

Response to Review Comments (Round 2)

Dear Editor and Referees,

Thank you for offering us the opportunity to further improve our revised manuscript (**egusphere-2025-2706**). We sincerely appreciate the constructive and insightful comments provided by the referees in this round of review. These comments have been highly valuable for improving the methodological clarity, reproducibility, interpretation, and overall presentation of the manuscript.

Following the editorial decision (**major revision**) and the referees' comments, we have substantially revised the manuscript. The main revisions are summarized as follows:

1) Clarification of the ARI formulation and methodological core. We have rewritten the description of ARI, reformatted the equation, explicitly defined all variables, and clarified that ARI is an event-driven, grid-based, context-weighted relative indicator rather than a complete physical-statistical measure of disaster risk.

2) Improvement of data-processing transparency and reproducibility. We have expanded the description of EM-DAT event filtering, the rationale for using the mortality threshold, representative coordinate assignment, spatial aggregation, grid-resolution selection, and sensitivity analyses.

3) More cautious interpretation and stronger limitation discussion. We have toned down overgeneralized claims, especially those concerning existing risk-assessment frameworks, and added clearer discussion of EM-DAT reporting bias, mortality-threshold bias, spatial simplification, country-level WRI assumptions, and the interpretive boundaries of ARI.

4) Clearer linkage between ARI and the 3H remote sensing dataset. We have revised the manuscript to explain that ARI serves as the hotspot-screening mechanism guiding the spatial selection and organization of the 3H Dataset. We have also clarified the dataset construction workflow, image-quality screening criteria, geographic imbalance, and potential future machine-learning applications.

5) Revision of figures, tables, references, and textual presentation. We have checked and revised figure captions, table descriptions, colour-scale explanations, missing figure issues, citation formatting, reference consistency, typographical errors, and other presentation details throughout the manuscript.

All modifications are clearly marked using track changes in the revised manuscript, and the corresponding revisions are explicitly referenced throughout this response letter for ease of review.

For clarity and ease of cross-reference, we have organized the review comments into 25 numbered points in this response letter. Because the comments from referees were provided mainly in paragraph form rather than as individually numbered comments, we have carefully separated them into specific points according to their content, without changing their meaning. Specifically, Comments 1~15 correspond to the comments from Referee #3, Comment 16 corresponds to the comment from Referee #4, and Comments 17~25 correspond to the comments from Referee #5.

In the following point-by-point response, the referee comments are reproduced in **blue** text, our responses are provided in **black** text, and the corresponding revised manuscript text is highlighted in **red** text with location references where appropriate. We hope that the revisions meet the expectations of the editor and referees, and we are grateful for your time and consideration.

➤ Review Comments of Referee #3:

Comment 1:

This manuscript presents a novel and well-structured approach to disaster risk assessment. The authors propose an Accumulated Risk Index (ARI) that combines recent disaster event data (from EM-DAT) with the established World Risk Index (WRI) to identify spatial hotspots.

Some issues that need further clarification.

The index is defined effectively as a weighted count of recent high-fatality events within a grid cell, where the weight is the country-level WRI. The authors refer to this as accumulated risk, but this definition is single dimensional and as they acknowledge in Equation 1 do not account the elements hazard, exposure, and vulnerability. ARI does not measure the magnitude or severity of each event, nor does it account for the spatial extent of the hazard footprint. The WRI is used as a constant weight per country, not as a dynamic measure of vulnerability.

Response 1:

We sincerely thank the referee for the positive evaluation of our manuscript and for this highly insightful comment. We fully agree that the original presentation of the ARI did not sufficiently distinguish between a classical structural risk model and the more limited, event-driven role that ARI is actually intended to play. The referee is correct that, in its current formulation, ARI does not explicitly reconstruct hazard, exposure, and vulnerability as independent and dynamically interacting components, does not differentiate included events by exact severity, and does not represent the full spatial footprint of each hazard. We are grateful for this comment because it prompted us to sharpen the conceptual scope of the index, moderate our terminology, and improve the theoretical consistency of the manuscript.

In response, we have revised the manuscript to clarify three closely related points.

First, we now define ARI much more explicitly as a complementary, event-driven, grid-based, context-weighted relative index, rather than as a full disaster risk model in the classical sense. Our intention was not to replace structural frameworks such as the WRI, nor to claim that ARI fully operationalizes the conventional hazard-exposure-vulnerability formulation. Rather, ARI was proposed to capture a more specific dimension that is often less visible in country-level frameworks: the recent spatial accumulation of major disaster shocks. We have therefore revised the text to make clear that Equation (1) is used only as a conceptual introduction to the broader disaster-risk field, not as the formal computational basis of ARI.

Second, we have clarified the role of the WRI in the ARI framework. We fully agree with the referee that the WRI is used here as a static country-level contextual weighting factor, not as a dynamic subnational vulnerability surface. This was a deliberate methodological choice. Conceptually, many dimensions central to vulnerability, such as institutional capacity, infrastructure, governance, and adaptive capacity, are most consistently available at national scale in current global datasets. Practically, globally comparable subnational vulnerability data remain very limited, especially in many developing countries that are central to this study. For this reason, we use WRI to provide broad contextual information on exposure and vulnerability conditions under which repeated disaster shocks occur, while avoiding the misleading impression that ARI reconstructs a dynamic vulnerability field at grid level. We have revised the relevant sections to state this role more directly and cautiously.

Third, we have made the current limitations of ARI more explicit. We now state more clearly that, once an event satisfies the major-event criterion, it is treated as one accumulated shock in the present implementation. Accordingly, ARI does not distinguish included events by exact mortality, economic loss, indirect impacts, or other severity measures. We also now state more directly that ARI does not reconstruct the true spatial footprint of hazards, particularly for spatially extensive

events such as floods and storms. These simplifications were not intended to be hidden assumptions. Rather, they reflect the constraints of globally comparable event data and the exploratory purpose of the framework. At the same time, we agree that they should be acknowledged much more transparently in the manuscript, and we have revised the Discussion and Conclusions accordingly.

To address this comment, we have made substantive revisions in the manuscript, especially in Sections 2.3, 3.3, 5, and 6. In particular:

- 1) In Section 2.3, we now state explicitly that Equation (1) is a simplified conceptual expression and that ARI should be interpreted as an event-driven, grid-based, context-weighted relative index, rather than a complete physical-statistical representation of disaster risk.
- 2) In Section 2.3.1, we clarify that the WRI is adopted as a reference index because it provides country-level structural context, but is inherently insufficient on its own for identifying recent grid-level clustering of major disaster impacts.
- 3) In Section 2.3.2, we rewrite the ARI formulation and variable definitions more clearly, and specify that the event-level association with WRI preserves the link between individual disaster records and their national risk contexts.
- 4) In Section 3.3, we further moderate our claims by stating that the strongest conclusion supported by the current evidence is not that ARI has been definitively validated as a complete measure of true disaster risk, but that it shows meaningful external plausibility and internal robustness as a comparative hotspot-identification framework.
- 5) In Section 5 and Section 6, we now state more clearly that ARI does not seek to replace long-term structural risk metrics, but to complement them by making the recent spatial accumulation of major disaster shocks more explicit.

For ease of review, the revised manuscript now includes the following related formulations:

1) *Section 2.3 Risk indices:*

“Equation (1) is a simplified schematic expression used here to illustrate the joint influence of physical processes and socio-economic fragility.”

2) *Section 2.3.1 World Risk Index (reference index):*

“In the present study, WRI is adopted as a reference index because it provides a well-established and annually updated country-level representation of exposure and vulnerability conditions. Its comprehensive indicator system and multi-hazard orientation make it suitable for contextualizing disaster events across regions with different socio-economic backgrounds (Frege et al., 2023; Mucke et al., 2022). However, WRI is not designed to capture the short-term spatial accumulation of recent disaster shocks at subnational or transboundary scales. Methodologically, it is constrained by its country-level assessment unit, and its relatively broad statistics for accumulated impacts, including “Internally displaced persons due to natural disasters” and “Population affected by disasters in the last 5 years” from a perspective of vulnerable groups, which may compromise data reliability as discussed in Section 2.1. Therefore, while WRI provides valuable structural context, it remains inherently insufficient as a standalone metric for identifying recent grid-level clustering of major disaster impacts.”

3) *Section 2.3.2 Accumulated Risk Index (proposed index):*

“Instead, it is adopted as a country-level reference index and contextual weighting factor, encapsulating the broader exposure-vulnerability setting in which disaster events occur. The proposed ARI builds on this reference framework by combining recent major disaster events with national WRI values at the grid scale. Conceptually, ARI integrates two components: 1) the number of major disaster events assigned to each grid cell from 2013 to 2023, based on the event filtering strategy in Section 2.1 and the spatial standardization procedure in Section 2.2; and 2) the WRI value of the country associated with each event, which adjusts for differences in exposure and vulnerability among affected regions. In this way,

ARI emphasizes the short-term accumulation of major disaster shocks in contexts with varying underlying risk conditions. The index is defined as:

$$ARI_j = \sum_{i=1}^{n_j} WRI_{c(i)}, \quad n_j \geq 1, \quad (3)$$

where ARI_j is the ARI of grid cell j , n_j is the number of major disaster events assigned to grid cell j from 2013 to 2023, i indexes the events occurring in that grid cell, $c(i)$ denotes the country associated with the i -th event according to the EM-DAT record, and $WRI_{c(i)}$ is the WRI value of that country. When $n_j = 0$, ARI_j is strictly defined as 0. In this formulation, ARI represents the accumulated sum of country-level WRI weights associated with recent major disaster events occurring within each grid cell, rather than a direct spatial aggregation of WRI itself. ARI should therefore be interpreted as an event-driven, grid-based, context-weighted relative index, rather than a complete physical-statistical representation of disaster risk.”

“This design keeps the interpretation of ARI transparent: once an event satisfies the major-event criterion defined in Section 2.1, it is counted as one accumulated shock, and differences in background risk are reflected through the associated country-level WRI value.”

“The selection of WRI as the reference framework for ARI is supported by four main considerations: 1) WRI is updated regularly through WRR and remains methodologically current; 2) its indicator system spans social, economic, political, and environmental dimensions, making it a comprehensive country-level contextual index; 3) it adopts a multi-hazard perspective rather than focusing on a single hazard; and 4) its quantitative scoring system facilitates comparison and further computation (Frege et al., 2023; Mucke et al., 2022). In ARI, these advantages are retained while shifting the analytical focus from long-term structural risk to the recent accumulation of major disaster events. Although WRI and the present study do not use exactly the same hazard taxonomy, both concentrate on major natural hazards affecting large populations. Given the aggregate role of WRI and the exploratory purpose of ARI, these differences are considered acceptable for the present analysis.”

“This event-level association preserves the link between individual disaster records and their national risk contexts, instead of assigning one single WRI value to the entire grid cell in advance.”

4) Section 3.3 External plausibility assessment and robustness checks:

“The strongest claim supported by the current evidence is therefore not that ARI has been definitively validated as a complete measure of true disaster risk, but that it shows meaningful external plausibility when compared with the spatial occurrence of subsequent major disaster events and remains robust under reasonable variation in key methodological parameters. These results support the use of ARI as a comparative hotspot-identification framework, while detailed local interpretation should still consider the methodological uncertainties discussed in Section 5.”

5) Section 5 Discussion:

“The central contribution of this study lies in reframing global multi-disaster assessment from a purely structural or country-based perspective toward a spatially explicit view of recent accumulated disaster pressure. Existing global indices such as the WRI remain highly valuable because they provide a broad and systematic representation of exposure and vulnerability conditions at the national scale. However, the results of this study suggest that such structural risk context alone may not fully express where major disaster shocks have clustered most intensely in recent years. In this sense, the ARI does not seek to replace long-term structural risk metrics, but to complement them by emphasizing a dimension that is often muted in country-level frameworks: the short-term accumulation of major disaster events across subnational and transboundary space. This complementary role is most clearly reflected in the sharper hotspot contrasts observed in ARI, particularly along the Himalayan belt, South and Southeast Asia, and other cross-border corridors where repeated

high-impact events have occurred within a relatively short period.

From a broader disaster-risk perspective, this distinction is important because resilience deterioration is not only a function of background vulnerability, but also of repeated shock exposure over time. A location may be embedded in a high-vulnerability national context without having experienced concentrated recent shocks, just as a location with only moderate country-level background risk may nonetheless emerge as a hotspot when multiple major events accumulate in close spatial proximity. The comparison between ARI and Equivalent WRI illustrates precisely this point: ARI contributes an additional temporal-spatial layer to global disaster assessment by identifying where recent disaster accumulation may have intensified pressure beyond what long-term structural indices alone would suggest. The added value of the proposed framework therefore lies less in redefining disaster risk in a universal sense than in revealing where recent accumulated impacts may warrant heightened analytical and practical attention.”

“These considerations also make the next steps for future research more concrete. A first priority is to move from the present static formulation toward dynamic or temporally weighted ARI variants, so that more recent shocks, or repeated shocks occurring before recovery is complete, can be represented more realistically.”

“In summary, the present study suggests that recent disaster accumulation deserves more explicit attention in global risk assessment than it typically receives in long-term country-based frameworks. ARI does not resolve the full complexity of global disaster risk, nor is it intended to do so. Its contribution is more focused: to identify where major disaster shocks have recently clustered under different national risk contexts, and to translate that identification into a spatial framework that can support subsequent observation-based analysis. Under that more disciplined interpretation, the value of ARI lies not in replacing existing indices, but in extending the analytical vocabulary available for understanding how repeated disaster shocks are distributed across space and where their accumulated pressure may be most consequential.”

6) *Section 6 Conclusions:*

“This study proposes the ARI as a grid-based and event-driven framework for examining recent accumulated disaster impacts at the global scale. By combining major disaster events recorded in EM-DAT during 2013-2023 with country-level WRI values, ARI adds explicit sensitivity to the spatial accumulation of recent high-impact shocks across subnational and transboundary space. The results reveal clear hotspot concentration, especially across the Himalayan belt, South Asia, and Southeast Asia, and show that ARI provides stronger spatial differentiation than a country-based reference framework. In this sense, ARI is not intended to replace long-term structural risk indices, but to complement them by highlighting where recent major disaster shocks have clustered most strongly.”

We are very grateful for this comment. It has been extremely helpful in prompting us to define the conceptual position of ARI more rigorously, to moderate our claims, and to align the theoretical interpretation of the index more closely with what the current data and methodology can genuinely support.

Comment 2:

The spatial processing of EM-DAT data is a major methodological weakness. The authors state that for events with textual location descriptions, they select a single representative location from the affected area. For events spanning multiple administrative units, they select one representative coordinate to avoid double-counting. This approach introduces several unresolved challenges:

- The loss of spatial information: e.g. a flood that affects an entire river basin (e.g., the 2022 Pakistan floods) is reduced to a single point.
- The criteria for selecting the representative location are not sufficiently detailed. Which site is chosen when multiple

locations are affected? How is the location of an affected area determined for a spatially extensive event?

- By artificially assigning one coordinate, the true spatial footprint is suppressed, and neighboring high-impact areas may be incorrectly assigned zero events.

the examples show that for the 2019 Bahamas storm, the selected coordinate is for Great Abaco Island, but the text description includes both Great Abaco and Grand Bahama. This illustrates the problem of using a single point.

Response 2:

We sincerely thank the referee for this important and highly constructive comment. We fully understand and accept the concern. The spatial standardization of EM-DAT events is a critical step in our framework, because it determines how heterogeneous event records can be linked to a consistent grid-based analytical structure. We agree that, for events recorded only through textual location descriptions, assigning a single representative coordinate inevitably compresses part of the original spatial complexity. This is particularly relevant for spatially extensive hazards such as floods and storms, and we appreciate the referee's request for a clearer explanation of both the rationale and the limitations of this choice.

At the outset, we would like to emphasize that the current point-based allocation method was not adopted because we believe a single coordinate can adequately represent the true spatial footprint of a disaster. Rather, it was adopted as a pragmatic standardization strategy under the constraints of a global multi-hazard dataset. In EM-DAT, while all earthquake events and a limited number of other records include precise coordinates, many events, especially floods and storms, are described only through heterogeneous place names, administrative units, or affected areas in text form. These descriptions vary substantially in precision, spatial scale, and format across countries and disaster types. Without a standardization procedure, it would be extremely difficult to aggregate all events consistently within one global comparative framework.

