Wikimpacts 1.0: A new global climate impact database based on automated information extraction from Wikipedia

Response to Reviewer 1

January 6, 2026

Contents

1 Reviewer 1

We would like to thank Reviewer #1 for their dedicated time in reviewing the manuscript and for their useful and constructive suggestions. We propose to implement the following changes in the database and manuscript:

1. Filter the multi-event article entries in Wikimpacts 1.0 and fix the geo-parsing issues identified.

2. Update the figures along with the updated database.

3. Add explanation for the location definitions across different levels and interpretation of error scores.

4. Improve the Discussion section.

Many suggestions point to potential additional research, which incited us to rework the manuscript, and which indicate that our study could ignite and inspire new research. In addition to addressing the Reviewers' feedback, we revised the manuscript such that it now complies with the Policy Forum article format.

Below, we would like to clarify our changes regarding all comments, which are repeated in grey boxes. This response letter contains numbered figures and references to these illustrations. Moreover, the following convention is applied to denote modifications that we propose to implement in the original manuscript: new text.

Comment 1

General assessment
The topic is timely and relevant. A global, open, LLM-based impact database is of clear interest to NHESS readers. The manuscript reads as a data-and-methods paper describing a new dataset, its structure, extraction pipeline, evaluation, and example applications. This fits well within the journal's scope for data-oriented contributions. I recommend major revisions to improve clarity, strengthen the discussion of limitations, and help users understand how to interpret and apply the dataset.

Response

We thank Reviewer #1 for the positive assessment of the study topic and for the thoughtful and constructive review. Below, we carefully address each comment (with some comments addressed jointly) and describe the corresponding changes made to the manuscript.

Comment 2

Regarding event article mapping and potential misclassification.
While inspecting the public database, I noticed several cases where impacts appear to be drawn from broad multi-event Wikipedia pages (e.g. "Tropical cyclones in 2017") even when a dedicated single-event article exists (e.g. "Cyclone Numa"). This can lead (it does actually) to incorrect country lists, the inclusion of non-impacted areas, or duplicate entries for the same event. The current manuscript does not fully explain how cross-references within multi-event articles are handled (for example, when one system contributes to another) nor how potential double-counting or mis-allocation of impacts is prevented. I recommend adding a subsection clarifying the filtering logic, giving examples of typical failure modes, and explaining whether any automatic or rule-based deduplication is applied during consolidation.
Specific comments on the article
Regarding the example of the 2011 European floods, the manuscript notes that the main event was categorised as an extratropical cyclone, although a flood category may be more appropriate. This is a useful illustration of hazard-type ambiguity. It would help if the authors commented briefly on how common such cases are and whether simple rule-based corrections might reduce them.

Response

We agree with the reviewer that there is potential misclassification of event article mapping and event types in the database. During the document selection process, we first distinguished between single-event and multi-event articles. For each event, we selected the best available article among those classified as describing it. However, as the reviewer correctly illustrated, some articles still cover multiple events, such as "Tropical cyclones in 2017". In the updated database, we therefore filtered out 195 multi-event article entries on tropical storms/cyclones from Wikimpacts 1.0. For other major event types, such as floods and wildfires, we did not remove articles with plural titles (e.g., 2021 European floods", 2015 Russian wildfires"), because these articles are not cross-linked to separate entries for individual flood or wildfire events. Notably, in Wikimpacts 1.0, we do not process tables or list items within articles; consequently, for articles such as "Tulsa tornadoes of 2017", information contained solely in tables is currently not included in our database. This procedure ensures that the Wikimpacts 1.0 database contains only single-event articles and avoids duplicated information arising from multi-event articles. Although we consider the likelihood of duplicated information within single-event articles to be relatively low, we will nevertheless address and resolve these potential issues in the next version of the database. We address this update in Section 3.1, L170: Moreover, within the 3,368 event articles, we identify 195 multi-event article entries on tropical storm/cyclone through the keyword search. We further identify a number of multi-event articles that remained in the database. To detect these, we perform a keyword search using the terms ["list", "season", "cyclones", "hurricanes", "typhoons", "tornadoes"] to identify potentially misclassified multi-event articles in the Wikimpacts 1.0 database. This procedure yields 191 entries containing the keyword "season" (e.g., 1939 Pacific hurricane season) and 4 entries containing the keyword "cyclones" (e.g., Tropical cyclones in 2013). For other keywords, such as "tornadoes", we find articles like "July 2009 Mid-Atlantic tornadoes" that describe multiple tornadoes but for which no separate

