Beyond the surface: leveraging NLP to map global natural hazard impacts

Anonymous ACL submission

Abstract

Understanding how natural hazards such as floods, droughts, and storms turn into disasters requires robust impact data. This paper develops a comprehensive global dataset of natural hazard impacts from peer-reviewed literature, synthesizing existing in-depth knowledge. Leveraging advances in natural language processing (NLP) and large language models (LLMs), we mapped over 12,000 open-access articles published since 1980 on climatological, hydrological, and meteorological disasters. Evaluation results show that precision ranges from 0.85 to 1 for the extraction of quantitative impact information. Our novel method using Retrieval-Augmented Generation captures detailed impact data on a wide range of sectors and systems, significantly improving the granularity and geographical coverage compared to existing global datasets. As such, this work fills critical gaps in natural hazard research, providing information on both direct and indirect disaster consequences.

1 Introduction

Climatic, meteorological, and hydrological disasters lead to significant adverse costs for individuals, private companies, and governments, affecting the normal functioning of societies worldwide (Winter et al., 2024). In 2023 alone, extreme weather events caused more than US$ 250 bn in damage, claiming at least 74,000 lives (Re, 2023). While information on direct losses such as fatalities and economic impacts is often available, data on indirect societal impacts of disasters remains fragmented (Jones et al., 2022, 2023; Mahecha et al., 2020).

To effectively manage risk, a comprehensive understanding of the multiple sectors and systems impacted is needed (de Brito et al., 2024). Accurate impact data can support authorities in directing resources to the most vulnerable areas. By analyzing historical data on past impacts, researchers can identify trends, lessons learned, and best adaptation practices (Kreibich et al., 2017). Moreover, impact data is essential for estimating future impacts and developing impact-based early warning systems (Hagenlocher et al., 2023; Hobeichi et al., 2022).

Although freely accessible datasets with global coverage (e.g., (Delforge et al., 2023) and DesInventar (UNISDR, n.d.)) offer valuable insights into the socio-economic impacts across various hazard types, they face a number of limitations. They often prioritize quantitative impact metrics, such as the number of deaths and insured losses, which are straightforward to measure compared to intangible or indirect impacts (de Brito et al., 2024). However, the long-term effects of disaster include social impacts such as migration and forced displacement (McMichael, 2020), psychological trauma (Cianconi et al., 2020), and a loss of trust in institutions (Akbar and Aldrich, 2017). While these impacts may not be as immediately visible, they can have enduring consequences (Cochrane, 2004), potentially overshadowing the immediate tangible losses in the long run.

To address these gaps, we developed a novel approach to derive a global dataset of natural hazard impacts from the full text of peer-reviewed articles, extending beyond traditional impact classes. Leveraging the power of large language models (LLMs) and advanced Retrieval-Augmented Generation (RAG) to extract information from unstructured text, we derive quantitative data and qualitative descriptions of various impact classes (e.g. society, water availability, agriculture, health, infrastructure, and economy). We provide a structured dataset, with spatial resolution at the sub-national administrative level, previously unattainable.

2 Data and methods

2.1 Article screening and selection

To develop a global impact database of natural hazard impacts, we analyzed the full text of peer-

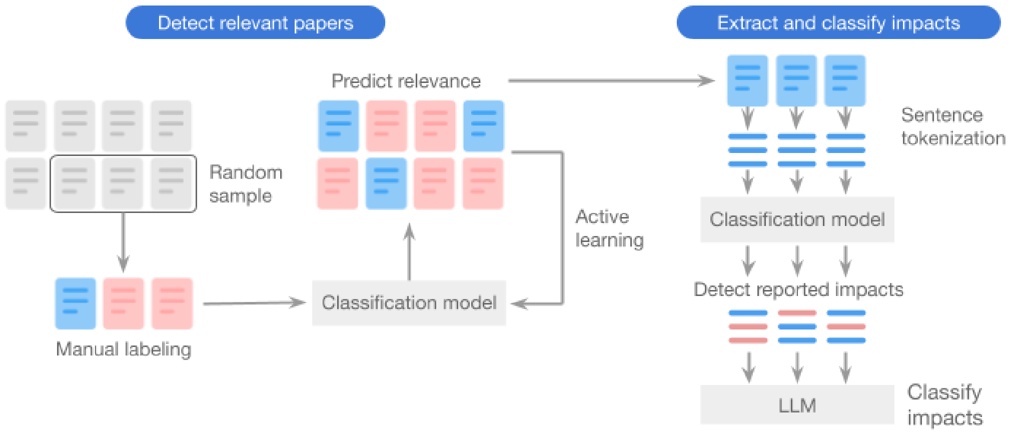

Figure 1: Methodological workflow.

