Reply to the comment about the paper

Content Analysis of Multi-Annual Time Series of Flood-Related Twitter (X) Data

Open Review Process, NHESS

In preparing this response, we marked the original reviewers' comments in black while the authors' replies are in red. Line numbers refer to the original manuscript.

Referee #1:

R1.0: This review is concerned with the article titled "Content Analysis of Multi-Annual Time Series of Flood-Related Twitter (X) Data". It is divided into three categories, namely, general comments, specific comments and technical comments.

General comments: The title of the article "Content Analysis of Multi-Annual Time Series of Flood-Related Twitter (X) Data" clearly reflects the contents of the paper, and the abstract provides a concise, complete, and unambiguous summary of the work done and the results obtained. Both these sections are pertinent and easy to understand. The manuscript is well-written and well-structured, delivering the idea, methodology, and results clearly and concisely. The figures are descriptive and of high quality, and the tables are informative. It is well-referenced with proper credit attributed to previous and/or related works, and the authors indicate each of their contributions and competing interests. Crediting the use of AI tools such as ChatGPT is fantastic, we are conducting research in the age of the AI revolution. The paper presents a comprehensive and innovative approach to using social media data from Twitter (X) to understand human behaviour and perceptions during several types of flooding events in Germany. The study develops an approach using advanced natural language processing (NLP) techniques, leveraging pre-existing and accessible tools, including transformer-based models like SBERT and clustering algorithms such as HDBSCAN, to automatically extract flood-related topics from large social media datasets. Several steps to clean and filter the data have been presented. This allows for a nuanced analysis of public response to various flood events. The paper's relevance is clear, given the increasing reliance on real-time social media data for disaster risk management and the potential to enhance flood preparedness and response strategies. Thus, this manuscript has good scientific significance, scientific quality, and presentation quality.

We thank this Reviewer for their constructive evaluation of our work. The suggested amendments have been considered carefully and are addressed individually below.

Following are a few of the concerns that require clarification.

R1.1:

Specific Comments:

1. **Clarification on Data Filtering**: The process for removing irrelevant tweets is well-explained. However, more detail on the limitations of this filtering process could be helpful.
2. **Interpretation of Topic Groups**: The clustering approach is well-explained. However, further discussion on the specific implications of the topics identified (such as "disaster management" or "fatalities") could be more elaborated.
3. **Comparisons with Traditional Data Sources**: The paper highlights Twitter (X) data as an alternative to traditional flood impact assessments. What would be the difference between the results from social media and conventional data sources?

Clarification on Data Filtering:

This topic was also raised by another reviewer so we decided to add a new subsection in Section 2 titled "2.1 Data Collection", which is placed before the already existing section on “Data Preparation and Filtering”. The new section will include the following information to provide clarity on the data source and therefore also on the necessary filtering and associated uncertainties:

"The specifics of data collection can be found in de Bruijn et al. (2017) and de Bruijn et al. (2019). The following section describes the processing performed by de Bruijn et al. (2017) and de Bruijn et al. (2019) followed by an overview of the additional processing performed in this study, which is described in detail in Section 2.2. The full data was collected based on the former Twitter (X) API in eleven languages (Bruijn et al. 2019). The data collection and processing involve three main types of input data. First, the authors of de Bruijn et al. (2017, 2019) used a database of known geo-locations, which contains over 4 million geographical locations including cities, towns, villages, and administrative divisions, along with alternative names and translations. Second, they collected tweets and associated metadata in real-time through the Twitter (X) streaming API using flood-related keywords in eleven languages, gathering 55.1 million tweets between July 2014 and July 2017. The keywords included terms like "flood," "flooding," and "inundation" and their equivalents in other languages. Third, they utilized GIS shapefiles of global time zones and analyzed Wikipedia articles to obtain lists of the 1000 most commonly used words per language (excluding location names with populations over 100,000). The data processing involved matching tweet text to the gazetteer through toponym recognition, scoring candidate locations based on spatial indicators, grouping related tweets, and using a voting process for toponym resolution. The system processes tweets in 24-hour windows and maintains a toponym resolution table to enable real-time geoparsing of new incoming tweets. Relevance to flooding was further ensured by classification and pre-selection based on BERT.

