Nat. Hazards Earth Syst. Sci. Discuss., 3, 4231–4264, 2015
www.nat-hazards-earth-syst-sci-discuss.net/3/4231/2015/
doi:10.5194/nhessd-3-4231-2015

Natural Hazards and Earth System Sciences
Open Access
Discussions

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]

[1]Center for Disaster Management and Risk Reduction Technology (CEDIM), Potsdam and Karlsruhe, Germany
[2]Section Geoinformatics, German Research Centre for Geosciences (GFZ), Potsdam, Germany
[3]Section Hydrology, German Research Centre for Geosciences (GFZ), Potsdam, Germany
[4]Humboldt University Berlin, Geography Department, Berlin, Germany

Received: 08 June 2015 – Accepted: 17 June 2015 – Published: 01 July 2015

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

Published by Copernicus Publications on behalf of the European Geosciences Union.

Abstract

During and shortly after a disaster data about the hazard and its consequences are scarce and not readily available. Information provided by eye-witnesses via social media are 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 case of floods. The novelty of this approach is the utilization of quantitative data that are derived from photos from eye-witnesses extracted from social media posts and its integration with established data. Due to the rapid availability of these posts compared to traditional data sources such as remote sensing data, for example areas affected by a flood 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 their timely interpretation and integration in mapping procedures. To support rapid inundation mapping we propose a methodology and develop a tool to filter geo-located 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 Internet can improve the information basis for disaster response after natural disaster (Poser and Dransch, 2010; Dransch et al., 2013). During the wildfire event in 2007 in California affected people posted information about the wildfires in their own neighborhood to an internet page to inform 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 eye-witnesses document their observations directly and as such improve situation awareness and assessment. Additionally, people may contribute information which cannot be captured by sensors since the phenomenon is not measurable or sensors are not available. Information from citizens can be conducive to rapidly describe the extent and intensity of the hazardous event as well as resulting impacts.

Rapid evaluations of hazardous events are needed for efficient response both in emergency management and in financial compensation and reconstruction planning. Estimates of monetary loss to be expected by 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 provides 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: the reduction of the high amount of information posted in social media to that information relevant for inundation mapping.

  One major challenge related to utilization of information posted in social media is the huge volume of information created continuously to all kind of topics. Concepts and tools are required to facilitate extracting the information 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 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 (e.g. 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 for 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 that combines various filtering methods with regard to selective contextual information reduction. It 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 complement the traditionally collected data. Additionally, they evaluated how it improves rapid inundation mapping.

2 Filtering social media information

Information from social media comes along with several challenges (Abel et al., 2012): the filtering of relevant information, their provision to people who need them, and quality assessment of the data. The presented research focuses predominately on 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 harvesting social media posts for analysing user's response during or after a disastrous event, mostly the keyword search of the appropriate social media service is used. Vieweg et al. (2010), for instance, used terms like "grass fire" or "red river" to collect Tweets that contain terms related to the Oklahoma Grassfires and Red River Floods in spring 2009. To identify the impact area of earthquakes by detecting the spatial and temporal characteristics of Twitter activity, Crooks et al. (2012) gathered Tweets through keyword search of "earth" and "quake". The selection of the keywords affects the amount 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 for disaster related and language-specific keywords for filtering. 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, contains more useful information capable to improve 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 be facilitated already during creation of social media posts by explicit user-driven assignment of predefined thematic keywords or hashtags, for instance #flood or #need, to allow easy extraction, filtering and computational evaluation of contained information (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 a 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, also automatic approaches such as supervised classification and natural language processing (NLP) techniques based on machine learning are used. For instance, Sakaki et al. (2010) used a support vector machine (SVM) based on linguistical and statistical features such as keywords, number of words and context of target-event words for 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 number of hashtags and user mentions. Other important features as observed by Verma et al. (2011) are subjectivity and sentiment that can help also 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 filtering social media posts, putting focus on interactive visualisation and exploration rather than on completely automated machine learning methods. MacEachren et al. (2011) presented a geovisual analytics approach for collecting and filtering of geocoded Tweet content within a visual interface to support crisis management to organise and understand spatial, temporal and thematic aspects of the evolving crisis situations. Also Morstatter et al. (2013) used visualisation techniques for organising 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 can be filtered that are both temporally and spatially related to the concrete disaster event under study. A visual interface facilitates 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 amount 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 analysed 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 allow for permanent access to all posts, and the continuous retrieval for social media that provides posts only for limited time. The event-related on demand retrieval enables retrieval of posts 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 use case 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 estimating the inundation depth. Photos show also context information, for instance existing means for mobile flood prevention. This context information supports the interpretation of derived information. The photo context allows the verification of the geolocation of the tweet. Apparent mismatches between the photo contents and its location according to the geo-location information of the post can in most cases be recognized from the spatial position of the tweet in a map. Means are required to visually explore the selected photos and derive meaningful information.

