Comment 1: *The PRISMA guidelines are cited but not rigorously followed. Key elements of a systematic review—such as a quality assessment of included studies, a risk of bias analysis, etc —are missing.*

We thank the reviewer for this methodological observation, and we agree that certain elements of a fully formal systematic review, in particular a numerical quality scoring instrument and an explicit risk of bias analysis, were absent from the manuscript. In line with other reviews carried out in machine learning and climate risk assessment (e.g, Zennaro et al. (2021); Salcedo-Sanz et al. (2022)), quality and relevance were assessed qualitatively during full-text screening along three criteria: methodological rigour (evaluated through the presence and type of model validation, e.g., cross-validation, independent test sets, or benchmark comparisons), relevance to the research questions, and diversity across data sources, geographical coverage, hazard types, and ML approaches. The third criterion was applied explicitly to avoid over-representation of any single method or region in the final corpus. We have clarified limitations and scope of the reviews explicit in a new methodological section (2.3, Limitations and scope of the review).

See lines 296-303:

While this review follows the PRISMA guidelines for search strategy, screening, and reporting, a formal numerical quality scoring of individual studies was not applied, consistent with standard practice in PRISMA-based reviews of computational methods in geoscience and climate risk (e.g.,(Ghaffarian et al., 2023; Salcedo-Sanz et al., 2022; Zennaro et al., 2021)). Instead, quality and relevance were assessed qualitatively during full-text screening based on three criteria: methodological rigour (evaluated through the presence and type of model validation, e.g., cross-validation, independent test sets, or benchmark comparisons), relevance to the research questions, and diversity across data sources, geographical coverage, hazard types, and ML approaches. The latter criterion was applied explicitly to avoid over-representation of any single method or region in the final corpus, and is documented in Appendix B.

Comment 2: *the work uses key terms (like multi-hazard, multi-risk, compound, cascading, etc.) largely descriptively and not in accordance with any existing frameworks (e.g., IPCC, UNDRR). As this is a common feature of this diverse field, how to the authors assess potential inconsistencies in this besides sematic key-word matching?*

We thank the reviewer for this valuable comment: this is a well-documented challenge of multi-hazard and multi-risk research and has been recognised by previous literature (Gill et al., 2022; Gill & Malamud, 2014; Tilloy et al., 2019). Our primary mitigation was the explicit adoption of Zschau (2017) as a definitional anchor in Section 2.1, which provides which provides a detailed review of multi-hazard and multi-risk assessment and proposes a scalar hierarchy from multi-layer single hazard through multi-hazard and multi-hazard risk to multi-risk, mapped onto increasing complexity of hazard and vulnerability interactions. This framework is particularly useful as an operational classification scheme for reviewing and classifying modelling studies, because it moves beyond purely descriptive terminology and offers a structured basis for distinguishing levels of interaction and complexity. In this sense, it was more directly applicable to our review purposes than IPCC or UNDRR terminology, which is highly valuable but more strongly oriented toward governance and policy communication than methodological categorisation. Accordingly, we used the Zschau framework to reclassify studies during full-text screening, independently of the labels used by the original authors.

We acknowledge, however, that this reclassification process has inherent limits, and this is explicitly recognised in Section 2.3 of the manuscript. Papers captured by our keyword search span a wide range of actual practices, particularly for the terms "compound" and "cascading," which are used variously to denote statistically co-occurring events (Zscheischler et al., 2018), multi-driver events, physical triggering sequences, and socioeconomic impact chains. Because paper selection was necessarily based on author-assigned terminology, the final corpus reflects this heterogeneity. However, the subsequent classification was based on full-text qualitative assessment, guided by the Zschau (2017) framework, rather than semantic keyword matching alone. Nevertheless, the Zschau hierarchy does not resolve all variation of usage, and the resulting thematic categories should therefore be interpreted as analytical rather than strictly taxonomic.

Comment 3: *The work treats vulnerability and exposure largely as static in mapping exercises, with limited discussion of dynamic, social justice, governance, or human behavior.*

We thank the reviewer for this observation. We agree that vulnerability and exposure are treated largely as static inputs in the majority of the reviewed studies, and that dynamic vulnerability, social justice, governance, and human behaviour are underrepresented in the data-driven multi-hazard and multi-risk literature.

The scarcity of data-driven approaches that operationalise dynamic vulnerability, account for adaptive behaviour, or integrate governance and social justice dimensions is itself a substantive finding that we now report and discuss in Section 3.3.1. We identified this as one of the most significant structural gaps in the field: the reviewed literature overwhelmingly treats vulnerability and exposure as static spatial layers, and the methodological tools to represent their dynamic, socially differentiated, and governance-mediated dimensions within ML or statistical frameworks remain largely undeveloped.