We fully agree with the referee's first concern, namely the loss of spatial information. A basin-wide flood or a large storm cannot be faithfully represented by a single point. The representative coordinate used in this study should therefore be understood only as a standardized spatial anchor for grid-cell attribution, not as a reconstruction of the actual hazard footprint. We have revised the manuscript to state this distinction much more explicitly and to clarify that the resulting ARI is a relative hotspot-identification tool under simplified spatial assignment, rather than a footprint-based spatial risk product.

We also agree with the referee's second concern that the original manuscript did not describe the representative-point selection logic in sufficient detail. In the revised manuscript, we now explain the procedure more explicitly. When precise latitude-longitude coordinates are already provided in EM-DAT, those coordinates are retained directly. When only textual place descriptions are available, all reported affected locations are first reviewed, and one representative site is selected according to two priority principles: proximity to the geographic center of the overall affected area, and spatial coincidence with the most severe observed impact when such information is available. The coordinates of the selected site are then derived from available administrative boundary data. We agree that this logic needed to be stated much more clearly so that readers can better evaluate the transparency and reproducibility of the procedure.

The referee is also correct to highlight the trade-off involved in assigning only one coordinate to events that affect multiple administrative units. We did not allocate one event to multiple grid cells because doing so would create a different methodological problem: double-counting. If a single disaster were copied into several adjacent cells, one event would no longer count as one event in the ARI framework. Instead, it would contribute multiple times to the global event total, thereby changing the mathematical meaning of the index and weakening comparability across locations and hazard types. We therefore chose a single-cell attribution rule to preserve event-count consistency. At the same time, we fully acknowledge that this choice suppresses part of the true spatial extent of broad events and may underrepresent spillover impacts in neighboring grid cells. We have revised the manuscript to make this trade-off much more explicit.

The referee's example of the 2019 Bahamas storm is especially helpful in clarifying this point. We agree that the event description includes both Great Abaco and Grand Bahama, and that selecting Great Abaco as the representative point compresses a multi-location event into a single spatial anchor. Although, at the scale of our $5^{\circ} \times 5^{\circ}$ grid, both islands fall within the same grid cell and therefore the final ARI assignment is unchanged in this specific case, we fully agree that the broader methodological concern remains valid. The Bahamas example is therefore useful not because it invalidates the criticism, but because it clearly illustrates the boundary between grid-level standardization and true event-footprint representation.

More broadly, this issue is closely tied to the analytical objective of the study. ARI is not intended to reconstruct the detailed spatial geometry of individual disasters. Its purpose is to identify the relative accumulation of major disaster events at the global grid scale under a consistent comparative framework. Because the analytical target is comparative hotspot identification rather than event-specific spatial modeling, a standardized point-based allocation was adopted as a compromise between global consistency and spatial realism. We fully agree, however, that such a choice is only defensible if its limitations are stated plainly and its selection criteria are described transparently.

In response to this comment, we have revised the manuscript in several places:

- 1) In Section 2.2, we now explain the hierarchical representative-point selection procedure more explicitly and state directly that the point-based standardization is intended for uniform global grid allocation rather than footprint reconstruction.
- 2) In Section 2.4, we further clarify that the choice of a $5^{\circ} \times 5^{\circ}$ grid reflects a methodological trade-off: the grid must be large enough to absorb some positional uncertainty and broad event footprints, while still being refined enough to reveal subnational and transboundary clustering.
- 3) In Section 5, we now state more directly that the use of one representative coordinate per event makes global grid-based aggregation feasible, but inevitably compresses the true footprint of geographically extensive hazards such as floods and storms. We also identify finer geocoding, better footprint allocation, and Earth-observation-derived hazard extents as important directions for future improvement.

For ease of review, the revised manuscript now includes the following related formulations:

1) *Section 2.2 Spatial standardization:*

“In addition, certain disaster events, notably floods and storms, may affect broad or spatially discontinuous areas, which makes direct spatial comparisons inherently difficult when only textual location descriptions are available.

To support consistent grid-based aggregation, this study employs a systematic approach to location data processing, converting each major disaster event into a single representative coordinate. When precise latitude-longitude coordinates are already provided within EM-DAT, they are retained directly. For events recorded only through textual descriptions, all reported affected locations are first reviewed, and one representative site is then selected according to two priority principles: proximity to the geographic center of the overall affected area, and spatial coincidence with the most severe disaster impacts. The latitude-longitude coordinates of the selected site are subsequently derived from available administrative boundary data, providing an approximate but standardized basis for disaster location identification.

For events spanning multiple administrative units, the assignment is strictly limited to one representative coordinate to ensure that each event is allocated to a single grid cell, thereby avoiding double-counting across adjacent grids. This point-based standardization is intended to enable uniform global allocation for ARI calculation, rather than to reconstruct the detailed spatial footprint of each disaster event. Therefore, inherent spatial simplifications and uncertainties are unavoidable, particularly for geographically extensive events, and this limitation is further discussed in Section 5.”

2) Section 2.4 Spatial aggregation for ARI computation:

“The statistical computation of ARI depends on the selection of grid cell, which in principle could correspond to different types of spatial assessment units. In this study, however, the objective is to identify recent multi-disaster accumulation patterns at the global scale while retaining a spatial resolution finer than national boundaries. This requires a grid size that is large enough to accommodate the positional uncertainty and broad footprint of some disaster events, but still sufficiently refined to reveal subnational and transboundary clustering.

The choice of spatial unit is therefore a methodological trade-off. If the grid size is excessively large, the main advantage of moving beyond country-level assessment units is weakened, and localized accumulated patterns may be obscured. If the grid size is overly small, two problems become more pronounced. First, due to the extensive impact range of large-scale hazards, which can extend over thousands of square kilometers, it becomes challenging to adequately characterize hazard types (e.g. floods and storms) that affect large areas. This limitation inhibits an effective representation of the independence of regional disaster events. Second, the spatial uncertainty and locational error introduced by representative event coordinates, as described in Section 2.2, will be relatively amplified as cell size decreases.”

“For these reasons, this study adopts a $5^{\circ} \times 5^{\circ}$ latitude-longitude grid as the minimum 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.”

3) Section 5 Discussion:

“At the spatial level, the use of one representative coordinate per event makes global grid-based aggregation feasible, but inevitably compresses the true footprint of geographically extensive hazards such as floods and storms.”

“A second priority is to improve event representation itself through finer geocoding, better footprint allocation, and the integration of Earth-observation-derived hazard extents, especially for spatially extensive disasters.”

We are very grateful for this comment. It has helped us recognize more clearly where the manuscript needed to be more transparent, more precise, and more explicit about the spatial boundaries of the current approach.

Comment 3:

The paper lacks any form of validation. While the sensitivity analyses (Section 3.3) are a solid point, they do not answer whether the ARI actually identifies meaningful “risk” or “accumulated impacts.” The authors do not compare ARI hotspots to any independent measure of disaster impact or risk.

Response 3:

We sincerely thank the referee for raising this fundamental and highly important point. We fully agree with the referee’s core judgment. The sensitivity analyses in the original manuscript were useful in showing that the ARI hotspot pattern was not purely an artifact of one specific parameter choice, but they were not sufficient to demonstrate whether the ARI corresponded to meaningful real-world “risk” or “accumulated impacts”. Strictly speaking, those analyses were robustness checks rather than external validation. We are very grateful to the referee for drawing this distinction so clearly, because it prompted us to improve the evidentiary structure of the manuscript in a more disciplined and transparent way.

We also agree that the comparison between ARI and Equivalent WRI, while useful for interpretation, does not constitute independent validation in the strict sense. Equivalent WRI remains derived from the WRI framework and therefore serves primarily as a reference for comparative interpretation rather than as an external benchmark fully independent of the ARI construction logic. In response to this comment, we have therefore revised the manuscript to distinguish much more clearly

among: 1) internal robustness evidence, 2) comparative support against a related reference framework, and 3) external plausibility assessment.

At the same time, we would like to note why validation is particularly challenging in the present context. ARI is an event-driven, grid-based, multi-hazard index intended to characterize the recent accumulation of major disaster shocks under macro-level national risk conditions. At present, there is no widely accepted global gridded benchmark dataset that directly measures “accumulated disaster impact” or “accumulated resilience erosion” across multiple hazard types in a way that is simultaneously independent, spatially explicit, and globally comparable. For this reason, a conventional one-to-one validation against an undisputed global “ground truth” is difficult. However, as the referee rightly points out, this does not remove the need for stronger external support. Rather, it requires the manuscript to frame the available evidence more carefully and to incorporate a more convincing plausibility assessment.

Motivated by this comment, we have strengthened the manuscript by adding a new out-of-sample external plausibility assessment. Specifically, because the ARI in this study is calculated from disaster events during 2013-2023, we now compare the resulting ARI hotspot pattern against the spatial distribution of major disaster events that actually occurred in 2024, using these post-period events as an external reference set. To keep the analysis focused and tractable while remaining representative, this comparison is conducted for Asia, which is also the region where the ARI map shows the strongest and most continuous hotspot pattern.

Following the same event-selection logic used in the main analysis, major disaster events in Asia during 2024 were extracted from EM-DAT using the criterion of Total Deaths ≥ 50 . Each event was then matched to its corresponding $5^\circ \times 5^\circ$ grid cell. For each such cell, the ARI ranking derived from the 2013-2023 analysis was recorded and compared with the corresponding country-level WRI ranking. The resulting event list is now presented in Table 4, and the spatial correspondence between the 2024 major events and the Asia ARI map is shown in Figure 7.

This additional assessment responds to the referee’s concern more directly than the previous presentation. Importantly, we do not claim that the 2024 events constitute a perfect “ground truth”, nor that this exercise provides definitive validation in the strongest predictive sense. Instead, we present it as an external plausibility assessment: if ARI is meaningfully capturing zones of recent accumulated disaster pressure, then it is reasonable to expect that a substantial proportion of subsequent major disaster events would occur in areas that already exhibit relatively high ARI values.

The result of this comparison is encouraging while also clarifying the conceptual boundary of the index. Most major disaster events in Asia during 2024 occurred in grid cells with relatively high ARI rankings, which supports the interpretation that ARI identifies locations subject to meaningful recent accumulated disaster pressure. In addition, the comparison suggests that ARI can reveal locally important or transboundary hotspot structures that are less readily apparent from country-level WRI alone. For example, some 2024 events in Nepal occurred in a grid cell with a very high ARI ranking despite Nepal’s comparatively modest WRI ranking, because that grid also accumulates major events from neighboring India and therefore captures a transboundary accumulated-impact corridor that a country-based index alone cannot easily express.

At the same time, the new comparison also clarifies what ARI is not intended to do. A small number of major 2024 events occurred in grid cells with no prior ARI value. These cases do not necessarily contradict the logic of the index; rather, they show that ARI is not designed as a universal predictor of all future major disasters, nor as a complete proxy for general disaster risk in the broadest sense. It is a backward-looking indicator of recent accumulated shock pressure. If a grid experiences a major event in 2024 but had no qualifying major events during 2013-2023, then the absence of a prior ARI signal is consistent with the way the index is defined. In this sense, accumulated risk and general disaster risk are related but not identical concepts.

In response to this comment, we have therefore revised the manuscript in three main ways. First, we have renamed and

reframed Section 3.3 as “External plausibility assessment and robustness checks”, so that the role of the evidence is described more accurately. Second, we have added the new out-of-sample comparison using 2024 major disaster events in Asia as an external plausibility assessment. Third, we have further tightened the interpretive boundary of ARI in the Discussion and Conclusions, making clear that the strongest claim supported by the current evidence is not that ARI has been definitively validated as a complete measure of true disaster risk, but that it shows meaningful external plausibility and internal robustness as a comparative hotspot-identification framework.

For ease of review, the revised manuscript now includes the following related formulations:

1) *Section 3.3 External plausibility assessment and robustness checks:*

“Given that ARI is a newly proposed index, it is important to examine both whether its hotspot pattern shows meaningful external plausibility and whether its main spatial features remain stable under reasonable methodological variation. Accordingly, this section consists of two complementary parts. First, an out-of-sample comparison is conducted using major disaster events that occurred in Asia during 2024, after the study period used to construct ARI, which is intended as an external plausibility assessment. Second, robustness checks are performed to evaluate the sensitivity of ARI to the fatality threshold used to define major disaster events and to the spatial resolution of the latitude-longitude grid.”

2) *Section 3.3.1 Out-of-sample plausibility assessment:*

“Because the ARI in this study is calculated from disaster events during 2013-2023, major disaster events that occurred in 2024 provide a natural post-period reference set for evaluating whether later major events tend to occur in areas already identified as ARI hotspots. To keep this additional analysis focused and tractable while remaining representative, the comparison is conducted for Asia, which is also the region where ARI exhibits the strongest and most continuous hotspot pattern in Figure 3.

Following the same event-selection logic used in the main analysis, major disaster events in Asia during 2024 are extracted from EM-DAT using the criterion of Total Deaths ≥ 50 . Each event is then matched to its corresponding $5^\circ \times 5^\circ$ grid cell. For each such cell, the ARI ranking derived from the 2013-2023 analysis is recorded and compared with the corresponding country-level WRI ranking. The resulting event list is summarized in Table 4, and the spatial correspondence between the 2024 events and the Asia ARI map is shown in Figure 7.

Table 4: Major disaster events in Asia during 2024 from EM-DAT and their corresponding ARI and WRI rankings.

Disaster subgroup	Disaster type	Disaster subtype	Country	Total deaths	ARI ranking	WRI ranking
Geophysical	Earthquake	Ground movement	Japan	699	54	24
Meteorological	Storm	Tropical cyclone	Myanmar	460	11	6
Hydrological	Mass movement (wet)	Landslide (wet)	India	378	15	3
Hydrological	Flood	Flood (General)	Pakistan	368	3	11
Hydrological	Flood	Flood (General)	Afghanistan	357	32	98
Meteorological	Storm	Tropical cyclone	Viet Nam	345	-	15
Hydrological	Flood	Flood (General)	Nepal	268	2	138
Hydrological	Flood	Flood (General)	Yemen	240	43	15
Meteorological	Storm	Tropical cyclone	Philippines	181	4	1
Hydrological	Flood	Flash flood	Afghanistan	150	106	98
Meteorological	Storm	Tropical cyclone	China	129	20	10
Hydrological	Mass movement (wet)	Landslide (wet)	Philippines	128	24	1
Hydrological	Flood	Flood (General)	India	125	2	3
Hydrological	Flood	Flood (General)	Indonesia	87	-	2
Hydrological	Flood	Flood (General)	China	77	20	10

Hydrological	Flood	Flash flood	Bangladesh	71	11	9
Hydrological	Flood	Flood (General)	China	65	20	10
Hydrological	Mass movement (wet)	Landslide (wet)	Nepal	62	2	138
Hydrological	Flood	Flood (General)	Yemen	62	43	15
Hydrological	Mass movement (wet)	Landslide (wet)	Indonesia	59	6	2
Hydrological	Flood	Flash flood	Afghanistan	58	3	98
Meteorological	Storm	Tropical cyclone	Philippines	53	4	1
Meteorological	Storm	Tropical cyclone	India	51	1	3

The comparison shows that most major disaster events in Asia during 2024 occurred in grid cells with relatively high ARI rankings, which supports the interpretation that ARI identifies locations subject to meaningful recent accumulated disaster pressure. Figure 7 further illustrates that many of the post-period events are spatially concentrated within or near the principal ARI hotspot belts identified from the 2013-2023 event set. This assessment also suggests that ARI can reveal spatial structures that are not readily apparent from country-level WRI alone. For example, some 2024 events in Nepal occurred in a grid cell with a very high ARI ranking, despite Nepal’s comparatively modest country-level WRI ranking. This reflects the fact that the corresponding grid accumulates major events from both Nepal and neighboring India, thereby capturing a transboundary accumulated-impact corridor that is less visible in a country-based index.