2.3 Components

Our approach to select relevant posts and to extract required information consists of three components: the PostCrawler for retrieval of the posts, the PostStorage for persistent storage, and the PostExplorer for exploration and extraction of information from single posts (Fig. 1). PostCrawler and PostStorage are generic components, the PostExplorer is adapted to the use case flood inundation mapping.

The PostCrawler retrieves 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 case of continuous retrieval of posts general disaster-relevant search terms are applied; they cover the type of hazard, e.g. “flood”, the perceptible triggers, e.g. “heavy rain”, and its impacts, e.g. “destructions” 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 in the configuration of the system. After retrieval the posts are pre-processed. Duplicates, caused by forwarding of already published posts, are removed. Also the features “date” and “location” are harmonized. Date and location can be contained at 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. Harmonization involves the identification of attributes containing the requested information (e.g. geo-coordinates) by using pattern matching, as well as standardization of different data formats. For standardization the appropriate attributes are individually parsed and converted to a common format. The results of the harmonization are added as extra attributes to the original post and saved in the PostStorage.

The collected posts from various social media services are permanently saved by the PostStorage in a common database. The persistent storage enables offline analysis of the collected posts. 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. The posts are stored as they are delivered by each service. By means of the harmonized attributes date and location, posts can be selected from the PostStorage by the same query independently of the social media service they are coming from.

The PostExplorer facilitates to explore the posts stored in the PostStorage and to extract the required information for impact assessment. Via multi-parameter filtering relevant posts can be selected from the data base. 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 considering the presence of links to extra media, like photos or videos. In addition, event-related text filters can be used to filter posts referring to concrete effects of a disaster, such as dike breaches.

A visual interface presents the selected posts for analysis. The interface is configured for photo analyses and consists of three main components which are shown in Fig. 2: a component for a quick overview about the whole number of filtered posts and related photos, a component to analyse single posts and related photos with respect to extract information about inundation, and a third component to localize the posts and photos in a map. The overview component allows for browsing through the filtered posts/photos. The analysis component depicts a single post/photo together with the information attached which is author, publication time, location, and content. It also provides fields to store the results of analysis in the PostStorage database as additional expert information to a single post. The expert can add the following information: the relevance of the post/photo for inundation mapping, if the presented situation is wet or dry, the inundation depth estimate, and an indication of the estimated reliability of the derived information. The localization component shows a map with the location of filtered posts. It facilitates verifying if the coordinates from the post’s metadata match with the place and context depicted in the photo.

2.4 Implementation

The implementations for PostCrawler and PostStorage are independent of specific disaster types, the PostExplorer is adapted for application during flood events as an example. In our use case we have chosen the social media platforms Twitter and Flickr as information source. Both services are characterized by open interfaces, moderate access restrictions and widespread use.

PostCrawler: we use the micro-blogging service Twitter for continuous collection and the content-sharing service Flickr for retrieval on demand. Twitter is a social-media service that allows retrieval of posts on demand, but the result of a query is not necessarily complete. Thus, we retrieve posts from Twitter continuously to capture all potentially disaster-relevant posts. The 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. The PostCrawler connects to Twitter’s freely available Streaming API and receives consecutively Tweets matching given filter predicates. For this purpose the PostCrawler performs the authentication procedures required by Twitter and requests the stream of Tweets, by giving appropriate disaster specific search terms, such as “flood”, “inundation” or “storm” as well as general damage relevant terms like “damage”, “victim”, or “destruction”. These search terms can be customized by the user to limit the amount of data while collecting.