Based on this data we additionally performed a combination of keyword and geolocation searches during the data pre-processing to obtain tweets related to flooding events in our study areas. We analyze a sample of Twitter (X) posts (n=43,287) collected from 2014 to 2022. Our sample includes all tweets posted during this time containing one or more of the three flood-related keywords (Hochwasser, Überflutung, Flut) written in German and geotagged within Germany. The table for all keywords in other languages is available at: https://www.nature.com/articles/s41597-019-0326-9/tables/2160 (de Bruijn et al., 2019).”

Based on your comment we will add the following sentence on the limitations introduced by the initial filtering steps in section 4 after Line 299:

"A limitation to the applicability of our model to different platforms and circumstances is the need for manual filtering and the associated uncertainties. The manual steps limit the transferability and may introduce a bias due to the individual variability of keyword selection. This limitation can be addressed by improved or combined embedding models (Laskar, 2020) or an embedding-based pre-selection"

Laskar, M. T. R., Huang, J. X., & Hoque, E. (2020, May). Contextualized Embeddings based Transformer Encoder for Sentence Similarity Modeling in Answer Selection Task. In N. Calzolari, F. Béchet, P. Blache, K. Choukri, C. Cieri, T. Declerck, … S. Piperidis (Eds.), Proceedings of the Twelfth Language Resources and Evaluation Conference (pp. 5505–5514). Retrieved from https://aclanthology.org/2020.lrec-1.676

Interpretation of Topic Groups:

We understand that the reviewer is referring to the lack of clarity and depth in topic interpretation.

We would like to point the reviewer to Table S2 and Table S1 in the Supporting Information. The two tables contain topic descriptions for all topics represented in Figure 1 and Figure 2 in the main manuscript. To make the contained information more accessible we will translate the two tables to English in the revised manuscript, as the top words and representative text are currently in German. The additional information in the Supplement contains a representative tweet as well as 10 keywords associated with the topic. We chose to represent the topics in the main text by only one keyword to improve readability and conciseness. While this probably leads to an oversimplification of topic representation, we tried to counteract this with a few instances of stating representative topic tweets. In the revised manuscript we will improve the topic interpretation by (i) adding the representative tweet in the text more often, (ii) extending the description of the topics when mentioned for the first time, and (iii) referring to the supplementary tables more frequently in the text where helpful.

Comparison with Traditional Data Sources:

To discuss the reviewer's question, we will add the following sentence in the Introduction after Line 22:

"Social media captures immediate personal experiences and emotional impacts that might be overlooked in conventional assessments, but lacks the standardized methodology and detailed technical measurements found in traditional sources. Therefore, analyses of social media data should not be seen, but as complementary analyses that enhance traditional flood impact assessments by providing rapid situational awareness and capturing the social dimensions of flood impacts that might otherwise go undocumented."

R1.3:

Technical Comments:

1. **Grammar and Style**:
   - Line 98: "The The Second" should be "The second."

2. **Figure Labels and Descriptions**: The figures provide valuable visual insights, but some (esp. Fig 3 & 4) would benefit from clearer labels or captions, particularly where technical details like clustering results or topic distributions are involved.
3. **In-text Citations Formatting**: Ensure that citations within the text follow a consistent format. There are some minor inconsistencies in how sources are referenced throughout the manuscript.

We Technical Comments:

1. Grammar and Style:
   We will correct the duplicate "The" on line 98 and conduct another thorough proofreading of the manuscript.
2. Figure Improvements:
   We will enhance Figures 3 and 4 by:

   Adding more detailed axis labels, including legend explanations in the caption, expanding captions to better explain technical elements (e.g., whether the represented results are an output of the clustering or other phases in our methodology).

3. Citation Formatting:
   We will review all citations to ensure consistent formatting throughout the manuscript according to the journal's guidelines.