Response Letter Accompanying the Revised Manuscript

Manuscript ID: egusphere-2026-937
Manuscript type: Review article

Dear Editors of Natural Hazards and Earth System Sciences,

We sincerely thank you for inviting us to submit a revised version of our manuscript. We are grateful to you and the two referees for the careful assessment of our work and for the constructive comments and suggestions, which have helped us improve the quality and clarity of the manuscript.

We have carefully revised the manuscript and addressed all comments raised by the Editor and the two referees. Our point-by-point responses are presented in blue below. Detailed changes can be found in the revised manuscript, and a marked-up version showing all modifications relative to the previous version has also been provided.

We have also addressed the Journal Editor's comments concerning the formatting of the citations and reference list and the availability of information on the publications included in the analysis.

On behalf of all authors, we hope that the revised manuscript has adequately and explicitly addressed the comments and suggestions raised during the review process. We sincerely thank the Editor and the two referees again for their time and valuable contributions, and we welcome any further comments.

All author responses are shown in blue.

Response to the Journal Editor

Comment

> There are some inconsistencies in the formatting of the references. In particular, the citation style does not appear to follow the NHESS guidelines, and the reference list is also not formatted accordingly. The authors should also consider how best to inform potential readers about the articles included in their analysis. Clearly, it is not feasible to include a reference list of more than 50,000 items, nor would a corresponding table be suitable for the manuscript itself. An online supplement or external repository providing this information may therefore be an appropriate solution.

Response to the Journal Editor

We thank the Journal Editor for these important comments. We have checked and reformatted the in-text citations and reference list to follow the NHESS author–year style consistently. Duplicate or inconsistent entries have been consolidated, author names and publication details have been standardised, and DOI/URL information has been presented consistently where available.

We have also addressed the need to make the evidence base transparent without placing thousands of entries in the manuscript. The complete bibliographic database of 1,967 peer-reviewed articles and the event-level dataset of 95 documented mountain flash-flood cases, including the 42 events with quantitative rainfall amount and duration, have been prepared as machine-readable supplementary data. The regional evidence workbook provides the paper-level coding, matched rules, geographic audit, system profiles and country–mountain-system crosswalk used in Sect. 2.2 and Table A1. Supplementary Sects. S1–S3 document the search, screening, modelling and coding procedures, and the Data availability statement has been revised accordingly.

Changes made in the revised submission: reference list and in-text citations throughout; Sect. 2.1, lines 95–143; Sect. 2.2, lines 145–190; Data availability, lines 952–955; Supplementary Sects. S1–S3.

Response to Referee #1

General comments

The presented article has a very broad scope and considerable synthetic value. It attempts to combine physical mechanisms, environmental conditions, social and economic impacts, risk management, as well as technical measures and nature-based solutions. This is fully justified given the interdisciplinary nature of the problem. The extensive literature base of 1,967 publications provides potential for identifying global trends. The paper describes well the problem of non-stationarity, highlighting the limited transferability of rainfall thresholds, emphasizing the initial state of the catchment, and pointing out the non-stationarity of climate. The causes of flash floods are correctly presented as a combination of rainfall, soil moisture, topography, hillslope–channel connectivity, sediment, and anthropogenic changes. The article also successfully links science and practice by emphasizing the importance of warning systems, spatial planning, catchment management, and hybrid retention solutions.

Response to the general comments

We sincerely thank the reviewer for the careful and constructive assessment of the manuscript and for recognising its interdisciplinary scope, synthetic value, treatment of non-stationarity, and connection between process understanding and risk-reduction practice. We have completed a substantial revision that directly addresses the concerns identified in the review.

The revised manuscript now: (i) defines the hazard–exposure–vulnerability–capacity–impact sequence and uses these terms consistently; (ii) clarifies the construction and limits of the bibliometric corpus and event subsets; (iii) introduces a GMBA mountain-system comparison of regional evidence and connects that evidence structure to the process synthesis; (iv) expands uncertainty propagation, model validation and machine-learning limitations; (v) adds life-cycle costs, implementation constraints and trade-offs; and (vi) narrows and clarifies the scope of the hazard-chain discussion. We have also replaced overly broad claims of global comparability with a more cautious description of a structured synthesis of available cross-regional evidence.

These revisions are distributed across Sects. 1–6, Appendix A, and Supplementary Sects. S1–S3. The detailed changes relevant to each point are described below.

Comment 1

However, the excessively broad scope of the analyzed problem comes at the expense of conceptual precision and clarity. The authors use the terms hazard, disaster, risk, impact, vulnerability, and resilience in a manner that is not fully differentiated. In a study of this type, clearer conceptual structuring would be expected.

Response to Comment 1

We agree that the previous version did not distinguish these concepts consistently. We have now added an explicit conceptual framework in the Introduction. Flood hazard is defined as the physical event and its frequency, intensity, duration and spatial extent; disaster risk is defined as the interaction of hazard with exposure, vulnerability and capacity; and realised impacts are the observed economic, social and ecological consequences after an event occurs. Capacity is treated as the ability to anticipate, respond and recover.