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/YFdItwOr7t”. Forwarded Tweets, so-called retweets, usually contain appropriate markings 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. The Tweets have attributes that are relevant for later event-related filtering; they include publication date and location as well as links to extra media such as photos or videos. As the location of the Tweet can be found in the metadata in various attributes and in various formats, they are transformed and merged if necessary. In order to also utilize Tweets without explicit location information in form of geo-coordinates, the open source software package CLAVIN (http://clavin.bericotechnologies.com/) (Cartographic Location And Vicinity INdexer) is used for geolocation. It facilitates to extract the local entities from text related attributes and to find associated geo-coordinates using the OpenStreetMap dataset (http://wiki.openstreetmap.org/wiki/Planet.osm) and GeoNames database (http://www.geonames.org/).

Flickr is a content sharing platform that provides full availability of posts over time. Thus, there is no need for continuous collecting of disaster-relevant posts; they are only retrieved when the disaster event under investigation is already known. For this purpose the Flickr-specific implementation of the PostCrawler connects to the representation state transfer (REST) interface of Flickr, authenticates itself and requests posts that contains 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. In response event-related documents

are returned also as JSON documents from Flickr. The PostCrawler for Flickr was programmed in Python. Access to the Flickr API is provided by the software library flickrapi (http://stuvel.eu/flickrapi).

Data Harmonization between both services is accomplished by parsing particular attributes, which include the location (in twitter: “coordinates”, and in Flickr: “location”) and creation date of the post (“created_at” and “datetaken”), and mapping each to a new common attribute (“coordinates” and “creation_date”) which enables a joint querying of both data sources for further analyses.

PostStorage: to save harmonized posts from various social media services, the open-source document-oriented database system MongoDB (http://www.mongodb.org/) is used as backend for the PostStorage. MongoDB allows storage of JSON-like documents without using a complex database scheme. By this means, it is possible to store posts from several services without doing additional data conversion. Nevertheless, each attribute is indexable and queryable. Beyond indexes for numeric, text and date attributes, it supports 2-D geospatial indexing, which facilitates spatial queries and makes it possible to retrieve posts from defined areas easily. Furthermore, taking advantage of the full-text search of MongoDB, a text filter can be used to restrict the common disaster-related posts collected by continuous retrieval to those referring to specific effects of the event under study by employing textual elements such as search terms like “flooded road” or keywords/hashtags like “#gauge” or “#waterlevel”.

The PostExplorer is implemented as a web-based user interface to select relevant posts for a known event, to browse content of the results and to enable deriving impact-related data. For the selection of posts, appropriate filter parameters can be set by the user by drop-down boxes that allow choosing a predefined event-type, 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 05 May 2013 until 21 June 2013). The multi-parameter filtering is realized as queries to the PostStorage that are parametrized by the corresponding inputs: temporal filtering selects posts that are published in the chosen time period. Spatial filtering selects all posts based on whether associated position is located within the chosen river basin that is described internally by a 2-dimensional 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 Twitters own service.

The resulting dataset of photos and text messages related to the selected event is presented in the visual interface detailed view. The post's coordinates are linked with the map view, which shows the chosen river basin district and the location of the selected observation in an interactive map. The photos are listed in a sliding list in the media view shown in Fig. 2. The sliding list shows four scaled-down versions of the filtered images at a time. By selecting a certain photo in this list, the corresponding post is displayed; including an enlarged version of the photo as well as the metadata associated. In addition, the location is highlighted in the map view. The combination of photo contents and map location allows for a verification of the geolocation of the tweet, i.e. that the locations of tweet and photo coincide. The visual interface enables the expert to attach attributes to the photo concerning the relevance of the photo content, classifying the flooding situation at the location as wet or dry and, if possible, estimating the inundation depth. Further the reliability of the depth estimate and be rated. The attributes assigned to the photos and the geo-coordinates are exported in a GIS compatible format for further processing.

As a web application the PostExplorer is a client-server application that is displayed in the user's web browser and is executed on a web server. On the server side, the Python-based web application framework Flask (http://flask.pocoo.org) is used. Flask is kept simple and minimal, but allows easy integration of existing libraries, such as for the interaction with MongoDB and to process and deliver documents in JSON through the Hypertext Transfer Protocol. In addition to the languages Hypertext Markup Language, Cascading Style Sheets and JavaScript used to implement the web interface,

the JavaScript library Leaflet (http://leafletjs.com) for the interactive map in the map view, the CSS framework Bootstrap (http://getbootstrap.com) and the plugin DataTables (https://datatables.net) for the JavaScript library jQuery (https://jquery.com/) for the presentation of the table in the text view are utilized on the client side.

