

Response to Review Comments (Round 3)

Dear Editor and Referee #3,

Thank you for providing us with another opportunity to further improve our revised manuscript (egusphere-2025-2706). We sincerely appreciate the referee's careful, rigorous, and constructive assessment in this round of review. The comments have been especially valuable in prompting us to clarify the conceptual status and evidentiary boundaries of the ARI, strengthen the quantitative comparison and sensitivity evidence, improve the transparency of the spatial-standardization procedure, and align the remote sensing component and manuscript claims more closely with what the data and methods support.

In response to the twelve comments, we have undertaken a focused but substantive revision of the manuscript. The principal revisions are summarized as follows:

1) Terminological and conceptual clarification of ARI. We have changed the full designation from “Accumulated Risk Index” to “Accumulated-Event Risk Indicator” while retaining the abbreviation ARI. The revised manuscript now defines ARI as a static, retrospective, event-driven, grid-based, and context-weighted relative indicator for hotspot prioritization. It explicitly distinguishes the arithmetic accumulation of qualifying major-event contributions over the fixed 2013–2023 period from dynamic recovery, temporal decay, evolving vulnerability, event sequencing, or resilience loss.

2) Addition of direct global diagnostic comparisons. We have introduced Equivalent WRI, unweighted major-event count, and death-sum as three direct reference indicators constructed using the same 2013–2023 event period and 5°×5° grid framework. Their relationships with ARI are evaluated through Spearman rank correlations, Top-14 Jaccard similarities, and tied-value characteristics. These analyses clarify the deliberate dependence of ARI on both WRI context and event recurrence while showing that its upper-priority pattern does not simply reproduce any one reference indicator. The comparisons are interpreted as construct-level diagnostics rather than as independent validation against a ground-truth risk surface.

3) Expansion and reframing of the post-period comparison. The previous 2024-only comparison has been extended to include 44 Asian major disaster events from 2024–2025. ARI is compared directly with the three reference indicators using global historical ranks, paired event-level comparisons, and a descriptive equal-grid reference defined over a 76-cell Asian comparison universe. The analysis is now consistently presented as a post-period consistency and benchmark comparison rather than predictive validation.

4) Quantified methodological sensitivity analyses. The ≥ 50 -fatality criterion is now described as a study-specific operational screening rule, and its influence is evaluated through local threshold changes to 40, 60, and 70 fatalities. The resulting effects are quantified through non-zero-cell counts, Top-14 retention, mean absolute rank displacement, and Spearman rank correlation. Grid-resolution sensitivity is evaluated through explicit parent–child correspondence between the 5° and 2.5° grids, while the 10° result is retained as a descriptive coarser-resolution comparison. We have also added a weight-conserving spatial-allocation sensitivity analysis for the 2022 Pakistan floods. These analyses distinguish preservation of the broader priority geography from exact invariance of ARI values, ranks, or membership near the selection boundary.

5) More transparent treatment of event-location standardization. The spatial-processing procedure is now described as rule-guided manual spatial standardization. The manuscript acknowledges that more than one plausible event-location anchor may exist for records containing several affected locations, that the manual selection step may involve qualitative

judgement, and that each anchor remains constrained by the geography documented in the corresponding EM-DAT record. It also states explicitly that assigning one event-location anchor to a single grid cell simplifies the spatial footprints of geographically extensive events. The influence of this simplification on the resulting hotspot pattern is further examined through the alternative allocation of the 2022 Pakistan flood event.

6) Clearer linkage between ARI and the 3H Dataset. The revised manuscript presents ARI as the macro-scale hotspot-screening and spatial-inclusion framework: it determines the 14 priority grids within which imagery is searched and data gaps are assessed, while imagery compilation and supplementation constitute the subsequent data-development stage. In this response, we further clarify that the relationship is one-directional and downstream: the imagery does not enter, validate, recalibrate, or update the present ARI calculation. The coverage and role of the supplementary imagery for Grids 3 and 7 are also quantified, including the approximately 0.019 % within-grid coverage of each acquisition and the resulting total of 19,835 standardized image tiles. The 3H Dataset is accordingly positioned as a spatially partial, hotspot-oriented observational resource rather than as complete or representative coverage of the 14 grids.

7) Revision of the manuscript’s evidentiary language and technical presentation. The Abstract, Results, Discussion, and Conclusions have been revised to distinguish diagnostic comparison, post-period benchmark evidence, and sensitivity analysis from independent validation, predictive performance, or universal robustness. A new Appendix B provides detailed event-level and grid-level supporting results. Figure 5 has been reinserted and checked, and figure and table captions, symbols, references, and related typographical details have been reviewed. The Data availability statement has also been clarified to distinguish the documentation provided in the manuscript from the planned repository deposition of the finalized dataset.

The principal new or substantially revised materials are located in Sections 2.1–2.6, Sections 3.2 and 3.3.1–3.3.4, Section 4, Sections 5 and 6, and Appendix B. The response tables and figures provide additional details, while the corresponding methods, summary results, interpretations, and limitations are incorporated directly into the revised manuscript.

All manuscript modifications are marked in the revised manuscript for ease of review. In the following point-by-point response, the referee’s comments are reproduced in blue text, our responses are provided in black text, and excerpts quoted directly from the revised manuscript are shown in red text with the corresponding section, table, figure, or appendix location.

We have carefully considered each comment and provide our detailed responses below. We are grateful for the referee’s time and for the opportunity to further improve the clarity, transparency, and evidentiary discipline of the manuscript.

Comment 1:

The ARI remains conceptually circular because it sums WRI values for events already selected by a mortality threshold, meaning high-WRI countries contribute more to ARI and high-ARI grids are therefore partly a reflection of WRI itself rather than an independent measure of “accumulated pressure”.

Response 1:

We sincerely thank the referee for raising this important conceptual concern. We agree with the factual premise that country-level WRI values enter explicitly into the ARI calculation: a qualifying event occurring in a country with a higher WRI necessarily makes a larger contribution to the ARI value of the grid to which that event is assigned. This dependence must be stated transparently, and ARI should not be interpreted as a disaster-risk measure that is independent of WRI.

However, we respectfully distinguish deliberate construct dependence from conceptual circularity. ARI was intentionally designed to combine two operationally distinct inputs that enter the calculation through different data sources and perform

different analytical roles: the spatial accumulation of mortality-screened major disaster events recorded in EM-DAT during 2013–2023 and the broader country-level structural risk context represented by WRI. Its purpose is not to reconstruct risk independently of WRI, but to contextualize the recent accumulation of major events according to the exposure–vulnerability conditions represented by WRI.

The two components enter the analysis through different data sources and perform different analytical roles. EM-DAT determines which events satisfy the study-specific screening criterion and provides the location information used to assign those events to grid cells. WRI does not determine which events enter the analysis or the grid cells to which they are assigned. It is introduced only after event selection and spatial allocation, as the country-level contextual weight associated with each qualifying event. EM-DAT therefore supplies the event-occurrence and location component, whereas WRI supplies the broader structural context in which those events occurred. The explicit use of WRI consequently makes ARI WRI-dependent by design, but it does not make the construction circular.

A circular validation argument would arise if WRI were used to construct ARI and then treated as an independent external benchmark validating the resulting indicator. We agree that our previous presentation did not distinguish this point sufficiently clearly. In particular, the comparison with Equivalent WRI should be understood as a component-based diagnostic comparison that examines how strongly the ARI pattern reflects its WRI-derived component and how the addition of historical event accumulation changes the resulting grid prioritization. It is not an independent validation of ARI and does not establish that ARI is intrinsically more accurate or universally superior to WRI. We have revised the manuscript accordingly and now describe ARI consistently as an event-driven, grid-based, context-weighted relative indicator for hotspot prioritization, with WRI explicitly serving as the country-level contextual weight.

Conceptual clarification alone, however, would not determine whether ARI merely reproduces WRI, whether it is effectively equivalent to an unweighted event-count surface, or whether it simply identifies grids with the largest cumulative fatalities. We therefore added a global diagnostic comparison among ARI and three reference indicators constructed using the same 2013–2023 study period and 5°×5° grid framework. Equivalent WRI represents the country-level WRI background transferred to the grid framework without historical event accumulation; unweighted major-event count represents the number of qualifying events assigned to each grid without WRI weighting; and death-sum represents the total reported fatalities of the qualifying events assigned to each grid. These reference indicators directly test the principal alternative interpretations of ARI: that it is primarily a remapping of WRI, an event-count surface, or a cumulative-fatality surface.

The comparison was conducted across the same 153 grid cells with non-zero ARI values. Spearman’s rank correlation coefficients were used to evaluate overall monotonic correspondence among the grid rankings, while Jaccard similarity coefficients were used to evaluate overlap among the tie-inclusive Top-14 high-value grid sets. The Top-14 definition was retained because these 14 grids constitute the principal hotspot set used subsequently in the study and in the construction of the 3H Dataset.

As shown in Response Table R1, ARI is associated with both its WRI-derived contextual component and its event-recurrence component. Its Spearman correlations with Equivalent WRI and unweighted major-event count are 0.698 and 0.697, respectively, while its correlation with death-sum is lower, at 0.546. These results are consistent with the mathematical construction of ARI and show directly that both WRI context and recent qualifying-event accumulation influence its ranking.

Overall rank association, however, does not imply identical hotspot prioritization. The Top-14 Jaccard similarities between ARI and Equivalent WRI, unweighted major-event count, and death-sum are only 0.148, 0.450, and 0.333, respectively. The particularly low overlap with Equivalent WRI shows that adding historical event accumulation materially reorganizes the upper-priority portion of the WRI-only spatial pattern. ARI is therefore WRI-contextualized by construction, but it is not a simple spatial remapping of WRI.

Response Table R1. Global diagnostic comparison between ARI and three reference indicators across the 153 grid cells with non-zero ARI values.

Indicator pair	Spearman's rank correlation coefficient	Jaccard similarity coefficient, Top-14 ^a
ARI vs. Equivalent WRI	0.698	0.148
ARI vs. unweighted major-event count	0.697	0.450
ARI vs. death-sum	0.546	0.333
Equivalent WRI vs. unweighted major-event count	0.234	0.067
Equivalent WRI vs. death-sum	0.231	0.069
Unweighted major-event count vs. death-sum	0.742	0.318

^a: All grid cells tied at the Top-14 cutoff were retained when constructing the high-value sets.

Similarly, the correlation between ARI and unweighted event count confirms that event recurrence is an important driver of ARI, as intended. Their incomplete Top-14 overlap nevertheless shows that WRI weighting changes which recurrent-event grids occupy the highest-priority positions. Two grids with the same number of qualifying events may receive different ARI values because their contributing events occurred under different country-level WRI contexts, and WRI weighting may alter the relative positions of grids with different event counts.

The lower correlation and limited hotspot overlap between ARI and death-sum further clarify the indicator's conceptual role. Mortality determines whether an event satisfies the study-specific inclusion criterion, but fatalities above that threshold are not used as a continuous event weight in the ARI calculation. Death-sum, by contrast, may be dominated by one or several exceptionally fatal events. ARI therefore represents mortality-screened event accumulation under different WRI contexts rather than cumulative mortality itself.

The relationships among the three reference indicators also demonstrate why these components should not be treated as interchangeable. Equivalent WRI has Spearman correlations of only 0.234 with event count and 0.231 with death-sum, while the corresponding Top-14 Jaccard similarities are 0.067 and 0.069. Country-level structural risk context, qualifying-event recurrence, and cumulative fatalities therefore provide substantially different ranking signals within the present comparison framework. ARI was specifically designed to integrate WRI context with event recurrence rather than to substitute one dimension for the other.

Event count and death-sum likewise capture different aspects of the historical record. Although their overall Spearman correlation is relatively high, at 0.742, their Top-14 Jaccard similarity is only 0.318. A grid characterized by repeated qualifying events is not necessarily the same as one dominated by the greatest cumulative fatality total. This distinction supports the use of multiple direct baselines rather than treating any single simple indicator as an equivalent representation of ARI.

We also examined tied values because extensive ties affect the practical ability of an indicator to distinguish among candidate hotspot grids. Across the same 153-cell comparison set, ARI produced 104 distinct values, compared with 44 for Equivalent WRI, 9 for unweighted event count, and 120 for death-sum. The proportions of grid cells belonging to tied-value groups were 49.7 % for ARI, 90.2 % for Equivalent WRI, 98.7 % for event count, and 36.6 % for death-sum. The largest event-count tie alone contained 80 grid cells. Equivalent WRI inherits extensive ties because multiple grids share the same country-level value, whereas event count is restricted to a small number of integer values. By combining event recurrence with heterogeneous WRI weights, ARI provides substantially finer differentiation among grids than either Equivalent WRI or unweighted event count and permits a more selective upper-priority set to be identified without retaining very large tied groups.

This differentiation is a practical prioritization advantage rather than evidence of greater absolute accuracy. Death-sum, for

example, produces more distinct values than ARI but represents a different quantity and may be strongly influenced by the magnitude of individual fatal events. The ranking-differentiation analysis therefore concerns the resolution with which candidate hotspot grids can be prioritized, not correspondence with an independently established representation of “true” disaster risk.

Taken together, the new analysis does not support the stronger interpretation that ARI is conceptually circular or merely reproduces WRI. ARI is intentionally dependent on WRI because WRI is incorporated as its contextual weight, and it is intentionally related to event recurrence because qualifying-event accumulation is its historical input. At the same time, the rank correlations, hotspot overlaps, and tied-value results show that its upper-priority pattern is not identical to the WRI-only background, the unweighted event-count surface, or cumulative reported fatalities. Its specific contribution lies in integrating recent mortality-screened major-event recurrence with country-level WRI context to support differentiated grid-level prioritization of recent major-event accumulation.

Accordingly, the revised manuscript now defines ARI consistently as an event-driven, grid-based, context-weighted relative indicator for hotspot prioritization, with WRI explicitly incorporated as its contextual weight; distinguishes deliberate construct dependence from circular validation; treats Equivalent WRI as a diagnostic reference rather than an independent validation benchmark; and incorporates the global quantitative comparisons reported above. Language implying that greater spatial contrast or ranking differentiation proves greater accuracy, robustness, or universal superiority has also been revised. The resulting interpretation aligns the claims of the study with the mathematical construction of ARI while retaining its intended contribution as a transparent framework for differentiated global hotspot prioritization.

For ease of review, the principal corresponding revisions in the manuscript are reproduced below:

1) Section 2.3.2 Accumulated-Event Risk Indicator (proposed indicator)

“In this formulation, ARI represents the accumulated sum of country-level WRI weights associated with recent major disaster events occurring within each grid cell. ARI should therefore be interpreted as an event-driven, grid-based, context-weighted relative indicator for hotspot prioritization.”

“For grid cells intersecting multiple countries, each event is linked to the WRI of the country recorded in EM-DAT as part of the affected area, preserving the association between individual disaster records and their national risk contexts. To facilitate comparison with the original country-based WRI, an Equivalent WRI is calculated by assigning each grid cell the maximum WRI of all overlapping countries. This provides a conservative grid-level representation of the highest WRI background within the common spatial framework, without incorporating historical event accumulation. Two additional event-based reference indicators are calculated using the same qualifying events and grid assignments: unweighted major-event count, defined as the number of qualifying events in each grid, and death-sum, defined as the total reported fatalities of those events. Their diagnostic comparisons with ARI are described in Section 2.6 and reported in Section 3.2.”

2) Section 3.2 Global ARI distribution, hotspot identification, and diagnostic comparison

“To characterize its relationships with the WRI-derived and event-based reference patterns, ARI is compared with Equivalent WRI, unweighted major-event count, and death-sum across the same 153 cells with non-zero baseline ARI. Table 4 reports pairwise Spearman rank correlations and Jaccard similarities among the tie-inclusive Top-14 high-value sets.”

“ARI has similar overall rank correlations with Equivalent WRI and unweighted major-event count, at 0.698 and 0.697, respectively, and a lower correlation with death-sum, at 0.546. These relationships are consistent with the formulation of ARI, in which WRI is deliberately incorporated as the contextual weight and qualifying-event recurrence supplies the historical event component.”

“The corresponding upper-priority overlaps are substantially lower than the overall rank associations, with Top-14 Jaccard

similarities of 0.148, 0.450, and 0.333 for Equivalent WRI, unweighted major-event count, and death-sum, respectively. These results indicate that historical event accumulation reorganizes the upper-priority part of the WRI-only pattern, WRI weighting changes the ordering of recurrent-event grids, and death-sum produces a different priority pattern from mortality-screened event accumulation.”

“The four indicators also differ in ranking differentiation. Across the 153-cell comparison set, ARI produces 104 distinct values, compared with 44 for Equivalent WRI, 9 for unweighted major-event count, and 120 for death-sum. The proportions of cells belonging to tied-value groups are 49.7 %, 90.2 %, 98.7 %, and 36.6 %, respectively. By combining event recurrence with heterogeneous WRI weights, ARI provides finer differentiation than the WRI-only and integer-valued event-count references within the present dataset, supporting more selective grid prioritization.”

3) Section 5 Discussion

“The principal contribution of ARI is to provide a spatially explicit complement to country-level structural risk indices. WRI characterizes national exposure and vulnerability conditions, whereas ARI highlights where mortality-screened major disaster events accumulated during 2013–2023 under those broader risk contexts. The diagnostic comparisons show that both event recurrence and WRI contextualization shape the ARI pattern. Unweighted major-event count remains a strong comparator, particularly as the screening set broadens, but WRI weighting changes the ordering and composition of the most selective upper-priority grids. It also reduces the extensive ties associated with country-level WRI values and integer-valued event counts, thereby supporting more differentiated prioritization when only a limited number of regions can be selected for further assessment.”

Comment 2:

“Accumulated risk” is a misleading label because the index measures the number of major events in a grid multiplied by national WRI, with no recovery period, no temporal decay, and no resilience erosion, so it is not “accumulated risk” in any dynamic sense.

Response 2:

We sincerely thank the referee for identifying the potential mismatch between the former terminology and the processes explicitly represented by the current formulation. We agree that, when “accumulated risk” is understood as a dynamic process involving post-disaster recovery, temporal decay of earlier impacts, event sequencing, evolving vulnerability, interactions among successive shocks, or resilience deterioration, the present ARI does not constitute such a model.

We respectfully clarify, however, that these dynamic mechanisms were not intended to define the analytical object of the present study. The original use of “accumulated” referred to a narrower and explicitly retrospective operation: the arithmetic aggregation of qualifying major-event contributions within a fixed historical period. The substantive purpose of ARI is to identify grid cells in which mortality-screened major disasters repeatedly occurred during 2013–2023 and to contextualize that event accumulation using the country-level structural risk conditions represented by WRI.

The current formulation is therefore retrospective and static rather than dynamically cumulative. For each grid cell, ARI sums the country-level WRI values associated with qualifying major disaster events assigned to that cell during the fixed 2013–2023 study period. All included events receive the same temporal treatment. A qualifying event recorded near the beginning of the study period contributes in the same manner as an otherwise equivalent event recorded near the end of the period. Event order, spacing, and recency therefore do not affect the calculated ARI value, and the formulation does not estimate recovery duration, temporal attenuation of earlier impacts, interactions among successive shocks, time-varying

vulnerability, or resilience loss.