The manuscript structure has also been aligned with this terminology. Section 3 examines physical hazard formation; Sect. 4 first separates exposure, vulnerability and response capacity and then analyses realised impacts and hazard chains; Sect. 5 explains how different interventions modify hazard, exposure, vulnerability, capacity or residual losses; and Sect. 6 organises governance around prevention of new risk, reduction of existing risk and management of residual risk. This creates a consistent analytical sequence from physical process to disaster outcome and risk governance.

Changes made in the revised manuscript: Sect. 1, lines 40–93; Sect. 4, lines 435–580; Sect. 5 opening synthesis, lines 584–595; and Sect. 6, lines 797–845.

Comment 2

> The criteria for selecting bibliometric cases are also unclear, and there is no discussion of their representativeness. The study covers the years 2000–2025, yet some conclusions regarding publication trends, risk growth, and climate change are formulated as if all periods had equal data quality. In reality, the increase in the number of publications may also reflect the development of science rather than solely the growing importance of the problem, and the rise in the number of documented events may partly result from improved reporting.

Response to Comment 2

We agree and have revised both the Methods and the interpretation of the results. Section 2.1 now distinguishes three evidence layers: the full bibliometric corpus of 1,967 peer-reviewed articles; 95 documented events for which location, occurrence time, causal processes and impacts could be identified; and the 42-event subset for which quantitative rainfall amount and duration were reported. Supplementary Sect. S1 provides the complete search query, database fields, inclusion and exclusion criteria, deduplication, staged screening and event-matching rules. Missing rainfall values were not inferred from unrelated datasets.

We have also revised the claims based on these data. Publication volume and study-location counts are now interpreted as indicators of research visibility, evidence availability and knowledge accumulation, not as direct measures of global flash-flood frequency, risk growth or climate-change effects. The text explicitly identifies the effects of language, database coverage, monitoring density, institutional research capacity and reporting practice. Climate-change statements are supported by the cited observational or modelling literature and are no longer inferred from bibliometric growth.

The 42-event rainfall subset is described as geographically uneven and dominated by Asian cases. Its distributions are used only to demonstrate reported variability and state dependence, not to derive a representative global climatology or transferable threshold.

Changes made in the revised manuscript: Sect. 2.1, lines 95–143; Sect. 3.1 and Fig. 3, lines 256–273; Supplementary Sect. S1, lines 15–61.

Comment 3

> There is also some lack of clarity in the presentation of the analyzed regions. Examples are drawn from China, India, and Europe, but there is no systematic comparison between areas with similar climate, geomorphology, or levels of urbanization.

Response to Comment 3

We agree that regional comparison is necessary, but the available literature does not support a uniform event-by-event quantitative comparison across all combinations of climate, geomorphology and urbanisation. Many regions have very few studies with comparable rainfall, runoff, sediment and loss data, and the evidence types are uneven. We therefore adopted a two-level comparison that is systematic without implying false equivalence or precision.

First, Sect. 2.2 now uses the GMBA Mountain Inventory v2.0 major mountain systems as a consistent physical spatial framework. We assigned 506 publications to major mountain systems and compared seven non-exclusive evidence categories using counts, within-system percentages and location quotients. Table A1 shows, for example, the dominance of impact and risk-reduction evidence in the Himalaya, observation-based evidence in the European Alps and Appalachians, historical reconstruction in the Carpathians, and remote-sensing/model-based evidence in the East Sahara Mountains. Supplementary Sect. S3 gives the coding rules, denominators, LQ equation, thresholds and audit trail.

Second, Sect. 3 provides the process-level comparison requested by the reviewer. Rather than forcing sparse data into a regional matrix, it compares how storm intensity and duration, antecedent moisture and cryospheric state, soil and vegetation, geology and disturbance, urbanisation, tributary synchronisation, and hydrological and sediment connectivity shift flood response across humid monsoon, arid wadi, Mediterranean, high-mountain, tropical and urbanising settings. This process-based synthesis is the appropriate level at which the available heterogeneous evidence can be compared. Section 6.3 then explains why intervention transfer must be conditioned on monitoring capacity, sediment regime, available land, maintenance resources and institutional context.

Changes made in the revised manuscript: Sect. 2.2, lines 145–190; Sect. 3, lines 242–434; Sect. 6.3, lines 870–910; Table A1; Supplementary Sect. S3, lines 253–282.

Comment 4

> The article also lacks discussion on uncertainty propagation, input data errors, model validation in poorly monitored regions, and the issue of overfitting in machine learning models.