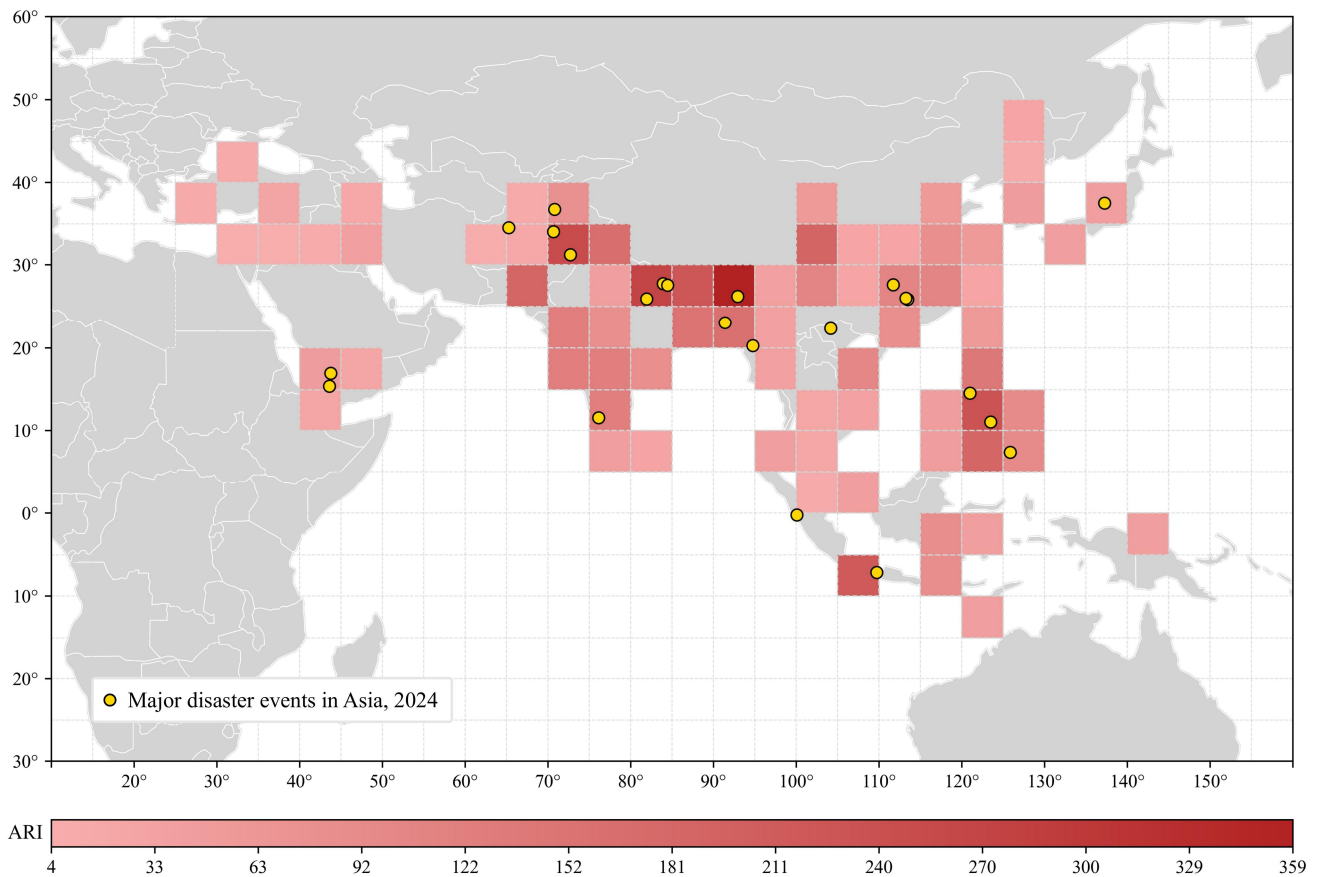


Figure 7: Asia ARI in 5°x5° latitude-longitude grids and major disaster events in Asia during 2024.

At the same time, this comparison also clarifies the conceptual boundary of ARI. Only two of the 2024 major events occurred in grid cells with no prior ARI value, such as the cases shown with “-” in Table 4. These cases do not necessarily contradict the index logic. Instead, they indicate that ARI is not intended as a universal predictor of all future major disasters, but as a backward-looking indicator of recent accumulated shock pressure. If a grid experiences a major event in 2024 but had no qualifying major events during 2013-2023, the absence of a prior ARI signal is consistent with the way ARI is defined. In this sense, accumulated risk and general disaster risk are related but not identical concepts. Therefore, the 2024

Asia comparison should be interpreted as an external plausibility assessment rather than definitive validation in a strict predictive sense. It strengthens confidence that ARI captures meaningful hotspot structures of recent accumulated disaster pressure, while also underscoring the need to interpret the index within its intended temporal and conceptual scope.”

3) *Section 3.3.2 Sensitivity to the fatality threshold:*

“These values represent a reasonable neighborhood around the baseline setting and are used to evaluate whether moderate threshold variation changes the principal hotspot pattern.”

“These changes affect the absolute magnitude of ARI and the number of included events, but they do not fundamentally alter the dominant hotspot geography. In all scenarios, the highest-ranking grids remain concentrated in South and Southeast Asia, particularly in regions covering northern India, Bangladesh, and the Himalayan belt. Moreover, many of the top 10 grids under the alternative thresholds coincide with those in the baseline configuration, often with only moderate changes in rank. This pattern indicates that the identification of the principal high-risk clusters is not strongly dependent on one exact threshold value, even though the event counts and ARI magnitudes vary as expected.”

4) *Section 3.3.3 Sensitivity to the grid resolution:*

“Figure 8 compares the resulting ARI distributions under these alternative resolutions.”

“This indicates that the broad spatial concentration of ARI is not an artifact of the single baseline resolution, but persists across reasonable changes in grid size.”

“Taken together, the out-of-sample plausibility assessment and the robustness checks indicate that ARI identifies spatially coherent and comparatively stable hotspot patterns at the global scale. The strongest claim supported by the current evidence is therefore not that ARI has been definitively validated as a complete measure of true disaster risk, but that it shows meaningful external plausibility when compared with the spatial occurrence of subsequent major disaster events and remains robust under reasonable variation in key methodological parameters. These results support the use of ARI as a comparative hotspot-identification framework, while detailed local interpretation should still consider the methodological uncertainties discussed in Section 5.”

5) *Section 5 Discussion:*

“At the same time, the present results also clarify the evidentiary status of ARI. The analyses in Section 3.3 indicate that the index exhibits meaningful external plausibility and internal robustness, but they do not justify stronger claims than that. The 2024 Asia comparison shows that many post-period major disaster events occurred within or near the principal hotspot belts identified from 2013-2023, which supports the interpretation that ARI captures spatially meaningful structures of recent accumulated disaster pressure. Likewise, the sensitivity tests on fatality threshold and grid resolution show that the main hotspot geography remains broadly stable under reasonable methodological variation. Taken together, these findings strengthen confidence in ARI as a comparative hotspot-identification framework. They do not, however, establish ARI as a definitive or exhaustive measure of disaster risk. This distinction is critical. ARI is best interpreted as a relative indicator of where recent major disaster shocks have accumulated under different country-level risk contexts, not as a complete substitute for structural risk models, nor as a universal predictor of future disasters.”

6) *Section 6 Conclusions:*

“The analyses further indicate that ARI shows meaningful external plausibility and internal robustness as a comparative hotspot-identification framework, while remaining a relative indicator rather than a complete measure of absolute disaster risk.”

We are very grateful for this comment. It helped us recognize the need to distinguish more carefully between robustness,

comparative interpretation, and external plausibility, and it significantly improved the scientific rigor, transparency, and credibility of the manuscript.

Comment 4:

The remote sensing dataset (3H) is a valuable contribution, but its connection to ARI is not fully justified. The dataset includes imagery for all areas within the 14 high-risk grids, not specifically for the locations where major disasters occurred. For a grid cell spanning hundreds of kilometers, this includes many areas that may not have been directly affected. The supplementary imagery (Grids 3 and 7) is centered on the location of the highest-fatality event, which is a good approach, but this is not done for the other grids.

The dataset is presented as a tool for future research. However, the manuscript does not demonstrate how the imagery correlates to the ARI or to the disaster events themselves.

Response 4:

We sincerely thank the referee for this important and highly constructive comment. We fully understand the concern and agree that, in the previous version of the manuscript, the relationship between the ARI framework and the 3H remote sensing dataset was not explained clearly enough. As a result, the remote sensing component could appear more detached from the core risk-assessment framework than we intended. We are grateful to the referee for identifying this issue so precisely, because it has helped us clarify both the logic of the dataset construction and the actual boundary of its contribution.

At the outset, we would like to make one key clarification. The purpose of constructing the 3H Dataset was not to create an event-footprint-level disaster imagery dataset, nor to use remote sensing imagery as a direct validation tool for ARI in the present manuscript. Rather, the 3H Dataset was designed as a hotspot-oriented remote sensing resource guided by the ARI results. In other words, ARI first identifies which $5^{\circ}\times 5^{\circ}$ grid cells deserve priority attention because they exhibit relatively high accumulated disaster pressure, and the 3H Dataset then provides standardized high-resolution imagery resources within those hotspot grids to support subsequent finer-scale Earth observation studies. The connection between ARI and 3H is therefore grid-level and hotspot-oriented, not event-by-event and footprint-specific.

We fully accept the referee's point that a $5^{\circ}\times 5^{\circ}$ grid cell covers a very large area and therefore inevitably includes locations that were not directly affected by the specific major disaster events recorded in EM-DAT. This is correct, and we do not dispute it. In fact, we believe the referee's comment usefully exposes the exact scale boundary of the 3H Dataset: it is not a disaster-footprint imagery archive, but an imagery collection for ARI-identified hotspot grids. Because the ARI itself is defined at the grid-cell level, the remote sensing component follows the same spatial logic. We have therefore revised the manuscript to state this design choice much more explicitly so that readers do not misinterpret the 3H Dataset as a direct image representation of the precise spatial extent of each disaster event.

At the same time, we would also like to clarify that the imagery in the 3H Dataset is not arbitrary coverage of these grids. Most of the source imagery was drawn from open remote sensing datasets such as DigitalGlobe and RAMP, both of which are strongly associated with disaster response, humanitarian mapping, or built-environment analysis in vulnerable regions. Nevertheless, we fully agree with the referee that this does not remove the underlying methodological point: even when the source imagery is disaster-relevant, the available coverage does not always correspond exactly to the footprint of the major events used in the ARI calculation. The manuscript now makes this boundary much clearer.

The referee also raises an important point regarding the different treatment of Grids 3 and 7. We agree that this was not sufficiently justified in the previous version. The reason is that the current 3H Dataset was constructed through a hybrid data-assembly strategy. For most high-risk grids, we relied on existing open-image coverage from available sources, whose spatial extents were determined by those archives. For a small number of hotspot grids where suitable open imagery was

lacking, we carried out supplementary data acquisition. In those specific cases, we intentionally centered the acquisitions on the location of the highest-fatality event in order to strengthen the link between the supplementary imagery and severe disaster impact. We have revised the manuscript to make this two-track logic transparent and to state clearly that the event-centred refinement used for Grids 3 and 7 does not imply that the entire 3H Dataset was constructed according to a uniform event-centred principle.

We also fully agree with the referee's broader criticism that the manuscript did not previously demonstrate how the imagery correlates to ARI or to the disaster events themselves. The present paper does not provide a quantitative remote sensing analysis linking ARI values to image-observed damage, land-cover change, recovery trajectories, or other image-derived indicators. We now state this much more clearly. The current manuscript supports the following claim: ARI provides a principled way to prioritize hotspot grids for remote sensing data compilation, and the 3H Dataset provides a standardized imagery basis for future studies in those prioritized areas. What the manuscript does not yet demonstrate is a direct quantitative correspondence between ARI and image-observed disaster impacts. We have revised the text accordingly so that the contribution chain is presented more precisely and more honestly: ARI is the hotspot-selection framework; 3H is the data-support layer for subsequent Earth observation analyses; and the coupling between ARI and directly observed image-derived damage remains a task for future work.

At the same time, we would like to note that compiling imagery at the hotspot-grid level, rather than only clipping narrow event-specific cores, also has a practical rationale for downstream remote sensing research. Although broader coverage weakens direct event-level correspondence, it preserves contextual and non-damaged areas within the same environmental and urban setting, which may be valuable for future applications such as building damage detection, exposure mapping, recovery monitoring, and other machine-learning workflows. We do not present this as a rebuttal to the referee's criticism, but as an additional explanation of why a hotspot-grid-oriented imagery collection can still be scientifically useful despite not being event-footprint-specific.

More broadly, the referee's comment helped us recognize that two different dataset concepts were insufficiently separated in the previous manuscript: 1) an event-footprint-oriented disaster imagery dataset, which would require precise spatial matching to individual disasters; and 2) a hotspot-oriented imagery dataset, which organizes available high-resolution imagery around macro-scale high-risk grids for subsequent detailed study. The present 3H Dataset belongs to the second category, not the first. We have revised the manuscript to make this distinction explicit in the workflow description, in Section 4, and in the Discussion and Conclusions.

In response to this comment, we have revised the manuscript to make four points much clearer: 1) the 3H Dataset is hotspot-oriented rather than event-footprint-oriented; 2) inclusion of entire hotspot grids means that not all images correspond directly to major disaster locations, and this is an acknowledged scale-related limitation rather than an overlooked issue; 3) the different treatment of Grids 3 and 7 reflects a hybrid data-assembly strategy under data-availability constraints, not a uniformly applied event-centred rule; and 4) the current study does not yet demonstrate a direct quantitative correlation between ARI and image-observed impacts, so the 3H Dataset is now presented as a complementary operational extension and future research resource, rather than as a direct validation of ARI.

For ease of review, the revised manuscript now includes the following related formulations:

1) *Section 1.1 Background:*

“Consequently, while remote sensing is invaluable for fine-scale disaster analysis, its direct and uniform deployment within a global risk-assessment framework remains difficult. This practical constraint further motivates the need for an intermediate spatial screening strategy to identify priority areas for subsequent Earth observation analysis.”

2) *Section 1.2 Motivation and Objectives:*

“In addition, in order to facilitate further remote sensing-based research, this study systematically integrates and supplements existing sub-metre visible-spectrum remote sensing imagery for ARI-identified high-risk grids, and publishes them as an open-source dataset after standardization. Within the present framework, the ARI dictates spatial prioritization, while the resulting 3H Dataset provides a foundational resource for potential fine-scale applications, including building damage detection, exposure mapping, and post-disaster recovery monitoring.”

3) *Section 2.5 Analytical workflow:*

“In this sense, ARI serves as the hotspot-screening mechanism that guides where remote sensing resources are prioritized, while the 3H Dataset provides the geospatial data foundation for subsequent finer-scale studies in those selected high-risk areas.

This workflow demonstrates that the remote sensing component is not independent of the risk model, but is developed as an operational extension of it. Specifically, ARI identifies where recent accumulated disaster pressure is concentrated, and the 3H Dataset translates those hotspot results into a standardized imagery resource that can support future applications. Together, these methodological stages constitute an integrated framework linking historical disaster records, contextual risk information, and Earth observation data.”

4) *Section 4 Remote sensing data integration for ARI hotspots:*

“In this study, the remote sensing component is developed as a complementary operational extension of the ARI framework. ARI first identifies hotspot grids where recent accumulated disaster pressure is concentrated, and remote sensing data are then compiled and supplemented within those ARI-identified locations. The resulting 3H Dataset is therefore organized at the hotspot-grid level and is intended to support subsequent finer-scale Earth observation and machine-learning studies in these areas. Within the present manuscript, its role is to provide a standardized geospatial data foundation rather than to serve as a direct image-based validation of ARI.”

5) *Section 4.2 Supplementing data gaps in high-risk grids:*

“The centre of each acquisition area was selected according to the most severe major event within the corresponding hotspot grid, specifically the event with the maximum “Total Deaths” recorded in that cell. This event-centred refinement was applied only to these supplementary acquisitions in data-scarce grids and does not alter the overall hotspot-oriented design of the 3H Dataset.”

6) *Section 4.3 Standardization and construction of the 3H Dataset:*

“By focusing explicitly on ARI-identified hotspot grids in developing countries and by providing imagery in a standardized format, the 3H Dataset is intended to support future finer-scale studies such as building damage detection, exposure mapping, and post-disaster recovery monitoring. In this way, it complements the macro-scale ARI framework by making high-risk, data-scarce locations more accessible for subsequent Earth observation analysis.”