These distinctions define the question that ARI is designed to answer; they do not make the indicator conceptually incomplete relative to that objective. ARI does not seek to reconstruct the evolving temporal state of disaster risk. It provides a standardized global representation of where recent mortality-screened major-event contributions accumulated spatially under different country-level WRI contexts.

We acknowledge that the previous designation, “Accumulated Risk Index”, did not communicate this distinction adequately. Expressions such as “accumulated risk”, “accumulated disaster pressure”, and “resilience deterioration” could reasonably be interpreted as claims that ARI quantified the dynamic consequences of repeated disasters. Such an interpretation would extend beyond the information contained in the present data and mathematical formulation. Simply explaining the intended meaning of “accumulated” while retaining the former name would not fully remove this ambiguity.

To retain continuity with the established abbreviation ARI while making the object of aggregation explicit, we have therefore changed the full designation from “Accumulated Risk Index” to “Accumulated-Event Risk Indicator”. The revised name emphasizes that it is qualifying major-event contributions that are aggregated, rather than implying that the temporal evolution of disaster risk, vulnerability, recovery, or resilience has been reconstructed. We also use “indicator” rather than “index” because ARI provides a comparative and screening-oriented representation within a specific event-selection, weighting, and grid framework, rather than a comprehensive measure of absolute disaster risk.

The revised Methods now define the retrospective formulation and interpretive scope of ARI explicitly. This clarification is placed directly in the methodological definition of ARI so that the temporal interpretation is established at the point where the indicator is introduced rather than being left only to a later limitations discussion. Corresponding revisions have been made throughout the Abstract, Introduction, Methods, Results, Discussion, and Conclusions. We now preferentially use more precise expressions such as “recent major-event accumulation”, “major-event accumulation under country-level WRI contexts”, “ARI-based hotspot prioritization”, and “concentrated recent major-event activity”.

Statements implying that repeated events have already caused measurable deterioration of local resilience have also been removed or revised. These revisions preserve the legitimate implication that repeatedly affected locations may warrant closer investigation of recovery and resilience without claiming that ARI itself quantifies those processes.

The mathematical description now states explicitly that reported fatalities determine event eligibility, after which each qualifying event contributes its associated country-level WRI value once, and that all events receive the same temporal treatment throughout 2013–2023. It further states that their chronological order does not affect ARI. Accordingly, no recency weighting or temporal decay enters the current calculation. This property defines the static retrospective design of ARI.

A genuinely dynamic formulation would constitute a different methodological extension and would require additional data and assumptions concerning event recency, recovery duration, intervals between successive events, time-varying exposure and vulnerability, and the treatment of shocks occurring before or after recovery. Such a formulation may be valuable for studying temporal disaster-pressure accumulation or resilience change, but it addresses a different research question from the present global hotspot-screening analysis.

The revised terminology therefore aligns the name and interpretation of ARI with the operation actually performed. ARI identifies where qualifying major-event contributions accumulated during 2013–2023 under different country-level WRI contexts. It is not a dynamic model of recovery or resilience loss, but it remains a clearly defined retrospective framework for identifying and differentiating global hotspots of recent major-event accumulation. The change from “Accumulated Risk Index” to “Accumulated-Event Risk Indicator” removes the former ambiguity while preserving and more clearly communicating this intended contribution.

For ease of review, the principal corresponding revisions in the manuscript are reproduced below:

1) Terminology used throughout the manuscript

“Accumulated-Event Risk Indicator (ARI)”

2) Section 2.3.2 Accumulated-Event Risk Indicator (proposed indicator)

“In the current implementation, reported fatalities determine event eligibility, after which each qualifying event contributes the associated country-level WRI value once. All events receive the same temporal treatment throughout 2013–2023, so their chronological order does not affect ARI. The resulting formulation provides a static retrospective representation of event accumulation over the fixed study period.”

3) Section 1.2 Motivation and Objectives

“Locations experiencing recurrent major events within a relatively short period may warrant closer assessment of preparedness, recovery conditions, and resilience, particularly in developing countries (Ridder et al., 2020; Rentschler et al., 2022; Bündnis Entwicklung Hilft and IFHV, 2023).”

4) Section 5 Discussion:

“Finally, the equal temporal treatment of all qualifying events makes ARI a static retrospective representation of event accumulation.”

“Future research can develop the framework in three main directions. First, dynamic or temporally weighted formulations could incorporate event recency, recovery periods, and evolving vulnerability.”

Comment 3:

The mortality threshold of 50 deaths is arbitrary, and sensitivity analysis shows the top 10 grids change rank under thresholds of 40, 60, and 70 deaths, so the claim of “stability” is overstated.

Response 3:

We sincerely thank the referee for raising two related issues concerning the fatality threshold: the rationale for selecting the 50-death baseline and the interpretation of the previous sensitivity results. We agree that our earlier use of broad expressions such as “stable” and “robust” did not distinguish sufficiently between exact rank invariance and preservation of the principal hotspot structure. We have therefore replaced the previous qualitative discussion with a quantified threshold-sensitivity analysis.

However, we respectfully distinguish a study-specific operational threshold from a threshold selected without methodological rationale. We do not present 50 deaths as a universally optimal or uniquely correct boundary. It was adopted in relation to the specific objective of the study: identifying global hotspots of mortality-intensive major-event accumulation while retaining a sufficiently large and geographically distributed sample for comparative 5°×5° grid analysis.

Reported fatalities were selected as the screening variable because they provide a transparent event-level measure that can be applied consistently across the selected EM-DAT records. The ≥ 50 -death criterion was adopted as a study-specific operational screening rule to focus the analysis on mortality-intensive events while retaining a sufficiently large and geographically distributed sample for global grid-level comparison. We do not claim that 50 deaths is a theoretically unique, statistically optimized, or universally applicable boundary between major and non-major disasters, nor is it an EM-DAT inclusion criterion or a general definition of disaster importance. It is a study-specific operational choice designed for the

present mortality-screened event-accumulation framework.

The revised manuscript now explains the study-specific operational rationale for this threshold and explicitly acknowledges that mortality-based screening influences the composition of the analysed event set. The resulting sample emphasizes mortality-intensive events and underrepresents economic-loss-dominated disasters, indirect cascading impacts, and slow-onset hazard processes. These selection effects define the type of major-event accumulation represented by ARI; they do not make the threshold devoid of methodological rationale.

The relevant sensitivity question is therefore not whether the 50-death threshold can be proven universally optimal, but whether moderate changes around this operational baseline materially alter the principal hotspot output of the study. The previous manuscript addressed this question only through a qualitative comparison of alternative Top-10 rankings. We agree that such a comparison was insufficient to support an unqualified stability claim and have replaced it with a quantitative local threshold-perturbation analysis.

The alternative thresholds of 40, 60, and 70 reported fatalities represent changes of –20 %, +20 %, and +40 %, respectively, relative to the 50-death baseline. They provide one less restrictive and two progressively more restrictive event-selection configurations. These values are now described explicitly as local perturbations of the operational baseline rather than as an undefined “reasonable neighborhood”. The analysis is intended to determine whether the study’s principal hotspot result is strongly dependent on the exact choice of 50 deaths within the tested local range. It is not intended to search exhaustively across every possible threshold, identify a statistically optimal cutoff, or establish invariance under all conceivable event-selection rules.

For each threshold configuration, we reconstructed the qualifying event set and recalculated the corresponding ARI values. We then evaluated the number of grid cells with non-zero ARI, overlap with the baseline Top-14 hotspot set, the retention rate of the baseline Top-14 grids, the mean absolute rank displacement of those grids, and the Spearman rank correlation with the baseline ARI ordering.

The Top-14 set was selected as the primary sensitivity frame because it represents the actual principal hotspot output of the study. These 14 grids were identified in the baseline analysis and subsequently used as the spatial basis of the 3H Dataset. The sensitivity analysis therefore evaluates the persistence of the study-defined priority-grid set rather than introducing an unrelated hotspot definition solely for diagnostic purposes.

Response Table R2. Quantified sensitivity of the principal high-ARI hotspot set to local changes in the 50-death event-selection threshold.

Fatality threshold	Change from baseline	Non-zero ARI grid cells	Overlap with baseline Top-14	Retention rate ^a	Mean absolute rank displacement ^b	Spearman’s rank correlation ^c
50	Baseline	153	14 / 14	100.0 %	0.0	1.000
40	–20%	172	13 / 14	92.9 %	2.0	0.890
60	+20%	132	12 / 14	85.7 %	3.1	0.907
70	+40%	120	12 / 14	85.7 %	3.4	0.880

^a: Retention rate is the proportion of the baseline Top-14 grids that remain within the Top-14 set under each threshold configuration.

^b: Mean absolute rank displacement was calculated for the 14 baseline high-ARI grids.

^c: Spearman rank correlations were calculated over a common grid universe defined as the union of grid cells with non-zero ARI under at least one of the four threshold configurations; cells absent under an individual configuration were assigned an ARI value of zero.

The results show that the event-selection threshold has a measurable effect on the spatial extent and ranking of the ARI results. Lowering the threshold from 50 to 40 deaths increases the number of grid cells with non-zero ARI from 153 to 172, whereas raising it to 60 and 70 deaths reduces the number to 132 and 120, respectively. Individual ARI values and grid

positions are therefore not invariant to the event-selection rule.

At the same time, the principal hotspot structure is largely retained across the tested configurations. Under the 40-death threshold, 13 of the 14 baseline high-ARI grids remain within the alternative Top-14 set, corresponding to a retention rate of 92.9 %. Under both the 60- and 70-death thresholds, 12 of the 14 baseline grids remain, corresponding to retention rates of 85.7 %. The associated mean absolute rank displacements are 2.0, 3.1, and 3.4 positions, while the Spearman rank correlations with the baseline remain between 0.880 and 0.907.

These results reveal a clear distinction between rank-level sensitivity and preservation of the principal hotspot structure. Changing the threshold alters the qualifying event set and therefore changes individual values, exact ranks, and some membership near the Top-14 boundary. It does not, however, completely reorganize the principal hotspot geography within the tested range. Most of the study-defined priority set remains identifiable, and the overall rank correspondence remains high.

Because the referee specifically refers to the Top-10 results reported in the previous manuscript, we also recalculated and retain those results here as a supplementary continuity check. Nine of the ten baseline Top-10 grids remain in the Top-10 set under both the 40- and 60-death thresholds, while eight remain under the 70-death threshold. The corresponding retention rates are 90.0 %, 90.0 %, and 80.0 %, and the mean absolute rank displacements are 1.7, 1.6, and 2.6 positions, respectively. These results are consistent with the primary Top-14 analysis. However, the Top-10 results are reported here specifically to address the referee's reference to the previous analysis; Top-14 is used as the principal sensitivity definition in the revised manuscript because it corresponds directly to the hotspot set on which the subsequent 3H Dataset construction is based.

The evidence therefore does not support exact threshold invariance, but neither does it support the implication that moderate changes around the 50-death baseline substantially overturn the study's principal hotspot result. Within the tested local range, 85.7–92.9 % of the baseline Top-14 grids remain in the corresponding alternative high-value sets, the mean rank displacement remains limited to 2.0–3.4 positions, and the overall Spearman correlations remain between 0.880 and 0.907. The appropriate conclusion is that the principal hotspot structure is largely retained, although individual ranks and exact upper-set membership change.

Accordingly, the revised manuscript no longer describes the 40-, 60-, and 70-death configurations as a vaguely defined “reasonable neighborhood” or implies that they establish the optimality of the 50-death baseline. They are now defined explicitly as local perturbations of a study-specific operational threshold, and the resulting effects are reported quantitatively through changes in spatial coverage, hotspot retention, rank displacement, and rank correlation. Broad claims of exact “stability” or general “robustness” have been replaced by the evidence-based conclusion that the principal hotspot structure is largely retained under the tested threshold changes. This revision recognizes genuine threshold sensitivity while clarifying that the 50-death criterion is an operational choice with an explicit study-based rationale, rather than a universally optimal cutoff.

For ease of review, the principal corresponding revisions in the manuscript are reproduced below:

1) Section 2.1 Disaster event dataset and filtering strategy

“Using a threshold of 50 fatalities (Total Deaths), this study identifies 344 major disaster events for detailed statistical analysis. Mortality is adopted as a study-specific operational screening variable because population-based indicators in EM-DAT, such as “No. Affected” and “No. Homeless”, are often influenced by differences in national definitions and estimation practices (UNDP, 2004; Newman and Noy, 2023; Peduzzi et al., 2009), while economic indicators such as “Total Damage” are strongly shaped by uneven valuation systems, reporting capacity, and development levels across countries (UNDP, 2004; Peduzzi et al., 2009). Mortality therefore provides relatively high cross-country comparability while allowing

a sufficiently large and geographically distributed sample for global grid-level analysis. The influence of the ≥ 50 -fatality criterion is examined through local threshold perturbations of 40, 60, and 70 reported fatalities, as described in Section 2.6 and reported in Section 3.3.2. Consequently, the resulting sample emphasizes mortality-intensive events and underrepresents economic-loss-dominated disasters, indirect cascading impacts, and slow-onset hazard processes.”

2) Section 3.3.2 Sensitivity to the fatality threshold

“The baseline ARI configuration uses a threshold of ≥ 50 fatalities (Total Deaths) to select major disaster events (Section 2.1). To test the sensitivity of ARI to this choice, alternative thresholds of 40, 60, and 70 fatalities are considered. These values represent local changes of -20% , $+20\%$, and $+40\%$, respectively, relative to the operational baseline. For each threshold, the qualifying event set and ARI values are recalculated, and the resulting patterns are compared through the number of non-zero cells, retention of the baseline Top-14 hotspot set, mean absolute rank displacement, and Spearman rank correlation.”

“The event-selection threshold has a measurable effect on the extent of the non-zero ARI surface and on individual grid values and ranks. Lowering the threshold to 40 deaths increases the number of non-zero cells from 153 to 172, whereas thresholds of 60 and 70 deaths reduce the number to 132 and 120, respectively. Nevertheless, 13 of the 14 baseline hotspot cells remain in the alternative Top-14 set under the 40-death configuration, and 12 remain under both more restrictive configurations. The corresponding retention rates are 92.9% , 85.7% , and 85.7% , with mean absolute rank displacements of 2.0–3.4 positions and Spearman correlations of 0.880–0.907. Within the tested local range, the principal hotspot structure remains largely retained, while exact rankings and some membership near the upper-priority boundary vary with the threshold.”

3) Section 5 Discussion

“The sensitivity analyses further show that the broad hotspot geography persists across the tested fatality thresholds, grid resolutions, and Pakistan flood-allocation scenarios. At the same time, changes in exact ranks and boundary membership confirm that ARI values remain dependent on event selection and spatial processing.”

Comment 4:

Point-based geocoding remains a fatal flaw for floods, as demonstrated by the 2022 Pakistan floods, which were basin-wide covering thousands of square kilometers but were reduced to a single coordinate, meaning ARI cannot distinguish between a localized landslide and a catastrophic basin flood.

Response 4:

We sincerely thank the referee for returning to this important spatial-representation issue and for identifying the 2022 Pakistan floods as a particularly demanding example. We agree with the central factual premise that a basin-wide flood cannot be represented as a complete physical footprint by a single coordinate. The point-based procedure used in the present ARI does not preserve inundation extent, affected area, or the internal spatial distribution of impact severity.

However, we respectfully distinguish this acknowledged spatial simplification from the stronger characterization of point-based geocoding as a “fatal flaw” for the intended application of ARI. ARI was not designed to reconstruct disaster footprints or distinguish a localized landslide from a basin-wide flood according to physical extent alone. Its purpose is to translate heterogeneous event-location records into a common global grid framework and identify macro-scale concentrations of qualifying major-event contributions. The relevant question is therefore not whether one coordinate can reproduce every spatial characteristic of an extensive flood—it cannot—but whether concentrating the event contribution in

one grid materially overturns the principal hotspot output of the framework.

The two consequences are now stated explicitly in the Methods and Discussion, while their implications for the sensitivity of the principal priority sets are summarized in the Conclusions. First, a geographically extensive event is reduced to one event-grid association and therefore does not contribute directly to every part of its affected footprint. Second, once an event satisfies the mortality-based inclusion criterion, its ARI contribution is not further weighted according to inundated area, fatalities above the threshold, affected population, economic losses, or another continuous measure of event severity. ARI consequently does not compare qualifying events according to their physical extent or total impact magnitude above the inclusion threshold.

The representative coordinate was not intended to imply that one point reconstructs the hazard footprint. It was adopted as a standardized event-location anchor for translating heterogeneous EM-DAT location descriptions into a common 5°×5° grid framework. ARI should therefore be interpreted as a macro-scale hotspot-screening indicator constructed from event-grid associations rather than as a flood-inundation map, hazard-footprint product, or spatially distributed loss model.

Clarifying the intended scope alone, however, would not determine whether point concentration materially affects the hotspot results. We therefore added an allocation-sensitivity case study for the specific 2022 Pakistan flood event raised by the referee.

We also corrected our previous explanation of multi-grid allocation. Assigning the complete event contribution independently to every affected grid would replicate the same event weight and artificially increase the total ARI. A valid distributed comparison must conserve the original contribution. The revised analysis therefore compares the original point allocation with both an event-excluded diagnostic and a weight-conserving equal fractional-allocation scenario.

The alternative spatial coverage was derived strictly from the affected-location descriptions recorded for the 2022 Pakistan flood event in EM-DAT. The reported provinces and districts were mapped to the ten 5°×5° cells intersecting Pakistan. Nine cells contained at least one location supported by the EM-DAT description, whereas no corresponding affected location was identified for Grid 3.

Response Table R3. EM-DAT-supported grid correspondence for the 2022 Pakistan floods.

Grid	Latitude range (°)	Longitude range (°)	Affected location(s) supported by the EM-DAT description
1	20–25	65–70	Lasbela district (Balochistan province); Sindh province
2	20–25	70–75	Sindh province
3	25–30	60–65	— ^a
4	25–30	65–70	Lasbela, Jhal Magsi, Noshki, Kachhi, Khuzdar and Kalat districts (Balochistan province); Sindh province
5	25–30	70–75	Sindh and Punjab provinces
6	30–35	65–70	Killa Saifullah, Pishin and Chaman districts (Balochistan province); Khyber Pakhtunkhwa province
7	30–35	70–75	Khyber Pakhtunkhwa, Punjab, Azad Jammu and Kashmir provinces
8	30–35	75–80	Punjab, Azad Jammu and Kashmir provinces
9	35–40	70–75	Khyber Pakhtunkhwa, Gilgit Baltistan and Kashmir provinces
10	35–40	75–80	Gilgit Baltistan and Kashmir provinces

^a: No corresponding affected location was identified from the EM-DAT description.

Response Table R3 is not intended to reconstruct the actual flood boundary, imply uniform impacts throughout every listed grid, or estimate the internal distribution of flood severity. Its purpose is to provide a transparent correspondence between the available EM-DAT location information and the grid cells included in the sensitivity calculation.

