



Social media as an information source for rapid flood inundation mapping

J. Fohringer^{1,2}, D. Dransch^{1,2,4}, H. Kreibich^{1,3}, and K. Schröter^{1,3}

¹Center for Disaster Management and Risk Reduction Technology (CEDIM), Potsdam and Karlsruhe, Germany

²Section Geoinformatics, German Research Centre for Geosciences (GFZ), Potsdam, Germany

³Section Hydrology, German Research Centre for Geosciences (GFZ), Potsdam, Germany

⁴Humboldt University Berlin, Geography Department, Berlin, Germany

Correspondence to: K. Schröter (kai.schroeter@gfz-potsdam.de)

Received: 8 June 2015 – Published in Nat. Hazards Earth Syst. Sci. Discuss.: 1 July 2015

Revised: 27 October 2015 – Accepted: 2 December 2015 – Published: 21 December 2015

Abstract. During and shortly after a disaster, data about the hazard and its consequences are scarce and not readily available. Information provided by eyewitnesses via social media is a valuable information source, which should be explored in a more effective way. This research proposes a methodology that leverages social media content to support rapid inundation mapping, including inundation extent and water depth in the case of floods. The novelty of this approach is the utilization of quantitative data that are derived from photos from eyewitnesses extracted from social media posts and their integration with established data. Due to the rapid availability of these posts compared to traditional data sources such as remote sensing data, areas affected by a flood, for example, can be determined quickly. The challenge is to filter the large number of posts to a manageable amount of potentially useful inundation-related information, as well as to interpret and integrate the posts into mapping procedures in a timely manner. To support rapid inundation mapping we propose a methodology and develop “PostDistiller”, a tool to filter geolocated posts from social media services which include links to photos. This spatial distributed contextualized in situ information is further explored manually. In an application case study during the June 2013 flood in central Europe we evaluate the utilization of this approach to infer spatial flood patterns and inundation depths in the city of Dresden.

1 Introduction

Information provided by citizens via the internet can improve the information basis for disaster response after natural disasters (Poser and Dransch, 2010; Dransch et al., 2013). During a wildfire event in 2007 in California, affected people posted information about the wildfires in their own neighborhood to an internet page to inform others about the expansion and behavior of the fires (Sutton et al., 2008; Goodchild and Glennon, 2010). Another popular example is “Ushahidi” (<http://www.ushahidi.com/>), a content-sharing platform that collects and provides eyewitness reports of natural disasters, like earthquakes (Haiti and Chile 2010, Christchurch 2011), winter storms (Washington, D.C. 2010), wildfires (Russia 2010) and floods (Missouri 2011 and Thailand 2011). Hirata et al. (2015) used the Ushahidi platform to test a procedure for collaborative flood mapping in Sao Paulo based on information provided by people equipped with mobile devices providing location data. Information provided by citizens offers several benefits for disaster assessment and management. It is highly up-to-date since eyewitnesses document their observations directly and as such improve situation awareness and assessment. Additionally, people may contribute information which cannot be captured by sensors if the phenomenon is not measurable or sensors are not available. Information from citizens can be conducive to a rapid description of the extent and intensity of the hazardous event as well as the resulting impacts.

Rapid evaluations of hazardous events are needed for efficient response both in emergency management and in finan-

cial compensation and reconstruction planning. Estimates of monetary loss to be expected in a certain hazard scenario can be provided by damage models. In the presented research we follow the hypothesis that social media contain additional and/or even exclusive information which can be used to reasonably infer spatial flood patterns and inundation depths and thus provide an important basis for the estimation of flood damage. We investigate how information from social media, such as Twitter or Flickr, can contribute data for this task. The following challenges will be addressed.

- Challenge 1: filtering information relevant for inundation mapping from the high amount of information posted in social media.

One major challenge related to the utilization of information posted in social media is the huge volume of information created continuously about all kind of topics. Concepts and tools are required to facilitate the extraction of information that is suitable for inundation mapping.

- Challenge 2: the utilization of the relevant information from social media for inundation mapping.

Information about the inundation situation and particularly about the flooding intensity in terms of inundation depth is essential for rapid loss estimation in the case of floods. Inundation depth data are typically provided after the flood by terrestrial surveys of flood marks, evaluation of aerial or satellite images, or ex-post hydrodynamic–numeric simulations of the flood. Near-real-time information on inundation depths is, if at all, available from in situ data (e.g., from water level gauges) or derived from remote sensing products (e.g., satellite images) in combination with terrain elevation data. It has to be investigated if social media can provide relevant and rapid information on the flood inundation area and depth.

Both challenges were addressed in a close and fruitful cooperation of flood experts and computer scientists. The computer scientists developed a tool, “PostDistiller”, that combines various filtering methods with regard to selective contextual information reduction. The functionalities, components and implementation of PostDistiller are detailed in Sect. 2.3. This tool also provides a visual interface to assess the filtered information and to derive suitable data for flood inundation mapping. The flood experts investigated and evaluated the utilization of information provided by this tool. They examined how the information derived from social media complements traditionally collected data. Additionally, they evaluated how it supports rapid inundation mapping.

2 Challenge 1: filtering social media information

Information from social media comes with several challenges (Abel et al., 2012): the filtering of relevant information, its provision of information to people who need it, and quality assessment of the data. The presented research focuses predominantly on the filtering and provision of data; data quality is treated implicitly.

2.1 State of the art and related work

To find meaningful information in the large amount of data, several approaches have been pursued so far: (1) filtering by keywords or by geographic queries, (2) filtering by crowdsourcing, (3) automatic filtering utilizing machine learning and natural language processing and (4) interactive visual spatiotemporal analysis/geovisual analytics.

For gathering social media posts for analyzing users’ response during or after a disastrous event, the keyword search of the appropriate social media service is mostly used. Vieweg et al. (2010), for instance, used terms like “grass fire” or “red river” to collect tweets that contained terms related to the Oklahoma grass fires and Red River floods in spring 2009. The selection of the keywords affects the number and quality of the returned tweets; this has been shown by e.g., Rogstadius et al. (2013) or Joseph et al. (2014). The geolocation of posts from social networks can be used as an alternative to filtering using disaster-related and language-specific keywords. Herfort et al. (2014) examined the spatial relation between location-based social media and flood events. Their results show that tweets, which are geographically closer to flood-affected areas, contain more useful information capable of improving situational awareness than others.

Another approach to determine relevant social media posts is to classify them manually by crowdsourcing and to filter irrelevant declared data. Crowdsourcing, as introduced by Howe (2006), means to distribute a specific task to an unknown set of volunteers to solve a problem by harnessing collective contribution rather than by an individual. Another form is to perform data processing tasks like translating, filtering, tagging or classifying this content by voluntary crowd workers (Rogstadius et al., 2011). These operations can already be facilitated during the creation of social media posts by explicit user-driven assignment of predefined thematic keywords or hashtags, for instance “#flood” or “#need”, to allow information that is contained in the posts to be easily extracted, filtered and computationally evaluated (Starbird and Stamberger, 2010). In this approach, in particular, the scalability is a problem because due to the volume and velocity of the posts, during the disaster response, high-speed processing is difficult, even for a large group of volunteers (Imran et al., 2014).