3 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 interests are spatial flood patterns and inundation depths. Social media content is promising to improve disaster response capabilities by adding supplementary information to improve situation awareness and assessment. However, the utility of this information source depends on the possibility to reasonably infer quantitative data on inundation depths. This will be tested within the use case 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 this light, the availability of data in space and time as well as the reliability of data sources are 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 (DEM), 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 as for instance LiDAR has significantly improved the resolution and vertical accuracy of DEMs within the lower range of decimetres (Mandlburger et al., 2009; Bates, 2012).

Water level sensors are usually installed with tens of kilometres distance along a river course and only a fraction is 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 flood level is the inundation depth. However, this approach neglects non-stationary hydrodynamic processes, limitation of flow volume and 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 boundary and initial conditions. The spatial detail of the simulated inundation depths depends on the discretization level of the model set-up 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 for the detection of inundated areas by comparing before and during flood images (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 short time for the flood situation and the region of interest. Further, the acquisition of images synchronously with the occurrence of flood peak, 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 this light, social media show promise to 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 on the flooding situation is valuable since remotely sensed information and flood inundation models experience difficulties 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 concerns the filtering of relevant information, the availability and the quality of information. As social media posts are not controlled or actively inquired there is no guarantee for their availability during the flood. The content and spatial coverage of the posts is very much depending 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 June flood 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. 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 which has brought severe impacts from major floods, most notably the recent events in August 2002, April 2006 and June 2013. Therefore, flood awareness in Dresden is on a high level and comprehensive flood management concepts have been put into practice (Landeshauptstadt Dresden, 2011).

During the June 2013 flood the peak water level at the gauge Dresden (Fig. 3) was registered on 06 June 2013 with 876 cm above gauge datum (i.e. 111.3 mNN). Due to an elongated flood wave the water level remained above 850 cm (ca. HQ20) from 05 to 07 June 2013 which is a critical level for flooding in several quarters of Dresden e.g. Laubegast, Kleinschachwitz upstream and Pieschen Süd downstream of the city centre (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 water and shipment administration (WSV)) and photos retrieved from Twitter and Flickr as information basis for 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 Pleiades HR1A multispectral image taken on 05 June 2013 with a horizontal resolution of 50 cm. In this product a SPOT 5 multi-spectral 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 water level observations and social media photo posts and DEM terrain data.

Inundation depth maps are derived for two scenarios: (a) online water level observations at the gauge Dresden and (b) information inferred from photos filtered from Twitter and Flickr services using the 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.

3.2.2 Results

Within scenario (a), the water level observation for the flood peak at the gauge Dresden 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 the 06 and 07 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 Pirna and downstream Gauge 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 extract information from the posts collected by PostCrawler and stored in the PostStorage which runs as a permanent service. The keywords used for the PostCrawler are terms related to flood and associated impacts both in English and in German language: flood: “Hochwasser”, “Flut”, “flood”, “floods”, “flooding”, “inundation”, “Sturzflut”, “Überflutung”, “Unterspülung”, “Regen”, “undermining”.

Triggered by the decision to analyse the June flood 2013 in the Dresden region this data base is automatically filtered based on event related features which include the definition of the flood period of interest (05 May 2013 until 21 June 2013), the availability of geo-location information attached to the posts, and the location within the target study region. Within seconds the relevant posts are provided to the PostExplorer which enables subsequent manual filtering and estimation of inundation depth based on the photo content. Within a GIS environment the plausibility of photo locations and derived inundation depths are checked. In this step, the time window for acquisition time of photos is narrowed to the period from 05 to 07 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 timeframe and the resulting numbers of tweets in each step are compiled in Fig. 4 for this specific application example.

For the Dresden example a number of 84 geo-located 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 5 inundation depth estimates are derived for subsequent flood inundation mapping. To give an impression on the challenge to estimate inundation depth based on photo content the 5 useful photo posts, their location and the inundation depths estimates in the Dresden study region are shown in Fig. 5.

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 of the photo. 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. The water level surface is around 50 cm higher than in scenario a (Fig. 6a). All GIS processing tasks are conducted using the GRASS software (GRASS, 2014).