Wikipedia articles exist for the individual events. For the remaining keywords, we do not identify any corresponding articles in the Wikimpacts 1.0 database. For other major event types, such as floods and wildfires, we do not remove articles with plural titles (e.g., 2021 European floods, 2015 Russian wildfires), because these articles are not cross-linked to distinct entries for individual flood or wildfire events. It is important to note that Wikimpacts 1.0 does not process tables or list items within articles. Consequently, for articles such as "Tulsa tornadoes of 2017", information that appears only in tables is currently excluded from the database. This filtering procedure ensures that the Wikimpacts 1.0 database is restricted to single-event articles and mitigates duplication arising from multi-event entries. The risk of duplicated information within single-event articles is expected to be relatively low, and we plan to verify this more systematically in future updates of the database. In total, we identify 195 misclassified multi-event articles and remove these entries from the database.

In response to the comment regarding the misclassification of the main event type, we found that 27 articles with titles containing the word "flood", such as "2011 European floods", were classified as extratropical storm/cyclone events. For example, the article "2011 European floods" states, "The 2011 floods in Europe were caused by a series of storms in the fall, including Cyclone Meeno and Tropical Storm Rolf." Consequently, our model considers this event as an extratropical storm/cyclone. In our database, the hazard column lists "flood" for all of these 27 events, except for the article "1999 Blayais Nuclear Power Plant flood". We acknowledge that this ambiguous classification is not an error or a limitation of our model; even human experts find it difficult to classify such events decisively. Therefore, we decided not to apply a keyword- or rule-based approach to modify the database for these events, and we have retained the original classification. Instead, we have revised the text in Section 5.1, lines 417-421:

However, a few entries are present for tropical Africa. Further investigation reveals that the latter may be an ambiguous classification between the main event categories of flood and extratropical cyclone. For example, the event "2011 European floods" - despite the article's title â also impacted North Africa. The floods were caused by a series of storms, and in our database this main event is categorised as extratropical cyclone. We identify 27 such cases in our database, notably, the hazard column lists "flood" for all of these 27 events, except for the article "1999 Blayais Nuclear Power Plant flood". We acknowledge that this kind of unavoidable ambiguous classification is not an error or a limitation of our model.

Comment 3

Specific comments on the article
Regarding the role of Wikimpacts 1.0 relative to existing datasets, it would help to clarify early in the Introduction whether the authors position this work as complementary to curated impact databases such as EM-DAT and DesInventar, or as an alternative source. It would also be useful to specify for which types of analyses the dataset is most appropriate (for example, global multi-hazard comparisons or exploratory sub-national studies) and which applications require caution (such as completeness-sensitive national loss or other variable accounting).
Coverage and Wikipedia reliance: the requirement that an English Wikipedia article must exist implies notability and language biases. Small-scale or local events, or those in regions with limited Wikipedia activity, are surely under-represented. I suggest adding a short paragraph addressing this and explaining how users should interpret the absence of events. This also helps contextualise the patterns shown in Fig. 4.

Response

We position Wikimpacts as a complementary impact database to EM-DAT and DesInventar, with its multi-level georeferencing as a clear distinctive feature. However, the inclusion criteria and event definitions differ across these databases. As a result, not all events can be directly matched between Wikimpacts and EM-DAT/DesInventar, although a subset of events can be aligned and compared, as illustrated in Fig.10 of the preprint. Wikimpacts is therefore particularly useful as a complementary source of fine-scale impact information, for cross-database impact comparisons, and for exploring sub-national impact and damage functions. For comprehensive national loss estimates, however, the impact information extracted from Wikimpacts should be validated against nationally reported data or other established impact databases.