Response to Comment 4

We have substantially expanded this discussion. Section 5.1 now traces uncertainty through the complete observation–model–warning chain. It addresses sparse and low-frequency gauge networks, radar beam blockage and ground clutter, satellite underestimation of short-duration rainfall peaks, gauge representativeness, incomplete event inventories, uncertain post-event records, DEM and parameter assumptions, and the propagation of these uncertainties into rainfall thresholds, flood simulation, inundation mapping and loss assessment.

The machine-learning discussion has also been made more critical. The revised text distinguishes predictive performance from transferability and explains the risks of overfitting when training data are small, imbalanced or spatially clustered. It notes that random train–test splits can leak shared event or spatial characteristics, that black-box models can be difficult to interpret, and that models may extrapolate poorly under changing climate, land use and catchment conditions. We therefore recommend spatial and temporal cross-validation, testing on independent events and catchments, sensitivity analysis, uncertainty ranges, and checks against hydrological process understanding. Physics-guided models and real-time observations are presented as useful supports, not as substitutes for validation.

Section 6.1 extends this reasoning to warning thresholds and structural designs, treating their performance as conditional and explicitly considering exceedance, redundancy, safe failure and post-event learning.

Changes made in the revised manuscript: Sect. 5.1, lines 598–684, especially lines 607–662; Sect. 6.1, lines 824–845.

Comment 5

> Furthermore, there is no analysis of costs, cost-effectiveness, trade-offs between objectives, or institutional and political constraints.

Response to Comment 5

We agree and have added a dedicated treatment of costs and trade-offs. The structural-measures section now discusses sediment filling, structural deterioration, maintenance, service life, design exceedance and the possibility that protective infrastructure can amplify consequences when it fails. It includes quantitative evidence from the Wartschenbach torrent, where maintaining the estimated protection level under projected increases in debris-flow activity would require a 57–63% increase in maintenance budgets and approximately 50% higher intervention costs.

The Nature-Based Solutions section now addresses establishment time, implementation scale, storage availability, performance under high-magnitude events and the limited evidence on decadal maintenance costs. It contrasts the more immediate, design-specific protection of structural measures with the slower and spatially variable regulation provided by NbS, while noting that both require maintenance and can create ecological or sediment-connectivity trade-offs.

Section 6.3 then synthesises the wider institutional and political constraints: recurrent expenditure, land scarcity, development pressure, enforcement capacity, relocation and compensation, insurance affordability and equity, long-term funding, local acceptance, and transboundary coordination. Because the available studies do not support a globally standardised cost-effectiveness ranking, we explicitly present this as a qualitative evidence synthesis and identify the lack of comparable life-cycle cost data as a remaining research gap.

Changes made in the revised manuscript: Sects. 5.2.1–5.2.2, lines 705–796; Sect. 6.3, lines 870–910; Conclusion, lines 938–943.

Comment 6

> The article strongly promotes integrated, hybrid, and multi-scale approaches; however, greater integration does not always necessarily lead to better outcomes. The intention to include events of different origins that may trigger flash floods—such as classical flash floods, flood cascades, dam-break floods, and glacial lake outburst floods—is understandable. Nevertheless, this approach tends to blur the main object of the study.

Response to Comment 6

We agree that integration is not inherently beneficial and that the scope must remain centred on mountain flash floods. We have therefore removed language implying that more integration automatically produces better outcomes. The revised manuscript treats coordinated portfolios as conditional: the appropriate combination depends on local hazard processes, exposure, sediment connectivity, implementation scale, land availability, monitoring capacity, funding, maintenance, institutions and community acceptance. Section 6 further distinguishes prevention of new risk, reduction of existing risk and management of residual risk, so interventions are linked to specific risk functions rather than combined indiscriminately.

We have also clarified the review object in the Introduction. Rainfall- or runoff-driven mountain flash floods remain the principal focus. “Mountain torrent” is retained only where it reflects source terminology; hyperconcentrated flow and debris flood are treated as sediment-rich transitional regimes; and debris flow is treated as a related but distinct gravity-driven mass-flow process. GLOFs, landslide-dam breaches, dam failures, wildfire and earthquake legacies, and infrastructure failures are included only where they trigger, transform or amplify mountain flash-flood hazard chains.

The revised hazard-chain section is structured around cross-event disturbance legacies, within-event transformations and downstream amplification. This preserves a clear core object while explaining the multi-hazard pathways that materially change mountain flash-flood magnitude or consequences.

Changes made in the revised manuscript: Sect. 1, lines 40–45 and 78–93; Sect. 4.2.2, lines 508–580; Sects. 5–6, lines 584–910.

Response to Referee #2

General comments

This manuscript offers a broad synthesis of flash-flood research in mountainous regions, drawing on 1,967 peer-reviewed studies published between 2000 and 2025. The combination of bibliometric mapping, dynamic topic modelling, process-based synthesis, and governance perspectives is genuinely interdisciplinary, and the overall structure — from global trends to hydrological mechanisms, cascading impacts, and adaptive governance — is logical and well organised.

