# Fusion of geospatial information from remote sensing and social media to prioritise rapid response actions in case of floods

Marc Wieland[1*], Sebastian Schmidt[2], Bernd Resch[2,3], Andreas Abecker[4], Sandro Martinis[1]

[1] *German Aerospace Center (DLR), German Remote Sensing Data Center (DFD)*

[2] *University of Salzburg, Department of Geoinformatics – Z_GIS*

[3] *Harvard University, Center for Geographic Analysis*

[4] *Disy Informationssysteme GmbH*

* *Correspondence: marc.wieland@dlr.de*

**Abstract**

Efficiently managing complex disasters relies on having a comprehensive understanding of the situation at hand. Immediately after a disaster strikes, it is crucial to quickly identify the most impacted areas to guide rapid response efforts and prioritise resource allocation effectively. Utilising early-stage estimations of impacted regions, derived from indicators such as building distribution, hazard zones or geo-social media reports, can aid in planning data collection initiatives to enhance situational awareness. Consequently, there is a need to improve the availability and accuracy of early-stage impact indicators and to integrate them into a coherent spatial and temporal analysis framework that enables identification of disaster-affected areas. In this study, a method is proposed that is tailored to quickly identifying disaster hotspots, especially in situations where detailed damage assessments or very high-resolution satellite images are not readily available. The approach leverages the H3 discrete global grid system and uses a log-linear pooling method coupled with an unsupervised hyperparameter optimization routine to fuse information on flood hazard extracted from medium-resolution satellite images with disaster-related data from Twitter and freely available supplementary geospatial data on exposed assets. The performance of the method is evaluated by comparing its outcomes against detailed damage assessments conducted during five real-world flood disasters. The results indicate that it is possible to determine the areas most affected by a flood solely based on readily available proxy information. Code and test data are available from: *https://github.com/MWieland/h3h*

**Keywords**

Information fusion; Geo-social media; Remote sensing; Flood monitoring; Disaster response

# 1 Introduction

Effective and efficient management of complex disaster situations relies heavily on comprehensive situational awareness. The lessons learned from recent disasters such as the 2021 floods in Western Germany underscore the paramount importance of timely geoinformation for safeguarding communities (Holzheimer et al. 2022). In the response phase, it is essential to quickly identify the most affected areas in order to guide rapid response actions and to decide where to focus limited resources. Commonly, this is an iterative process with continuous updates whenever new or more accurate information becomes available. While initial estimates of disaster impacts may be based on fuzzy and incomplete information, they are still relevant for providing an initial situational picture before more detailed and spatially refined damage assessments become available. Early-stage estimates of most severely affected areas based on indirect proxy information (e.g., population distribution, hazard zones, etc.) can be used to plan data acquisition campaigns to improve the situational awareness and hence the general understanding of where to focus further efforts in a targeted and efficient manner.

A plethora of studies exists that target damage assessment from multi-temporal remote sensing images (Adriano et al. 2021; Zheng et al. 2021; Wiguna et al. 2024). These approaches provide detailed per-building assessments and are almost exclusively based on very high-resolution aerial or satellite images (with ground sampling distance <=1m) that have been acquired on-demand over specifically targeted areas (Voigt et al. 2016). Similarly, the rapid damage mapping protocols employed by the Copernicus Emergency Mapping Service (EMS) (Ajmar et al. 2017) are built on the premise that a preliminary selection of the most impacted regions has already been conducted before initiating the rapid mapping activation. However, particularly in the initial phase of a disaster, this assumption is often improbable, leading to delays between the onset of a disaster and the availability of map products. This raises the fundamental question: How can the most likely affected areas be detected before initiating on-demand satellite data acquisitions and planning aerial or in-situ surveys?

A viable solution is to analyse remote sensing images from systematically acquiring satellites such as Landsat, Sentinel-1 or Sentinel-2 (with ground sampling distance between 10 and 30 m) (Aimaiti et al. 2022). Putri et al. (2022) conclude that damage grading can only be accurately achieved using higher spatial resolution images. Hence, studies based on medium-resolution imagery commonly focus on the large-scale delineation of the hazard extent (Chini et al. 2021; Martinis et al. 2022; Krullikowski et al. 2023). Although high accuracies are being reported for flood hazard delineation and systematically acquiring satellite missions enable timely product delivery, solely focusing on hazard assessment for prioritisation is inadequate. In order to identify areas that require rapid response actions, one must also consider the elements at risk, such as buildings and population. In addition, not all disaster impacts can directly be observed from the top-view of remote sensing data.

Geo-social media data, i.e., social media posts containing a geographic reference such as a GNSS position or a location mention, can complement this by providing ground-level information. This can involve both text and image content, which can offer information about the start, course and aftermath of an event. A major advantage of the data is its high temporal resolution, which is characterised by a constant data stream and its near real-time availability

(Fohringer et al. 2015). Most of the research focuses on the automatic extraction of topics from social media posts. Traditionally, keyword-based methods were employed for this purpose. Later, strategies based on supervised classification or unsupervised topic modelling were proposed. For the latter, Latent Dirichlet Allocation (LDA) (Zhou and Chen 2014; Resch et al. 2018; Havas and Resch 2021) was often used. With the recent surge of deep-learning methods, more complex methods have become more frequent. In this context, Convolutional Neural Networks (CNN) (Huang et al. 2020), Bidirectional Encoder Representations from Transformers (BERT) (Adwaith et al. 2022), and Graph Neural Networks (GNN) (Papadimos et al. 2023) have become popular. Classification tasks generally range from simple binary categorisation (Hanny et al. 2024) to more complex cases, such as relevance classification of textual content (Powers et al. 2023; Blomeier et al. 2024). While there is a clear focus on textual data, there is also a considerable amount of research on automated image analysis from social media, also in the disaster management context (Barz et al. 2021; Kamoji and Kalla 2023). However, the structural and content-related diversity of the data still poses methodological challenges. In the case of disaster management, a potential problem is data scarcity in affected areas (Guo et al. 2023).

The fusion of multiple modalities from remote sensing, social media and other geodata is still a largely unsolved research problem. Havas et al. (2017) propose a conceptual approach towards leveraging information based on remote sensing, geo-social media and crowdsourcing to support disaster management in near real time. A central research question in this context is how information layers with different semantic meaningfulness, spatial resolutions and temporal delays can be merged (Li et al. 2017). While several studies fuse multi-modal data to improve landcover classifications (Cao et al. 2020) or flood hazard delineation (Rosser et al. 2017; Huang et al. 2018), only few studies aim at identifying disaster impacts. (Florath et al. 2024) combine Twitter (now officially X) data with hurricane trajectories in an Extremely Randomized Tree regression approach to estimate the impact areas of two hurricanes. Wang et al. (2018) propose an information fusion approach from remote sensing and social media for the identification of very large floods. For this, they homogenise their data using the Least effort and Maximum Entropy principle and test their methodology on the 2013 Boulder flood. Liu et al. (2021) propose a method called geographic optimal transport to align spatial representations of remote sensing-derived indices and keyword-filtered Twitter data. Yang et al. (2022) superimpose SAR imagery with posts from Sina Weibo to identify disaster affected areas.

Geostatistical methods, such as composite indicators, have been used to fuse geospatial information layers, and particularly to generate proxies for risk estimation at regional and global scales (Nardo et al. 2005; Nadim et al. 2006; Peduzzi et al. 2009). Pittore 2015 extends the concept of composite indicators and introduces focus maps as a means of prioritising data collection for efficient seismic risk assessment. Focus maps aim to implement a joint density of sampling probability for information collection that is conditionally dependent on the indicators themselves. The concept is based on a probability pooling of normalised and weighted input indicators. Also, supervised machine learning approaches (Snidaro et al. 2015; Poria et al. 2017; Avgerinakis et al. 2018) and methods based on Bayesian networks (Muesing et al. 2019) have been proposed in this context. Supervised machine learning methods are, however, constrained to often limited availability of training data and tend to lack transparency, which hampers their generalisation ability and application in disaster management.

Consequently, there is a need to enhance the availability and quality of early-stage impact proxy information layers and their integration into a spatially and temporally consistent framework that enables the rapid identification of areas that are most affected by a disaster. The aim of this study is to cover this need and propose an unsupervised method to identify disaster hotspots that is specifically targeted towards situations where detailed damage assessments and on-demand very high-resolution satellite or aerial images are not (yet) available. The work centres around the following research questions (RQ):

***RQ1****: Is it possible to determine the areas most affected by a flood solely based on readily available proxy information to guide on-demand satellite data acquisitions, aerial or in-situ survey campaigns?*

***RQ2****: Can flood-related information layers with different semantic meaningfulness, spatial resolutions and temporal delays be fused in a spatially and temporally consistent framework?*

***RQ3****: What is the contribution of different proxy information layers from remote sensing, social media and other geodata to the identification of flood hotspots?*

The proposed method fuses information about flood hazard derived from systematically acquired medium-resolution Sentinel-1 (radar) and Sentinel-2 (multi-spectral) satellite images with disaster-relevant, georeferenced posts from Twitter and freely available ancillary geodata about exposed assets. The concept of focus maps (Pittore 2015) is combined with the H3 grid system (Sahr 2011) for a spatially explicit multi-scale information fusion. The performance of the method is evaluated against detailed damage assessments from rapid mapping activations of diverse real-world flood disasters. Code and test data are published openly and can be accessed from the following source: *https://github.com/MWieland/h3h*

# 2 Study areas and data

The study areas were globally sampled and covered a large variety of land-cover classes and languages (Figure 1). For five distinct flood disasters, Twitter data (Section 2.1), flood maps from remote sensing imagery (Section 2.2) and crowd-sourced building locations (Section 2.3) were collected. For subsets of the study areas, per-building damage data from Copernicus EMS and an openly available damage detection benchmark dataset have been acquired (Section 2.4). Table 1 provides an overview of data availability and disaster extent per study area. While the selection of study areas was generally limited by the availability of high-quality damage data (for evaluation purposes), the aim was to test the global applicability of the methods under varying environmental, cultural and geographic conditions.

**USA, August 2017**: On 25 August 2017, hurricane Harvey made landfall on the coast of Texas. Substantial rainfall resulted in widespread flooding across south-eastern Texas, affecting regions such as Fort Bend, Brazoria, Galveston and Harris County along with extensive areas of Houston. The Copernicus EMS (EMSR229) was activated on 15 March 2019 for selected areas in the greater Houston area. Beyond rapid mapping efforts, extensive damage mapping has been conducted as part of compiling the xBD remote sensing benchmark dataset (Gupta et al. 2019).

**Mozambique, March 2019**: During the night of 14 March, tropical cyclone Idai made landfall near Beira, Mozambique causing major infrastructural damages as a result of the combined effects of heavy rainfall, strong winds and storm surge. Following the cyclone, a significant humanitarian crisis emerged, demanding urgent assistance for hundreds of thousands of people in Mozambique and Zimbabwe. The Copernicus EMS (EMSR348) was activated on 15 March 2019 to produce damage maps for the most severely impacted areas in and around the city of Beira.