Figure 6c shows the flood footprint based on remote sensing data recorded on 05 June 2013. This reference inundation map indicates inundations in Dresden in the district of Laubegast upstream and in Pieschen Süd downstream of the city centre (cf. Fig. 3) the pattern of which reflects the former course of ancient river branches. From the comparison of the outcomes of the inundation depth mapping scenarios with the reference flood footprint it is visible that both scenarios overestimate inundated areas. This applies for the inundation mapping based on water level observations (scenario a) for the part downstream of the gauge in Dresden which is located in the city centre. For the upstream part no inundations are detected. In contrast, for the inundation mapping based on social media data (scenario b) also areas upstream of the gauge in Dresden are classified as inundated and provided with inundation depth data. This is due to the inundation depth data available from the social media photos in the district of Laubegast. However, in this scenario the extent of inundated areas in the target area is overestimated even stronger than using solely water level observations. This can be explained, first, by the higher elevation of the water level surface derived from the social media data. Second, both inundation depths 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 in place. Moreover, the spatial interpolation procedures neither account for hydraulic flow paths nor correct for puddles, i.e. low lying areas that are behind dams or walls and hence are not flooded. In this regard, the remote sensing flood footprint could be used as a mask in order to constrain the inundation depth maps. This information update would be available several hours later (at best 24 h after image acquisition).

4 Discussion

The methodology and tool for filtering the massive amounts 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 to manually filter and evaluate the photos as well as GIS processing and thus depends on the amount of pictures to be analysed. Still, in comparison to alternative data sources for inundation depth mapping, e.g. remote sensing or hydraulic numeric modelling, 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 gauge exists 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 to the water level observations from the gauge in Dresden that the inundation depth estimates derived from the photos for specific locations in Dresden in combination with the base height levels from the DGM10 provide useful water level elevations. The differences are in the order of decimetres which is sufficient for the purpose of rapid inundation mapping. 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 borne 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 photo shoot 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 digital elevation models which may achieve vertical ac-

curacies for terrain data in the range of ±15 cm (Mandlburger et al., 2009) and thus also include details about dike crests, or from 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 modelling 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 to 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 general for improving the 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 a first step the processing chain allows for filtering a manageable amount of potentially interesting social media posts within seconds. In the Dresden application case 84 potentially interesting posts were selected out of almost 16 Mio posts. In a second step the tool supports to manually assess and filter 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 three to four hours 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 procedure proposed are that information for the estimation of inundation depths 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 depths 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 behaviour of human sensors. Hence, the uncertainty of inundation depth data derived and the uncontrollable availability of the information source are major threats to the utility of the approach. Another disadvantage is related to the fact that the more photos 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 an added value. In this light, further research is required to (i) 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 DEM or 3d-city models, (iii) to use ancillary data such as flood footprints or hydrodynamic numeric modelling 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.

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 Thieken, 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.

Crooks, A., Croitoru, A., Stefanidis, A., and Radzikowski, J.: #Earthquake: twitter as a distributed sensor system, T. GIS, 17, 124–147, doi:10.1111/j.1467-9671.2012.01359.x, 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., Hundecha, 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 In460 formation 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.

Mandlburger, 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 TweetXplorer, in: Proceedings of the 19th ACMSIGKDD 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, 2011.

Discussion Paper | Discussion Paper | Discussion Paper | Discussion Paper |

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 onWorld-WideWeb,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.

