Tracing online flood conversations across borders: A watershed level analysis of geo-social media topics during the 2021 European flood

Sébastien Dujardin[1,2*], Dorian Arifi[3,4*], Sebastian Schmidt[3,4], Catherine Linard[1,2], Bernd Resch[3,4,5]

[1] Department of Geography, University of Namur, Belgium
[2] ILEE - Institute of Life, Earth and Environment, University of Namur, Belgium
[3] Department of Geoinformatics – Z_GIS, University of Salzburg, Austria
[4] IT:U Interdisciplinary Transformation University, Linz, Austria
[5] Center for Geographic Analysis, Harvard University, USA
** Equally contributing authors*

Correspondence to: Sébastien Dujardin (sebastien.dujardin@unamur.be) and Dorian Arifi (dorian.arifi@plus.ac.at)

Abstract. In the face of rapid population growth, urbanisation, and accelerating climate change, ~~the need for~~ rapid and accurate disaster detection has become critical to minimising human and material losses. In this context, geo-social media data has proven to be a sensible data source for tracing disaster-related conversations, especially during flood events. However, current research often neglects the relationship between information from social media posts and their corresponding geographical context. In this paper, we examine the emergence of disaster-related social media topics in relation with hydrological and socio-environmental features on watershed level during the 2021 Western European flood, while focusing on transboundary river basins. Building upon an advanced machine learning-based topic modelling approach, we show the emergence of flood-related geo-social media topics both in river ~~-~~basin~~-~~-specific and cross-basin contexts. Our analysis reveals distinct spatio-temporal dynamics in the public discourse, showing that timely topics describing heavy rainfalls~~s~~ or flood damages were closely tied to immediate environmental conditions in upstream areas, while post-disaster topics about helping victims or volunteering were more prevalent in less affected areas located in both upstream and downstream areas. These findings highlight how social media responses to disasters differ spatially across watersheds and underscore the importance of integrating geo-social media analysis into disaster coordination efforts, opening new opportunities for transboundary collaborations and the coordination of emergency response along border-crossing rivers.

1. Introduction

Rapid climate change is altering precipitation patterns, leading to more intense and frequent climate-related disasters (IPCC, 2012, 2021). The increasing number and severity of flood events can be attributed to both climate and non-climate-related drivers (Clarke et al., 2022), including urbanisation in areas exposed to flood hazards (Ionita and Nagavciuc, 2021; UNISDR, 2015). In this context, effective flood risk management requires understanding how communities respond to flood events, especially in transboundary river basins, i.e. basins that cross political and administrative jurisdiction borders between countries, provinces, or cities (Rahayu et al., 2024). However, upstream and downstream areas are often managed by different governance structures and authorities, each with its own policies, priorities, and response frameworks, which ~~is making~~makes it difficult to align coordination efforts and resources across borders (Clegg et al. 2023). This paper contributes to addressing this issue by examining the role of social media data as a tool for capturing and analysing the responses of communities within transboundary river basins during flood events.

Recent advances in social media analytics offer new tools for monitoring and analysing public responses to disasters (Kryvasheyeu et al., 2016; Resch et al., 2018; Wang and Ye, 2018; Florath et al., 2024; Fohringer et al., 2015). Platforms such as Twitter or Weibo provide user-generated content that can be analysed to reveal public perceptions, behaviours, and sentiments during and after disasters (Beigi et al., 2016; Karmegam and Mappillairaju, 2020). Numerous studies have demonstrated the value of social media data for disaster management (Acikara et al., 2023; Yu et al., 2018), improving situational awareness (Yin et al., 2012), facilitating emergency response (Huang and Xiao, 2015), improving damage estimates (Zou et al., 2018), and even predicting the impacts of flooded areas (Bruneau et al., 2021). Specifically, the analysis of georeferenced social media posts (hereafter: geo-social media posts) enables the mapping of online information onto geographic spaces, making it particularly useful for early detection and damage classification in flood events (Tan and Schultz, 2021).

One of the most devastating recent flood events in transboundary river basins occurred in Western Europe during the summer of 2021, triggered by cyclone "Bernd", which brought long-lasting precipitation over the Eifel mountains due to orographic effects and dynamic uplift (Junghänel et al., 2021). This event caused severe flooding across Germany, Belgium, Luxembourg, France, and the Netherlands, resulting in over 200 fatalities and significant material damage (Kahle et al., 2022; Fekete and Sandholz, 2021; Schüttrumpf et al., 2022). Despite severe weather warnings, communication deficiencies hindered effective disaster response (Fekete and Sandholz, 2021). In this regard, effective, swift communication remains a challenge for situational awareness across borders. Geo-social media data could help address these issues, but few studies have examined how digital traces reflect the interconnectedness of upstream and downstream communities in transboundary river basins during flood events.

While several studies have applied social media and natural language processing (NLP) methods to analyse the 2021 European floods (Blomeier et al., 2024; Hanny and Resch, 2024; Moghadas et al., 2023), they primarily focused on specific regions, such as the Ahr Valley, and did not provide a comprehensive semantic analysis ~~on~~ of how online conversations emerged across borders and along the mainly impacted river basins. Moreover, most studies examining floods through social media data have focused on urban or regional scales (e.g. Wang et al. 2018, Tan and Shultz 2021)~~)~~, without considering the specificities of the river basins crossing these areas. A watershed approach is particularly relevant for transboundary flood risk management because it accounts for the interconnectedness of upstream and downstream communities, which often span across national boundaries (UNECE, 2009). Flooding in one part of a river basin can have complex spatial and temporal cause–effect relationships depending on both hydrographic (Gunnell et al., 2019) ~~(Gunnell et al. 2019)~~ and socio-environmental characteristics (Lorenz et al., 2001), which require a comprehensive view ~~over~~ of upstream and downstream response efforts for effective disaster management.

Despite the clear importance of this approach, there is limited evidence in the context of transboundary river basins, where international collaboration is often necessary but challenging due to differences in language, governance, and disaster management practices (Polese et al., 2024; Mehta and Warner, 2022). Although transboundary water management in European river basins, such as the Rhine, Danube, and Iberian rivers, has a long-established history (UNECE, 2009), effective cooperation among riparian countries still remains complex (Rahayu et al., 2024; Aall et al., 2023). Achieving successful collaboration therefore requires a deeper understanding of the natural and social processes driving these shared risks.

In this paper, we seek to identify geo-social media users' responses to heavy rainfall and subsequent flooding events in a transboundary river basin context. Specifically, our analysis aims to identify the emergence of different online topics throughout the flood event, with a particular focus on the identification of topics that dominate online conversations across upstream and downstream areas of river basins. To our knowledge, such a watershed-based analysis of up- and downstream differences in flood-related geo-social media topic emergence has not been considered in previous studies. Therefore, we aim at answering the following two research questions:

1. Which geo-social media topics can be observed before, during and after flood disasters in a transboundary river basin?
2. Which difference can be observed in the emergence of flood-related geo-social media topics across upstream and downstream areas within a river basin?

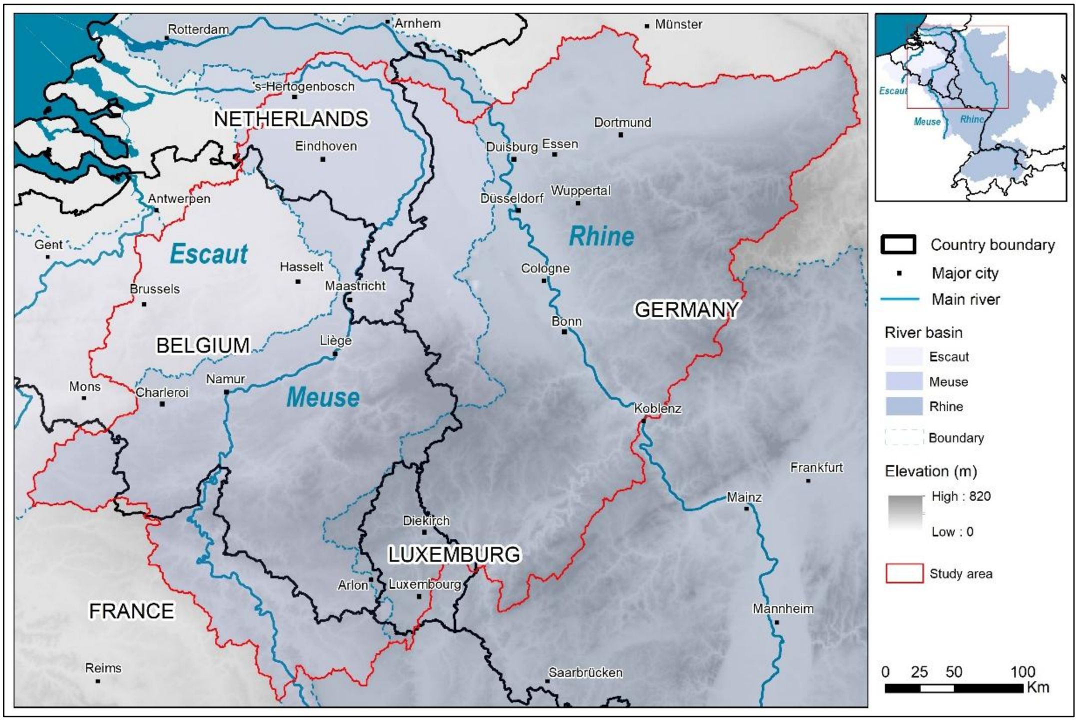
2. Data and method

The following sections describe the research area, data and methodological steps taken for our analysis in detail.

2.1. Research Area

Our study area includes regions in France, Belgium, the Netherlands, Germany, and Luxembourg (cf. Figure 1). Within these regions are the Ardennes and the Eifel, which comprise a low mountain range, incised by several fluvial valleys (Dietze et al., 2022) that are part of the catchment areas of the Lower Meuse and Lower Rhine rivers. The most important cities crossed by the Meuse river are Namur, Liège, Maastricht and Hertogenbosch. In Germany, the Rhine river flows through the main cities of Bonn, Cologne, and Düsseldorf. In the Northeastern part of this study area is the Escaut river which reaches the sea near the city of Antwerp.

The time frame of this study was ~~a three-week period~~three weeks from 11 to 31 July 2021, covering the precipitation peak on 14 July with one week prior and two weeks after. This timeframe was selected in line with the disaster phases commonly ~~recognized~~ recognised in the literature (e.g. Kruspe et al. 2021, Zou et al. 2018)~~,)~~. Also, it allowed to capture sufficient time both ~~prior~~ before and after the flood event to capture notable changes in precipitation patterns and online discussions.


Rotterdam
Arnhem
Münster
's-Hertogenbosch
NETHERLANDS
Eindhoven
Dortmund
Duisburg
Essen
Antwerpen
Dusseldorf
Wuppertal
Gent
Escaut
Rhine
Hasselt
Maastricht
Cologne
Brussels
GERMANY
Bonn
BELGIUM
Liège
Mons
Namur
Charleroi
Meuse
Koblenz
Frankfurt
Mainz
Diekirch
LUXEMBURG
Arlon
Luxembourg
FRANCE
Mannheim
Reims
Saarbrücken
Escaut
Meuse
Rhine
Country boundary
Major city
Main river
River basin
Escaut
Meuse
Rhine
Boundary
Elevation (m)
High : 820
Low : 0
Study area
0 25 50 100
Km