For classification of text content in posts into relevant and non-relevant information, automatic approaches such as supervised classification and natural language process-

ing (NLP) techniques based on machine learning are also used. For instance, Sakaki et al. (2010) used a support vector machine (SVM) based on linguistic and statistical features such as keywords, the number of words and the context of target-event words for the detection of earthquake events in Japan. Yin et al. (2012) developed a classifier that automatically identifies tweets including information about the condition of certain infrastructure components like buildings, roads or energy supplies during the Christchurch earthquake in February 2011 by utilizing additional Twitter-specific statistical features like the number of hashtags and user mentions. Other important features as observed by Verma et al. (2011) are subjectivity and sentiment that can also help to find information contributing to situational awareness. However, Imran et al. (2013) have shown that pre-trained classifiers are suitable for the classification for a specific disaster event, but achieve significantly inferior results for another event of the same type. As a consequence, classifiers have to be adjusted for each disaster event and for each task, e.g., event detection or damage assessment, in order to achieve the best possible accuracy in classifying relevant posts.

To connect the benefits of crowdsourcing (ad-hoc classification without need for training classifiers) and machine learning (scalability and automatic processing), Imran et al. (2014) present a system to use volunteers to manually classify part of the incoming data as training data for an automatic classification system.

Geovisual analytics approaches also allow social media posts to be filtered, putting focus on interactive visualization and exploration rather than on completely automated machine learning methods. MacEachren et al. (2011) presented a geovisual analytics approach for the collection and filtering of geocoded tweet content within a visual interface to support crisis management to organize and understand spatial, temporal and thematic aspects of evolving crisis situations. Also Morstatter et al. (2013) used visualization techniques for organizing tweets by these aspects, for instance time graphs, which show the number of tweets matching a query per day, network graphs, showing which matching tweets were propagating most and heat maps, showing the spatial distribution of these tweets.

The approach presented in this paper combines filtering and visualization methods. Keywords are used, as in most works presented here, for the retrieval of generally disaster-related data. From the collected subset of posts, those that can be filtered are both temporally and spatially related to the concrete disaster event under study. A visual interface facilitates the exploration of filtered posts with the purpose of deriving specific quantitative or qualitative data. Compared to the methods and procedures discussed in this paragraph, neither training classifiers (machine learning/natural language processing) nor a sufficiently large number of volunteers (crowdsourcing) are necessary in our approach.

2.2 Requirements

Rapid impact assessment requires quick information about a specific hazardous event. This includes the type of impact, such as inundation, the affected area and the time when the effect was observed. All posts containing such information have to be selected from the high amount of information posted to social media. Additionally, the selected posts have to be analyzed to extract qualitative and quantitative information about the impact either from text, photos or videos which are enclosed in the post.

The selection of all relevant posts for a specific disaster event should be possible at any time when it is needed. Since not all social media services provide full retrieval of all posts at any time, two types of retrieval have to be available: the event-related on-demand retrieval for social media that allows all posts to be permanently accessed, and the continuous retrieval for social media that only provides posts for a limited time. The event-related on-demand retrieval enables posts to be retrieved by an accurately fitting query. In contrast, in the continuous retrieval, the event is not known in advance, therefore posts must be retrieved that generally refer to several types of natural hazards and their impacts, such as “flooding” or “inundation”. Continuous retrieval results in a collection of posts covering a variety of disasters; therefore, additional filters are necessary to select those posts that are relevant for the specific event under study.

The extraction of information related to the impact is dependent on the type of event. In our case study we focus on inundation mapping after floods; thus, information about inundation area and water depth have to be extracted. We focused on photos to extract this information, since photos have the following advantages. They show the relation of water level and parts of the environment, such as windows, roofs or traffic signs; this facilitates estimation of the inundation depth. Photos also show contextual information, for instance existing means for mobile flood prevention or nearby buildings. This contextual information supports the interpretation and verification of derived information. For example, the photo’s context allows the post’s geolocation to be verified. Apparent mismatches between the photo contents and its location can in most cases be recognized by locating posts on a map. Means are required to visually explore the selected photos and derive meaningful information.

2.3 PostDistiller

2.3.1 Components

Our approach to select relevant posts and to extract required information consists of three components: (1) PostCrawler for the retrieval of the posts, (2) PostStorage for persistent storage and (3) PostExplorer for the exploration and extraction of information from single posts (Fig. 1). PostCrawler



Figure 1. PostDistiller system architecture.

and PostStorage are generic components; PostExplorer is adapted to the case study flood inundation mapping.

PostCrawler retrieves and preprocesses disaster-related posts from social media services. Depending on the temporal availability of posts provided by the social media service, the posts are collected by either retrieving a data stream continuously (e.g., in Twitter), or an event-related set of data on-demand (e.g., in Flickr). In the case of continuous retrieval of posts, general disaster-relevant search terms are applied for retrieval; they cover the type of hazard, e.g., “flood”, the perceptible triggers, e.g., “heavy rain”, and its impacts, e.g., “destruction” and “damage”. For event-related on-demand retrieval, these search terms are stated more precisely regarding observable effects and consequences of the specific event, like “overflowing rivers” or “flooded roads”, as well as the affected area and corresponding time period. In both cases, the search terms can be customized by the user. After retrieval the posts are automatically preprocessed regarding duplicates, georeferencing and harmonization of the data format. Duplicates of posts, caused by the forwarding of posts that have already been published, are removed. Posts without explicit location information in the form of geocoordinates are automatically georeferenced if possible. The features “date” and “location” are harmonized with respect to their formal description. This becomes necessary since the date and loca-

tion can be contained in different attributes within the same post; for example, the location in a tweet can be given either in designated attribute “coordinates” or in the user profile. Using various social media services simultaneously, these features can also appear in different formats or encodings, e.g., as geographical coordinates as longitude and latitude or vice versa as latitude and longitude. The results of georeferencing and harmonization are added as extra attributes to the original post and saved in PostStorage.

The collected posts from various social media services are permanently saved by the *PostStorage* in a database. The database stores all attributes of a post which are text, links to external media (images and videos), location, creation date, user profile, URLs and others. Data selection is possible by means of the harmonized attributes “date” and “location” and other attributes of the posts.