reviewed articles reporting on the consequences of disasters. In order to identify relevant documents, we created a search string that integrates English terms associated with hazards and their potential impacts across various sectors (Appendix B, Table 4). Our search terms were based on the EM-DAT hazard classes (Delforge et al., 2023) and refined with input from domain experts and existing papers on impact assessment (e.g. (Sodoge et al., 2023)). With this, we aimed to capture major direct and indirect impacts caused by single or compound weather extreme events, including droughts, floods, heatwaves, cold waves, fires, storms, and mass movements.

The search was conducted in two databases: Science Direct and Pubmed, both of which allow for full-text retrieval. The first is a broad scientific database, while the latter focuses mainly on biomedical and life sciences. We restricted our search to open-access scientific articles under CC-BY licenses published between 1980 and 2023. The search was performed on titles, abstracts, and keywords, focusing on research articles and reviews while excluding book chapters, conference papers, reports, data articles, and short communications.

To identify relevant articles for full-text retrieval, we developed a transformer-based classification model. We manually labeled a randomly selected sample of 10% of the articles (n = 4,765), with an inter-annotator agreement of 96%. Among these, 173 were classified as relevant. An article was deemed relevant if the title and abstract described a hazard event (e.g. a flood) during a given time and location (i.e. country, region, city, water basin). We encoded the text with SciBERT (Beltagy et al., 2019), as it has been pre-trained on scientific articles. To address class imbalances, we employed an active learning approach (Callaghan et al., 2021) and used a weighted random sampler to select the training and testing splits. After applying the classification model, we further re-evaluated the 173 documents selected as relevant.

2.2 Disaster description

For each relevant article, we used a RAG method (Gao et al., 2024) to extract descriptive information about the investigated disaster from the article's title and abstract. This information includes the location, time, and type of hazard:

Location: The location affected by the disaster is specified across various geographical units, namely: country, state, city, water basin, latitude, and longitude. Named locations extracted with RAG were resolved into latitude and longitude using OpenStreetMap (OSM) (OpenStreetMap, 2017). Latitude and longitude (*numeric*) are within the interval of <-20,037,508.34, 20,037,508.34> and use the Mercator projection (EPSG 3857). Countries are specified with country names (*string*) and ISO 3166-1 alpha-3 codes (*string*). State and city are described with standardized international names from OSM (*string*) and the Global Administrative Areas (GADM) unique IDs (*integer*) (Areas, 2012). Water basins were cross-referenced with the data from HydroSHEDS (Lehner and Grill, 2013) and the World Meteorological Organization (WMO) Basins and Sub-basins (Global Runoff Data Centre, GRDC, 2020) to identify their unique basin ID (*integer*).

Time: The disaster time is represented by the start and end dates when available, both specified in YYYY-MM-DD format.

Hazard type: The hazard class is specified by a string selected from the following set: drought, flood, heatwave, cold wave, storm, mass movement, wildfire. These classes are compatible with existing global datasets (Delforge et al., 2023). To ensure consistency, we mapped the words predicted by the LLM to the predefined classes using a manual verbalizer (Schick et al.). The verbalizations incorporate the same terms used in the query search (e.g. "superstorm" would be mapped to "storm").

2.3 Sentence classification

To identify sentences describing disaster consequences, we performed a second labeling task in the article's full text. This step was necessary because, although impacts are often mentioned in the results section, they can also appear in other parts of the paper, such as the introduction or study area description. Additionally, papers may have different structures, making it difficult to automatically divide sections.

