Supplement of Flash Floods in Mountainous Regions: Global Research Trends, Process Mechanisms, and Control Measures

Zhenghang Chen[1,2], Meili Feng[1*], Matthew F. Johnson[4*], Nigel Wright[5], Ying Weng[6], Faith Ka Shun Chan[1,3*], Feng Wu[2*]

[1]Faculty of Science and Engineering, University of Nottingham Ningbo China, 315100 Ningbo, China
[2]Key Laboratory of Land Surface Pattern and Simulation, Institute of Geographic Sciences and Natural Resources Research, CAS, Beijing, 100101, China
[3]Water@Leeds Research Institute and School of Geography, University of Leeds, LS2 9JT, UK
[4]School of Geography, University of Nottingham, NG7 2RD, UK
[5]School of Engineering, University of Birmingham, Edgbaston, Birmingham B15 2TT, United Kingdom
[6]School of Computer Science, University of Nottingham Ningbo China, 315100 Ningbo, China

Correspondence to: Faith Ka Shun Chan (Faith.Chan@nottingham.edu.cn), Meili Feng (Meili.Feng@nottingham.edu.cn), Feng Wu (wufeng@igsnrr.ac.cn), Matthew F. Johnson (m.johnson@nottingham.ac.uk)

S1 Data collection

This review employs a systematic, transparent approach to identify, screen, and synthesise academic literature on flash floods in mountainous environments. The objective is to capture a representative body of peer-reviewed evidence while minimising selection bias through a predefined search strategy, explicit inclusion/exclusion criteria, and a staged screening procedure. The overall process was designed to be reproducible and scientifically robust, and it follows established systematic review reporting principles (PRISMA 2020; Page et al., 2021).

An exhaustive keyword search was conducted in the Web of Science Core Collection and Scopus databases for the period 2000–2025. The search strategy combines (i) disaster-related vocabulary (e.g., flash flood, cloudburst, rapid-onset flood) and (ii) terrain-related descriptors (e.g., mountain, alpine, upland, wadi, arroyo) to cover heterogeneous terminology across disciplines and regions. By jointly constraining hazard type and terrain setting, the search focuses on mountainous flash floods while reducing the inclusion of non-target contexts such as coastal flooding or urban waterlogging. The Boolean query used was:

("flash flood*" OR "flash-flood*" OR "pluvial flash flood*" OR
"rapid-onset flood*" OR "short-duration flood*" OR "cloudburst*" OR
"cloud burst" OR "pluvial flood*" OR "gully flood*" OR "torrent flood*" OR
"mountain torrent*" OR (flood* NEAR/3 (flash OR rapid OR short-duration OR pluvial OR torrent OR gully)))
AND
(mountain* OR alpine OR "hilly area*" OR hill* OR upland* OR highland* OR
plateau* OR "high-elevation" OR "steep slope*" OR foothill* OR headwater* OR
"small catchment*" OR orographic OR wadi* OR arroyo* OR ravine* OR torrent* OR gulch*)

In Web of Science, the query was applied to the Topic field (TS). In Scopus, it was applied to Title–Abstract–Keywords (TITLE-ABS-KEY). Search outputs were restricted to peer-reviewed journal research articles published in English between 2000 and 2025.

To be included, studies had to meet the following criteria:

(1) Publication type: peer-reviewed journal research article;

(2) Time window: 2000–2025;

(3) Study setting: flash floods in mountainous, hilly, or alpine environments;

(4) Topical relevance: addressing flash-flood processes, impacts, modelling, or prevention and control/mitigation;

(5) Language: English.

Studies were excluded if they met any of the following conditions:

(1) Non-article types: reviews, editorials, conference abstracts, news, briefings, or letters;

(2) Non-peer-reviewed materials: reports and grey literature;

(3) Off-target hazard contexts: coastal floods, large-river floods, storm surges, or lowland/urban waterlogging;

(4) Non-flash-flood hazards: disasters not directly addressing flash floods (e.g., landslides, glacier-lake outburst floods, debris flows);

(5) Duplicates across databases.