We have strengthened the discussion in Section 3.3.1 to make this finding more explicit, connecting it to the broader literature on dynamic vulnerability (e.g., (Albulescu & Armaș, 2024a; Stolte et al., 2024)) to contextualise the gap. We also highlight it in the conclusion as one of the most pressing future research directions for the field, where closer integration of data-driven methods with agent-based modelling, participatory approaches, and socially-informed frameworks could produce meaningful advances (see also comment 9).

See lines 767-780 (results and discussion):

A critical limitation of the studies reviewed in this section is the static treatment of vulnerability. Most applications use fixed proxies – building footprints, land-use classifications, census-derived population density – that do not evolve in response to hazard occurrence, adaptation measures, or broader socio-economic change (de Ruiter & van Loon, 2022). This static framing can substantially underestimate risk in contexts where vulnerability is shaped by governance failures, structural inequalities, or rapid urban expansion (Šakić Trogrlić et al., 2024; Ward et al., 2022)*. A particularly underexplored challenge in multi-hazard risk assessment is that vulnerabilities do not simply add up across hazards: they interact. Synergies and asynergies between vulnerabilities mean that the combination of hazards can fundamentally alter how exposed elements are affected. For instance, adaptation measures designed to reduce risk from one hazard may increase vulnerability to another, and damage caused by a first hazard event can leave a system more vulnerable to a subsequent one* (Albulescu & Armaș, 2024; de Ruiter & van Loon, 2022)*. Stolte et al., (2024) further demonstrate through a global systematic review of urban vulnerability that the drivers of vulnerability differ substantially across hazard types,* *and explicitly call for research into multi-hazard vulnerability dynamics as a necessary step beyond the current dominant paradigm of treating multiple hazards in parallel rather than in interaction. Despite growing conceptual recognition of this problem, it remains essentially unaddressed in the data-driven literature reviewed in this study, where vulnerability interactions are neither modelled nor discussed.*

See lines 1069-1076:

A further methodological consideration for future research is the development of data-driven frameworks that move beyond static representations of vulnerability and exposure. The reviewed literature overwhelmingly treats these components as fixed spatial layers, with limited engagement with their dynamic, socially differentiated, and governance-mediated dimensions. Addressing this gap will require closer integration of ML and statistical methods with approaches capable of representing how vulnerability evolves over time, including agent-based modelling, participatory data collection, and socially-informed frameworks that explicitly account for adaptive behaviour, equity, and governance processes. Progress in this direction would not only improve the realism of multi-risk assessments but also strengthen their relevance for policy and decision-making in contexts where social vulnerability is itself a driver of risk (Bankoff & Hilhorst, 2022; Cannon, 2017).

Comment 4: *The review heavily biases towards ML while important statistical approaches (e.g., Bayesian networks, spatial statistics, agent-based modeling) are mentioned only briefly.*

We thank the reviewer for this observation. We acknowledge that the original manuscript gave insufficient attention to the broader landscape of quantitative methods for multi-hazard and multi-risk assessment. However, we want to clarify that this reflects a deliberate scoping decision: the review was designed specifically to assess the state of data-driven, and in particular ML-based, approaches, and was not intended to provide a comprehensive treatment of all quantitative frameworks in the field.

That said, we agree that situating the review within this broader methodological landscape is important, both to contextualise the scope and to acknowledge the complementary value of approaches such as Bayesian networks, agent-based models, and spatial statistics. We have therefore added a dedicated paragraph in Section 2 (lines 304–322) that explicitly acknowledges these complementary frameworks, summarises their distinctive strengths, and directs the reader to the relevant dedicated reviews.

See lines 304 – 322:

Moreover, while this review focuses on ML and copula-based methods as the primary data-driven approaches for multi-hazard and multi-risk assessment, it is important to acknowledge that several complementary quantitative frameworks exist and have been the subject of dedicated reviews that fall outside the scope of the present work. Bayesian networks (BNs) provide a probabilistic framework for multi-hazard causal modelling, capturing conditional dependencies between risk drivers through directed acyclic graphs and propagating uncertainty in a transparent, interpretable way; they are particularly valuable in data-sparse contexts where causal structure can be informed by expert knowledge, and their application to climate multi-risk assessment has been reviewed in depth by Sperotto et al. (2017). Agent-based models (ABMs) simulate the adaptive behaviour of individuals and institutions under hazard scenarios, making them suited to capturing dynamic vulnerability, evacuation dynamics, and community resilience processes that purely data-driven models cannot represent; comprehensive reviews of their application in disaster management are provided by Anshuka et al., (2022) and Zhuo & Han, (2020). More broadly, the full landscape of quantitative methods for modelling hazard interrelationships, including *stochastic, empirical, and mechanistic approaches, is systematically covered by Tilloy et al. (2019), providing a valuable complement to the ML-focused perspective of the present review.*

Comment 5: The authors should also provide statistics on the application domains (as presently literature appears limited to "Scopus, English-only, 2010–2024") and especially highlight if any potential regional bias in selected studies with a tendency to minimize the contribution of the Global South where risk dynamics differ.