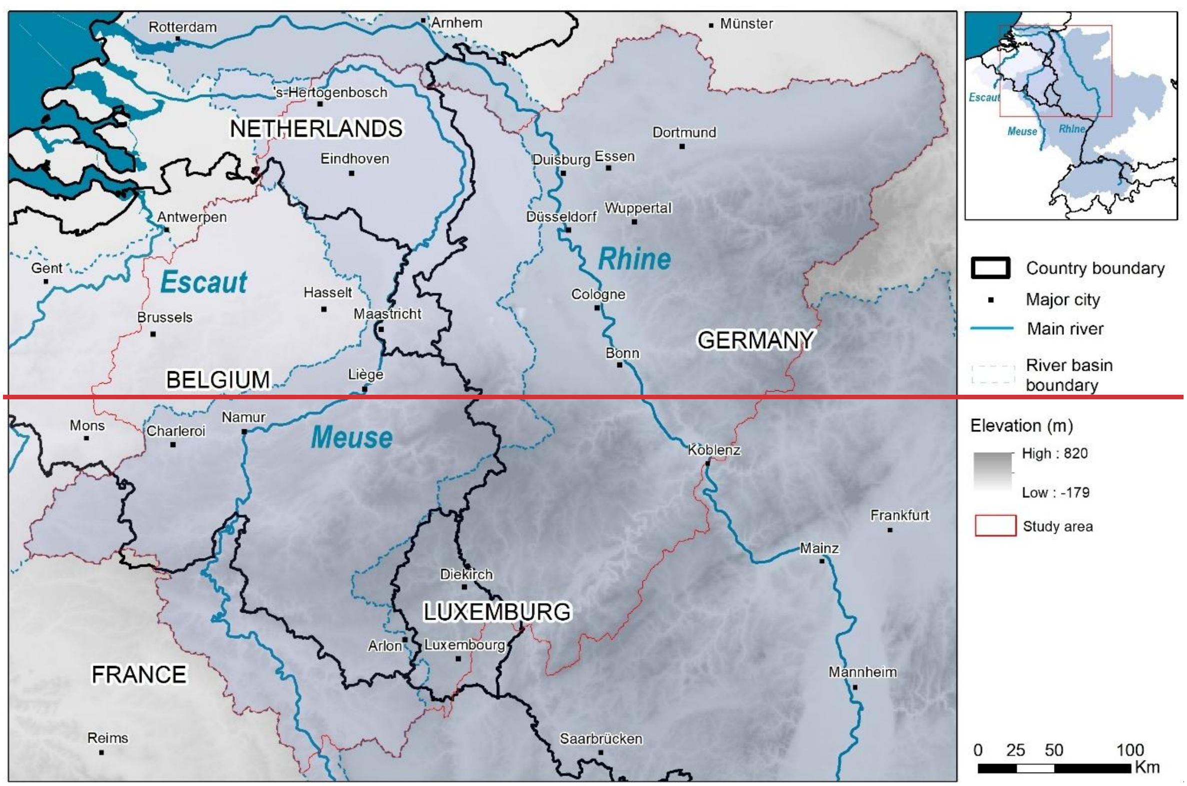

Rotterdam
Arnhem
Münster
's-Hertogenbosch
NETHERLANDS
Dortmund
Eindhoven
Duisburg
Essen
Antwerpen
Düsseldorf
Wuppertal
Gent
Escaut
Rhine
Hasselt
Cologne
Brussels
Maastricht
Bonn
GERMANY
BELGIUM
Liège
Mons
Charleroi
Namur
Meuse
Koblenz
Frankfurt
Mainz
Diekirch
LUXEMBURG
Arlon
Luxembourg
Mannheim
FRANCE
Reims
Saarbrücken
Escaut
Meuse
Rhine
Country boundary
Major city
Main river
River basin boundary
Elevation (m)
High : 820
Low : -179
Study area
0 25 50 100
Km

Figure 1 Study area map showing the main river streams (Escaut, Meuse, Rhine) and their corresponding transboundary river basins (shaded in blue). The main river sections examined in this study (in red) span five Western European countries: France, Belgium, the Netherlands, Germany, and Luxembourg.

2.2. Data

2.2.1. Precipitation and flood data

The first dataset selected for delineating the spatial extent of our study area was precipitation data generated using MAR (*Modèle Atmosphérique Régional*) (Wyard et al., 2021). MAR consists of simulated precipitation forced by ERA5 reanalyses, i.e. the fifth-generation atmospheric reanalysis of the global climate carried out by the European Centre for Medium-Range Weather Forecasts (ECMWF). Data was provided at 5km spatial resolution and 60 minutes temporal resolution, for the period June 14 to September 30, 2021. It contained total precipitation in mm and latitude-longitude variables in NetCDF format. This type of regional model allows for the downscaling of global models to finer temporal and spatial scales, providing reliable meteorological data for mapping summer rainfalls at the regional level (Doutreloup et al., 2022).

To further identify the most impacted watersheds, we used two complementary layers of information. The first layer contained the sections of the river network monitored by the European Flood Awareness System (EFAS) where six-hour averaged simulated river discharge exceeded the 20-year flood return period thresholds over the 11 to 31 July 2021. This information was produced by the Copernicus Emergency Mapping Service (EMS) model-derived river discharge and made readily available by the Copernicus EMS website (CEMS, 2021b).

A second layer of information was used to identify the ~~extend~~ extent of the flooded zones across our study area. It was retrieved from the Mapping Portal of the Copernicus EMS (Wania et al., 2021). Activated upon request from the German, Belgian, and Dutch authorities, the service provided mapping outputs (the EMSR517, EMSR518, and EMSR520 dataset) that contain remote sensing-based information regarding the flooding extent over these countries (CEMS, 2021a). We selected the vector packages of the flood delineation products across the period from 14 to 16 July 2021 and merged the different layers to delineate the extent of flooded areas along the Meuse and the Rhine Rivers.

2.2.2. Hydrographic and socio-environmental data

The main data source used to describe the hydrographic component of river basins was the HydroBASIN database from HydroSHEDS (Hydrological data and maps based on SHuttle Elevation Derivatives at multiple Scales), a global dataset that provides high-resolution digital data on river networks and watersheds (Lehner et al., 2008). Building upon NASA's Shuttle Radar Topography Mission (SRTM) elevation data, HydroBASIN offers a series of vectorised polygon layers that depict sub-basin boundaries at a global scale (Lehner, 2014). The data is organised into 12 hierarchically nested sub-basin breakdowns globally, allowing for the analysis of river basins at various scales, from small streams to large river systems. This standardised dataset allowed for running a consistent analysis across the five countries studied.

The HydroBASIN database also contains relevant hydrographic attributes that we used to describe the size of catchment area (in km²) and distance to sink (in km). The former is a metric that describes the potential quantity of water that can be drained into the watershed (Lehner, 2014). The latter provides an indication of the distance from the watershed outlet to the outlet of the main river basin (i.e. the North Sea) along the river network.

Three complementary datasets were used to describe additional watershed characteristics. First, we used a digital elevation model (DEM) from the Shuttle Rada Topography Mission (SRTM) (Rabus et al., 2003) with a resolution of 1 arc-second (~30m) to describe the average altitude of each watershed polygon. Second, we used a 1 km² population grid from EUROSTAT derived from the 2021 population and housing census (EUROSTAT, 2021). Third, we used the degree of urbanisation layer from EUROSTAT, which categorises local administrative units as cities, towns and suburbs or rural areas based on a combination of geographical continuity and population density (EUROSTAT, 2019). The dataset selected dated from 2020 with a scale resolution of 1 m (EUROSTAT, 2020).

2.2.3. Twitter data

Throughout this study, we refer to user-generated geo-social media posts from the platform formerly known as Twitter (now X) as 'Tweets', and for consistency with our dataset, we continue to refer to the platform as Twitter. This choice of terminology ~~is intended to~~better reflects the historical context of the data collection process, including specific content moderation practices and data accessibility, that set the original dataset apart from the data available on X today. The georeferenced posts from Twitter ~~(now X)~~ were gathered following previous methods (Havas and Resch, 2021; Schmidt et al., 2023) through the official APIs of the social media network. For each Tweet, we ~~attained~~ extracted the text, the timestamp at which it was posted and its geo-location. This geo-location can be manually set by the user and is provided in the form of coordinates or a bounding box referring to ~~of~~ a so-called “place”. Extracting the dataset for our timeframe and area of interest yielded a total of 14.423 Tweets on which we applied a disaster-related classification, which finally left us with 7.223 Tweets for the subsequent analysis steps.~~The xxx initial geo-social media posts werwe filtered for our timeframe, and area of interest and disaster relatedness, which left us with 7,223 Tweets for the subsequent analysis steps.~~ A summary of the study area characteristics per main basin can be found in Table 1.

Table 1. Summary of study area characteristics per main basin.

	Escaut	Meuse	Rhine
Watershed area (km²)	16.498	71.008	74.093
Precipitations (mm)	91,7	115,0	106,0
Flooded area (km²)	0.01	57.9	63.7

Population density (pers./km²)	251	114	193
Geotagged tweets (N)	1.419	3.090	2.714

2.3. Methodology

Our methodology consisted of several steps, including the semantic analysis of Twitter data and the identification of spatio-temporal patterns. Figure 2 provides an overview of our workflow.

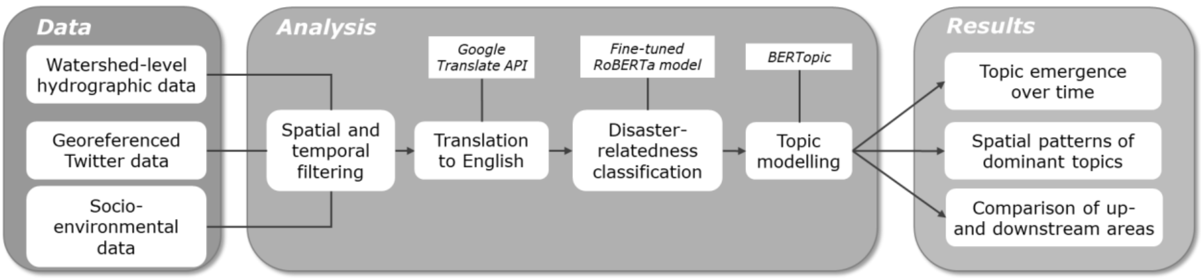


Figure 2. Workflow implemented in this study to extract disaster-related Tweets, ~~and~~ identify ~~topics about flood~~flood-related topics at the watershed level, and highlight patterns of dominant topic occurrence across up- and downstream watersheds.

2.3.1. Delineation of main river basins and daily precipitations per watershed

The most important river catchment areas affected by flooding were delineated using a precipitation dataset in a two-stage process. First, we calculated the total precipitation for the entire study area over the period from 7 to 27 July 2021 using the MAR dataset, which provides a 5 km spatial resolution and hourly temporal resolution. This allowed us to map overall precipitation patterns across the region using a quantile classification. Next, we manually selected the sections of the main river basins that contained areas with more than 100 mm of rain. The delineation of sections from the main river basins was performed using the HydroBASIN delineation at level 7, which represents an intermediate watershed size. This level of detail was ideal for capturing sub-regional hydrographic basins, effectively reflecting precipitation patterns at both regional and country scales. To identify smaller river basins and detect more localised variations of precipitations, we utilised the ~~HydroSHEDS~~ HydroBASIN delineation level 12, which served as our smallest spatial unit of analysis, referred to hereafter as "watershed." This level was used for aggregating Twitter and precipitation data to provide detailed spatial insights. Daily precipitation values were then aggregated at the watershed level using a zonal statistics approach. Daily precipitation data were aggregated at the watershed level using a zonal statistics method. We employed a coverage fraction technique (weighted sum) to summarise the raster precipitation values within each watershed polygon. We choose the weighted sum method that multiplies the precipitation amount of each grid cell by the fraction of the cell contained within the watershed, thereby refining sub-estimates of total precipitations per watershed.~~We applied the coverage fraction (weighted sum) to summarise the~~

~~precipitation values from the raster dataset within each watershed polygon.~~ All data processing except for the topic modelling was conducted using the *dplyr*, *stars*, *sf*, and *exactextractr* packages in R (v 4.3.1) and R Studio (v 2024.04.2).

2.3.2. Semantic classification of social media data

To identify flood-related Tweets and topics, we first translated all Tweets from different languages to English using the Google Translate API. This was done to prevent later topic formation from being mainly influenced by different language characteristics and not the actual contents of the Tweets. Furthermore, we relied on the Google Translate API due to its extensive language support, including regional dialects, which reflect geo-social media discussions across diverse communities. This also offers a higher likelihood that languages beyond official national tongues, such as Turkish or Arabic variants, are included, minimis~~z~~ing the risk of excluding or misrepresenting sub-community discussions. Second, we employed a fine-tuned Twitter-XLM-RoBERTa-base model developed by Hanny et al. (2024) to identify Tweets which were disaster-related, i.e. with content that refers to the occurrence or consequences of both natural and human-induced hazards. It classifies Tweets based on their texts into the categories “unrelated” and “related” with ~~regards~~ regard to any type of disaster event.

In a third step, we identified different topics in the disaster-related Tweets, utilising the state-of-the-art machine learning model BERTopic (Grootendorst, 2022), which consists of five main steps to identify topics in the textual input data. ~~First, it created embeddings for the individual Tweet texts using a BERT-based algorithm (in our case *multi-qa-distilbert-cos-v1*).~~ First, it converted the individual Tweet texts into numerical representations by creating embeddings using a BERT-based algorithm (in our case, multi-qa-distilbert-cos-v1), which maps words into a vector space designed to preserve semantic relationships. Second, the algorithm reduced these embeddings from a 768-dimensional space into a 5-dimensional space. For this, we used the~~a~~ UMAP (Uniform Manifold Approximation and Projection) dimensionality reduction algorithm with five components. In this lower dimensional space, a clustering algorithm identified texts with similar embeddings. We achieved the best results in terms of topic coherence by utilising a K-Means clustering algorithm, where the number of topics identified corresponds to the predetermined number of clusters. ~~The fourth step was a vectorisation step in which we specifically decided to exclude English stop words to allow for more meaningful topic formation.~~ In the fourth step, we used a CountVectorizer from scikit-learn to transform a list of stop words into word vectors, explicitly excluding them to allow for more meaningful topic formation. In the last step, the most relevant words per cluster were identified with the help of a class-based Term Frequency-Inverse Document Frequency (c-TF-IDF) method. More detailed descriptions for each step can be found in Grootendorst (2022).