Records retrieved from the two databases were merged, and duplicates were removed. The remaining records underwent staged screening based on titles/abstracts and, where necessary, full-text checks against the criteria above. Following this search and screening process, 1,967 papers were retained and included in the final analysis.

S2 TAC-DTM framework

To analyse an interdisciplinary and large-volume corpus on mountainous flash floods spanning hydrology, meteorology, geomorphology and sediment dynamics, ecology, disaster management, and socioeconomics, we adopted a reproducible thematic workflow that combines dynamic topic modelling with manual thematic consolidation and trend mapping. The goal was to extract interpretable themes from the literature, quantify how their prominence evolves over time, and generate stable keyword sets that can be inspected and reported in a transparent manner. The modelling corpus was built from bibliographic records containing ID, year, title, and abstract. Because the same publication could appear in multiple rows, records were consolidated into a single document prior to modelling by grouping on the document ID when available; when IDs were missing, a fallback key based on title and year was used. During consolidation, the earliest year was retained if discrepancies occurred, titles were kept as the first non-empty entry, and abstract text was concatenated after removing duplicate fragments and publisher trailing notes.

Dynamic Topic Model was employed because static topic models assume time-invariant topic–word distributions and therefore cannot represent shifts in vocabulary and research emphasis over a multi-decade window. In contrast, DTM estimates topic–word distributions for successive time slices and constrains their evolution to be gradual, which enables the recovery of both thematic structure and its temporal dynamics. All documents were assigned to yearly timepoints spanning 2000–2025, so that model outputs include time-varying topic keywords and document-level topic mixtures that support trend analysis across years.

Text preprocessing was designed to preserve domain semantics while suppressing noise. Titles and abstracts were concatenated into a single text field and normalised by removing punctuation artefacts, excessive whitespace, and copyright or publisher tail notes. Tokenisation was performed using NLTK; lemmatisation was enabled in the main runs, corresponding to the setting DO_LEMMA = 1 and WordNet-based lemmatisation, with required NLTK resources downloaded when absent. Standard English stop words were removed using the NLTK stop-word list, and an expanded stop list was additionally applied to remove publisher and formatting artefacts, weak-information academic filler terms, measurement units and symbols, pure numbers and years, common abbreviations, Roman numerals, and high-frequency geographic names that otherwise dominate global corpora. In parallel, regex guards were used to filter purely numeric tokens, percentages and ratio-like forms, and explicit year strings. Documents were filtered by minimum length to reduce instability from very short texts, using a minimum token threshold MIN_TOK = 15.

Because flash-flood research relies heavily on technical collocations, phrase preservation was treated as a core step. Two-word phrases were merged into single tokens using a combined strategy consisting of a curated binary-phrase whitelist and automatic bigram mining. The whitelist enforced the retention of domain collocations as single tokens, while automatic mining identified additional collocations in the corpus under minimum frequency and association constraints; in the implementation, the automatic mining required a minimum bigram frequency of 15 and a minimum pointwise mutual information threshold of 3.0. The final phrase set was defined as the union of the whitelist phrases and the automatically mined phrases, and merging was applied greedily from left to right to ensure consistent tokenisation of recurrent two-word terms.

To ensure that the topic modelling results were reproducible and not sensitive to implementation randomness, deterministic training controls were enforced throughout. In all model runs, we applied inverse document frequency (IDF) term weighting, fixed the random seed (SEED = 2025), and trained the model in a single thread (workers = 1) to avoid non-determinism from parallel updates. The DTM was trained using tomotopy with the number of topics set to K_TOPICS = 15 as the selected topic granularity, the number of training iterations set to ITER = 900, the minimum corpus frequency for vocabulary inclusion set to MIN_CF = 15, and the number of globally high-frequency terms removed set to RM_TOP = 8. Training was performed in incremental steps and monitored via log-likelihood per word when supported by the library version to confirm stable convergence behaviour.