However, the manuscript does not fully deliver on several stated ambitions. The most significant issue is the incomplete integration between the bibliometric component and the qualitative synthesis. The Dynamic Topic Modelling results function largely as a standalone section rather than shaping the subsequent process discussion. Rainfall thresholds, early warning systems, and cascade mechanisms are often framed in general terms, despite being derived from a spatially unrepresentative corpus.

A related tension concerns the claim to establish a "globally comparable baseline." Semi-arid mountain environments, African highland systems, and Andean catchments remain underrepresented, meaning several key findings rest on a corpus more regionally specific than the framing implies. The two conceptual contributions highlighted in the abstract — the "state-dependent triggering function" and the "threshold surface" framework — are promising but remain underformalised and disconnected from the operational sections that follow.

In my view, the manuscript could become suitable for publication in NHESS, but it still requires major revision. The comments below aim to help the authors strengthen its conceptual depth, geographic balance, and critical rigour.

Response to the general comments

We sincerely thank the reviewer for the detailed and constructive assessment. We have completed a substantial revision addressing the integration of the bibliometric and qualitative components, methodological transparency, geographic representativeness, sediment connectivity, cascading hazards, machine-learning limitations and Nature-Based Solutions.

The claim to establish a “globally comparable baseline” has been removed. The revised contribution is described as a “structured synthesis of available cross-regional evidence” that explicitly identifies geographic and thematic imbalances. The six thematic domains identified by the topic analysis are now used as an organising bridge to the later synthesis of meteorological forcing, arid wadi flooding, sediment processes, special geological and topographic settings, risk management and eco-climatic impacts. The state-dependent triggering and threshold-surface concepts have also been developed as a process framework linking storm forcing, antecedent catchment state and connectivity to dynamic warning practice; they are not presented as a universally calibrated predictive model.

The detailed responses below identify the specific revisions in the manuscript and Supplement.

Comment 1

Machine learning discussion is insufficiently critical

In Section 4.1, machine learning approaches are presented mainly through performance improvements. Important limitations receive little attention: model transferability across hydro-climatically contrasted basins, uncertainty propagation in data-sparse environments, interpretability constraints, and dependence on large training datasets. These issues are especially relevant for the poorly gauged mountain regions discussed elsewhere. The claim that deep-learning models can "skilfully predict extreme events in ungauged watersheds" (ll. 480–482) deserves more cautious framing. A more balanced treatment of capabilities and limitations would considerably strengthen this section.

Response to Comment 1

We agree and have rewritten the machine-learning discussion to provide a balanced assessment. Machine learning is now presented as a conditional tool for regionalisation, susceptibility mapping and forecasting, not as a stand-alone solution for ungauged basins. Although global deep-learning studies demonstrate potential, the revised text explicitly states that high predictive scores do not by themselves establish transferability to hydro-climatically different mountain catchments.

The new discussion covers overfitting in small, imbalanced and spatially clustered datasets; optimistic performance caused by random train–test splits; dependence on representative training data and complete event inventories; interpretability limitations; and poor extrapolation under non-stationary climate, land use and catchment conditions. It recommends spatial and temporal cross-validation, independent events and catchments, uncertainty reporting and consistency checks against hydrological processes. The section also traces uncertainty from radar, satellite and gauge inputs through thresholds, model simulation, inundation maps and loss estimates.

The earlier statement that deep-learning models can "skilfully predict extreme events in ungauged watersheds" is now contextualised as evidence of potential in large, well-curated datasets, immediately followed by explicit transferability and validation limitations. Physics-guided learning and real-time observations are presented as ways to improve robustness, while their continued dependence on input and validation quality is made clear.

Changes made in the revised manuscript: Sect. 5.1, lines 598–662; Sect. 6.1, lines 824–845.

Comment 2

> Methodological transparency of the DTM framework
>
> The Dynamic Topic Modelling approach is informative, but its limitations are not discussed. The sensitivity of topic classification to keyword selection, corpus composition, and differential coverage between WoS and Scopus is not addressed. Although the authors mention "deterministic training protocols" (l. 86), the implications for topic boundaries and temporal trends are unexplored. Greater methodological transparency would improve reproducibility and allow readers to assess the robustness of the thematic structure underpinning the review.

Response to Comment 2

We agree that reproducibility required more than the earlier statement about deterministic training. Supplementary Sect. S2 now documents the complete TAC-DTM workflow: record consolidation, preprocessing, lemmatisation, minimum document length, standard and expanded stop-word handling, regex filtering, curated phrase locking, automatic bigram mining, annual time slicing, topic modelling, stabilisation, manual thematic consolidation and trend calculation.

The principal model settings are now reported explicitly: K_TOPICS = 15, SEED = 2025, workers = 1, ITER = 900, MIN_CF = 15 and RM_TOP = 8. The Supplement also describes the ten-round stabilisation procedure, the inspection of anchor vocabularies and the sensitivity checks in which topic number and vocabulary filters were varied. The phrase whitelist and expanded stop-word dictionary are reproduced so that the preprocessing can be inspected and rerun.