Three scenarios were then evaluated:

- 1) Original point allocation: the complete Pakistan WRI contribution associated with the event, 26.45, is assigned to the original representative cell, Grid 4 at 25–30°N and 65–70°E.
- 2) Event-excluded diagnostic: the event is removed from the calculation to isolate its contribution to the original ARI pattern.
- 3) Equal fractional allocation: the same total contribution of 26.45 is conserved and divided equally among the nine EM-DAT-supported cells, with each receiving approximately 2.94.

The event-excluded scenario is used only as a diagnostic reference and is not proposed as an alternative geocoding procedure. Likewise, equal fractional allocation is not presented as a reconstruction of inundation extent, mortality distribution, affected population, or spatially varying damage. It provides a deliberately simple, weight-conserving test of allocation sensitivity across the nine supported cells by replacing the maximally concentrated point assignment with a distributed multi-grid alternative.

Response Table R4. Within-Pakistan ARI values and ranks under the three allocation scenarios.

Grid	Original ARI	Original Pakistan rank	Event-excluded ARI	Event-excluded Pakistan rank	Fractional ARI	Fractional Pakistan rank
1	0	7	0	7	2.94	7
2	124.56	4	124.56	4	127.5	4
3	0	7	0	7	0	10
4	185.15	2	158.7	3	161.64	3
5	0	7	0	7	2.94	7
6	20.1	6	20.1	6	23.04	6
7	253.12	1	253.12	1	256.06	1
8	166.08	3	166.08	2	169.02	2
9	80.03	5	80.03	5	82.97	5
10	0	7	0	7	2.94	7

The results confirm that point allocation produces a measurable concentration effect in the representative grid. Under the original configuration, Grid 4 has an ARI value of 185.15 and ranks second among the Pakistan-related cells. Removing the event reduces its ARI to 158.70 and changes its rank to third. Under fractional allocation, its ARI becomes 161.64 and it also ranks third.

The redistribution therefore changes the relative ordering of Grids 4 and 8, but it does not reorganize the principal within-Pakistan hotspot structure. Grid 7 remains the highest-ranked Pakistan cell under all three scenarios, and the membership of the Pakistan Top-5 set remains unchanged. Grids 1, 5, and 10 receive small ARI values of approximately 2.94 under fractional allocation because the EM-DAT description supports affected locations within them. Grid 3 remains at zero because no corresponding affected location was identified.

The total ARI across the ten Pakistan-related cells is approximately 829.04 under the original allocation, 802.59 when the event is excluded, and 829.05 under fractional allocation. The difference of 26.45 between the original and event-excluded scenarios isolates the contribution of the Pakistan flood event, while the negligible difference between the original and fractional totals results only from rounding. The fractional procedure therefore redistributes rather than duplicates the event weight.

Because ARI is a global grid-ranking framework, within-Pakistan persistence alone is not sufficient to evaluate the broader consequence of the allocation assumption. We therefore also recalculated the global ranks of the Pakistan-related cells with non-zero baseline ARI.

The alternative allocation affects several individual global ranks. Grid 4 changes from ninth under the original allocation to

tenth under fractional allocation, while Grid 8 changes from tenth to ninth. Grid 6 moves from 106th to 93rd after receiving a fractional contribution, and Grid 9 moves from 32nd to 30th. The spatial-allocation assumption is therefore consequential at the level of individual grid values and positions.

Response Table R5. Global-rank changes of Pakistan-related grid cells under the three allocation scenarios.

Grid	Original global rank	Event-excluded global rank	Fractional global rank	Change from original to fractional
2	15	15	15	0
4	9	11	10	↓ 1
6	106	106	93	↑ 13
7	3	3	3	0
8	10	9	9	↑ 1
9	32	32	30	↑ 2

Only Pakistan-related cells with non-zero baseline ARI are included in Response Table R5. An upward arrow indicates movement to a higher global position, corresponding to a lower numerical rank; a downward arrow indicates movement to a lower global position.

The principal global hotspot membership, however, is not reorganized across the three configurations. Grid 7 remains third globally, and Grid 2 remains fifteenth. Most importantly, the predefined global Top-14 set remains unchanged under the original, event-excluded, and fractional-allocation configurations. The global Top-5, Top-10, and Top-20 sets also remain unchanged. Grids 4 and 8 exchange their relative positions within the Top-10 set, but neither leaves the set and no new grid enters it.

Grid 4 also remains a high-value location when the concentrated contribution is either removed or redistributed. It ranks third within Pakistan under both alternative scenarios, eleventh globally when the event is excluded, and tenth globally under fractional allocation. The original point assignment therefore amplifies the ARI value of an already high-ranking location rather than creating a hotspot that disappears when the event contribution is redistributed.

The case study consequently establishes a clear distinction between a spatial concentration effect and invalidation of the macro-scale hotspot result. Point-based allocation changes individual ARI values and ranks, as expected. For the specific 2022 Pakistan floods, however, replacing the point assignment with a weight-conserving distribution across nine EM-DAT-supported cells preserves the Pakistan Top-5 set and the global Top-5, Top-10, Top-14, and Top-20 sets. The demanding example raised by the referee therefore does not support the stronger conclusion that point-based geocoding is fatal to the intended hotspot-screening application of ARI.

This single-event analysis is not presented as proof that every extensive flood, storm, or mass-movement event would produce the same degree of priority-set persistence. Equal fractional allocation is itself a stress-test scenario rather than an event-footprint reconstruction. These qualifications define the reach of the case study, but they do not negate its direct finding: in the cited Pakistan case, the point representation produces measurable local concentration without overturning the principal national or global hotspot structure.

Accordingly, the revised manuscript describes the representative coordinate consistently as an event-location anchor, distinguishes full-weight replication from weight-conserving fractional allocation, and reports the Pakistan analysis as an allocation-sensitivity test. It also makes clear that ARI is intended for macro-scale grid-level hotspot screening rather than event-footprint, inundation, hazard-intensity, or spatially distributed loss estimation. Within that intended analytical scope, the new results show that the point-based simplification has identifiable local effects but does not invalidate the principal hotspot output in the challenging case specifically raised by the referee.

For ease of review, the principal corresponding revisions in the manuscript are reproduced below:

1) Section 2.2 Spatial standardization

“For events spanning multiple administrative units, one event-location anchor is assigned to a single grid cell, thereby avoiding double-counting across adjacent grids. This point-based standardization enables uniform global allocation for ARI calculation but simplifies the spatial footprints of geographically extensive events. Table 3 documents representative examples by linking the original location descriptions with the selected sites and coordinates. The implications of this spatial simplification are discussed in Section 5, and its influence on the resulting hotspot pattern is evaluated through an alternative allocation of the 2022 Pakistan flood event in Section 2.6.”

2) Section 2.6 Comparative and sensitivity analyses

“Finally, the influence of point allocation for a geographically extensive event is examined using the 2022 Pakistan floods. The EM-DAT location description supports affected locations across nine 5°×5° cells. Three scenarios are evaluated: 1) the original point allocation, in which the full Pakistan WRI contribution of 26.45 is assigned to the original representative cell; 2) an event-excluded diagnostic, in which the event is removed to isolate its contribution; and 3) an equal fractional allocation, in which the same total contribution is conserved and divided equally among the nine supported cells. The fractional scenario therefore provides a weight-conserving test of allocation sensitivity across the supported cells. ARI values and ranks are recalculated within the Pakistan-related grid set and within the global ranking, and changes in the predefined Top-5, Top-10, Top-14, and Top-20 sets are recorded.”

3) Section 3.3.4 Allocation sensitivity for the 2022 Pakistan floods

“The 2022 Pakistan floods provide a case for examining how point-based event assignment affects the ARI result for a geographically extensive disaster. The original calculation assigns the complete Pakistan WRI contribution of 26.45 to the representative cell at 25–30° N and 65–70° E. Two diagnostic alternatives are examined: removing the event from the calculation and distributing the same total contribution equally among the nine 5° cells supported by the EM-DAT location description. Under the fractional scenario, each of these cells receives approximately 2.94, conserving the total event contribution across the nine cells. Detailed grid-level results are provided in Table B3.”

“The original assignment produces a measurable concentration effect in the representative cell. Its ARI value is 185.15, placing it second among the Pakistan-related cells and ninth globally. Removing the event reduces the value to 158.70, with corresponding ranks of third within Pakistan and eleventh globally. Under equal fractional allocation, its ARI becomes 161.64 and it ranks third within Pakistan and tenth globally.”

“The redistribution changes the relative ordering of this cell and the neighbouring Pakistan-related cell originally ranked tenth globally: the two exchange the ninth and tenth global positions. Several lower-ranked cells also receive small contributions and change rank. The highest-ranked Pakistan-related cell nevertheless remains first nationally and third globally under all three configurations.”

“Most importantly, the Pakistan Top-5 set remains unchanged, as do the global Top-5, Top-10, Top-14, and Top-20 sets. The original point assignment therefore increases the ARI value of an already high-ranking location but does not create a principal hotspot that disappears when the same event contribution is redistributed across the nine supported cells. For this case, the allocation assumption affects individual values and positions without reorganizing the principal national or global priority sets.”

4) Section 5 Discussion

“The sensitivity analyses further show that the broad hotspot geography persists across the tested fatality thresholds, grid

resolutions, and Pakistan flood-allocation scenarios. At the same time, changes in exact ranks and boundary membership confirm that ARI values remain dependent on event selection and spatial processing. Grid refinement primarily localizes the strongest signal within broader hotspot regions, while the Pakistan case illustrates how point allocation can affect local values without necessarily reorganizing the principal priority sets.”

“Assigning one source-constrained event-location anchor to each event enables uniform global grid aggregation, but compresses the spatial extent of geographically extensive hazards and retains some coordinate-level judgement for text-based records.”

5) Table B3 notes

“a: No corresponding affected location was identified from the EM-DAT description.”

“Notes: Only Pakistan-related cells with non-zero baseline ARI are included. An upward arrow indicates movement to a higher global position, corresponding to a lower numerical rank; a downward arrow indicates movement to a lower global position.”

Comment 5:

The representative-point selection logic is not reproducible because criteria such as “proximity to geographic center” and “spatial coincidence with most severe impacts” are subjective, with no algorithm, no inter-rater reliability, and no validation against true footprints.

Response 5:

We sincerely thank the referee for raising the reproducibility of the representative-point selection procedure. We agree that the expressions “proximity to the geographic centre” and “spatial coincidence with the most severe impacts”, as used in the previous manuscript, described the considerations applied during processing but did not constitute a fully deterministic geocoding algorithm. Where an event record contained several affected locations, the available textual information could support more than one plausible event-location anchor, and an independent analyst would not necessarily select exactly the same coordinate.

However, we respectfully distinguish the absence of complete coordinate-level determinism from the stronger implication that the spatial assignments were unconstrained, arbitrary, or unsuitable for the intended analysis. The procedure followed a consistent decision logic designed to translate heterogeneous EM-DAT location information into a common 5°×5° grid framework. It should therefore be characterized as a rule-guided manual spatial-standardization procedure rather than either an automated geocoder or an unconstrained subjective selection process.

This issue is related to, but distinct from, the spatial-footprint question addressed in our response to Comment 4. Comment 4 concerns whether any single coordinate can preserve the complete spatial extent and internal impact distribution of a geographically extensive event. It cannot. The present comment concerns whether the same textual record would necessarily produce the same selected coordinate when processed independently. A procedure may be mechanically repeatable while being spatially implausible, or spatially plausible at the intended scale without being fully deterministic at the coordinate level. We distinguish these questions explicitly in this response. Correspondingly, the revised manuscript separately acknowledges the qualitative judgement that may remain in selecting an event-location anchor and the simplification of geographically extensive event footprints produced by assigning each event to one grid cell.

The representative-point procedure was required because the spatial information available in EM-DAT is heterogeneous. Some selected records, particularly earthquakes, contain explicit latitude–longitude coordinates. Other events, especially

floods, storms, and mass movements, are described through place names, administrative units, or lists of affected areas with substantially different levels of spatial precision.

Where an explicit coordinate was reported in EM-DAT, that coordinate was retained. Where only textual location information was available, the reported affected places and administrative units were reviewed and one practical event-location anchor was selected with reference to their geographic distribution and, where clearly documented in the source information, the principal area of reported impact. The selected location was then converted to geographic coordinates using available geographic and administrative-boundary information and assigned to the corresponding $5^{\circ} \times 5^{\circ}$ grid cell.

This procedure constrains the selection through the reported event geography. It does not guarantee that every ambiguous record has only one defensible point solution, but neither does it permit a coordinate to be selected independently of the locations documented in the source record.

We agree that the text-based selection step may retain qualitative judgement. “Most severe impacts” could not be defined using one globally uniform quantitative variable because the EM-DAT location descriptions do not consistently report fatalities, affected population, physical damage, or economic loss separately for every listed place. Likewise, geographic centrality was used as a practical consideration where the available location description was spatially dispersed rather than as a formally optimized centroid or weighted spatial statistic. We have therefore removed wording suggesting that the procedure was fully objective, automated, or independently validated.

The revised Methods now describe the procedure more accurately. A formally deterministic rule would improve mechanical repeatability, but it would not necessarily improve spatial validity. For example, always assigning the geometric centroid of the reported administrative units would ensure that the same input generated the same coordinate. Where impacts were unevenly distributed within or among those units, however, the centroid might fall far from the principal affected area and imply a level of precision unsupported by the source record. Similar problems could arise from automatically selecting the first listed location or calculating an unweighted mean of all place coordinates. Exact algorithmic reproducibility and spatial plausibility under incomplete information are therefore related but not interchangeable objectives.

The present approach prioritizes a documented and geographically constrained event-location anchor over a mechanically exact point that may be misleadingly precise. Its purpose is not to establish the unique “true” coordinate of a disaster but to support a standardized event-to-grid association at the coarse global scale used by ARI.

No formal inter-rater reliability assessment was conducted for the 344 selected events, and we therefore do not report an operator-agreement statistic or claim exact coordinate-level invariance. The selected points have likewise not been compared globally with one consistent multi-hazard archive of complete event footprints. We state these limitations explicitly in this response. In the revised manuscript, the corresponding methodological boundaries are reflected by acknowledging that more than one plausible anchor may exist, that the manual selection step may involve qualitative judgement, and that point-based standardization simplifies the footprints of geographically extensive events. These limitations do not, however, demonstrate that the implemented decisions were arbitrary or that rule-guided manual standardization is inappropriate for a 5° macro-scale analysis. Inter-rater agreement would measure whether analysts made identical selections; it would not by itself establish which point most accurately represented an extensive event footprint.

The event examples included in the manuscript are therefore presented as illustrations of the spatial-standardization logic rather than as formal global footprint validation. The $5^{\circ} \times 5^{\circ}$ grid scale is also relevant to the practical effect of coordinate-level variation. Where several plausible event anchors fall within the same 5° cell, their coordinate differences do not alter the grid assignment or ARI contribution. The coarse grid can therefore absorb some within-cell variation, although cross-grid uncertainty remains relevant for events whose reported locations span multiple cells.

The allocation-sensitivity analysis presented in our response to Comment 4 addresses a different but complementary question: whether a substantially different spatial allocation for a demanding extensive event materially changes the downstream hotspot output. For the 2022 Pakistan floods, replacing the original point assignment with a weight-conserving allocation across nine EM-DAT-supported cells changes individual ARI values and ranks. However, the Pakistan Top-5 and the predefined global Top-14 sets remain unchanged across the original, event-excluded, and fractional-allocation configurations; the global Top-5, Top-10, and Top-20 sets also remain unchanged. This analysis does not establish that the original point is uniquely reproducible, but it does show that a substantially different plausible allocation does not overturn the macro-scale priority structure in the case raised by the referee.

Coordinate-level determinism, inter-operator agreement, physical footprint validity, and persistence of the grid-level hotspot output are therefore distinct evaluation criteria. The present method does not claim exact point-level reproducibility or complete footprint reconstruction. It does provide a consistent, source-constrained and documented procedure for converting heterogeneous event-location records into standardized 5° grid assignments, which is the function required by the present macro-scale hotspot-screening framework.

Accordingly, the revised manuscript characterizes the spatial-processing procedure as a rule-guided manual standardization method, states explicitly where qualitative judgement may remain, and removes language implying algorithmic or independently validated geocoding. In this response, we further explain why imposing a deterministic centroid or similar mechanical rule would not necessarily produce a more spatially defensible event-location anchor. The Pakistan case is used only to examine the sensitivity of the resulting hotspot output to an alternative spatial allocation. Further operator-agreement or footprint-based assessments could evaluate additional properties of the geocoding procedure, but the absence of those separate evaluations does not invalidate the transparent event-grid standardization implemented for the present study objective.

For ease of review, the principal corresponding revisions in the manuscript are reproduced below:

1) Section 2.2 Spatial standardization

“To support consistent grid-based aggregation, this study employs a rule-guided manual spatial-standardization procedure that converts each major disaster event into a single representative event-location anchor. When precise latitude–longitude coordinates are already provided within EM-DAT, they are retained directly. For events recorded only through textual descriptions, the reported affected places and administrative units are reviewed, and one practical event-location anchor is selected with reference to their geographic distribution and, where clearly documented in the source information, the principal area of reported impact. The selected location is then converted to latitude–longitude coordinates using available geographic and administrative-boundary information.”

“The source records vary considerably in spatial detail, and more than one plausible anchor may exist for an event described through several affected locations. The manual selection step may therefore involve qualitative judgement. Nevertheless, the selected anchor is constrained by the geography documented in the corresponding EM-DAT record and is applied consistently to establish a standardized event–grid association at the analytical scale used in this study.”

“For events spanning multiple administrative units, one event-location anchor is assigned to a single grid cell, thereby avoiding double-counting across adjacent grids. This point-based standardization enables uniform global allocation for ARI calculation but simplifies the spatial footprints of geographically extensive events. Table 3 documents representative examples by linking the original location descriptions with the selected sites and coordinates. The implications of this spatial simplification are discussed in Section 5, and its influence on the resulting hotspot pattern is evaluated through an alternative allocation of the 2022 Pakistan flood event in Section 2.6.”

2) Section 5 Discussion

“Assigning one source-constrained event-location anchor to each event enables uniform global grid aggregation, but compresses the spatial extent of geographically extensive hazards and retains some coordinate-level judgement for text-based records.”

Comment 6:

The 2024 out-of-sample check is not validation but merely a consistency check, because ARI is built on 2013-2023 events and showing that 2024 events also occur in high-ARI grids does not test whether ARI outperforms a null model such as random assignment or population-weighted assignment.

Response 6:

We sincerely thank the referee for drawing this important distinction. We agree that the post-2023 event comparison should not be described as validation in a strict predictive or inferential sense. Demonstrating that later events occur in grids with high historical ARI values establishes temporally separated spatial correspondence, but without an explicit reference it does not determine whether that correspondence exceeds an unstructured spatial expectation or differs from simpler historical indicators.