Similarly, to address the bias associated with English-language Wikipedia, we have added the following to the Introduction from L63:

The Wikimpacts database is designed to complement existing global disaster impact databases such as EM-DAT and DesInventar, notably by providing georeferenced information at sub-national level. Although the inclusion criteria and event definitions differ across databases, matching events by type, date, and location allows us to identify a set of shared events that can be jointly analysed. These overlapping records can support global multi-hazard comparisons and sub-national studies, particularly when combined with climate and other geospatial data. We find relatively few climate event articles in other languages that are not reported in English. Although we rely solely on English Wikipedia articles in the Wikimpacts 1.0 database, the English bias may exist, but it is not the main issue. We acknowledge that regions with limited Wikipedia activity and small or highly localized events are likely to be under-reported. Consequently, the Wikimpacts 1.0 database should not be used in isolation for complete or highly sensitive national loss assessments. For such applications, we recommend benchmarking Wikimpacts against other databases such as EM-DAT and officially reported government statistics.

Comment 4

Specific comments on the article
L1-L3 definitions (Sect. 2 and Sect. 3.3.3): the manuscript would benefit from a clearer and more consistent description of what "location" means at each level. At L80, levels are defined as event (L1), country (L2), and sub-national (L3). Later wording in Sect. 3.3.3 is less precise. Restating the definitions once, with consistent terminology, will help users interpret the later evaluation of location accuracy.

Response

The "location" definition differs across 3 levels; therefore, we would prefer to keep the technical definition table in the main text, rather than moving it to the Supplementary Information, as Reviewer #1 suggested in comment 14. Table 1 in the preprint illustrates the definitions of each location-related item in the database and corresponding levels. To make the user better interpret the definition of location-related information, we add more explanation in Section 2, without additional information in Section 3.3.3.

From L83 in Section 2 (Database Structure), we add The location-related information is summarized in Table 1. For L1 and L2, the locations are specified in Administrative_Areas_Norm, which contains a list of affected countries. For L3, the location information comprises both Administrative_Areas_Norm and Locations_Norm, indicating one affected country and a list of affected sub-locations such as cities. In the remainder of the text, we use the general term "location" or "locations" to refer to both these fields.

Comment 5

Specific comments on the article
Hazard types (Sect. 3.4): the exclusion of events that cannot be mapped to the seven main hazard categories, such as landslides, should be made explicit as a limitation. It would help readers to understand whether landslide impacts are entirely lost or whether some are absorbed under parent storm or flood events.

Response

To address this concern, we have added the following sentence at L258 in Section 3.4: Events which cannot be mapped to any of the main event types are not directly reported in the database. Nonetheless, they may be implicitly accounted for in the impact data, for example, if the reported impacts for a flood include the impacts of a landslide triggered by the flood.

Comment 6

Specific comments on the article
Phrase "extensive spatio-temporal coverage" (L58): this wording may overstate completeness given Wikipedia's known biases. Moderating this statement or directing the reader to the evaluation and limitations sections would avoid possible misinterpretation.

Response

We partially agree with the reviewer and recognize the limitation for extending the Wikimpacts database coverage within Wikipedia limits. However, we do not claim that extensive coverage will resolve the bias issue in the data. We have therefore revised the sentence in L58 to: extensive spatio-temporal coverage (within Wikipedia limits)

Comment 7

Specific comments on the article
Temporal coverage: the database extends from 1034 to 2024. The sharp rise in event counts after the 19th century makes it clear that older entries are sparse. Adding one sentence advising users how to interpret pre-1900 records (highly incomplete, not suitable for quantitative trend analysis) would improve clarity.

Response We thank the reviewer for this helpful suggestion and have added the following sentence in L386 in Section 5.1.1: Records before 1900 are sparse due to limited reporting, and are therefore not suitable for quantitative trend analyses.

Further, we also acknowledge the value of including more pre-1900 records, and have added the following sentence to Section 6.3 (Limitations), L575: Moreover, the pre-1900 records in Wikimpacts 1.0 database are limited, and future versions of Wikimpacts database will aim to include available historical events prior to 1900.