We thank the reviewer for raising this important point. We fully agree that transparency about the geographical distribution of the reviewed literature – both in terms of authorship and application domains – is essential for interpreting the findings of the review and for acknowledging the potential influence of regional bias on the methods and priorities that dominate the corpus.

In the revised manuscript, we have added a dedicated analysis of the geographical distribution of the 152 reviewed papers, examining three dimensions simultaneously: lead author institution country, co-author countries, and case study regions. The results are reported in Section 2 (lines 323–336) and visualised through a Sankey diagram (Figure 4) and summary statistics (Figure 5), both provided in Appendix B. The analysis reveals a marked concentration of lead authorship in Europe (35.5%), East Asia (27.0%), and North America (21.1%), while Africa, South America, and Oceania collectively contribute less than 5%. Notably, we also document a systematic decoupling between authorship geography and case study geography: South/SE Asia, the Middle East, and Africa appear more frequently as study areas than as sources of authorship, suggesting that data-driven methods are in several cases developed in data-rich institutional contexts and subsequently applied to regions with different risk dynamics, data availability, and governance structures.

We acknowledge explicitly that this pattern is partly an artefact of the Scopus, English-only, 2010–2024 search scope, which may systematically underrepresent research published in other languages or in regional journals not indexed by Scopus. This limitation is discussed in the conclusions (lines 1077–1085), where we also call attention to the implications for the transferability of methods and the relevance of findings for Global South contexts. We hope this analysis addresses the reviewer's concern and provides the transparency needed to interpret the geographical scope of the review.

See lines 323-336:

Finally, some considerations need to be taken on the geographical distribution of the 152 papers included in this review, which reveal a marked concentration in a small number of regions. In terms of lead authorship, Europe (35.5%) and East Asia (27.0%) together account for nearly two thirds of the corpus, followed by North America (21.1%), while the Global South is substantially underrepresented: Africa, South America, and Oceania collectively contribute less than 5% of lead authors. A similar pattern holds for co-authorship, though with a slight broadening of participation: South/SE Asia rises to 6.4% and Middle East to 7.3%, suggesting that researchers from these regions participate more frequently as collaborators than as lead investigators. The most pronounced shift occurs in the case study column: Global studies account for 14.9% of the corpus, and South/SE Asia (10.6%), Middle East (6.8%), and Africa (5.0%) are more represented as study areas than as sources of authorship, indicating that data-driven methods developed in high-income regions are frequently applied to, rather than developed within, lower-income contexts. The full breakdown of lead author institution country, co-author countries, and case study regions, together with a Sankey diagram illustrating the flows between these three dimensions, is provided in Appendix B. These geographical imbalances should be borne in *mind when interpreting the findings of this review, as the methods, datasets, and risk priorities that dominate the literature inevitably reflect the institutional contexts in which the research was produced.*

See lines 1077-1085:

Finally, the geographical distribution of the reviewed studies, visualised in the Sankey diagram in Appendix B, points to an imbalance that is worth acknowledging explicitly. Europe, North America and East Asia together account for more than 80% of lead authorships, while Africa, Latin America, and Small Island Developing States contribute less than 5%. This pattern partly reflects the Scopus, English-only, 2010–2024 scope of the search strategy, which may systematically underrepresent research published in other languages or in regional journals not indexed by Scopus. At the same time, the decoupling between authorship geography and case study geography, with South/SE Asia, the Middle East, and Africa appearing more frequently as study areas than as sources of authorship, suggests that data-driven methods are in several cases developed in data-rich institutional contexts and subsequently applied to regions with different risk dynamics, data availability, and governance structures (Tiggeloven et al., 2025).