The *PostExplorer* assists the flood experts in various ways. It enables relevant posts to be selected from the database according to different post attributes. It supports the exploration of the information inherent in the post, and it enables extracted information to be captured and stored in the database. Data selection is realized by multi-parameter filtering. As natural disasters affect a limited region within a limited period of time, the posts will be filtered based on their publication date and location. Further filtering is achieved by con-

performs the authentication procedures required by Twitter and requests the stream of tweets, by entering appropriate disaster-specific search terms, such as “flood”, “inundation” or “damage”. These search terms can be customized by the user to limit the amount of data during collection. PostCrawler for Twitter has been implemented in Java. To access Twitter’s Streaming API, the Hosebird Client (hbc) (<https://github.com/twitter/hbc>) is used. Tweets are received as documents in JavaScript Object Notation (JSON) consisting of attribute value pairs, like “text”: “The flood cannot impress us...” or “url”: “<http://t.co/YFdltwOr7t>”. The Flickr-specific implementation of PostCrawler connects to the representation state transfer (REST) interface of Flickr, authenticates itself and requests posts that contain corresponding event-related search terms in appropriate metadata (title, description or tags), for example “elbe”, “water level” or “gauge”. Time and area of the event are also included in the request. Selected documents are also returned in JSON format from Flickr. PostCrawler for Flickr was programmed in Python. Access to the Flickr API is provided by the software library flickrapi (<http://stuvcl.eu/flickrapi>). The preprocessing of collected posts is implemented as follows. Duplicate removal: forwarded tweets, so-called retweets, are identified by appropriate markings that exist either in the text or the metadata, e.g., a preceding “RT” or in the attributes “retweeted” or “retweeted_status”. Those retweets are stored separately in order to avoid duplication. Data harmonization: data harmonization between both services is accomplished by parsing attributes, which include the location (in Twitter: “coordinates”, and in Flickr: “location”) and creation date of the post (“created_at” and “date_taken”) and the mapping of each of these to a new shared attribute (“coordinates” and “creation_date”). Georeferencing: to add geocoordinates to posts without explicit location information, the open-source software package CLAVIN (<http://clavin.bericotechnologies.com/>) (Cartographic Location And Vicinity INdexer) is used. It helps to extract the local entities from text-related attributes and to find associated geocoordinates using the OpenStreetMap data set (<http://wiki.openstreetmap.org/wiki/Planet.osm>) and GeoNames database (<http://www.geonames.org/>).

PostStorage: to save preprocessed posts, the open-source database system MongoDB (<http://www.mongodb.org/>) is used as a back-end database for PostStorage. MongoDB is a document-oriented database that allows storage of JSON-like documents in the form that they are delivered by Twitter and Flickr. This is different to common relational databases which need predefined data schemes. By these means, it is possible to store posts from several social media services without doing additional data conversion. Each attribute of the posts is indexable and queryable. The database supports indexes for numeric, text and date attributes; it also supports 2-D geospatial indexing. These indexes facilitate post selection from the database in various ways. Spatial queries allow posts to be easily retrieved from defined areas. The full-text

search of MongoDB allows text to be filtered according to search terms like “flooded road” or keywords/hashtags like “#waterlevel”.

PostExplorer: PostExplorer’s functionality regarding data selection, data exploration and data capture is implemented as follows. Data selection is realized by multi-parameter filtering. Temporal filtering selects posts that are published in the chosen time period. Spatial filtering selects all posts based on whether the associated position is located within the chosen target area, e.g., a river basin that is described internally by a 2-D multipolygon. The media filtering is done by selecting all posts that contain one or more URLs in either the text itself or in the corresponding metadata. As we are interested in photos attached to collected posts, it is determined whether embedded URLs point to images of popular photo-sharing services Instagram (<http://instagram.com>), TwitPic (<http://twitpic.com>), Path (<https://path.com>) or Twitter’s own service. Appropriate filter parameters can be set by drop-down boxes that allow a predefined event-type to be chosen, e.g., “flood”. Depending on this selection, the user chooses the river basin to be examined from a predefined list (e.g., Elbe) as well as the time period of considered posts (e.g., from 5 May 2013 until 21 June 2013). For data exploration and data capture, an interactive visual interface has been set up to allow the database and the selected posts to directly interact. The four components of the visual interface are realized as follows. The data set of photos and text messages resulting from the filtering is presented in the visual interface. For an overview, the photos are listed in a sliding list. The sliding list shows four scaled-down versions of the filtered images at a time (component 1). By selecting a certain photo in this list, the corresponding post is displayed; it presents an enlarged version of the photo as well as the attributes associated with the post (component 2). The photo’s location derived from its coordinates is highlighted in the map view (component 3). Information that has been extracted from the photo by an expert is captured via input boxes and stored in the database (component 4). The visual interface is implemented as a web-based user interface. As a web application, PostExplorer is a client-server application that is displayed in the user’s web browser and is executed on a web server. Regarding the server, the Python-based web application framework, Flask (<http://flask.pocoo.org>), is used. Flask is kept simple and minimal, but allows existing libraries to be easily integrated, e.g., such that they can interact with MongoDB, or to process and deliver documents in JSON through the Hypertext Transfer Protocol. In addition to the Hypertext Markup Language, we applied several means: Cascading Style Sheets and JavaScript to implement the web interface, and the JavaScript library Leaflet (<http://leafletjs.com>) to implement the interactive map.

3 Challenge 2: the utilization of the information from social media for rapid inundation mapping

One challenge for rapid flood impact assessment is to obtain an overview of the flooding situation in which the main topics of interest are spatial flood patterns and inundation depths. Social media content shows promise, in that it improves disaster response capabilities by adding supplementary information to improve situation awareness and assessment. However, the utility of this information source depends on the possibility of reasonably inferring quantitative data on inundation depths. This will be tested within the case study of the June 2013 flood in the city of Dresden (Germany).

3.1 State of the art and related work

Given the aim to rapidly provide flood inundation depth maps, a pragmatic attitude towards data sources and quality is needed, meaning that any suitable information should be exploited as soon as it becomes available and might be discarded or updated when further data become available with time. In light of this, the availability of data in space and time, as well as the reliability of data sources is of particular importance.

Data sources which are usually used for inundation mapping are water level observations at river gauges, operational hydrodynamic–numeric model results or remote sensing data. In combination with topographic terrain data, which are available from topographic maps or digital elevation models (DEMs), the inundation depth within the flooded areas can be estimated. The requirements for topographic data are considerable. This particularly concerns the accuracy of ground levels as well as the realistic representation of hinterland flow paths and flood protection schemes, since these details locally control flooding. The advent of airborne laser altimetry, for instance lidar, has significantly improved the resolution and vertical accuracy of DEMs within the lower range of decimeters (Mandlburger et al., 2009; Bates, 2012).

Water level sensors are usually installed for tens of kilometers along a river course and only a fraction are equipped with online data transmission features. Depending on the sampling interval of the measurement network, water level values are available online within minutes to hours or days. Hence, during floods, only limited point information of water levels is available for inundation mapping. Linear interpolation of water levels between gauging stations is straightforward to obtain an estimate of the flood level (Apel et al., 2009). The intersection of this level with a DEM then yields a map of inundated areas. The difference between ground levels and the flood level is the inundation depth. However, this approach neglects non-stationary hydrodynamic processes, the limitation of flow volume and the effects of hydraulic structures. A higher spatial data density would be needed to approximate the actual characteristics of the water level gradient along a river more realistically.

Hydrodynamic–numeric models compute floodplain inundations by solving the hydrodynamic equations of motion for given geometric and hydraulic boundaries and initial conditions. The spatial detail of the simulated inundation depths depends on the discretization level of the model setup which is usually below 100 m horizontal resolution (Horritt and Bates, 2002; Falter et al., 2014). The near-real-time application of hydrodynamic–numeric models is hampered by the need to provide appropriate estimates of initial and boundary conditions, to assimilate model simulations and observations (Matgen et al., 2007) and by considerable computational costs (Di Baldassarre et al., 2009). Computation time depends particularly on the size of the computational domain and its spatial resolution (Falter et al., 2013) and the complexity level of model equations (Horritt and Bates, 2002). Alternatively, the inundated areas and inundation depths can be calculated in advance for a set of flood scenarios. However, the underlying assumptions of such scenarios might differ from the actual situation of a real event, e.g., dike breaches. The consideration of such unforeseen incidents is not feasible.