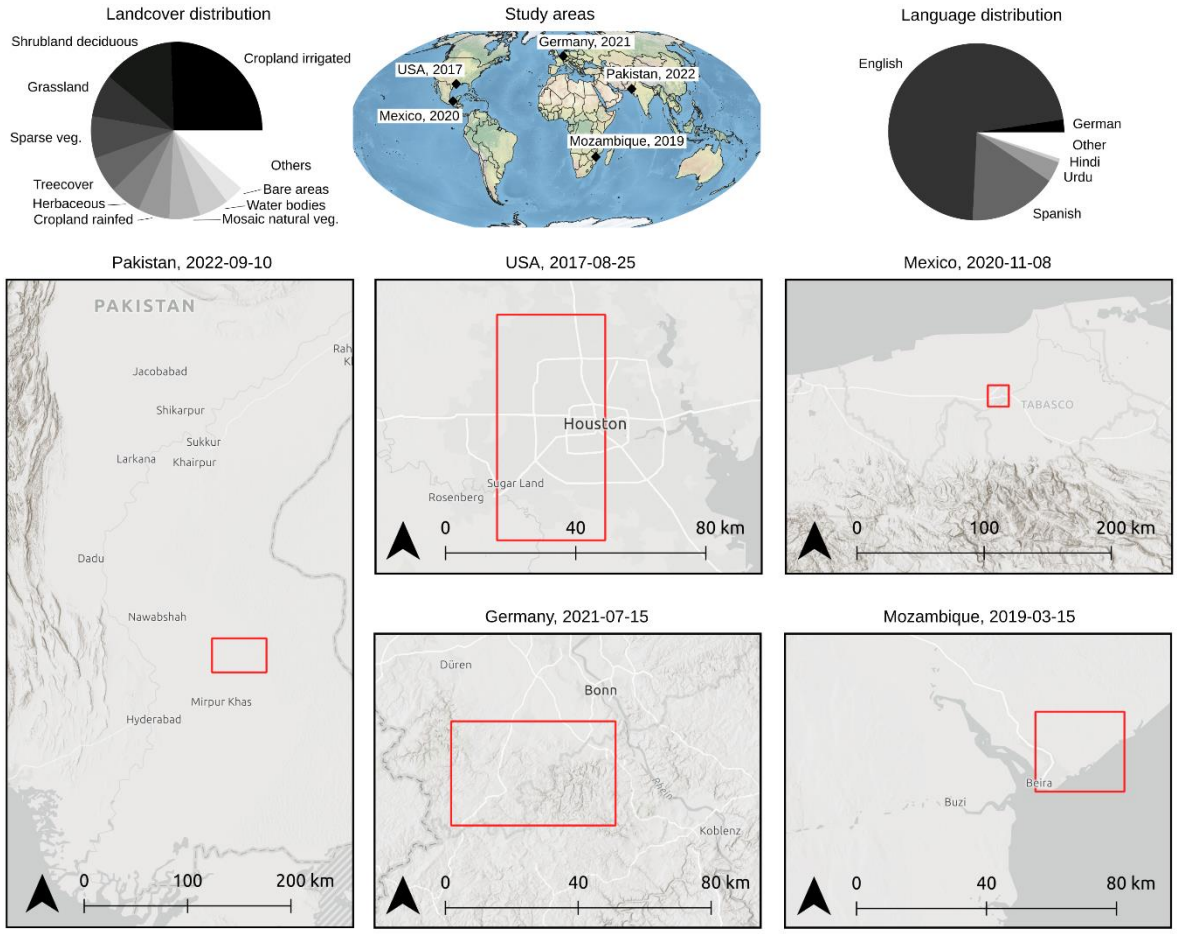


*Figure 1: Overview of the location of the study areas with dates of respective flood disasters as well as distribution of predominant land-cover and languages. For subsets of the study areas per-building damage data is available for evaluation (red bounding boxes).*

**Mexico, November 2020**: During October and November 2020, a sequence of cold fronts accompanied by two cyclones led to significant flooding in Chiapas, Tabasco, and Veracruz, Mexico. These events submerged vast areas, impacting approximately 800,000 people, damaging more than 200,000 houses and flooding thousands of hectares of farmland. The Copernicus EMS (EMSR479) was activated on 8 November 2020 to produce damage maps for selected areas in Villahermosa.

**Germany, July 2021**: From 14 to 15 July 2021, torrential rainfall caused severe flooding in the German states of North Rhine-Westphalia and Rhineland-Palatinate. Particularly affected was the Ahrweiler district in Rhineland-Palatinate, where the Ahr river surged, devastating numerous structures and claiming more than 110 lives. In response to this, emergency

mapping was conducted jointly by DLR's Centre for Satellite-based Crisis Information (ZKI) and the Copernicus EMS (EMSR517).

**Pakistan, September 2022**: From June to October 2022, heavier than usual monsoon rains and melting glaciers due to a severe heat wave caused some of the most severe floods in Pakistan's recent history. The Copernicus EMS (EMSR631) was activated on 10 September 2022 to produce damage maps for selected areas in Sindh province. The area of interest centres around the most affected Sindh and Balochistan provinces.

*Table 1: Data availability per study area.*

| Study area | Date<br>Mapping activation | Coverage<br>km² | Damage<br>Mapped buildings | Social media<br>Relevant tweets | Remote sensing<br>Flooded km² | Buildings<br>Total |
|---|---|---|---|---|---|---|
| **USA** | 2017-08-25 | 6,477 | 37,373 | 12,571 | 134 | 1,714,895 |
| **Mozambique** | 2019-03-15 | 5,820 | 17,019 | 49 | 825 | 236,825 |
| **Mexico** | 2020-11-08 | 40,045 | 11,706 | 2,434 | 214 | 2,435,987 |
| **Germany** | 2021-07-15 | 5,130 | 7,721 | 548 | 35 | 569,314 |
| **Pakistan** | 2022-09-10 | 183,000 | 9,480 | 2,800 | 20,584 | 11,772,478 |

## 2.1 Geo-social media

The social media platform Twitter provides georeferenced data via various Application Programming Interface (API) endpoints. Following the approach of Havas et al. (2021), Tweets were retrieved via the REST and streaming API. In a first step, this data was filtered based on the bounding box of the respective area of interest (Figure 1) and a predefined time frame. Table 2 shows an overview of the filtering criteria and data. The important attributes for the study provided by the APIs were the date, the text and the georeference of each Tweet. The latter can be represented as a point or as the bounding box of a "place", which is a location that the user assigned purposefully.

*Table 2: Overview of acquired Twitter data.*

| Study area | Date<br>Mapping activation | Date range<br>Time-series (data acquisition) | Tweets<br>Total | Disaster-related Tweets<br>Total |
|---|---|---|---|---|
| **USA** | 2017-08-25 | 2017-08-01 - 2017-09-14 | 160,975 | 12,571 |
| **Mozambique** | 2019-03-15 | 2019-03-02 - 2019-03-31 | 95 | 49 |
| **Mexico** | 2020-11-08 | 2020-10-25 - 2020-11-29 | 48,604 | 2,434 |
| **Germany** | 2021-07-15 | 2021-07-01 - 2021-07-31 | 11,177 | 548 |
| **Pakistan** | 2022-09-10 | 2022-09-03 - 2022-09-16 | 33,752 | 2,800 |

## 2.2 Remote sensing

With the launch of Sentinel-1 (April 2014 and April 2016) and Sentinel-2 (June 2015 and March 2017) satellites, comprehensive monitoring of surface water dynamics became feasible on a large scale, offering high spatial resolution (approximately 10 to 20 m ground sampling distance), frequent revisits (approximately every 2 to 6 days depending on location), and wide

swath coverage (exceeding 250 km). Sentinel-1 satellites are equipped with a synthetic aperture radar (SAR) C-band sensor that allows them to penetrate through clouds and image day and night. Sentinel-2 satellites carry a multispectral instrument that captures data across 13 spectral bands, spanning visible, near-infrared, and short-wave infrared ranges. In this study, data from both Sentinel-1 and Sentinel-2 satellites was used to delineate surface water bodies and to differentiate between temporary flooded areas and permanent water bodies. For each flood, the satellite images that were closest to the date of the respective rapid mapping activation were acquired and used as basis to outline the maximum observed water extent. A time-series of satellite images acquired over a two-year period prior to the disaster date was used to identify permanent water bodies under consideration of seasonal variations. Table 3 provides an overview of the satellite images used for each study area.

*Table 3: Overview of acquired remote sensing data.*

| Study area | Date<br>Mapping activation | Date<br>Satellite image (flood) | Date range<br>Time-series (permanent water) | Sentinel-1<br>Total number of satellite images | Sentinel-2<br>Total number of satellite images |
|---|---|---|---|---|---|
| **USA** | 2017-08-25 | 2017-08-30 | 2015-01-01 - 2016-12-31 | 364 | 518 |
| **Mozambique** | 2019-03-15 | 2019-03-19 | 2017-01-01 - 2018-12-31 | 339 | 535 |
| **Mexico** | 2020-11-08 | 2020-11-08 | 2018-01-01 - 2019-12-31 | 387 | 594 |
| **Germany** | 2021-07-15 | 2021-07-15 | 2019-01-01 - 2020-12-31 | 375 | 601 |
| **Pakistan** | 2022-09-10 | 2022-09-11 | 2020-01-01 - 2021-12-31 | 1362 | 4681 |

## 2.3 Crowd-sourcing and open geodata

Crowd-sourced datasets like OpenStreetMap (OSM) (OpenStreetMap 2024) can provide invaluable high-quality information about location and characteristics of exposed buildings and infrastructure worldwide. However, large variations in completeness and quality of crowd-sourced datasets across geographic regions often hamper their usability for analyses that require globally consistent data quality and completeness (Barron et al. 2014; Herfort et al. 2023). Therefore, building footprints from OSM were extended with Microsoft's Building Footprints (Microsoft building footprints 2024) and Google's Open Buildings (Google open buildings 2024) datasets on a case-by-case basis for each study area. After thorough visual quality and completeness checks between building footprints and reference satellite imagery prior to the respective flood disaster, OSM data was extended with Google's Open Buildings for the study areas in Pakistan and Mexico. For the study area in the USA, OSM data was extended with Microsoft's Building Footprints. The study areas in Germany and Mozambique were already completely covered by OSM. To combine different datasets, a fuzzy location matching criterion was applied on the centroids of the OSM building geometries. All buildings of the supplementary dataset whose centroid fell outside of a 10 m buffer to the OSM buildings were selected. An overview of the acquired buildings per study area is provided in Table 1.

## 2.4 Per-building damage data

Building damage mapping by means of in-situ surveys or visual interpretation of very high-resolution optical remote sensing data is a common procedure to estimate the impact of a disaster. In this study, readily available damage data from Copernicus EMS (Copernicus

Emergency Management Service - Mapping 2022) and the xBD dataset (Gupta et al. 2019) was used as reference to validate and test the method. For the study areas in Germany, Mexico, Pakistan and Mozambique reference data from Copernicus EMS was derived. For the USA study site, an excerpt of the xBD dataset is used. Per-building damages in these datasets are categorised on the basis of ordinal classification schemes (e.g., 1: possibly damaged, 2: damaged, 3: destroyed).

# 3 Method

Figure 2 provides an overview of the methodological framework of this study. Twitter data is used to identify geo-social media hotspots (Section 3.1), flood maps are computed on the basis of Sentinel satellite images (Section 3.2), and OpenStreetMap as well as other building footprint data define the distribution of exposed assets. Copernicus EMS and xBD provide observed damage data as independent reference. All data layers are aggregated to the hexagonal H3 grid system (Section 3.3), normalized, and then combined using log-linear pooling to predict flood hotspots (Section 3.4). Methods to choose hyperparameters with and without reference data are introduced (Section 3.5) and the predicted hotspots are finally compared with observed hotspots for performance evaluation (Section 3.6).

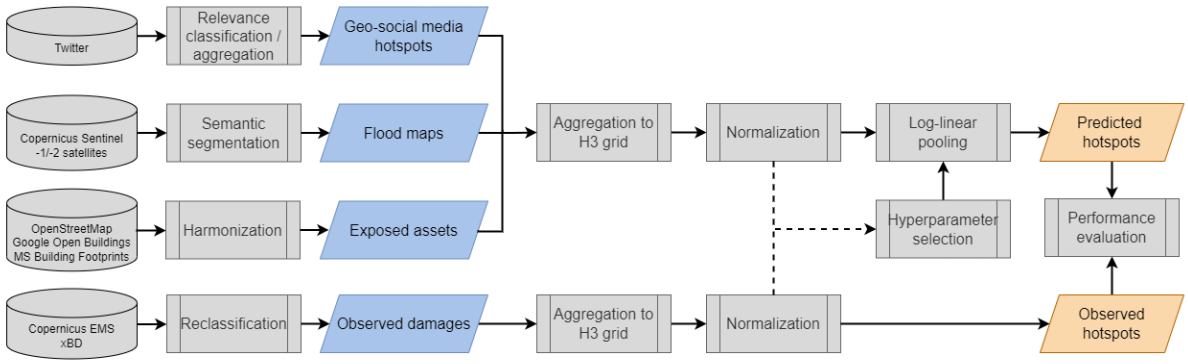


*Figure 2: Flowchart of the methodological framework.*

## 3.1 Hotspot detection from geo-social media

To identify Tweets directly related to flooding, the model developed by Hanny et al. (2024) was used. It is a fine-tuned Twitter-XLM-RoBERTa-base model that was improved by applying an active learning approach. It classifies Tweets solely on the basis of their texts into the categories “unrelated” and “related” with regard to natural disasters. It yielded an accuracy of 0.94 on a test data set based on CrisisLex. The results of the classification were then aggregated on regular grids with a spatial resolution of 5 km by calculating the sum of all Tweets and all disaster-related Tweets within the respective cell. Based on these aggregated results, a spatial hot spot analysis was performed. For this, the *Getis-Ord* $G_i^*$ algorithm (Equation 1) was used which considers the values of spatially adjacent observations. In this study, spatial adjacency was defined by a Queen contiguity spatial weights matrix, to detect hot and cold spots, i.e. statistically significant areas with particularly high or low occurrences of the phenomenon under investigation. With respect to the application of this study, this referred to the ratio of disaster-related Tweets on all Tweets in a region.

$$G_i^* = \frac{\sum_{j=1}^{n} w_{i,j} x_j - \bar{X} \sum_{j=1}^{n} w_{i,j}}{S \sqrt{\frac{\left[ n \sum_{j=1}^{n} w_{i,j}^2 - \left( \sum_{j=1}^{n} w_{i,j} \right)^2 \right]}{n-1}}} \quad (1)$$

where $x_j$ is the attribute value of the feature $j$, $w_{i,j}$ is the spatial weight between feature $i$ and $j$, $n$ is the sample size, $X$ is the mean and $S$ the standard deviation of the data set.