7) *Section 5 Discussion:*

“Seen from this perspective, the broader significance of the study is not only methodological but also infrastructural. ARI provides a macro-scale screening logic for identifying where recent accumulated pressure is concentrated, while the 3H Dataset offers a micro-scale observational basis for examining what those hotspots look like on the ground. The two components therefore operate at different but connected scales. Together, they suggest a possible pathway for linking global disaster statistics, contextual risk conditions, and Earth observation resources in a more operationally coherent way than is commonly achieved in either purely index-based or purely image-based studies.”

“Finally, the ARI-guided 3H Dataset opens an avenue for subsequent machine-learning applications, including building damage detection, exposure mapping, recovery monitoring, and domain adaptation in data-scarce regions. Such finer-scale outputs may, in turn, provide new evidence for further refining and evaluating accumulated-risk indicators.”

8) *Section 6 Conclusions:*

“Building on the ARI-identified hotspot grids, this study also develops the 3H Dataset, which contains more than 7.5 million standardized sub-metre remote sensing image tiles across 14 high-risk grids in 10 developing countries. Together, ARI and the 3H Dataset provide a connected basis for linking global disaster statistics, contextual risk information, and Earth observation resources, thereby supporting subsequent research and practice in multi-disaster hotspot identification and finer-scale geospatial analysis.”

We are very grateful for this comment. It has substantially improved the logical coherence of the manuscript and helped us define the true contribution boundary of the remote sensing component much more clearly and transparently.

Comment 5:

The abstract is quite dense. Consider shortening it to focus more clearly on the ARI methodology, the key findings and the intended contribution.

Response 5:

We sincerely thank the referee for this helpful suggestion. We agree that the previous abstract was too dense and attempted to accommodate too much background, methodological detail, results, and application-oriented discussion within a limited space. As a consequence, the central message of the study was not presented as clearly or as efficiently as it should have been.

In response, we have substantially revised and streamlined the abstract. In the revised version, we reduced the amount of general background material and reorganized the text so that it now focuses more directly on three core elements: 1) the basic construction logic of the ARI, 2) the principal findings regarding the spatial pattern of identified hotspot grids, and 3) the actual contribution of the study, stated in a more careful and appropriately delimited way.

More specifically, the revised abstract now presents the ARI methodology more explicitly and succinctly, highlights the main hotspot-identification results, and avoids overly expansive phrasing regarding the scope of contribution. We also moderated the wording related to the remote sensing dataset so that the abstract better matches the actual evidentiary strength of the manuscript. In this sense, the revision is also consistent with related comments from other referees, particularly those requesting a clearer statement of the ARI formulation and a more cautious presentation of the role of the 3H Dataset.

For ease of review, the revised abstract now reads as follows:

“This study introduces an Accumulated Risk Index (ARI) to map recent global multi-disaster hotspots, leveraging Emergency Events Database (EM-DAT) records from 2013 to 2023. By aggregating World Risk Index (WRI) values for major disaster events within 5°×5° latitude-longitude grids, the ARI couples recent hazard occurrences with country-level contextual risk conditions. Results show that 153 grids (5.9% of the global total) have non-zero ARI values, predominantly concentrated across South and Southeast Asia and along the Himalayan belt. Compared to a spatially equivalent WRI, the ARI captures sharper contrasts among neighbouring grids, illuminating the accumulated disaster burden at both transboundary and subnational scales. To facilitate ongoing research in these critical zones, this study additionally presents the 3H Dataset, a standardized compilation of sub-metre remote sensing imagery derived from existing open sources and supplementary acquisitions. This unified dataset offers a complementary resource to support finer-scale, observation-based analyses in high-risk, data-scarce developing regions.”

We believe this revision has substantially improved the readability, focus, and precision of the abstract, and we are grateful to the referee for drawing our attention to this issue.

Comment 6:

The intended utilization of ARI is listed for policymakers, practitioners, and researchers, but the manuscript does not later return to these specific purposes.

Response 6:

We sincerely thank the referee for pointing out this important issue of structural consistency. We agree that, in the previous version of the manuscript, the intended user groups of the ARI framework were introduced in the early part of the paper, but the later sections did not return to these categories in a sufficiently explicit and organized way. As a result, the link between the proposed framework and its practical or research-oriented use cases was not as clearly closed as it should have been.

In response, we have revised the manuscript to make this “method-results-use” connection more explicit. In particular, we now return to the three intended user groups, policymakers, practitioners, and researchers, in the Discussion, where the practical meaning and appropriate scope of the ARI-3H framework are summarized more clearly.

More specifically, the revised manuscript now distinguishes these user groups as follows:

- 1) For policymakers, ARI is presented as a macro-scale screening tool that can help identify broad regions where repeated high-impact events may have intensified disaster pressure and where preparedness planning, resilience-building, or capacity support may warrant priority attention.
- 2) For practitioners and humanitarian actors, ARI is not presented as a substitute for local operational assessment; rather, it is now described more clearly as a spatial cue indicating where finer-scale assessment, field verification, or observation-based monitoring may be most needed.
- 3) For researchers, the role of the framework is also stated more explicitly: ARI provides a comparative hotspot-identification structure at macro scale, while the 3H Dataset extends that framework into a standardized imagery resource for subsequent finer-scale Earth observation and machine-learning studies in high-risk, data-scarce regions.

For ease of review, the revised manuscript now includes the following passage in *Section 5 Discussion*:

“Within those boundaries, the combined ARI-remote sensing framework still offers a meaningful contribution. For policymakers, ARI provides a way to screen for broad regions where repeated high-impact events may have intensified disaster pressure and where preparedness, resilience-building, or capacity support may need to be prioritized. For practitioners and humanitarian actors, the identified hotspot grids can serve as a spatial cue for where finer-scale assessment, field verification, or observation-based monitoring may be most warranted. For researchers, the 3H Dataset extends the ARI framework from hotspot identification to data infrastructure by making sub-metre imagery available in a standardized form for high-risk, data-scarce regions.”

We believe this revision helps close the “method-results-use” loop more effectively and makes the practical meaning of the study clearer to readers from different communities. We are grateful to the referee for identifying this gap, because it has helped us present the contribution of the manuscript in a more coherent and complete way.

Comment 7:

Please ensure the source citations are correctly formatted in Table 1 and that the "statistic year" for the Natural Disaster Hotspots Project is accurate.

Response 7:

We sincerely thank the referee for this careful and important check. We fully agree that the accuracy and consistency of the source information in Table 1 are essential, especially because this table is intended to provide a comparative overview of representative global disaster risk assessment frameworks.

In response, we have re-examined all entries in Table 1 and revised the table to improve both citation consistency and bibliographic accuracy. In particular, we checked the correspondence among the cited source, the publication year, and the reported “statistic year” for each framework, and we standardized the source presentation so that the table is more internally consistent and easier to interpret.

Following the referee’s specific suggestion, we also revisited the entry for the Natural Disaster Hotspots Project and verified the corresponding source information and “statistic year” more carefully. The table has been corrected where necessary so that the distinction between publication time and statistical period is now presented more accurately and transparently.

More broadly, we appreciate this comment because it highlights how even relatively small inconsistencies in a summary table can affect the overall traceability and credibility of the manuscript. We believe the revised Table 1 is now more accurate and better aligned with the journal’s expectations for clarity and consistency.

Comment 8:

Add a note in the text acknowledging this limitation explicitly.

Response 8:

We sincerely thank the referee for this important reminder. We fully agree that the relevant limitation should be stated explicitly in the manuscript itself, rather than being explained only in the response letter. Methodological boundaries and interpretive limits should be transparent to all readers without requiring access to the review correspondence.

In response, we have added an explicit statement in the main text acknowledging this limitation more directly and visibly. Rather than leaving it only implied or scattered across different sections, we now state it more clearly in the manuscript’s discussion of methodological boundaries and limitations, so that readers can more readily understand the appropriate scope of interpretation.

More specifically, this point has now been incorporated into the revised Discussion section, where the principal limitations of the present framework are summarized in a more systematic way. This addition helps make the manuscript more transparent and ensures that the relevant limitation is presented as part of the paper itself, not only as an explanation in the response letter.

We appreciate this suggestion because it improves not only the completeness of the manuscript, but also its clarity and scientific caution.

Comment 9:

Several typos throughout the text should be double checked, as for example the ARI formula.

Response 9:

We sincerely thank the referee for this careful observation, and we apologize for these oversights. We fully agree that typographical, notational, and formatting errors, especially in equations, variable definitions, and methodological descriptions, can negatively affect the clarity, professionalism, and reproducibility of the manuscript.

In response, we have carefully proofread the manuscript again, with particular attention to the ARI formulation, the

presentation of mathematical notation, variable definitions, superscripts and subscripts, and the consistency of figure and table cross-references. We also revised the surrounding explanatory text so that the meanings of the symbols and the interpretation of the ARI are now stated more clearly and consistently.

In particular, because the ARI equation lies at the core of the methodological contribution, we have revised the relevant part of Section 2.3.2 to improve both the presentation and the interpretability of the formula. The equation is now written more clearly, all variables are explicitly defined, and the accompanying text now explains more directly that ARI should be interpreted as an event-driven, grid-based, context-weighted relative index, rather than as a complete physical-statistical representation of disaster risk. We believe these revisions have substantially improved the clarity and internal consistency of the manuscript.

We appreciate the referee's reminder, because it helped us identify and correct issues that, although small in appearance, are important for the precision and readability of the paper.

Comment 10:

Figure 2 Please explain what “N-Na” means and how overlap degree is calculated.

Response 10:

We sincerely thank the referee for pointing out this lack of clarity. We agree that the notation used in Figure 2 was not sufficiently self-explanatory in the previous version, and that this could hinder the reader's interpretation of the figure.

The primary purpose of Figure 2 is qualitative rather than strictly quantitative. It is intended to show that the selected major disaster events from 2013 to 2023 are not randomly distributed across the globe, but instead exhibit evident spatial clustering, particularly along tectonically active belts and in monsoon-influenced regions. In this sense, the figure provides empirical motivation for adopting grid-based spatial units in the ARI framework.

In the previous version, the notations N_1 - N_4 were intended to indicate the maximum overlap level of plotted event points in the corresponding subpanels, that is, the highest visible local concentration of events within each disaster-type-specific map. However, we fully agree with the referee that this notation was too ambiguous, because it did not make clear whether it referred to an exact count, a normalized value, or a purely visual indicator.

In response, we have revised both the main text and the figure caption to explain this more explicitly. We now state that the “overlap degree” is derived from the visual superposition of semi-transparent event points in the corresponding map panels. Darker areas emerge where multiple events overlap spatially, so the overlap degree is intended as an approximate visual proxy for local event concentration, rather than as a formal geospatial statistic or a quantitative analytical variable comparable to ARI.

We have also revised the wording so that the visual and approximate nature of this indicator is now much clearer. The intention is to prevent readers from interpreting N_1 - N_4 as a formal statistical metric, while still allowing the figure to communicate relative clustering intensity in an intuitive way.

For ease of review, the revised manuscript now includes the following formulations:

Section 3.1 Spatiotemporal characteristics of recent disaster events:

“In panels (b)-(e), semi-transparent event points are superposed to indicate local clustering intensity. The notations N_1 - N_4 represent the maximum spatial overlap of plotted points in the corresponding subpanels, and the overlap degree is directly derived from this visual superposition as an approximate proxy for local event concentration rather than a formal geospatial statistic.”

“Figure 2: Geographic spatial distribution of selected major disaster events. (a) Global overview. Panels (b)-(e) visualize local event clustering using semi-transparent points. The notations N1-N4 denote the peak point superposition density, which serves as an approximate indicator of local disaster concentration within respective subpanels.”

We are grateful to the referee for highlighting this issue, because it has helped us improve both the clarity and the interpretability of Figure 2.

Comment 11:

Figure 3 and 4 The color scales for ARI and Equivalent WRI are different (ARI uses a continuous color bar, Equivalent WRI uses discrete categories). For direct comparison (especially in Figure 4), it would be helpful to use a consistent color scale, or to clearly explain why different scales are used.

Response 11:

We sincerely thank the referee for this very helpful comment. We agree that, when two indicators are presented side by side for comparison, differences in visual encoding can increase the reader’s interpretive burden. We also understand the referee’s concern that the comparison between ARI and Equivalent WRI, especially in Figure 4, may appear less straightforward than intended if the logic of the colour design is not explained clearly enough.

The comparison between ARI and Equivalent WRI is indeed one of the key interpretive components of the manuscript. Figure 3 presents this comparison at the global scale, while Figure 4 serves as a zoomed regional continuation of the same comparison. We therefore agree that the rationale behind the colour presentation needed to be made much more explicit.

At the same time, we would like to clarify why the two maps were not forced into one identical absolute numeric colour scale. ARI and Equivalent WRI differ substantially in both numerical range and analytical meaning. ARI is an event-driven accumulation index whose values result from the summation of WRI-weighted major disaster events within each $5^{\circ} \times 5^{\circ}$ grid. By contrast, Equivalent WRI is a grid-level reference indicator derived from country-level WRI using a conservative maximum assignment rule. The two indices are therefore not directly comparable in an absolute numeric sense. If a fully identical absolute scale were imposed, the visual variation within Equivalent WRI would become highly compressed, and much of its internal spatial contrast would be lost.

For this reason, the revised figures retain different numerical ranges while using an analogous colour progression. The intention of the comparison is not to imply one-to-one equivalence between colour values in the two panels, but rather to allow readers to compare relative spatial pattern, hotspot structure, and degree of spatial differentiation. In other words, the comparison asks whether the two indicators identify similar or different hotspot belts, boundary transitions, and subnational contrasts, rather than whether the same colour corresponds to the same absolute value. We agree that this comparison logic was not sufficiently clear in the previous version, and we have revised the manuscript accordingly.

In response to this comment, we have improved the manuscript in two ways. First, we have revised the captions and main text of Figures 3 and 4 to state explicitly that the two panels use analogous colour progression but different numerical ranges because ARI and Equivalent WRI serve different analytical roles. Second, we now explain more directly in the main text that the purpose of the comparison is to highlight differences in spatial pattern and relative contrast, rather than exact equivalence of colour values.

For ease of review, the revised manuscript now includes the following formulations:

Section 3.2 Global ARI distribution and hotspots:

“Figure 3: Global distributions of (a) ARI and (b) Equivalent WRI in $5^{\circ} \times 5^{\circ}$ latitude-longitude grids. The two panels use analogous colour progression but different numerical ranges.”

“Although Figure 3(a) and Figure 3(b) use analogous colour progression, they retain different numerical ranges because ARI and Equivalent WRI serve different analytical roles. ARI is a grid-based index of recent event accumulation, whereas Equivalent WRI is a grid-level reference indicator derived from country-level background risk. Therefore, the comparison is intended to emphasize differences in spatial pattern and relative contrast rather than direct equivalence of colour values between the two maps. From this perspective, ARI reveals sharper subnational and transboundary differentiation than Equivalent WRI, especially in areas affected by repeated recent disaster shocks.”

“The figure is designed for a grid-aligned comparison: the full 5°×5° gridlines are shown in the regional window, and the two highlighted sub-windows (A and B) are reproduced on the right as actual cell-by-cell mosaics for Equivalent WRI and ARI using the same colour progression as in Figure 3.”

“Taken together, Figure 4 makes the contrast in Figure 3 more explicit: incorporating near-term event accumulation does not simply reproduce the country-based background pattern, but adds spatial discrimination along transboundary hazard corridors and within large countries where subnational variability is important.”

We are grateful for this comment, because it helped us improve not only the presentation of Figures 3 and 4, but also the clarity with which the manuscript explains the analytical relationship between ARI and Equivalent WRI.

Comment 12:

Figure 5 This figure is not present in the provided PDF. It is referenced in the text (page 11) but the figure file is missing.

Response 12:

We sincerely thank the referee for pointing out this issue, and we apologize for the inconvenience caused. We fully agree that the absence of Figure 5 from the review PDF was problematic, especially because the figure was explicitly referenced in the main text.

In response, we carefully checked the revised manuscript and ensured that Figure 5 is now correctly included, embedded, and displayed in the manuscript file. We also rechecked the full set of figures and their cross-references to reduce the risk of similar technical omissions in the revised submission.

We are grateful to the referee for identifying this technical problem, because it helped us improve the completeness and consistency of the revised manuscript.

Comment 13:

Table 4 shows that the number of events changes with threshold, but the authors conclude that the main spatial clusters remain stable.