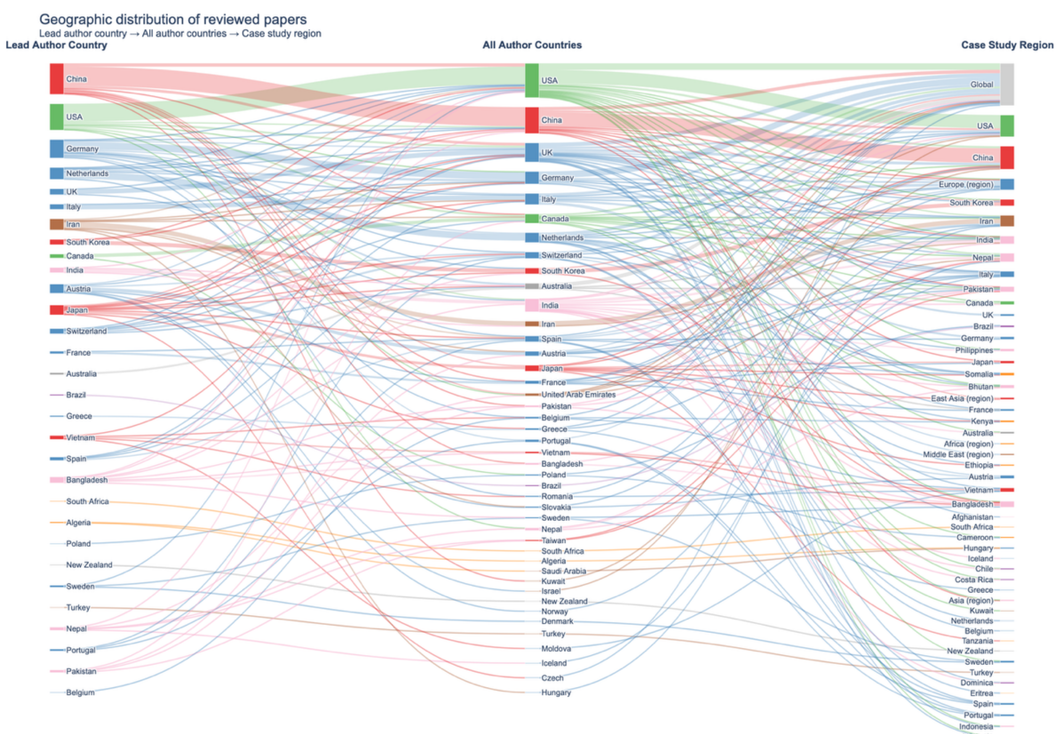

Figure 1: Geographic distribution of reviewed papers: Sankey diagrams between main authors countries, co-author countries and analysed case studies

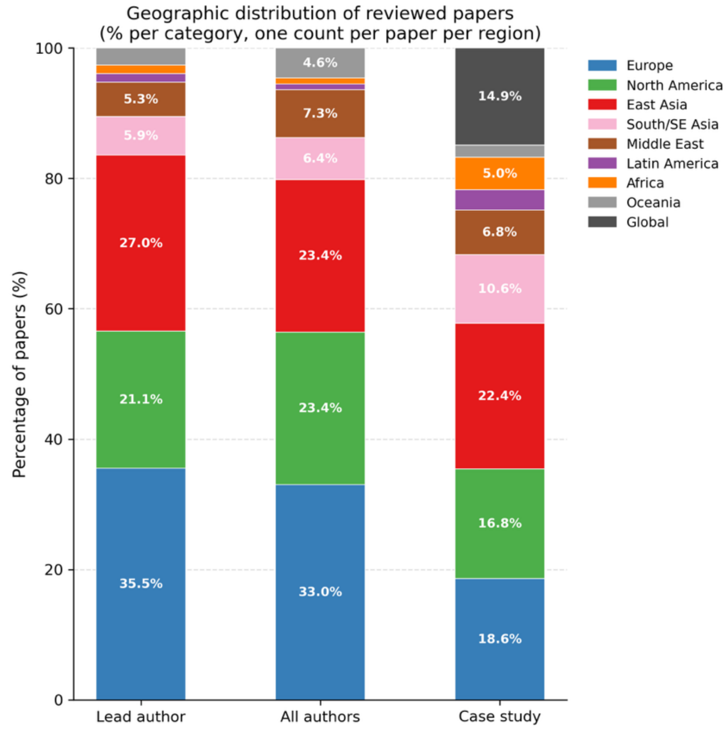

Figure 2: Summary statistics of geographic distribution of reviewed papers

Finally, the excel table (LR_geo_distribution_final.xlsx) used to create these figures is presented in the supplementary materials to the reviewers' response.

Comment 6: Although uncertainty quantification (UQ) is mentioned, the work does not systematically address how this is (or isn't) integrated across the modeling chain.

We thank the reviewer for this observation. We agree that the original manuscript did not systematically address how uncertainty quantification is (or is not) integrated across the modelling chain. We have addressed this in two places in the revised manuscript.

In the results section (lines 784–790), we now document explicitly that end-to-end UQ across the full hazard–exposure–vulnerability–risk chain is absent from the reviewed corpus, even in single-hazard contexts, and that the closest existing examples (Dawkins et al., 2023; Kropf et al., 2022) do not extend to multi-hazard interactions. In the conclusions (lines 1045–1052), we synthesise this as a structural gap in the field, distinguishing between aleatory and epistemic uncertainty and noting that current practice addresses UQ only in fragments, with copula-based approaches covering only statistically correlated hazard pairs and leaving cascading and triggered hazards unaddressed. We identify the development of a genuinely integrated UQ framework, propagating uncertainty continuously from inputs through multi-hazard interactions to the final risk estimate, as one of the most important open methodological challenges for the field.