Comment 8

Specific comments on the article
Evaluation (Sect. 4): Its structure and intent are valuable. It would help the reader if Sect. 3 indicated that the extraction quality is assessed in Sect. 4. Furthermore, Sect. 4 would benefit from a short explanation of how the 70 + 156 gold-standard events were sampled (randomly or stratified). This may affect the generalisability of the error rates (?).

Response

To address the first part of this comment, we added a brief paragraph at the end of Section 1 (Introduction) that outlines the structure of the paper as follows:

The remainder of this paper is organized as follows. Section 2, *Database Structure*, presents an overview of the database design and the technical definitions of all fields. Section 3, *Wikimpacts Processing Pipeline*, describes in detail the methodology used to construct the database, while Section 4, *Evaluation of the Pipeline*, reports the evaluation methods and results. Section 5, *Wikimpacts 1.0: Content of the Database*, presents the spatial and temporal distribution of the database content and compares it with EM-DAT. Section 6, *Discussion*, provides a detailed assessment of the pipeline and database, the comparison with EM-DAT, and the limitations of the database. Finally, Section 7, *Conclusion*, summarizes the main contributions of this work.

We address the second part of the comment jointly with the related comments from another reviewer concerning the annotation process, as follows in L279:

The gold standard events are randomly sampled from the full set of 5,046 classified climate events, with the constraint that their event-type distribution is representative of that of the entire database. Only single-event articles are annotated for processing in the Wikimpacts 1.0 database. The annotation was conducted over the course of one year by two postdoctoral researchers in climate science and two researchers with a master's degree in water engineering.

Comment 9

Specific comments on the article
Regarding the interpretation of field-specific error rates, Table 6 reveals that location has a much higher error rate than other fields. The manuscript would be strengthened by explaining what types of errors dominate (for example, administrative-level mismatches, NULL penalties, or coordinate issues) and by offering guidance on how users should interpret L2 and L3 location fields. A short paragraph identifying which extracted fields are robust for most applications and which require caution would be particularly helpful.
The sentence "Within L1, event and timing data are highly accurate, while location data is less robust," is not intuitive because L1 represents the aggregated event level. It would be useful to clarify what "location" means at L1 and why its accuracy is lower.

Response

We address this point jointly with the corresponding comments provided by other reviewers on this section. We have updated the interpretation of the error scores across different fields in the discussion section, reflecting on the evaluation and analysis of the results and providing targeted recommendations for potential application use cases. We prefer not to add further information to Section 4.3 (Evaluation Results), to make the quantitative evaluation easy to follow. We instead direct readers to Section 6.1 (Database Quality Assessment), where these aspects are discussed in detail. We also refer the reviewer to our reply to Comment 4, where we clarify the meaning of locations at each level.

We propose the following edits: L370 in Section 4.3: For an overview of the error score analysis, we refer the reader to Section 6.1. In Section 6.1, we implement the changes as follows: In L502, we add Overall, for L1, the error rate for location information is high, with a score of 0.48 in Administrative_Areas_Norm across 156 events. We find that 35 NULL penalties, each scoring 1 for the corresponding attribute, account for nearly 46% of the total error score. Similarly, NULL penalties in the L2 and L3 location information also lead to high error scores in these location-related fields. During the consolidation process, we filter out these entries. Moreover, we do not compare the results obtained after consolidation with the gold standard, because, as described in Section 3.4, the data are processed from L3 to L2 to L1, and the resulting processed data no longer match the originally annotated data.

Comment 10

Specific comments on the article
Comparison with EM-DAT: the manuscript would be improved by a more cautious framing of discrepancies. Differences may arise not only from extraction errors but also from different event definitions, thresholds, and loss components. I recommend explicitly positioning Wikimpacts as complementary to curated sources and offering guidance on how users might use them together.