Furthermore, we limited the number of topics to 30, of which we found 19 topics to be ~~flood-related~~flood related. We further aggregated these 19 topics into 11 main topics that shared similar overarching themes. To cope with randomness in the topic formation due to random starting points in the dimensionality and clustering algorithms, we also performed a topic stability analysis across several topic modelling iterations. For this, we reran the BERTopic algorithm 20 times and compared which topics were replicated across most model iterations, i.e. were the most stable ones. ~~To assess differences between keywords defining each topic across iterations, we employed the string edit distance. Finally, we chose the topic model iteration which exhibited the most stable topics for our subsequent analysis (see Figure S2 for more details).~~ To compare how frequently a specific topic appeared across iterations, we sought to identify a threshold for the maximum allowable difference between topics. The reasoning behind this ~~being~~is that the keywords defining topic A in iteration 1 can be split across several different topics in other iterations. Hence, topic A in iteration 1 could in theory be matching with several to almost all topics in iteration 2. As a result, some matches will be very weak or even m~~ismatches~~is aligned, especially if we also allow for the characters of words to ~~also~~ change slightly ~~change~~. Thus, to mitigate the likelihood of mismatches, we defined an upper bound threshold for how many changes are allowed between to topics of different iterations to be classified as a match. We found that a 17% difference provided such an upper bound, accounting for slight changes in defining words, while ensuring that one topic was only matched to one other topic per iteration. To assess differences between keywords defining each topic, we employed the string edit distance. Finally, we chose the topic model iteration for our subsequent analysis which exhibited the most stable topics (see Figure S2 for more details).

2.3.3. Identification of daily dominant topics per watershed

Once we ~~identified~~ classified the Tweets into~~the~~ various flood-related topic~~s~~ categories, we further analysed which topic was most frequently discussed daily within each watershed. First, ~~Using a spatial join method, w~~we associated the ~~disaster-related~~ Tweets ~~to~~ with their corresponding watersheds based on the XY coordinates of ~~each~~ social media posts using a spatial join method. Next, ~~we calculated the daily occurrence of each topic per watershed and compared their relative importance across river basins using percentage values.~~

~~T~~we ~~he~~ identified~~cation of~~ the most frequently discussed topics ~~daily dominant topics~~using a heuristic approach able to track the evolution of online conversations and trace their dominant character over time. The first rule consisted of counting the number of Tweets belonging to a given topic per day and watershed and selecting the topic category with the maximum Tweet occurrence. ~~involved determining the topic with the highest daily occurrence in each watershed,~~This was calculated following Eq. (1):

$$Topic\_max_{HYBAS,date} = \max\left(Topic\_count_{HYBAS,date}\right) \qquad (1)$$

where *Topic_max* represents the daily dominant topic per watershed, and *Topic_count* refers to the number of times a particular topic appeared in the conversation for a specific watershed (*HYBAS*) and date (*date*). The second rule consisted of discarding all topics with the same maximum occurrence. If two different topics or more had the same maximum value, we removed them from the analysis because we considered those to be equally discussed and thus not representative of the most important conversation taking place in the watershed. ~~calculated following Eq. (1):~~

$$\cancel{Topic\_max_{HYBAS,date} = \max(Topic\_count_{HYBAS,date}) \qquad (1)}$$

~~where *Topic_max* represents the daily dominant topic for each watershed, and *Topic_count* refers to the number of times a particular topic appeared in the conversation for a specific watershed (*HYBAS*) and date (*date*). If~~e ~~multiple topics had the same maximum occurrence in a given watershed on a particular day, we discarded all those topics to ensure that only one unique dominant topic was included in the analysis. This method ensured that the dominant topic for each watershed and day was selected without ambiguity.~~ While an exact solution (e.g. weighted averages of topics) would have been more difficult to interpret, this approach allowed the selection of topics that stood out from online conversations. ~~Additionally, t~~It ~~his~~ also~~selection process~~ helped to ~~reduce~~ reduce the bias of over-representation affecting~~from~~ areas with higher social media activity because a ~~by ensuring that~~ dominant topic~~s~~ was systematically ~~were~~ ~~chosen~~ extracted ~~from watersheds~~ regardless of whether the number of ~~classified~~ generated Tweets was low or high.

2.3.4. Comparison of topic locations with flood and watershed characteristics

We assessed the relationship between dominant flood-related topics and their location across river basins by computing several key variables describing the flood and watershed characteristics.

First, we identified two variables to analyse whether topics emerged in areas affected by the flood. We summarised precipitation values by watershed by computing the average amount of total daily precipitation from the MAR dataset over the period from 7 to 27 July 2021. The percentage of flooded areas per watershed was assessed by dividing the extent of flooded areas delineated and the Copernicus Emergency Mapping Service with the total area of the watershed.

Second, we employed five main watershed characteristics to identify co-occurrences of flood-related online conversation with hydrographic and socio-environmental characteristics. These characteristics included the size of the catchment area, the sink proximity, the elevation, the population density and the degree of urbanisation at the watershed level (see maps in Figure S3). The catchment area and sink proximity values were built upon the HydroBASIN database. A low catchment area indicates a low drainage surface and thus is associated with small river streams, while a high catchment area means greater drainage surface and larger river streams (Chorley, 2019). The sink proximity was computed using the inverse value of distance to sink provided in the HydroBASIN database. A low value means a long distance between the source and the sea outlet, while a high score indicates ~~a close proximity~~a proximity to the sea. The average elevation value per watershed was computed based on the 30 m resolution SRTM elevation data. We also used the inverted value and labelled this variable low elevation to associate high scores with lowlands and low scores with uplands. Using inverted value for these two variables facilitated plots' readability and interpretability. The population density was computed by averaging the 1 km EUROSTAT population grid cell value per watershed polygon. The degree of urbanisation was computed by selecting the local administrative units (LAU) of the category 'Cities'. While these represented large cities mainly located in the riparian zones of main rivers, we measured the coverage fraction of this layer as to provide a percentage of the city class per watershed. The scores describing watershed characteristics graded from low to high, in line with an overall intuitive upstream-downstream logic. This association was supported by a close inspection of variable maps showing urban and densely populated watersheds concentrated close to the sea where elevation is ~~lower~~lower, and the size of catchment areas is larger.

Ridgeline plot~~s~~ is~~are~~ a data visualisation technique we used next to display the distribution of the different topics across the continuous variable attributes~~ across the different topics~~. Ridgeline plots rely upon a kernel density function that estimates the probability density of a variable by smoothing out the distribution using a kernel, which is a continuous and symmetric function (Wilke, 2019). To visualise the spatial variability of dominant topics, we classified watersheds based on their attribute values by creating 100 quantiles and counted the number of times a given topic dominated the conversations over the period from 7 to 27 July. Each distribution plot thus represents the occurrence of a given dominant topic (Y-axis) based on the variable attribute calculated at the watershed level (X-axis). Variable attributes were normalised to a 0-100 scale for comparability. ~~All values were normalised to a 0-100 scale for comparability. Separate plots were created for each topic and each variable to compare central tendencies and variability across upstream and downstream areas. Ridgeline plots were created in R (*ggridges* package) and rely upon a kernel density function that estimates the probability density of a variable by smoothing out the distribution using a kernel, which is a continuous and symmetric function (Wilke, 2019).~~ Ridgeline plots were created in R (*ggridges* package – Wilke, 2024) and ~~S~~separate plots were created for each topic and ~~each~~ variable to compare central tendencies and variability across upstream and downstream areas. For variables describing watershed characteristics, a single peak in the middle (unimodal pattern) suggests that topic occurrence is most frequent in areas corresponding to midstream river sections. Two peaks on the left and right (bimodal distribution) ~~indicates~~ indicate that the topic is more often dominant at both extremes of the basin, with a low occurrence in mid-basin areas. A peak on the right or left ~~suggest~~ suggests that the topic is most relevant in areas associated with downstream or upstream areas. Finally, a flat or even distribution indicates that the topic is equally relevant across the entire basin and consistent across the different parts of the river. The median value was indicated for each distribution to show the extent to which a topic falls into either downstream ~~of~~ or upstream locations. A delineation was also drawn for each watershed ~~characteristics~~ characteristic at a score of 50 ~~in order to~~to mark the separation between upstream and downstream locations.

3. Results

3.1. Precipitation patterns across main river basins

The total amount of precipitation across our study area during the period from 7 to 27 July ranged from 30 to 199 mm (Figure 3). The Meuse River basin recorded abundant and widespread rainfall, particularly in watersheds connected to the Lesse, Ourthe, Amblève, Vesdre and Rur rivers with amounts of precipitations greater than 130 mm. In the Rhine basin, while the extent of high cumulative precipitation was less widespread, significant rainfall was observed in watersheds along the Moselle, Ahr, Erft, and Ruhr rivers. Lastly, in the Escaut river basin, higher precipitation levels were observed in the eastern region, with a lower maximum of 129 mm recorded over the Dyle River. The portions of the main river basins and the watersheds that ~~best~~ covered the best areas with high precipitation levels consisted of a total of 479 watersheds covering an area of 6,000, 28,000 and 29,000 km² for the Escaut, Meuse and Rhine rivers, respectively. Watersheds presented an average size of 131.8 km², 135.0 km², and 135.9 km² respectively, providing a comparable unit of analysis across the three main river basins.

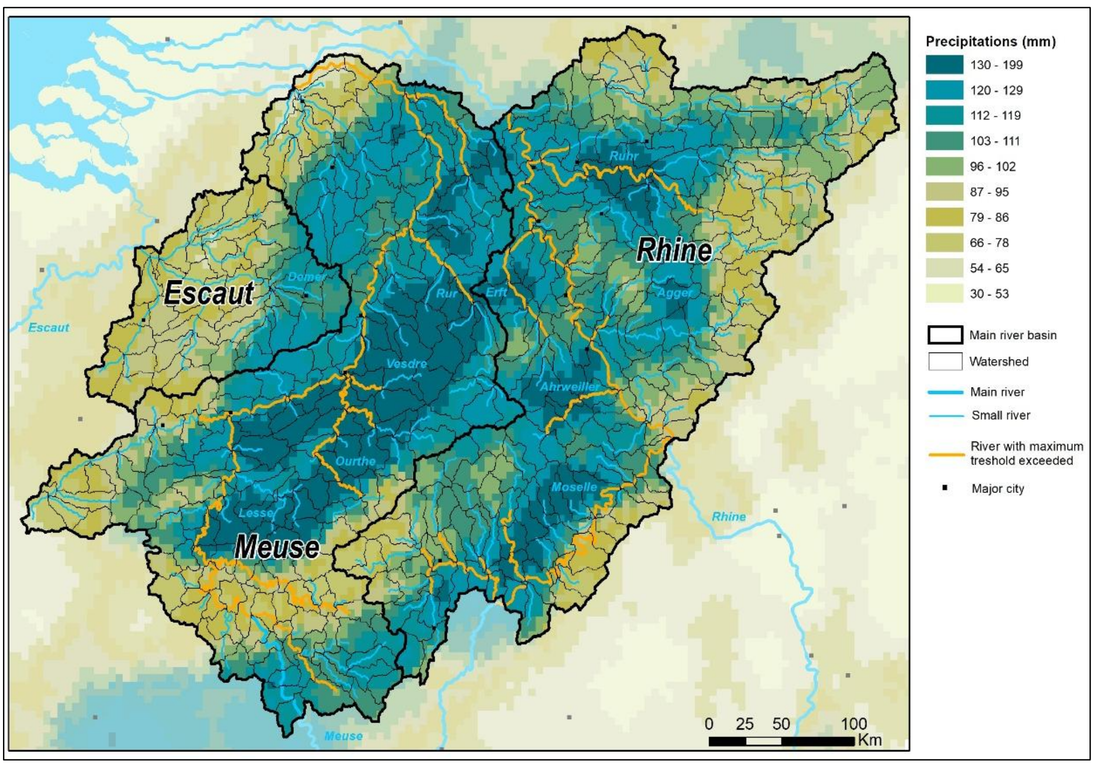


Figure 3. Map showing the study area delineation based on the total precipitations computed at 5 km resolution for the period from 7 to 27 July 2021. The selected portions of the main river basins (HydroBASIN delineation level 7) contain lower size watersheds (HydroBASIN delineation level 12) covering smaller rivers and their tributaries. Rivers where the maximum water threshold

exceeded between the period from 11 to 31 July 2021, and with drainage areas larger than 500 km² identified by CEMS (2021b), are overlayed in thick ~~blue~~ orange. These are the ~~and show the~~ rivers the most impacted by the precipitations.