## 3.2 Flood mapping from satellite images

A modular processing chain was deployed for surface water monitoring from Sentinel-1 and Sentinel-2 satellite images that handles satellite data search, preprocessing, analysis and dissemination over a predefined area of interest (Wieland and Martinis 2019). Images are analysed using sensor-specific pre-trained CNN for semantic segmentation to extract binary water masks. The network architectures utilize a U-Net decoder and combine it with an Efficientnet-B4 encoder. The networks were trained on an extended version of the freely available S1S2-Water dataset (Wieland et al., 2023) with an AdamW optimizer, initial learning rate of 1e-3, weight decay of 1e-2 and a weighted combination of binary cross entropy and Lovász Hinge loss. As input feature space for Sentinel-1, VV and VH polarizations as well as slope information derived from the Copernicus Digital Elevation Model (DEM) were used. Similarly, for Sentinel-2 six spectral bands (Red, Green, Blue, Near-Infrared, Shortwave-Infrared-1, Shortwave-Infrared-2) and slope information were used as input features to train the network. As reported by Bereczky et al. (2022) and confirmed by Wieland et al. (2023), models trained on data showing no distinct inundation perform well in mapping the water extent during flood events, reaching Intersection over Union (IoU) scores of >0.8 for Sentinel-1 and >0.9 for Sentinel-2. Time-series analysis was used to further distinguish permanent and seasonal water bodies from temporary, potentially hazardous flooded areas. In this study, followed Martinis et al. (2022) who specifically consider seasonality of surface water dynamics. The seasonal reference water masks for the study areas were computed over a reference time period of two years, for which all available Sentinel-1 and Sentinel-2 images were analysed.

$$y_{flood} = y_{water} \backslash r \quad (2)$$

The set difference of the water mask $y_{water}$ derived from the satellite image that depicts the flood event and the reference water mask $r$ produces the flood mask $y_{flood}$ that is being used for further analysis (Equation 2).

## 3.3 Aggregation on H3 discrete global grid system

In this study, the H3 grid system that features a hierarchical structure with 16 spatial resolutions and a 64-bit unsigned unique integer index for each cell is used (Sahr 2011). The grid cells have a hexagonal shape, with uniform distances between the centroid of each hexagon and the centres of its neighbouring cells. The choice of the H3 resolution level

depends on the specific application as well as the extent and scale of the study area under consideration. Since this study focused on regional to national scale analysis, H3 resolution levels 6, 7 and 8 were used, which correspond to average edge lengths of approximately 3.7 km (roughly city size), 1.4 km (roughly neighbourhood size) and 0.5 km (roughly block size). To convert the proxy information layers into H3 grid cells at the desired resolution levels, specific import routines were devised depending on the format of the input data. Flood masks were stored as binary raster layers and as such were converted by summing the positive raster values inside each H3 grid cell, therefore providing a count of flooded pixels per grid cell. Twitter hotspot maps were provided as vector polygon layers and converted to H3 by assigning the polygon value at the centroid of each H3 grid cell. Exposed buildings were provided as vector point layers and converted to H3 by counting the number of buildings per H3 grid cell. Similarly, the per-building damage datasets were provided as vector point layers with ordinal damage categories. Summing up the per-building damage grades provides an aggregate measure of damage per H3 grid cell. The H3-Pandas Python package (H3-Pandas 2024) was used to handle all H3 specific geoprocessing operations. Converted input data were stored in a H3-indexed Geopandas dataframe for further analysis.

## 3.4 Information fusion

Following the concept of focus maps, as outlined by Pittore (2015), the fusion of heterogeneous proxy information layers from remote sensing, social media and other geodata was approached with the overall aim to derive a joint probability of sampling that is conditionally dependent on the input layers themselves. Within this sampling framework, let $D_i$ be the set of proxy information layers that are relevant for the prioritization of disaster hotspots and that are aligned on a two-dimensional geographical grid ($G$) with the same extent, origin and grid resolution (Equation 3). In this study, $G$ is defined by the H3 grid cells that span an area of interest at a predefined resolution level.

$$D_i = D_i(x, y) \in [0, \infty], (x, y) \in G \tag{3}$$

The conditional probability of sampling given a single proxy layer $P(S|D_i)$ at a specific location can be defined by a simple normalization of its values to the range [0,1], where larger values indicate higher relevance of a location for the disaster under consideration. Here a *minmax* normalization (Equation 4) was applied and different truncations were tested based on quantile intervals to account for outliers.

$$P(S|D_i) = \frac{D_i - \min(D_i)}{\max(D_i) - \min(D_i)} \tag{4}$$

To approximate the joint probability of sampling and thus the prioritization of disaster hotspots given more than one proxy layer $D_i$, an appropriate pooling operator $PG$ is required. Here a log-linear pooling operator is used, which performs multiplicative pooling that emphasizes the locations where all input layers indicate a higher probability of sampling (Equation 5).

$$P(S|D_0, D_1, \dots, D_n) \approx \ln PG\big(P(S|D_0), P(S|D_1), \dots, P(S|D_n)\big) = \ln Z + \sum_{i=1}^{n} w_i \ln P(S|D_i) \quad (5)$$

The property of a log-linear pooling that it directly prioritizes locations with matching input layers makes it particularly suitable for the identification of disaster hotspots. Furthermore, by defining pooling weights $w_i$ one is able to control the relative importance of proxy information layers on the result.

## 3.5 Hyperparameter optimization

The main hyperparameters that are required by the information fusion method are the weights $w_i$ of the proxy information layers that are used as input to the log-linear pooling (Equation 5). The weights basically control the relative importance of the input layers on the result and allow to vary the discriminating effect of the pooling, making the results more or less selective on the specific input. Other hyperparameters to consider were the lower and upper quantiles $[q_l, q_u]$ at which the values during normalization of the input layers were truncated (Equation 4). Using truncation with a quantile interval can enhance the robustness of the normalization. Values falling outside the defined interval were adjusted to the boundary values. This effectively removed the extreme tails of the distribution, which would otherwise have exerted a disproportionate influence on the resulting probability $P(S|D_i)$.

In this study, different methods were compared to select hyperparameters depending on the availability of reference data (e.g., observed damage distributions). The base assumption of this work was that in a real disaster situation no reference data would be available and the information fusion method would be deployed in an unsupervised manner (Section 3.5.1). However, also the possibility to learn hyperparameters prior to an application based on observed damage data from other disasters was tested (Section 3.5.2).

### 3.5.1 Selecting hyperparameters without reference data

Optimizing log-linear pooling weights without reference data typically involves relying on theoretical or subjective criteria. A widely used approach is to assign equal weights to each input layer, which assumes that each source is equally reliable and valuable. Equal weights ensure a neutral and balanced combination of probability distributions, especially when there is no reason to favour one source over another. Commonly, equal weights are assigned such that they sum up to one. Applying weights that sum up to greater than one is possible in the context of log-linear pooling and effectively increases the discriminating effect of the pooling, thus making the results more selective (Pittore 2015). If prior knowledge about the reliability or relevance of each input layer is available, pooling weights can also be assigned subjectively with higher weights referring to more reliable or relevant layers.

An objective method to optimize log-linear pooling weights without reference data is to minimize the sum of the Kullback-Leibler (KL) divergences from each input layer distribution to the pooled distribution (de Carvalho et al. 2023). The KL divergence from an input distribution $P(S|D_i)$ to the pooled distribution $P(S|D_1, D_2, \dots, D_n)$ is defined as follows:

$$D_{KL}(P(S|D_i)||P(S|D_1, D_2, \dots, D_n)) = \sum_{x \in X} P(S|D_i)(x) log \frac{P(S|D_i)(x)}{P(S|D_1, D_2, \dots, D_n)(x)} \quad (6)$$

The objective is to find the weights $w_1, w_2, \dots, w_n$ that minimize the sum of the KL divergences from each input distribution to the pooled distribution, while $w_i \geq 0$ (Equation 7).

$$\min_w \sum_{i=1}^{n} D_{KL}(P(S|D_i)||P(S|D_1, D_2, \dots, D_n)) \quad (7)$$

This objective function was chosen over maximizing entropy, because it leads to a unique solution as detailed and proofed in de Carvalho et al. (2023).

### 3.5.2 Selecting hyperparameters with reference data

A supervised approach was tested to optimize hyperparameters according to a prior distribution from observed damage data. In this setup, the available reference data is split into training and validation sets and cross-validation is used to find the weights resulting in the best predictive performance on the validation set. Grid search is deployed as a transparent and reproducible approach to find optimal hyperparameters for normalization and log-linear pooling. The results of the grid search were, moreover, used for an explorative analysis of the hyperparameter space. For the task at hand, the search space for $w_i$ is defined as follows:

$$w_i \in \{0.0, 0.25, 0.5, 0.75, 1.0, 1.25, 1.5, 1.75, 2.0\} \quad (8)$$

For the normalization quantile interval, the following search space is tested:

$$[q_l, q_u] \in \{[q_{0.02}, q_{0.98}], [q_{0.01}, q_{0.99}], [q_{0.0}, q_{1.0}]\} \quad (9)$$

The actual calculation of $q_l$ and $q_u$ depends on the specific distribution of the random variable under consideration (the input layers) and can be calculated as follows:

$$q_l = F^{-1}(p_l);\ q_u = F^{-1}(p_u) \quad (10)$$

where $F^{-1}$ is the inverse cumulative distribution function of the random variable, and $p_l$ and $p_u$ the probabilities corresponding to the lower and upper quantiles.

## 3.6 Performance evaluation

Performance evaluation in this study was conducted through a quantitative approach by comparing the predicted disaster hotspots resulting from the information fusion with normalized damage distributions from reference data. Evaluation metrics were derived from an ordinary least squares (OLS) regression between predicted and true hotspot values. The adjusted coefficient of determination $R^2_{adj}$ is used to measure the proportion of variance in the dependent variable that is explained by the independent variables, with $R^2$ being the regular coefficient of determination, $n$ being the number of observations and $k$ the number of independent variables (Equation 11).

$$R^2_{adj} = 1 - \left( \frac{(1 - R^2)(n - 1)}{n - k - 1} \right) \tag{11}$$

To prepare the reference data, the true damage grades from Copernicus EMS or xBD were first converted into a compatible H3 format as described in Section 3.3. Subsequently, the damage grades were normalized with Equation 4 to ensure uniformity across the dataset. The spatial extent of the comparison is then restricted to the extent of the reference data, allowing for a focused evaluation that only considers areas where the extent of damages is known.

# 4 Results

## 4.1 Explorative analysis of hyperparameters

To explore the influence of hyperparameters on the predictions, a grid search was run across a large variety of parameter combinations, hotspot maps were computed for each of them and compared the predictions against a prior distribution from observed damage data. Figure 3 shows an example of such a comparison for a single hyperparameter combination at H3 resolution 7 over the Germany study area. Input layers in this example were normalized without truncation ($[q_{0.0}, q_{1.0}]$) and log-linear pooling was performed with equal weights ($[w_{Buildings}=1.0, w_{Flood}=1.0, w_{Tweets}=1.0]$).