Response 13:

We sincerely thank the referee for this important comment. We agree that, in the previous version of the manuscript, the relevant statement was not sufficiently precise. The point of the sensitivity analysis was not that the absolute number of selected events remains unchanged across different fatality thresholds. On the contrary, as shown in Table 5, the number of included major disaster events changes substantially when the threshold is adjusted, and this is entirely expected. Our intended point was instead that the dominant spatial concentration of high-ARI grids remains broadly similar across reasonable threshold settings.

In other words, the sensitivity analysis was designed to assess the robustness of the relative hotspot pattern, not the invariance of event counts or absolute ARI magnitudes. We agree that this distinction needed to be made much clearer. In response, we have revised the relevant text so that the robustness claim now refers explicitly to the persistence of the

principal hotspot belts and the substantial overlap among the top-ranked grids, rather than to stability in the total number of events.

More specifically, the revised manuscript now explains that lowering or raising the fatality threshold changes both the number of selected events and the absolute ARI values, but does not fundamentally alter the dominant hotspot geography. Across the alternative threshold settings examined, the highest-ranking grids remain concentrated in broadly similar regions, especially in South Asia, Southeast Asia, and the Himalayan belt. In addition, many of the top-ranked grids under alternative settings overlap with those identified under the baseline threshold, even if their exact rankings shift to some extent. We believe this is a more accurate and appropriately cautious interpretation of the evidence.

For ease of review, the revised manuscript now includes the following formulations in *Section 3.3.2 Sensitivity to the fatality threshold*:

“These values represent a reasonable neighborhood around the baseline setting and are used to evaluate whether moderate threshold variation changes the principal hotspot pattern.”

“These changes affect the absolute magnitude of ARI and the number of included events, but they do not fundamentally alter the dominant hotspot geography. In all scenarios, the highest-ranking grids remain concentrated in South and Southeast Asia, particularly in regions covering northern India, Bangladesh, and the Himalayan belt. Moreover, many of the top 10 grids under the alternative thresholds coincide with those in the baseline configuration, often with only moderate changes in rank. This pattern indicates that the identification of the principal high-risk clusters is not strongly dependent on one exact threshold value, even though the event counts and ARI magnitudes vary as expected.”

“Taken together, the out-of-sample plausibility assessment and the robustness checks indicate that ARI identifies spatially coherent and comparatively stable hotspot patterns at the global scale. The strongest claim supported by the current evidence is therefore not that ARI has been definitively validated as a complete measure of true disaster risk, but that it shows meaningful external plausibility when compared with the spatial occurrence of subsequent major disaster events and remains robust under reasonable variation in key methodological parameters. These results support the use of ARI as a comparative hotspot-identification framework, while detailed local interpretation should still consider the methodological uncertainties discussed in Section 5.”

We have also ensured that the concluding wording of Section 3.3 is now more disciplined and does not overstate what the sensitivity analyses can demonstrate. The revised text makes clear that the current evidence supports the interpretation of ARI as a framework that identifies spatially coherent and comparatively stable hotspot patterns under reasonable methodological variation, rather than implying invariance in event counts or definitive validation of “true” disaster risk.

We appreciate this comment because it helped us sharpen an important distinction between absolute event-number variation and relative hotspot persistence, and we believe the revised wording now presents the result in a clearer and more scientifically cautious way.

Comment 14:

The authors should clarify how the tiling was performed (e.g., was there any overlap? Were tiles that contain only water or non-vegetated areas included? Was there any filtering for cloud cover beyond visual inspection?).

Response 14:

We sincerely thank the referee for raising this important point. We fully agree that the previous version of the manuscript described the preprocessing and tiling workflow of the 3H Dataset too briefly, and that additional procedural detail is necessary to improve methodological transparency and reproducibility.

In response, we have revised the dataset-construction description to make the workflow more explicit and operational. The revised manuscript now explains the 3H preprocessing pipeline as a sequence of steps: ARI-guided hotspot selection, source imagery collection, visual quality screening, spatial standardization and tiling, and metadata assignment / final dataset organization. This revision was intended to make clearer how the final 256×256-pixel TIF patches were generated from heterogeneous source imagery and how unsuitable inputs were filtered before inclusion in the final dataset.

More specifically, we now state that candidate scenes were first screened through manual visual inspection. Scenes with severe artefacts, substantial cloud, cloud-shadow, or haze contamination, or with very limited meaningful built-up content, were excluded because they were poorly aligned with the intended downstream applications of the dataset. We agree with the referee that this point was previously too vague, especially with respect to cloud screening and non-target areas. The revised text now makes it explicit that the present workflow relies primarily on visual quality screening rather than a fully automated cloud-masking or land-cover-filtering pipeline, and that this should be understood as a practical but partly subjective preprocessing step.

The referee also asks specifically about the tiling procedure. In the revised workflow, the retained scenes were spatially standardized and tiled into fixed-size 256×256 pixel patches. The purpose of this step is dataset standardization across heterogeneous sources rather than event-footprint reconstruction. We also clarify in our revised description that the dataset is organized at the hotspot-grid level, with each tile stored in TIF format and annotated with the corresponding hotspot-grid identifier and data source, so that each standardized unit can be linked back to both its originating imagery and its ARI-identified grid location.

With respect to the referee’s specific concern about water- or non-vegetation-dominated areas, we would like to clarify the intended scope of the dataset. The 3H Dataset was not designed as a perfectly curated building-only imagery set, and no fully automated semantic land-cover mask was applied before tiling. However, the revised manuscript now states more clearly that scenes with little or no meaningful built-up content were excluded during the visual screening stage, because they were not well aligned with the intended downstream applications such as building damage detection, exposure mapping, and post-disaster recovery analysis. In this sense, the current workflow seeks to improve usability through practical screening rather than through a strict automated land-cover selection scheme.

We appreciate the referee’s specific question regarding cloud cover beyond visual inspection. The revised manuscript now makes clear that cloud contamination was screened primarily through manual visual inspection, rather than through an automated cloud-detection algorithm. We agree that this should be stated explicitly, because it is both a practical design choice and a limitation of the present preprocessing pipeline.

For ease of review, the revised manuscript now includes the following description in *Section 4.3 Standardization and construction of the 3H Dataset*:

“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 visually inspected to remove imagery with severe artefacts, substantial cloud, cloud-shadow, or haze contamination that prevented reliable interpretation of built-up targets. In addition, scenes with little or no meaningful built-up content were excluded because they were poorly aligned with the intended downstream applications of the dataset. At the present stage, this screening procedure was based primarily on manual visual inspection rather than a fully automated filtering framework. 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 so that each standardized unit could be linked back to its originating imagery and ARI-identified grid location. This

workflow was applied consistently across the 14 high-risk grids.”

We are grateful for this comment, because it helped us make the dataset description substantially clearer and more reproducible. We believe the revised manuscript now communicates the preprocessing logic of the 3H Dataset in a much more transparent way.

Comment 15:

The references are generally well-formatted, but there are a few issues that need attention.

Response 15:

We sincerely thank the referee for this careful check. We appreciate the positive assessment that the reference list is generally well formatted, and we fully agree that the remaining issues should be corrected to ensure that the manuscript meets the journal’s standards of accuracy, consistency, and presentation quality.

In response to this comment, we have re-examined the reference list in detail and revised it to improve overall consistency. In particular, we checked the formatting of author names, journal titles, publication years, volume and issue information, page ranges, DOI presentation, and consistency between the in-text citations and the reference list. We also rechecked references to reports, datasets, and online resources so that they are presented in a more consistent and traceable way, especially for sources that are important to the methodological background of this study.

In addition, we cross-checked the bibliography against the main text to ensure that cited references are included in the reference list, that uncited entries are avoided, and that the overall formatting is more closely aligned with the NHESS reference style.

We are grateful to the referee for drawing attention to these details, because this helped us improve the overall rigor and professionalism of the revised manuscript.

➤ Review Comments of Referee #4:

Comment 16:

The study presents a promising and relevant contribution, but currently underspecifies its methodological core. Strengthening clarity and reproducibility—particularly in how the ARI is constructed and how data are processed—will significantly improve the manuscript’s rigor and impact.

Response 16:

We sincerely thank the referee for this constructive and well-balanced overall assessment. We greatly appreciate the recognition that the study addresses a relevant problem and has the potential to make a meaningful contribution. At the same time, we fully agree with the referee’s central concern: in the previous version, the manuscript did not explain the methodological core with sufficient clarity, particularly with respect to how the ARI is constructed, how key variables are operationalized, and how the underlying data are processed in a reproducible manner.

This comment has been especially valuable because it captures the main issue that underlies several of the more specific referee comments: the study does not primarily need more results, but a clearer and more transparent explanation of how the framework works, what assumptions it makes, and what kind of conclusions it can legitimately support. In response, we have substantially revised the manuscript to strengthen both clarity and reproducibility.

First, we have made the construction logic of the ARI much more explicit. The revised manuscript now clarifies that ARI is

an event-driven, grid-based, context-weighted relative index, rather than a full structural risk model in the classical hazard-exposure-vulnerability sense. We have rewritten the core methodological description, clarified the conceptual role of the WRI as a country-level contextual weighting factor, and made the interpretation boundary of the ARI more explicit. In particular, the revised Section 2.3.2 now presents the ARI formulation more clearly, defines all variables explicitly, and explains more directly what the index is intended to capture and what it is not intended to represent.

Second, we have revised the data-processing description to make the workflow more operational and reproducible. For the EM-DAT component, the revised manuscript now explains more clearly: how disaster records were filtered, why Total Deaths ≥ 50 was adopted as the operational threshold, how purely textual event locations were standardized into representative coordinates, and how multi-location events were assigned to grid cells while avoiding double-counting. These steps are now described more systematically in Sections 2.1 and 2.2, and the associated assumptions and limitations are discussed more directly later in the manuscript.

Third, we have improved the explanation of the spatial aggregation logic in Section 2.4. The manuscript now states more explicitly why a $5^\circ \times 5^\circ$ grid was adopted, what methodological trade-off this choice reflects, and how alternative grid sizes are used in the sensitivity analysis. This helps connect the index design more transparently to the spatial uncertainty of the underlying data and to the global comparative purpose of the study.

Fourth, we have strengthened the methodological coherence of the full framework by revising Section 2.5 Analytical workflow. The revised text now presents the three stages of the study, major-event identification, ARI construction, and ARI-guided remote sensing data integration, as one connected workflow. This addition makes it clearer that the remote sensing component is not an independent add-on, but an operational extension of the hotspot-identification logic established by ARI.

Fifth, we have also revised the 3H Dataset description to improve reproducibility. The revised Section 4 now explains more clearly how imagery sources were selected, how hotspot-oriented coverage was assembled, how supplementary acquisitions were used in data-scarce grids, and how the preprocessing pipeline was carried out through visual quality screening, standardization, tiling, and metadata assignment. This revision was intended to ensure that the dataset is presented not only as a resource for future applications, but also as a transparent methodological component of the study.

Finally, we have improved the consistency between the methods, results, and interpretation sections. The revised manuscript now aligns the methodological description with the evidentiary framing adopted later in the paper. In particular, Section 3.3 now distinguishes more clearly between external plausibility assessment and robustness checks, while Sections 5 and 6 define the contribution and limits of the ARI in a more disciplined way. We believe this is important because reproducibility is not only a matter of procedural detail, but also of ensuring that the claims made in the results, discussion, and conclusions are fully consistent with what the method actually does.

For ease of review, the revised manuscript now reflects this strengthened methodological core in the following key locations:

- 1) Abstract: clearer and more focused summary of the ARI formulation, principal hotspot findings, and the role of the 3H Dataset.
- 2) Sections 1.1 and 1.2: stronger methodological positioning, clearer motivation, and a more explicit research question.
- 3) Sections 2.1-2.4: more explicit treatment of event filtering, spatial standardization, ARI formulation, and grid selection.
- 4) Section 2.5: integrated analytical workflow connecting ARI and the remote sensing component.
- 5) Section 3.3: clearer evidentiary structure distinguishing plausibility assessment from robustness checks.
- 6) Section 4: more transparent description of the 3H Dataset construction and preprocessing workflow.

7) Sections 5 and 6: more consistent interpretation of the ARI as a complementary comparative hotspot-identification framework with explicit methodological boundaries.

Overall, we are very grateful for this comment. It helped us recognize that the most important way to improve the manuscript was to make the methodological core more explicit, reproducible, and internally consistent. We believe the revised manuscript now presents the ARI framework, the data-processing logic, and the evidentiary scope of the study in a substantially clearer and more rigorous way.

➤ Review Comments of Referee #5:

Comment 17:

Dear authors,

Thank you for your efforts. I have included my suggestions below.

Best

For Abstract

The ARI formulation is not described explicitly. Please briefly include the core mathematical idea (example: accumulation of WRI-weighted events).

Please soften the language “valuable resource” and “precise analyses” to “potentially useful for” like phrases.

Response 17:

We sincerely thank the referee for these specific and helpful suggestions regarding the Abstract. We fully agree that the Abstract should present the methodological core of the study more directly, and that the contribution of the 3H Dataset should be described in a more cautious and proportionate manner.

In response to the first point, we have revised the Abstract to state the core formulation idea of the ARI more explicitly. The revised text now explains, in concise form, that the ARI is constructed by aggregating the WRI values associated with major disaster events within $5^{\circ} \times 5^{\circ}$ latitude-longitude grids, so that readers can grasp the central methodological idea of the paper immediately at the Abstract level.

In response to the second point, we have also moderated the wording used to describe the remote sensing dataset and its role. We agree that expressions such as “valuable resource” and “precise analyses” were stronger than warranted by the current scope of the manuscript. The revised Abstract therefore uses more restrained language and presents the 3H Dataset as a complementary resource intended to support future finer-scale, observation-based analyses, rather than implying a stronger demonstrated capability than the paper currently provides.

For ease of review, the revised Abstract now reads as follows:

“This study introduces an Accumulated Risk Index (ARI) to map recent global multi-disaster hotspots, leveraging Emergency Events Database (EM-DAT) records from 2013 to 2023. By aggregating World Risk Index (WRI) values for major disaster events within $5^{\circ} \times 5^{\circ}$ latitude-longitude grids, the ARI couples recent hazard occurrences with country-level contextual risk conditions. Results show that 153 grids (5.9% of the global total) have non-zero ARI values, predominantly concentrated across South and Southeast Asia and along the Himalayan belt. Compared to a spatially equivalent WRI, the ARI captures sharper contrasts among neighbouring grids, illuminating the accumulated disaster burden at both transboundary and subnational scales. To facilitate ongoing research in these critical zones, this study additionally presents

the 3H Dataset, a standardized compilation of sub-metre remote sensing imagery derived from existing open sources and supplementary acquisitions. This unified dataset offers a complementary resource to support finer-scale, observation-based analyses in high-risk, data-scarce developing regions.”

We appreciate this comment because it helped us align the Abstract more closely with both the methodological contribution and the evidentiary boundaries of the manuscript. We believe these revisions improve the precision, balance, and overall readability of the Abstract.

Comment 18:

Introduction

The review is descriptive but not critical enough. For example, explicitly compare methodological paradigms (indicator-based vs statistical vs hybrid). Clarify why existing indices fail quantitatively.

The paper implies a gap but does not state a clear research question. Add a statement like “This study addresses whether a grid-based accumulation framework improves global multi-hazard risk differentiation.

Response 18:

We sincerely thank the referee for this important and constructive suggestion. We fully agree that the Introduction in the previous version was more descriptive than analytical, and that the methodological positioning of the study could be stated more clearly. We also agree that the manuscript benefited from a more explicit statement of the central research question.

In response, we have revised the Introduction in two main ways.

First, we have strengthened the literature review by organizing it more explicitly around three broad methodological paradigms: 1) indicator-based approaches, which characterize disaster risk through composite measures of exposure, vulnerability, and coping capacity; 2) statistical or event-based approaches, which rely more directly on historical disaster outcomes such as mortality and economic loss; and 3) hybrid approaches, which combine historical disaster records with broader socio-economic context.

The revised text now makes clear that these paradigms are complementary rather than mutually exclusive, but that they are not equally well suited to all analytical purposes. In particular, for the specific objective of this study, identifying where recent major multi-disaster shocks have accumulated at a globally comparable yet spatially differentiated scale, several limitations remain important. Long-term national indices may smooth out recent high-impact event accumulation; country-based assessment units may obscure subnational heterogeneity and transboundary corridors; and purely event-based approaches may lack sufficient contextual information for interpretation. We have therefore revised the Introduction to frame the contribution of the study more critically, while avoiding the implication that existing indices are unsuccessful or quantitatively invalid. Our intention is not to suggest that existing frameworks “fail”, but rather to clarify that they were developed for different purposes and may be less sensitive to the particular phenomenon addressed here.