Remote sensing data allow inundated areas to be detected by comparing images from before and during floods (Wang, 2002). In combination with a DEM, the approximation of flood water levels, and thus the estimation of inundation depth, is feasible by detecting the flood boundary and extracting height information from the DEM (Zwenzner and Voigt, 2009; Mason et al., 2012). However, image acquisition is largely dependent on the revisiting time of orbital platforms, which in turn is inversely related to spatial resolution (Di Baldassarre et al., 2009). During a flood it is not guaranteed that suitable remote sensing images are available within a short time for the flood situation and the region of interest. Further, the synchronous acquisition of images with the occurrence of flood peaks, in order to capture maximum flood extent, is hard to achieve. This particularly applies for large areas due to dynamic flood processes. Usually, image delivery and processing is feasible within 24–48 h (Schumann et al., 2009).

In light of this, social media show promise, in that social media fill the time gap until inundation depth information from other data sources might become available. The derivation of inundation depths from photos could complement observations from water level gauges with additional distributed in situ information and support the inundation mapping process. Schnebele and Cervone (2013) show the complementary value of information extracted from photos and videos which have been compiled from a search on the internet for flood extent mapping. In urban areas, the additional micro-level evidence of the flooding situation is valuable since there are difficulties in utilizing remotely sensed information and flood inundation models in these areas (Zwenzner and Voigt, 2009; Apel et al., 2009). Despite these obvious opportunities of social media for rapid flood damage estimation, there are a number of challenges to overcome. This

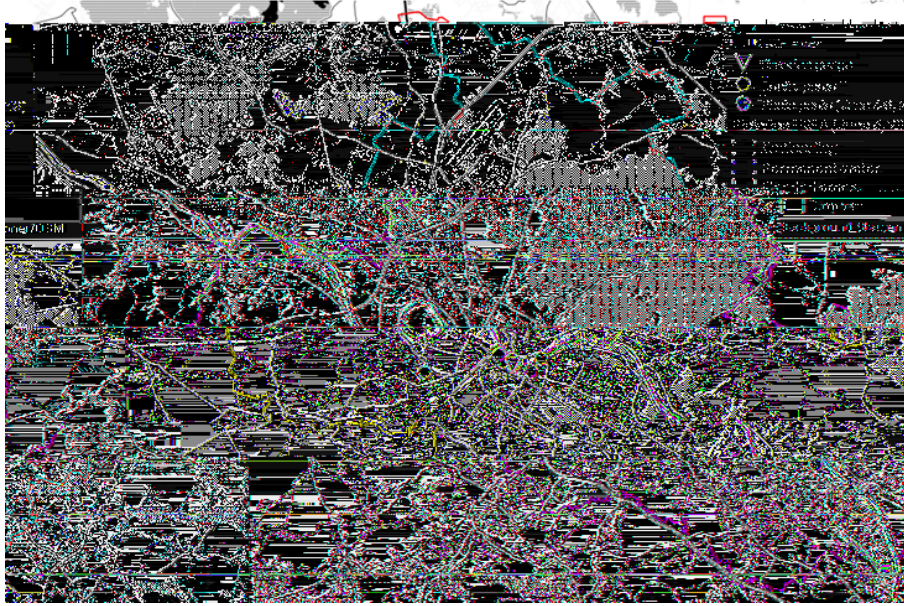


Figure 3. Study region and data sources for flood inundation depth mapping.

concerns the filtering of relevant information and the availability and quality of information. As social media posts are not controlled or actively investigated, there is no guarantee for their availability during the flood. The content and spatial coverage of the posts very much depends on the caprice of tweeters. Data quality, credibility of information and uncertainty concerning location and inferred inundation depth are important issues (Poser and Dransch, 2010).

3.2 Case study: Dresden flood, June 2013

We investigate the usefulness of photos posted via Twitter and Flickr as an information source for rapid inundation depth mapping within the city of Dresden during the flood in June 2013 using PostDistiller. Urban areas are of specific interest because on the one hand potential flood damage is high, and on the other hand, the number of social media activists is large. The city of Dresden (Saxony, Germany), with almost 800 000 inhabitants, is located on the banks of the river Elbe of which major flooding has caused severe impacts, most notably the recent events of August 2002, April 2006 and June 2013. Therefore, there is a high level of flood awareness in Dresden and comprehensive flood management concepts have been put into practice (Landeshauptstadt Dresden, 2011).

During the June 2013 flood, the peak water level at the Dresden gauge (Fig. 3) was registered on 6 June 2013 with 876 cm above the gauge datum (i.e., 111.3 mNN). Due to an elongated flood wave, the water level remained above 850 cm (ca. HQ20) from 5 to 7 June 2013 which is a critical level for flooding in several districts of Dresden e.g., Laubegast and Kleinzschachwitz upstream, and Pieschen Süd downstream

of the city center (see Fig. 3, Landeshauptstadt Dresden, 2011).

3.2.1 Data and inundation mapping scenarios

Within the Dresden case study, we use data from the water level gauge in Dresden (operated by the Waterways and Shipping Administration (WSV)) and photos retrieved from Twitter and Flickr as information sources for rapid inundation mapping. Information on ground level is available from the DGM10 (Federal Agency for Cartography and Geodesy) which has a vertical accuracy of ± 0.5 to ± 2 m.

Further, for this study, a footprint of flooded areas in Dresden is available from Perils AG (www.perils.org) which is based on Pléiades HR1A multispectral image taken on 5 June 2013 with a horizontal resolution of 50 cm. In this product, a SPOT 5 multispectral image from 21 August 2011 has been used as a reference to classify flooded areas and permanent water surfaces. Even though this footprint has been released as a rapid inundation mapping product and might not meet the requirements of a careful documentation of flooded areas, in the context of this study, it is a useful reference to evaluate the outcomes of the rapid inundation mapping procedures based on either water level observations or social media photo posts and DEM terrain data.

Inundation depth maps are derived for two scenarios: (a) online water level observations at the Dresden gauge and (b) information inferred from photos filtered from Twitter and Flickr services using PostCrawler, PostStorage and PostExplorer implementation presented in this paper. The satellite-based flood footprint is used to evaluate the mapping results in terms of inundation extent.