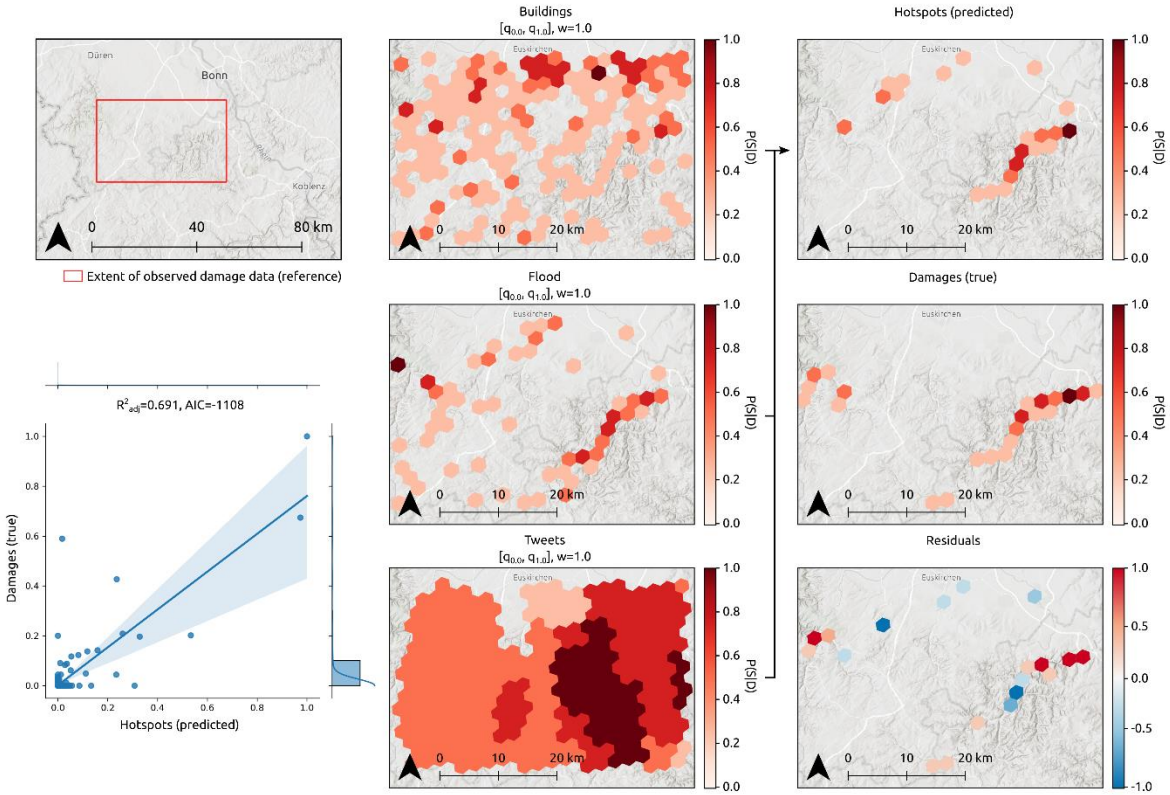


*Figure 3: Comparison of predicted hotspot map and prior distribution from observed damage data. Input layers were prepared at H3 resolution 7 and normalized without truncation.*

Across all study areas, 10,935 parameter combinations were evaluated for each of three H3 resolutions, which resulted in a total of 32,805 test runs. Figure 4 shows the distribution of $R^2_{adj}$ values over all tested hyperparameter combinations in all study areas grouped by H3 resolution. Please note that a direct comparison of regression metrics like $R^2_{adj}$ across resolutions is not meaningful, due to different spatial aggregations. For data aggregation in a different manner, the underlying data structure changes. This can affect the variability and relationships between variables, leading to different $R^2_{adj}$ values. However, for each resolution separately it was observed that a large variance of the dependent variable (here the true damages) could be predicted from the independent variables (here buildings, flood and Tweets). With the right hyperparameter settings, it was possible to achieve $R^2_{adj}$ values above 0.9 (at resolution 6), 0.8 (at resolution 7) and 0.6 (at resolution 8) and hence a very good fit of the model to the observations. The choice of the normalization quantile interval for truncation during normalization of the input layers only had minor effects on the predictive power of the models, suggesting robustness to this parameter. The spread was similar across all quantiles, indicating that normalization quantiles have minimal impact on model performance variability at each resolution. At H3 resolution 6, it was observed that normalization without truncation ($[q_{0.0}, q_{1.0}]$) resulted in slightly better fits between model predictions and observations. At resolutions 7 and 8, normalization with truncation ($[q_{0.01}, q_{0.99}]$) lead to minor improvements compared to stronger truncation ($[q_{0.02}, q_{0.98}]$) or no truncation ($[q_{0.0}, q_{1.0}]$).

Figure 5 shows the distribution of $R^2_{adj}$ values across varying log-linear pooling weights for the input layers (buildings, flood and Tweets) at different H3 resolutions. Input layers were normalized with truncation ($[q_{0.01}, q_{0.99}]$). The results indicate that the choice of pooling weights significantly impacted the $R^2_{adj}$ values for different input layers, with each layer showing a

unique pattern. For buildings (first column), the $R^2_{adj}$ values increase with higher pooling weights at all resolutions. This suggested that increasing the discriminating effect of the pooling and therefore making the results more selective for this layer improved model performance. The $R^2_{adj}$ values for the flood input layer (second column) improved with increasing pooling weights up to a point (between 0.25 and 0.75 depending on resolution) and then started to decline. This indicated an optimal level of pooling weights beyond which model performance may decrease. The results for Tweets (third column) showed a relatively stable pattern with less variation across different pooling weights, suggesting that model performance for this layer was less sensitive to the choice of pooling weights.

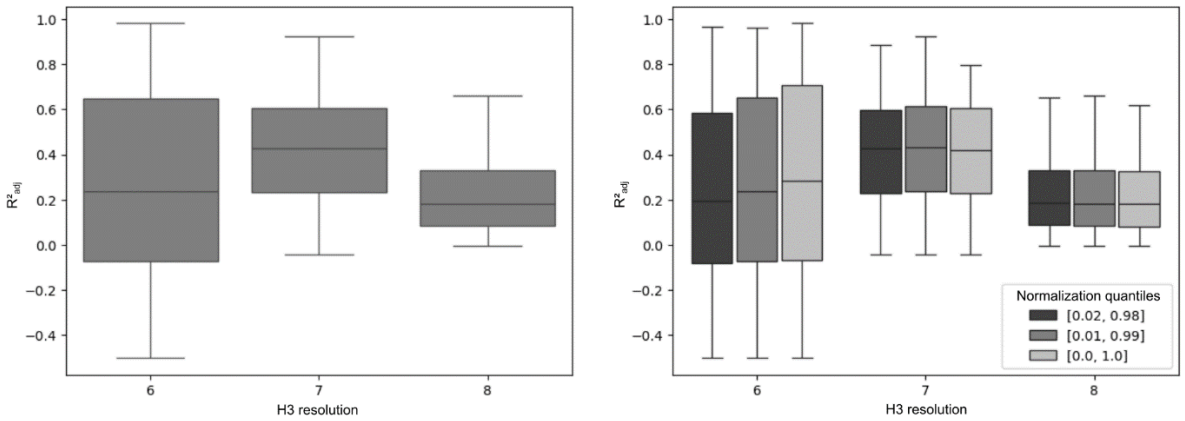


*Figure 4: Results of grid search across all tested hyperparameter combinations in all study areas grouped by H3 resolution (left) and further categorized by normalization quantiles (right).*

Figure 5 can also provide valuable insights into the relative importance of input layers on the prediction results. Layers with consistently higher $R^2_{adj}$ values were generally more important to the prediction. In this case, buildings showed the most substantial increase in $R^2_{adj}$ values with higher pooling weights, therefore suggesting high relative importance. Significant changes in $R^2_{adj}$ values with pooling weights indicated sensitivity and importance. Buildings and flood layers both showed sensitivity, with buildings improving consistently and flood having an optimal pooling weight. An input layer that maintains stable $R^2_{adj}$ values across different pooling weights suggests robustness but not necessarily importance. Tweets showed stability but less variation of $R^2_{adj}$ values across pooling weights, which indicates that they were reliable but of lesser relative importance for the prediction compared to the other layers. In summary, buildings appeared to be the most important input layer for predictions, followed by flood, with Tweets being the least impactful but still providing consistent predictive value.

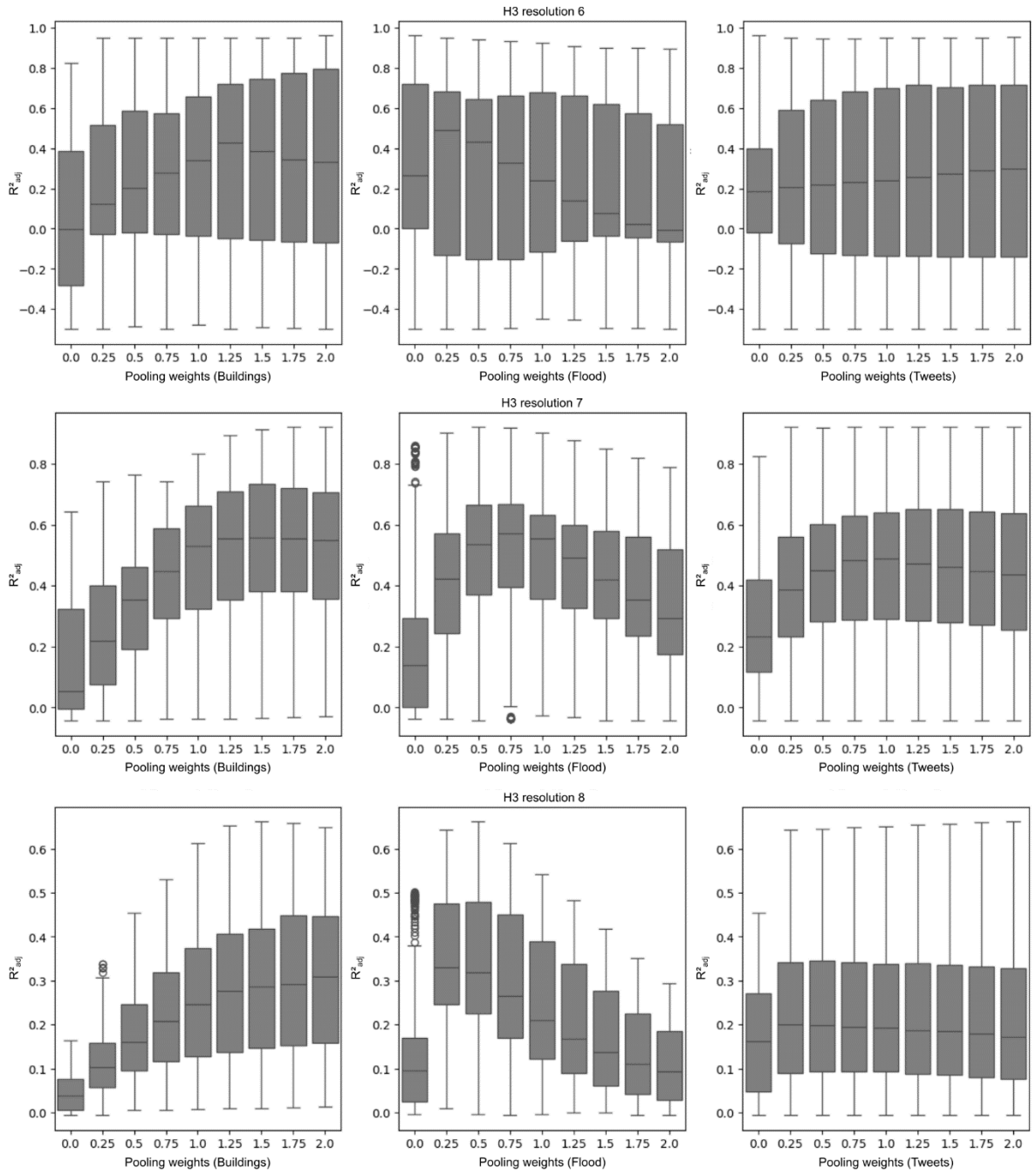


*Figure 5: Results of grid search across different log-linear pooling weights for the input layers (buildings, flood and tweets) at different H3 resolutions.*

## 4.2 Optimization of pooling weights

Figure 6 compares the performance of different methods (with and without reference) to choose log-linear pooling weights for the input layers (buildings, flood and Tweets) at H3 resolution 7. According to the results in Section 4.1, input layers were normalized with truncation ($[q_{0.01}, q_{0.99}]$) before pooling. Various approaches were tested: optimizing hyperparameters specifically for each study using grid search, assigning equal weights (0.33 and 1.0), employing supervised methods with cross-validation optimized on different study areas, and using an unsupervised method based on KL divergence.