Discussion Paper | Discussion Paper | Discussion Paper | Discussion Paper |

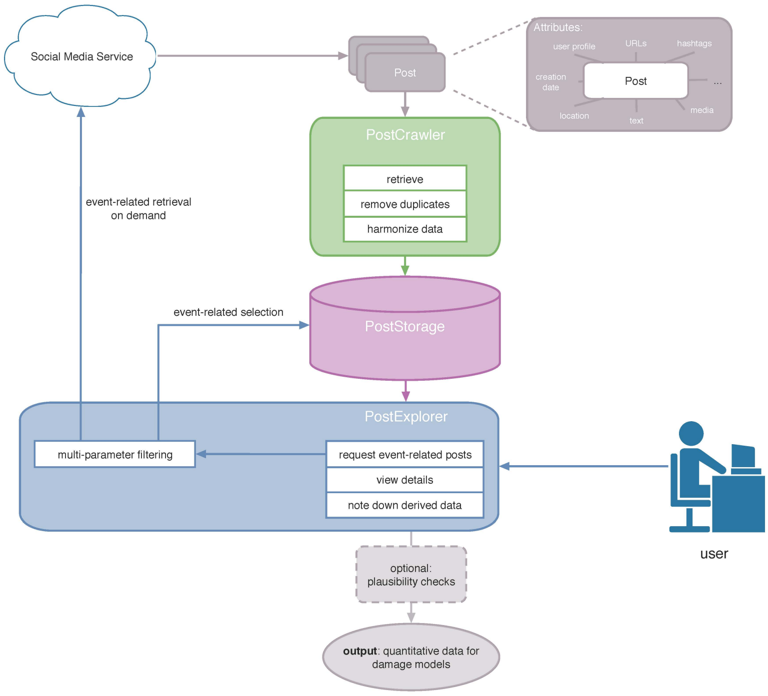

Figure 1. System architecture.

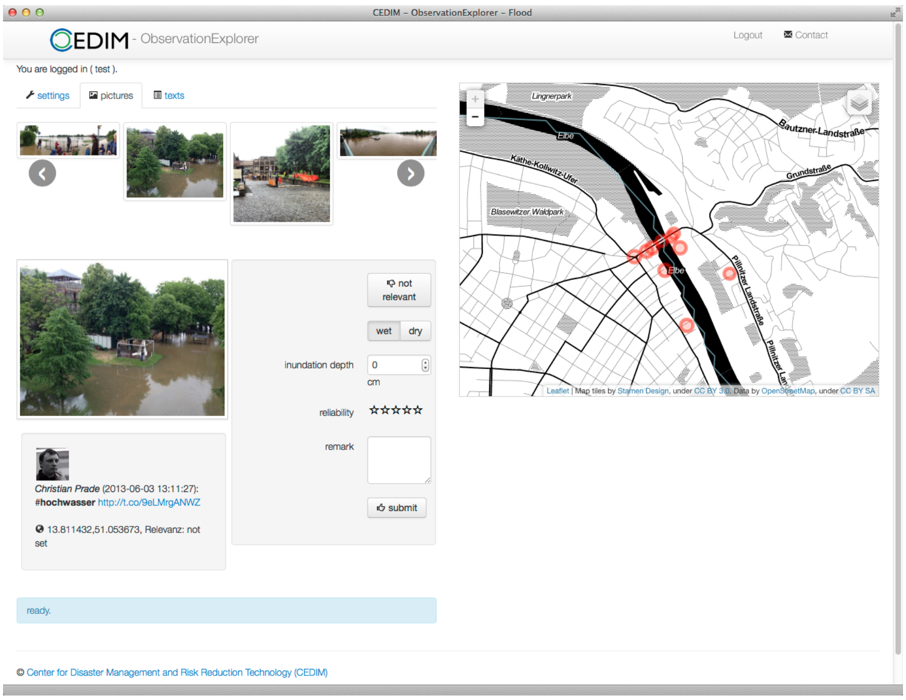

Figure 2. PostExplorer: media view and map view (map tiles by Stamen Design), under a Creative Commons Attribution (CC BY 3.0) license. Data by OpenStreetMap, under Open Data Commons Open Database license (ODbL).

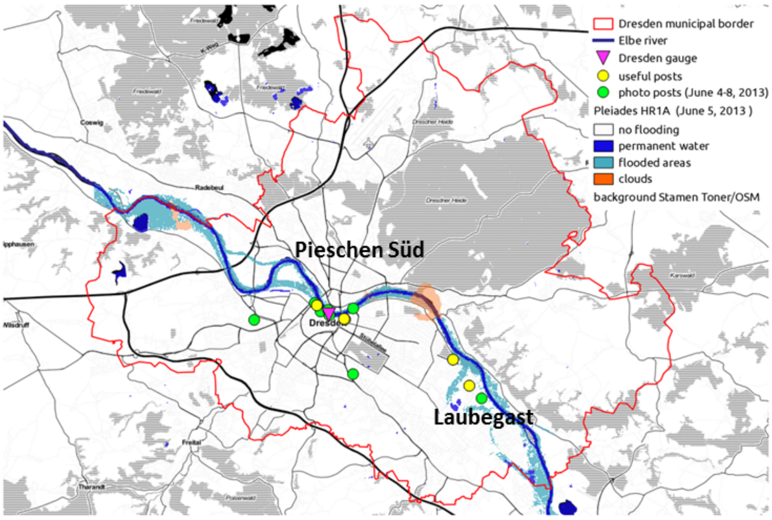

Figure 3. Study region and data sources for flood inundation depth mapping.