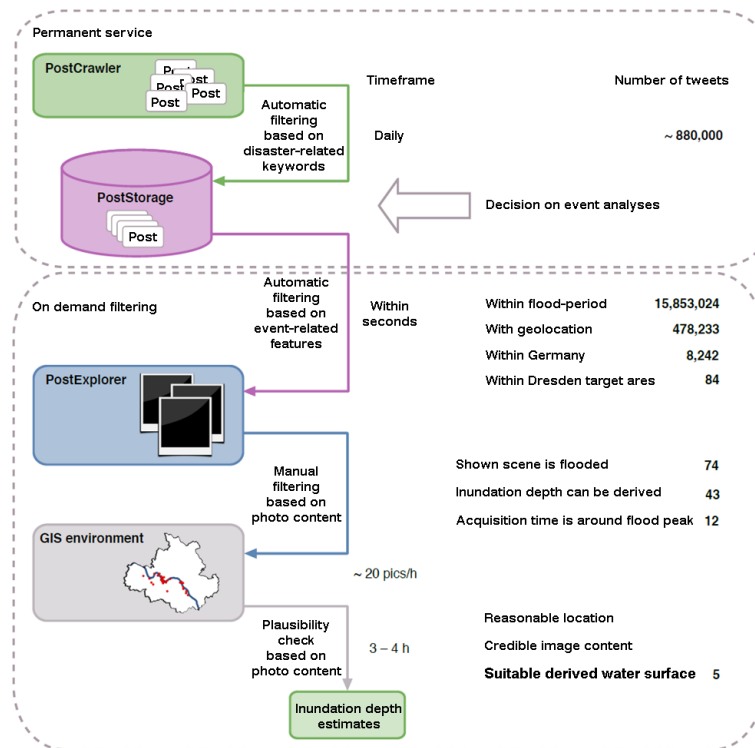


Figure 4. Process chain, time frame and number of tweets for the June 2013 Dresden flood, processed within PostCrawler, PostStorage, PostExplorer and GIS environment for the automatic and manual filtering of tweets.

3.2.2 Results

Within scenario (a), the water level observation for the flood peak at the Dresden gauge retrieved online is intersected with the DEM10. Considering hydrodynamic flow processes, the water level is not horizontal but inclined along the flow direction. In view of the elongated flood wave which led to almost constant high flood levels during 6 and 7 June 2013, it is reasonable to assume quasi-stationary flow conditions in the time period around the flood peak. Therefore, we assume that the gradient of the water level along the river is approximately parallel to the bottom slope (on average 0.27‰ between the upstream gauge at Pirna and downstream gauge in Meissen). The inclined water level surface is intersected with the DEM in such a way that all areas below the water level are assumed to be inundated. The difference between the water surface and ground level is the inundation depth. The resulting inundation depth map is shown in Fig. 6a.

In scenario (b) we utilize PostDistiller to extract information from the posts collected by PostCrawler and stored in PostStorage which runs as a permanent service. The keywords used for PostCrawler are terms related to floods and associated impacts, both in English and in German: flood: “Hochwasser”, “Flut”, “flood”, “floods”, “flooding”, “inundation”, “Sturzflut”, “Überflutung”, “Unterspülung”, “Regen”, “undermining”.

Triggered by the decision to analyze the June 2013 flood in the Dresden region, this database is automatically filtered based on event-related features which include the definition of the flood period of interest (5 May 2013 until 21 June 2013), the availability of geolocation information attached to the posts and the location within the target study region. Within seconds, the relevant posts are passed on to PostExplorer which enables subsequent manual filtering and visual inspection of photo contents for the estimation of inundation depth. Within a geographic information system (GIS) environment, the plausibility of photo locations and derived inundation depths are checked. In this step, the time window for the acquisition time of photos is narrowed to the period from 5 to 7 June 2013 to exclusively capture the inundation situation around the occurrence of the flood peak at the water level gauge in Dresden. The process chain, the time frame and the resulting number of tweets in each step are compiled in Fig. 4 for this specific application example.

For the Dresden example, 84 geolocated posts with photos attached are available within the target time and area. As a result of plausibility checks and expert image evaluation, a total number of five inundation depth estimates are derived for subsequent flood inundation mapping. To give an impression of the challenge in estimating inundation depth based on photo content, the five useful photo posts, their location and the inundation depth estimates in the Dresden study region

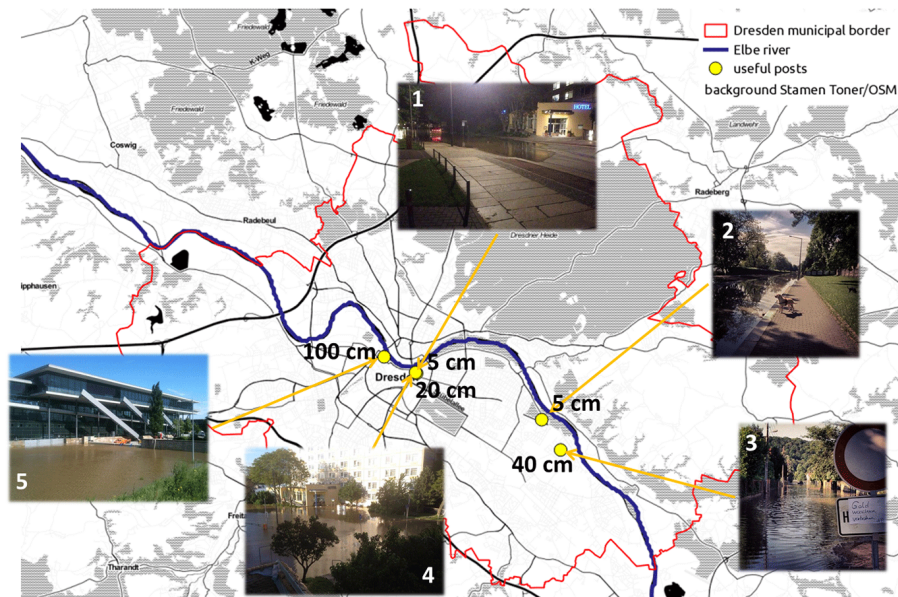


Figure 5. Location of useful photos retrieved with PostDistiller and inundation depths estimates. (Photos by Denny Tumlrirsch (@Flitzpatrick), @ubahnverleih, Sven Wernicke (@SvenWernicke) and Leo Käbner (@leokaesner).)

are shown in Fig. 5. For instance, photos 1 and 2 in Fig. 5 show inundated roads but a dry sidewalk. This context enables the analyst to estimate inundation depth in the order of approximately 5 cm. Photo 4 in Fig. 5 shows flood water on an open space between residential buildings. The orange waste bin which can be seen in this photo is not yet touched by the flood water which provides an indication to estimate inundation depth in the order of 20 cm.

Next, these point estimates of inundation depth are converted into water levels with reference to the base height level (mNN). This is achieved by adding the inundation depth to the ground level height available from the DGM10 at the location in which the photo was taken. The resulting heights are sample points of the spatial continuous water level surface. Given the origin of these points, they obviously do not show a regular spatial structure as for instance an equidistant grid. Further, the sample size of data points is rather small. Given these properties we follow the recommendations of Li and Heap (2014) for the selection of spatial data interpolation methods, and apply a bilinear spline interpolation to obtain an estimate of the water level surface within the target area. Finally, the water level surface is intersected with the DEM10 and the difference between the water surface and ground level provides the inundation depth within the inundated area. The resulting inundation depth map is shown in Fig. 6b. All GIS processing tasks are conducted using GRASS software (GRASS, 2014).

3.2.3 Evaluation

The inundation maps derived for both mapping scenarios using either (a) online water level observations at the Dresden gauge (Fig. 6a) or (b) information inferred from photos filtered from social media (Fig. 6b) are compared in terms of spatial inundation extent and inundation depths.