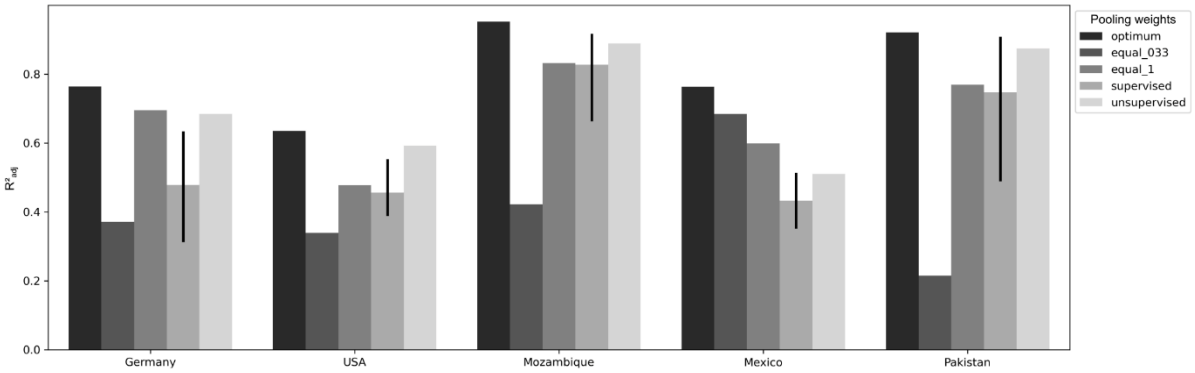


*Figure 6: Comparison of methods to choose log-linear pooling weights for the input layers (buildings, flood and Tweets) grouped by study area. Input layers were prepared at H3 resolution 7 and normalized with truncation ($[q_{0.01}, q_{0.99}]$) before pooling.*

The results indicated that while equal weight methods provided a straightforward and relatively effective approach, methods that involved optimization tended to offer better performance in terms of the $R^2_{adj}$ values. This suggests the importance of context-specific tuning for achieving the best results in log-linear pooling. Optimizing hyperparameters specifically for each study area (*optimum*) may be unrealistic for real applications due to the lack of appropriate reference data during an ongoing disaster. The results for this method, however, showed that a very good fit with observed damage distributions can be achieved solely based on the provided proxy information layers. Assigning equal weights that sum up to one (*equal_033*) showed the worst performance of all tested methods. Using equal weights of one per input layer (*equal_1*), however, performed relatively well and even outperformed optimized methods in some study areas. Cross-validation with hyperparameters optimized on different study areas (*supervised*) was competitive and outperformed the equal weights method in some cases. However, it showed large variations depending on the study areas it had been trained on. The unsupervised method based on KL divergence (*unsupervised*) showed the least performance variability and achieved good $R^2_{adj}$ values for all study areas. It was, therefore, the top-performing method in terms of best fit and stability of results, together with the equal weights method (*equal_1*).

Figure 7 shows hotspot maps that compare an observed damage distribution (*reference*) to predicted distributions using different methods for selecting log-linear pooling weights. Each map illustrates the joint probability of sampling and thus the prioritization of disaster hotspots given input proxy information layers $P(S|D_{Buildings}, D_{Flood}, D_{Tweets})$ across a geographic extent, with varying intensities of red indicating higher probabilities. Observed damages (top left) serve as the reference for comparison, where the most significant damage concentrates in the eastern part of the map, particularly around the Ahr valley and Bad Neuenahr-Ahrweiler. Optimizing weights on the same area with reference (*optimum*) closely matches the reference and further highlighted that the input proxy layers can well explain an actual damage distribution. All other methods were able to capture the general damage patterns with highest concentrations of sampling probabilities around the Ahr valley. Only the prediction using equal weights of 0.33 (*equal_033*) showed a clearly diffused and less concentrated distribution compared to the reference. Overall, the maps underlined that while equal weighting methods provided a baseline, methods that involved some form of optimization or learning (*supervised* or *unsupervised*) delivered better performance in predicting damage distributions similar to the observed reference.

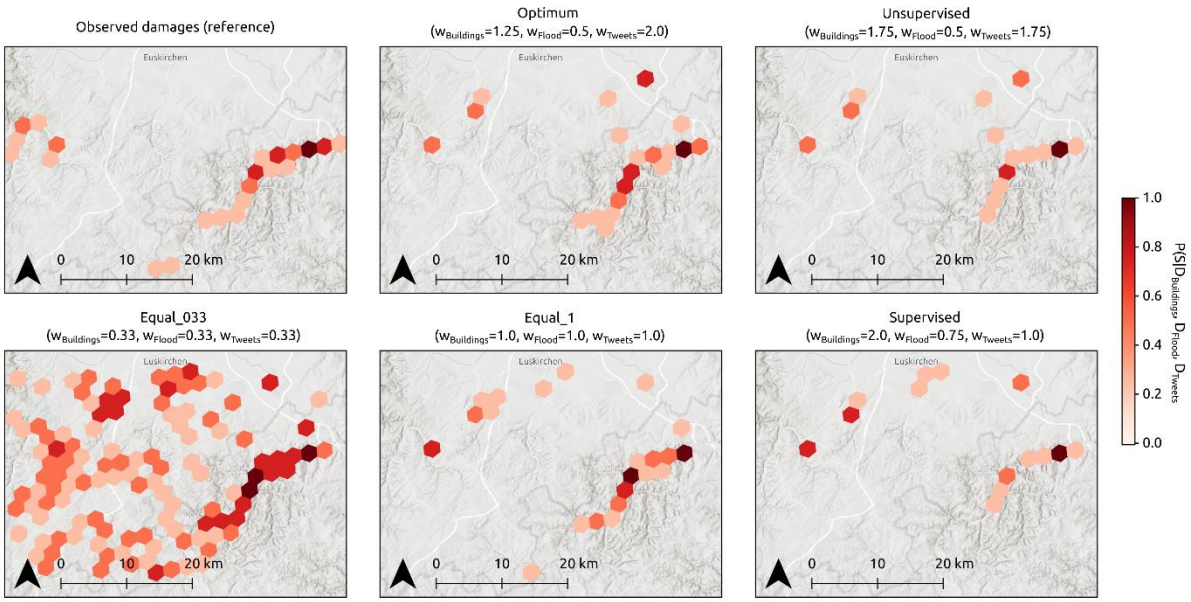


*Figure 7: Map comparison of methods to choose log-linear pooling weights for the Germany study area. Input layers were prepared at H3 resolution 7 and normalized with truncation ($[q_{0.01}, q_{0.99}]$) before pooling.*

## 4.3 Applications

In the following, the log-linear pooling method was applied to the study areas using each of them as example to test specific features that are relevant for an application in a real-world disaster situation. These include, testing the method's spatial (Section 4.3.1) and temporal (Section 4.3.2) consistency as well as its flexibility to incorporate different thematic input layers (Section 4.3.3). The whole area of interest in each study area was used, unsupervised pooling weights optimization and normalization of input layers with truncation ($[q_{0.01}, q_{0.99}]$) were applied.

### 4.3.1 Spatial consistency (Pakistan)

Figure 8 shows the Pakistan study area at different H3 resolutions. The columns represent input layers and respective predicted hotspot maps. At resolution 6, the map shows significant hotspots in the northern part of Sindh province, which aligns with the locations of Larkana and Jacobabad. There were some predicted hotspots in the central area, consistent with reported flood impact in Dadu. The southern area showed distinct hotspots centred around the highly populated greater Hyderabad region. The most prominent hotspot around Jacobabad is depicted in further detail at H3 resolution 7 and further resolved at resolution 8 for the city of Dera Allah Yar. The predicted hotspots aligned well with the reported locations of major impact during the Pakistan floods of 2022, as reported in the news and documented by the United Nations (UN-OCHA 2023). The maps, moreover, showed spatial consistency between resolutions with increasing levels of details of flood impact being revealed at higher resolution levels.

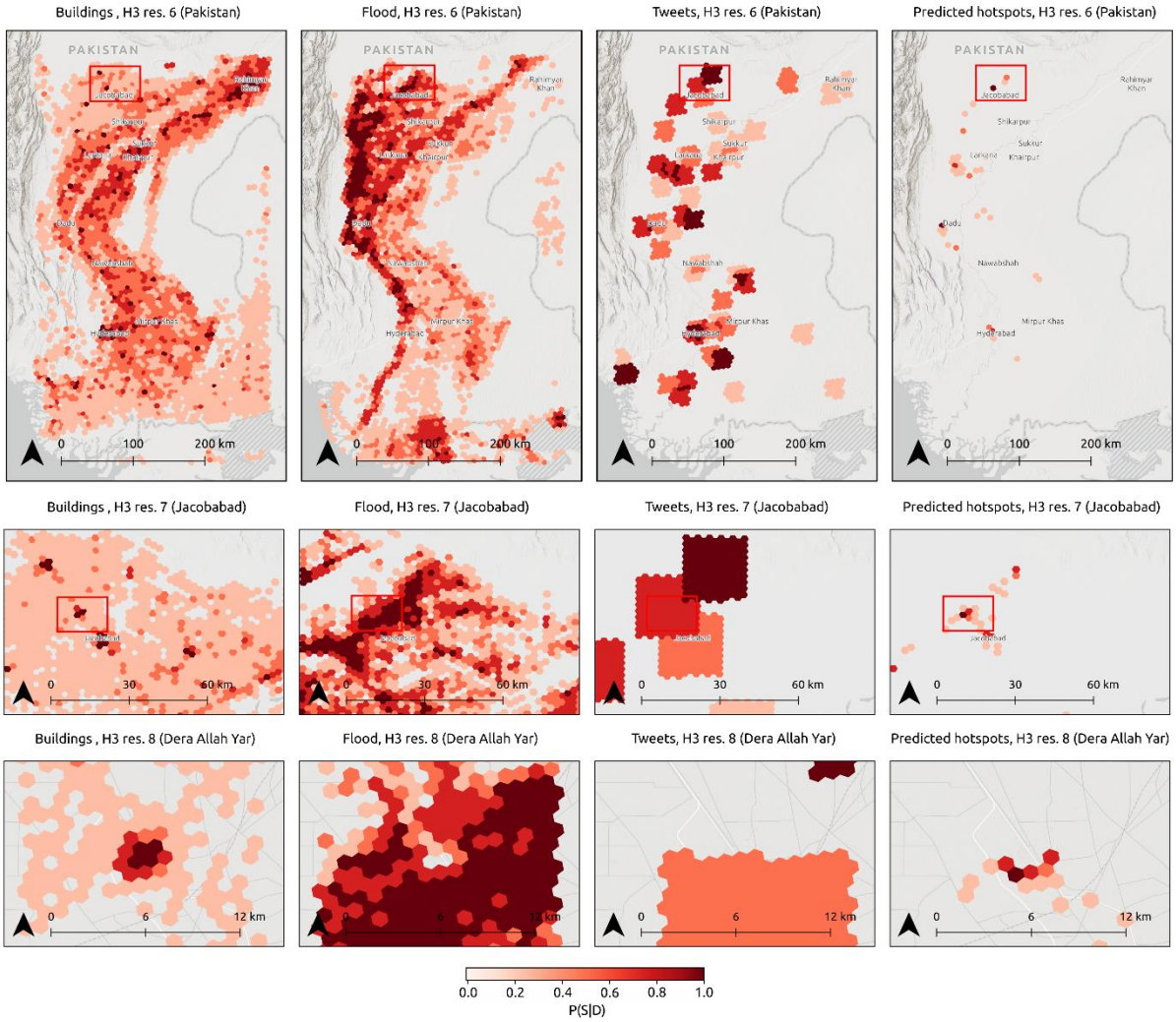


*Figure 8: Predicted hotspot maps at different H3 resolutions for the Pakistan study area.*

### 4.3.2 Temporal consistency (Germany)

Figure 9 shows the dynamic nature of the floods in Western Germany from 13 to 18 July 2021. The maps highlight the spread and intensity of the flood event at H3 resolution 7, while the time series plot captures the quantitative temporal trends in both flood extent and public attention through social media activity.