We have also added the limitation requested by the reviewer. Topic boundaries and annual trends are explicitly described as conditional on corpus composition, search terms, keyword and dictionary choices, vocabulary filtering, and differential disciplinary and regional coverage in Web of Science and Scopus. Deterministic controls ensure reproducibility for a fixed corpus and configuration; they do not remove these sources of structural sensitivity. Consequently, topic trends are interpreted only as changes within the indexed corpus, not as exact measures of global research activity.

Finally, Sect. 2.3 now links each consolidated thematic domain to the process and management sections that follow, preventing the topic model from operating as a detached bibliometric result.

Changes made in the revised manuscript: Sects. 2.1–2.3, lines 95–239; Supplementary Sect. S2, lines 64–251.

Comment 3

Representativeness of the 42 flash-flood events

The synthesis of 42 events in Figure 3 is visually effective, but the representativeness of this sample is insufficiently discussed. Given the dominance of Asian case studies, it is unclear whether the reported intensity-duration relationships capture hydro-climatic variability across semi-arid Mediterranean, North African, or South American environments. The selection criteria for these 42 events from 1,967 studies are not stated, and potential biases are not acknowledged. This limitation directly affects the validity of the derived thresholds and their transferability.

Response to Comment 3

We agree and have clarified the construction, purpose and limitations of the event evidence. The revised analysis distinguishes the 95 documented events shown geographically in Fig. 3a from the 42-event rainfall subset used in Fig. 3b–c. Supplementary Sect. S1 states that an event record required an identifiable location and occurrence time and evidence of causal processes and impacts; reports of the same disaster were matched by location and time. The 42-event subset required quantitative rainfall amount and corresponding duration reported in the source publication. Missing rainfall information was not inferred.

The text and caption now explicitly state that the subset is geographically uneven, dominated by Asian cases, and contains relatively few events from semi-arid Mediterranean basins, North Africa and South America. The 2022–2024 summer-mean MIDI layer is described only as broad spatial context and not as the antecedent moisture state of each historical event.

Most importantly, no globally transferable threshold is derived from these 42 cases. The distributions are treated as descriptive evidence of variability among documented events and are used to support the state-dependent threshold framework. The manuscript now states that operational thresholds must be locally calibrated and dynamically updated using intensity, duration, antecedent moisture and other catchment conditions.

Changes made in the revised manuscript: Sect. 2.1, lines 101–104; Sect. 3.1 and Fig. 3, lines 256–278; Supplementary Sect. S1, lines 15–61.

Comment 4

Nature-Based Solutions lacks comparative evidence

Section 4.2.2 presents NbS positively but remains largely conceptual. Key dimensions are underdeveloped: long-term effectiveness under increasing rainfall extremes, maintenance costs over decadal timescales, scale dependency of flood attenuation effects, and comparative performance relative to structural measures. The manuscript acknowledges that plant growth takes time (l. 619) but does not explore how this time-lag interacts with near-term flood risk. A more evidence-based treatment would improve the practical relevance of this section.

Response to Comment 4

We agree and have substantially revised the Nature-Based Solutions discussion. It now presents NbS as complements to structural protection whose performance depends on catchment position, implementation extent, available storage, vegetation maturity, maintenance and event magnitude. The revised section includes scale-dependent evidence from Belo Horizonte, where simulated peak-flow reductions varied from approximately 60% to 5% as implementation decreased from 100% to 10% of suitable areas, together with evidence on wetland storage and small-catchment attenuation.

The time lag is now treated as an operational constraint. The manuscript states that plant growth and soil recovery are gradual and that newly implemented vegetation cannot be assumed to provide the function of mature cover immediately. Warning, emergency response and existing protection therefore remain necessary during establishment. The text also notes that passive blue–green infrastructure can provide insufficient storage during extreme rainfall and that forest expansion does not remove all multi-peak flood responses.

We have added a direct comparison with structural measures. Structural works can offer more immediate and design-specific protection, but they require life-cycle maintenance, lose storage through sedimentation and can alter channel and ecological conditions. NbS provide distributed regulation and co-benefits, but their effects are slower and scale dependent. Because empirical evidence on decadal performance and maintenance cost remains sparse, the revised manuscript does not claim a universal cost-effectiveness advantage for either approach; it identifies this as an evidence gap and supports context-specific grey–green portfolios.

Changes made in the revised manuscript: Sect. 5.2.1, lines 705–760; Sect. 5.2.2, lines 762–796; Sect. 6.3, lines 870–884.

Comment 5

Sediment connectivity is never conceptualised

Hillslope–channel connectivity is repeatedly cited as a key control on flash-flood initiation in Sections 3.1.1–3.1.2, yet sediment connectivity is never developed as a conceptual framework. A clearer distinction between hydrological and sediment connectivity would strengthen the process synthesis. Integrating frameworks such as the sediment cascade concept, the connectivity index, and source–transfer–sink organisation would help clarify the conditions under which flash floods transition toward hyperconcentrated flows or debris flows — a distinction that remains underspecified throughout.