3.2. Geo-social media topics

Table 2Table 2 illustrates the flood-related topics ~~we~~ identified in our geo-social media data and the corresponding number of Tweets per topic. Each topic was manually assigned a short abbreviation for subsequent analysis. The most straightforward flood-related topics included the *Heavy Rain* topic (540 Tweets), which focused primarily on precipitation events, and the *Help to Victims* topic (594 Tweets), which discussed support for those affected. The *Volunteering and Donations* topic (245 Tweets) highlighted community assistance during the flood. Other topics related to traffic disruptions due to heavy rain and flooding appeared to be closely related in space and content. Thes~~, and~~ e were thus aggregated into a single *Roads Blocked* topic (617 Tweets). Since Topics 2 (Belgian flood) and 13 (Limburg flood) both focused on the same flooded areas, they were merged into the *Meuse Flood* topic, comprising a total of 1,108 Tweets. Similarly, the *Rhine Flood* topic, with 905 Tweets, incorporated topics 25 and 9, which covered overlapping areas within the Rhine river basin. The *Damages* topic (440 Tweets) also reflected the immediate impacts of heavy rain and flooding and comprised two subtopics about water damages and power ~~outage~~outages. The last aggregated topic was the *Compassion* topic (576 Tweets) for which both subtopics were concerned with expressing compassion for the victims. Beyond these, we also found three more politically loaded topics. The *Mourning Victims* topic (358 Tweets) corresponded to a national victim day in Belgium, acknowledging the human cost of the floods. The *German Politics* topic (190 Tweets) focused on the discussion surrounding ~~politician's~~ politicians' management or mismanagement of the flood response in Germany. The *Climate Crisis* topic (231 Tweets) captured discussions on climate change as a contributing factor to the flooding and called to take ~~actions~~ action ~~for limiting~~to limit its impacts. Overall, these topics provided a comprehensive view of the public discourse during the flooding event, highlighting both immediate flood-related concerns and broader socio-political debates (see Table 2). A complete list of all topics and their dominant words can be found in Table S1.

Table 22 Topic overview and most important words.

Topic Abbreviation	**Relevant Terms**	**Number of Tweets**
Heavy Rain	Topic 28: flood, rain, floods, water, weather, flooding, storm, heavy, flooded, like	540
Roads Blocked	Topic 29: direction, near, lummen[2], blocked, accident, lane, brussels[2], closed, ranst[2], stop	235, 63,

	Topic 20: ring, inner, accident, near, lane, outer, blocked, zellik[2], left, tervuren[2] Topic 16: hotton[2], tohogne[2], ardenne[2,3], roche[2], travel, direction, blocked, towards, flooding, accident Topic 15: samson[1], gesves[2], closed, towards, flooding Topic 4: towards, direction, closed, near, blocked, bastogne[2], li, flooding, charleroi[2], travel	10, 8, 301
Meuse Flood	Topic 2: limburg[2,3], water, maas[1], high, flooding, venlo[2], valkenburg[2], maastricht[2], watersnood, south Topic 13: belgium, floods, liege[2,3], li[2,3], namur[2,3], meuse[1], dinant[2], water, flooding, city	719, 389
Rhine Flood	Topic 25: germany, flood, rhine, rain, heavy, water, erftstadt[2], nrw[3], cologne[2], wuppertal[2] Topic 9: ahrweiler[2], flood, help, germany, donations, bonn_district_ahr[2], fire, people	460, 445
Damages	Topic 26: water, basement, high, see, damage, flooded, dry, photo, house, cellar Topic 10: electricity, power, warning, diesel, disaster, siren, areas, lives, without, outage	301, 139
Help to Victims	Topic 5: people, affected, flood, disaster, many, victims, floods, solidarity, help, thanks	594
Volunteering and Donations	Topic 23: donations, help, donate, aid, flood, donation, thank, money, volunteers, distance	245
Compassion	Topic 18: bless, god, amen, living, dead, lord, condolences, relatives, flees, crawls Topic 30: good, strength, thank, family, luck, everyone, keep, people, fingers, thanks	207, 369
Mourning Victims	Topic 21: belgium, national, day, victims, mourning, solidarity, floods, silence, minute, netherlands	358
German Politics	Topic 22: laschet[4], germany, merkel[4], chancellor, cdu[4], german, catastrophe, people, climate, nrw[3]~~nrw~~	190

Climate Crisis	Topic 1: climate, change, crisis, climatecrisis, catastrophe, protection, energy, extreme, heat, climateactionnow	231

[1]River, [2]City or municipality, [3]Province or region, [4]Policitican or political party; nrw=North Rhine-Westphalia, CDU=*Christlich Demokratische Union Deutschlands*.~~The following words represent names of politicians, political parties, cities, rivers, and geographic regions mentioned in the table above.~~

~~Politicians: Merkel, Laschet~~

~~Political Parties: CDU~~

~~Rivers: Maas, Meuse, Rhine, Samson, Ahr~~

~~Cities: Brussels, Ranst, Zellik, Tervuren, Hotton, Tohogne, Charleroi, Bastogne, Venlo, Valkenburg, Maastricht, Liège, Namur, Dinant, Cologne, Wuppertal, Bonn, Ahrweiler, Erfstadt~~

~~Regions: Lummen, Ardenne, Limburg, NRW, Ahrweiler, GesvesPolitical Parties: CDU~~

~~Politicians: Merkel, Laschet~~

3.3. Emergence of flood-related topics per main river basin

Figure 4 shows stacked bar plots for each river basin, which depict the percentage of daily flood-related Tweet counts per topic over all Tweets, revealing three key findings: First, topics were either river basin-specific (*Meuse Flood*, *Rhine Flood*, *Roads Blocked*, *Mourning Victims*, *German Politics*) or stretching across basins (*Heavy Rain*, *Damages*, *Help to Victims*, *Volunteering and Donations*, *Compassion*, *Climate Crisis*). Second, the timing of the topic emergence varied compared to the timing of the precipitation peak. Some topics, such as *Heavy Rain*, *Meuse Flood*, *Rhine Flood*, and *Damages*, peaked during or shortly after the precipitation maximum in their respective basins, while others, including *Roads Blocked*, *Volunteering and Donations*, *Mourning Victims*, and *German Politics*, reached their highest activity levels a few days later. Third, the relative importance of certain topics varied significantly across river basins. For instance, the *Help to Victims* and *Mourning Victims* topics were more prominent in the Escaut ~~River~~ river basin. In contrast, in the Meuse and Rhine River basins, the dominant topics were *Meuse Flood* and *Rhine Flood*, respectively, coinciding with the more severe flooding conditions in these areas.

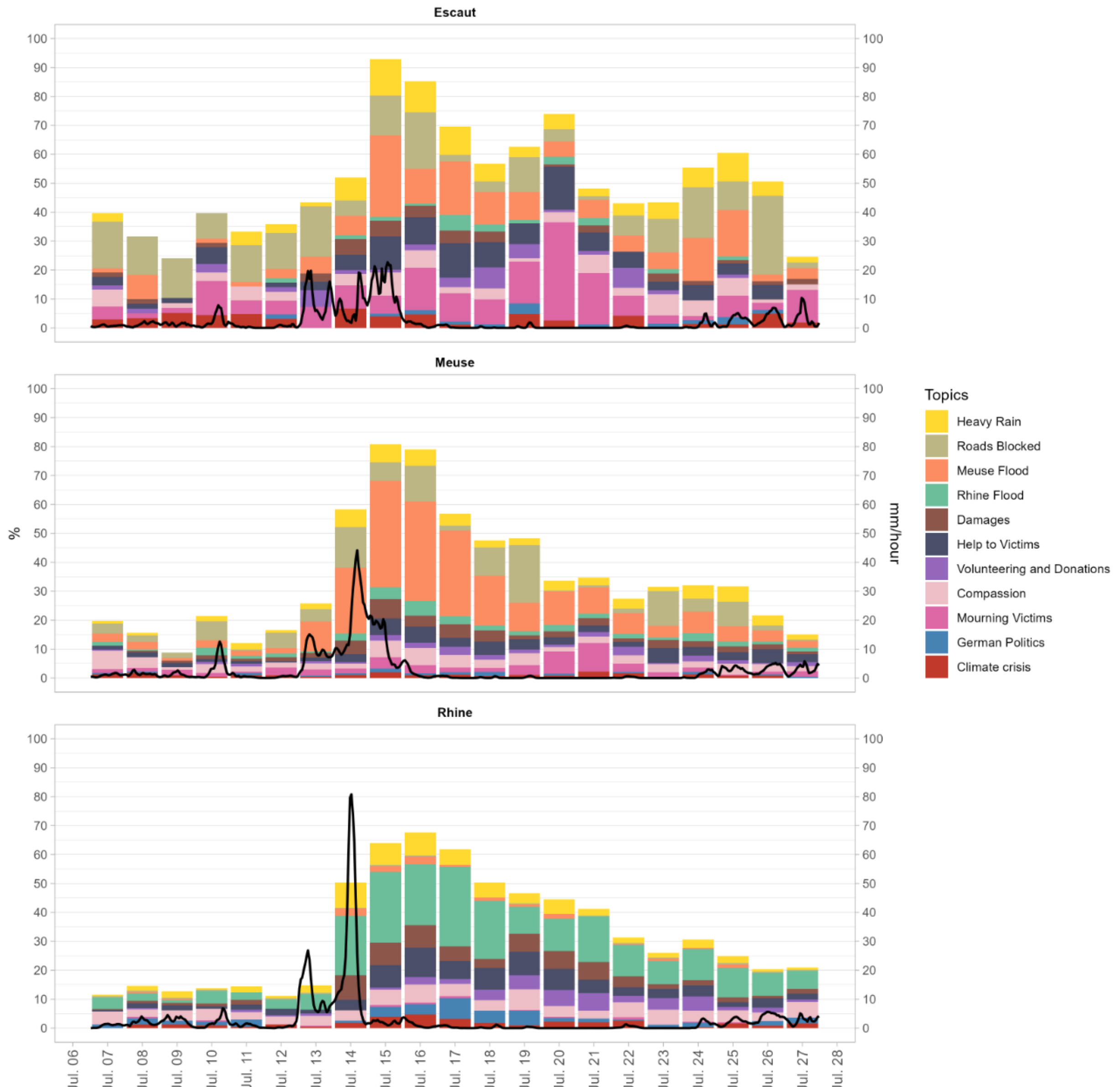


Figure 4. Bar plot reporting the percentage of daily flood-related Tweet counts per topic and main river basin as identified by BERTopic. The remaining percentage represents the share of Tweets unrelated to flooding. Hourly precipitation rates (mm/hour) averaged per main basin (black line) show variations of precipitation intensity and peak time.

3.4. Spatial distribution of dominant geo-social media topics

To assess the spatial distribution and temporal dominance of flood-related geo-social media topics, we analysed the number of days each topic was dominant (i.e., had the highest number of Tweets) within different watersheds for the period from 7 to

27 July 2021 (cf. Figure 5). This analysis identified places of sustained topic dominance, i.e. where certain topics were central to online conversations over an extended period. Results showed several cross-basin topics, such as *Heavy Rain*, *Damages*, *Help to Victims*, *Volunteering and Donations*, *Compassion*, and *Climate Crisis*, which were relatively evenly distributed across river basins. These topics had a low maximum number of dominant days, ranging from 3 to 4 days, with the exception being the *Compassion* topic, which remained dominant for 10 days. In contrast, river basin-specific topics, such as *Meuse Flood* and *Rhine Flood*, were concentrated along the main river courses and transcended national boundaries. They dominated online conversations for the longest periods, with sustained dominance reaching up to 10 days in areas such as Maastricht and Bonn. The *Roads Blocked* topic was notably concentrated in the Meuse and Escaut river basins, where it maintained dominance across large portions of the river basins and, in some watersheds, lasted up to 8 days. Similarly, the *Mourning Victims* topic, which also spanned the Meuse and Escaut river basins, had its longest duration of dominance in Brussels, where it remained central for 6 days. The *German Politics* topic was particularly relevant in major German cities along the Rhine River, but had a shorter dominance period, lasting no more than 3 days. Finally, the N/A topic highlights areas with no geo-social media posts, showing a lack of data in the watersheds on the outskirts of our study area. These regions, primarily in the southern and eastern parts of our study area, are more remote and less urbanised compared to the northwestern areas, which had higher levels of online engagement.