- **July 13-14, 2021**: The region experienced an unusually high amount of rainfall, with some areas recording more than 150 mm in 48 hours. The maps for July 13 and 14 show no significant hotspots, aligning with the fact that heavy rainfall and severe flooding had not yet reached critical levels.
- **July 15, 2021**: The floods reached their peak, especially affecting areas around Euskirchen (northern parts of the study area) and the Ahr Valley (central parts of the study area). The map for July 15 shows the appearance of several hotspots, corresponding to significant flooding in these areas. Respectively, the time series plot shows a strong increase in flood pixel counts and relevant Tweets. This indicated the growing severity of the situation and increasing public awareness.

- **July 16, 2021**: While the flood water extent began to decrease in most areas, the extent of damage became fully apparent. The hotspot map for July 16 keeps pointing to the most severely affected regions.
- **July 17-18, 2021**: Disaster response efforts were active in the affected areas and public attention remained high as the extent of the disaster was reported in the media. The maps for July 17 and 18 show fewer hotspots, indicating that the flood water extent further decreased while focus remained on high impact areas in the Ahr valley. The time series plot shows a significant drop in flood pixel counts, while Tweet counts remained high, indicating ongoing public engagement with the disaster.

The results confirmed that the method was able to produce temporally consistent outputs, which aligned well with the spatial distribution of actual events that happened during the disaster.

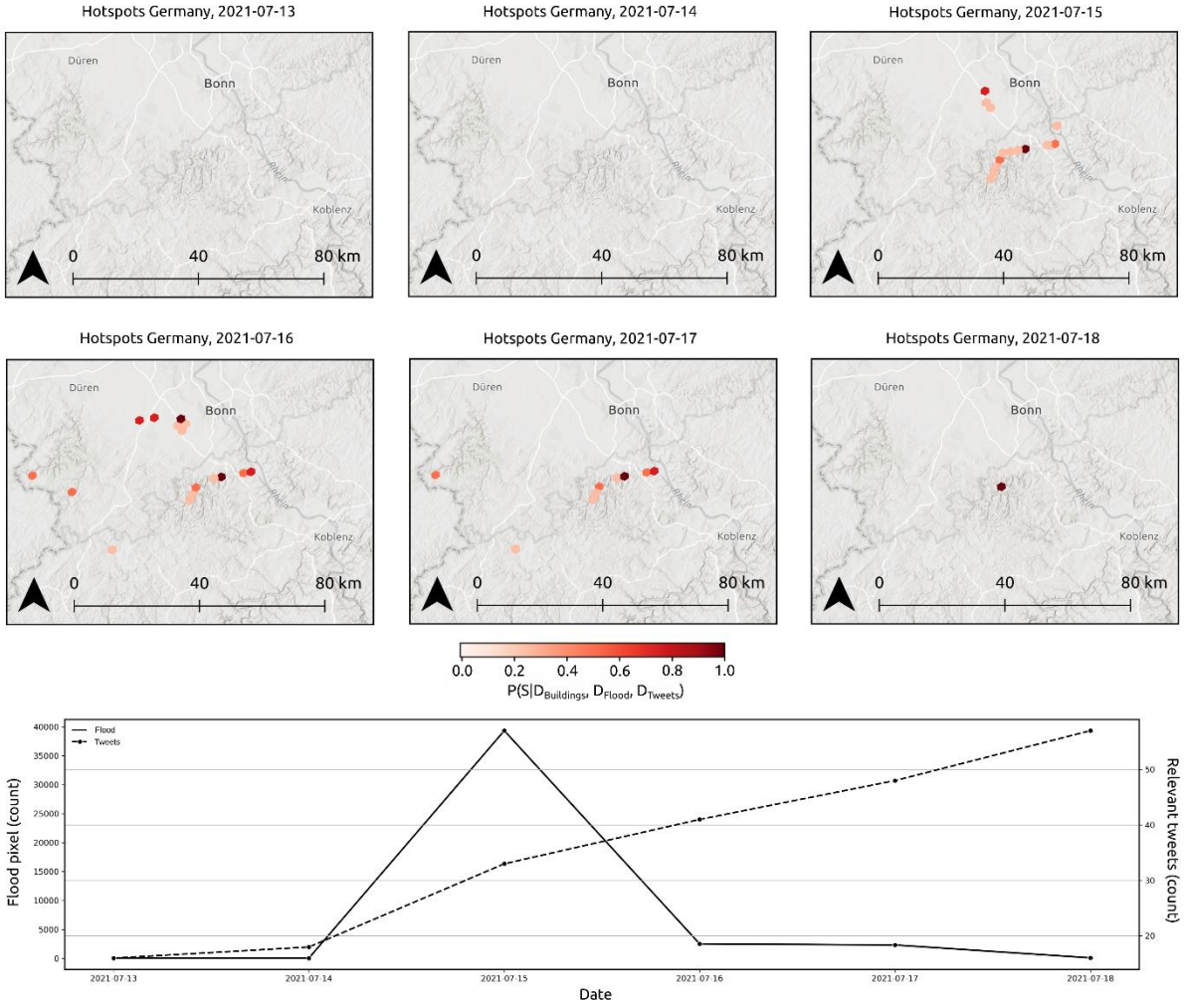


*Figure 9: Time series of predicted hotspot maps at H3 resolution 7 for Germany.*

### 4.3.3 Influence of different exposure layers (Mozambique)

Figure 10 shows the influence of choosing different datasets for mapping exposed assets using the study area of Mozambique as an example. The same flood and Tweets input layers were applied, while the source of the exposure layer was varied. It compares per-building

information from OSM with the aggregated built-up area product GHS-BUILT-S (building counts per 100 m grid cells) of the Global Human Settlement Layer (GHS) (Pesaresi and Politis 2023). Moreover, population datasets from GHS-POP (population counts per 100 m grid cells) (Schiavina et al. 2023) and Kontur (population counts per H3 resolution 7 hexagons) (Tarakanov 2020) were compared for their usability as input layer to predict hotspots. Overall, the figure indicates that the city of Beira was the most critically affected hotspot in the study area, primarily due to the high building and population densities. Despite a general agreement across the different input layers regarding most exposed regions, differences in the predicted hotspot maps became visible. In particular, when using the Kontur population dataset as input, the predicted hotspot maps were more dispersed compared to the hotspots with GHSL population input. Spatial distribution and intensity of hotspots derived with GHSL buildings and population input were well aligned with each other. The GHSL-based hotspots, moreover, closely matched the predicted hotspots on the basis of OSM data.

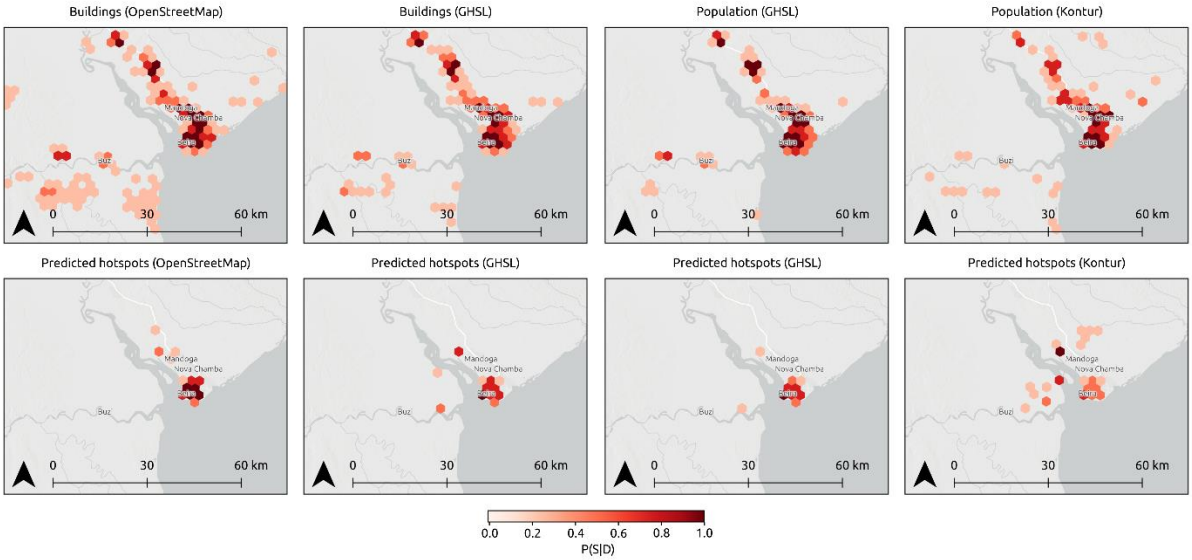


*Figure 10: Predicted hotspot maps at H3 resolution 7 derived with different exposure input layers. Building locations from OSM (first column), building counts from GHSL Global Human Settlement Layer (second column), population counts from GHSL (third column), and population counts from Kontur (last column).*

# 5 Discussion

Several studies developed methods to fuse information from remote sensing and social media to improve accuracy of flood hazard maps (Rosser et al. 2017; Huang et al. 2018; Guo et al. 2023). Contrary to these, the study at hand aimed at predicting the sampling probability for additional information collection that is conditionally dependent on the input proxy information layers themselves. To achieve this, a spatially and temporally consistent log-linear pooling framework was developed based on the H3 discrete global grid system. The fundamental approach was similar to what has been proposed by Pittore (2015) in the context of seismic risk assessment. Major differences to their work, however, include the utilization of a discrete global grid system for scalable harmonization of input layers, the use of geo-social media data, the development of an unsupervised optimization of pooling weights, and a quantitative performance evaluation against observed damage distributions. The latter, moreover, allowed

to perform an extensive hyperparameter exploration to better understand the sensitivity of the method to variations in input layers, normalization and pooling weights. Lastly, the application of this study focussed on flood rapid response rather than seismic risk assessment. To this regard, it aimed at determining the most likely disaster affected areas before initiating on-demand satellite data acquisitions and planning aerial or in-situ survey campaigns.

*Information fusion*

This study demonstrated that the proposed information fusion method effectively predicted disaster hotspots in comparison to observed damage distributions. Despite the absence of an absolute ground reference, the proposed approach offered a reliable comparative analysis against damage distributions derived as part of rapid emergency mapping activations by the Copernicus EMS. This is an advancement over other studies that use probability pooling methods for example in the field of risk assessment, which commonly lack appropriate reference data for validation (Nadim et al. 2006; Pittore et al. 2015). Through extensive hyperparameter tuning and comparing various methods to approximate optimal pooling weights, it was identified that equal weight methods are straightforward and relatively effective, but optimization-based methods tend to perform better. This underscored the significance of context-specific tuning in achieving optimal results in log-linear pooling. An unsupervised method minimizing the KL divergence between input distributions and predicted distributions showed superior performance, addressing the limitations of supervised methods which require extensive training data (Wang et al. 2018). The proposed method leverages the H3 discrete global grid system, fuses medium-resolution satellite data, social media inputs, and geospatial data on exposed assets, offering a coherent and scalable framework for disaster hotspot identification. This integrative approach contrasts with the methods of previous studies (Yang et al. 2022; Florath et al. 2024) by providing a more flexible and comprehensive solution adaptable to various disaster types and data sources. This capability is critical in improving situational awareness during disaster response.

*Reliability and semantic classification of geo-social media data*

In the social media analysis, solely georeferenced data from Twitter are employed. The reliability of such data has been shown in several studies. For instance, Ferner et al. 2020 compare their Twitter-based analysis with satellite-derived damage information for the 2014 Napa Valley earthquake and the 2017 Hurricane Harvey, showing that the footprints coincide significantly. Similarly, Yang et al. 2019 confirm the reliability of Twitter data for Hurricane Harvey. Liu et al. 2020 compare flood-related Tweets to meteorological alerts for a flood in Boulder, USA and find that texts and imagery from Tweets can contain plenty useful information for disaster management purposes. Additionally, Twitter's Academic API, which was also used in our study, has been shown to provide a representative sample of Tweets (Pfeffer et al. 2023). In another context, Tweets have also been found to be a reliable data source for influenza surveillance (Aslam et al. 2014). For Sina Weibo, a comparable platform to Twitter in China, a strong correlation to human mobility has been identified, i.e. a reliability of this data as a representation of human activity (Liu et al. 2022). However, as our study covered several areas of interest, it cannot be assumed that this reliability would be the same for all regions. In particular, more rural areas are more likely to be less suitable for social media-based analyses. To mitigate this issue and increase the amount of available data,

content from other social media platforms (e.g. Mastodon, TikTok) could be used in future work. However, such posts often contain spatially less accurate georeferences.