Response We agree that differences in inclusion criteria, as well as in event and impact definitions, can contribute to discrepancies in the comparison. In line with the comments concerning the EM-DAT comparison raised by another reviewer, we have added the following text in L560 Section 6.2 to address this point:

Notably, the substantial differences between Wikimpacts and EM-DAT arise not only from extraction errors in our pipeline, but also from divergent event definitions, inclusion thresholds and criteria for impact data, differences in impact sources, and differences in the definitions and components of impact categories. Moreover, in this study we did not conduct a systematic comparison with other existing databases, such as DesInventar, but previous work shows large discrepancies between DesInventar and EM-DAT (Worou and Messori, 2025; Panwar and Sen, 2019). We therefore recommend using our database as a complementary resource to existing databases. Furthermore, when matching events across different databases, we suggest testing different matching algorithms, for example based on event names and by allowing temporal buffers in event dates.

Comment 11

Specific comments on the article
Fig. 4: The distribution of events confirms that Wikimpacts 1.0 reflects high-impact, media-visible disasters rather than a full record of climate extremes. Making this explicit in the Discussion would prevent users from interpreting the dataset as complete.

Response We partly agree with this statement, although we note that the diffuse contributor base to Wikipedia means that also events of local relevance that may not make it to the international news outlets may be reported. It is nonetheless true that Wikimpacts does not provide a complete record of climate extremes. Accordingly, we have added the following sentence to Section 5 L379 to clarify this point:

Notably, the Wikimpacts 1.0 database is constructed from events recorded in the English-language Wikipedia, and does not constitute an exhaustive record of all climate extremes.

Comment 12

Specific comments on the article
Discussion: The existing section is strong, but could more directly address certain limitations:

1. Notability and language biases linked to Wikipedia;
2. The deliberate exclusion of particular hazards (for example, landslides) and implications for multi-hazard studies;
3. Systematic weaknesses in LLM extraction for multi-country or compound hazards beyond aggregated error rates.

A short concluding paragraph offering explicit user guidance, indicating suitable and unsuitable use cases, would be valuable. For example, the database appears well-suited for global comparative studies, exploratory sub-national analyses where data exist, or cross-hazard synthesis, but less suitable for completeness-sensitive applications or studies focused on small, local events.

Response

Together with the comments from the other reviewers on this section, we have added the following text to Section 6.3.

1. In L573, we add We find that most of the reported events in other language editions of Wikipedia are also reported in English. While the English language bias exists in our database, it is not the main issue regarding the overall bias. We also note that including articles about climate events from other language Wikipedias, where there is no English reporting, would not completely address the location bias issue.

2. In L578, we add Moreover, certain hazards, such as landslides, are not included in our database.

3. In L579, we add Systematic errors introduced by the LLM-based extraction, such as misclassification of L2 and L3 information, may lead to duplicate entries in the database.

4. In L583, we add Nevertheless, the Wikimpacts 1.0 database is suitable for global impact-database benchmarking, exploratory sub-national impact assessment, and risk modeling in data-rich contexts (e.g., tropical storm events in the United States). However, it should be used with caution in applications that are highly sensitive to completeness or that focus on local or small-scale events.

Comment 13

Minor comments
Regarding language and clarity, a few specific examples illustrate areas where editing would improve readability:

1. The sentence at L32 about geolocating EM-DAT events is grammatically unclear and should be revised.
2. The phrase "administrative units at the same level can also be highly variable" (L35) needs clarification as to what variability matters for impact analysis.
3. The sentence beginning "Due to the categorization based on single hazards" (L40) would benefit from rephrasing for grammar and conceptual clarity.
4. The reference to DesInventar at L42 would be clearer if the specific shared limitations were briefly stated.
5. The last part of the Introduction (L55-70) blends methodological details that belong in Methods or Data Availability. Ending the Introduction with a clearer statement of objectives and contributions would strengthen the structure.

Response We adapt the text as follows based on the comments and suggestions above:

1. In L32, we rephrase the sentence to Researchers have attempted to geolocate disaster events from EM-DAT, yet the resulting data comes with limited temporal coverage (events from 1960 to 2018) and include some geoparsing errors (Worou and Messori, 2025; Lindersson and Messori, 2025). Therefore, mapping the aggregated impact from national level to subnational scales remains challenging (Rosvold and Buhaug, 2021; Delforge et al., 2025).