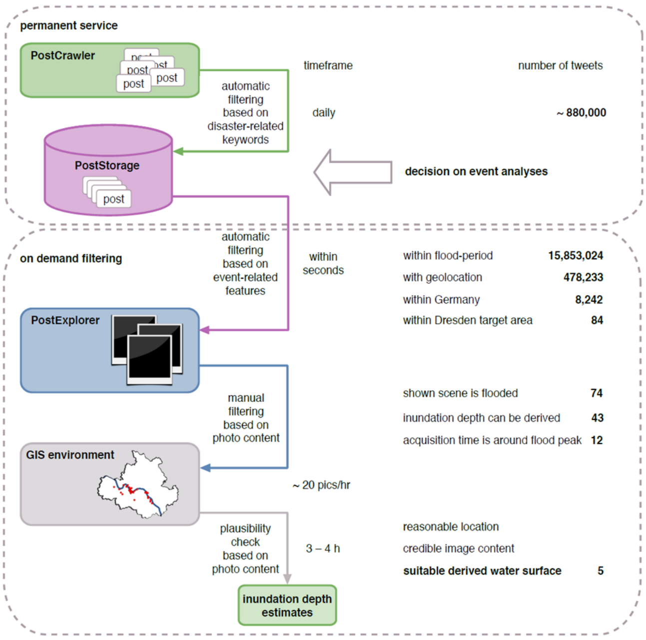

Figure 4. Process chain, timeframe and number of tweets for the Dresden flood in June 2013 handled within PostCrawler, PostStorage, PostExplorer and GIS environment for automatic and manual filtering of tweets.

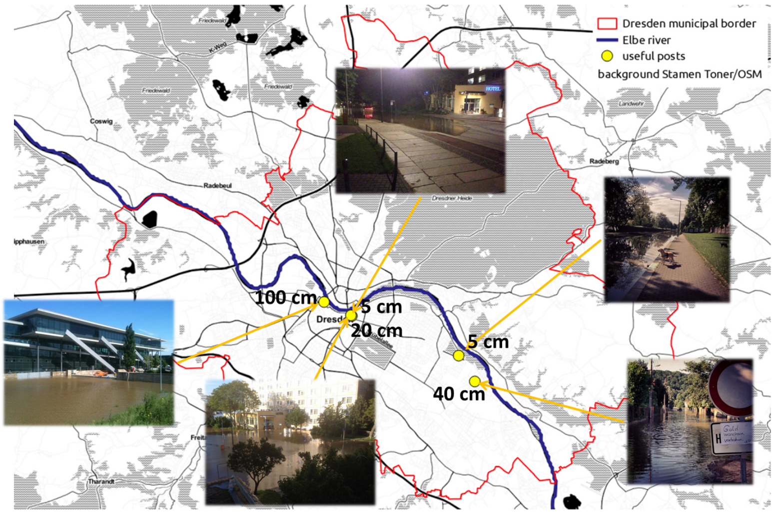

Figure 5. Location of useful photos retrieved from filter and inundation depths estimates (Photos by Denny Tumlirsch (@Flitzpatrick), @ubahnverleih, Sven Wernicke (@SvenWernicke),Leo Käßner (@leokaesner)).

Discussion Paper | Discussion Paper | Discussion Paper | Discussion Paper |

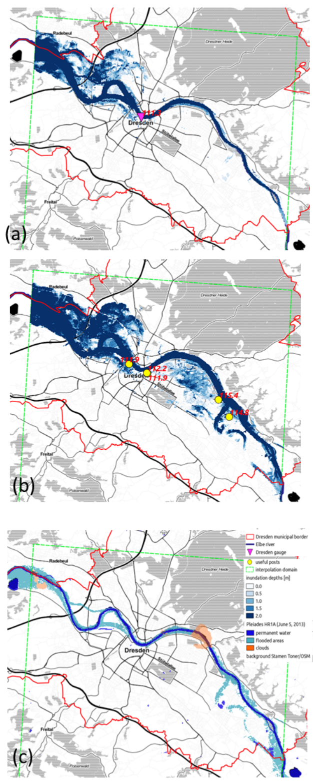

Figure 6. Inundation maps and inundation depths derived from online water level observations (top panel), social media content (middle panel) and inundated area for reference remote sensing flood footprint (bottom panel).