The water level surface of the inundation map derived from social media photos is on average 1.5 m above the water level surface of the online water level observation. Accordingly, the inundated area, based on social media photos, resulting from the intersection with the DEM is larger than using the online water level observation, as can be seen from Fig. 6a and b. The spatial distribution of the differences between both inundation depth estimates is given in Fig. 6d. The differences have been calculated by subtracting scenario (a) from scenario (b) within the overlapping areas. This map of differences illustrates that the agreement of inundation depth estimates is best in the city center of Dresden with differences smaller than 1 m. In contrast, in the upper part of the Elbe River in the region where additional point estimates of inundation depth are derived from social media photos, the differences amount up to 4 m. This comparison reveals that first, in the case that no water level observations are available, social media may provide useful alternative inundation depths estimates and, second, that through the spatial distribution of information sources, social media may also provide additional information to the inundation mapping process.

Both mapping approaches are contrasted with a flood footprint which is based on remote sensing data recorded on 5 June 2013, given in Fig. 6c. This reference inundation map

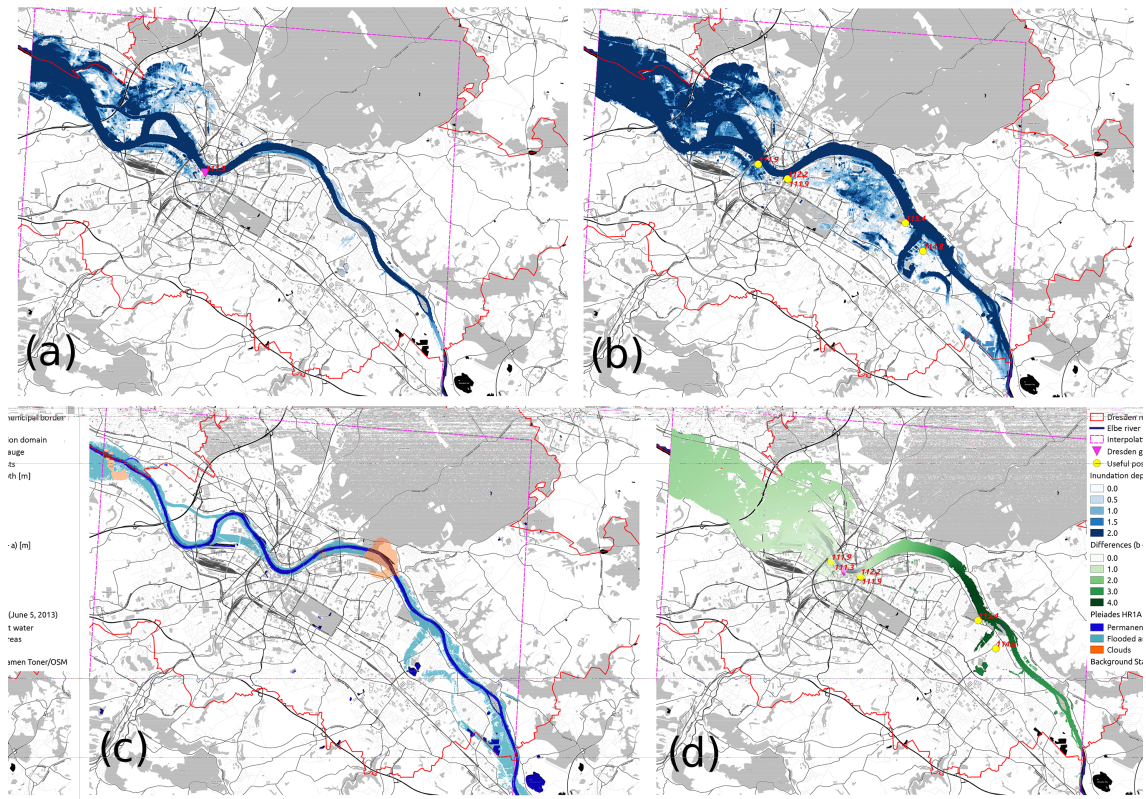


Figure 6. Inundation maps and inundation depths derived from online water level observations (a) and social media content (b); inundated area derived from the reference remote sensing flood footprint (c); and differences between inundation depths for overlapping areas in scenarios (a) and (b) (panel d).

indicates inundations in Dresden in the district of Laubegast upstream and in Pieschen Süd downstream of the city center (cf. Fig. 3), the pattern of which reflects the former course of ancient river branches. From this comparison it is apparent that both mapping scenarios clearly overestimate inundated areas. This applies for the inundation mapping based on water level observations (scenario a) for the section downstream of the gauge in Dresden which is located in the city center. In this scenario, for the upstream section, no inundations are detected. In contrast, for the inundation mapping based on social media data (scenario b) areas upstream of the gauge in Dresden are also classified as inundated and provide inundation depth information which is the outcome of the inundation depth estimates available from the social media photos in the district of Laubegast (cf. photos 2 and 3 in Fig. 5). However, in scenario (b), the extent of inundated areas in the target area is overestimated even stronger than using solely water level observations in scenario (a). Both inundation depth mapping scenarios intersect the estimated water level with a 10 m DEM. This level of detail for the topographic terrain does not map dike crests, mobile flood protection walls and other flood protection schemes that are in place. Moreover, the spatial interpolation procedures neither account for hydraulic flow paths nor correct for puddles, i.e., low-lying ar-

reas that are behind dams or walls and hence are not flooded in reality. To overcome the weaknesses of the spatial interpolation schemes, the remote-sensing-based flood footprint could be used as a mask in order to spatially constrain the inundation depth maps. In our case study, such an information update would have been available several hours later (24 h after image acquisition at best).

4 Discussion

The methodology and tools for filtering the massive amount of social media data described in this paper proved to be robust and effective within the application example for inundation mapping during the June 2013 flood in the city of Dresden. The filtering and data processing chain effectively supports the target-oriented evaluation of photo content. For this example, the temporal demand of the processing chain to provide inundation depth maps is in the range of 2 to 8 h. This expenditure of time essentially stems from the effort of manually filtering and evaluating the photos as well as GIS processing, and thus depends on the number of photos to be analyzed. Still, in comparison to alternative data sources for inundation depth mapping, e.g., remote sensing or hydraulic

numeric modeling, the inundation depth map derived from social media is more rapidly available and proved a useful complement to water level gauge observations, or are even an exclusive in situ data source for the case that no water level gauges exist in the target area. Technical possibilities to further limit the preselection of photos for instance by making use of automatic image analyses should be investigated to further improve the efficiency of the manual filtering and evaluation.