We selected 38 articles covering at least one of the investigated hazard types and diverse geographic areas. We tokenized the text into sentences using spaCy (Honnibal et al., 2020) and used INCEpTION (Klie et al., 2018) to annotate them. Each of the sentences (n = 6,585) was assigned to one or more of 17 classes, including "hazard impact".(Appendix C, Figure 2). A detailed description of the criteria for classifying the sentences is provided in Appendix C, Table 5. We then fine-tuned SciBERT (Beltagy et al., 2019) to identify sentences reporting impacts in unseen text. We trained the model with a stratified k-fold cross-validation approach.

2.4 Impact data extraction and classification

The impact classes were adapted from the classification used by the Intergovernmental Panel on Climate Change (IPCC) to report impacts of climate change on human systems (Shukla et al., 2022), and include impacts to (i) society, (ii) water availability, (iii) agriculture, (iv) health, (v) infrastructure and (vi) economic sectors. For each class, we included a qualitative (list of *strings*) and a quantitative description of the impacts reported. Quantitative information also includes some of the classes encompassed by the EM-DAT database, such as Human deaths, Injured and Affected.We detail below the quantitative information extracted for each of the new proposed impact classes.

Society: These might include an increase in violence, crime rates, and conflicts. Hazards can also incur losses of livelihood (e.g. by reducing the number of job opportunities and/or income) and internal migration (Kondo and Lizarralde, 2021). We do not extract any quantitative information for this class. We focus instead on extracting sentences that mention their occurrence.

Water availability: Hazards can alter water availability directly (e.g. prolonged droughts can lead to depleted reservoirs and reduced groundwater recharge) or indirectly (e.g. storms can cause the contamination of water sources). To quantify these impacts, we computed the number of water sources affected by the hazard. This is represented by an *integer* associated with the total number of affected sources and a list of tuples describing them *source* = *<sourceType, sourceName>*.

Agriculture: Impacts on agriculture were quantified by the production loss associated with the hazard. This loss is represented by the percent reduction (*numeric*) and the description of the crop type, which corresponds to the crop name (*string*) and a unique ID (*string*), extracted from the Indicative Crop Classification developed by the Food and Agriculture Organization (Som, 2010).

Health: Besides physical injuries and fatalities, hazards can lead to an increase in the incidence of infectious diseases, mental health issues, under-nutrition, and obesity rates (e.g. excess respiratory-related cases associated with thunderstorms). Health impacts are coded to unique IDs (*string*) from the list of three-character categories of the International Classification of Diseases from the World Health Organization (Organization, 2004).

Infrastructure: These include impacts to transportation, sanitation, health, IT, communication, energy, and housing. Quantitative measures extracted include the number of affected infrastructures, described by the value (*integer*) and a list containing the types of the affected infrastructures (*string* selected from the closed set: transportation network, energy, public, residential, IT and communication, health, sanitation and hygiene, education).

Economic sectors: Monetary losses are described by the value (*numeric*), currency (*string* following the ISO 4217 standard), associated sector (*string* selected from the closed set: industry and commerce, agriculture, forestry, tourism).

2.5 Validation

To validate the article selection and sentence classification models, we calculated accuracy, recall, and F1 scores on the datasets annotated by our team. Both models were trained with a stratified k-fold approach (k was set to 5), wherein each split, the subsets maintain the proportion of 50% of class labels. To validate the qualitative description of the impacts, we measured the similarity between real and predicted instances. To validate the quantitative measures extracted with RAG, we calculated the precision, i.e. the percentage of correctly extracted values out of all extracted values. For the impact data extraction and classification tasks, we compared two models: GPT-4 (Achiam et al., 2023) and Mistral (Jiang et al., 2023). Our evaluation in this paper is limited to the following representative fields: human deaths, number of affected people, health and residential infrastructures, and qualitative water and social impacts.

3 Results

Our query search returned 47,629 results from Science Direct and 13,933 from Pubmed. After duplicate removal, we ended up with a sample of 60,923 articles. The classification model estimated that 12,070 of these report hazard impacts, with an accuracy of 96% (Table 1). The sentence classification model identified more than 182,000 of the sentences that reported impacts of natural hazards, with an accuracy of 92%. The hazards investigated in these papers are distributed unequally worldwide: more than 50% of them happened in Europe and Asia, while only 10% affected Oceania and South America. This is in agreement with the findings of previous knowledge synthesis papers, which find an overrepresentation of hydro-hazard studies in the Global North compared to the Global South (Stein et al., 2024). This bias is likely caused due to the choice of English as the language of analysis.