Although the previous manuscript referred to this analysis as an “out-of-sample plausibility assessment” rather than definitive validation, expressions such as “meaningful external plausibility” could imply a stronger evidentiary status than the analysis established. We have therefore replaced this terminology throughout the revised manuscript with “post-period consistency and benchmark comparison”.

ARI is not intended as a probabilistic model of where the next disaster will occur. It is constructed retrospectively from mortality-screened major disaster events recorded during 2013–2023 and identifies grid cells in which qualifying events accumulated under different country-level WRI contexts. A temporally separated comparison can examine whether later major events remain concentrated in previously identified priority regions and whether this concentration differs from relevant alternatives. It does not estimate future-event probabilities or constitute causal or predictive validation.

We also distinguish between the two reference strategies mentioned by the referee. The equal-grid reference implemented in the revised analysis asks whether the observed concentration within a selected high-value set exceeds the descriptive rate implied by equal selection among the cells in a defined spatial universe. It corresponds to the referee’s concern about an unstructured spatial expectation, but it is not implemented as a stochastic permutation test or formal significance test. A population-weighted reference would instead introduce a substantive exposure model in which expected event concentration is conditioned on population. These are not equivalent reference frameworks. For a multi-hazard event set, disaster location is structured by tectonic setting, river systems, topography, climate, coastlines, and storm tracks, as well as population exposure. Population-weighted event density may therefore be useful for a separate exposure-normalized question, but it is not a necessary substitute for the equal-grid reference implemented here and does not directly test whether ARI differs from WRI, historical event count, or cumulative fatalities.

Changing the terminology alone would not adequately address the underlying concern. We therefore expanded the post-period assessment in three substantive ways. The comparison period was extended from 2024 alone to the complete 2024–2025 period; ARI was compared directly with Equivalent WRI, unweighted major-event count, and death-sum; and an explicitly defined equal-grid reference was introduced using several nested high-value definitions.

- **Expanded post-period event set and global-rank comparison**

The expanded analysis contains 44 major disaster events recorded in Asia during 2024–2025, comprising 23 events in 2024 and 21 events in 2025. The same event-screening criterion used in the principal analysis—at least 50 reported fatalities—was applied. These 44 events occurred in 22 unique 5°×5° grid cells.

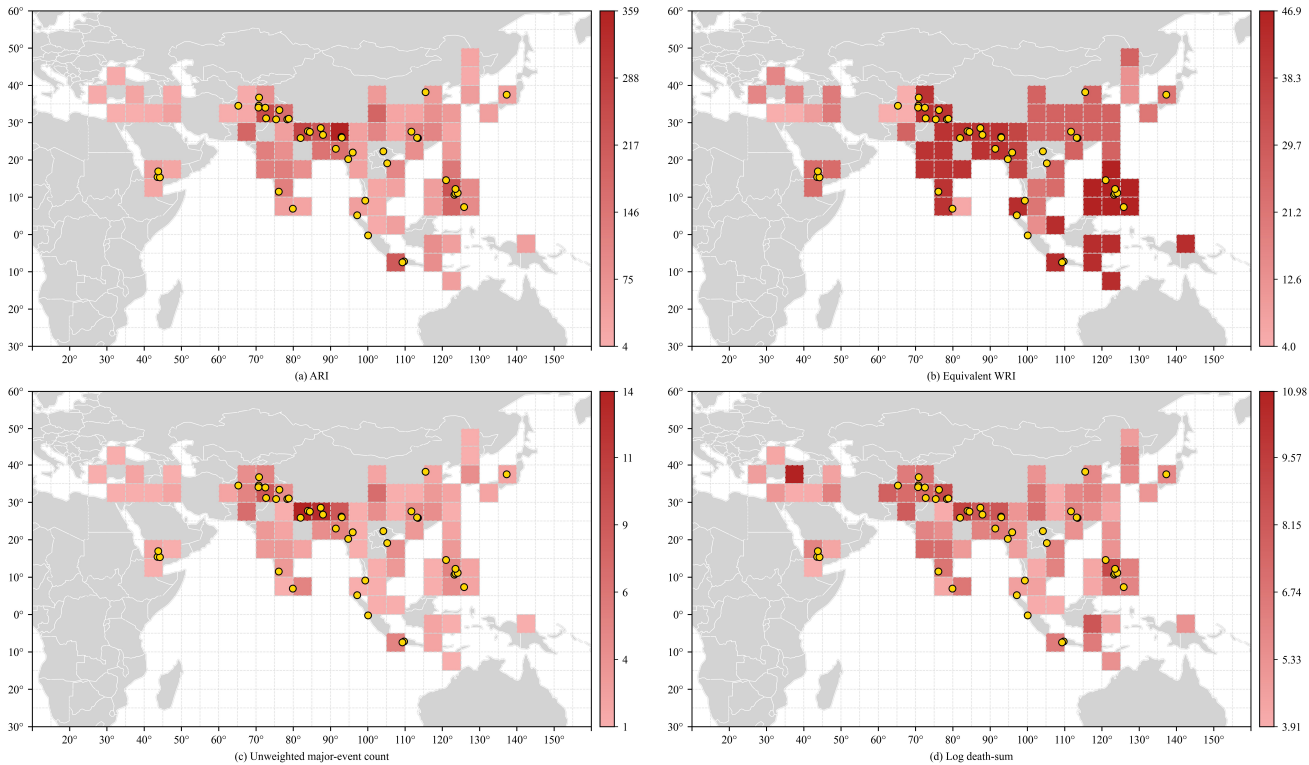
Of the 44 events, 42 occurred in cells included in the common 2013–2023 grid-ranking universe used for the four-indicator comparison. The remaining two occurred in post-period-only cells outside that common historical ranking universe. The complete event-level results are reported in Response Table R6. An en dash indicates that the corresponding post-period event grid was outside the common historical ranking universe and therefore had no comparable prior rank in the table. It does not indicate missing post-period event information or the absence of an underlying country-level WRI value.

Response Table R6. Global baseline grid ranks of the 2024–2025 Asian major disaster events under ARI and three reference indicators constructed from the 2013–2023 data.

Total deaths	Disaster type	Country	ARI rank	Equivalent WRI rank	Unweighted major-event count rank	Death-sum rank
3820	Earthquake	Myanmar	62	42	74	89
2200	Earthquake	Afghanistan	3	18	3	14
1109	Storm	Indonesia	49	8	74	106
1037	Flood	Pakistan	3	18	3	14
823	Storm	Sri Lanka	56	18	74	122
699	Earthquake	Japan	54	97	42	56
460	Storm	Myanmar	11	18	9	33
378	Mass movement	India	15	18	28	23
372	Storm	Philippines	4	1	9	4
368	Flood	Pakistan	3	18	3	14
357	Flood	Afghanistan	32	52	9	22
345	Storm	Viet Nam	–	–	–	–
276	Flood	Thailand	49	8	74	106
268	Flood	Nepal	2	18	1	3
240	Flood	Yemen	43	78	42	47
181	Storm	Philippines	4	1	9	4
150	Flood	Afghanistan	106	69	9	18
140	Flood	India	10	18	16	5
129	Storm	China	20	52	16	42
128	Mass movement	Philippines	24	1	42	51
126	Earthquake	China	5	18	2	8
125	Flood	India	2	18	1	3
105	Mass movement	India	10	18	16	5
98	Flood	India	10	18	16	5
91	Flood	China	37	52	42	58
87	Flood	Indonesia	–	–	–	–
82	Flood	Yemen	43	78	42	47
79	Earthquake	Philippines	4	1	9	4
77	Flood	China	20	52	16	42
71	Flood	Bangladesh	11	18	9	33
65	Flood	China	20	52	16	42
62	Mass movement	Nepal	2	18	1	3
62	Flood	Yemen	43	78	42	47
59	Mass movement	Indonesia	6	8	9	32

58	Flood	Afghanistan	3	18	3	14
57	Flood	Nepal	5	18	2	8
57	Storm	Viet Nam	23	52	16	45
55	Storm	Philippines	4	1	9	4
54	Earthquake	Thailand	62	42	74	89
53	Storm	Philippines	4	1	9	4
52	Flood	India	10	18	16	5
51	Storm	India	1	18	3	6
51	Mass movement	Indonesia	6	8	9	32
50	Flood	India	1	18	3	6

The accompanying four-panel map compares the spatial patterns of ARI, Equivalent WRI, unweighted major-event count, and death-sum with the locations of the 2024–2025 Asian major disaster events.



Response Figure R1. Spatial comparison of (a) ARI, (b) Equivalent WRI, (c) unweighted major-event count, and (d) death-sum constructed from 2013–2023 data, together with the locations of major disaster events recorded in Asia during 2024–2025. Each panel uses a separately normalized colour scale for within-indicator interpretation.

The colour scale in each panel is normalized separately according to the corresponding indicator values. The panels should therefore be interpreted as within-indicator spatial comparisons rather than as direct numerical comparisons of colour intensity across indicators.

The occurrence of 42 of the 44 post-period events in historically ranked cells indicates temporal continuity in the geography of mortality-intensive major disasters. This is relevant descriptive evidence because the indicators were constructed exclusively from 2013–2023 records, whereas the comparison events occurred subsequently. That proportion alone, however, cannot distinguish the contribution of ARI from persistent event geography or unweighted recurrence. We therefore introduced the direct indicator comparisons and equal-grid reference described below.

For each of the 42 post-period events occurring in a historically ranked cell, we recorded the corresponding global rank

under ARI, Equivalent WRI, unweighted major-event count, and death-sum. All four indicators were calculated exclusively from the 2013–2023 data, preserving temporal separation between indicator construction and the 2024–2025 comparison. Equivalent WRI represents the country-level WRI background mapped to the grid framework without historical event accumulation. Event count represents unweighted recurrence of qualifying events, while death-sum represents their cumulative reported fatalities.

Response Table R7. Event-level summary of global baseline grid ranks for the 2024–2025 Asian major disaster events.

Indicator	Mean grid rank	Median grid rank	Events located in globally ranked Top-14 grids, n (%)
ARI	21	10	24 (57.1 %)
Equivalent WRI	28	18	10 (23.8 %)
Unweighted major-event count	21.48	9	22 (52.4 %)
Death-sum	31.31	20	20 (47.6 %)

The Top-14 comparison was retained because it corresponds to the principal global hotspot set used throughout the study and subsequently formed the spatial basis of the 3H Dataset. It therefore evaluates whether later Asian events were concentrated in the same predefined global priority set rather than introducing a new benchmark solely for this comparison.

ARI produced the lowest mean rank, at 21.00, compared with 21.48 for event count, 28.00 for Equivalent WRI, and 31.31 for death-sum. Twenty-four of the 42 rankable events, or 57.1 %, occurred in globally ranked Top-14 ARI grids. The corresponding totals were 22 for event count, 20 for death-sum, and 10 for Equivalent WRI.

Event count produced a slightly better median rank, at 9 compared with 10 for ARI. The results therefore do not indicate that ARI is uniformly more favourable under every summary statistic. They do show that ARI combines the lowest mean rank with the largest number of later events located in the predefined global Top-14 set.

We also conducted paired event-level comparisons between ARI and each baseline.

Response Table R8. Paired event-level comparison of ARI grid ranks with three reference indicators for the 2024–2025 Asian major disaster events.

Paired comparison	ARI has a better rank (lower numerical value)	Equal rank	ARI has a worse rank (higher numerical value)
ARI vs Equivalent WRI	30	0	12
ARI vs. unweighted major-event count	21	4	17
ARI vs. death-sum	31	5	6

ARI assigned a better historical rank than Equivalent WRI for 30 of the 42 rankable events and a better rank than death-sum for 31. The comparison with event count was closer: ARI produced a better rank for 21 events, an equal rank for 4, and a worse rank for 17.

Taken together, Response Tables R7 and R8 show that ARI differs clearly from both the WRI-only and cumulative-fatality patterns. Relative to event count, the comparison is more balanced. This is consistent with the construction of ARI because historical event recurrence is an intended component of the indicator and should therefore provide a strong reference. The contribution of WRI contextualization is not that it removes the recurrence signal, but that it changes the ordering and composition of the upper-priority set. We therefore treat event count as a strong direct comparator rather than requiring ARI to dominate it under every summary measure.

- **Equal-grid reference**

The predefined global Top-14 benchmark and the Top-5, Top-10, and Top-20 definitions used in the equal-grid-reference

analysis serve different purposes. Top-14 maintains continuity with the study’s principal global hotspot set and the 14 grids used in the 3H Dataset. Top-5, Top-10, and Top-20 are nested diagnostic definitions applied within an Asia-specific comparison universe and represent highly selective, intermediate, and broader screening levels.

The Asian comparison universe was defined as the union of the 74 Asian grid cells with non-zero ARI during 2013–2023 and the two additional post-period-only cells containing 2024–2025 major disaster events. This produced 76 unique 5°×5° cells and ensured that later events outside the prior ARI-positive area were retained rather than excluded by construction.

For each indicator, Top-5, Top-10, and Top-20 high-value sets were identified within this universe. All cells tied at a nominal ranking boundary were retained. The expected hit rate under equal grid selection was calculated as the actual number of cells retained in the tie-inclusive high-value set divided by the 76-cell Asian comparison universe.

We report two complementary observed rates. The event-level hit rate is the proportion of all 44 post-period events occurring within the high-value set and therefore retains repeated events in the same grid. The unique-grid hit rate is the proportion of the 22 unique post-period event grids included in the high-value set and counts each grid once. The latter evaluates spatial coverage without amplification from repeated events, while the former is more directly aligned with the event-accumulation objective of ARI.

Response Table R9. Post-period hit rates of high-ARI grid sets relative to an equal-grid reference.

High-ARI definition	Equal-grid expected hit rate	Event-level observed hit rate	Event-level observed-to-expected ratio	Unique-grid observed hit rate	Unique-grid observed-to-expected ratio
Top-5	6.58 %	36.36 %	5.53	22.73 %	3.45
Top-10	13.16 %	50.00 %	3.80	31.82 %	2.42
Top-20	26.32 %	63.64 %	2.42	45.45 %	1.73

Across all three definitions, the observed concentration of post-period events in high-ARI sets exceeded the equal-grid reference expectation under both event-level and unique-grid accounting.

For the Top-5 set, the equal-grid expected rate was 6.58 %, while the observed event-level and unique-grid rates were 36.36 % and 22.73 %. The corresponding observed-to-expected ratios were 5.53 and 3.45. For the Top-10 set, the expected rate was 13.16 %, whereas the observed event-level and unique-grid rates reached 50.00 % and 31.82 %, producing observed-to-expected ratios of 3.80 and 2.42. The Top-10 ARI set therefore contained half of all 44 post-period events while representing only 13.16 % of the defined Asian comparison universe.

The difference between the event-level and unique-grid results is also informative. The ARI Top-10 set contains 7 of the 22 unique post-period event grids, but these 7 grids account for 22 of the 44 events. The unique-grid result shows that the above-reference correspondence is not produced solely by repeatedly counting several events in the same grid. The higher event-level rate additionally shows that the selected high-ARI grids contain a disproportionate share of repeated post-period major-event occurrence.

For the broader Top-20 set, the expected rate was 26.32 %, compared with observed event-level and unique-grid rates of 63.64 % and 45.45 %. The corresponding observed-to-expected ratios remained above the equal-grid reference at 2.42 and 1.73. The progressive decline in observed-to-expected ratios as the retained set expands is consistent with the reduced selectivity of broader screening definitions.

The same equal-grid comparison procedure was applied to Equivalent WRI, unweighted major-event count, and death-sum.

The cross-indicator results reveal a screening-breadth-dependent pattern. At Top-5, ARI produced the highest observed-to-expected ratio among the four implemented indicators under both event-level and unique-grid accounting, at

5.53 and 3.45, respectively. At Top-10, ARI retained the highest event-level observed-to-expected ratio, at 3.80, compared with 3.45 for death-sum, 3.17 for event count, and 1.08 for Equivalent WRI. Under unique-grid accounting, event count was slightly higher than ARI, at 2.59 compared with 2.42. At Top-20, event count was slightly higher under both measures, while ARI remained above Equivalent WRI and death-sum.

Response Table R10. Observed-to-expected ratios for four indicators under the equal-grid reference.

High-value definition	ARI observed-to-expected ratio		Equivalent WRI observed-to-expected ratio		Unweighted major-event count observed-to-expected ratio		Death-sum observed-to-expected ratio	
	Event-level	Unique-grid	Event-level	Unique-grid	Event-level	Unique-grid	Event-level	Unique-grid
Top-5	5.53	3.45	1.48	0.99	3.17	2.30	4.84	2.76
Top-10	3.80	2.42	1.08	0.86	3.17	2.59	3.45	2.07
Top-20	2.42	1.73	1.52	1.26	2.47	1.97	1.99	1.55

These results identify the screening range in which the contextual weighting contributes most visibly. ARI shows its clearest comparative concentration in the most selective upper-priority tail. As the retained set broadens, unweighted event recurrence becomes increasingly competitive. This pattern is consistent with the intended role of ARI: WRI contextualization is most consequential when a limited number of priority grids must be distinguished, whereas the recurrence signal shared by ARI and event count naturally becomes more prominent under broader spatial coverage.

- **Tie-inclusive selection and ranking differentiation**

Because tied values cause the actual number of cells in some nominal Top-N sets to exceed N, we report the tie-inclusive set sizes used in calculating the expected hit rates.

Response Table R11. Actual grid counts retained under the tie-inclusive Top-N definitions.

Indicator	Actual grid count under Top-5	Actual grid count under Top-10	Actual grid count under Top-20
ARI	5	10	20
Equivalent WRI	7	16	33
Unweighted major-event count	6	12	21
Death-sum	5	10	20

ARI retained exactly 5, 10, and 20 cells under the three nominal definitions. By contrast, the Equivalent WRI sets expanded to 7, 16, and 33 cells, while the event-count sets expanded to 6, 12, and 21 because of ties at the ranking boundaries. The expected rates and observed-to-expected ratios in Response Tables R9 and R10 were calculated using these actual retained set sizes rather than assuming that every nominal Top-N definition contained exactly N cells.

The tie structure also has practical relevance for hotspot prioritization. Equivalent WRI inherits extensive ties because multiple grids share the same country-level value, while event count is limited to a small set of integer values. Within the present dataset, ARI provides finer differentiation among the highest-value grids and therefore permits a more selective upper-priority set without retaining large numbers of tied cells.

This is a prioritization-resolution advantage rather than a claim of absolute accuracy. Greater numerical differentiation does not establish correspondence with an independently defined “true” risk surface, but it provides practical value when the objective is to distinguish among candidate priority grids under a selective screening threshold.

- **Interpretation and evidentiary boundary**

The revised analysis implements an equal-grid reference corresponding to the referee’s concern about an unstructured or

uniformly assigned spatial expectation. Across the Top-5, Top-10, and Top-20 definitions, the concentration of 2024–2025 Asian major disaster events in high-ARI grid sets exceeds this descriptive reference under both event-level and unique-grid accounting. The comparison is not presented as a formal stochastic null-model test or statistical significance test.