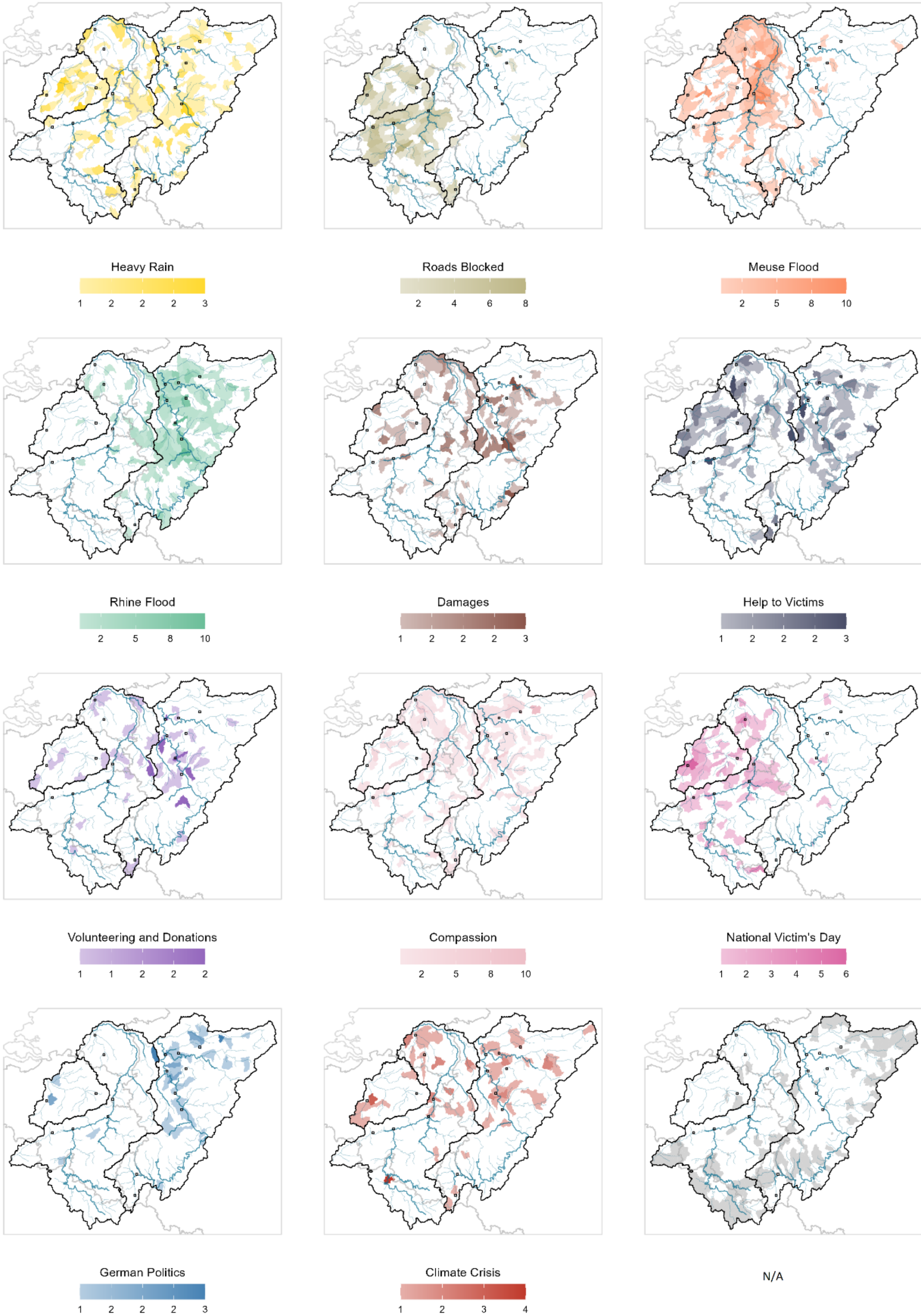


Figure 5. Number of days a topic dominates flood-related conversations in a watershed during the period from 7 to 27 July 2021. Large rivers with maximum threshold exceeded are represented by thick blue lines.

Figure 6 summarises the most dominant topic per watershed over the entire study period, highlighting a distinction between river basin-specific topics that sustained prolonged dominance in areas severely affected by flooding, and cross-basin topics that were broadly distributed but short-lived. In particular, the *Rhine Flood* and *Meuse Flood* topics were most dominant along their respective river courses, spanning multiple countries: France, Luxembourg, Belgium, and the Netherlands for the river Meuse, and Germany and the Netherlands for the Rhine river. In contrast, cross-basin topics such as *Heavy Rain*, *Damages*, and *Compassion* were more ephemeral and mainly dominated peripheral areas outside the main river courses. Specifically, the *Heavy Rain* topic was dominant in the headwaters of river basins, while the *Damages* topic was more prevalent in watersheds associated with secondary rivers, mainly in the Rhine and Meuse basins. The *Compassion* topic also dominated secondary river areas but was primarily dominant in regions with lower precipitation levels.

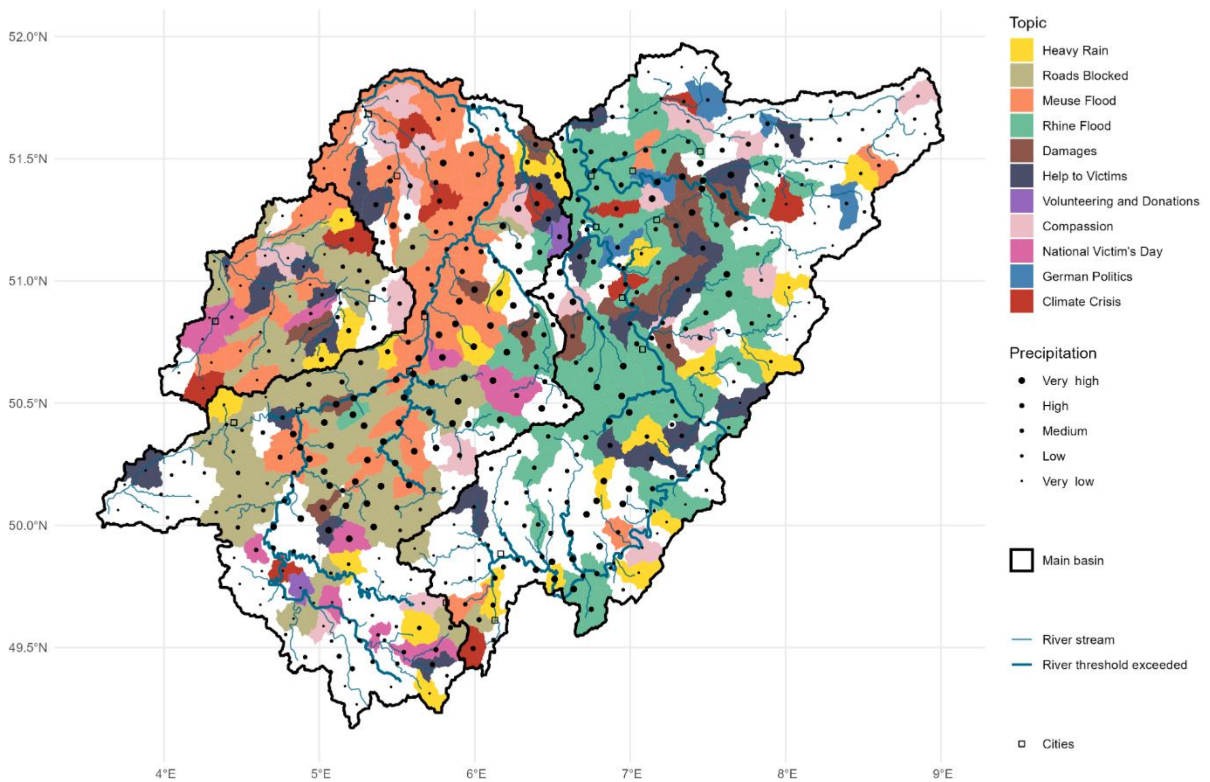


Figure 6. Map showing the overall dominant geo-social media topics per watershed during the period from 7 to 27 July 2021. Dots represent mean precipitation per watershed (5 quantile classes). Large rivers with maximum threshold exceeded between 11 to 31 July 2021 are represented by thick blue lines.

3.5. Comparison of topic occurrence across upstream and downstream areas

To further assess whether dominant topics emerged at specific locations across upstream and downstream areas of the river basins, we examined the spatial distribution of topic occurrence in the light of~~with respect to the~~ flood and watershed characteristics (Figure 7). The goal of this analysis was to determine whether some topics were more prevalent in specific areas within the river basins depending on watersheds' varying precipitation, flood extent, catchment size, elevation, population density, and urbanisation levels.

Our results revealed distinct patterns in topic occurrence across different socio-environmental conditions. The *Heavy Rain* topic was most frequent in regions with medium to high precipitation levels (A1) and in watersheds characterised by less flooded areas (B1) and smaller catchments (C1), suggesting that this topic was driven more by rainfall events than by the flood extent. In contrast, the *Roads Blocked* topic showed more nuanced distributions, appearing in both high and low precipitation areas (A2), but peaking in less flooded regions (B2) and midstream sections with medium-sized catchments and elevations (C2). This topic was also present across areas with varying population densities (F2) and levels of urbanisation (G2), indicating its broader relevance across urban and rural environments.

The *Meuse Flood* topic was mainly dominant in watersheds which recorded medium precipitation levels (A3). Interestingly, this topic was prominent across two distinct ranges of flooded areas (B3), catchment sizes (C3), and urbanisation levels (G3), underscoring their importance in both urban and rural environments located upstream and downstream of the river basin. The *Rhine Flood* topic followed a similar trend except that it was more frequently discussed in places with high precipitations (A4) and flooded areas (B4), reflecting the differences in flood characteristics between the two basins.

The topic ~~about~~ reporting damages exhibited a bimodal distribution~~,~~. It mainly ~~emerging~~ emerged ~~frequently~~ in both low- and ~~high precipitation~~high-precipitation areas (A5) and across both flooded and non-flooded regions (B5). ~~With respect to~~Concerning the river basin's characteristics, the *Damages* topic distribution showed important similarities with the *Rhine* and *Meuse Flood* topics (e.g. C3-4-5), but peaks of topic occurrence did not appear at the same ~~exact~~ locations. The *Damages* topic often emerged in different parts of the river basins including smaller catchment areas (C5), higher elevations(E5), and areas with further distances from the sink (D5), suggesting a greater occurrence in smaller rivers in upstream areas where runoff flooding occurred.

Topics about *Help to Victims* and *Volunteering* dominated areas, which recorded medium to low precipitation~~s~~ levels (6A, 7A). Likewise, the distribution of these topics showed a demarcated peak in areas with low flooded areas (B6, B7). These topics were mainly located in downstream areas with ~~a~~ similar sink proximity (D6, D7), elevation (6E, 7E), population density (6F, 7F), and urbanisation level (G6, G7), suggesting that these discussions stemmed from less affected regions. Similar patterns were observed for the *Compassion* topic, although this topic showed no significant peak across catchment areas (C8), with an even distribution and no clear trend in favour of either upstream or downstream areas.

Finally, the *Mourning Victims* topic was concentrated in low precipitation (A9) and flood (B9) regions, but with a marked downstream bias (D9, E9), indicating that this topic ~~was~~ mainly ~~appearing~~ appeared in the downstream portion of the river basins in highly populated (F9) and urbanised areas (G9). Both the *German politics* and *Climate change* topics showed no marked peaks aligned with ~~in relation to~~ precipitation (A10, A11) or flooded ~~extent~~ area (B10, B11), but were predominantly discussed in low elevation (E10, E11), densely populated (F10, F11), and highly urbanised regions (G10, G11), indicating that these conversations were more tied to socio-political factors than to direct environmental conditions.