Response to Comment 5

We agree and have added a dedicated Sect. 3.3 entitled "Hydrological and sediment connectivity in mountain flash-flood processes." Hydrological connectivity is now defined as the event-scale activation and synchronisation of runoff pathways from hillslopes and tributaries to the main channel. Sediment connectivity is defined separately as the ability to transfer erodible material from sources through pathways to downstream sinks.

The revised section distinguishes structural connectivity—the potential spatial coupling of sources and pathways—from functional connectivity, which describes how strongly that coupling is activated during a particular event. It introduces connectivity indices and a source–transfer–sink organisation that includes landslides, gullies, wildfire- or earthquake-derived material and channel deposits as sources; steep gullies and confined channels as transfer pathways; and fans, floodplains, low-gradient reaches, reservoirs and check dams as sinks.

This framework clarifies that rapid hydrological response does not necessarily imply a sediment-rich flood. The transition toward hyperconcentrated flow, debris flood or debris flow depends on the simultaneous availability of erodible material, connection to the channel network, transport capacity and event-scale remobilisation. The text also explains why no single sediment-concentration threshold is imposed across heterogeneous catchments and includes coupled hydro-sediment-morphological modelling evidence.

Changes made in the revised manuscript: Sect. 3.3, lines 388–434.

Comment 6

Cascading hazard interactions need structured treatment

Section 3.2 introduces cascading impacts but treats interacting hazard types largely as separate illustrative examples rather than components of coupled hazard chains. The multi-hazard relevance of the review would be strengthened by explicit synthesis of cascade pathways such as wildfire-induced hydrophobicity driving debris flow generation, GLOF triggering sediment-rich flood waves, landslide dam breach flooding, or flash flood-driven infrastructure failure. Figure 4, intended to represent cascading impacts, does not illustrate dynamic coupling between hazard components, limiting its conceptual value.

Response to Comment 6

We agree and have completely restructured the cascading-hazard treatment. Section 4.2.2 now distinguishes: (i) cross-event cascades, in which wildfire, earthquake, volcanic or cryospheric disturbance leaves persistent hydrological or sedimentary legacies; (ii) within-event cascades, in which an initial process such as a landslide, rock–ice avalanche, moraine or landslide-dam breach rapidly generates or transforms a flood; and (iii) downstream amplification through channel blockage, temporary barrier failure, engineered dam failure, tailings-dam failure and critical-infrastructure disruption.

The revised synthesis explicitly analyses the pathways identified by the reviewer, including post-fire hydrophobicity and sediment mobilisation, GLOF generation and downstream dam failure, landslide-dam breach flooding, earthquake-derived sediment remobilisation, large-wood blockage, and flood-related failure of roads, bridges, power, water-treatment and communication systems. It also distinguishes causal chains from simple spatial co-occurrence and notes that most quantitative evidence still comes from individual event reconstructions and site-specific models.

Figure 4 has been redesigned as a hazard-chain and cascading-impact framework. It now links cross-event disturbance legacies, within-event flood generation and transformation, downstream hazard amplification, direct impacts and longer-term economic, social and ecological consequences.

Changes made in the revised manuscript: Sect. 4.2.2 and Fig. 4, lines 508–580.

Comment 7

Geographic imbalance is not integrated into the synthesis

The manuscript recognises the underrepresentation of Africa and South America but does not integrate this observation into process findings or management recommendations. This creates an internal inconsistency: the Boolean query explicitly targets data-scarce environments through terms such as "wadi" and "steep slope", yet the synthesis remains dominated by European and Asian evidence. Figure 1b documents events in Morocco, Ethiopia, and South Africa, but these regions receive minimal attention in subsequent sections. Mountain environments such as the Atlas ranges, Maghrebian semi-arid basins, and Andean catchments — characterised by ephemeral channels, high sediment availability, sparse gauging networks, and extreme rainfall intermittency — offer important complementary perspectives necessary to substantiate the claim of a "globally comparable baseline."

Response to Comment 7

We agree that acknowledging geographic imbalance is insufficient unless it affects interpretation. We have therefore integrated this limitation at three analytical levels.

First, Sect. 2.2 replaces a purely country-based description with a reproducible GMBA mountain-system comparison. The 506 system-resolved papers are compared by evidence source and applied focus using counts, within-system percentages and location quotients. This analysis shows that observation-rich profiles in parts of Europe cannot be assumed for the East Sahara, Atlas, Andes or East African Highlands, where evidence is smaller, more fragmented, and often more dependent on remote sensing, reconstruction or site-specific modelling.