The direct reference-indicator comparisons also clarify the specific contribution of ARI. Relative to Equivalent WRI, ARI produces substantially greater post-period concentration and finer grid-level differentiation. Relative to death-sum, ARI produces higher observed-to-expected ratios under all tested Top-N definitions. Relative to event count, ARI shows its clearest comparative concentration in the selective Top-5 set and retains the highest event-level concentration at Top-10, while event count is slightly higher under the Top-10 unique-grid measure and the broader Top-20 comparisons.

The evidence therefore supports the following conclusion. The 2024–2025 Asian major disaster events are concentrated in high-ARI grid sets at rates above the implemented equal-grid reference under both event-level and unique-grid accounting. ARI shows its strongest comparative concentration in the selective Top-5 set, retains the highest event-level concentration at Top-10, and remains broadly comparable to unweighted major-event count at Top-20, while producing higher observed-to-expected ratios than Equivalent WRI and death-sum under all tested definitions.

This constitutes temporally separated descriptive benchmark evidence for the intended hotspot-screening function of ARI. It shows that later major events are disproportionately concentrated in the historically identified priority sets and that the resulting pattern is not simply identical to WRI alone, cumulative fatalities, or unweighted recurrence.

The analysis remains a post-period consistency and benchmark comparison rather than predictive validation. This distinction defines the kind of inference that can be made; it does not imply that the analysis lacks evidentiary value. A retrospective hotspot indicator need not predict every individual future event to demonstrate useful prioritization. The relevant finding is that later events are concentrated disproportionately in its historically identified upper-priority sets, particularly in the most selective tail.

The equal-grid reference deliberately addresses the referee’s random-assignment question through a descriptive equal-grid expectation within the defined 76-cell universe. A population-weighted assignment would answer a different exposure-oriented question and is not an equivalent missing reference. Developing such a benchmark would require an explicit choice among total population, population density, population within an event footprint, or another exposure denominator, each of which would produce a different analytical construct. We therefore do not claim comparison with a population-normalized event model, but its absence does not affect the conclusions drawn from the implemented equal-grid reference and the three direct reference indicators.

The two post-period-only event cells further show that ARI does not function as a universal binary predictor and does not imply that previously inactive or lower-ranked cells cannot subsequently experience a major disaster. This is consistent with its stated purpose as a comparative prioritization framework rather than a deterministic forecast.

Accordingly, the revised manuscript replaces validation-like terminology with “post-period consistency and benchmark comparison”, expands the comparison to the complete 2024–2025 period, adds direct comparisons with Equivalent WRI, unweighted major-event count, and death-sum, introduces the explicitly defined 76-cell equal-grid reference, and reports both event-level and unique-grid results using tie-inclusive high-value sets. The revised evidence shows that later Asian major events are concentrated above the equal-grid reference in historically high-ARI sets, with the strongest relative concentration in the selective upper-priority tail and increasing competitiveness of unweighted event count as the retained set broadens.

For ease of review, the principal corresponding revisions in the manuscript are reproduced below:

1) Section 2.1 Disaster event dataset and filtering strategy

“Considering the timeliness of disaster statistics and reliability of historical records, this study uses 2013–2023 as the fixed historical period for constructing ARI and the associated baseline indicators. In addition, earlier EM-DAT records (particularly pre-2000) are known to suffer from reporting inconsistencies and missing entries, especially in developing countries. Therefore, restricting the temporal scope to the past decade ensures higher data completeness and comparability, consistent with the objective of capturing recent major-event accumulation. Records from 2024–2025 are excluded from the construction of ARI and are used only for the temporally separated post-period consistency and benchmark comparison described in Section 2.6.”

2) Section 2.6 Comparative and sensitivity analyses

“Second, post-period consistency and benchmark comparison are evaluated using 44 Asian events recorded during 2024–2025 and meeting the same ≥ 50 -fatality screening criterion as the baseline analysis. These events comprise 23 records from 2024 and 21 from 2025, and occur in 22 unique $5^\circ \times 5^\circ$ grid cells. ARI and the three reference indicators used for this comparison are calculated exclusively from 2013–2023 data. For the 42 post-period events located in cells belonging to the common historical ranking universe, the corresponding global ranks under the four indicators are recorded and compared through mean and median rank, coverage by the predefined global Top-14 sets, and paired event-level rank comparisons. The remaining two events occur in post-period-only cells and are retained in the analysis without assigning them a historical rank.”

“An equal-grid reference is additionally implemented within an Asia-specific comparison universe defined as the union of the 74 Asian cells with non-zero ARI during 2013–2023 and the two post-period-only event cells, producing 76 unique cells. For each of the four indicators, nominal Top-5, Top-10, and Top-20 sets are identified within this universe, with all boundary ties retained. The expected hit rate under equal grid selection is calculated as the actual number of cells in each tie-inclusive high-value set divided by 76. Two observed rates are reported: the event-level hit rate, defined as the proportion of all 44 events occurring within the selected set, and the unique-grid hit rate, defined as the proportion of the 22 event-containing cells included in that set. The observed-to-expected ratio is calculated by dividing each observed rate by its corresponding equal-grid expected rate. These quantities are used as descriptive benchmarks of post-period spatial correspondence.”

3) Section 3.3.1 Post-period consistency and benchmark comparison

“The post-period comparison contains 44 major disaster events recorded in Asia during 2024–2025 and meeting the same ≥ 50 -fatality criterion as the baseline analysis. These comprise 23 events in 2024 and 21 in 2025 and occur in 22 unique $5^\circ \times 5^\circ$ cells. All four historical indicators—ARI, Equivalent WRI, unweighted major-event count, and death-sum—are constructed exclusively from the 2013–2023 records. Of the 44 later events, 42 occur in cells included in the common historical ranking universe, while two occur in post-period-only cells without comparable prior ranks. Detailed event-level information is provided in Table B1.”

“The equal-grid comparison uses an Asia-specific universe consisting of the 74 Asian cells with non-zero ARI during 2013–2023 and the two post-period-only event cells, yielding 76 cells. Under the tie-inclusive Top-10 ARI definition, the selected 10 cells represent 13.16 % of this universe but contain 22 of the 44 later events and 7 of their 22 unique event cells. The resulting event-level and unique-grid observed-to-expected ratios are 3.80 and 2.42, respectively.”

“Across the nested screening definitions, ARI produces event-level and unique-grid observed-to-expected ratios of 5.53 and 3.45 at Top-5, 3.80 and 2.42 at Top-10, and 2.42 and 1.73 at Top-20. At Top-5, these are the highest ratios among the four indicators under both accounting approaches. At Top-10, ARI retains the highest event-level ratio, whereas event count is slightly higher under unique-grid accounting. At Top-20, event count is slightly higher under both measures. The complete cross-indicator results and actual tie-inclusive set sizes are reported in Table B2.”

“Overall, the 2024–2025 Asian major events are concentrated above the equal-grid reference in historically high-ARI sets, with the strongest relative concentration in the selective upper-priority tail and increasing competitiveness of unweighted event count as the retained set broadens.”

4) Table B1 note

“Note: The en dash (–) identifies an event located in a post-period-only cell outside that universe and therefore indicates the absence of a comparable prior grid rank; it does not indicate missing event information or the absence of an underlying country-level WRI value.”

5) Table B2, Panel D note

“Note: Expected rates were calculated separately for each indicator using its actual tie-inclusive set size reported in Panel E.”

Comment 7:

No comparison against any existing index is provided, so it is unknown whether ARI identifies different hotspots than a simple event-count grid, population-weighted event density, or the WRI itself.

Response 7:

We sincerely thank the referee for emphasizing the need for a more systematic comparison between ARI and simpler or established reference indicators. We agree that such comparisons are important for determining whether ARI merely reproduces the country-level WRI background, an unweighted event-count surface, or a cumulative-fatality surface.

We respectfully clarify that the previous manuscript was not entirely without comparison to an existing index. Figures 3 and 4, together with the associated discussion, compared ARI with Equivalent WRI, which transferred the country-level WRI background to the same $5^{\circ}\times 5^{\circ}$ grid framework. Nevertheless, we accept the broader substance of the comment: that comparison was primarily visual and descriptive and did not quantitatively determine how strongly the ARI pattern reflected WRI weighting, how strongly it reflected event recurrence, or whether its upper-priority grid set differed from those generated by simpler event-based representations.

We have addressed this issue through two complementary analyses. The global diagnostic comparison reported in our response to Comment 1 evaluates ARI, Equivalent WRI, unweighted major-event count, and death-sum across the same 153 grid cells with non-zero ARI during 2013–2023. The expanded 2024–2025 Asian analysis reported in our response to Comment 6 then examines how the priority grids produced by these four indicators correspond to later major-event locations. The first analysis evaluates construct dependence and baseline differentiation within the historical study period; the second provides a temporally separated post-period benchmark comparison.

The three reference indicators were selected because they correspond directly to the principal alternative interpretations of ARI. Equivalent WRI tests whether ARI merely reproduces its WRI-derived contextual component. Unweighted major-event count tests whether ARI is effectively an event-recurrence surface without a meaningful contribution from WRI weighting. Death-sum tests whether the hotspot pattern primarily reflects cumulative reported fatalities. All four indicators were constructed using the same 2013–2023 period, qualifying event records, event-location anchors, and $5^{\circ}\times 5^{\circ}$ grid framework, so the comparisons isolate differences in indicator construction rather than differences in the underlying event sample or spatial processing.

The complete global results are reported in Response Table R1 under Comment 1 and are not duplicated here. Across the 153-cell comparison set, ARI has Spearman rank correlations of 0.698 with Equivalent WRI, 0.697 with unweighted event

count, and 0.546 with death-sum. These associations are consistent with the mathematical formulation of ARI. WRI is deliberately incorporated as the contextual weight, while qualifying-event recurrence is its historical input. ARI is therefore expected to retain signals from both components.

The corresponding Top-14 hotspot overlaps, however, are substantially lower than the overall rank associations. The Jaccard similarities between ARI and Equivalent WRI, event count, and death-sum are 0.148, 0.450, and 0.333, respectively. The low overlap with Equivalent WRI shows that adding historical event accumulation materially reorganizes the upper-priority portion of the WRI-only pattern. The incomplete overlap with event count shows that WRI weighting changes which recurrent-event grids occupy the highest-priority positions. The comparison with death-sum further shows that mortality-screened event accumulation is not equivalent to summing fatalities, which may be dominated by one or several exceptionally fatal events.

The relationships among the reference indicators reinforce this distinction. Equivalent WRI has Spearman correlations of only 0.234 with event count and 0.231 with death-sum, while the corresponding Top-14 Jaccard similarities are 0.067 and 0.069. Country-level structural risk context, qualifying-event recurrence, and cumulative reported fatalities therefore provide non-interchangeable ranking signals within the implemented framework. ARI was specifically designed to integrate the first two of these dimensions.

Event count and death-sum likewise represent different properties of the historical record. Although their overall Spearman correlation is relatively high, at 0.742, their Top-14 Jaccard similarity is only 0.318. A grid characterized by repeated qualifying events is not necessarily the same as one dominated by the largest cumulative fatality total. This distinction is central to the interpretation of ARI as an indicator of mortality-screened major-event accumulation under different country-level WRI contexts rather than as a cumulative-mortality measure.

We also quantified ranking differentiation because extensive tied values affect the practical ability of an indicator to distinguish among candidate hotspot grids. Across the 153-cell comparison set, ARI produced 104 distinct values, compared with 44 for Equivalent WRI, 9 for unweighted event count, and 120 for death-sum. The proportions of grid cells belonging to tied-value groups were 49.7 % for ARI, 90.2 % for Equivalent WRI, 98.7 % for event count, and 36.6 % for death-sum. The largest event-count tie alone contained 80 grid cells.

Equivalent WRI inherits extensive ties because multiple grids share the same country-level value, while event count is restricted to a small number of integer values. By combining event recurrence with heterogeneous WRI weights, ARI provides substantially finer differentiation than either Equivalent WRI or unweighted event count within the present dataset. This permits a more selective upper-priority set to be identified without retaining very large tied groups.

Finer differentiation should not be confused with greater absolute accuracy. Death-sum, for example, produces more distinct values than ARI but represents a different quantity and may be strongly influenced by the magnitude of individual fatal events. The relevant contribution of ARI is therefore not that it generates the largest possible number of distinct values, but that it combines two substantively different dimensions while providing sufficient ranking resolution for selective hotspot prioritization.

The 2024–2025 Asian comparison under Comment 6 provides complementary evidence. Among the 42 later events occurring in historically ranked grid cells, ARI assigned a better rank than Equivalent WRI for 30 events and a better rank than death-sum for 31. Relative to unweighted event count, ARI assigned a better rank for 21 events, the same rank for 4, and a worse rank for 17. Using the predefined global Top-14 benchmark, ARI covered 24 of the 42 rankable events, compared with 22 for event count, 20 for death-sum, and 10 for Equivalent WRI.

These findings show the clearest differences between ARI and the WRI-only and cumulative-fatality reference patterns. The

comparison with event count is appropriately closer because event recurrence is an intended component of ARI. This does not make the two indicators equivalent. Rather, it shows that WRI contextualization modifies the membership and ordering of the upper-priority grid set while preserving a strong historical recurrence signal.

The equal-grid comparison further identifies where this modification is most consequential. At Top-5, ARI produced the highest event-level and unique-grid observed-to-expected ratios among the four indicators, at 5.53 and 3.45. At Top-10, ARI retained the highest event-level ratio, at 3.80, while event count was slightly higher under the unique-grid measure, at 2.59 compared with 2.42 for ARI. At Top-20, event count was slightly higher under both accounting approaches, although ARI remained above Equivalent WRI and death-sum.

The post-period evidence therefore supports a specific rather than universal comparative conclusion. ARI provides its clearest additional differentiation in the most selective upper-priority tail, whereas unweighted event recurrence becomes increasingly competitive as the retained hotspot set broadens. This is consistent with the intended role of ARI as a context-weighted prioritization framework: the influence of WRI weighting is most visible when a small number of candidate priority grids must be distinguished.

We also carefully considered the referee's suggestion of a population-weighted event-density indicator. Such an indicator could be useful for examining event occurrence or mortality relative to a specified population-exposure denominator. It would not, however, constitute a direct component reference indicator for ARI or an equivalent neutral reference.

A population-weighted indicator would change the analytical quantity being assessed. ARI evaluates the accumulation of qualifying major-event contributions under different country-level WRI contexts. A population-weighted event-density model would instead evaluate event occurrence or mortality relative to total population, population density, affected population, population within an estimated event footprint, or another exposure measure. These alternatives are not interchangeable and would lead to different interpretations.

This distinction is particularly important for a multi-hazard dataset. Population shapes exposure and the potential for a hazard to produce substantial human impacts, but it does not independently determine where earthquakes, floods, storms, or mass movements occur. Their spatial distributions are also governed by tectonic setting, river systems, terrain, climate, coastlines, and storm tracks. Population weighting is therefore an exposure-oriented alternative model rather than a general null expectation for multi-hazard event geography.

The present event-location-anchor procedure also limits the defensibility of a simple grid-population normalization. Dividing point-assigned event counts by the total population of a 5° cell would be computationally straightforward, but for broad floods, storms, and other geographically extensive events, the denominator would not necessarily represent the population exposed to or affected by the event. Such a calculation could introduce an appearance of precision that is not supported by the underlying spatial records.

For these reasons, we have not added a population-weighted indicator solely to expand the number of reference indicators. This is a deliberate construct and scope decision rather than an omission in the set of direct, consistently constructed comparisons implemented here. A defensible population- or exposure-normalized analysis would require an explicitly defined exposure concept and, preferably, more spatially distributed event-footprint information. It would answer a related but different research question.

The implemented comparisons directly address the central question raised by the referee: whether ARI simply reproduces WRI, unweighted event count, or cumulative reported fatalities. ARI remains intentionally associated with WRI and event recurrence by construction, but its upper-priority hotspot set, ranking differentiation, and post-period concentration are not identical to those of any of the three direct reference indicators.

The specific contribution supported by these analyses is contextualized and differentiated hotspot prioritization. WRI weighting changes which recurrent-event grids are elevated to the highest priority, while the historical event signal remains visible. The resulting ARI surface provides finer differentiation than the WRI-only and event-count-only alternatives within the present dataset and shows its clearest comparative post-period concentration in the most selective hotspot sets.

Accordingly, the revised manuscript now supplements the original visual ARI–Equivalent WRI comparison with quantitative global and post-period comparisons against Equivalent WRI, unweighted major-event count, and death-sum. It reports overall rank association, tie-inclusive Top-14 overlap, ranking differentiation, paired later-event ranks, Top-14 event coverage, and equal-grid-reference concentration. The Discussion and Conclusions now present ARI as a spatially explicit complement to country-level structural risk indices whose added role lies in selective grid-level prioritization, rather than as a universally superior replacement for existing indices or for every possible exposure-normalized model.

The global diagnostic comparison reproduced under Response 1 and the post-period four-indicator comparison reproduced under Response 6 also directly address this comment. The definitions and interpretive role of the three reference indicators are stated in the revised manuscript as follows:

1) Section 2.3.2 Accumulated-Event Risk Indicator (proposed indicator)

“For grid cells intersecting multiple countries, each event is linked to the WRI of the country recorded in EM-DAT as part of the affected area, preserving the association between individual disaster records and their national risk contexts. To facilitate comparison with the original country-based WRI, an Equivalent WRI is calculated by assigning each grid cell the maximum WRI of all overlapping countries. This provides a conservative grid-level representation of the highest WRI background within the common spatial framework, without incorporating historical event accumulation. Two additional event-based reference indicators are calculated using the same qualifying events and grid assignments: unweighted major-event count, defined as the number of qualifying events in each grid, and death-sum, defined as the total reported fatalities of those events. Their diagnostic comparisons with ARI are described in Section 2.6 and reported in Section 3.2.”

2) Section 5 Discussion

“The principal contribution of ARI is to provide a spatially explicit complement to country-level structural risk indices. WRI characterizes national exposure and vulnerability conditions, whereas ARI highlights where mortality-screened major disaster events accumulated during 2013–2023 under those broader risk contexts. The diagnostic comparisons show that both event recurrence and WRI contextualization shape the ARI pattern. Unweighted major-event count remains a strong comparator, particularly as the screening set broadens, but WRI weighting changes the ordering and composition of the most selective upper-priority grids. It also reduces the extensive ties associated with country-level WRI values and integer-valued event counts, thereby supporting more differentiated prioritization when only a limited number of regions can be selected for further assessment.”

The revised manuscript does not include a population-weighted event-density indicator because such an indicator would represent a distinct exposure-normalized construct rather than a direct reference indicator for the present ARI formulation. The reasons for this decision are explained in the response above, whereas the manuscript reports the three direct and consistently constructed reference indicators.

Comment 8:

Sensitivity analysis is incomplete because only thresholds of 40, 60, and 70 are tested, and the authors do not justify the “reasonable neighborhood” claim or quantify what percentage of high-ARI grids at 5° resolution remain high-ARI at 2.5°

resolution.