For identifying relevant geo-social media posts, a binary RoBERTa-based model was used. Alternative approaches, such as specific relevance classification (Derczynski et al. 2018; Blomeier et al. 2024) or topic modelling (Havas and Resch 2021; Hanny and Resch 2024), could be explored in future studies to understand their impact on the results of the information fusion. Initially, attempts were carried out to directly aggregate relevant Tweets to the H3 grid, but the resolution significantly influenced the results of geo-social media analysis, particularly when the number of relevant Tweets was limited and not well georeferenced. This issue could be mitigated using an independent spatial hotspot analysis before conversion to H3, which helped address the Modifiable Areal Unit Problem (MAUP) by incorporating a chosen spatial neighbourhood. However, a meaningful spatial delimitation for aggregating Tweets remains a limitation of the approach. Additionally, changes in social media platforms' policies may impact academic data access and user structure, as was shown for Twitter (Schmidt et al. 2023), which also affects the reliability of the approach. Therefore, replicating this study with different data sources is essential for moving the findings forward into a real-world application.

*Flood mapping from satellite images*

The experiments highlighted the sensitivity of the input flood water mask on the results, emphasizing the need to understand the accuracy and limitations of the respective inputs. Errors in the source data directly affect the resulting hotspot map, as the fusion method does not handle uncertainties or compensates for input errors. Bayesian information fusion methods that incorporate uncertainty could provide significant benefits in this context (Pittore et al. 2018). To ensure high quality input flood masks, Martinis et al. (2022) stressed the importance of using an up-to-date reference water mask that considers seasonality, as static masks may lead to unreliable flood extent representations. This become particularly relevant in regions with dynamic hydrological conditions. In this study, reference water masks with a seasonal component were derived to reduce potential over-estimations of flood inundation. Fichtner et al. (2023) advanced this concept by identifying hazardous flood areas through spatio-temporal anomaly detection in a time series of water maps. This approach could be a robust alternative for continuous monitoring and should be further explored in future studies. If archived water extents are not readily available and/or computational resources are limited, it is suggested to use a single pre-event (non-flood) water extent of the same season as reference.

*Exposure information*

Crowd-sourced per-building datasets were used to define exposure. The main reason to focus on per-building data was to ensure the same geographical core entity as the reference dataset (per-building observed damages). Such detailed datasets can provide high-quality information about location and characteristics of single buildings, roads and other infrastructure. Variations in data completeness, however, may limit their usability for analyses that require globally consistent data quality and completeness (Herfort et al. 2023). Therefore, it was shown that globally consistent gridded representations of built-up area (GHSL-Builtup) and population distribution (GHSL-Pop or Kontur) at large geographical units can provide a valid alternative.

*H3 discrete global grid system*

The H3 Discrete Global Grid System (DGGS) was chosen to combine heterogeneous input proxy information layers, showing advantages over traditional GIS methods (Li and Stefanakis 2020). A DGGS enables consistent multiscale analysis by dividing the Earth's surface into distinct grid cells, facilitating the integration of diverse data types cohesively (Purss et al. 2019). This approach mitigated computational challenges by spatially aggregating pixel information and allowed for the observation of successive phenomena at the same geographic locations across scales (Chaudhuri et al. 2021). Moreover, the H3 grid offered enhanced performance, scalability, and operational flexibility, making it a suitable choice for disaster hotspot identification and response planning.

# 6 Conclusions

In this study, a method was proposed to quickly identify disaster hotspots, especially in situations where detailed damage assessments from in-situ surveys or analysis of very high-resolution satellite images are not readily available. The method utilizes the H3 discrete global grid system for spatially consistent integration of heterogenous geoinformation layers. It combines flood hazard data derived from systematically acquired medium-resolution satellite imagery with disaster-related information from geo-social media and freely accessible geospatial data on exposed assets. The fusion is based on a log-linear pooling of normalized proxy information layers coupled with an optimization of pooling weights. With respect to the main research questions, we can draw the following conclusions.

***RQ1****: Is it possible to determine the areas most affected by a flood solely based on readily available proxy information to guide on-demand satellite data acquisitions, aerial or in-situ survey campaigns?*

In five diverse study areas, it was shown that it is possible to determine the areas most affected by a flood solely based on readily available proxy information. An extensive hyperparameter search and a comparison of different methods to approximate optimal pooling weights revealed that while equal weight methods provided a straightforward and relatively effective approach, methods that involved optimization tended to offer better performance. This suggests the importance of context-specific tuning for achieving the best results in log-linear pooling. An unsupervised method that minimizes the KL divergence between input distributions and predicted distribution showed superior performance and can overcome the training data limitations of supervised methods.

***RQ2****: Can flood-related information layers with different semantic meaningfulness, spatial resolutions and temporal delays be fused in a spatially and temporally consistent framework?*

While being simple and transparent, the method was capable to incorporate geoinformation layers with different semantic meaningfulness and spatial resolutions into a spatially and temporally consistent framework. This means it can be adapted to other natural hazards (e.g., landslides, fire, earthquakes, etc.) or exposed assets (e.g., roads, railways, critical infrastructure, etc.). In this study, flood hazard extents derived from remote sensing, exposed buildings from crowd-sourcing and relevant text messages from Twitter were used as input. The methods to derive these information layers from the respective data sources, are established, extensively tested and highly accurate.

***RQ3**: What is the contribution of different proxy information layers from remote sensing, social media and other geodata to the identification of flood hotspots?*

The experiments revealed that exposed buildings appear to be the most important input layer for predictions, followed by flood maps, with Tweets being the least impactful but still providing consistent predictive value. Using other data sources (e.g., public datasets, field reports, etc.) to derive proxy information layers is possible. However, it is important to understand the accuracy and limitations of the respective input, as errors in the source directly transfer into to the resulting hotspot map. The fusion method itself does not handle uncertainties, nor is it able to compensate for errors in the input. Therefore, next steps focus on exploring uncertainty aware Bayesian information fusion methods for this application. Ongoing and future efforts aim to incorporate these methods into rapid mapping processes to support prioritisation and early-triggering of data acquisition (Mühlbauer et al. 2024) and to contribute to data-driven decision-making in early phases of disaster response.

# References


Adriano B, Yokoya N, Xia J, et al (2021) Learning from multimodal and multitemporal earth observation data for building damage mapping. ISPRS Journal of Photogrammetry and Remote Sensing 175:132–143. https://doi.org/10.1016/j.isprsjprs.2021.02.016

Adwaith D, Abishake AK, Raghul SV, Sivasankar E (2022) Enhancing multimodal disaster tweet classification using state-of-the-art deep learning networks. Multimedia Tools Appl 81:18483–18501. https://doi.org/10.1007/s11042-022-12217-3

Aimaiti Y, Sanon C, Koch M, et al (2022) War related building damage assessment in Kyiv, Ukraine, using Sentinel-1 radar and Sentinel-2 optical Images. Remote Sensing 14:6239. https://doi.org/10.3390/rs14246239

Ajmar A, Boccardo P, Broglia M, et al (2017) Response to flood events: the role of satellite-based emergency mapping and the experience of the Copernicus Emergency Management Service. In: Flood Damage Survey and Assessment: New Insights from Research and Practice. American Geophysical Union, pp 213–228

Aslam AA, Tsou M-H, Spitzberg BH, et al (2014) The Reliability of Tweets as a Supplementary Method of Seasonal Influenza Surveillance. Journal of Medical Internet Research 16:e3532. https://doi.org/10.2196/jmir.3532

Avgerinakis K, Moumtzidou A, Andreadis S, et al (2018) A multimodal approach in estimating road passability through a flooded area using social media and satellite images. Sophia Antipolis, France, pp 1–3

Barron C, Neis P, Zipf A (2014) A comprehensive framework for intrinsic OpenStreetMap quality analysis. Transactions in GIS 18:877–895. https://doi.org/10.1111/tgis.12073

Barz B, Schröter K, Kra A-C, Denzler J (2021) Finding relevant flood images on Twitter using content-based filters. In: Del Bimbo A, Cucchiara R, Sclaroff S, et al. (eds) Pattern Recognition. ICPR International Workshops and Challenges. Springer International Publishing, Cham, pp 5–14

761 Bereczky M, Wieland M, Böhnke C, et al (2022) Water mapping for flood detection using
762 SAR data and convolutional neural networks. IEEE Journal of Selected Topics in
763 Applied Earth Observations and Remote Sensing 15:2023–2036

764 Blomeier E, Schmidt S, Resch B (2024) Drowning in the information flood: machine-learning-
765 based relevance classification of flood-related Tweets for disaster management.
766 Information 15:149. https://doi.org/10.3390/info15030149

767 Cao R, Tu W, Yang C, et al (2020) Deep learning-based remote and social sensing data
768 fusion for urban region function recognition. ISPRS Journal of Photogrammetry and
769 Remote Sensing 163:82–97. https://doi.org/10.1016/j.isprsjprs.2020.02.014

770 Chaudhuri C, Gray A, Robertson C (2021) InundatEd-v1.0: a height above nearest drainage
771 (HAND)-based flood risk modeling system using a discrete global grid system.
772 Geoscientific Model Development 14:3295–3315. https://doi.org/10.5194/gmd-14-
773 3295-2021

774 Chini M, Pelich R, Li Y, et al (2021) SAR-based flood mapping: where we are and future
775 challenges. In: 2021 IEEE International Geoscience and Remote Sensing
776 Symposium IGARSS. pp 884–886

777 Copernicus Emergency Management Service - Mapping (2022).
778 https://emergency.copernicus.eu/mapping/list-of-activations-rapid. Accessed 14 Feb
779 2024

780 de Carvalho LM, Villela DAM, Coelho FC, Bastos LS (2023) Combining probability
781 distributions: extending the logarithmic pooling approach. Bayesian Anal 18:.
782 https://doi.org/10.1214/22-BA1311

783 Derczynski L, Bontcheva K, Meesters K, Maynard D (2018) Helping crisis responders find
784 the informative needle in the Tweet haystack. In: Proceedings of the 15th ISCRAM
785 Conference. New York, USA

786 Ferner C, Havas C, Birnbacher E, et al (2020) Automated Seeded Latent Dirichlet Allocation
787 for Social Media Based Event Detection and Mapping. Information 11:376.
788 https://doi.org/10.3390/info11080376

789 Fichtner F, Mandery N, Wieland M, et al (2023) Time-series analysis of Sentinel-1/2 data for
790 flood detection using a discrete global grid system and seasonal decomposition.
791 International Journal of Applied Earth Observation and Geoinformation 119:103329.
792 https://doi.org/10.1016/j.jag.2023.103329

793 Florath J, Chanussot J, Keller S (2024) Rapid natural hazard extent estimation from twitter
794 data: investigation for hurricane impact areas. Nat Hazards 120:6775–6796.
795 https://doi.org/10.1007/s11069-024-06488-2

796 Fohringer J, Dransch D, Kreibich H, Schröter K (2015) Social media as an information
797 source for rapid flood inundation mapping. Natural Hazards and Earth System
798 Sciences 15:2725–2738. https://doi.org/10.5194/nhess-15-2725-2015

799 Google open buildings (2024). https://sites.research.google/open-buildings/. Accessed 13
800 Feb 2024

Guo K, Guan M, Yan H (2023) Utilising social media data to evaluate urban flood impact in data scarce cities. International Journal of Disaster Risk Reduction 93:103780. https://doi.org/10.1016/j.ijdrr.2023.103780

Gupta R, Goodman B, Patel N, et al (2019) Creating xBD: a dataset for assessing building damage from satellite imagery. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. Long Beach, USA, pp 10–17

H3-Pandas (2024). https://github.com/DahnJ/H3-Pandas. Accessed 11 Feb 2024

Hanny D, Resch B (2024) Clustering-based joint topic-sentiment modeling of social media data: a neural networks approach. Information 15:200. https://doi.org/10.3390/info15040200