It is recognized from the comparison with the water level observations from the gauge in Dresden that the inundation depth estimates derived from photos for specific locations in Dresden in combination with the base height levels from the DGM10 provide decent water level elevations. The differences are in the order of decimeters which is acceptable for the purpose of rapid inundation mapping, particularly when no other information source is available. In this context, both the vertical accuracy level of the DGM10 which is around ± 0.5 to ± 2 m and the vagueness of referencing inundation depth from the photo content, as illustrated in Fig. 5, have to be kept in mind. Further, the potential difference of the location of the photo and the geolocation of the tweet, as well as the offset from the location of where the photo was taken and the photo contents, involves uncertainties concerning the horizontal location. Improvements concerning the vertical and horizontal accuracy of inundation depth estimates can be expected from using higher resolution lidar DEMs which may achieve vertical accuracies for terrain data in the range of ± 15 cm (Mandlbürger et al., 2009) and thus also include details about dike crests. Improvements can also be expected as a result of the integration of more detailed information about the reference environment as for instance available from 3-D-city models (Gröger et al., 2012).

To reduce the inaccuracies concerning the spatial extent of flooding, more exact topographic terrain data should be used and appropriately considered within spatial interpolation. Making use of ancillary data such as remote sensing flood footprints, or hydraulic numeric modeling results based on detailed topographic terrain data within spatial interpolation, e.g., external drift Kriging (Goovaerts, 1997), should be investigated.

The availability and the spatial coverage of information support from social media within the target area cannot be controlled but depends on the random activity of social media users. Crowdsourcing, i.e., distributing the task of inundation depth estimation, solves this problem by harnessing collective contributions (Howe, 2006) by actively pushing the acquisition of information via social media, and thus could enhance the reliability of this data source and improve the coverage.

In summary, the results obtained from the application case in Dresden support the initial hypothesis that social media contain additional and potentially even exclusive information that is useful for inundation depth mapping as a basis for rapid damage estimation, but also more generally for im-

proving situation awareness and assessment during the flood event.

5 Conclusions

A methodology and tool to automatically filter and efficiently support the manual extraction of information from social media posts for rapid inundation mapping has been presented. In the first step, the processing chain allows a manageable number of potentially interesting social media posts to be filtered within seconds. In the Dresden application case, 84 potentially interesting posts were selected out of almost 16 million posts. In the second step, PostDistiller supports the process by manually assessing and filtering the automatically derived posts according to relevance and plausibility of their content. Finally, information about inundation depth is extracted. All in all, estimates on inundation depth could be derived within 3 to 4 h in the Dresden example. In comparison to traditional data sources such as satellite data, social media can provide data more rapidly.

The outcomes of the application case are encouraging. Strengths of the proposed procedure are that information for the estimation of inundation depth is rapidly available, particularly in urban areas where it is of high interest, and of great value because alternative information sources like remote sensing data analysis do not perform very well. The photos provided represent a snapshot of the current situation and thus also help to improve situation awareness and assessment. In contrast, the detail of location information that can be extracted from social media posts is limited and inundation depth estimates are associated with uncertainty concerning the timing, location and magnitude. Another weakness is that appropriate social media information is not reliably available as it depends on the random behavior of human sensors. Hence, the uncertainty of derived inundation depth data and the uncontrollable availability of the information sources are major threats to the utility of the approach. Another disadvantage is related to the fact that the more photos that are available, the longer it takes to manually evaluate the photo contents and finally derive an inundation depth map. Automation and integrated quality assessment are crucial for any operational application of the tool.

Nevertheless, social media as an information source for rapid inundation mapping provide the opportunity to close the information gap when traditional data sources are lacking or are sparse. In particular, the joint usage of different data streams seems to provide added value. In light of this, further research is required (i) to investigate technical possibilities to improve the preselection of photos by making use of automatic image analyses, (ii) to integrate more detailed information on the reference environment as for instance provided by high-resolution lidar DEMs or 3-D-city models, (iii) to use ancillary data such as flood footprints or hydrodynamic numeric modeling results in order to constrain

inundated areas and to continuously update inundation depth maps and (iv) for the purpose of quality control, to develop a probabilistic mapping framework that accounts for the uncertainties involved in the different data sources and the final result.

Acknowledgements. The study was undertaken and financed under the framework of CEDIM – Center for Disaster Management and Risk Reduction Technology.

The article processing charges for this open-access publication were covered by a Research Centre of the Helmholtz Association.

Edited by: D. Molinari

Reviewed by: two anonymous referees

References

- Abel, F., Hauff, C., Houben, G.-J., Stronkman, R., and Tao, K.: Semantics + filtering + search = twitcident exploring information in social web streams, in: Proceedings of the 23rd ACM Conference on Hypertext and Social Media, HT '12, ACM, New York, NY, USA, 285–294, doi:10.1145/2309996.2310043, 2012.
- Apel, H., Aronica, G., Kreibich, H., and Thielen, A.: Flood risk analyses – how detailed do we need to be?, *Nat. Hazards*, 49, 79–98, doi:10.1007/s11069-008-9277-8, 2009.
- Bates, P. D.: Integrating remote sensing data with flood inundation models: how far have we got?, *Hydrol. Process.*, 26, 2515–2521, doi:10.1002/hyp.9374, 2012.
- Di Baldassarre, G., Schumann, G., and Bates, P.: Near real time satellite imagery to support and verify timely flood modelling, *Hydrol. Process.*, 23, 799–803, doi:10.1002/hyp.7229, 2009.
- Dransch, D., Poser, K., Fohringer, J., and Lucas, C.: Volunteered geographic information for disaster management, in: Citizen E-Participation in Urban Governance: Crowdsourcing and Collaborative Creativity, Advances in electronic government, digital divide, and regional development (AEGDDRD) book series, edited by: Silva, C. N., Information Science Reference, Hershey, Pa., 98–118, 2013.
- Falter, D., Vorogushyn, S., Lhomme, J., Apel, H., Gouldby, B., and Merz, B.: Hydraulic model evaluation for large-scale flood risk assessments, *Hydrol. Process.*, 27, 1331–1340, doi:10.1002/hyp.9553, 2013.
- Falter, D., Dung, N. V., Vorogushyn, S., Schröter, K., Huncheda, Y., Kreibich, H., Apel, H., Theisselmann, F., and Merz, B.: Continuous, large-scale simulation model for flood risk assessments: proof-of-concept, *J. Flood Risk Manage.*, doi:10.1111/jfr3.12105, in press, 2014.
- Goodchild, M. F. and Glennon, J. A.: Crowdsourcing geographic information for disaster response: a research frontier, *Int. J. Digital Earth*, 3, 231–241, 2010.
- Goovaerts, P.: *Geostatistics for Natural Resources Evaluation*, OUP, USA, 1997.
- GRASS Development Team: *Geographic Resources Analysis Support System (GRASS) Software*, available at: <http://grass.osgeo.org> (last access: 03 February 2015), 2014.
- Gröger, G. and Plümer, L.: CityGML – interoperable semantic 3D city models, *ISPRS J. Photogramm.*, 71, 12–33, doi:10.1016/j.isprsjprs.2012.04.004, 2012.
- Herfort, B., de Albuquerque, J. P., Schelhorn, S.-J., and Zipf, A.: Exploring the geographical relations between social media and flood phenomena to improve situational awareness, in: *Connecting a Digital Europe Through Location and Place*, edited by: Huerta, J., Schade, S., and Granell, C., Lecture Notes in Geoinformation and Cartography, Springer International Publishing, Cham, doi:10.1007/978-3-319-03611-3_4, 55–71, 2014.
- Hirata, E., Giannotti, M. A., Larocca, A. P. C., and Quintanilha, J. A.: Flooding and inundation collaborative mapping – use of the crowdmap/ushahidi platform in the city of Sao Paulo – Brazil, *J. Flood Risk Manage.*, doi:10.1111/jfr3.12181, 2015.
- Horritt, M. S. and Bates, P. D.: Evaluation of 1D and 2D numerical models for predicting river flood inundation, *J. Hydrol.*, 268, 87–99, doi:10.1016/S0022-1694(02)00121-X, 2002.
- Howe, J.: The Rise of Crowdsourcing, <http://www.wired.com/wired/archive/14.06/crowds.html> (last access: 25 July 2014), 2006.
- Imran, M., Elbassuoni, S., Castillo, C., Diaz, F., and Meier, P.: Practical extraction of disaster-relevant information from social media, in: Proceedings of the 22nd International Conference on World Wide Web Companion, WWW '13 Companion, Rio de Janeiro, Brazil, 1021–1024, 2013.
- Imran, M., Castillo, C., Lucas, J., Meier, P., and Vieweg, S.: AIDR: artificial intelligence for disaster response, in: Proceedings of the Companion Publication of the 23rd International Conference on World Wide Web Companion, WWW Companion '14, Seoul, Republic of Korea, 159–162, 2014.
- Joseph, K., Landwehr, P. M., and Carley, K. M.: An approach to selecting keywords to track on twitter during a disaster, in: Proceedings of the 11th International Conference on Information Systems for Crisis Response and Management (ISCRAM), State College, Pennsylvania, USA, 2014.
- Landeshauptstadt Dresden: *Plan Hochwasservorsorge Dresden*, Landeshauptstadt Dresden, Geschäftsbereich Wirtschaft, Umweltamt, Dresden, 2011.
- Li, J. and Heap, A. D.: Spatial interpolation methods applied in the environmental sciences: a review, *Environ. Modell. Softw.*, 53, 173–189, doi:10.1016/j.envsoft.2013.12.008, 2014.
- MacEachren, A. M., Jaiswal, A., Robinson, A. C., Pezanowski, S., Savelyev, A., Mitra, P., Zhang, X., and Blanford, J.: SensePlace2: GeoTwitter analytics support for situational awareness, in: 2011 IEEE Conference on Visual Analytics Science and Technology (VAST), Providence, Rhode Island, USA, doi:10.1109/VAST.2011.6102456, 181–190, 2011.
- Mandlbürger, G., Hauer, C., Höfle, B., Habersack, H., and Pfeifer, N.: Optimisation of LiDAR derived terrain models for river flow modelling, *Hydrol. Earth Syst. Sci.*, 13, 1453–1466, doi:10.5194/hess-13-1453-2009, 2009.
- Mason, D. C., Schumann, G. J.-P., Neal, J. C., Garcia-Pintado, J., and Bates, P. D.: Automatic near real-time selection of flood water levels from high resolution Synthetic Aperture Radar images for assimilation into hydraulic models: a case study, *Remote Sens. Environ.*, 124, 705–716, doi:10.1016/j.rse.2012.06.017, 2012.
- Matgen, P., Schumann, G., Henry, J.-B., Hoffmann, L., and Pfister, L.: Integration of SAR-derived river inundation areas, high-