See lines 784 – 790:

Another aspect to consider is uncertainty and its propagation across the risk modelling chain: attempts to propagate it formally across the hazard–exposure–vulnerability–risk chain are rare even in single-hazard contexts: (Kropf et al., 2022) introduced a sensitivity and uncertainty analysis framework within the CLIMADA platform that varies hazard, exposure, and vulnerability inputs simultaneously, and (Dawkins et al., 2023) extended this *to formally quantify uncertainty contributions from each component, with an application using GAM for heat-stress risk assessment, but neither study addresses multi-hazard interactions. However, no study in the reviewed corpus achieves end-to-end UQ in a multi-hazard risk context, propagating uncertainty from input data through hazard modelling and ML or statistical methods to the final risk estimate (Beven et al., 2018).*

Lines 1045-1052:

Another methodological gap identified by this review is the absence of end-to-end uncertainty quantification frameworks for multi-hazard risk assessment. Current practice addresses UQ in fragments: aleatory uncertainty in input data is handled at the start of the chain, epistemic uncertainty in ML models is occasionally addressed through Bayesian or ensemble methods at the hazard stage, and copula-based approaches characterise joint uncertainty for statistically correlated hazard pairs, but these efforts are rarely connected, and they do not extend to the full multi-hazard concept, which encompasses cascading and triggered hazards beyond the reach of shared statistical distributions. An integrated framework would propagate both aleatory and epistemic uncertainty continuously from input data through multi-hazard interactions, ML and statistical model outputs, and exposure and vulnerability components, to the final risk estimate (Beven et al., 2018).

Comment 7: Presently, Figure 3 is loosely integrated into the narrative of the work.

We thank the reviewer for this observation. We agree that Figure 3 was not sufficiently integrated into the narrative of the manuscript. Rather than restructuring the main text around it, we have moved the figure to Supplementary Materials, where it remains available as a reference overview of the methods used across the four research themes. The key methodological patterns it illustrates, the dominance of CNNs in EO processing, Random Forest and ensemble methods in risk and future risk assessment, and LSTM and deep learning architectures in hazard prediction, are now synthesised directly in the conclusions, ensuring that the analytical content is carried by the text rather than delegated to the figure.

Comment 8: Finally. The work places emphasis on machine learning without adequate discussion of its limitations, overhyped applications, or failures in real-world risk management. Which should have been part of the robustness of the work.

We thank the reviewer for this important observation. We fully agree that a rigorous review of data-driven methods for multi-hazard risk assessment must engage critically with the limitations of those methods, not only catalogue their methodological advances. In response, we have added a new Section 3.5 ("Limitations and research gaps") positioned at the end of the results, immediately before the conclusions.

The section addresses four structural limitations that we consider most consequential for the operational uptake of data-driven multi-hazard approaches: climate non-stationarity and the validity of training distributions under future conditions; the persistent gap between hazard prediction and actionable impact information; the barriers to operational adoption posed by the black-box nature of deep learning and the current immaturity of XAI integration; and the reproducibility and spatial validation concerns that affect the reliability of reported model performance across the reviewed literature. This discussion is also connected to the revised conclusions, which bring these limitations forward as the central challenges the field must address to move from proof-of-concept demonstrations to genuine operational uptake.

Lines 1020 - 1044

Beyond the methodological advances documented in the preceding sections, this review also identifies a set of persistent limitations and structural gaps in the field that are directly relevant to the operational uptake of data-driven approaches in multi-hazard risk management. A first and fundamental limitation is climate non-stationarity, as highlighted in Section 3.4: ML models trained on historical hazard records implicitly assume that the statistical relationships between predictors and outcomes will remain stable into the future. Relying on past norms and training distributions will prove inappropriate under non-stationary risk conditions, where projected increases in hazard frequency and severity, combined with shifting exposure and vulnerability, create conditions that fall outside the range of any historical training set. This is particularly acute for compound and cascading events, which are by definition rare in the historical record yet are precisely the configurations that climate change is projected to intensify.

A second limitation is the gap between hazard prediction and impact prediction. Most ML applications reviewed optimise for hazard or susceptibility metrics, but impact prediction requires integrating physical hazard outputs with dynamic exposure and vulnerability data at sub-kilometre scales, a challenge that the reviewed papers largely sidestep by using static proxies. This gap between technical model performance and actionable risk information represents one of the most important unresolved challenges in translating ML-based risk assessment into operational decision-making (Reichstein et al., 2025; Tiggeloven et al., 2025).

A third set of concerns relates to interpretability and trust. The black-box nature of deep learning models creates well-recognised barriers to adoption in high-stakes regulatory and emergency management contexts, where stakeholders need not only a prediction but a justification they can interrogate and contest. However, current XAI applications remain predominantly proof-of-concept and are rarely integrated into operational early warning or risk assessment workflows(Ghaffarian et al., 2023). Moreover, reproducibility and validation remain a persistent concern. In geoscientific applications, spatial autocorrelation means that random train-test splits routinely inflate apparent model skill relative to genuinely independent spatial holdouts (Sweet et al., 2023), and the reviewed literature shows limited adoption of spatially blocked cross-validation or independent regional test sets. These limitations do not invalidate the contributions reviewed here, but they do underscore the need for more rigorous validation protocols, realistic appraisal of out-of-sample performance, and explicit discussion of the conditions under which ML approaches can be expected to generalise beyond their training contexts.