Second, Sect. 3 uses a process-based comparison across the available cases. Arid wadi, Mediterranean, tropical, high-mountain, humid monsoon and urbanising settings are compared through storm

intermittency and spatial concentration, antecedent state, infiltration and storage, sediment availability, ephemeral or confined channels, tributary synchronisation and connectivity. Because quantitative event data are sparse and evidence types differ, we did not impose a region-by-region threshold table that would suggest unsupported comparability. Instead, regional evidence limitations are carried into the interpretation of threshold transferability and causal mechanisms.

Third, Sects. 5–6 condition management recommendations on monitoring density, warning lead time, sediment regime, land availability, maintenance capacity, institutional coordination and local acceptance. The revised conclusion specifically prioritises high-resolution monitoring, long-term observatories and open event inventories in under-monitored mountain regions.

Changes made in the revised manuscript: Sect. 1, lines 63–93; Sect. 2.2, lines 145–190; Sect. 3, lines 242–434; Sects. 5–6, lines 584–910; Conclusion, lines 938–943; Table A1; Supplementary Sect. S3.

Comment 8

> Terminological inconsistency
>
> The terms "flash flood", "debris flow", "hyperconcentrated flow", and "mountain torrent" are used interchangeably in several passages. A brief terminological framework in the Introduction would improve consistency throughout the manuscript.

Response to Comment 8

We agree and have added a concise terminological framework to the Introduction. “Mountain flash flood” is the principal term for rapid-onset flooding in steep catchments. “Mountain torrent” is retained only when it reflects terminology in the source literature and is not treated as a separate hazard class. “Hyperconcentrated flow” and “debris flood” denote sediment-rich flood regimes, whereas “debris flow” is treated as a related but distinct gravity-driven mass-flow process. GLOFs, landslide-dam breaches and dam failures are discussed only as triggers or cascading pathways connected to the main review object.

This distinction is carried into Sect. 3.3, where hydrological and sediment connectivity explains the conditions under which clear-water flooding becomes sediment rich, and into Sect. 4.2.2, where related hazards are positioned within explicit causal chains rather than merged terminologically.

Changes made in the revised manuscript: Sect. 1, lines 40–45; Sect. 3.3, lines 388–434; Sect. 4.2.2, lines 508–580.

Comment 9

> Figure 2 readability
>
> The thematic pillar figure is visually dense. Improving label readability and clarifying hierarchical relationships between sub-topics would make this figure more accessible to readers.

Response to Comment 9

We have redesigned Figure 2 as a two-part synthesis. The figure now combines a quantitative overview of the six consolidated thematic domains and their publication counts with a simplified card-based presentation of the principal subtopics. Crowded radial text has been removed, label size and spacing have been increased, and the hierarchy between each domain and its subtopics is now explicit.

The accompanying text and caption also explain how the six domains connect to the subsequent review: risk management and adaptation, meteorological forcing, arid wadi flooding, sediment processes, special geological and topographic settings, and eco-climate impacts become organising links to the later process, impact and governance sections.

Changes made in the revised manuscript: Sect. 2.3 and Fig. 2, lines 192–240.

Comment 10

Figure 4 impact levels

The circular layout of Figure 4 does not clearly differentiate between primary flood impacts and secondary cascading effects. A clearer visual distinction between impact levels would improve its analytical value.

Response to Comment 10

We agree and have replaced the former circular impact inventory with a layered hazard-chain framework. The revised Figure 4 distinguishes cross-event disturbance legacies, within-event flood generation and transformation, downstream amplification, exposed receptors, realised direct impacts and longer-term consequences.

In the accompanying text, direct economic, social and ecological impacts are first treated in Sect. 4.2.1. Section 4.2.2 then explains secondary and cascading effects, including temporary or engineered barrier failure, contamination, loss of access, delayed evacuation, failure of power or communication systems and prolonged disruption of essential services. This structure makes the causal order and the distinction between impact levels explicit.

Changes made in the revised manuscript: Sects. 4.2.1–4.2.2 and Fig. 4, lines 478–580.

Comment 11

Transition between Sections 3 and 4

The transition between process mechanisms and management measures is abrupt. A short linking paragraph connecting process understanding to governance implications would improve the logical flow of the manuscript.

Response to Comment 11

We have added explicit transitions at both relevant points in the revised structure. The opening of Sect. 4 now explains how storm forcing, runoff, sediment and debris become disaster risk only when they intersect with exposure, vulnerability and response capacity. It also previews the progression from these conditions to realised impacts and cascading failures.

The opening of Sect. 5 then links the hazard–risk–impact synthesis to intervention. It explains how the evidence on process and consequence identifies distinct intervention points: modifying hazard processes, reducing exposure, addressing vulnerability, strengthening response capacity and managing residual losses. These bridges make clear why particular monitoring, planning, structural, Nature-Based and response measures follow from the mechanisms analysed earlier.

Changes made in the revised manuscript: Sect. 4 opening, lines 435–443; Sect. 5 opening, lines 584–595.