We validated the information extraction method based on RAG (Table 2) with the dataset of 6,585 labeled sentences. Out of these, 664 reported impacts. Both models performed well in extracting specific quantitative information from text (precision > 0.8). Qualitative information was less accurate (similarity < 0.6). When comparing GTP-4 with Mistral, we found that Mistral performs better in identifying quantitative information in scientific text. In terms of the investigated impact classes, we found that both models tend to perform better for infrastructure classes.

Model	Accuracy	Recall	F1-score
Document selection	0.96	0.92	0.93
Sentence classification	0.92	0.93	0.91

Table 1: Classification model performance.

Measure	Model	Similarity	Precision
Affected	GPT-4	-	0.85
	Mistral	-	0.91
Deaths	GPT-4	-	0.86
	Mistral	-	0.96
Society	GPT-4	0.34 ±0.11	-
	Mistral	0.36 ±0.09	-
Water	GPT-4	0.57 ±0.19	-
	Mistral	0.55 ±0.20	-
Households	GPT-4	-	1
	Mistral	-	1
Health infra.	GPT-4	-	1
	Mistral	-	1

Table 2: Performance of information extraction using RAG. Society and Water indicate qualitative impacts on society and water availability. Households and Health infra. denote the affected infrastructure numbers. Similarity measures are shown as mean ± standard deviation.

4 Conclusion

In this paper, we presented an NLP pipeline to automatically extract and systematize information about the impacts of climatic, meteorological, and hydrological hazards from scientific text. Our approach has the added value of including both direct and indirect impact classes, which are often ignored in existing impact databases. Results showed that in order to achieve high accuracy in data extraction, it is necessary to combine several strategies, including transfer learning, fine-tuning pre-trained models, and retrieval augmentation processes. Mistral and GPT-4 show similar performances in RAG, which indicates that open-source models are suitable for such tasks. By adding the sentence classification as a pre-retrieval step, we reduced context length and the need for text chunking.

Ethics Statement

All text data used in this study is open access and freely available. We included exclusively articles published under CC-BY license. LLMs were used only in excerpts of the full text selected by a classification model.

References

Josh Achiam, Steven Adler, Sandhini Agarwal, Lama Ahmad, Ilge Akkaya, Florencia Leoni Aleman, Diogo Almeida, Janko Altenschmidt, Sam Altman, Shyamal Anadkat, et al. 2023. Gpt-4 technical report. *arXiv preprint arXiv:2303.08774*.

Muhammad Siddique Akbar and Daniel P Aldrich. 2017. Determinants of post-flood social and institutional trust among disaster victims. *Journal of Contingencies and Crisis Management*, 25(4):279–288.

Global Administrative Areas. 2012. Gadm database of global administrative areas, version 2.0. http://www.gadm.org. Accessed: 2024-05-22.

Iz Beltagy, Kyle Lo, and Arman Cohan. 2019. SciBERT: A pretrained language model for scientific text. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 3615–3620, Hong Kong, China. Association for Computational Linguistics.

Max Callaghan, Carl-Friedrich Schleussner, Shruti Nath, Quentin Lejeune, Thomas R Knutson, Markus Reichstein, Gerrit Hansen, Emily Theokritoff, Marina Andrijevic, Robert J Brecha, et al. 2021. Machine-learning-based evidence and attribution mapping of 100,000 climate impact studies. *Nature climate change*, 11(11):966–972.

Paolo Cianconi, Sophia Betrò, and Luigi Janiri. 2020. The impact of climate change on mental health: a systematic descriptive review. *Frontiers in psychiatry*, 11:490206.

Harold C Cochrane. 2004. Indirect losses from natural disasters: measurement and myth. In *Modeling spatial and economic impacts of disasters*, pages 37–52. Springer.

Mariana Madruga de Brito, Jan Sodoge, Alexander Fekete, Michael Hagenlocher, Elco Koks, Christian Kuhlicke, Gabriele Messori, Marleen de Ruiter, Pia-Johanna Schweizer, and Philip J Ward. 2024. Uncovering the dynamics of multi-sector impacts of hydrological extremes: A methods overview. *Earth's Future*, 12(1):e2023EF003906.