Second, we have added a more explicit research-question statement in *Section 1.2 Motivation and Objectives*. The revised manuscript now states directly that the study investigates whether a grid-based accumulation framework, coupling recent major disaster events with country-level contextual risk information, can improve the spatial differentiation of global multi-disaster hotspots beyond conventional national indices. We agree that this explicit formulation makes the logic of the Introduction much clearer and helps connect the literature review more directly to the Methods, Results, and Discussion.

For ease of review, the revised manuscript now includes the following related formulations:

1) *Section 1.1 Background:*

“As summarized in Table 1, representative global multi-disaster risk assessment frameworks broadly fall into three methodological paradigms. Indicator-based approaches (e.g., WRI and INFORM) emphasize exposure, vulnerability, and coping-capacity conditions through composite indicators and are therefore effective for portraying relatively stable structural risk patterns. Statistical or event-based approaches (e.g., DRI and Natural Disaster Hotspots) rely more directly on historical disaster outcomes, such as mortality and economic loss, and are therefore more sensitive to realized impacts. Hybrid approaches (e.g., TRI and MhRI) attempt to combine historical disaster records with broader socio-economic context. These paradigms are complementary rather than mutually exclusive, but they are not equally suitable for all analytical purposes. For the specific objective of this study, identifying where recent major multi-disaster shocks have accumulated at a globally comparable yet spatially differentiated scale, several limitations remain noteworthy.

First, in terms of data timeliness, many existing frameworks are designed to characterize long-term or slowly changing risk trends. Even with regular updates, they may not respond strongly to the concentration of major disaster shocks within a relatively short period, such as the most recent decade. Second, in terms of assessment unit, country-based frameworks facilitate the integration of socio-economic indicators, but they are often too coarse to represent subnational heterogeneity or transboundary disaster corridors, limiting their compatibility with subsequent remote-sensing-based analysis (Shi and Kasperson, 2015; Dilley, 2005; Paprotny et al., 2018; Mucke et al., 2011). Third, in terms of hazard representation, some frameworks focus on a limited set of natural hazards or mix natural and human-induced hazards, which complicates direct comparison in a unified multi-hazard setting. Fourth, in terms of impact indicators, the choice of unified parameters varies widely across studies, from mortality-centred indicators to broader social and economic composites, and each choice involves a trade-off between comparability, interpretability, and multidimensional completeness (Yarveysi et al., 2023; Dilley, 2005; UNDP, 2004; Marin et al., 2021; Frege et al., 2023; Mucke et al., 2022). These considerations suggest that existing frameworks provide important reference points, but they may not fully capture the recent accumulated effects of major disaster events at subnational and transboundary scales.”

2) Section 1.2 Motivation and Objectives:

“Despite substantial progress in existing risk indices, capturing recent and spatially concentrated disaster impacts at a globally comparable scale remains challenging. First, long-term averages of hazard occurrence and socio-economic indicators may obscure the accumulated impacts of recurrent disaster events observed in the most recent decade. When major events cluster within a relatively short period, their compounded effects may place repeated pressure on local coping capacity and recovery processes, particularly in developing countries (Ridder et al., 2020; Rentschler et al., 2022; Frege et al., 2023). Capturing these near-term accumulated effects therefore requires a complementary index that explicitly emphasizes recent, high-impact disaster events rather than multi-decade averages. Second, the prevalent use of country-level assessment units restricts the ability to identify subnational or transboundary hazard clusters. Natural hazards follow geophysical processes rather than political borders, and their spatial footprint often spans multiple administrative units. These concerns were also highlighted in previous research (Paprotny et al., 2018; Mucke et al., 2011) and reinforce the need for methodological innovation. In particular, a more spatially explicit framework is needed to better capture cross-border corridors, within-country heterogeneity, and localized accumulated disaster pressure.

Accordingly, this study investigates whether a grid-based accumulation framework, coupling recent major disaster events with country-level contextual risk information, can improve the spatial differentiation of global multi-disaster hotspots beyond conventional national indices. Motivated by this objective, this study introduces the ARI, a global multi-disaster risk assessment index designed to specifically capture the accumulated effects of major disaster events from 2013 to 2023. The index employs 5°×5° latitude-longitude grids as its basic assessment unit and incorporates WRI as a contextual weighting coefficient. The proposed ARI aims to provide a complementary perspective that links recent disaster dynamics with the underlying socio-economic environment.”

We appreciate this comment because it helped us sharpen both the critical framing of the literature review and the clarity of the manuscript's central research question. We believe these revisions have substantially improved the methodological positioning of the paper.

Comment 19:

Data acquisition

Over 50 deaths, this threshold introduces selection bias toward high-mortality events.

It ignores economic loss from disasters and slow-onset hazards.

Please provide a sensitivity analysis or a justification for why 50? What happens at 10 or 100?

You dismiss other indicators too strongly. Please acknowledge limitations, like “This study prioritizes mortality but does not capture economic or indirect impacts.”

Response 19:

We sincerely thank the referee for this important and highly constructive set of comments. We fully agree that using Total Deaths ≥ 50 as the screening threshold introduces a degree of selection bias toward high-mortality events, and that such a choice cannot represent the full spectrum of disaster consequences, especially economic losses, indirect or delayed impacts, and many slow-onset hazard processes. We also appreciate the referee's observation that our earlier wording placed too much emphasis on the limitations of alternative indicators without being sufficiently explicit about the limitations of mortality itself. This criticism has been very helpful in prompting us to clarify both the rationale and the boundaries of our approach more explicitly.

First, we have revised the manuscript to state more clearly that the use of mortality in this study does not imply that deaths are the only, or the most comprehensive, representation of disaster impact. Rather, mortality is adopted here as a pragmatic screening variable under global data constraints. Among the impact variables available in EM-DAT, several alternatives, such as No. Affected, No. Homeless, and especially economic-loss indicators, often involve substantial cross-country inconsistency, missingness, heterogeneous estimation procedures, and uneven reporting quality. These issues are particularly pronounced in many developing countries, which are central to the present study. By comparison, although mortality is certainly imperfect and influenced by many contextual factors, it remains one of the more consistently reported and comparatively interpretable indicators across countries. We therefore now state more explicitly that mortality is used not as a comprehensive definition of disaster risk, but as a globally comparable proxy for identifying recent high-impact events within the specific scope of this study.

Second, we have strengthened the manuscript's acknowledgement of what this choice does not capture. The revised text now states much more directly that the present framework prioritizes mortality-screened major events and therefore does not fully capture economic-loss-dominated disasters, indirect or cascading impacts, prolonged disruption, or many slow-onset hazard processes. This clarification now appears both in the data-filtering discussion and again in the broader limitations discussion, so that readers can understand this as a core interpretive boundary of the ARI rather than as a minor caveat.

Third, we have clarified the rationale for the threshold of 50 fatalities. We agree that this choice required a more explicit justification. In the revised manuscript, we now describe the threshold as a pragmatic methodological compromise rather than an absolute theoretical cut-off. A lower threshold would incorporate many smaller events and thereby increase sensitivity to uneven reporting bias, whereas a much higher threshold would reduce the sample size too severely for meaningful global spatial comparison. The threshold of 50 was therefore retained as an operational middle ground: high enough to focus on major disruptive shocks, but still broad enough to preserve sufficient sample size for grid-based

accumulation analysis.

We also appreciate the referee's question about what happens at 10 or 100, because it highlights the importance of understanding the behavior of the framework at both ends of the threshold range. In the revised manuscript, our formal sensitivity analysis remains focused on 40, 60, and 70 fatalities, which represent a reasonable neighborhood around the baseline setting and allow us to test whether moderate threshold variation changes the principal hotspot pattern. We have not added a full empirical analysis for 10 or 100 in the manuscript. However, we now explain the threshold logic more clearly: conceptually, a threshold close to 10 would move much nearer to the minimum event-entry logic of EM-DAT and would likely admit a much larger number of lower-severity events, thereby diluting the specific focus of the study on accumulated major shocks and increasing sensitivity to uneven reporting of smaller events. By contrast, a threshold as high as 100 would focus narrowly on only the most extreme events, but at the cost of substantially reducing the sample size and spatial coverage of the resulting hotspots. We therefore present 50 as a practical comparability-inclusiveness compromise, rather than as a uniquely correct value.

In addition, we have revised the manuscript so that the role of the threshold is now linked more explicitly to the evidentiary structure of the study. Section 3.3.2 now clarifies that the sensitivity analysis is intended to evaluate whether moderate threshold variation changes the principal hotspot pattern, not whether event counts remain constant. The revised wording makes clear that alternative thresholds affect the absolute number of included events and the magnitude of ARI values, but do not fundamentally alter the dominant hotspot geography.

For ease of review, the revised manuscript now includes the following related formulations:

1) *Section 2.1 Disaster event dataset and filtering strategy:*

“Accordingly, the present framework mainly targets sudden-onset hazard processes for which major event accumulation can be more directly represented within EM-DAT records and more plausibly connected to Earth observation applications.”

“The selection of “Total Deaths” as the primary criterion does not imply that mortality fully captures disaster risk or impact. Rather, it is adopted here as a pragmatic and comparatively more consistent proxy for recent high-impact events under global data constraints. Population-based indicators in EM-DAT, such as “No. Affected” and “No. Homeless”, are often influenced by differences in national definitions and estimation practices, which depend heavily on other metrics or overall event assessments (UNDP, 2004; Newman and Noy, 2023; Peduzzi et al., 2009). Similarly, 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). Consequently, mortality is used in this study as an operational screening variable with relatively stronger cross-country comparability.

Similarly, the threshold of ≥ 50 fatalities represents a pragmatic methodological compromise rather than an absolute theoretical cut-off. A lower threshold would incorporate numerous small-scale events, thereby exacerbating sensitivity to uneven reporting biases, whereas a much higher threshold would severely restrict the sample size required for robust global spatial comparisons. Therefore, this study employs the ≥ 50 criterion to retain a sufficient number of events for meaningful grid-based accumulation analysis, while selectively emphasizing disasters that cause major socio-economic disruption. The sensitivity of this choice is further evaluated in Section 3.3.2. Inherently, this also means that the present framework prioritizes mortality-screened major events and does not fully capture economic-loss-dominated disasters, indirect cascading impacts, or slow-onset hazard processes.”

“At the same time, this outcome may bias the final event composition toward hazard subtypes prone to generating short-term, high-mortality impacts.”

2) *Section 2.3.2 Accumulated Risk Index (proposed index):*

“In this formulation, ARI represents the accumulated sum of country-level WRI weights associated with recent major disaster events occurring within each grid cell, rather than a direct spatial aggregation of WRI itself. ARI should therefore be interpreted as an event-driven, grid-based, context-weighted relative index, rather than a complete physical-statistical representation of disaster risk.”

“This design keeps the interpretation of ARI transparent: once an event satisfies the major-event criterion defined in Section 2.1, it is counted as one accumulated shock, and differences in background risk are reflected through the associated country-level WRI value.”

3) *Section 3.3 External plausibility assessment and robustness checks:*

“Given that ARI is a newly proposed index, it is important to examine both whether its hotspot pattern shows meaningful external plausibility and whether its main spatial features remain stable under reasonable methodological variation.”

“Second, robustness checks are performed to evaluate the sensitivity of ARI to the fatality threshold used to define major disaster events and to the spatial resolution of the latitude-longitude grid.”

4) *Section 3.3.2 Sensitivity to the fatality threshold:*

“These values represent a reasonable neighborhood around the baseline setting and are used to evaluate whether moderate threshold variation changes the principal hotspot pattern.”

“These changes affect the absolute magnitude of ARI and the number of included events, but they do not fundamentally alter the dominant hotspot geography. In all scenarios, the highest-ranking grids remain concentrated in South and Southeast Asia, particularly in regions covering northern India, Bangladesh, and the Himalayan belt. Moreover, many of the top 10 grids under the alternative thresholds coincide with those in the baseline configuration, often with only moderate changes in rank. This pattern indicates that the identification of the principal high-risk clusters is not strongly dependent on one exact threshold value, even though the event counts and ARI magnitudes vary as expected.”

“Taken together, the out-of-sample plausibility assessment and the robustness checks indicate that ARI identifies spatially coherent and comparatively stable hotspot patterns at the global scale. The strongest claim supported by the current evidence is therefore not that ARI has been definitively validated as a complete measure of true disaster risk, but that it shows meaningful external plausibility when compared with the spatial occurrence of subsequent major disaster events and remains robust under reasonable variation in key methodological parameters. These results support the use of ARI as a comparative hotspot-identification framework, while detailed local interpretation should still consider the methodological uncertainties discussed in Section 5.”

5) *Section 5 Discussion:*

“This interpretation also helps place the limitations of the study in a more constructive light. The current framework is necessarily selective. Its mortality-based event filtering improves cross-country comparability and reduces the noise of under-reported small events, but it also privileges short-onset, high-fatality shocks and does not fully capture economic losses, indirect impacts, prolonged disruption, or many slow-onset processes. Similarly, the final hazard composition is shaped by the study period and inclusion rule, which means that the resulting event set should not be mistaken for a hazard-neutral representation of all disaster processes recorded in EM-DAT.”

“A third direction is to enrich the impact basis of the framework by exploring composite severity formulations that combine mortality with other dimensions such as affected population, economic disruption, infrastructure damage, or remotely sensed damage proxies when data quality allows.”

We sincerely appreciate this comment. It has helped us present the study in a more balanced and methodologically

transparent way, and it has significantly improved how the scope, assumptions, and limitations of the ARI are communicated to readers.

Comment 20:

You exclude some subtypes because there are no major events. This may distort hazard representativeness. Please clarify whether exclusion is methodological or data-driven, and the potential bias introduced.

Response 20:

We sincerely thank the referee for raising this important point. We fully agree that the omission of certain disaster subtypes from the final analysis may affect the apparent representativeness of the multi-hazard framework, and that it is therefore necessary to clarify whether this omission is methodological or data-driven, as well as what type of bias it may introduce.

In the present study, the exclusion of these subtypes is primarily data-driven rather than a priori methodological. In other words, we did not predefine these subtypes as irrelevant, nor did we intentionally remove them from the hazard taxonomy before conducting the analysis. Instead, all disaster types and subtypes within the scope of the study were initially considered under the same screening framework. However, under the selected study period (2013-2023) and the operational definition of a major event (Total Deaths ≥ 50), some subtypes, such as derecho, extra-tropical storm, storm surge, and coastal flood, did not have any recorded events that satisfied the inclusion threshold. As a result, they were not carried forward into the subsequent ARI calculation. Their absence in the final ARI analysis therefore reflects the outcome of a uniform filtering rule applied to the available data, rather than a separate methodological decision to exclude them.

At the same time, we fully accept the referee's broader concern that, even when such exclusion is data-driven, it can still introduce bias into hazard representativeness. This is an important point. Because the current framework identifies "major events" through a mortality-based threshold, it is more likely to retain hazard subtypes that tend to generate acute, high-fatality impacts over relatively short time periods. By contrast, subtypes that may produce substantial consequences through other pathways, such as major economic loss, prolonged disruption, repeated moderate impacts, or lower-mortality but still highly destructive processes, are less likely to appear prominently in the final sample if they do not cross the selected fatality threshold during the study window. We therefore agree that the resulting ARI should not be interpreted as a fully balanced representation of all natural hazard subtypes in EM-DAT. Rather, it is more accurately understood as a framework that captures the spatial accumulation of recent, mortality-screened, high-impact events under the current data and filtering conditions.

This distinction is especially important because the apparent dominance of the final event set by earthquakes, storms, floods, and mass movements could otherwise be misread as implying that these are the only globally relevant hazards. That is not our intention. What the current results show is only that, under the chosen temporal window and inclusion rule, these hazard groups accounted for the events that met the study's operational definition of "major". We agree that this point needed to be stated much more explicitly so that readers do not interpret the final ARI outputs as equally representative of all hazard subtypes or as a hazard-neutral summary of multi-hazard processes.