- precision topographic data and a river flow model toward near real-time flood management, *Int. J. Appl. Earth Obs.*, 9, 247–263, doi:10.1016/j.jag.2006.03.003, 2007.
- Morstatter, F., Kumar, S., Liu, H., and Maciejewski, R.: Understanding twitter data with TweetExplorer, in: Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '13, ACM, New York, NY, USA, 1482–1485, doi:10.1145/2487575.2487703, 2013.
- Poser, K. and Dransch, D.: Volunteered geographic information for disaster management with application to rapid flood damage estimation, *Geomatica*, 64, 89–98, 2010.
- Rogstadius, J., Kostakos, V., and Laredo, J.: Towards real-time emergency response using crowd supported analysis of social media, in: Proceedings of CHI Workshop on Crowdsourcing and Human Computation, Systems, Studies and Platforms, Vancouver, British Columbia, Canada, 1–3, 2011.
- Rogstadius, J., Vukovic, M., Teixeira, C. A., Kostakos, V., Karapanos, E., and Laredo, J. A.: CrisisTracker: crowdsourced social media curation for disaster awareness, *IBM J. Res. Dev.*, 57, 4:1–4:13, doi:10.1147/JRD.2013.2260692, 2013.
- Sakaki, T., Okazaki, M., and Matsuo, Y.: Earthquake shakes twitter users: real-time event detection by social sensors, in: Proceedings of the 19th International Conference on World-Wide Web, WWW'10, ACM, New York, NY, USA, 851–860, doi:10.1145/1772690.1772777, 2010.
- Schnebele, E. and Cervone, G.: Improving remote sensing flood assessment using volunteered geographical data, *Nat. Hazards Earth Syst. Sci.*, 13, 669–677, doi:10.5194/nhess-13-669-2013, 2013.
- Schumann, G., Di Baldassarre, G., and Bates, P. D.: The utility of spaceborne radar to render flood inundation maps based on multialgorithm ensembles, *IEEE T. Geosci. Remote*, 47, 2801–2807, doi:10.1109/TGRS.2009.2017937, 2009.
- Starbird, K. and Stamberger, J.: Tweak the tweet: leveraging microblogging proliferation with a prescriptive syntax to support citizen reporting, in: Proceedings of the 7th International Conference on Information Systems for Crisis Response and Management (ISCRAM), Seattle, Washington, USA, 2010.
- Sutton, J., Palen, L., and Shklovski, I.: Backchannels on the front lines: emergent uses of social media in the 2007 southern California wildfires, Proceedings of the 5th International ISCRAM Conference, Delft, the Netherlands, 624–632, 2008.
- Verma, S., Vieweg, S., Corvey, W. J., Palen, L., Martin, J. H., Palmer, M., Schram, A., and Anderson, K. M.: Natural Language Processing to the rescue? Extracting “situational awareness” tweets during mass emergency, in: Proceedings of the Fifth International Conference on Weblogs and Social Media, Barcelona, Catalonia, Spain, 1–8, 2011.
- Vieweg, S., Hughes, A. L., Starbird, K., and Palen, L.: Microblogging during two natural hazards events: what twitter may contribute to situational awareness, in: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '10, Atlanta, Georgia, USA, 1079–1088, doi:10.1145/1753326.1753486, 2010.
- Wang, Y.: Mapping extent of floods: what we have learned and how we can do better, *Nat. Hazards Rev.*, 3, 68–73, doi:10.1061/(ASCE)1527-6988(2002)3:2(68), 2002.
- Yin, J., Lampert, A., Cameron, M., Robinson, B., and Power, R.: Using social media to enhance emergency situation awareness, *IEEE Intell. Syst.*, 27, 52–59, doi:10.1109/MIS.2012.6, 2012.
- Zwenzner, H. and Voigt, S.: Improved estimation of flood parameters by combining space based SAR data with very high resolution digital elevation data, *Hydrol. Earth Syst. Sci.*, 13, 567–576, doi:10.5194/hess-13-567-2009, 2009.