Damien Delforge, Valentin Wathelet, Regina Below, Cinzia Lanfredi Sofial, Margo Tonneliere, Joris van Loenhout, and Niko Speybroeck. 2023. EM-DAT: The emergency events database.

Yunfan Gao, Yun Xiong, Xinyu Gao, Kangxiang Jia, Jinliu Pan, Yuxi Bi, Yi Dai, Jiawei Sun, Meng Wang, and Haofen Wang. 2024. Retrieval-augmented generation for large language models: A survey. Version Number: 5.

Global Runoff Data Centre, GRDC. 2020. *WMO Basins and Sub-Basins*, 3rd, rev. ext. ed. edition. Federal Institute of Hydrology (BfG), Koblenz, Germany.

Michael Hagenlocher, Gustavo Naumann, Isabel Meza, Veit Blauhut, Davide Cotti, Petra Döll, Katrin Ehlert, Franziska Gaupp, AF Van Loon, JA Marengo, et al. 2023. Tackling growing drought risks—the need for a systemic perspective. *Earth's Future*, 11(9):e2023EF003857.

Sanaa Hobeichi, Gab Abramowitz, Jason P Evans, and Anna Ukkola. 2022. Toward a robust, impact-based, predictive drought metric. *Water Resources Research*, 58(2):e2021WR031829.

Matthew Honnibal, Ines Montani, Sofie Van Landeghem, and Adriane Boyd. 2020. spacy: Industrial-strength natural language processing in python.

Albert Q Jiang, Alexandre Sablayrolles, Arthur Mensch, Chris Bamford, Devendra Singh Chaplot, Diego de las Casas, Florian Bressand, Gianna Lengyel, Guillaume Lample, Lucile Saulnier, et al. 2023. Mistral 7b. *arXiv preprint arXiv:2310.06825*.

Rebecca Louise Jones, Debarati Guha-Sapir, and Sandy Tubeuf. 2022. Human and economic impacts of natural disasters: can we trust the global data? *Scientific data*, 9(1):572.

Rebecca Louise Jones, Aditi Kharb, and Sandy Tubeuf. 2023. The untold story of missing data in disaster research: a systematic review of the empirical literature utilising the emergency events database (em-dat). *Environmental Research Letters*.

Jan-Christoph Klie, Michael Bugert, Beto Boullosa, Richard Eckart de Castilho, and Iryna Gurevych. 2018. The INCEpTION platform: Machine-assisted and knowledge-oriented interactive annotation. In *Proceedings of the 27th International Conference on Computational Linguistics: System Demonstrations*, pages 5–9, Santa Fe, New Mexico. Association for Computational Linguistics.

Tamiyo Kondo and Gonzalo Lizarralde. 2021. Maladaptation, fragmentation, and other secondary effects of centralized post-disaster urban planning: The case of the 2011 “cascading” disaster in japan. 58:102219.

Heidi Kreibich, Giuliano Di Baldassarre, Sergiy Vorogushyn, Jeroen CJH Aerts, Heiko Apel, Giuseppe T Aronica, Karsten Arnbjerg-Nielsen, Laurens M Bouwer, Philip Bubeck, Tommaso Caloiero, et al. 2017. Adaptation to flood risk: Results of international paired flood event studies. *Earth's Future*, 5(10):953–965.

Bernhard Lehner and Günther Grill. 2013. Global river hydrography and network routing: baseline data and new approaches to study the world's large river systems. 27(15):2171–2186.

Miguel D Mahecha, Debarati Guha-Sapir, Jeroen Smits, Fabian Gans, and Guido Kraemer. 2020. Data challenges limit our global understanding of humanitarian disasters triggered by climate extremes. In *Climate Extremes and Their Implications for Impact and Risk Assessment*, pages 243–256. Elsevier.

Celia McMichael. 2020. Human mobility, climate change, and health: Unpacking the connections. *The Lancet Planetary Health*, 4(6):e217–e218.

OpenStreetMap. 2017. Tmc/location code list/location types. https://wiki.openstreetmap.org/wiki/TMC/Location_Code_List/Location_Types. Accessed: 2024-05-22.

World Health Organization. 2004. *International Statistical Classification of Diseases and related health problems: Alphabetical index*, volume 3. World Health Organization.

