Venice Was Flooding ... One Tweet at a Time

VALERIO LORINI, European Commission - Joint Research Centre, Ispra, Italy and Universitat Pompeu Fabra, Barcelona , Spain
PAOLA RUFOLO, Fincons Spa, Milan, Italy
CARLOS CASTILLO, Universitat Pompeu Fabra, Barcelona, Spain

Before urban flooding actually happens, weather forecasts with varying degrees of precision are available to emergency managers. In the aftermath of the event, authoritative information including Earth Observation (EO) data can be used to estimate precisely the flood extent, possibly after several hours. This study aims to determine how social media information can reduce the inherent uncertainty of the information in the immediate aftermath of an urban flood event. Specifically, the study investigates how to collect relevant social media images and to interpolate such data in order to create a map.

The premise of the study is that social media platforms, when combined with digital surface models, can provide control points for creating a reliable near real-time estimate of the flood extent. In the study, we compared a flood extent map derived from social media with that derived from authoritative altimetry data during one of the worst floods to hit Venice, which occurred in November 2019.

The results of the experiments show a good overall accuracy using several digital surface models. Given the global coverage of such models and the low resources required, we think the methodology proposed could be beneficial for emergency managers. Specifically, we describe how a flood extent map can be made available within 24 h, or even less, after urban flooding strikes a densely inhabited area, where data generated by the public are available.

CCS Concepts: • **Information systems → Information retrieval**.

Additional Key Words and Phrases: social media, disaster risk management, satellite mapping, image classification, machine learning

ACM Reference Format:
Valerio Lorini, Paola Rufolo, and Carlos Castillo. 2022. Venice Was Flooding ... One Tweet at a Time. *Proc. ACM Hum.-Comput. Interact.* 6, CSCW2, Article 382 (November 2022), 16 pages. https://doi.org/10.1145/3555107

1 INTRODUCTION

Citizens constantly use social media, including for broadcasting content during disasters and emergencies. The vast amount of such data originating from the public can be used to provide access to timely and relevant information, offering additional decision-making support to emergency managers. Recent advances in Natural Language Processing (NLP) and computer vision technologies provide an opportunity to improve situational awareness for disaster management and response teams. Images extracted from social media can be vital in the immediate aftermath of an event when authoritative data and products based on Earth Observation (EO) are not yet available. Mapping ground truth information is crucial for the early assessment of impacts, in terms of their intensity and spatial distribution.

Authors' addresses: Valerio Lorini, European Commission - Joint Research Centre, Ispra, Italy, Universitat Pompeu Fabra, Barcelona , Spain, valerio.lorini@ec.europa.eu; Paola Rufolo, Fincons Spa, Milan, Italy, paolarufolo@gmail.com; Carlos Castillo, Universitat Pompeu Fabra, Barcelona, Spain, chato@acm.org.

2573-0142/2022/11-ART382
https://doi.org/10.1145/3555107

The EU's Copernicus Emergency Management Service (CEMS) On-Demand Mapping, which has operated since February 2015, consists of a set of information services funded by the European Commission. As part of the On-Demand Mapping component of the CEMS, satellite imagery and other geospatial data are used to provide mapping service free of charge for natural disasters, human-made emergencies and humanitarian crises throughout the world. Only authorized users such as civil protection, entitled emergency response organizations or international charters can activate the service. The maps are available in two temporal modes: Rapid Mapping (RM), and Risk and Recovery Mapping (RRM). The former provides geospatial information within hours or days of the activation following a disaster, while the latter provides geospatial information supporting disaster management activities not related to immediate response.[1]

Floods represent 36% of activations of the CEMS RM service.[2] Flood extent is normally delineated using both Synthetic Aperture Radars (SAR) satellite images (especially those from the Sentinel-1 satellite) – useful for detecting water-covered areas even at night or in the presence of clouds - and image data from optical EO sensors, which allow the identification of damages for impact assessment. Unfortunately, the effectiveness of such satellite image analysis is of limited use in urban areas, at the point that these are commonly masked and not analyzed. For optical sensors, the limitations are due to shadows cast by buildings, trees or narrow streets. For SAR sensors, the side-looking viewing geometry and the multiple scattering in built-up areas do not allow the presence of water to be properly distinguished.