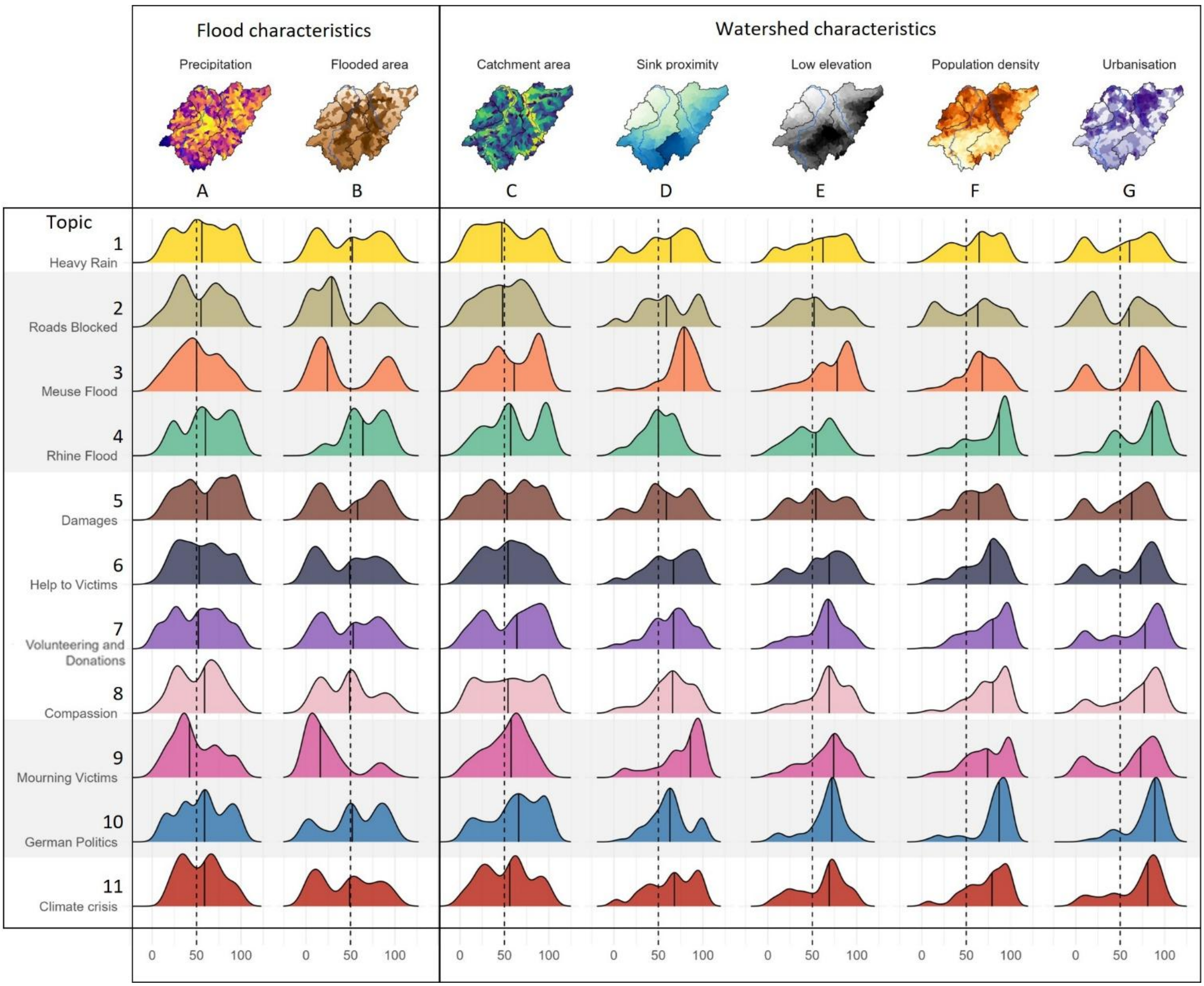


Figure 7. Distribution plots of dominant topic occurrences~~-~~ across the study area based on flood characteristics (A, B) (precipitation and flooded area) and five watershed characteristics (C:G) (catchment area, sink proximity, elevation, population density, and degree of urbanisation). The X-axis represents the variables' low (0) to high (100) values. The Y-axis shows the estimated kernel density, reflecting how often a topic dominated discussions over the period from 7 to 27 July 2021. The black line indicates the median of each distribution, while the dashed line marks the separation between upstream and downstream locations. River basin-specific topics are highlighted with a grey background.

4. Discussion

The results of our study revealed distinct spatio-temporal and semantic patterns in social media responses to flood events in transboundary river basins. Key findings show that cross-basin topics generated prior and during the precipitation peaks~~,~~ – such as *Heavy Rain* and *Damages*~~,~~ – were short-lived and spatio-temporally closely associated with precipitation levels and flood impacts. This suggests that these online conversations were mainly driven by social media users' immediate responses to changing environmental conditions in their respective watersheds. In contrast, river basin-specific topics such as the *Meuse Flood* and *Rhine Flood* demonstrated sustained prominence along the respective river courses throughout the flood event, reflecting the long-lasting impact of flooding on social media user activity in these areas. Specific topics such as the *Roads Blocked* topic highlighted disruptions in infrastructure, with a nuanced presence outside the main river streams. Post-disaster topics such as *Help to Victims* and *Volunteering* were concentrated in less severely affected either upstream or downstream areas, suggesting a larger focus on the emergency and needs of the affected areas located nearby. Overall, our analysis indicated that the nature and focus of online conversations varied significantly depending on user locations in the watershed and the severity of flood impacts. This provides new insights into how social media user communities engage with flood-related discussions in both upstream and downstream parts of a transboundary river basin.

4.1. The contribution of a watershed-based approach ~~for~~ to tracing flood conversations

The watershed-based approach implemented in this research highlighted the relevance of using geo-social media information at the watershed scale, especially along transboundary rivers. While previous studies showed the relevance of social media data ~~to provide~~in providing timely information at the national or regional scales (Tan and Schultz, 2021; Wang and Ye, 2018), our findings ~~demonstrates~~ demonstrate that similar flood-related topics can emerge in neighbouring countries, providing a broader transboundary perspective on flood-related discussions.

Further, we found that flood-related conversations can be associated with specific major river basins. Specifically, our results showed a clear distinction between broadly distributed cross-basin topics and river basin-specific topics in regions heavily affected by flooding. This was especially apparent for the river ~~basin specific~~basin-specific topics *Meuse Flood* and *Rhine Flood*. Interestingly, these two main topics displayed a bimodal pattern when looking at their frequency distribution across the river basin characteristics (Figure 7). This indicates that these flood-related topics were occurring across countries at two distinct levels of their respective river profile, suggesting that the nature and location of flood reports online can vary depending on the hydrographic context.

Indeed, a manual inspection of sample Tweets of the *Meuse Flood* and the *Rhine Flood* topics showed that upstream topics described severe flooding in specific regions, like the Vesdre (Belgium) and Ahr (Germany) watersheds. Yet, topics were also numerous downstream, but in contrast, these were mainly focused on either reporting the flood occurring upstream or providing water level updates regarding the lower section of the Meuse river where flood defences did not breach (Koelewijn et al., 2023). This underlines the value of a watershed-based approach but also shows that geo-social media posts emerging in different hydrographic contexts can reflect different realities – one where the flood's impact is real and another where the flood is only discussed remotely and anticipatively.

Besides, a high level of social media activity may also indicate that the flood impact was less severe, or that the most critical phase of the event has already been mitigated. This was confirmed by the important number of topics about *Help to Victims* and *Volunteering* topics that dominated areas located in less affected regions. Such a type of evidence therefore highlights the importance ~~to take~~of taking caution when interpreting peaks of flood-related topics as indicators of an actual flood-related response and suggests that information generated from social media ~~should always be use in complementarity~~ should always be used as a complementary source alongside verified ~~with other~~ traditional sources of information to provide a comprehensive assessment of situational awareness along the river profiles.

The topics detected using a watershed-based approach contribute to the existing literature on the potential of geo-social media as a valuable tool for monitoring and understanding public responses during disasters (Kryvasheyeu et al., 2016; Silver and Andrey, 2019; Zou et al., 2018; Resch et al., 2018; Fohringer et al., 2015) and for supporting emergency management and reconstruction efforts (Tan and Schultz, 2021; Shan et al., 2023). We advance this body of work by showing that multiple online discussions can be detected as shown by the variety of flood-related topics identified (Table 2). Such a type of topic extraction aligns with some recent research, such as the work of Zander et al. (2023) in Germany, who also identified similar topics. However, by using a transformer-based topic modelling approach (BERTopic), we leveraged word embeddings, allowing us to extract even more nuanced and fine-grained topics specific to each river basin~~s~~ and relevant across the five countries studied.

4.2. Social media data biases and study limitations

4.2.1. Selection bias

One important limitation is the selection bias inherent to social media data, which means our results do not fully capture the broader diversity of public responses across different socio-demographic groups (Petutschnig et al., 2021; Jiang et al., 2019). Twitter data represents a non-uniform sample of the population (Mislove et al., 2011) and exhibits significant biases towards specific age groups, often male and urban populations (Malik et al., 2015), influenced by various factors including user behaviours (Rzeszewski and Beluch, 2017). Besides, considering that our study area included several countries and languages, cultural differences in the use of social media were expected. In fact, our results ~~clearly~~ reflected semantic differences across countries in geo-social media topics which appeared sometimes to be mainly bound to the language spoken within country and regional boundaries (e.g. the *Meuse Flood* topic~~s~~ was more frequent in the ~~Dutch speaking~~Dutch-speaking region of Flanders in Belgium and the Province of Limburg in the Netherlands). Social media activity also tends to concentrate in highly urbanised and ~~–~~populated areas (Fan et al., 2020), leading to the underrepresentation of remote and more vulnerable regions (Karimiziarani et al., 2022; Forati and Ghose, 2022; Xiao et al., 2015). This limitation was verified in our analysis with multiple dominant topics being the most frequent in densely populated and urbanised watersheds (Figure 7; F1:11 and G F1:11).

~~Thus,~~Such a type of underrepresentation of rural areas in social media data can affect the interpretation of flood impacts. For instance, rural towns such as Schuld (BBC, 2021) and Pepinster (DW, 2021) were severely devastated by the floods, ~~with houses being swept away by the floods,~~ resulting in a high~~er~~ proportion of casualties per capita ~~than in urban areas~~. Residents in these areas might have been less likely to tweet updates due to power outages (Reuters, 2021), mobile network failures (Koks et al., 2022), or simply because of lower digital engagement rates. Therefore, the implications for emergency response are that an overreliance on social media signals could lead emergency responders to underestimate the severity of flooding in low-social-media-usage regions and prioriti~~z~~se urban relief efforts over rural recovery needs. To address this drawback, future studies could triangulate additional types of data sources (when available across countries), including remote sensing-based data and questionnaire field surveys for detailed damage assessments or official news media sources from press articles for verified ground-level information (Vicari et al., 2019). An information fusion approach (Wieland et al., 2025) would also help to identify disaster hotspots and evaluate potential cross-border biases in geo-social media data during crisis management situations.

~~To address this drawback, future studies could integrate additional data sources, such as traditional field surveys or official news media sources such as press article (Vicari et al., 2019) in order to provide a more comprehensive view of public response to further evaluate the reliability and potential cross-border biases of geo-social media data in crisis management situationss.~~

4.2.2. Geolocation limitations

Another limitation of this study is the potential spatial bias in crowdsourced data. Our watershed-based approach relied on accurately extracting geo-tagged Tweets within watershed boundaries. In our dataset, most Tweet locations were provided as polygons (81%), with an average polygon size of 185 km² (median size of 119 km²). However, this polygon size corresponds to the scale of European cities such as Brussels (161 km²) or Düsseldorf (217 km²), and additional visual analysis indicated that most of these polygons were concentrated around major urban centres. Consequently, we assumed that most Tweets with polygon locations within a city's watershed area originated from these cities. Nevertheless, this spatial discrepancy should be considered when interpreting sustained topic dominance in watersheds which encompass large urban areas. To mitigate this issue, future research could incorporate a higher proportion of precisely geotagged Tweets when available and refine watershed boundary delineations by integrating the spatial extent of urban areas.

4.2.3. Topic stability

~~An additional consideration in our analysis was the inherent variability of the semantic modelling algorithm (BERTopic) (Grootendorst, 2022b), which is not entirely deterministic and depends on randomness in identifying topic clusters. To mitigate this issue, we ran the algorithm 20 times to assess the stability of the topics, distinguishing between stable and unstable clusters. However, some topics relevant to the flood analysis, such as *Damages*, displayed low stability, appearing in only 25% of the iterations. Similarly, topic 18, which was aggregated into the *Compassion* topic, was identified in only five iterations (25%), and the *Help to Victims* topic was stable across nine iterations (45%). This variability needs to be considered when interpreting these less stable topics, as it indicates potential inconsistencies in their reproducibility. Despite this, the iteration we used for the final analysis exhibited the highest overall topic stability, with the majority of topics remaining consistent in terms of defining keywords across multiple runs. Future studies could enhance topic stability by incorporating ensemble approaches, combining results from multiple iterations to form a consensus topic structure, or by exploring alternative deterministic algorithms which offer higher stability in semantic modelling.~~ An additional consideration in our analysis was the inherent variability of the semantic modelling algorithm (BERTopic) (Grootendorst, 2022b), which is not entirely deterministic and depends on randomness in identifying topic clusters. To mitigate this issue, we ran the algorithm 20 times to assess the topics stability ~~of the topics~~, distinguishing between stable and unstable clusters. We found that the keyword sets defining the Topic 26 and Topic 10 of the Damages topic only occurred across 20% and 30% of the iterations when allowing for a maximal difference of 17%. Similarly, topic 18, which was ~~-~~aggregated ~~part of~~ into the Compassion topic, was identified in only five iterations (25%), ~~,~~ and the Help to Victims topic was stable across nine iterations (45%).

This is because we applied a highly restrictive maximal difference threshold of just 17% between topics across iterations, which, in some cases, corresponded to a difference of fewer than eight characters. Therefore, this does not imply that other iterations lacked topics related to damages, victim assistance, or compassion.~~This is due to the fact that we applied a highly restrictive maximal difference threshold of just 17% between topics across iterations, which, in some cases, corresponded to a difference of fewer than eight characters. Thus, this does not imply that other iterations were missing topics related to damages, help to victims, or compassion.~~ Rather, it means that the defining keywords for these topics changed by more than 17%, exceeding our threshold and resulting in their classification as distinct topics when comparing across iterations. This variability in keywords needs to be considered when interpreting these less stable topics. However, future studies could enhance topic stability analysis by incorporating ensemble approaches~~,~~ that combin~~ing~~e results from multiple iterations to form a consensus topic structure, or by exploring more sophisticated embedding-based similarity comparisons, which allow ~~to~~ captur~~e~~ing the underlying meaning of ~~the~~ keywords.