Munich Re. 2023. Natural disasters in 2022, munich re natcatservice, münchen, germany.

Timo Schick, Helmut Schmid, and Hinrich Schütze. Automatically identifying words that can serve as labels for few-shot text classification. Version Number: 1.

P. R. Shukla, J. Skea, R. Slade, A. Al Khourdajie, R. van Diemen, D. McCollum, M. Pathak, S. Some, P. Vyas, R. Fradera, M. Belkacemi, A. Hasija, G. Lisboa, S. Luz, and J. Malley, editors. 2022. *Climate Change 2022: Mitigation of Climate Change. Contribution of Working Group III to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press, Cambridge, UK and New York, NY, USA.

Jan Sodoge, Christian Kuhlicke, and Mariana Madruga de Brito. 2023. Automatized spatio-temporal detection of drought impacts from newspaper articles using natural language processing and machine learning. *Weather and Climate Extremes*, 41:100574.

Hiek Som. 2010. World programme for the census of agriculture 2010. *Rome, Italy: FAO Statistics Division*.

Lina Stein, S Karthik Mukkavilli, Birgit M Pfitzmann, Peter WJ Staar, Ugur Ozturk, Cesar Berrospi, Thomas Brunschwiler, and Thorsten Wagener. 2024. Wealth over woe: global biases in hydro-hazard research.

UNISDR. n.d. Desinventar. https://www.desinventar.net. United Nations Office for Disaster Risk Reduction.

Samantha C Winter, Mark R Winter, LaNae Plaxico, Anna K Balakrishnan, Millicent Dzombo, Loni Philip Tabb, Ebuka Ukoh, Chloe Lincoln, Lena Moraa Obara, Stephanie Achieng Otieno, et al. 2024. Extreme weather should be defined according to impacts on climate-vulnerable communities. *Nature Climate Change*, pages 1–6.

A Schema example

```
hazardType: Hurricane,
location: {country: {countryName: ”Nicaragua”, countryCode: ”NIC”},
      state: NA,
      city: NA,
    waterBasin: {basinName: NA, basinCode: NA},
      latitude: 12.865416,
      longitude: -85.207229},
time: {startDate: NA, endDate: NA},
affected: {200000},
humanDeaths: {2000},
. . .
healthImpacts: {qualitative: {NA},
quantitative: {{disease: ”trauma”, icdCode: ”F38”,
value: 60.1, unit: ”percentage increase”,
location: {country: {countryName: ”Nicaragua”, countryCode: ”NIC”},
state: NA, city: ”Quezalguaque”,
latitude: 12.50683000, longitude: -86.90292000},
annotation: “When comparing the different villages included in the study. . . ”}
}},
healthImpacts: {qualitative: {NA},
quantitative: {{disease: ”trauma”, icdCode: ”F38”,
value: 60.1, unit: ”percentage increase”,
location: {country: {countryName: ”Nicaragua”, countryCode: ”NIC”},
state: NA, city: ”Quezalguaque”,
latitude: 12.50683000, longitude: -86.90292000},
annotation: “When comparing the different villages included in the study. . . ”}
}},
source: {title: ”Psychological impact of the hurricane Mitch in Nicaragua
in a one-year perspective”,
doi: 10.1007/s001270050298,
journal: ”Social Psychiatry and Psychiatric Epidemiology”,
yearPublication: 2001,
Source: ”PubMed”}
```

B Query search terms

Hazard	Query
General terms	multi-hazard OR “several hazards” OR “compound hazard!”
Drought	drought! OR dry spell!
Flood	flood! OR inundation! OR Glacial lake outburst
Storm	storm! OR superstorm! OR wind?storm! OR snow?storm! OR blizzard! OR derecho OR winter?storm! OR hail OR extra?tropical?storm OR thunderstorm! OR tornado! OR tropical?cyclone OR storm surge! OR hurricane! OR typhoon
Heatwave	heat?wave OR heat episode! OR ((heat OR hot) AND spell!) OR heat?stress
Coldwave	cold?wave! OR severe winter conditions OR cold spell
Mass movement	landslide! OR rock?fall OR mudslide OR mass movement
Wildfire	forest?fire! OR wild?fire! OR land?fire OR bush?fire

Table 3: Query search terms used to search for papers reporting meteorological, hydrological and climatic hazards published from 1990 to 2024. ! = any subsequent letters; ? = any letter or space to replace the question mark.