Comment 12

North African and Mediterranean references

Despite the presence of wadi-related systems in the Boolean query, references from the Moroccan High Atlas, Algerian Tell, and Tunisian Dorsale are largely absent. A more balanced citation base would strengthen the geographic credibility of the review.

Response to Comment 12

We agree that North African and Mediterranean evidence needed clearer treatment. We re-screened the corpus and revised the regional synthesis using a consistent spatial hierarchy rather than mixing countries, small subranges and individual basins. Section 2.2 now uses GMBA major mountain systems as the primary comparison unit. It includes the East Sahara Mountains (40 papers) and the Atlas

Mountains (7 papers), with explicit discussion of remote-sensing and modelling dominance in the East Sahara and of available work from the Moroccan High Atlas. It also states directly that the Atlas evidence base is too small to define a stable regional evidence profile.

The re-screening did not yield a sufficiently robust and comparable body of eligible studies to support separate quantitative profiles for the Algerian Tell or Tunisian Dorsale. Creating standalone categories for these small subranges would have mixed spatial scales and implied a level of evidential precision that the corpus cannot support. We therefore retained the consistent major-mountain-system framework and made the remaining gap explicit rather than filling it with weakly comparable material.

At the process level, Mediterranean and semi-arid evidence is integrated throughout Sect. 3, including short-duration rainfall extremes and changing rainfall rates in southern Italy and Genoa, terrain-modulated convection, arid wadi runoff in Egypt, Mediterranean urbanisation effects, and sediment/connectivity controls. Management evidence from Mediterranean mountain torrents and semi-arid catchments is also incorporated in the structural, NbS and trade-off sections. This provides substantive process and management context while transparently acknowledging that North African coverage remains limited.

Changes made in the revised manuscript: Sect. 2.2, lines 145–190; Sect. 3, lines 242–434; Sects. 5.2–6.3, lines 705–910; Table A1; Supplementary Sect. S3.

Comment 13

Future research priorities

The conclusion summarises findings effectively but does not include a structured discussion of future research directions. A short subsection identifying priorities — high-resolution monitoring in data-scarce regions, coupled hydro-sedimentary modelling, uncertainty quantification in ML-based systems, and integration of AI with physically based models — would add significant value for the research community.

In relation to the NHESS review criteria, I consider the manuscript to address a relevant topic within the scope of the journal and to have clear scientific potential. However, its scientific quality is currently limited by insufficient integration between the bibliometric and qualitative components, incomplete methodological transparency, and an underdeveloped treatment of geographic representativeness, sediment connectivity, and cascading hazards. The presentation is generally clear, but some figures and sections require clarification and restructuring.

Response to Comment 13

We agree and have added a structured future-research paragraph to the Conclusion. It prioritises high-resolution rainfall and discharge monitoring, longer-term observatories and open event inventories in under-monitored mountain regions. It also calls for stronger uncertainty quantification, spatial and temporal validation and out-of-sample testing of machine-learning warning systems, together with closer integration of artificial intelligence, physical models and real-time observations.

The revised Conclusion further identifies the need for comparative quantitative evaluation of intervention performance, life-cycle costs and implementation conditions so that transferability can be assessed rather than assumed. The need to couple runoff generation with sediment transfer and morphological change is developed in Sect. 3.3 through the hydrological/sediment connectivity framework and coupled hydro-sediment-morphological modelling evidence. Together these additions cover the monitoring, hydro-sedimentary, ML uncertainty, hybrid AI–physics and intervention-evaluation priorities identified by the reviewer.

Changes made in the revised manuscript: Sect. 3.3, lines 388–434; Conclusion, lines 938–943.

Recommendation

major revision

This manuscript is scientifically ambitious and addresses a topic of clear relevance for the NHESS community. The bibliometric foundation is solid, the governance framework is well structured, and the interdisciplinary scope is commendable. However, several structural and conceptual weaknesses require substantive revision before the manuscript can be considered for publication. These include the incomplete integration of bibliometric findings into the qualitative synthesis, the underdevelopment of key conceptual frameworks, the insufficient treatment of sediment connectivity and cascading hazard interactions, and the geographic imbalance that currently undermines the claim of global comparability. Addressing these issues carefully would substantially strengthen a manuscript that has clear potential for significant impact in the field.

Overall response to the recommendation

We sincerely thank the reviewer for the clear recommendation and constructive guidance. The manuscript has undergone the substantive revision requested. The bibliometric and qualitative components are now connected through the six thematic domains and the regional evidence framework; TAC-DTM methods and limitations are documented in detail; geographic imbalance is quantified and carried into process and management interpretation; sediment connectivity and cascading hazards have dedicated conceptual sections; and Figures 2 and 4 have been redesigned.

We have also narrowed claims of global comparability, clarified terminology and scope, expanded uncertainty and machine-learning limitations, and added critical treatments of NbS, costs, trade-offs and implementation conditions. We believe these changes have substantially improved the manuscript's scientific rigour, transparency and relevance to NHESS.