Response 8:

We sincerely thank the referee for identifying two aspects of the previous sensitivity analysis that required clearer treatment. First, the rationale and intended scope of the alternative fatality thresholds were not defined precisely enough. Second, the previous grid-resolution comparison relied mainly on visual interpretation and did not directly quantify the correspondence between the principal high-ARI grid set at 5° and the high-value pattern obtained at 2.5°.

We agree that these points required revision. However, we respectfully distinguish between a sensitivity analysis that is insufficiently defined and one that is incomplete merely because it does not test every possible threshold or resolution. Sensitivity analysis must be proportionate to the methodological choices and substantive output being examined. The purpose here is to determine whether the study's principal hotspot result is substantially altered by specified changes to the event-selection threshold and spatial aggregation unit, not to exhaust the full parameter space or identify a universally optimal configuration.

We have therefore strengthened the analysis in two corresponding ways. The fatality-threshold assessment is now defined explicitly as a local perturbation analysis around the 50-death operational baseline, rather than as an examination of an undefined "reasonable neighborhood". The previous primarily visual grid-resolution assessment has been supplemented with a quantitative parent-child comparison between the 5° and 2.5° grid systems.

Both analyses use the predefined Top-14 high-ARI set as their primary reference. This definition was not introduced retrospectively for the sensitivity tests. The 14 grids constitute the principal hotspot output of the baseline analysis and subsequently define the spatial framework used in constructing the 3H Dataset. Evaluating this same set therefore maintains consistency among the main hotspot result, its sensitivity assessment, and the downstream imagery component.

- **Fatality-threshold sensitivity**

As explained in our response to Comment 3, the ≥ 50 -death criterion is a study-specific operational screening rule rather than a theoretically unique, statistically optimized, or universally applicable boundary between major and non-major disasters. Its rationale is to focus the analysis on mortality-intensive major events while retaining a sufficiently large and geographically distributed event set for global grid-level comparison.

The alternative thresholds of 40, 60, and 70 deaths represent changes of -20 %, +20 %, and +40 %, respectively, relative to the baseline. They provide one less restrictive and two progressively more restrictive event-selection configurations. We have removed the expression "reasonable neighborhood" because it did not define either the tested range or the purpose of the analysis adequately. The revised manuscript now describes these values explicitly as local perturbations of the operational baseline.

These configurations are intended to test whether moderate changes around the chosen threshold materially alter the actual principal hotspot set used in the study. They are not presented as an exhaustive search across all conceivable thresholds, an optimization procedure for selecting 50 deaths, or evidence of invariance over an unrestricted parameter range. The analysis therefore addresses local sensitivity around the operational baseline rather than attempting to establish a universally optimal threshold.

The complete quantified results are reported in Response Table R2 under Comment 3 and are cross-referenced rather than duplicated. Lowering the threshold from 50 to 40 deaths increases the number of grid cells with non-zero ARI from 153 to 172, whereas raising it to 60 and 70 deaths reduces the number to 132 and 120, respectively. The event-selection rule therefore has a measurable effect on the extent of the non-zero ARI surface and on individual grid values and rankings.

At the same time, the principal hotspot structure is largely retained under the tested configurations. Thirteen of the 14 baseline hotspot grids remain within the alternative Top-14 set under the 40-death threshold, corresponding to a retention rate of 92.9 %. Twelve remain under both the 60- and 70-death thresholds, corresponding to retention rates of 85.7 %. The mean absolute rank displacement ranges from 2.0 to 3.4 positions, while the Spearman rank correlations with the baseline range from 0.880 to 0.907.

All 14 baseline hotspot grids retain non-zero ARI values under every alternative threshold. The grids that leave an alternative Top-14 set therefore do not disappear from the ARI surface. Their relative positions shift around the upper-priority boundary as the composition of the qualifying event set changes.

These results distinguish rank-level sensitivity from preservation of the principal hotspot structure. Individual values, exact ranks, and some boundary membership change, as expected when different events enter the calculation. Nevertheless, 85.7–92.9 % of the study-defined Top-14 set remains within the alternative high-value sets, all baseline hotspots remain ARI-positive, and the overall rank correspondence remains high.

The tested configurations therefore do not demonstrate exact threshold invariance, but neither do they substantially reorganize the principal hotspot geography. The relevant evidence-based conclusion is that the main priority structure is largely retained under moderate local changes to the fatality threshold, although individual ranks and exact upper-set membership vary.

• **Quantified 5°–2.5° grid-resolution comparison**

We agree that the previous grid-resolution analysis did not provide a direct quantitative answer to the referee’s question concerning how many of the principal 5° hotspots retain finer-resolution high-value support after refinement to 2.5°.

The revised analysis again focuses on the predefined Top-14 hotspot set because this represents the study’s actual principal output. It is the same set used in the baseline analysis, the fatality-threshold sensitivity assessment, and construction of the 3H Dataset. Introducing numerous unrelated fixed-count or percentile definitions could answer additional diagnostic questions, but it is not necessary for evaluating the persistence of the specific priority set on which the study is based.

The number of cells with non-zero ARI differs between the two resolutions: 153 at 5° and 218 at 2.5°. Absolute ARI values are also not directly comparable across the two configurations because the spatial aggregation units differ. Refining one 5° cell into four 2.5° cells changes both the distribution of event contributions and the ranking universe. We therefore compare relative high-value sets through their parent–child spatial relationships rather than apply one absolute ARI threshold at both scales.

At 2.5° resolution, several cells share the same ARI value at the nominal Top-14 boundary. Applying the tie-inclusive rule used consistently elsewhere in the analysis retains 16 high-ARI 2.5° cells and avoids arbitrarily excluding cells with identical values.

Each 5° cell contains four 2.5° child cells. The child-grid retention ratio is defined as the proportion of high-ARI 5° cells containing at least one high-ARI 2.5° child. This directly evaluates whether the principal coarse-resolution hotspots retain a finer-resolution high-value core. Requiring all four children to remain in the high-value set would test a different proposition—whether the ARI signal is spatially uniform throughout the complete 5° parent cell—rather than whether the broad hotspot remains supported after refinement.

The parent-grid containment ratio evaluates the relationship in the opposite direction: it is the proportion of high-ARI 2.5° cells located within high-ARI 5° parent cells. It therefore determines whether the finer-resolution hotspots remain geographically nested within the baseline priority regions or emerge primarily in unrelated locations.

The results are presented in Response Table R12.

Of the 14 high-ARI 5° cells, 12 contain at least one high-ARI 2.5° child cell. The resulting child-grid retention ratio is 85.7 %, directly quantifying the proportion of the principal coarse-resolution hotspots that retain finer-resolution high-value support.

Viewed in the opposite direction, 14 of the 16 tie-inclusive high-ARI 2.5° cells are located within high-ARI 5° parent cells, producing a parent-grid containment ratio of 87.5 %. The two measures provide complementary evidence: most of the predefined 5° hotspots retain a finer-resolution high-value core, and most of the highest-value 2.5° cells remain embedded within the original 5° priority regions.

Response Table R12. Cross-resolution spatial correspondence of the predefined high-ARI hotspot set between the 5° baseline and the 2.5° grid.

High-ARI definition	High-ARI 5° cells	High-ARI 2.5° cells	Parent-grid containment ratio ^a	Child-grid retention ratio ^b	Mean high-ARI child share ^c
Top-14	14	16	87.5 %	85.7 %	25.0 %

^a: Proportion of high-ARI 2.5° cells located within high-ARI 5° parent cells.

^b: Proportion of high-ARI 5° cells containing at least one high-ARI 2.5° child cell.

^c: Mean proportion of the four constituent 2.5° child cells classified as high ARI across the 14 high-ARI 5° parent cells.

The refined pattern is not identical to the baseline pattern. Two of the 14 high-ARI 5° cells do not contain a Top-14 2.5° child, while two of the 16 high-ARI 2.5° cells occur outside the baseline Top-14 parent regions. This shows that spatial refinement changes some upper-set membership and reveals finer concentrations that are not represented identically at the coarse resolution.

The mean high-ARI child share is 25.0 %, indicating that, on average across the 14 baseline hotspot cells, one of the four constituent 2.5° cells belongs to the tie-inclusive finer-resolution high-value set. This result should not be interpreted as weak support for the 5° hotspots. A 5° hotspot aggregates event contributions across a broad area, and those contributions may be concentrated in only one part of the cell. Refinement is therefore expected to localize the highest-value signal rather than reproduce the complete area of each broad parent cell.

Because Top-14 is applied separately at the two resolutions, the two high-value sets are not area-equivalent. Sixteen 2.5° cells cover a much smaller total area than fourteen 5° cells. The reported measures consequently assess geographical nesting and finer-scale localization, not preservation of an identical high-value area or equality of absolute ARI magnitudes across resolutions.

The cross-resolution results show substantial but incomplete spatial persistence without implying resolution independence. The 5° hotspot set largely identifies the broader regions within which the strongest 2.5° concentrations are located, while the finer grid provides additional localization within and, in two cases, outside those broader regions.

Taken together, the two sensitivity analyses provide a consistent quantitative result. Under the tested fatality-threshold perturbations, 85.7–92.9 % of the baseline Top-14 set remains in the corresponding alternative high-value sets and the rank correlations remain between 0.880 and 0.907. Under refinement to 2.5°, 85.7 % of the baseline 5° hotspots retain a high-value child, while 87.5 % of the high-value 2.5° cells remain nested within baseline hotspot parents.

These findings indicate that the principal hotspot geography is largely retained under the two methodological variations examined. Sensitivity is visible in movements near the ranking boundary and in finer spatial localization, rather than in wholesale displacement of the study's priority regions. This is a stronger and more precise conclusion than either an unqualified claim of “robustness” or the opposing implication that the analysis is inadequate because it does not exhaust

every possible threshold and resolution.

Accordingly, the revised manuscript defines the alternative fatality thresholds as local perturbations around the operational baseline, reports their effects quantitatively through hotspot retention, rank displacement, spatial coverage, and rank correlation, and supplements the former visual grid-resolution comparison with explicit parent–child measures. The resulting evidence shows that the broad hotspot geography is largely retained under the tested configurations while exact ranks, boundary membership, and finer spatial localization remain sensitive to the methodological choices.

The quantified fatality-threshold revisions reproduced under Response 3 also address the first part of this comment. The additional grid-resolution revisions are reproduced below:

1) Section 2.4 Spatial aggregation for ARI computation

“For these reasons, this study adopts a $5^{\circ} \times 5^{\circ}$ latitude–longitude grid as the baseline assessment unit for ARI computation. This resolution is intended to provide a practical balance among global coverage, spatial differentiation, data reliability, and compatibility with remote sensing applications. To examine the sensitivity of this choice, alternative grid sizes of $2.5^{\circ} \times 2.5^{\circ}$ and $10^{\circ} \times 10^{\circ}$ are evaluated in Section 3.3. The quantitative cross-resolution assessment focuses on the exactly nested 5° and 2.5° grids and uses parent–child spatial correspondence measures defined in Section 2.6; the 10° result is retained as a descriptive coarser-resolution comparison.”

2) Section 2.6 Comparative and sensitivity analyses

“Grid-resolution sensitivity is evaluated by recalculating ARI using $2.5^{\circ} \times 2.5^{\circ}$ and $10^{\circ} \times 10^{\circ}$ grids. Because absolute ARI values depend on the aggregation unit, the quantitative 5° – 2.5° comparison uses relative high-value sets and their exact parent–child spatial correspondence. Each 5° cell contains four 2.5° child cells. The child-grid retention ratio is defined as the proportion of the 14 baseline high-ARI 5° cells containing at least one tie-inclusive high-ARI 2.5° child. The parent-grid containment ratio is the proportion of tie-inclusive high-ARI 2.5° cells located within baseline high-ARI 5° parents. The mean high-ARI child share is the average proportion of the four child cells classified as high ARI across the 14 baseline hotspot parents. The 10° calculation provides a descriptive comparison of the effect of spatial coarsening.”

3) Section 3.3.3 Sensitivity to the grid resolution

“Refinement from 5° to 2.5° increases the number of non-zero ARI cells from 153 to 218 and reveals finer spatial variation within the broader hotspot regions. Because the spatial aggregation units and ranking universes differ, cross-resolution comparison focuses on the relative high-value sets and their spatial correspondence. At the nominal Top-14 boundary, tied values result in a tie-inclusive set of 16 high-ARI 2.5° cells.”

“Of the 14 baseline high-ARI 5° cells, 12 contain at least one high-ARI 2.5° child, producing a child-grid retention ratio of 85.7 %. Viewed in the opposite direction, 14 of the 16 high-ARI 2.5° cells lie within baseline high-ARI 5° parents, producing a parent-grid containment ratio of 87.5 %. The mean proportion of high-ARI children across the 14 baseline hotspot parents is 25.0 %, equivalent on average to one of the four constituent 2.5° cells. This pattern indicates that refinement tends to localize the strongest event-accumulation signal within the broader 5° hotspot cells rather than reproduce their complete areas as uniformly high-value zones.”

“The 10° map provides a complementary visual representation of spatial coarsening: fewer non-zero cells and smoother regional gradients are obtained, while the broad concentration across South and Southeast Asia remains visible.”

4) Section 5 Discussion

“The sensitivity analyses further show that the broad hotspot geography persists across the tested fatality thresholds, grid resolutions, and Pakistan flood-allocation scenarios. At the same time, changes in exact ranks and boundary membership

confirm that ARI values remain dependent on event selection and spatial processing. Grid refinement primarily localizes the strongest signal within broader hotspot regions, while the Pakistan case illustrates how point allocation can affect local values without necessarily reorganizing the principal priority sets.”

Comment 9:

The 3H dataset is largely detached from ARI because ARI identifies 14 high-risk grids and the authors then compile imagery within those grids from pre-existing RAMP and DigitalGlobe datasets, meaning ARI merely selected which grids to look at rather than generating new data.

Response 9:

We sincerely thank the referee for raising the relationship between ARI and the 3H Dataset. We agree with the factual premise that ARI does not generate the original satellite observations included in the dataset. The primary satellite observations were produced by the pre-existing RAMP and DigitalGlobe/Maxar open-data resources, while supplementary visible-spectrum imagery was acquired and processed specifically for Grids 3 and 7, for which suitable imagery had not been identified from those selected open-data sources.

We also agree that the imagery does not enter the present ARI calculation and is not used to validate, recalibrate, or update the ARI values. We therefore describe the methodological relationship in this response as one-directional and downstream rather than computationally bidirectional. In the revised manuscript, this sequence is expressed by stating that ARI determines the spatial inclusion framework and that imagery compilation constitutes the subsequent data-development stage.

We respectfully disagree, however, with the stronger characterization of the 3H Dataset as “largely detached” from ARI or with the implication that selecting where to assemble imagery is merely incidental. ARI provides the macro-scale spatial architecture of the dataset. It determines which regions are eligible for inclusion, defines the geographic domains within which imagery is searched, structures the assessment of coverage gaps, determines where supplementary acquisition is required, and supplies the hotspot-grid framework through which the resulting resource is organized and documented.

Taken together, Figure 1 and Section 4 of the revised manuscript present the workflow as follows: global major-event records and country-level WRI context → ARI construction and hotspot identification → imagery search and gap assessment within the selected grids → imagery compilation, supplementation, quality screening, standardization, and hotspot-grid-level organization.

Under this formulation, ARI addresses the macro-scale question of where mortality-screened major disaster events accumulated most strongly during 2013–2023 under different country-level WRI contexts. The 3H Dataset addresses the corresponding downstream question of what standardized high-resolution Earth observation resources can be assembled within those priority regions to support subsequent finer-scale investigation.

The contribution of ARI to the remote-sensing component is therefore spatial prioritization rather than image generation. The principal analysis evaluates 153 cells with non-zero ARI and identifies 14 as the principal high-ARI hotspot set. These 14 grids determine the complete spatial inclusion framework of the 3H Dataset. The imagery search, gap-identification procedure, supplementary acquisition decisions, data organization, and grid-level metadata are all conditioned by this hotspot set.

This relationship is operationally substantive. Without the ARI screening stage, the target regions would have to be selected according to another criterion, such as isolated crisis-data availability, existing open-data coverage, administrative preference, prior familiarity with particular countries, or an externally defined regional list. Each alternative would generate a different

geographic dataset. In the present study, ARI provides an explicit, documented, and consistently applied study-level rule for reducing the global analysis to a tractable set of priority regions.

This prioritization function is especially relevant because globally comprehensive and consistently processed sub-metre imagery is not readily available. High-resolution commercial imagery can be costly, public coverage is geographically uneven, and the available products differ in resolution, format, acquisition context, scene dimensions, and preprocessing status. Compiling equivalent imagery for every global $5^{\circ} \times 5^{\circ}$ grid would be neither technically nor practically proportionate to the scope of this study. ARI supplies the macro-scale basis for deciding where the finite imagery-compilation effort should be concentrated.

A downstream relationship does not require the two components to operate at the same spatial scale. A high-ARI 5° grid represents a broad region in which qualifying major-event contributions accumulated strongly under the associated WRI context. It does not imply that every location within the grid has identical local risk. Similarly, imagery assembled within a hotspot grid need not cover the complete cell or every historical disaster footprint contributing to its ARI value. Depending on source availability, it may represent selected urban, peri-urban, disaster-affected, or contextual areas within the broader priority region.

The 3H Dataset is therefore a hotspot-oriented observational resource rather than a complete areal census of the 14 grids. The revised manuscript now states explicitly that its coverage is targeted, heterogeneous, and spatially partial within and among the hotspot grids. This scale distinction does not weaken the ARI–dataset relationship; it defines how a macro-scale prioritization result is translated into a finer-scale data resource.

We also agree that most primary satellite observations were produced independently of the present study. We have consequently revised any wording that could imply that the authors generated the majority of the original sensor-level data. The data-development contribution instead lies in constructing an ARI-guided and hotspot-oriented standardized imagery resource through the selection, integration, supplementation, screening, processing, and organization of heterogeneous source observations.

The dataset-construction process comprises the ARI-based identification of the 14 target grids, review of available imagery within those regions, identification of suitable-coverage gaps, targeted supplementation in data-scarce grids, image-quality screening, conversion into standardized 256×256 -pixel units, common TIF formatting, and source- and grid-level metadata organization. The final manuscript summarizes this process in five stages: ARI-guided hotspot selection, source imagery collection within the selected grids, visual quality screening, spatial standardization and tiling, and metadata assignment and final dataset organization. These operations transform geographically and technically heterogeneous imagery into a consistent and traceable standardized resource.

The resulting dataset contains 7,583,094 standardized TIF image tiles with spatial resolutions of approximately 0.27–0.54 m, each measuring 256×256 pixels. It contains imagery from selected locations in 14 ARI-identified hotspot grids distributed across ten developing countries. The imagery provides heterogeneous and spatially partial coverage across the 14 grid cells. These characteristics represent a substantial data-integration and standardization contribution even though most underlying observations originated from existing platforms.

RAMP and DigitalGlobe/Maxar provide most of the source observations. The present study contributes their ARI-guided spatial selection, integration, gap supplementation, quality screening, tiling, common formatting, and source- and grid-level organization. The distinction between primary source observations and the standardized dataset assembled in this study is now stated clearly in the revised manuscript.