Lines 1100-1111

The gap between hazard prediction and impact prediction remains largely unresolved and bridging it will demand closer integration of data driven model outputs with dynamic representations of exposure and vulnerability, including human behaviour, adaptive responses, and the social and governance dimensions that determine how risk is distributed across communities. Explainability is a further priority: XAI methods need to move beyond their current role as exploratory tools and be embedded into operational early warning and risk assessment workflows, where their ability to illuminate driver interactions and build stakeholder trust is most consequential. End-to-end uncertainty quantification across the full modelling chain remains absent and developing integrated frameworks that propagate both aleatory and epistemic uncertainty from inputs through multi-hazard interactions to *the final risk estimate is one of the most important open methodological challenges for the field. Underlying all of these challenges is the problem of non-stationarity: as climate change intensifies hazard frequency and severity, shifts exposure and vulnerability, and increases the likelihood of compound and cascading configurations that fall outside any historical training set, the assumption that past conditions are a reliable guide to future risk becomes increasingly untenable, with direct consequences for the validity of ML-based projections of multi-risk evolution.*

Comment 9: The review should discuss any potential ethical, political, or justice dimensions of using AI/ML in risk assessment—such as data colonialism, algorithmic bias, or the risk of reinforcing inequalities.

We fully agree that ethical dimensions are increasingly important in the responsible application of AI/ML to risk assessment. These issues were insufficiently addressed in the original manuscript. In response, we have added discussion in two locations. In Section 2.3, the geographic analysis of the 152 reviewed papers documents a systematic decoupling between authorship geography and case study geography (see comment 5) — a pattern that raises concerns about data colonialism, the representativeness of training datasets, and whether methodological priorities reflect the needs of the most exposed populations. In Section 3.5, the discussion of dynamic vulnerability explicitly engages with social justice, noting that the absence of socially differentiated vulnerability representations in data-driven frameworks is not merely a technical gap but reflects broader inequities in data availability and research prioritisation (Bankoff & Hilhorst, 2022; Cannon, 2017). Both threads are brought together in the conclusions, which call for the adoption of ethical review standards for AI-based disaster risk research, building on frameworks developed for responsible AI in environmental science (McGovern et al., 2022) and on systematic assessments of the guardrails needed for trustworthy AI deployment in hazard risk contexts (Ghaffarian et al., 2023), and for collaborative, locally-grounded research initiatives in currently underrepresented regions as a necessary condition for developing multi-risk frameworks that are both scientifically robust and equitable (Naudé & Vinuesa, 2021; Tiggeloven et al., 2025).

Lines 764-767

Social justice dimensions also remain largely absent from the reviewed multi-risk literature: only few of the papers analysed explicitly consider vulnerability dimensions such as gender and age, while the question of how ML-based risk maps might inherit biases from historically underinvested impact datasets remains largely unaddressed (McGovern et al., 2022).

Lines 1053- 1076

A further methodological consideration for future research is the development of data-driven frameworks that move beyond static representations of vulnerability and exposure. The reviewed literature overwhelmingly treats these components as fixed spatial layers, with limited engagement with their dynamic, socially differentiated, and governance-mediated dimensions. Addressing this gap will require closer integration of ML and statistical methods with approaches capable of representing how vulnerability evolves over time, including agent-based modelling, participatory data collection, and socially-informed frameworks that explicitly account for adaptive behaviour, equity, and governance processes. Progress in this direction would not only improve the realism of multi-risk assessments but also strengthen their relevance for policy and decision-making in contexts where social vulnerability is itself a driver of risk (Cannon 2017; Bankoff and Hilhorst 2022).

Finally, the geographical distribution of the reviewed studies, visualised in the Sankey diagram in Appendix B, points to an imbalance that is worth acknowledging explicitly. Europe, North America and East Asia together account for more than 80% of lead authorships, while Africa and South America contribute less than 5%. This pattern partly reflects the Scopus, English-only, 2010–2024 scope of the search strategy, which may systematically underrepresent research published in other languages or in regional journals not indexed by Scopus. At the same time, the decoupling between authorship geography and case study geography, with South/SE Asia, the Middle East, and Africa appearing more frequently as study areas than as sources of authorship, suggests that data-driven methods are in several cases developed in data-rich institutional contexts and subsequently applied to regions with different risk dynamics, data availability, and governance structures (Tiggeloven et al., 2025). While drawing strong conclusions about data colonialism or algorithmic bias from a bibliometric analysis alone would go beyond the scope of this review, these patterns do raise questions that the community should engage with: whether training datasets and validation benchmarks are representative of the contexts in which models are ultimately deployed, and whether the priorities shaping methodological innovation reflect the needs of the most exposed populations. Future work in this area should pay closer attention to the transferability of data-driven multi-risk frameworks across different socio-economic and data environments (Tiggeloven et al., 2025), and collaborative initiatives fostering locally-grounded research in currently underrepresented regions would strengthen both the scientific robustness and the equity dimensions of the field (Naudé & Viluesa, 2021).