Subclass	**Query**
Social	
-	violen! OR crime! OR war! OR conflict! OR dispute! OR livelihood OR unemploy! OR poverty OR income
Water availability	
Physical	“water scarcity” OR “water supply” OR “water availability” OR “lack of water” OR “hydrological stress” OR “drinking water”
Water quality	(water AND (chlorophyll OR nitrogen OR phosphorus OR quality OR pollution OR heavy metal! OR pesticide!)) OR algae?bloom
Agriculture	
Crop production	(food AND (security OR supply OR food production)) OR famine
Animal and livestock health and productivity	livestock OR cattle! OR (animal AND (well-being OR husbandry OR welfare OR nutrition))
Fisheries yields	fishery! OR aquaculture OR fish stock
Health	
General terms	health! OR well?being OR ill OR illness OR disease! OR syndrome! OR infect! OR medical! OR disabilit!
Fatalities	death! OR fatalit! OR died OR casualties OR “loss of life”
Physical injuries	injur!
Infectious diseases	infectious disease! OR cholera OR giardiasis OR cryptosporidiosis OR leptospirosis
Nutrition and obesity	obes! OR over?weight OR under?weight OR hunger OR stunting OR wasting OR undernourish! OR undernutrition OR anthropometr! OR malnutrition OR malnour! OR anemia OR anaemia OR ""micro?nutrient!"" OR diabet!
Mental health	mental OR depress! OR !stress! OR anxi! OR ptsd OR psycho! OR psychiatric! OR !trauma! OR post-traumatic OR suicide! OR solastalgi!
Air pollution	“air quality” OR “air pollution” OR PM2.5 OR “fine particulate” OR asthma
Displacement	(displacem! OR relocation! OR migration OR refugee! OR homeless! OR emergency shelter)
Infrastructure	
General terms	infrastructure!
Transportation	bridge! OR road! OR highway! OR train! OR transport! OR rail! OR ship OR mobility
Sanitation	((water OR waste?water) AND treatment plant!) OR sewage! OR sewer! OR sewerage! OR waste OR landfill
Health infrastructure	hospital! OR care clinic! OR emergenc! OR pharmac!
IT and communication	digital infrastructure OR communication infrastructure OR ((mobile OR !phone OR internet) AND (network! OR system!))
Energy	energy OR electricity OR heating OR gas supply OR biogas OR ((wind OR hydro OR nuclear OR coal OR thermal) AND power)
Housing	propert! OR house! OR building! OR infrastructure!
Economic sectors	
Industry and commerce	(macroeconomic AND loss) OR economic assets OR capital OR companies OR business! OR industr! OR commerce
Agriculture	crop! losses OR crop yield! OR crop quality OR crop failure OR yield loss! OR agriculture
Forestry	forest dieback OR forest damage OR tree vitality OR tree growth OR tree dieback OR forestry OR die?off
Tourism	tourism OR tourist! OR hotel! OR museum! OR culture OR cultural OR recreation!

Table 4: Query terms used to search for papers reporting impacts of meteorological, hydrological and climate hazards published from 1990 to 2024. ! = any subsequent letters; ? = any letter or space to replace the question mark.

C Sentence labelling

Class	Criteria
Background	Sentence describing knowledge domain
Prior work	Description of previous work related to the study, presentation of results from previous papers that can be important to the study
Objective	Aims and goals of the study
Motivation	Motivation for performing the study
Method	Details of the research procedure
Data	Description of the dataset (source, size, type)
Equation	Mathematical equations
Impact	Social, economic, health, infrastructure, and environmental impacts of the natural hazard described in the study
Response	Response and adaptation measures to the impacts of the natural hazard
Results	Sentences describing study findings, consequences, and analysis of the results
Conclusion	Sentences on the support or rejection of hypotheses and suggestions for future work
Recommendation	Recommendations for extending the findings or method described in the paper, implications for policy making, stakeholders, and decision-makers
Metadata	Author's names, affiliations, funding agencies
Hazard cause	Climate and social causes of the hazard described in the paper
Hazard description	General description of the hazard, including date and location
Cascading hazard	Cascading hazard associated with the main hazard described in the study
Cascading impact	Cascading impact associated with the hazard described in the study

Table 5: Detailed sentence annotation characteristics.