For most hotspot grids, suitable imagery was identified from the RAMP Building Footprint Training Dataset or the

DigitalGlobe Open Data Program. Two hotspot grids lacked suitable imagery from the selected open-data sources and therefore required supplementary visible-spectrum acquisition. Approximately 50 km² was acquired for each of these grids, with the acquisition centre linked to the qualifying event reporting the highest number of fatalities within the corresponding cell. The coverage and intended role of these additions are addressed in detail in our response to Comment 10.

The supplementary acquisitions further demonstrate the operational role of ARI. These grids were not selected because commercial imagery happened to be available; rather, the lack of suitable imagery was identified after the grids had already been designated as priorities by ARI. Supplementary acquisition was then used to prevent the complete absence of observational material from two members of the predefined hotspot set. The decision to acquire additional imagery therefore followed directly from the ARI-based inclusion framework and subsequent gap assessment.

This event-centred supplementation applies only to the two data-scarce grids. It does not imply that the complete 3H Dataset is uniformly centred on individual disaster events or aligned with complete event footprints. The broader resource remains organized at the hotspot-grid level and reflects the imagery available or acquired within the ARI-prioritized regions.

The imagery does not currently feed quantitatively back into ARI, and the presence of imagery within a grid is not treated as evidence that its ARI value is correct. A direct image-based evaluation would require a different research design involving event-specific and temporally matched imagery, consistently derived damage or exposure variables, comparable spatial aggregation, and quantitative assessment against ARI and its components. Such an analysis is not the purpose of the present dataset-construction component.

This lack of feedback does not make the dataset detached from ARI. It establishes the direction of the linkage: ARI guides the spatial construction of the resource, while the resource provides a downstream observational foundation for subsequent research. One analytical output can legitimately govern the design of a later data product without that product being required to validate the upstream analysis.

The 3H Dataset may support subsequent localized Earth observation and model-development tasks, including built-environment mapping, exposure characterization, damage detection, and recovery monitoring. Section 4 also identifies finer-scale Earth observation and machine-learning studies as intended downstream uses. These are potential applications of the resource rather than image-based analyses completed in the present study.

We have therefore revised the positioning of the two components as follows: ARI is the macro-scale, WRI-contextualized hotspot-screening component of the study. The ARI-guided 3H Dataset extends this screening framework into a finer-scale observational resource by assembling standardized high-resolution imagery from selected locations within the 14 priority grids.

This formulation accepts the factual points that ARI does not create the original satellite observations and that the imagery does not presently validate or recalibrate ARI. It rejects, however, the inference that the dataset is therefore largely detached from the indicator. Its target regions, inclusion framework, search domain, gap-identification process, supplementary acquisition decisions, organizational units, and grid-level metadata all derive from the ARI hotspot results.

Accordingly, the revised manuscript distinguishes the primary satellite observations from the standardized imagery compilation produced through the present workflow and defines ARI as the spatial inclusion framework followed by imagery compilation as the subsequent data-development stage. In this response, we further clarify explicitly that this relationship is one-directional and downstream and does not involve image-based validation, recalibration, or updating of ARI. The combined contribution lies in translating a transparent global hotspot-prioritization framework into a large, standardized, and traceable high-resolution imagery resource for subsequent investigation of the identified regions.

For ease of review, the principal corresponding revisions in the manuscript are reproduced below:

1) Section 2.5 Analytical workflow

“Figure 1 summarizes the analytical workflow from disaster-event selection to ARI construction and remote sensing data integration. The workflow consists of three sequential stages: 1) Identification of major disaster events: EM-DAT records for 2013–2023 are filtered using the ≥ 50 -fatality threshold (Section 2.1), and each selected event is assigned a standardized event-location anchor (Section 2.2). 2) Construction of the ARI: Major disaster events are aggregated within $5^\circ \times 5^\circ$ grid cells and weighted using country-level WRI values (Section 2.3). The grid configuration is described in Section 2.4, and the comparative and sensitivity analyses are presented in Section 2.6. 3) Development of the remote sensing dataset: Existing sub-metre imagery within the high-ARI hotspot grids is compiled and supplemented where necessary to construct the 3H Dataset (Section 4), which supports subsequent finer-scale studies in the selected priority regions.”

2) Section 4 ARI-guided remote sensing data integration and the 3H Dataset

“The ARI results presented in Section 3.2 identify 14 high-ARI latitude–longitude hotspot grids in which mortality-screened major-event contributions accumulated most strongly during 2013–2023 under the associated country-level WRI contexts. These grids provide spatial focal points for finer-scale risk assessment and management, for which remote sensing is widely used to map exposed assets and post-disaster damage.”

“High-resolution remote sensing data, particularly sub-metre visible-spectrum imagery, remain costly and unevenly available, especially in developing countries. Existing open-data resources often focus on large cities or specific crisis events and vary in image type, size, and other attributes, complicating their integration and application. These constraints motivate a targeted effort to compile standardized imagery within the ARI-identified priority regions.”

“In this study, ARI defines the macro-scale hotspot set within which imagery is searched and data gaps are assessed. Remote sensing data are then compiled and supplemented within these locations to form the 3H Dataset, which is organized at the hotspot-grid level as a standardized resource for subsequent finer-scale Earth observation and machine-learning studies. Accordingly, ARI determines the spatial inclusion framework, while imagery compilation constitutes the subsequent data-development stage.”

3) Section 4.1 Data sources and spatial coverage

“Using these sources, sub-metre visible-spectrum images were compiled for selected areas within the 14 high-ARI hotspot grids, as summarized in Table A2 in Appendix A. Most grids are covered by one or several DigitalGlobe/Maxar crisis datasets, while some, especially those in Bangladesh, include multiple RAMP tiles covering urban and peri-urban areas. The spatial coverage reflects both the concentration of ARI hotspots mainly in South Asia, Southeast Asia, and the Himalayan belt and the geographically uneven availability of open high-resolution imagery. Consequently, coverage is targeted and heterogeneous within and among the hotspot grids.”

“The primary satellite observations were produced by RAMP and DigitalGlobe/Maxar.”

4) Section 4.3 Standardization and construction of the 3H Dataset

“To create a coherent and user-friendly dataset, all imagery from RAMP, DigitalGlobe/Maxar, and the supplementary acquisitions was subjected to a common standardization workflow. The overall process can be summarized as follows: 1) ARI-guided hotspot selection; 2) source imagery collection within the selected grids; 3) visual quality screening of candidate scenes; 4) spatial standardization and tiling; and 5) metadata assignment and final dataset organization.”

“First, candidate scenes were manually inspected, and imagery with severe artefacts, substantial cloud, cloud-shadow, or haze contamination that prevented reliable interpretation of built-up targets was removed. Scenes with little or no meaningful built-up content were also excluded because they were poorly aligned with the intended downstream applications of the

dataset. Second, the retained scenes were spatially standardized and tiled into fixed-size patches of 256×256 pixels. Third, all tiles were stored in TIF format and annotated with the corresponding hotspot-grid identifier and data source, preserving traceability to their source imagery and ARI-identified grid locations. This workflow was applied consistently across the 14 high-ARI hotspot grids.”

“The final standardized dataset contains 7,583,094 TIF image tiles, each of size 256×256 pixels, with spatial resolutions ranging from 0.27 m to 0.54 m. The complete compilation contains imagery from selected locations in 14 ARI-identified hotspot grids distributed across ten developing countries. The imagery provides heterogeneous and spatially partial coverage across the 14 grid cells. Figure 10 provides representative examples of the imagery, illustrating the diversity of urban forms, building densities, and environmental contexts captured by the dataset.”

“The final data compilation for each high-ARI hotspot grid is summarized in Table A3 in Appendix A. This standardized compilation is referred to as the “3H Dataset”, denoting high-resolution imagery assembled for high-ARI hotspots. By providing standardized imagery from selected locations within these grids, the dataset supports future finer-scale studies such as building damage detection, exposure mapping, and post-disaster recovery monitoring.”

5) Section 5 Discussion

“The ARI-guided 3H Dataset extends this screening framework into a finer-scale observational resource. ARI defines the 14-grid spatial inclusion framework, while the dataset assembles standardized high-resolution imagery from selected locations within those priority regions. Coverage remains heterogeneous because both open-data availability and supplementary acquisition are spatially uneven. The dataset is therefore most appropriate for localized Earth observation and model-development tasks, including built-environment mapping, exposure characterization, damage detection, and recovery monitoring. Its practical value lies in improving the accessibility and technical consistency of high-resolution observations in regions identified through a transparent global prioritization process.”

Comment 10:

The supplementary acquisitions for Grids 3 and 7 total only 100 km², which is vanishingly small relative to 5° × 5° grids that cover approximately 500,000 km² at the equator, and the authors do not explain why this tiny sample is adequate.

Response 10:

We sincerely thank the referee for drawing attention to the scale of the supplementary acquisitions relative to the 5°×5° hotspot grids. We agree with the factual premise that approximately 50 km² of imagery within each grid cannot be regarded as representative coverage of an entire 5°×5° cell. The previous manuscript did not distinguish sufficiently clearly between targeted imagery gap filling and grid-wide spatial representation, and we have revised the description accordingly.

However, we respectfully distinguish the absence of grid-wide representativeness from fitness for the acquisitions’ explicitly limited purpose. The supplementary imagery was not designed as an area-proportional sample of the full grids, a statistically representative survey of their internal conditions, or a comprehensive inventory of land cover, exposure, or disaster impacts. Its purpose was to address a specific source-imagery coverage gap: Grids 3 and 7 were members of the predefined ARI hotspot set, but no suitable imagery for these two grids had been identified through the selected RAMP and DigitalGlobe/Maxar open-data sources. The acquisitions therefore provided localized standardized high-resolution imagery in two priority grids that would otherwise have been entirely absent from the hotspot-oriented compilation.

- **Quantified spatial coverage and intended purpose**

We have now quantified the spatial relationship explicitly. Grids 3 and 7 both extend from 30° to 35°N and span five degrees of longitude. Using a spherical surface-area approximation, each grid covers approximately 260,600 km². An acquisition area of approximately 50 km² therefore represents about 0.019 % of the corresponding grid area.

The approximate 500,000 km² value cited in the comment is larger than the calculated area of either of these specific cells, principally because the east–west width of a five-degree longitude interval decreases with latitude. This numerical correction does not alter the substantive point: the supplementary acquisition covers only a very small fraction of each hotspot grid. The revised manuscript now reports both the approximate grid area and the corresponding coverage percentage so that the approximately 100 km² total cannot be interpreted as broad spatial coverage of the two cells.

The acquisitions were initiated after review of the RAMP and DigitalGlobe/Maxar resources identified no suitable imagery for Grids 3 and 7. Within each grid, the approximately 50 km² acquisition area was centred on the representative location of the qualifying event with the largest reported number of fatalities. This event-based rule provided a transparent and locally relevant anchor for selecting one observational area within an otherwise very large hotspot region.

The acquisition anchor was not intended to identify a statistically representative point within the grid or reconstruct the complete footprint of the associated event. Its role was to provide a consistent and disaster-relevant basis for localized data acquisition where no suitable included imagery was otherwise available. The spatial extent, approximate coverage percentage, and event-centred acquisition logic are now stated in Section 4.2. The resulting standardized tile counts are reported separately in Table A3, while Section 4.3 explicitly characterizes the imagery coverage across the 14 grids as heterogeneous and spatially partial.

The acquisitions were not defined as a fixed percentage of each grid because a larger or proportionally specified area would not, by itself, guarantee representativeness. Even a substantially larger acquisition could remain spatially biased unless it were designed according to a separately defined sampling framework based on built-up area, population distribution, land-cover classes, event footprints, or another explicit spatial stratification.

There is also an important scale distinction between ARI and the imagery resource. The 5° grid is a macro-scale statistical unit used for global hotspot prioritization. It was not intended to be the unit over which complete sub-metre observational coverage would necessarily be assembled. Acquiring even a modest fraction of a roughly 260,600 km² grid at sub-metre resolution would require imagery volumes, acquisition resources, storage, and processing capacity far beyond those needed for the targeted downstream resource developed here.

This scale difference does not make the localized imagery representative of the complete grid. It explains the intended transition between the two components: ARI identifies broad priority regions, while the supplementary acquisitions provide high-resolution observational material at selected locations within two of the data-scarce priority regions.

- **Spatial coverage and standardized data volume**

Following standardization, the supplementary imagery for Grid 3 produced 9,853 tiles, while that for Grid 7 produced 9,982 tiles. Together, the two acquisitions contribute 19,835 standardized 256×256-pixel TIF tiles to the 3H Dataset. Table A3 reports a retained spatial resolution of 0.30 m for both supplementary acquisitions.

Tile count and spatial coverage describe different properties of the resource. The 19,835 tiles indicate the amount of processed high-resolution imagery available for localized analysis. They do not imply 19,835 spatially independent sampling locations, because many adjacent tiles originate from the same acquisition area. We therefore report tile volume and areal coverage separately.

We have removed the previous statement that the supplementation “ensures that the most data-scarce yet high-risk grids are adequately represented”. That wording could be understood as claiming that the imagery represents the complete grids,

which is not the purpose or evidentiary basis of the acquisitions.

The revised interpretation is more precise: the supplementary imagery addresses the absence of suitable included imagery from the selected open-data resources for Grids 3 and 7, demonstrates a transparent ARI-guided gap-filling procedure, and provides georeferenced, standardized, sub-metre imagery within both priority regions. It does not establish grid-wide spatial representativeness or complete internal coverage.

The distinction is important because adequacy must be evaluated relative to the stated objective. Approximately 50 km² is not adequate for characterizing the complete area or heterogeneity of a 5° grid. It is, however, fit for the narrower stated purpose of localized imagery gap filling: ensuring that two ARI-identified priority grids are not wholly excluded from the standardized imagery compilation and that localized high-resolution analysis is possible within them.

- **Position within the 3H Dataset**

As clarified in our response to Comment 9, the 3H Dataset is a hotspot-oriented imagery compilation rather than a spatially exhaustive or uniformly sampled remote-sensing benchmark. This characteristic applies not only to the supplementary imagery for Grids 3 and 7 but also to imagery assembled from the existing RAMP and DigitalGlobe/Maxar resources for the other hotspot grids.

Source availability and spatial coverage vary among the 14 grids. Some contain relatively extensive urban or crisis-event imagery, whereas others contain more localized coverage. The supplementary acquisitions reduce one specific imbalance—the complete absence of suitable included imagery for Grids 3 and 7—but they do not make all hotspot grids equivalent in areal coverage, acquisition timing, event correspondence, spatial resolution, or data volume.

The revised manuscript therefore states that the dataset is organized consistently through ARI hotspot-grid identifiers while the imagery coverage within and among those grids remains heterogeneous. More specifically, Section 4.3 characterizes the imagery as providing “heterogeneous and spatially partial coverage across the 14 grid cells”. The two supplementary acquisitions are identified as localized, event-anchored additions within that broader resource.

Acknowledging their small areal percentage does not make these additions scientifically negligible. Before supplementation, no suitable imagery for Grids 3 and 7 had been identified through the selected open-data sources. After supplementation, each grid contains a georeferenced and standardized sub-metre imagery resource selected through a transparent acquisition rule and incorporated into the same organizational and metadata framework as the remaining hotspot grids.

The contribution is therefore a concrete transition from no suitable imagery identified in the selected open-data resources to an initial observational resource within each of two priority regions. These data may support localized exploratory analysis, built-environment mapping, exposure-feature extraction, preprocessing development, and disaster-related investigation where the acquisition timing and imagery content are appropriate. More extensive spatial sampling would be required for questions specifically concerned with complete within-grid heterogeneity or grid-wide generalization, but those objectives are distinct from the targeted gap-filling function performed here.

Accordingly, the revised manuscript now quantifies the approximately 0.019 % areal coverage, describes the acquisitions explicitly as localized imagery gap filling, and presents the acquisition centres as transparent event-based anchors rather than complete impact-footprint representations. The standardized tile volumes and technical characteristics are reported in Table A3 rather than in Section 4.2. The revised interpretation neither overstates the spatial scope of the acquisitions nor dismisses their actual contribution: together, they provide 19,835 standardized high-resolution tiles and include localized observational material for two ARI-prioritized grids for which no suitable imagery had been identified from the selected open-data resources.

For ease of review, the corresponding revisions in Section 4.2 are reproduced below:

1) Section 4.2 Targeted supplementation in data-scarce hotspot grids

“Although RAMP and DigitalGlobe/Maxar provide imagery for most hotspot grids, two high-ARI hotspot grids (Grids 3 and 7 in Table A2) lack suitable imagery from these selected open-data resources. To address these gaps, additional visible-spectrum satellite images were acquired and processed specifically for these grids. For each of the two grids, imagery covering a rectangular area of approximately 50 km² was obtained, resulting in 100 km² of supplementary data in total. Grids 3 and 7 both extend from 30° to 35° N and span 5° of longitude. Based on a spherical surface-area approximation, each grid covers approximately 260,600 km². Each approximately 50 km² acquisition therefore represents about 0.019 % of the corresponding grid area and provides localized imagery gap filling within that grid.”

“The centre of each acquisition area was selected according to the representative location of the qualifying event reporting the highest number of fatalities within the corresponding hotspot grid. This event-centred rule was applied to the two supplementary acquisitions. The supplementary images were selected to match, as closely as possible, the spatial resolution and spectral characteristics of the existing source imagery and were processed through the same standardization workflow as the RAMP and DigitalGlobe/Maxar materials.”

2) Section 4.3 Standardization and construction of the 3H Dataset

“The final standardized dataset contains 7,583,094 TIF image tiles, each of size 256×256 pixels, with spatial resolutions ranging from 0.27 m to 0.54 m. The complete compilation contains imagery from selected locations in 14 ARI-identified hotspot grids distributed across ten developing countries. The imagery provides heterogeneous and spatially partial coverage across the 14 grid cells.”

Comment 11:

The abstract still overclaims, and the conclusion states that ARI “shows meaningful external plausibility and internal robustness as a comparative hotspot-identification framework”, but this is true only in the weakest sense because it is plausible by construction that high-ARI grids have had more disasters.

Response 11:

We sincerely thank the referee for returning to the wording of the Abstract and Conclusions. We agree that these sections require particularly close alignment with the evidence because they provide many readers with their principal understanding of the study. Although several claims had already been moderated in the previous revision, expressions such as “meaningful external plausibility” and “internal robustness” remained overly evaluative and could imply predictive validation or methodological invariance beyond what the analyses establish. We have therefore removed these expressions and substantially revised both sections.

However, we respectfully distinguish the in-sample property identified by the referee from the full body of evidence now reported in the revision. It is indeed true by construction that the 2013–2023 qualifying events contribute directly to the ARI values of the grids to which they are assigned. Their concentration in high-ARI cells cannot therefore be treated as independent validation. We fully accept this point, and neither the revised Abstract nor the revised Conclusions use the historical concentration of the construction-period events as validation of ARI.