Because topic modelling does not yield a unique solution, the modelling pipeline included an iterative stabilisation procedure to obtain consistent anchor vocabularies. Preprocessing and modelling were repeated across multiple rounds in which near-synonymous expressions were merged and cleaned based on co-occurrence structure and topic outputs were inspected for interpretability. Anchor words were treated as stable when repeated runs recovered the same core semantic signatures and separations, and in practice this stabilisation was reached after ten rounds. In addition, robustness was evaluated by varying key hyperparameters that control granularity and filtering, including the number of topics, global high-frequency term removal, and minimum corpus frequency, and by comparing stable versus edge components of topic-word sets across settings. Stable components were retained as the basis for thematic consolidation, whereas parameter-sensitive fragments were treated as candidates for pruning.

To operationalise the stabilised preprocessing and to make the modelling fully reproducible, we implemented two explicit domain dictionaries in the codebase. The first is a curated phrase whitelist used for binary phrase locking, ensuring that core technical collocations are always merged into a single token before modelling. The second is an expanded stop-word list that is applied in addition to the standard NLTK English stop words to suppress publisher artefacts, weak-information academic filler terms, units and formatting residues, common abbreviations, and high-frequency place names that would otherwise dominate topic–word distributions in a global corpus. The exact dictionaries used are reproduced below.

Domain phrase whitelist (force-merged into a single token; bigrams only) :

```
PHRASE_WHITELIST = {
    # Hazard/process/event
    "convective storm","extreme rainfall","intense rainfall","heavy rainfall","heavy precipitation",
    "extreme precipitation","rainfall event","storm event","rainfall threshold","rainfall intensity",
    "rainfall runoff","runoff generation","extreme event","outburst flood","flood hazard",
    "flood risk","flood protection","flood inundation","floodplain connectivity","early warning",
    "hazard map","risk map","risk assessment","vulnerability assessment","flood mitigation","flood control",

    # Geomorphology/terrain and hydrologic entities
    "alluvial fan","mountain torrent","mountain stream","landslide dam","glacial lake",
    "headwater catchment","small catchment","small watershed","river basin","drainage basin",
    "drainage network","channel network","flood plain","karst aquifer","karst landscape",
    "piedmont plain","river bed","ephemeral stream",

    # Hydrologic/sediment metrics
    "runoff coefficient","suspended sediment","sediment transport","sediment concentration",
    "sediment yield","sediment deposition","peak discharge","peak flow","soil moisture",
    "soil water","flow depth","stream power",
```

```
    # Data/observation/tools and models
    "radar rainfall","weather radar","rain gauge","gauge network","remote sensing",
    "numerical simulation","numerical model","hydrological model","hydrologic model",
    "hydrodynamic model","numerical weather","weather prediction","numerical weather prediction",
    "quantitative precipitation","precipitation estimation","precipitation forecast",
    "quantitative precipitation estimation","quantitative precipitation forecast",
    "distributed hydrological","hydrological model",
    "wrf model","lstm model","random forest","machine learning","deep learning",
    "support vector","vector machine","intensity duration",
    "land use","land cover","analytical hierarchy","hierarchy process",

    # Concepts/solutions
    "nature based","nature-based solution","climate change",
}
```