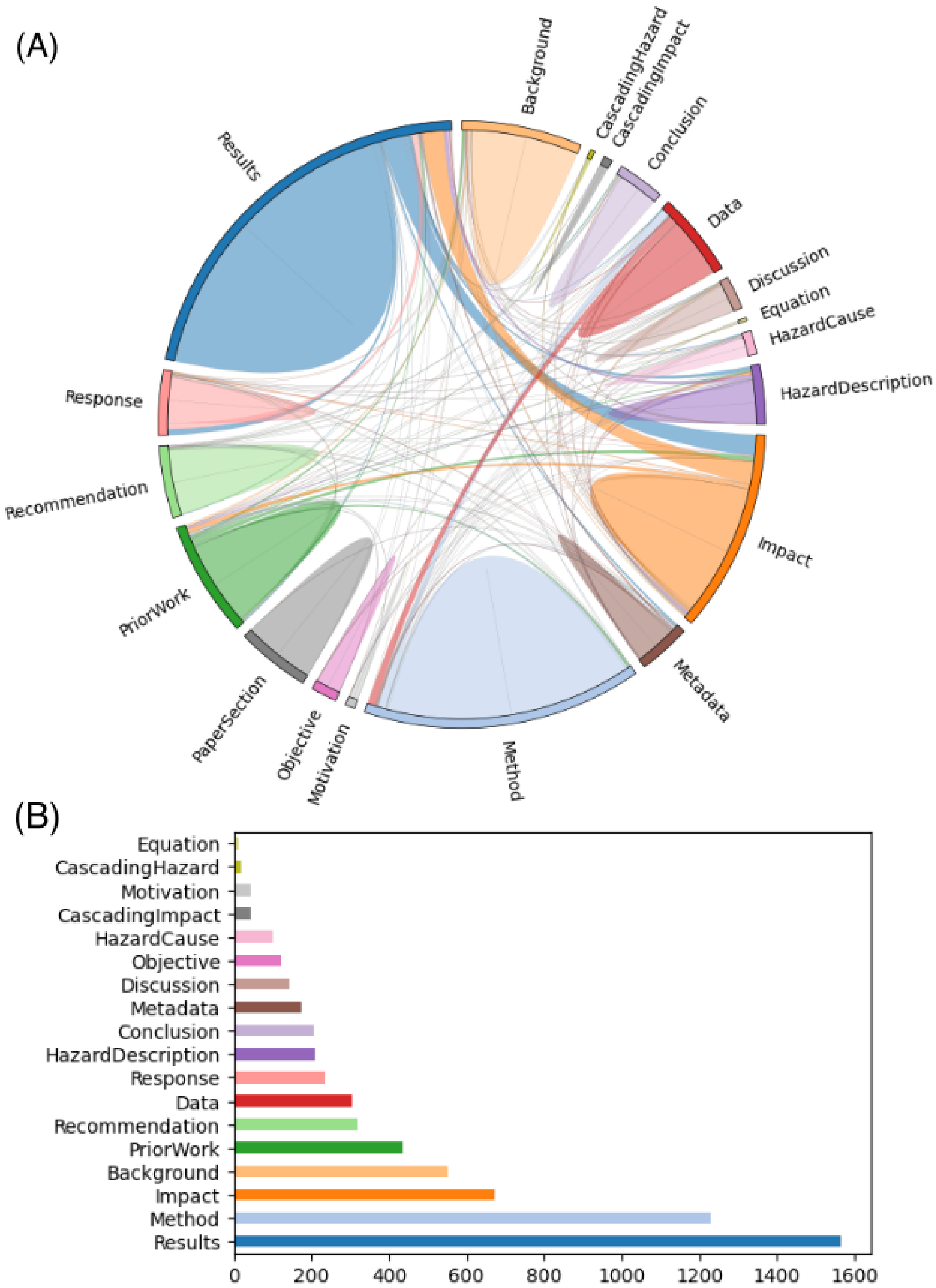

Figure 2: Characteristics of sentence annotations in scientific papers of the natural hazards research domain. (A)demonstrates sentence co-occurrences and (B) shows the frequencies of sentences.