The revised analyses extend beyond restating that in-sample relationship. They provide three analytically distinct forms of information: diagnostic comparison within the historical construction period, temporally separated correspondence with 2024–2025 events, and quantified sensitivity to alternative methodological configurations.

First, the global comparison with Equivalent WRI, unweighted major-event count, and death-sum is presented as a

diagnostic assessment of construct dependence, reference-pattern differentiation, and hotspot overlap. As reported in Response Table R1 under Comment 1, ARI is associated with both its WRI-derived contextual component and historical event recurrence, as expected from its formulation. However, its Top-14 hotspot set does not simply reproduce the Equivalent WRI, event-count, or death-sum pattern. This analysis establishes the specific differentiation introduced by the combined construction; it is not presented as independent validation against a ground-truth disaster-risk surface.

Second, the 2024–2025 Asian analysis is temporally separated from the 2013–2023 construction period and includes an explicitly defined equal-grid reference and three direct historical reference indicators. Later events were not used in calculating ARI. Their concentration in high-ARI sets is therefore not an in-sample result “by construction”.

As reported in Response Tables R6–R11 under Comment 6, the observed event-level and unique-grid hit rates exceed the equal-grid reference expectations under the Top-5, Top-10, and Top-20 definitions. ARI produces its clearest comparative concentration in the selective Top-5 set and the highest event-level concentration at Top-10. Unweighted event count remains a strong comparator and becomes slightly stronger under some broader or unique-grid definitions. These results provide temporally separated descriptive evidence for the hotspot-screening function of ARI, but they do not establish predictive validity, causal performance, or universal superiority over simpler indicators. Persistent hazard geography and historical event recurrence may also contribute to the observed post-period correspondence. The revised manuscript therefore describes this analysis as a post-period consistency and benchmark comparison rather than as validation.

Third, the fatality-threshold and grid-resolution analyses quantify how the principal hotspot result responds when two central methodological choices are changed. As reported in Response Tables R2 and R12, 85.7–92.9 % of the baseline Top-14 grids remain in the alternative threshold-based Top-14 sets, with Spearman correlations of 0.880–0.907. Under refinement to 2.5°, 85.7 % of the baseline 5° hotspots contain a high-value child cell, and 87.5 % of the high-value 2.5° cells remain within baseline hotspot parents.

These results do not establish exact rank invariance, threshold independence, or resolution independence. They show that the principal hotspot structure is largely retained under the tested configurations. We therefore replace the former term “internal robustness” with direct descriptions of hotspot retention, rank sensitivity, and cross-resolution spatial correspondence.

To make these distinctions explicit, the relevant Results subsection has been retitled “Post-period consistency, benchmark comparison, and sensitivity analyses”. The revised Abstract and Conclusions now state what ARI calculates, what the comparisons indicate, and the specific use supported by the evidence, without relying on broad evaluative terms such as “external plausibility” or “robustness”.

The revised Abstract retains the original manuscript’s concise method–result–resource structure while implementing the terminology and evidentiary distinctions established in this revision. It defines ARI as an Accumulated-Event Risk Indicator, identifies the study-specific fatality screening rule and WRI contextualization, and summarizes the diagnostic, post-period, and sensitivity evidence without presenting any of these analyses as independent validation or universal robustness. It states that ARI produces a distinct upper-priority pattern, that the post-period comparison indicates above-benchmark concentration, and that the principal hotspot geography is largely retained under the tested methodological changes. The supporting evidence is deliberately condensed so that the principal emphasis remains on ARI construction, the main hotspot result, and the ARI-guided 3H Dataset.

The revised Conclusions have likewise been shortened and recalibrated. They state the principal numerical findings, acknowledge unweighted event count as a strong comparator, report only that the broader priority pattern shows substantial geographical correspondence across the examined configurations, and explicitly recognize that exact ARI values, rankings, and boundary membership remain dependent on event-selection and spatial-processing choices. The contribution of the 3H Dataset is described as a downstream imagery resource rather than as validation of ARI.

Accordingly, the revision removes “meaningful external plausibility”, “internal robustness”, and related unqualified uses of “robust”, “stable”, and “valid”. Historical event concentration is identified as an in-sample property; the global reference-indicator analysis is presented as a construct-level diagnostic comparison; the 2024–2025 analysis is described as a post-period consistency and benchmark comparison; and sensitivity is reported through direct retention, rank-displacement, and cross-resolution measures. These changes calibrate the claims without accepting the stronger inference that the diagnostic, temporally separated, and sensitivity evidence is entirely reducible to the construction-period relationship identified by the referee.

For ease of review, the revised Abstract, Results subsection title, and Conclusions are reproduced below:

1) Revised Abstract

“This study introduces an Accumulated-Event Risk Indicator (ARI) to identify global hotspots of recent major-event accumulation using 2013–2023 records from the Emergency Events Database (EM-DAT). For each $5^{\circ} \times 5^{\circ}$ latitude–longitude grid cell, ARI sums the country-level World Risk Index (WRI) values associated with events meeting a study-specific threshold of at least 50 reported fatalities, thereby combining recent major-event recurrence with contextual risk conditions. The baseline analysis identifies 153 grid cells with non-zero ARI values and 14 principal hotspots, located mainly in South and Southeast Asia and along the Himalayan belt. Diagnostic comparisons with a grid-level WRI-only reference, unweighted major-event count, and summed reported fatalities show that ARI produces a distinct upper-priority pattern. A temporally separated comparison indicates that major events recorded in Asia during 2024–2025 are concentrated in high-ARI sets at rates above the defined equal-grid benchmark. Sensitivity analyses further show that the principal hotspot geography is largely retained under the tested changes to the fatality threshold, grid resolution, and spatial allocation. The study also presents the 3H Dataset, an ARI-guided compilation of standardized high-resolution remote-sensing imagery assembled from open-data resources and supplementary acquisitions for the 14 hotspot cells. The dataset serves as a downstream observational resource for finer-scale analyses in data-scarce priority regions.”

2) Revised Section 3.3 title

“3.3 Post-period consistency, benchmark comparison, and sensitivity analyses”

3) Revised Conclusions

“This study develops the ARI as a retrospective, grid-based framework for identifying where mortality-screened major disaster events accumulated during 2013–2023 under different country-level WRI contexts. ARI sums the WRI values associated with qualifying events assigned to each $5^{\circ} \times 5^{\circ}$ grid cell. The baseline analysis identifies 153 cells with non-zero ARI values and defines 14 principal hotspot cells, concentrated mainly across South and Southeast Asia and along the Himalayan belt. Diagnostic comparisons show that ARI reflects both historical event recurrence and country-level WRI context while producing an upper-priority pattern distinct from those derived from Equivalent WRI, unweighted major-event count, or cumulative reported fatalities alone. The 2024–2025 Asian comparison provides temporally separated evidence for this hotspot-screening function, with later major events concentrated in historically high-ARI sets at rates above the equal-grid reference, particularly in the selective upper-priority tail. Unweighted event count nevertheless remains a strong comparator as the retained priority set broadens. The sensitivity analyses indicate that the principal hotspot geography is largely retained across the tested fatality thresholds, grid resolutions, and alternative allocation of the 2022 Pakistan flood contribution. Exact ARI values, rankings, and membership near the selection boundary remain dependent on event-selection and spatial-processing choices, but the broader priority pattern shows substantial geographical correspondence across the examined configurations.”

“Building on the 14 ARI-identified hotspot cells, the study also presents the 3H Dataset, comprising 7,583,094 standardized

sub-metre image tiles from selected locations across ten developing countries. ARI provides the macro-scale spatial prioritization framework, while the imagery compilation supports subsequent finer-scale studies of buildings, exposure, disaster-related change, damage, and recovery. Together, these components connect recent major-event accumulation, country-level structural risk context, and high-resolution Earth observation resources within a common priority-region framework.”

Comment 12:

Figure 5 is missing from the PDF, the reference list has formatting inconsistencies, Table 4 uses “” to indicate no prior ARI but the footnote says “-” is used, and supplementary material is not provided despite the authors stating the 3H dataset will be on Zenodo.

Response 12:

We sincerely thank the referee for drawing attention to these technical, formatting, reference, and data-availability matters. We agree that Figure 5 must be reliably visible, that the notation in the post-period table must correspond exactly to its explanatory note, that the reference list must be internally consistent, and that the planned repository release of the 3H Dataset should be described accurately.

We have re-examined the relevant manuscript files and revised each element as described below.

- **Figure 5**

We carefully re-examined the editable manuscript and the files prepared for the present revision. Regardless of whether the earlier problem arose during PDF generation, submission processing, or document rendering, we accept the referee’s substantive observation that Figure 5 was not reliably accessible in the submitted material. The fact that it may have remained visible in an author-retained version does not resolve the problem experienced during review.

For the present revision, Figure 5 was reinserted directly into the editable manuscript. We then regenerated and inspected the document to confirm the visibility of the image, its placement, numbering, in-text citation, caption, and sequence relative to the surrounding figures. In the current revised manuscript, Figure 5 appears in full immediately before Figure 6. The final caption reads:

“Figure 5: Global distribution of ARI in 5°×5° latitude–longitude grids across continental regions.”

For direct inspection, Figure 5 from the revised manuscript is reproduced below.

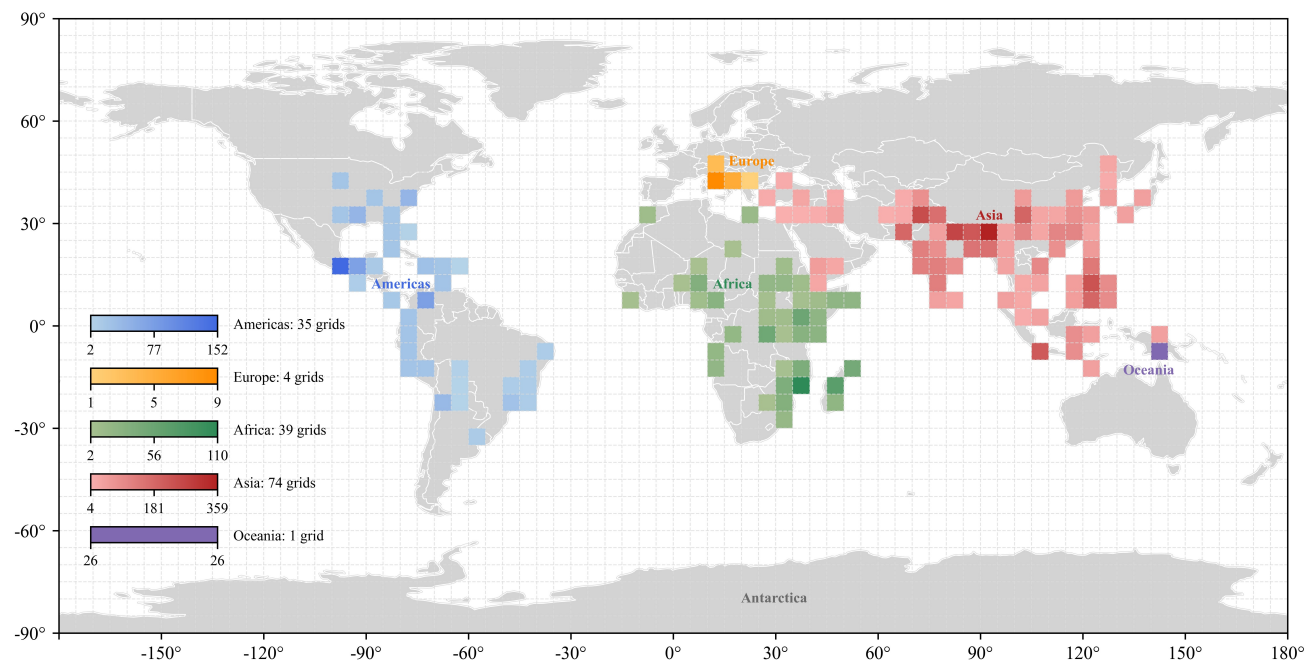
We have therefore addressed the issue as a document-accessibility problem rather than relying on the presence of the figure in an earlier author-retained file. The final submission files will again be checked after PDF generation to ensure that the figure remains visible to editors, referees, and readers.

- **Notation in the post-period results table**

We also re-examined the former Table 4. The post-period analysis has since been expanded from 2024 alone to 2024–2025, extended to four historical indicators, and moved to Table B1. In the author-retained manuscript, the applicable entries were displayed using a dash rather than quotation marks or an intentionally empty string. However, the previous keyboard hyphen was visually ambiguous, and the meaning of the symbol was not explained with sufficient precision.

All applicable entries have therefore been standardized using the typographic en dash “–”. The same en dash is now used in all four rank columns for the two events located in post-period-only cells. The revised table note clarifies that the symbol

concerns membership in the common historical ranking universe used for comparison. It does not indicate missing post-period event information, an absent country-level WRI value, or simply the absence of a qualifying historical event. Accordingly, the table entries and explanatory note now use exactly the same typographic symbol and interpretation.



Response Figure R2. Reproduction of Figure 5 from the revised manuscript.

• **Reference-list formatting**

We appreciate the referee’s continued attention to the reference list. We conducted another systematic review of the bibliography and its correspondence with the in-text citations.

The review covered author names, publication years, titles, journal or institutional-source information, volume and issue information where applicable, page ranges or article numbers, DOI presentation, and consistency between author–year citations in the manuscript and the corresponding reference-list entries. Remaining presentation details, including punctuation, capitalization, page-range en dashes, DOI formatting, line breaks, and the formatting of journal articles, reports, and books, were also harmonized.

Full author lists are retained in the reference list rather than shortened with “et al.” Institutional authors, edited books, technical reports, journal articles, article numbers, page ranges, and DOI formats were checked according to their respective source types. We also reviewed the spelling and ordering of author names and the agreement between publication years in the text and reference list.

The remaining inconsistencies identified during this review were corrected. Figure and table captions, numerical and date ranges, typographic en dashes, negative signs, terminology, and related presentation details were also checked throughout the manuscript.

• **3H Dataset documentation and repository statement**

We agree that the previous wording could be read as implying that a Zenodo repository record was already available. We have therefore revised the Data availability statement to distinguish the documentation included in the manuscript from the planned repository release of the finalized dataset.

The composition and technical characteristics of the 3H Dataset are documented in Section 4.3, Figure 10, and Tables A2–A3, which report the imagery sources and acquisition locations, representative examples, grid-level tile counts,

spatial-resolution ranges, standardized image dimensions and file format, and the principal processing workflow. Because the collection is large and hierarchically organized, it is intended for repository deposition rather than inclusion as conventional manuscript supplementary material.

The finalized 3H Dataset will be deposited in Zenodo before final publication, and the repository DOI and complete dataset citation will be added to the manuscript. Reuse and redistribution of the underlying imagery remain subject to the applicable terms and conditions of the original data providers.

Accordingly, the present revision addresses the four issues directly. Figure 5 has been reinserted, regenerated, rechecked, and reproduced in this response. The post-period table now uses the typographic en dash consistently and defines its meaning in the accompanying note. The reference list and corresponding in-text citations have undergone a systematic consistency review. The Data availability statement has been revised to distinguish the dataset documentation provided in the manuscript from the planned Zenodo deposition of the finalized resource and to state the applicable conditions of the original imagery providers.

For ease of review, the principal corresponding manuscript revisions are reproduced below:

1) Revised Figure 5 caption

“Figure 5: Global distribution of ARI in 5°×5° latitude–longitude grids across continental regions.”

2) Revised Table B1 note

“Note: The en dash (–) identifies an event located in a post-period-only cell outside that universe and therefore indicates the absence of a comparable prior grid rank; it does not indicate missing event information or the absence of an underlying country-level WRI value.”

3) References and technical presentation:

The reference list and corresponding in-text citations have been systematically checked and revised. Full author lists are provided rather than “et al.” in the reference list; author names, publication years, titles, journal or institutional-source information, report and book details, page ranges or article numbers, DOI presentation, and citation–reference correspondence have been reviewed. Figure and table captions, date and numerical ranges, en dashes, negative signs, terminology, and related typographical details have also been checked and standardized throughout the revised manuscript.

4) Revised Data availability statement

“The disaster-event records used in this study were obtained from EM-DAT and remain subject to the database’s applicable access and citation conditions. The composition, sources, and technical characteristics of the 3H Dataset are documented in Section 4.3, Figure 10, and Tables A2–A3. The finalized 3H Dataset will be deposited in Zenodo before final publication, and the repository DOI and complete dataset citation will be added to the manuscript. Reuse and redistribution of the underlying imagery remain subject to the applicable terms and conditions of the original data providers.”

We would like to sincerely thank the editor and Referee #3 once again for the time, care, and critical attention devoted to evaluating our manuscript. The comments raised in this round have helped us make several distinctions that are central to the correct interpretation of the study: deliberate WRI dependence versus circular validation; retrospective major-event accumulation versus dynamic recovery or resilience change; temporally separated benchmark evidence versus predictive validation; preservation of the broader priority geography versus exact rank, threshold, or resolution invariance; source-constrained event-location anchoring versus event-footprint reconstruction; and ARI-guided imagery compilation

versus image-based validation of ARI.

We have carefully considered all twelve comments and revised the manuscript accordingly. The revised version now defines the analytical object, mathematical construction, intended use, and evidentiary boundaries of ARI more precisely; introduces three consistently constructed reference indicators; extends and reframes the post-period comparison; quantifies sensitivity to the fatality threshold, grid resolution, and spatial allocation; and clarifies how ARI guides the spatial organization and downstream development of the 3H Dataset. The Abstract, Results, Discussion, and Conclusions have also been revised so that their claims correspond directly to the analyses reported in the manuscript.

We recognize that the present framework remains selective. It relies on mortality-screened EM-DAT events, country-level WRI context, a fixed historical period, and simplified event-grid associations. It does not reconstruct complete hazard footprints, dynamically changing vulnerability, recovery trajectories, event-specific impact severity above the inclusion threshold, or a universally applicable representation of absolute disaster risk. The post-period comparison is descriptive rather than predictive, and the sensitivity results do not establish exact methodological invariance. These boundaries are now stated explicitly and incorporated into the interpretation of the results rather than treated as secondary qualifications.

Within this clearly defined scope, the principal contribution of ARI is to provide a spatially explicit complement to country-level structural risk indices. It combines recent mortality-screened major-event recurrence with country-level WRI context to generate differentiated global grid-scale priorities, while retaining transparent dependence on both components. The ARI-guided 3H Dataset then extends this macro-scale screening into a downstream observational resource by assembling standardized high-resolution imagery from selected locations within the 14 priority grids. Together, the two components connect retrospective major-event accumulation, structural risk context, and finer-scale Earth observation resources without treating the imagery as validation of the indicator.

We hope that the conceptual clarifications, additional quantitative analyses, and more carefully calibrated interpretations and claims have addressed the referee's concerns and have strengthened the transparency and interpretability of the study. We respectfully submit the revised manuscript for further consideration by **Natural Hazards and Earth System Sciences** and remain grateful for the guidance provided throughout the review process.

Kind regards,

Qingzhao Kong and Erqi Zhu

August 5, 2026