4.2.4. Dominant topic selection

Instead of computing the relative importance of each topic, our method assigned the topic name that was most discussed per watershed on a given day based on two main heuristic rules: selecting the most discussed topic and removing topics with the same number of maximum occurrences. A detailed analysis of this dominant topic selection process revealed that this method proved to be robust in selecting the most relevant topics. Indeed, the less frequently discussed topics were primarily affected by the filtering process (see Table S3). For instance, 61% of topics mentioned only once on a given day and watershed were discarded. In contrast, none of the most frequently discussed topics (i.e., those occurring between 11 and 35 times) were removed from the analysis. Visually, we also observed that these smaller topics corresponded to more peripheral places located outside the most impacted areas, thus carrying some noise that was removed thanks to this method. Topic-wise, the filtering process impacted each topic category to a relatively similar degree, with the proportion of discarded topics ranging from 40% to 70%, except for the *Roads Blocked* topic, for which only 24.8% of occurrences were filtered out (see Table S2). We found that this topic was atypical, as Tweets were mainly generated from the *Touring Mobilis* Twitter account, a traffic information service active in Belgium and the Netherlands that provides real-time updates on road conditions. It described widespread rain- or flood-induced traffic problems occurring throughout our study area, which can explain why it was less frequently in conflict with other topics located closer to the flooded areas.

4.3. Implications for transboundary flood risk management

Despite these limitations, our findings provide~~have~~ meaningful implications for transboundary flood risk management. We show that social media analytics can support the detection, monitoring and prediction of human responses to flood by sharing information with stakeholders and action forces across interconnected regions and countries. This potential of georeferenced social media data for early warning purposes ~~has been confirmed~~contributes to ~~in~~ a variety of ~~other~~previous studies in the field of disaster risk reduction (Havas and Resch, 2021; Rossi et al., 2018; Schmidt et al., 2025; Stollberg and Groeve, 2012) ~~(Schmidt et al. 2025, Havas & Resch 2021, Rossi et al. 2018)~~.

First, topics dealing with heavy rainfalls can indicate problematic precipitations in upstream areas and thereby help to anticipate dangerous water flows in upstream areas or overflow flooding in downstream areas. Another topic that has the potential for disaster early warning ~~was~~is ~~concerned with~~the one informing on ~~is the one that refers to~~ blocked roads. This topic could be used to identify increases of road traffic issues, especially in remote, rural environments located in the upper parts of the watershed, ~~and~~ thereby allowing ~~a better assessment of affected areas and~~ for the improvement of emergency resource allocation. In both cases, this type of information complements traditional meteorological information ~~such as~~from radar images and ~~,~~ other satellite-based flood signals~~as has also been shown in prior research~~ (Jongman et al., 2015)~~(Jongman et la.~~

~~2015),~~ because it can inform about the on-site, problematic impact of heavy precipitations ~~for~~ on ~~road traffic and~~ human activities~~mobility~~.

Second, river ~~basin specific~~basin-specific topics might allow for ~~to identify~~the identification of the sections of the main river basins affected by overflow flooding and show the extent to which water levels are unusually high ~~downstream~~ during ~~the~~ disaster events. ~~This might assist in the protection of people living in lower parts of the river.~~ This becomes helpful~~could~~ for ~~help~~ policymakers and emergency responders when it comes to~~to identify~~target interventions in lower parts of the river where people ~~might be~~are the most at risk of stream overflow, enabling the organisation of timely emergency assistance whe~~n~~re needed. In this regard, Restrepo-Estrada et al. (2018) ~~Restrepo-Estrada et al. 2018~~ similarly emphasi~~z~~sed the capability of geo-social media data to improve streamflow estimation. D~~D~~amage-related topics on the other hand may be used for the detection of rapid damage assessments from run-off flooding. This topic might also help to detect smaller and indirect effects of the flood such as flooded basements and power failures which dominated conversations in different parts of the main river sections affected by the flood.

Third, post-disaster response topics focused on helping victims and volunteering initiatives and can further be used to identify where help is either called for or coming from in a transboundary river basin during the recovery phase. Once the location associated with the Tweet has been verified, ~~this~~ it can be leveraged to~~could~~ trigger faster emergency relief operations across regional or national borders. Finally, topics related to mourning victims, politics or climate change ~~might enable~~enable a deeper understanding of the concerns of those living outside the impacted areas and thereby provide a remote perspective on the causes and potential mismanagement of the flood disaster.

Future research could explore how different countries within a shared river basin can use such type of information to better communicate and coordinate emergency response in the face of a transboundary flood. This, however, must also be ~~done~~developed in ~~accordance with~~parallel with the creation of standardi~~z~~sed technical and legal frameworks for international disaster management (Gilga et al., 2024)~~(Gilga et al.) (Gilga et al. 2024)~~. ~~This is especially challenging considering that the~~Since the ex-post analysis provided here has not been tested in real-time for addressing ~~emergency situations~~emergency situations, ~~. S~~studies in the field should therefore focus on developing ~~the development of~~ methods able to identify dominant topics in near real-time and over shorter time windows. To meet this challenge, we emphasise the critical need for social media data providers to allow API access in emergency situations triggered by disaster events.

5. Conclusion

Our study provides a novel perspective on flood-related discussions on social media by adopting a watershed-based approach to analyse topic emergence and their distribution in transboundary river basins. Our findings reveal distinct spatio-temporal dynamics in the public discourse, showing how timely topics describing heavy rains or flood damages were closely tied to immediate environmental conditions in the upstream areas, while post-disaster topics about helping victims or volunteering were more prevalent in areas less affected by flooding located both upstream and downstream. This understanding of how social media conversations evolve in relation to~~-~~ flood severity and watersheds' socio-environmental characteristics offers new opportunities for integrating geo-social media analytics into transboundary flood risk management. By enhancing the understanding of how social media users engage with flood-related information along river streams, this approach provides a framework for future studies to explore the interplay between environmental conditions, social media engagement, and transboundary collaboration in disaster contexts. Ultimately, by incorporating insights from social media into traditional disaster management strategies and tools — such as early warning and monitoring services~~,~~ — future research and policy initiatives can ~~improve~~ enhance transboundary coordination in flood response and recovery efforts ~~as well as~~while more effectively ~~better~~ addressing the needs of populations increasingly exposed to climate~~-~~ risks.

Author contribution

Conceptualization, S.D, D.A, S.S, C.L. and B.R ; methodology, S.D and D.A.; software, S.D and D.A.; validation, S.D and D.A.; formal analysis, S.D, D.A and S.S; investigation, S.D and S.S; resources, B.R; data curation, D.A, S.S and S.D; writing---original draft preparation, S.D, D.A and S.S; writing---review and editing, S.D, D.A, S.S, C.L. and B.R; visualization, S.D and D.A; supervision, C.L., B.R; project administration, S.D., B.R; funding acquisition, S.D, B.R. All authors have read and agreed to the published version of the manuscript.

Competing interests

The authors declare that they have no conflict of interest.

Acknowledgments

We would like to warmly thank Xavier Fettweis and Sébastien Doutreloup from the Climate Lab at the Department of Geography, University of Liège, for providing ~~customized~~ customised precipitation data tailored to our study area. This work was funded by the Belgian *Fonds de la Recherche Scientifique – FNRS* under Dr. Sébastien Dujardin's postdoctoral fellowship 2020-23 and was approved by the UNamur's *Comité d'Ethique en Sciences Humaines* (2020/05). It also received funding from the European Commission—European Union under HORIZON EUROPE (HORIZON Research and Innovation Actions) under grant agreement 101093003 (HORIZON-CL4-2022-DATA-01-01). Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union—European Commission. Neither the European Commission nor the European Union can be held responsible for them.

Code availability statement

The code to reproduce the Tweet translation, topic modelling and the topic coherence analysis can be found in the corresponding GitHub repository: https://github.com/DorianZGIS/Tracing-online-flood-conversations-across-borders.git.

References

Aall, C., Holm, T. B., Cauchy, A., Rudolf, F., Harris, K., Jansen, M. K., Gobert, J., Lager, F., Arvis, B., and Bour, M.: Think global—act local: the challenge of producing actionable knowledge on transboundary climate risks at the sub-national level of governance, Frontiers in Climate, 5, 10.3389/fclim.2023.1170142, 2023.

Acikara, T., Xia, B., Yigitcanlar, T., Hon, C., Acikara, T., Xia, B., Yigitcanlar, T., and Hon, C.: Contribution of Social Media Analytics to Disaster Response Effectiveness: A Systematic Review of the Literature, Sustainability, 15, 10.3390/su15118860, 2023.

BBC: In pictures: Floods kill dozens in Germany and Belgium, 2021.

Beigi, G., Hu, X., Maciejewski, R., and Liu, H.: An Overview of Sentiment Analysis in Social Media and Its Applications, Sentiment Analysis and Ontology Engineering, 10.1007/978-3-319-30319-2_13, 2016.

Blomeier, E., Schmidt, S., Resch, B., Blomeier, E., Schmidt, S., and Resch, B.: Drowning in the Information Flood: Machine-Learning-Based Relevance Classification of Flood-Related Tweets for Disaster Management, Information, 15, 10.3390/info15030149, 2024.

Bruneau, P., Brangbour, E., Marchand-Maillet, S., Hostache, R., Chini, M., Pelich, R.-M., Matgen, P., and Tamisier, T.: Measuring the impact of natural hazards with citizen science: The case of flooded area estimation using Twitter, Remote Sensing, 13, 1153, 2021.

The Copernicus Emergency Management Service forecasts, notifies, and monitors devastating floods in Germany, Netherlands, Belgium and Switzerland.: https://emergency.copernicus.eu/mapping/ems/copernicus-emergency-management-service-forecasts-notifies-and-monitors-devastating-floods, last access: 04/10/2024.

Flooding in Europe: https://climate.copernicus.eu/esotc/2021/flooding-july, last

Chorley, R. J.: Introduction to physical hydrology, Routledge2019.

Clarke, B., Otto, F., Stuart-Smith, R., Harrington, L., Clarke, B., Otto, F., Stuart-Smith, R., and Harrington, L.: Extreme weather impacts of climate change: an attribution perspective, Environmental Research: Climate, 1, 10.1088/2752-5295/ac6e7d, 2022.

Doutreloup, S., Bois, B., Pohl, B., Zito, S., and Richard, Y.: Climatic comparison between Belgium, Champagne, Alsace, Jura and Bourgogne for wine production using the regional model MAR, OENO One, 56, 1-17, 10.20870/oeno-one.2022.56.3.5356, 2022.

DW: Belgians pick up the pieces after deadly floods, 2021.

EUROSTAT: Methodological manual on territorial typologies – 2018 edition, Publications Office, doi/10.2785/930137, 2019.

Degree of urbanisation: https://ec.europa.eu/eurostat/web/gisco/geodata/population-distribution/degree-urbanisation, last access: 04/10/2024.

GEOSTAT 1 km2 population grid: https://ec.europa.eu/eurostat/web/gisco/geodata/population-distribution/geostat, last access: 04/10/2024.

Fan, C., Esparza, M., Dargin, J., Wu, F., Oztekin, B., and Mostafavi, A.: Spatial biases in crowdsourced data: Social media content attention concentrates on populous areas in disasters, Computers, Environment and Urban Systems, 83, 101514, https://doi.org/10.1016/j.compenvurbsys.2020.101514, 2020.

Fekete, A. and Sandholz, S.: Here comes the flood, but not failure? Lessons to learn after the heavy rain and pluvial floods in Germany 2021, Water, 13, 3016, 2021.

Florath, J., Chanussot, J., and Keller, S.: Rapid natural hazard extent estimation from twitter data: investigation for hurricane impact areas, Natural Hazards, 120, 6775-6796, 2024.

Fohringer, J., Dransch, D., Kreibich, H., and Schröter, K.: Social media as an information source for rapid flood inundation mapping, Natural Hazards and Earth System Sciences, 15, 10.5194/nhess-15-2725-2015, 2015.

Forati, A. M. and Ghose, R.: Examining community vulnerabilities through multi-scale geospatial analysis of social media activity during hurricane Irma, International Journal of Disaster Risk Reduction, 68, 102701, 10.1016/j.ijdrr.2021.102701, 2022.

Gilga, C., Hochwarter, C., Knoche, L., Schmidt, S., Ringler, G., Wieland, M., and Resch, B.: Legal and ethical considerations for demand-driven data collection and AI-based analysis in flood response, Available at SSRN 4911032, https://doi.org/10.2139/ssrn.4911032, 2024.

Grootendorst, M.: BERTopic: Neural topic modeling with a class-based TF-IDF procedure, 10.48550/ARXIV.2203.05794, 2022.

Gunnell, K., Mulligan, M., Francis, R. A., and Hole, D. G.: Evaluating natural infrastructure for flood management within the watersheds of selected global cities, Science of The Total Environment, 670, 10.1016/j.scitotenv.2019.03.212, 2019.