Lines 1056-1060

Addressing these gaps, alongside the geographic and equity imbalances documented in this review, will require not only methodological innovation but also more inclusive research practices: collaborative frameworks that bring together physical scientists, social scientists, and communities in currently underrepresented regions, co-producing knowledge that is robust, transferable, and genuinely relevant to those most exposed to the evolving risks of a changing climate.

References

Albulescu, A.-C., & Armaș, I. (2024). An impact-chain-based exploration of multi-hazard vulnerability dynamics: The multi-hazard of floods and the COVID-19 pandemic in Romania. *Natural Hazards and Earth System Sciences*, *24*(8), 2895–2922. https://doi.org/10.5194/nhess-24-2895-2024

Anshuka, A., Van Ogtrop, F. F., Sanderson, D., & Leao, S. Z. (2022). A systematic review of agent-based model for flood risk management and assessment using the ODD protocol. *Natural Hazards*, *112*(3), 2739–2771. https://doi.org/10.1007/s11069-022-05286-y

Bankoff, G., & Hilhorst, D. (2022). *Why Vulnerability Still Matters: The Politics of Disaster Risk Creation* (1st edn). Routledge. https://doi.org/10.4324/9781003219453

Beven, K. J., Aspinall, W. P., Bates, P. D., Borgomeo, E., Goda, K., Hall, J. W., Page, T., Phillips, J. C., Simpson, M., Smith, P. J., Wagener, T., & Watson, M. (2018). Epistemic uncertainties and natural hazard risk assessment – Part 2: What should constitute good practice? *Natural Hazards and Earth System Sciences*, *18*(10), 2769–2783. https://doi.org/10.5194/nhess-18-2769-2018

Cannon, T. (2017). Social Vulnerability and Environmental Hazards. In D. Richardson, N. Castree, M. F. Goodchild, A. Kobayashi, W. Liu, & R. A. Marston (Eds), *International Encyclopedia of Geography* (1st edn, pp. 1–8). Wiley. https://doi.org/10.1002/9781118786352.wbieg0845

Dawkins, L. C., Bernie, D. J., Pianosi, F., Lowe, J. A., & Economou, T. (2023). Quantifying uncertainty and sensitivity in climate risk assessments: Varying hazard, exposure and vulnerability modelling choices. *Climate Risk Management*, *40*, 100511. https://doi.org/10.1016/j.crm.2023.100511

de Ruiter, M. C., & van Loon, A. F. (2022). The challenges of dynamic vulnerability and how to assess it. *iScience*, *25*(8), 104720. https://doi.org/10.1016/j.isci.2022.104720

Ghaffarian, S., Taghikhah, F. R., & Maier, H. R. (2023). Explainable artificial intelligence in disaster risk management: Achievements and prospective futures. *International Journal of Disaster Risk Reduction*, *98*, 104123. https://doi.org/10.1016/j.ijdrr.2023.104123

Gill, J. C., Duncan, M., Ciurean, R., Smale, L., Stuparu, D., Schlumberger, J., de Ruiter, M., Tiggeloven, T., Torresan, S., Gottardo, S., Mysiak, J., Harris, R., Petrescu, E.-C., Girard, T., Khazai, B., Claassen, J., Dai, R., Champion, A., Daloz, A. S., … Ward, P. (2022). D1.2 Handbook of multi-hazard, multi-risk definitions and concepts. *Myriad*, (101003276).

Gill, J. C., & Malamud, B. D. (2014). Reviewing and visualizing the interactions of natural hazards. *Reviews of Geophysics*, *52*(4), 680–722. https://doi.org/10.1002/2013RG000445

Kropf, C. M., Ciullo, A., Otth, L., Meiler, S., Rana, A., Schmid, E., McCaughey, J. W., & Bresch, D. N. (2022). Uncertainty and sensitivity analysis for probabilistic weather and climate-risk modelling: An implementation in CLIMADA v.3.1.0. *Geoscientific Model Development*, *15*(18), 7177–7201. https://doi.org/10.5194/gmd-15-7177-2022

McGovern, A., Ebert-Uphoff, I., Gagne, D. J., & Bostrom, A. (2022). Why we need to focus on developing ethical, responsible, and trustworthy artificial intelligence approaches for environmental science. *Environmental Data Science*, *1*, e6. https://doi.org/10.1017/eds.2022.5

Naudé, W., & Vinuesa, R. (2021). Data deprivations, data gaps and digital divides: Lessons from the COVID-19 pandemic. *Big Data & Society*, *8*(2), 20539517211025545. https://doi.org/10.1177/20539517211025545