In response to this comment, we have clarified this issue in two main places. First, in Section 2.1, we now state directly that the missing subtypes were not removed by design, but were absent from the final analysis because no events of those subtypes met the major-event threshold within the selected study period. Second, in Section 5, we now acknowledge more explicitly that the current framework is selective and that the final hazard composition is shaped by both the study period and the inclusion rule, which means that the resulting event set should not be mistaken for a hazard-neutral representation of all disaster processes recorded in EM-DAT.

For ease of review, the revised manuscript now includes the following formulations:

1) *Section 2.1 Disaster event dataset and filtering strategy:*

“Notably, four subtypes (derecho, extra-tropical storm, storm surge, and coastal flood) are excluded from the final analysis, as there are no records of major events with at least 50 fatalities during the study period. This exclusion is fundamentally data-driven within the present framework, rather than an a priori methodological deletion. At the same time, this outcome may bias the final event composition toward hazard subtypes prone to generating short-term, high-mortality impacts.”

2) *Section 5 Discussion:*

“This interpretation also helps place the limitations of the study in a more constructive light. The current framework is necessarily selective. Its mortality-based event filtering improves cross-country comparability and reduces the noise of under-reported small events, but it also privileges short-onset, high-fatality shocks and does not fully capture economic losses, indirect impacts, prolonged disruption, or many slow-onset processes. Similarly, the final hazard composition is shaped by the study period and inclusion rule, which means that the resulting event set should not be mistaken for a hazard-neutral representation of all disaster processes recorded in EM-DAT.”

We are very grateful for this comment. It has helped us articulate more clearly both the logic of the inclusion process and the interpretive boundary of the resulting ARI patterns. We believe that making this distinction explicit improves the transparency of the manuscript and helps prevent overgeneralization of the hazard representativeness of the current results.

Comment 21:

Mathematical clarity

Equation (3) is poorly formatted and unclear, and variables are not clearly defined. Please rewrite the equation cleanly and define all variables explicitly.

Response 21:

We sincerely thank the referee for pointing out this important issue. We fully agree that, in the previous version of the manuscript, Equation (3) was not presented with sufficient clarity. Because the ARI formulation is the methodological core of the study, ambiguity in the equation layout, notation, or variable definitions can directly weaken the transparency, rigor, and reproducibility of the framework. We therefore took this comment very seriously.

We agree that the problem was not only typographical, but also interpretive. Although the surrounding text already indicated that ARI is constructed by accumulating major disaster events within each grid cell and associating each event with a country-level WRI value, the equation itself did not communicate this logic clearly enough. As a result, readers could reasonably find it difficult to determine exactly what was being summed, what each index referred to, and whether the formula represented an aggregation of events, an aggregation of WRI values, or a spatial transformation of the WRI itself. We therefore revised both the equation and its accompanying explanation in a more systematic way.

First, we rewrote Equation (3) using a cleaner and more explicit summation form, so that the construction logic of ARI is immediately visible from the formula itself. The revised presentation now makes clear that the ARI for a given grid cell is obtained by summing the country-level WRI values associated with the major disaster events assigned to that grid during the study period. We also state the boundary condition explicitly by defining $ARI = 0$ when no major disaster events are assigned to the grid.

Second, we expanded the variable definitions immediately following the equation so that every symbol is explicitly identified. In the revised manuscript, ARI_j , n_j , the event index i , the country-association function $c(i)$, and $WRI_{c(i)}$ are each defined directly in connection with the formula. This makes it much clearer that the fundamental unit entering the summation is the event, while the WRI serves as a country-level contextual weight attached to that event, rather than as an

independently gridded variable.

Third, beyond the equation itself, we also strengthened the surrounding methodological explanation so that the interpretive boundary of the formula is clearer. The revised text now states more explicitly that, under the current implementation, all major events that pass the inclusion threshold enter the ARI calculation with equal event-count status, while differences in contextual weighting arise through the WRI value of the corresponding country. In this sense, ARI is now described more clearly as a WRI-contextualized event-accumulation index, rather than as a complete disaster-risk function that directly incorporates event severity gradients, physical footprints, economic losses, or multidimensional impact measures.

For ease of review, the revised manuscript now reads as follows in *Section 2.3.2 Accumulated Risk Index (proposed index)*:

“Instead, it is adopted as a country-level reference index and contextual weighting factor, encapsulating the broader exposure-vulnerability setting in which disaster events occur. The proposed ARI builds on this reference framework by combining recent major disaster events with national WRI values at the grid scale. Conceptually, ARI integrates two components: 1) the number of major disaster events assigned to each grid cell from 2013 to 2023, based on the event filtering strategy in Section 2.1 and the spatial standardization procedure in Section 2.2; and 2) the WRI value of the country associated with each event, which adjusts for differences in exposure and vulnerability among affected regions. In this way, ARI emphasizes the short-term accumulation of major disaster shocks in contexts with varying underlying risk conditions. The index is defined as:

$$ARI_j = \sum_{i=1}^{n_j} WRI_{c(i)}, \quad n_j \geq 1, \quad (3)$$

where ARI_j is the ARI of grid cell j , n_j is the number of major disaster events assigned to grid cell j from 2013 to 2023, i indexes the events occurring in that grid cell, $c(i)$ denotes the country associated with the i -th event according to the EM-DAT record, and $WRI_{c(i)}$ is the WRI value of that country. When $n_j = 0$, ARI_j is strictly defined as 0. In this formulation, ARI represents the accumulated sum of country-level WRI weights associated with recent major disaster events occurring within each grid cell, rather than a direct spatial aggregation of WRI itself. ARI should therefore be interpreted as an event-driven, grid-based, context-weighted relative index, rather than a complete physical-statistical representation of disaster risk.”

We are very grateful for this comment, because it prompted a necessary improvement to one of the most central parts of the manuscript.

Comment 22:

Remote Sensing Dataset

The dataset feels somewhat detached from the risk model. Please explicitly link: how ARI guides dataset selection and potential ML applications.

Response 22:

We sincerely thank the referee for this helpful suggestion. We agree that, in the previous version of the manuscript, the relationship between the remote sensing dataset and the ARI-based risk model was not stated explicitly enough. Although the 3H Dataset was presented after the ARI hotspot results, the manuscript did not sufficiently clarify how ARI guided the spatial selection of the dataset or how the dataset could support subsequent machine-learning applications. We therefore revised the manuscript to make this linkage more direct and methodologically coherent.

The key clarification is that the 3H Dataset was not constructed as an independent remote sensing add-on. Instead, it is designed as an ARI-guided operational extension of the risk assessment framework. In the revised manuscript, we now

explain that ARI first identifies $5^{\circ}\times 5^{\circ}$ grid cells where recent major disaster shocks have accumulated under different country-level WRI contexts. These ARI-identified hotspot grids then define the spatial scope within which high-resolution visible-spectrum remote sensing imagery is collected, integrated, supplemented where necessary, standardized, and organized. In this sense, ARI answers the question of where accumulated disaster pressure is concentrated, while the 3H Dataset provides a standardized Earth observation basis for later finer-scale analysis within those prioritized areas.

We have also clarified that this linkage operates primarily at the hotspot-grid level, rather than at the level of exact disaster footprints. The 3H Dataset is therefore not presented as direct image-based validation of ARI, nor as an event-footprint dataset. Rather, it is intended to make ARI-identified high-risk and data-scarce regions more accessible for future observation-based research. This distinction is now stated more explicitly to avoid overinterpreting the dataset as a direct spatial representation of each individual disaster event.

In response to the referee's suggestion regarding machine-learning applications, we also strengthened the discussion of the dataset's potential use. The revised manuscript now states that the standardized imagery may support future tasks such as building damage detection, exposure mapping, post-disaster recovery monitoring, disaster-related feature extraction, and domain adaptation in high-risk developing regions. We present these as potential downstream applications, not as analyses already completed in the present study. This revised framing better reflects the current contribution: ARI provides hotspot prioritization, while the 3H Dataset provides a data infrastructure for subsequent remote sensing and machine-learning studies in those prioritized locations.

The revised manuscript now includes the following clarifications:

1) *Abstract:*

“To facilitate ongoing research in these critical zones, this study additionally presents the 3H Dataset, a standardized compilation of sub-metre remote sensing imagery derived from existing open sources and supplementary acquisitions. This unified dataset offers a complementary resource to support finer-scale, observation-based analyses in high-risk, data-scarce developing regions.”

2) *Section 1.1 Background:*

“Consequently, while remote sensing is invaluable for fine-scale disaster analysis, its direct and uniform deployment within a global risk-assessment framework remains difficult. This practical constraint further motivates the need for an intermediate spatial screening strategy to identify priority areas for subsequent Earth observation analysis.”

3) *Section 1.2 Motivation and Objectives:*

“In addition, in order to facilitate further remote sensing-based research, this study systematically integrates and supplements existing sub-metre visible-spectrum remote sensing imagery for ARI-identified high-risk grids, and publishes them as an open-source dataset after standardization. Within the present framework, the ARI dictates spatial prioritization, while the resulting 3H Dataset provides a foundational resource for potential fine-scale applications, including building damage detection, exposure mapping, and post-disaster recovery monitoring.”

4) *Section 2.1 Disaster event dataset and filtering strategy:*

“Accordingly, the present framework mainly targets sudden-onset hazard processes for which major event accumulation can be more directly represented within EM-DAT records and more plausibly connected to Earth observation applications.”

5) *Section 2.5 Analytical workflow:*

“In this sense, ARI serves as the hotspot-screening mechanism that guides where remote sensing resources are prioritized, while the 3H Dataset provides the geospatial data foundation for subsequent finer-scale studies in those selected high-risk

areas.

This workflow demonstrates that the remote sensing component is not independent of the risk model, but is developed as an operational extension of it. Specifically, ARI identifies where recent accumulated disaster pressure is concentrated, and the 3H Dataset translates those hotspot results into a standardized imagery resource that can support future applications. Together, these methodological stages constitute an integrated framework linking historical disaster records, contextual risk information, and Earth observation data.”

6) *Section 4 Remote sensing data integration for ARI hotspots:*

“In this study, the remote sensing component is developed as a complementary operational extension of the ARI framework. ARI first identifies hotspot grids where recent accumulated disaster pressure is concentrated, and remote sensing data are then compiled and supplemented within those ARI-identified locations. The resulting 3H Dataset is therefore organized at the hotspot-grid level and is intended to support subsequent finer-scale Earth observation and machine-learning studies in these areas. Within the present manuscript, its role is to provide a standardized geospatial data foundation rather than to serve as a direct image-based validation of ARI.”

7) *Section 4.3 Standardization and construction of the 3H Dataset:*

“By focusing explicitly on ARI-identified hotspot grids in developing countries and by providing imagery in a standardized format, the 3H Dataset is intended to support future finer-scale studies such as building damage detection, exposure mapping, and post-disaster recovery monitoring. In this way, it complements the macro-scale ARI framework by making high-risk, data-scarce locations more accessible for subsequent Earth observation analysis.”

8) *Section 5 Discussion:*

“For researchers, the 3H Dataset extends the ARI framework from hotspot identification to data infrastructure by making sub-metre imagery available in a standardized form for high-risk, data-scarce regions.”

“Seen from this perspective, the broader significance of the study is not only methodological but also infrastructural. ARI provides a macro-scale screening logic for identifying where recent accumulated pressure is concentrated, while the 3H Dataset offers a micro-scale observational basis for examining what those hotspots look like on the ground. The two components therefore operate at different but connected scales. Together, they suggest a possible pathway for linking global disaster statistics, contextual risk conditions, and Earth observation resources in a more operationally coherent way than is commonly achieved in either purely index-based or purely image-based studies.”

“Finally, the ARI-guided 3H Dataset opens an avenue for subsequent machine-learning applications, including building damage detection, exposure mapping, recovery monitoring, and domain adaptation in data-scarce regions. Such finer-scale outputs may, in turn, provide new evidence for further refining and evaluating accumulated-risk indicators.”

9) *Section 6 Conclusions:*

“Building on the ARI-identified hotspot grids, this study also develops the 3H Dataset, which contains more than 7.5 million standardized sub-metre remote sensing image tiles across 14 high-risk grids in 10 developing countries. Together, ARI and the 3H Dataset provide a connected basis for linking global disaster statistics, contextual risk information, and Earth observation resources, thereby supporting subsequent research and practice in multi-disaster hotspot identification and finer-scale geospatial analysis.”

We are grateful for this comment, because it helped us clarify that the 3H Dataset should be understood not as a detached dataset or as direct validation of ARI, but as an ARI-guided data-support layer for future finer-scale remote sensing and machine-learning research.

Comment 23:

Image filtering criteria are vague, for instance, “removing damaged or irrelevant images.”

Please define quality criteria and a preprocessing pipeline.

Response 23:

We sincerely thank the referee for this important comment. We fully agree that the previous wording, particularly expressions such as “removing damaged or irrelevant images”, was too vague and could cause ambiguity. In a disaster-related remote sensing study, the term “damaged images” may be misunderstood as referring to images showing disaster-damaged ground objects, whereas our original intention was to refer to images with poor data quality or limited analytical suitability. We have therefore revised the manuscript to define the image filtering criteria and preprocessing pipeline more explicitly.

In the revised manuscript, we clarify that image filtering was based on image quality and suitability for downstream remote sensing analysis, not on whether disaster damage was visible on the ground. Specifically, candidate scenes were excluded when they met any of the following conditions: 1) severe image-quality defects, such as corrupted files, missing regions, black padding, striping, severe artefacts, or other visual defects that prevented reliable interpretation; 2) substantial cloud, cloud shadow, haze, or atmospheric contamination that obscured built-up targets; and 3) little or no meaningful built-up content, such as scenes dominated almost entirely by open water, dense vegetation, bare land, or other non-built surfaces, which were poorly aligned with the intended downstream applications of the dataset.

We also revised the manuscript to describe the preprocessing workflow as a structured pipeline rather than a brief narrative statement. The updated workflow consists of five steps: 1) ARI-guided hotspot grid selection; 2) collection of candidate imagery within the selected grids from DigitalGlobe Open Data, RAMP-related sources, and supplementary acquisitions; 3) visual quality screening of candidate scenes according to the criteria described above; 4) spatial standardization and tiling into fixed-size 256×256 pixel TIF patches; and 5) metadata assignment and final organization by hotspot-grid identifier and data source.

We further clarify that the quality-control process was primarily based on manual visual inspection rather than a fully automated filtering framework. This was a pragmatic choice given the heterogeneity of the source imagery and the different acquisition conditions of the open datasets used. We now acknowledge this as a limitation of the current preprocessing pipeline, while also making clear that the purpose of the screening step was to remove clearly unsuitable imagery before standardization and patch generation.

The revised manuscript now states:

Section 4.3 Standardization and construction of the 3H Dataset:

“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 visually inspected to remove imagery with severe artefacts, substantial cloud, cloud-shadow, or haze contamination that prevented reliable interpretation of built-up targets. In addition, scenes with little or no meaningful built-up content were excluded because they were poorly aligned with the intended downstream applications of the dataset. At the present stage, this screening procedure was based primarily on manual visual inspection rather than a fully automated filtering framework. 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

so that each standardized unit could be linked back to its originating imagery and ARI-identified grid location. This workflow was applied consistently across the 14 high-risk grids.”

We believe these revisions improve the transparency and reproducibility of the 3H Dataset construction process, and we are grateful to the referee for prompting us to replace vague terminology with clearer quality criteria and a more explicit preprocessing workflow.

Comment 24:

Also, please discuss geographic imbalance, not only focus on Asia.

Response 24:

We sincerely thank the referee for this important suggestion. We fully agree that the manuscript should discuss the geographic imbalance of the 3H Dataset more explicitly, rather than focusing mainly on the Asian hotspot regions. Without such clarification, readers may incorrectly interpret the dataset as a globally balanced remote sensing benchmark, which is not the intended positioning of this study.

In the revised manuscript, we clarify that the geographic imbalance arises from two related but distinct factors. First, the ARI results themselves are spatially uneven. Under the 2013–2023 study period and the adopted mortality-based event-screening rule, the highest ARI grids are concentrated mainly in South Asia, Southeast Asia, and the Himalayan belt. Therefore, because the 3H Dataset is designed to follow ARI-identified hotspot grids, its spatial distribution naturally reflects this hotspot pattern. Second, the availability of open sub-metre remote sensing imagery is also geographically uneven. Public high-resolution imagery often depends on crisis-response archives, humanitarian mapping projects, and existing open-data initiatives, which do not provide uniform coverage across all global high-risk regions. As a result, even where ARI identifies potential hotspots outside Asia, the ability to compile standardized open imagery may be constrained by data availability.