Hanny, D. and Resch, B.: Multimodal Geo-Information Extraction from Social Media for Supporting Decision-Making in Disaster Management, AGILE: GIScience Series, 5, 10.5194/agile-giss-5-28-2024, 2024.

Hanny, D., Schmidt, S., and Resch, B.: Active Learning for Identifying Disaster-Related Tweets: A Compar, Lecture Notes in Networks and Systems, 10.1007/978-3-031-66428-1_8, 2024.

Havas, C. and Resch, B.: Portability of semantic and spatial–temporal machine learning methods to analyse social media for near-real-time disaster monitoring, Natural Hazards, 108, 2939-2969, 10.1007/s11069-021-04808-4, 2021.

Huang, Q. and Xiao, Y.: Geographic Situational Awareness: Mining Tweets for Disaster Preparedness, Emergency Response, Impact, and Recovery, ISPRS International Journal of Geo-Information, 4, 1549-1568, 2015.

Ionita, M. and Nagavciuc, V.: Extreme floods in the Eastern part of Europe: Large-scale drivers and associated impacts, Water, 13, 1122, 10.3390/w13081122, 2021.

IPCC: Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation. A Special Report of Working Groups I and II of the Intergovernmental Panel on Climate Change [Field, C.B., V. Barros, T.F. Stocker, D. Qin, D.J. Dokken, K.L. Ebi, M.D. Mastrandrea, K.J. Mach, G.-K. Plattner, S.K. Allen, M. Tignor, and P.M. Midgley (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 582 pp.2012.

IPCC, Press, C. U. (Ed.): Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change, Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA2021.

Jiang, Y., Li, Z., and Ye, X.: Understanding demographic and socioeconomic biases of geotagged Twitter users at the county level, Cartography and Geographic Information Science, 46, 10.1080/15230406.2018.1434834, 2019.

Jongman, B., Wagemaker, J., Romero, B. R., and De Perez, E. C.: Early Flood Detection for Rapid Humanitarian Response: Harnessing Near Real-Time Satellite and Twitter Signals, 10.3390/ijgi4042246, 2015.

Kahle, M., Kempf, M., Martin, B., and Glaser, R.: Classifying the 2021 “Ahrtal” flood event using hermeneutic interpretation, natural language processing, and instrumental data analyses, Environmental Research Communications, 2022.

Karimiziarani, M., Jafarzadegan, K., Abbaszadeh, P., Shao, W., and Moradkhani, H.: Hazard risk awareness and disaster management: Extracting the information content of twitter data, Sustainable Cities and Society, 77, 103577, 2022.

Karmegam, D. and Mappillairaju, B.: Spatio-temporal distribution of negative emotions on Twitter during floods in Chennai, India, in 2015: a post hoc analysis, International Journal of Health Geographics, 19, 19, 10.1186/s12942-020-00214-4, 2020.

Koelewijn, A., Pol, J., and Schaijk, M. v.: Performance of flood defences in the Netherlands during the 2021 summer floods, Journal of Coastal and Riverine Flood Risk, 2, 10.59490/jcrfr.2023.0007, 2023.

Koks, E. E., van Ginkel, K. C. H., van Marle, M. J. E., and Lemnitzer, A.: Brief communication: Critical infrastructure impacts of the 2021 mid-July western European flood event, Nat. Hazards Earth Syst. Sci., 22, 3831-3838, 10.5194/nhess-22-3831-2022, 2022.

Kruspe, A., Kersten, J., and Klan, F.: Review article: Detection of actionable tweets in crisis events, Natural Hazards and Earth System Sciences, 21, 10.5194/nhess-21-1825-2021, 2021.

Kryvasheyeu, Y., Chen, H., Obradovich, N., Moro, E., Hentenryck, P. V., Fowler, J., and Cebrian, M.: Rapid assessment of disaster damage using social media activity, Science Advances, 2, doi:10.1126/sciadv.1500779, 2016.

Lehner, B.: HydroBASINS: Global watershed boundaries and sub-basin delineations derived from HydroSHEDS data at 15 second resolution — Technical documentation version 1.c, 2014.

Lehner, B., Verdin, K., and Jarvis, A.: New Global Hydrography Derived From Spaceborne Elevation Data, Eos, Transactions American Geophysical Union, 89, 10.1029/2008EO100001, 2008.

Lorenz, C., Gilbert, A., Vellinga, P., Lorenz, C., Gilbert, A., and Vellinga, P.: Sustainable management of transboundary river basins: a line of reasoning, Regional Environmental Change, 2, 10.1007/s101130100023, 2001.

Malik, M., Lamba, H., Nakos, C., and Pfeffer, J.: Population bias in geotagged tweets, Proceedings of the International AAAI Conference on Web and Social Media, 18-27,

Mehta, A. S. and Warner, J. F.: Multi-level hegemony in transboundary Flood Risk Management: A downstream perspective on the Maritsa Basin, Environmental Science & Policy, 129, 10.1016/j.envsci.2021.12.014, 2022.

Mislove, A., Lehmann, S., Ahn, Y.-Y., Onnela, J.-P., and Rosenquist, J.: Understanding the demographics of Twitter users, Proceedings of the International AAAI Conference on Web and Social Media, 554-557,

Moghadas, M., Fekete, A., Rajabifard, A., and Kötter, T.: The wisdom of crowds for improved disaster resilience: a near-real-time analysis of crowdsourced social media data on the 2021 flood in Germany, GeoJournal, 88, 4215-4241, 10.1007/s10708-023-10858-x, 2023.

Petutschnig, A., Resch, B., Lang, S., and Havas, C.: Evaluating the Representativeness of Socio-Demographic Variables over Time for Geo-Social Media Data, ISPRS International Journal of Geo-Information, 10, 323, 2021.

Polese, M., Tocchi, G., Babič, A., Dolšek, M., Faravelli, M., Quaroni, D., Borzi, B., Rebora, N., Ottonelli, D., Wernhart, S., Pejovic, J., Serdar, N., Lebar, K., Rusjan, S., Masi, R., Resch, C., Kern, H., Cipranić, I., Ostojic, M., and Prota, A.: Multi-risk assessment in transboundary areas: A framework for harmonized evaluation considering seismic and flood risks, International Journal of Disaster Risk Reduction, 101, 10.1016/j.ijdrr.2024.104275, 2024.

Rabus, B., Eineder, M., Roth, A., and Bamler, R.: The shuttle radar topography mission—a new class of digital elevation models acquired by spaceborne radar, ISPRS Journal of Photogrammetry and Remote Sensing, 57, 10.1016/S0924-2716(02)00124-7, 2003.

Rahayu, H. P., Zulfa, K. I., Nurhasanah, D., Haigh, R., Amaratunga, D., and Wahdiny, I. I.: Unveiling transboundary challenges in river flood risk management: learning from the Ciliwung River basin, Natural Hazards and Earth System Sciences, 24, 10.5194/nhess-24-2045-2024, 2024.

Resch, B., Usländer, F., and Havas, C.: Combining machine-learning topic models and spatiotemporal analysis of social media data for disaster footprint and damage assessment, Cartography and Geographic Information Science, 45, 362-376, 10.1080/15230406.2017.1356242, 2018.

Restrepo-Estrada, C., de Andrade, S. C., Abe, N., Fava, M. C., Mendiondo, E. M., and de Albuquerque, J. P.: Geo-social media as a proxy for hydrometeorological data for streamflow estimation and to improve flood monitoring, Computers & Geosciences, 111, 148-158, https://doi.org/10.1016/j.cageo.2017.10.010, 2018.

Reuters: Floods cut power to 200,000 households in western Germany, 2021.

Rossi, C., Acerbo, F. S., Ylinen, K., Juga, I., Nurmi, P., Bosca, A., Tarasconi, F., Cristoforetti, M., and Alikadic, A.: Early detection and information extraction for weather-induced floods using social media streams, International Journal of Disaster Risk Reduction, 30, 145-157, https://doi.org/10.1016/j.ijdrr.2018.03.002, 2018.

Rzeszewski, M. and Beluch, L.: Spatial Characteristics of Twitter Users—Toward the Understanding of Geosocial Media Production, ISPRS International Journal of Geo-Information, 6, 236, 2017.

Schmidt, S., Friedemann, M., Hanny, D., Resch, B., Riedlinger, T., and Mühlbauer, M.: Enhancing satellite-based emergency mapping: Identifying wildfires through geo-social media analysis, Big Earth Data, 1-23, 10.1080/20964471.2025.2454526, 2025.

Schmidt, S., Zorenböhmer, C., Arifi, D., Resch, B., Schmidt, S., Zorenböhmer, C., Arifi, D., and Resch, B.: Polarity-Based Sentiment Analysis of Georeferenced Tweets Related to the 2022 Twitter Acquisition, Information, 14, 10.3390/info14020071, 2023.

Schüttrumpf, H., Birkmann, J., Brüll, C., Burghardt, L., Johann, G., Klopries, E., Lehmkuhl, F., Schütttrumpf, A., and Wolf, S.: Herausforderungen an den Wiederaufbau nach dem Katastrophenhochwasser 2021 in der Eifel, Nachhaltigkeit im Wasserbau—Umwelt, Transport, Energie, Dresdner Wasserbauliche Mitteilungen 68, 2022, 5-16,

Shan, S., Zhao, F., Shan, S., and Zhao, F.: Social media-based urban disaster recovery and resilience analysis of the Henan deluge, Natural Hazards 2023 118:1, 118, 10.1007/s11069-023-06010-0, 2023.

Silver, A. and Andrey, J.: Public attention to extreme weather as reflected by social media activity, Journal of Contingencies and Crisis Management, 27, 10.1111/1468-5973.12265, 2019.

Stollberg, B. and Groeve, T. d.: The use of social media within the global disaster alert and coordination system (GDACS), Proceedings of the 21st International Conference on World Wide Web, Lyon, France, 10.1145/2187980.2188185, 2012.

Tan, L. and Schultz, D. M.: Damage classification and recovery analysis of the Chongqing, China, floods of August 2020 based on social-media data, Journal of Cleaner Production, 313, 10.1016/j.jclepro.2021.127882, 2021.

UNECE: Transboundary flood risk management: experiences from the UNECE region, 2009.

UNISDR: The human cost of weather related disasters 1995-2015, 2015.

Vicari, R., Tchiguirinskaia, I., Tisserand, B., and Schertzer, D.: Climate risks, digital media, and big data: following communication trails to investigate urban communities' resilience, Natural Hazards and Earth System Sciences, 19, 10.5194/nhess-19-1485-2019, 2019.

Wang, R.-Q., Mao, H., Wang, Y., Rae, C., and Shaw, W.: Hyper-resolution monitoring of urban flooding with social media and crowdsourcing data, Computers & Geosciences, 111, 139-147, 2018.

Wang, Z. and Ye, X.: Social media analytics for natural disaster management, International Journal of Geographical Information Science, 32, 10.1080/13658816.2017.1367003, 2018.

Wania, A., Joubert-Boitat, I., Dottori, F., Kalas, M., Salamon, P., Wania, A., Joubert-Boitat, I., Dottori, F., Kalas, M., and Salamon, P.: Increasing Timeliness of Satellite-Based Flood Mapping Using Early Warning Systems in the Copernicus Emergency Management Service, Remote Sensing, 13, 10.3390/rs13112114, 2021.

Wieland, M., Schmidt, S., Resch, B., Abecker, A., and Martinis, S.: Fusion of geospatial information from remote sensing and social media to prioritise rapid response actions in case of floods, Natural Hazards, 10.1007/s11069-025-07120-7, 2025.

Wilke, C. O.: Fundamentals of data visualization: a primer on making informative and compelling figures, O'Reilly Media2019.

Wilke, C. O.: ggridges: Ridgeline Plots in 'ggplot2' [code], 2024.

Wyard, C., Scholzen, C., Doutreloup, S., Hallot, É., and Fettweis, X.: Future evolution of the hydroclimatic conditions favouring floods in the south-east of Belgium by 2100 using a regional climate model, International Journal of Climatology, 41, 10.1002/joc.6642, 2021.

Xiao, Y., Huang, Q., and Wu, K.: Understanding social media data for disaster management, Natural Hazards, 79, 1663-1679, 10.1007/s11069-015-1918-0, 2015.

Yin, J., Karimi, S., Lampert, A., Cameron, M., Robinson, B., and Power, R.: Using social media to enhance emergency situation awareness, 2015,

Yu, M., Yang, C., and Li, Y.: Big data in natural disaster management: a review, Geosciences, 8, 165, 2018.

Zander, K. K., Nguyen, D., Mirbabaie, M., and Garnett, S. T.: Aware but not prepared: understanding situational awareness during the century flood in Germany in 2021, International Journal of Disaster Risk Reduction, 96, 10.1016/j.ijdrr.2023.103936, 2023.

Zou, L., Lam, N. S. N., Cai, H., and Qiang, Y.: Mining Twitter Data for Improved Understanding of Disaster Resilience, The Annals of the American Association of Geographers, 108, 1422-1441, 10.1080/24694452.2017.1421897, 2018.