Additional stop words (used together with NLTK stop words) :

```
EXTRA_STOP = {
    # Publisher/formatting noise
    "elsevier","rights","reserved","copyright","springer","wiley","taylor","francis",
    "permission","figure","table","issue","vol","doi","al","et_al","data_set","dataset",
    "ltd","university","section","supplementary","author","chapter","section","example","question","version",

    # English function words (supplementary)
    "that","this","these","those","are","is","was","were","be","being","been",
    "as","by","on","in","of","for","to","from","with","at","an","and","or","nor",
    "not","than","which","who","whom","into","over","under","between","among",
    "within","without","per",

    # Weak-information academic filler terms
    "value","values","show","shows","shown","significant","significantly","result","results",
    "analysis","analyses","based","using","used","use","method","methods","approach",
    "approaches","paper","study","studies","case","cases","also","well","general",
    "different","various","several","two","one","many","high","low","due","however",
```

```
    "therefore","including","present","presented","compared","available","type","types",
    "term","terms","period","factor","factors","parameter","parameters","effect","effects",
    "response","responses","state","states","region","regions","research","knowledge",
    "pattern","property","properties","feature","features","option","problem","solution",
    "practice","issue",

    # Units/dimensions and formatting residues
    "km","km2","mm","cm","m3","kg","ha","yr","yrs","°c","deg","degree","degrees",
    "mm/h","m3/s","/s","1.","2010.","2018.","r2",

    # Technical/writing abbreviations (high-frequency in this corpus)
    "inf","/inf","mc","bp","sr","la","le","el","ii","iii","iv","vi","vii","ix",
    "min","max","avg","etc","vs","ca","eg","ie","i.e","e.g","u.s.","us","b.v","e.","m.","ass","and/or",
    "mg","ni","john_son","hpa","sc","pb","gh","gi",

    # Broad place names (optional; remove this block if geographic topics are intended to be retained)
    "oman","saudi","pakistan","bhutan","romania","egypt","poland","germany","france",
    "california","switzerland","nepal","qena","uttarakhand","sikkim","beijing","korea",
    "vietnam","malaysia","bangladesh","algeria","argentina","mexico","spain","italy","japan",
    "america","american","usa","uk","china","india","taiwan","arabian","himalaya","tibetan_plateau",
    "pyrenees","carpathians","tatra","appalachian","texas","jordan","madeira","sinai","andes",

    # Other low-information nouns/blocks that tend to dominate
    "grid","class","category","score","character","service","program","platform","mission",
    "project","information","detail","details",
}
```

After an initial DTM run with 15 topics (K_TOPICS = 15) and repeated preprocessing iterations to stabilise anchor vocabularies, the machine-derived topics were manually consolidated into six macro-themes to obtain a tractable, literature-facing taxonomy. This consolidation was performed by jointly considering the time-varying top words of each topic across yearly slices and a set of representative documents with high topic shares (high-loading papers) that best expressed each topic's semantic core. Based on these materials, we drafted explicit thematic definitions and compiled theme-specific keyword cues and boundary rules, including exclusion rules for frequent borderline cases (e.g., debris-flow-only studies without a flash-flood component; large-river flood papers where flash flooding was peripheral). Once the six macro-themes were fixed, all documents were manually screened and assigned one or multiple theme labels using titles and abstracts as the primary evidence. Multi-label assignment was retained deliberately to preserve interdisciplinary overlap and to avoid forcing single-theme classification in studies that simultaneously address rainfall forcing, hydrologic response, sediment dynamics, and risk management.

On the premise of retaining vocabulary evidence and avoiding extrapolation, we summarized similar topics into six superordinate categories and sorted out their interrelationships. The final main research topics are: (1) Meteorological triggers and rainfall patterns define event-scale precipitation drivers, including monsoons, strong convection, orographic rain, stratiform precipitation and cyclone-related rainfall, and answer "how does the sky rain?" (2) Flash floods (Wadi) in ravines in arid areas. This category of research emphasizes the context of surges in intermittent ravines in arid/semi-arid regions. (3) This category studies the movement and accumulation of sediment in mountains. Focus on the debris flow-bedload-fan process and its quantitative characterization. (4) The impact of special geology and terrain on flash floods. This category of research brings together karst, moraine/alpine, volcanic, coastal boulder, and lithological terms as differential controls on the underlying surface and geological setting. (5) This category of research on flash flood risk management and adaptation strategies covers the entire process of flash flood risk management in mountainous areas, with the core focus being on strengthening the prediction capabilities, structural protection and resilience building in mountainous areas. Covers the parallel chain through hazard/risk maps, observation/assimilation/aggregation and nowcasting to alarm thresholds, engineering facilities (inspection dams/dams/embankments/reservoirs/drainage). (6) This category of research emphasizes ecological environmental impacts and climate change. Points to long-scale context such as water quality, lagoon/coast, habitat and vegetation cover, as well as regional comparisons and historical/archaeological sequences. In general, these six categories logically start from meteorological forcing, control the impact process and sediment production/accumulation through geomorphology and geology, and feed back to flash flood risk management through observation-forecasting-early warning and engineering response, and finally the ecological impact on the climate background on a longer time scale.