Hanny D, Schmidt S, Resch B (2024) Active Learning for Identifying Disaster-Related Tweets: A Comparison with Keyword Filtering and Generic Fine-Tuning. In: Arai K (ed) Intelligent Systems and Applications. Springer Nature Switzerland, Cham, pp 126–142

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

Havas C, Resch B, Francalanci C, et al (2017) E2mC: improving emergency management service practice through social media and crowdsourcing analysis in near real time. Sensors 17:2766. https://doi.org/10.3390/s17122766

Havas C, Wendlinger L, Stier J, et al (2021) Spatio-temporal machine learning analysis of social media data and refugee movement statistics. ISPRS International Journal of Geo-Information 10:498. https://doi.org/10.3390/ijgi10080498

Herfort B, Lautenbach S, Porto de Albuquerque J, et al (2023) A spatio-temporal analysis investigating completeness and inequalities of global urban building data in OpenStreetMap. Nat Commun 14:3985. https://doi.org/10.1038/s41467-023-39698-6

Holzheimer E, Kippnich U, Kippnich M, et al (2022) Erkundung im Ahrtal mit Unterstützung von Verfahren der Künstlichen Intelligenz. In: Die Flut im Juli 2021. Erfahrungen und Perspektiven aus dem Rettungsingenieurwesen und Katastrophenrisikomanagement. TH Köln, Köln, pp 22–26

Huang X, Li Z, Wang C, Ning H (2020) Identifying disaster related social media for rapid response: a visual-textual fused CNN architecture. International Journal of Digital Earth 13:1017–1039. https://doi.org/10.1080/17538947.2019.1633425

Huang X, Wang C, Li Z (2018) A near real-time flood-mapping approach by integrating social media and post-event satellite imagery. Annals of GIS 24:113–123. https://doi.org/10.1080/19475683.2018.1450787

Kamoji S, Kalla M (2023) Effective flood prediction model based on Twitter text and image analysis using BMLP and SDAE-HHNN. Engineering Applications of Artificial Intelligence 123:106365. https://doi.org/10.1016/j.engappai.2023.106365

Krullikowski C, Chow C, Wieland M, et al (2023) Estimating ensemble likelihoods for the Sentinel-1-based Global Flood Monitoring product of the Copernicus Emergency Management Service. IEEE Journal of Selected Topics in Applied Earth

Observations and Remote Sensing 16:6917–6930. https://doi.org/10.1109/JSTARS.2023.3292350

Li J, He Z, Plaza J, et al (2017) Social media: new perspectives to improve remote sensing for emergency response. Proceedings of the IEEE 105:1900–1912. https://doi.org/10.1109/JPROC.2017.2684460

Li M, Stefanakis E (2020) Geospatial operations of discrete global grid systems: a comparison with traditional GIS. J geovis spat anal 4:26. https://doi.org/10.1007/s41651-020-00066-3

Liu L, Wang R, Guan WW, et al (2022) Assessing Reliability of Chinese Geotagged Social Media Data for Spatiotemporal Representation of Human Mobility. ISPRS International Journal of Geo-Information 11:145. https://doi.org/10.3390/ijgi11020145

Liu X, Kar B, Montiel Ishino FA, et al (2020) Assessing the Reliability of Relevant Tweets and Validation Using Manual and Automatic Approaches for Flood Risk Communication. ISPRS International Journal of Geo-Information 9:532. https://doi.org/10.3390/ijgi9090532

Liu Z, Qiu Q, Li J, et al (2021) Geographic optimal transport for heterogeneous data: fusing remote sensing and social media. IEEE Transactions on Geoscience and Remote Sensing 59:6935–6945. https://doi.org/10.1109/TGRS.2020.3031337

Martinis S, Groth S, Wieland M, et al (2022) Towards a global seasonal and permanent reference water product from Sentinel-1/2 data for improved flood mapping. Remote Sensing of Environment 278:113077. https://doi.org/10.1016/j.rse.2022.113077

Microsoft building footprints (2024). https://www.microsoft.com/en-us/maps/bing-maps/building-footprints. Accessed 13 Feb 2024

Muesing J, Burks L, Iuzzolino M, et al (2019) Fully bayesian human-machine data fusion for robust dynamic target surveillance and characterization. In: AIAA Scitech 2019 Forum. American Institute of Aeronautics and Astronautics, San Diego, USA

Mühlbauer M, Friedemann M, Roll J, et al (2024) Improved satellite-based emergency mapping through automated triggering of processes. In: Proceedings of the ISCRAM Conference. Muenster, Germany

Nadim F, Kjekstad O, Peduzzi P, et al (2006) Global landslide and avalanche hotspots. Landslides 3:159–173. https://doi.org/10.1007/s10346-006-0036-1

Nardo M, Saisana M, Saltelli A, Tarantola S (2005) Tools for composite indicators building. EUR 21682 EN. JRC31473

OpenStreetMap (2024). http://www.openstreetmap.org. Accessed 11 Mar 2024

Papadimos T, Andreadis S, Gialampoukidis I, et al (2023) Flood-related multimedia benchmark evaluation: challenges, results and a novel GNN approach. Sensors 23:3767. https://doi.org/10.3390/s23073767

Peduzzi P, Dao H, Herold C, Mouton F (2009) Assessing global exposure and vulnerability towards natural hazards: the disaster risk index. Natural Hazards and Earth System Sciences 9:1149–1159

Pesaresi M, Politis P (2023) GHS-BUILT-S R2023A - GHS built-up surface grid, derived from Sentinel2 composite and Landsat, multitemporal (1975-2030). http://data.europa.eu/89h/9f06f36f-4b11-47ec-abb0-4f8b7b1d72ea

Pfeffer J, Mooseder A, Lasser J, et al (2023) This Sample Seems to Be Good Enough! Assessing Coverage and Temporal Reliability of Twitter's Academic API. Proceedings of the International AAAI Conference on Web and Social Media 17:720–729. https://doi.org/10.1609/icwsm.v17i1.22182

Pittore M (2015) Focus maps: a means of prioritizing data collection for efficient geo-risk assessment. Annals of Geophysics 58:S0107. https://doi.org/10.4401/ag-6692

Pittore M, Graziani L, Maramai A, et al (2018) Bayesian estimation of macroseismic intensity from post-earthquake rapid damage mapping. Earthquake Spectra 34:1809–1828. https://doi.org/10.1193/112517EQS241M

Pittore M, Wieland M, Errize M, et al (2015) Improving post-earthquake insurance claim management: a novel approach to prioritize geospatial data collection. ISPRS International Journal of Geo-Information 4:2401–2427. https://doi.org/10.3390/ijgi4042401

Poria S, Cambria E, Bajpai R, Hussain A (2017) A review of affective computing: From unimodal analysis to multimodal fusion. Information Fusion 37:98–125. https://doi.org/10.1016/j.inffus.2017.02.003

Powers CJ, Devaraj A, Ashqeen K, et al (2023) Using artificial intelligence to identify emergency messages on social media during a natural disaster: A deep learning approach. International Journal of Information Management Data Insights 3:100164. https://doi.org/10.1016/j.jjimei.2023.100164

Purss MBJ, Peterson PR, Strobl P, et al (2019) Datacubes: a discrete global grid systems perspective. Cartographica 54:63–71. https://doi.org/10.3138/cart.54.1.2018-0017

Putri AFS, Widyatmanti W, Umarhadi DA (2022) Sentinel-1 and Sentinel-2 data fusion to distinguish building damage level of the 2018 Lombok Earthquake. Remote Sensing Applications: Society and Environment 26:100724. https://doi.org/10.1016/j.rsase.2022.100724

Resch B, Usländer F, Havas C (2018) 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. https://doi.org/10.1080/15230406.2017.1356242

Rosser JF, Leibovici DG, Jackson MJ (2017) Rapid flood inundation mapping using social media, remote sensing and topographic data. Natural Hazards 87:103–120. https://doi.org/10.1007/s11069-017-2755-0

Sahr K (2011) Hexagonal discrete global grid systems for geospatial computing. Archives of Photogrammetry, Cartography and Remote Sensing 22:363–376

Schiavina M, Freire S, Carioli A, MacManus K (2023) GHS-POP R2023A - GHS population grid multitemporal (1975-2030). In: European Commission, Joint Research Centre (JRC). http://data.europa.eu/89h/2ff68a52-5b5b-4a22-8f40-c41da8332cfe

924 Schmidt S, Zorenböhmer C, Arifi D, Resch B (2023) Polarity-based sentiment analysis of
925 georeferenced tweets related to the 2022 Twitter acquisition. Information 14:71.
926 https://doi.org/10.3390/info14020071

927 Snidaro L, Visentini I, Bryan K (2015) Fusing uncertain knowledge and evidence for maritime
928 situational awareness via Markov Logic Networks. Information Fusion 21:159–172.
929 https://doi.org/10.1016/j.inffus.2013.03.004

930 Tarakanov A (2020) Kontur population dataset. In: Kontur Inc.
931 https://www.kontur.io/portfolio/population-dataset/. Accessed 15 Feb 2024

932 UN-OCHA (2023) Revised Pakistan 2022 Floods Response Plan Final Report. UN-OCHA,
933 Geneva

934 Voigt S, Giulio-Tonolo F, Lyons J, et al (2016) Global trends in satellite-based emergency
935 mapping. Science 353:247–252. https://doi.org/10.1126/science.aad8728

936 Wang H, Skau E, Krim H, Cervone G (2018) Fusing heterogeneous data: a case for remote
937 sensing and social media. IEEE Transactions on Geoscience and Remote Sensing
938 56:6956–6968. https://doi.org/10.1109/TGRS.2018.2846199

939 Wieland M, Fichtner F, Martinis S, et al (2023) S1S2-Water: A global dataset for semantic
940 segmentation of water bodies from Sentinel-1 and Sentinel-2 data. IEEE Journal of
941 Selected Topics in Applied Earth Observations and Remote Sensing 17:1084–1099

942 Wieland M, Martinis S (2019) A modular processing chain for automated flood monitoring
943 from multi-spectral satellite data. Remote Sensing 11:2330.
944 https://doi.org/10.3390/rs11192330

945 Wiguna S, Adriano B, Mas E, Koshimura S (2024) Evaluation of deep learning models for
946 building damage mapping in emergency response settings. IEEE J Sel Top Appl
947 Earth Observations Remote Sensing 17:5651–5667.
948 https://doi.org/10.1109/JSTARS.2024.3367853

949 Yang J, Yu M, Qin H, et al (2019) A Twitter Data Credibility Framework—Hurricane Harvey
950 as a Use Case. ISPRS International Journal of Geo-Information 8:111.
951 https://doi.org/10.3390/ijgi8030111

952 Yang T, Xie J, Li G, et al (2022) Extracting disaster-related location information through
953 social media to assist remote sensing for disaster analysis: the case of the flood
954 disaster in the Yangtze river basin in China in 2020. Remote Sensing 14:1199.
955 https://doi.org/10.3390/rs14051199

956 Zheng Z, Zhong Y, Wang J, et al (2021) Building damage assessment for rapid disaster
957 response with a deep object-based semantic change detection framework: From
958 natural disasters to man-made disasters. Remote Sensing of Environment
959 265:112636. https://doi.org/10.1016/j.rse.2021.112636

960 Zhou X, Chen L (2014) Event detection over twitter social media streams. The VLDB Journal
961 23:381–400. https://doi.org/10.1007/s00778-013-0320-3

962

# Declarations

## Funding

This work was supported in part by the German Federal Ministry of Education and Research (BMBF) and the Austrian Research Promotion Agency (FFG) through the project “Künstliche Intelligenz zur Analyse von Erdbeobachtungs- und Internetdaten zur Entscheidungs-unterstützung im Katastrophenfall” (AIFER) under grant numbers 13N15525-13N15529 and 879732. Additional support has been received from the European Union’s Horizon Europe research and innovation programme as part of the project “Trusted extremely precise mapping and prediction for emergency management” (TEMA) under grant number 101093003.

## Author contributions

All authors contributed to the study conception and design. Material preparation, data collection and analysis were performed by Marc Wieland and Sebastian Schmidt. The first draft of the manuscript was written by Marc Wieland and Sebastian Schmidt and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

## Competing interests

The authors have no relevant financial or non-financial interests to disclose.