
Regarding this latter point, we would like to clarify that, in our view, the potential issues arising in real instrumental catalogs are not related to the catalog size. In fact, the number of events is typically sufficient to ensure a highly accurate estimation of the b-value. Rather, the main limitations are likely associated with data heterogeneity and measurement uncertainties, which are often difficult to fully characterize and may introduce residual biases.
We hope that these additions satisfactorily address your concerns and improve the overall quality of the manuscript.
Thank you again for your support and for handling our submission.
Kind regards,
The authors