Reichstein, M., Benson, V., Blunk, J., Camps-Valls, G., Creutzig, F., Fearnley, C. J., Han, B., Kornhuber, K., Rahaman, N., Schölkopf, B., Tárraga, J. M., Vinuesa, R., Dall, K., Denzler, J., Frank, D., Martini, G., Nganga, N., Maddix, D. C., & Weldemariam, K. (2025). Early warning of complex climate risk with integrated artificial intelligence. *Nature Communications*, *16*(1), 2564. https://doi.org/10.1038/s41467-025-57640-w

Šakić Trogrlić, R., Reiter, K., Ciurean, R. L., Gottardo, S., Torresan, S., Daloz, A. S., Ma, L., Padrón Fumero, N., Tatman, S., Hochrainer-Stigler, S., de Ruiter, M. C., Schlumberger, J., Harris, R., Garcia-Gonzalez, S., García-Vaquero, M., Arévalo, T. L. F., Hernandez-Martin, R., Mendoza-Jimenez, J., Ferrario, D. M., … Ward, P. J. (2024). Challenges in assessing and managing multi-hazard risks: A European stakeholders perspective. *Environmental Science & Policy*, *157*, 103774. https://doi.org/10.1016/j.envsci.2024.103774

Salcedo-Sanz, S., Pérez-Aracil, J., Ascenso, G., Del Ser, J., Casillas-Pérez, D., Kadow, C., Fister, D., Barriopedro, D., García-Herrera, R., Restelli, M., Giuliani, M., & Castelletti, A. (2022). *Analysis, Characterization, Prediction and Attribution of Extreme Atmospheric Events with Machine Learning: A Review*.

Sperotto, A., Molina, J.-L., Torresan, S., Critto, A., & Marcomini, A. (2017). Reviewing Bayesian Networks potentials for climate change impacts assessment and management: A multi-risk perspective. *Journal of Environmental Management*, *202*, 320–331. https://doi.org/10.1016/j.jenvman.2017.07.044

Stolte, T. R., Koks, E. E., De Moel, H., Reimann, L., Van Vliet, J., De Ruiter, M. C., & Ward, P. J. (2024). VulneraCity–drivers and dynamics of urban vulnerability based on a global systematic literature review. *International Journal of Disaster Risk Reduction*, *108*, 104535. https://doi.org/10.1016/j.ijdrr.2024.104535

Sweet, L., Müller, C., Anand, M., & Zscheischler, J. (2023). Cross-Validation Strategy Impacts the Performance and Interpretation of Machine Learning Models. *Artificial Intelligence for the Earth Systems*, *2*(4). https://doi.org/10.1175/AIES-D-23-0026.1

Tiggeloven, T., Pfeiffer, S., Matanó, A., Van Den Homberg, M., Thalheimer, L., Reichstein, M., & Torresan, S. (2025). The role of artificial intelligence for early warning systems: Status, applicability, guardrails, and ways forward. *iScience*, *28*(11). https://www.cell.com/iscience/fulltext/S2589-0042(25)01950-9

Tilloy, A., Malamud, B. D., Winter, H., & Joly-Laugel, A. (2019). A review of quantification methodologies for multi-hazard interrelationships. *Earth-Science Reviews*, *196*, 102881. https://doi.org/10.1016/j.earscirev.2019.102881

Ward, P. J., Daniell, J., Duncan, M., Dunne, A., Hananel, C., Hochrainer-Stigler, S., Tijssen, A., Torresan, S., Ciurean, R., Gill, J. C., Sillmann, J., Couasnon, A., Koks, E., Padrón-Fumero, N., Tatman, S., Tronstad Lund, M., Adesiyun, A., Aerts, J. C. J. H., Alabaster, A., … de Ruiter, M. C. (2022). Invited perspectives: A research agenda towards disaster risk management pathways in multi-(hazard-)risk assessment.

Natural Hazards and Earth System Sciences, *22*(4), 1487–1497. https://doi.org/10.5194/nhess-22-1487-2022

Zennaro, F., Furlan, E., Simeoni, C., Torresan, S., Aslan, S., Critto, A., & Marcomini, A. (2021). Exploring machine learning potential for climate change risk assessment. *Earth-Science Reviews*, *220*, 103752. https://doi.org/10.1016/j.earscirev.2021.103752

Zhuo, L., & Han, D. (2020). Agent-based modelling and flood risk management: A compendious literature review. *Journal of Hydrology*, *591*, 125600. https://doi.org/10.1016/j.jhydrol.2020.125600

Zschau, J. (2017). Where are we with multihazards , multirisks assessment capacities ? *Science for Disaster Risk Management 2017: Knowing Better and Losing Less*, 98–115.

Zscheischler, J., Westra, S., van den Hurk, B. J. J. M., Seneviratne, S. I., Ward, P. J., Pitman, A., AghaKouchak, A., Bresch, D. N., Leonard, M., Wahl, T., & Zhang, X. (2018). Future climate risk from compound events. *Nature Climate Change*, *8*(6), 469–477. https://doi.org/10.1038/s41558-018-0156-3