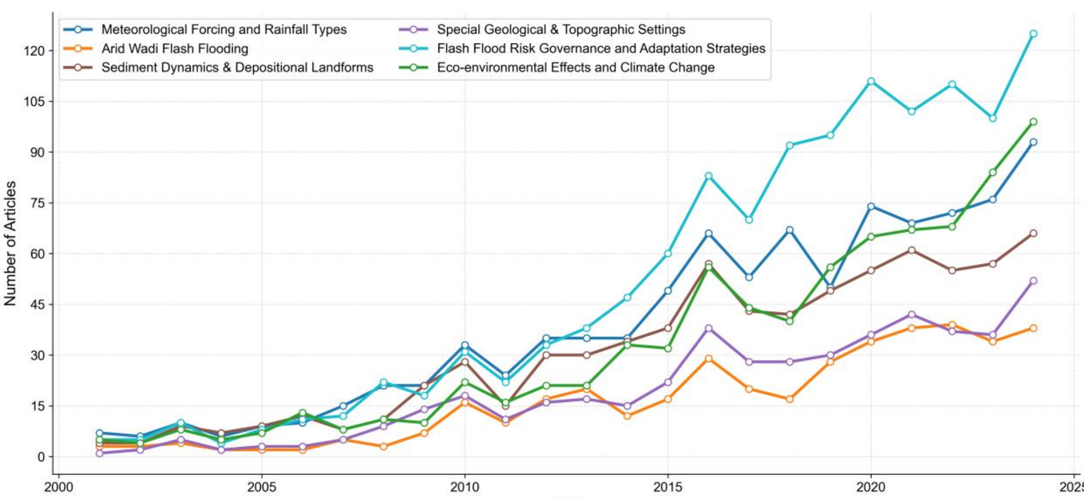

Fig. S1. Annual theme intensity time series (2001–2025). Lines with circle markers show the annual intensity of each theme, calculated as the sum of document-level topic papers for that theme in each year.

To refine within-theme heterogeneity and obtain more reliable keyword inventories for each macro-theme, we conducted a second modelling round after labelling. For each of the six themes, we extracted a theme-specific subset comprising all documents manually assigned to that theme, and trained a separate DTM on each subset, resulting in six within-theme DTM runs. The outputs were used to infer sub-topic structures and candidate keyword groups, which were then aggregated and filtered to form the final theme keyword inventories. Keyword aggregation drew on the top terms per topic and timepoint (TOP_N = 12), and consensus theme keywords were obtained by counting recurring terms across years and removing residual noise using the same token filters applied during preprocessing. Finally, the consensus keyword lists were manually screened to remove terms lacking clear semantic or domain relevance (e.g., residual generic or context-free tokens), while retaining informative terms as comprehensively as possible. The resulting, curated lists were reported as the final theme keyword inventories and used to compile theme summaries.

Topic prominence over time was quantified from the document-level topic mixtures estimated by the DTM. For each topic and year, annual prevalence was computed as the number of documents associated with that topic, yielding a yearly frequency series that captures shifts in thematic attention. These dynamics were visualised using line plots for selected topics supporting both interpretability checks and transparent reporting.