We have therefore revised the manuscript to state more clearly that the 3H Dataset is a hotspot-oriented and data-availability-constrained collection, rather than a geographically balanced global benchmark. This distinction is important for interpreting the dataset. The current coverage is useful for supporting finer-scale Earth observation studies in selected high-risk, data-scarce regions, but it should not be used to draw conclusions about global geographic representativeness. In particular, models developed or evaluated using the current dataset may still require further testing before being transferred to underrepresented regions such as Africa, the Americas, or other non-Asian disaster-prone areas.

The revised manuscript now states:

1) *Section 4.1 Data sources and spatial coverage:*

“The resulting spatial coverage is not globally balanced. This concentration is produced by two interacting factors: first, ARI hotspots during 2013-2023 are themselves concentrated mainly in South Asia, Southeast Asia, and the Himalayan belt; second, the global availability of open high-resolution imagery is geographically uneven. As a result, the 3H Dataset should be understood as a targeted hotspot-oriented collection rather than a globally balanced remote sensing benchmark.”

2) *Section 5 Discussion:*

“Importantly, this dataset should be understood in the same spirit as ARI itself: not as a globally balanced benchmark, but as a targeted resource oriented toward empirically identified hotspot zones, especially in developing countries where the gap between disaster exposure and data accessibility remains pronounced.”

3) *Section 6 Conclusions:*

“Building on the ARI-identified hotspot grids, this study also develops the 3H Dataset, which contains more than 7.5 million

standardized sub-metre remote sensing image tiles across 14 high-risk grids in 10 developing countries. Together, ARI and the 3H Dataset provide a connected basis for linking global disaster statistics, contextual risk information, and Earth observation resources, thereby supporting subsequent research and practice in multi-disaster hotspot identification and finer-scale geospatial analysis.”

We appreciate this comment because it helped us present the remote sensing component more carefully. The revised manuscript now makes clear that the 3H Dataset is not intended to provide uniform global coverage, but rather to serve as an ARI-guided first step toward linking macro-scale disaster hotspot identification with finer-scale Earth observation resources.

Comment 25:

For Conclusion

There is an overgeneralization. Claims about “bias in traditional methods” are not rigorously proven. Please tone down or support with a quantitative comparison.

For the limitation, please add data bias (EM-DAT), mortality threshold bias, spatial simplification, and ARI assumptions.

Your future work is too generic. Please add concrete directions, dynamic ARI, integration with climate projections, and machine learning applications.

Regards.

Response 25:

We sincerely thank the referee for these thoughtful and comprehensive suggestions regarding the Conclusion. We fully agree that the concluding section should be particularly cautious, well-bounded, and strictly aligned with the evidence actually provided in the manuscript. We also appreciate that this comment does not only point to a wording problem, but to a broader issue of scientific positioning: the Conclusion should clearly distinguish between what the present study has demonstrated, what it can reasonably suggest, and what remains beyond the current analytical scope. This comment has therefore helped us revise the final part of the manuscript in a more disciplined and transparent way.

First, regarding the referee’s concern about overgeneralization, we agree that the previous wording concerning possible “bias in traditional methods” was too strong. In the original manuscript, our intention was to indicate that long-term, country-level risk frameworks may not fully reveal the short-term spatial concentration of recent major disaster events at subnational or transboundary scales. However, we recognize that using the word “bias” could be interpreted as a general criticism of existing risk-assessment methods, which is not sufficiently supported by the evidence presented in this study. Demonstrating systematic bias in traditional or established methods would require a dedicated comparative evaluation, including multiple benchmark indices, independent validation datasets, and formal quantitative tests. Such an analysis is beyond the scope of the present manuscript.

Therefore, in the revised manuscript, we have chosen to tone down this claim rather than attempt to overextend the present comparison. We now avoid suggesting that traditional methods are “biased” in a general sense. Instead, we state more carefully that country-level structural indices, such as the WRI, remain highly valuable for representing broad exposure and vulnerability conditions, but may not fully express where recent major disaster shocks have clustered within a relatively short period. This revision better reflects the actual contribution of ARI: it is proposed not as a replacement for existing global risk indices, but as a complementary, event-driven perspective that highlights recent accumulated disaster pressure at the grid scale.

Second, we agree that the limitations of the study should be summarized more explicitly, especially in the final sections of the manuscript. Although several limitations were discussed in earlier methodological sections, we recognize that the

Conclusion should provide readers with a clear and consolidated understanding of the boundaries of ARI. In response, we have added a more systematic discussion of four major limitations:

- 1) The first limitation concerns data bias inherited from EM-DAT. EM-DAT provides one of the most important and widely used global disaster event databases, and it is indispensable for a study of this scale. Nevertheless, it is affected by reporting heterogeneity across countries, time periods, hazard types, and impact indicators. Smaller events, slow-onset processes, indirect losses, and economic-loss-dominated disasters may be unevenly recorded or underrepresented. In addition, the spatial information in EM-DAT is often based on textual location descriptions rather than detailed hazard footprints. Therefore, ARI necessarily inherits part of the uncertainty and unevenness of the underlying disaster-event records.
- 2) The second limitation concerns mortality-threshold bias. The use of “Total Deaths ≥ 50 ” was adopted as a pragmatic screening rule to identify recent major events with relatively stronger cross-country comparability. However, we fully agree that this choice gives priority to high-mortality events and may underrepresent disasters whose main impacts are expressed through economic loss, affected population, infrastructure disruption, long-term recovery burden, or slow-onset processes. Thus, ARI should not be interpreted as a comprehensive measure of all disaster impacts. It is more accurately described as an index of recent accumulated major disaster shocks under a mortality-based screening strategy.
- 3) The third limitation concerns spatial simplification. To enable global grid-based aggregation, each disaster event is assigned to a representative coordinate and then to a latitude-longitude grid cell. This standardization is necessary for implementing a consistent global framework, but it inevitably compresses the real spatial footprint of many events, especially floods, storms, and other geographically extensive hazards. Consequently, ARI may not fully represent the internal spatial heterogeneity of large-area disasters, and neighboring affected areas may be simplified when a single representative location is used. We now state this limitation more explicitly and emphasize that ARI is suitable for macro-scale hotspot identification, not for precise local-scale hazard mapping.
- 4) The fourth limitation concerns the simplifying assumptions embedded in ARI itself. In the current formulation, all events above the fatality threshold enter the calculation as discrete accumulated shocks, and their contextual weight is determined by the country-level WRI value associated with each event. This design makes the index transparent and reproducible, but it also means that ARI does not explicitly model event severity gradients above the threshold, physical hazard intensity, actual damage footprint, dynamic vulnerability, coping capacity after repeated shocks, or recovery processes over time. We therefore clarify that ARI is a relative, event-driven, context-weighted hotspot-identification framework, rather than a complete physical-statistical model of disaster risk.

Third, we agree that the future work described in the previous version was too generic. Following the referee’s suggestion, we have revised this part to provide more concrete and actionable directions:

- 1) One important direction is the development of a dynamic ARI. The current ARI is a static accumulation index over the 2013-2023 period. Future work could introduce temporal weighting, time-decay functions, or recovery-sensitive accumulation mechanisms. For example, more recent events could be assigned greater influence, or repeated shocks occurring before full recovery could be weighted differently from isolated events. Such extensions would allow ARI to better represent temporal sequences of disaster pressure and potential resilience erosion.
- 2) A second direction is to integrate ARI with climate projections and forward-looking hazard scenarios. The present study is retrospective, as it uses observed disaster records from EM-DAT. Future studies could connect ARI hotspots with projected changes in hazard frequency, intensity, or spatial distribution under different climate scenarios. This would help bridge recent accumulated disaster impacts with future risk trajectories and would make the framework more useful for anticipatory risk reduction and adaptation planning.

3) A third direction is to improve event representation and impact characterization. As data quality improves, future versions of ARI could incorporate additional dimensions such as affected population, economic loss, infrastructure damage, housing damage, recovery duration, or remotely sensed hazard footprints. This would help reduce the current dependence on a mortality-based threshold and make the framework more sensitive to hazards whose severity is not primarily reflected in fatalities.

4) A fourth direction is to develop machine-learning applications based on the ARI-guided 3H Dataset. In the revised manuscript, we clarify that the 3H Dataset is not intended to directly validate ARI. Instead, it provides a standardized remote sensing data foundation for finer-scale studies in ARI-identified high-risk and data-scarce regions. Future research can build on this dataset for building damage detection, exposure mapping, built-up area characterization, post-disaster recovery monitoring, domain adaptation, and weakly supervised learning in disaster-prone developing regions. Such Earth observation outputs may also provide new evidence for evaluating and refining accumulated-risk indicators in later studies.

Overall, we are grateful for this comment because it has helped us improve the scientific balance of the Conclusion. In the revised manuscript, we have therefore made three main changes: 1) we removed or toned down overgeneralized statements about “bias” in existing methods and repositioned ARI as a complementary framework; 2) we added a clearer summary of the main limitations, including EM-DAT data bias, mortality-threshold bias, spatial simplification, and ARI’s simplifying assumptions; and 3) we revised the future work discussion to include more specific directions, particularly dynamic ARI development, integration with climate projections, improved event-footprint and impact representation, and machine-learning applications based on the 3H Dataset.

Revised content now appears as follows:

1) *Section 5 Discussion:*

“The central contribution of this study lies in reframing global multi-disaster assessment from a purely structural or country-based perspective toward a spatially explicit view of recent accumulated disaster pressure. Existing global indices such as the WRI remain highly valuable because they provide a broad and systematic representation of exposure and vulnerability conditions at the national scale. However, the results of this study suggest that such structural risk context alone may not fully express where major disaster shocks have clustered most intensely in recent years. In this sense, the ARI does not seek to replace long-term structural risk metrics, but to complement them by emphasizing a dimension that is often muted in country-level frameworks: the short-term accumulation of major disaster events across subnational and transboundary space. This complementary role is most clearly reflected in the sharper hotspot contrasts observed in ARI, particularly along the Himalayan belt, South and Southeast Asia, and other cross-border corridors where repeated high-impact events have occurred within a relatively short period.

From a broader disaster-risk perspective, this distinction is important because resilience deterioration is not only a function of background vulnerability, but also of repeated shock exposure over time. A location may be embedded in a high-vulnerability national context without having experienced concentrated recent shocks, just as a location with only moderate country-level background risk may nonetheless emerge as a hotspot when multiple major events accumulate in close spatial proximity. The comparison between ARI and Equivalent WRI illustrates precisely this point: ARI contributes an additional temporal-spatial layer to global disaster assessment by identifying where recent disaster accumulation may have intensified pressure beyond what long-term structural indices alone would suggest. The added value of the proposed framework therefore lies less in redefining disaster risk in a universal sense than in revealing where recent accumulated impacts may warrant heightened analytical and practical attention.”

“This interpretation also helps place the limitations of the study in a more constructive light. The current framework is necessarily selective. Its mortality-based event filtering improves cross-country comparability and reduces the noise of

under-reported small events, but it also privileges short-onset, high-fatality shocks and does not fully capture economic losses, indirect impacts, prolonged disruption, or many slow-onset processes. Similarly, the final hazard composition is shaped by the study period and inclusion rule, which means that the resulting event set should not be mistaken for a hazard-neutral representation of all disaster processes recorded in EM-DAT. At the spatial level, the use of one representative coordinate per event makes global grid-based aggregation feasible, but inevitably compresses the true footprint of geographically extensive hazards such as floods and storms. At the contextual level, ARI depends on the strengths and weaknesses of both EM-DAT and WRI: the former is affected by reporting heterogeneity and imperfect geolocation, while the latter provides only country-level contextual risk information and aggregates diverse indicators with uneven temporal update cycles. These are not minor caveats. They define the scope within which the present results should be read. Yet they also point directly to the specific directions in which accumulated-risk assessment can be improved.”

“These considerations also make the next steps for future research more concrete. A first priority is to move from the present static formulation toward dynamic or temporally weighted ARI variants, so that more recent shocks, or repeated shocks occurring before recovery is complete, can be represented more realistically. A second priority is to improve event representation itself through finer geocoding, better footprint allocation, and the integration of Earth-observation-derived hazard extents, especially for spatially extensive disasters. A third direction is to enrich the impact basis of the framework by exploring composite severity formulations that combine mortality with other dimensions such as affected population, economic disruption, infrastructure damage, or remotely sensed damage proxies when data quality allows. A fourth direction is to link ARI more explicitly with forward-looking analysis, including projected hazard changes under climate scenarios and evolving socio-economic conditions, so that the current retrospective accumulation logic can be related to possible future risk trajectories. Finally, the ARI-guided 3H Dataset opens an avenue for subsequent machine-learning applications, including building damage detection, exposure mapping, recovery monitoring, and domain adaptation in data-scarce regions. Such finer-scale outputs may, in turn, provide new evidence for further refining and evaluating accumulated-risk indicators.

In summary, the present study suggests that recent disaster accumulation deserves more explicit attention in global risk assessment than it typically receives in long-term country-based frameworks. ARI does not resolve the full complexity of global disaster risk, nor is it intended to do so. Its contribution is more focused: to identify where major disaster shocks have recently clustered under different national risk contexts, and to translate that identification into a spatial framework that can support subsequent observation-based analysis. Under that more disciplined interpretation, the value of ARI lies not in replacing existing indices, but in extending the analytical vocabulary available for understanding how repeated disaster shocks are distributed across space and where their accumulated pressure may be most consequential.”

2) Section 6 Conclusions:

“This study proposes the ARI as a grid-based and event-driven framework for examining recent accumulated disaster impacts at the global scale. By combining major disaster events recorded in EM-DAT during 2013-2023 with country-level WRI values, ARI adds explicit sensitivity to the spatial accumulation of recent high-impact shocks across subnational and transboundary space. The results reveal clear hotspot concentration, especially across the Himalayan belt, South Asia, and Southeast Asia, and show that ARI provides stronger spatial differentiation than a country-based reference framework. In this sense, ARI is not intended to replace long-term structural risk indices, but to complement them by highlighting where recent major disaster shocks have clustered most strongly. The analyses further indicate that ARI shows meaningful external plausibility and internal robustness as a comparative hotspot-identification framework, while remaining a relative indicator rather than a complete measure of absolute disaster risk. Building on the ARI-identified hotspot grids, this study also develops the 3H Dataset, which contains more than 7.5 million standardized sub-metre remote sensing image tiles across 14 high-risk grids in 10 developing countries. Together, ARI and the 3H Dataset provide a connected basis for linking global disaster statistics, contextual risk information, and Earth observation resources, thereby supporting subsequent research and

practice in multi-disaster hotspot identification and finer-scale geospatial analysis.”

We sincerely thank the referee again for this valuable comment. It has helped us revise the Conclusion from a more cautious and methodologically transparent perspective, and we believe these changes will make the final positioning of the study more balanced, defensible, and useful for readers.

We would like to sincerely thank the editor and all referees once again for the time, effort, and constructive feedback devoted to evaluating our manuscript. The comments raised in this round of review have helped us identify critical areas for improvement in the previous version, particularly regarding the methodological clarity of ARI, the transparency of data processing, the interpretation of the results, and the connection between the risk-assessment framework and the remote sensing dataset.

We have carefully considered each comment and have revised the manuscript accordingly. Through these revisions, we believe that the manuscript now presents the proposed ARI framework in a clearer and more reproducible manner, acknowledges its assumptions and limitations more explicitly, and provides a more balanced interpretation of its contribution. Furthermore, the revised text now more explicitly articulates the role of the 3H Dataset as an ARI-guided remote sensing resource for future finer-scale analysis, rather than as an independent or detached component.

We fully recognize that the proposed framework remains exploratory and that several methodological limitations still require further investigation in future work. Nevertheless, we hope that the substantial revisions made in this version have adequately addressed the concerns raised by the editor and referees, and that the manuscript is now clearer, more rigorous, and more suitable for consideration by **Natural Hazards and Earth System Sciences**.

Thank you again for your valuable guidance and for reconsidering our revised manuscript.

Kind regards,

Qingzhao Kong and Erqi Zhu

April 27, 2026