2. In L34 and L35, we rephrase the sentence to Moreover, the level of administrative divisions used in the latter database varies between countries, and administrative units at the same level can also be highly variable due to the differing numbers and resolutions across intra- or inter admin levels, which hinders calibration of damage functions in impact assessment studies (Prahl et al., 2016).

3. In L40, we rephrase the sentence to Because impact entries are categorized based on single hazards, thus, impacts from co-occurring or multi-hazard events may not be captured appropriately (Lee et al., 2024; Mithal et al., 2024).

4. In L42, we rephrase the sentence to Similar limitations, such as the reporting of impacts as single-hazard categories, also affect other global multi-hazard impact databases, such as DesInventar (UNISDR, n.d.).

5. For L55-L70, we prefer to retain the current text, which provides a brief introduction to the method and explains how it addresses the database limitations discussed earlier in this section. With respect to the explicit statement of objectives and contributions, we have added a dedicated paragraph in the Introduction section, as indicated in Comment 3.

Comment 14

Minor comments

1. In Section 2, the opening paragraph focuses on repository and accessibility information rather than internal structure. Moving that material to a Data Availability section would allow Section 2 to begin more directly with the conceptual design. Similarly, some technical field-definition details could move to Supplementary Information.
2. Regarding abbreviations, SI at L88 should be defined on first use.
3. Regarding the example referring to 2025 Wikipedia data (L89), the manuscript could clarify how information "as of 2025" was obtained when the mining cutoff appears to be 2024. A brief explanation would prevent confusion. Consider a relevant note in Figure 1.

Response

We have addressed these comments as follows:

1. We move the first paragraph in Section 2 to the Data Availability section. Besides, the technical definitions are maintained in this section as reasons indicated in Comment 4.
2. We define Supplementary Material (SI) on L88.
3. In L89, we rephrase the sentence to In relation to the 2021 European flood event, the information on fatalities recorded in Wikipedia has been updated as of 2024 and indicates 196 deaths in Germany, 39 in Belgium, 2 in Romania, and 1 each in Italy and Austria as of 2025 (information manually extracted on 20 September 2025).

References

Delforge, D., Wathelet, V., Below, R., Sofia, C. L., Tonnelier, M., van Loenhout, J. A., and Speybroeck, N.: EM-DAT: the Emergency Events Database, International Journal of Disaster Risk Reduction, 124, 105 509, ISSN 2212-4209, https://doi.org/https://doi.org/10.1016/j.ijdrr.2025.105509, https://www.sciencedirect.com/science/article/pii/S2212420925003334, 2025.

Lee, R., White, C. J., Adnan, M. S. G., Douglas, J., Mahecha, M. D., O'Loughlin, F. E., Patelli, E., Ramos, A. M., Roberts, M. J., Martius, O., Tubaldi, E., van den Hurk, B., Ward, P. J., and Zscheischler, J.: Reclassifying historical disasters: From single to multi-hazards, Science of the Total Environment, 912, 169 120, 2024.

Lindersson, S. and Messori, G.: SHEDIS-Temperature: linking temperature-related disaster impacts to subnational data on meteorology and human exposure, Earth System Science Data, 17, 6379–6403, 2025.

Mithal, V., Sillmann, J., and Zscheischler, J.: Linking regional economic impacts of temperature-related disasters to underlying climatic hazards, Environmental Research Letters, 19, 124 010, 2024.

Panwar, V. and Sen, S.: Economic impact of natural disasters: An empirical re-examination, Margin: The Journal of Applied Economic Research, 13, 109–139, 2019.

Prahl, B. F., Rybski, D., Boettle, M., and Kropp, J. P.: Damage functions for climate-related hazards: unification and uncertainty analysis, Natural Hazards and Earth System Sciences, 16, 1189–1203, 2016.

Rosvold, E. L. and Buhaug, H.: GDIS, a global dataset of geocoded disaster locations, Scientific data, 8, 61, 2021.

UNISDR: DesInventar: United Nations Office for Disaster Risk Reduction, Retrieved in May 2024 from https://www.desinventar.net, n.d.

Worou, K. and Messori, G.: Compounding droughts and floods amplify socio-economic impacts, Environmental Research Letters, 20, 104 024, 2025.