On 13th November 2019, the mayor of the Italian city of Venice declared a state of emergency after an exceptionally high tide, recorded as the worst in 50 years, flooded the city.[3] A deep cyclonic circulation had affected the Mediterranean area the previous day, resulting in severe weather over the Italian peninsula. One of the most affected areas was North-Eastern Italy, particularly the Friuli Venezia Giulia and Veneto regions. A full moon (+26 cm surge), combined with the exceptionally high level of the Mediterranean sea in November 2019, and a deep small-scale atmospheric pressure moving rapidly northward and passing over the Venice lagoon just west of the city (+30/35 cm surge), led to a high tide with a maximum recorded value of 189 cm on 12th November at 10:50 p.m. (hereafter we refer to local time) - the highest recorded since a similar event in 1966 [6]. According to the altimetry available to the municipality of Venice, as a result of an *"Acqua Alta"* (high tide) of 189 cm, about 82% of the public pedestrian traffic areas were flooded. The impact on the city was dramatic, with two fatalities in the Pallestrina neighbourhood, severe damage to the crypt of the San Marco basilica, three ferries sunk[4] and 2,494 claims for economic loss of totalling 9 million euros in damages by residents and businesses, in the initial weeks following the event.[5]

The RM service of the CEMS was activated by the Italian National Civil Protection Department on 14th November at 01:15 p.m.[6] as other exceptional high tides were forecasted for the following day. Among the set of available images from space satellites, the closest to the time and area of the event was acquired by the GeoEye optical satellite sensor on 14th November at 11:13 a.m. Flooded areas were not adequately visible due to the narrow streets of the city and the high off-nadir acquisition angle of the image (not vertical). Thus the optical imagery was complemented by ancillary data: tide-levels combined with contour lines. The flood extent delineation that was delivered on 15th November at 09:18 p.m. shows the detected situation when the tide recorded had fallen to about 115 cm, with an estimated flooded area below 30%. This significantly failed to

[1] https://emergency.copernicus.eu/mapping/ems/service-overview
[2] https://emergency.copernicus.eu/mapping/ems/rapid-mapping-portfolio
[3] https://www.nytimes.com/2019/11/13/world/europe/venice-flood.html
[4] https://www.ilgazzettino.it/nordest/venezia/acqua_alta_allarme_notte_venezia-4858433.html
[5] https://live.comune.venezia.it/it/dati-richieste-danni-acqua-alta-12-novembre-2019-venezia
[6] https://bit.ly/3AzHJA4

capture the maximum extent flood later reported by the municipality.[7] What we learn from this is that, the use of remote sensing to delineate flood extent in a city can be incomplete or inaccurate, especially for fast-developing events such as urban floods.

Between 12th and 14th November 2019, we collected posts on Twitter using specific filters related to the Venice flood event. In order to fill the gap in information between the immediate aftermath of the event and the moment when authoritative data were available, we delineated a potential flood extent based on images that were classified as relevant to the event and data available for free to the public. In this way we were able to estimate a maximum flood extent similar to that recorded and validated by authorities. This work presents a scalable methodology for combining deep learning models for image classification with global or local Digital Elevation Models (DEMs) and other geospatial information, for a near real-time delineation of the flooded area. The results show how the use of social media information for urban floods can complement EO data and can help to improve situational awareness.

Emergency managers use flood maps based on either hydraulic models or remote sensing data. Hydraulic models require detailed digital information of the impacted area and forecasts that may not be available readily or at the desired spatial granularity. On average, the minimum time needed by CEMS RM service to provide crisis information after an activation request by an authorized user[8] is 24 h [22] and their temporal resolution is often limited depending on the timestamp of the image. Due to the technical issues described earlier in this section, remote sensing analysis is of limited use in urban areas to the point that these areas are commonly not analyzed and left out of the product map. We present a fast methodology complementing both approaches. It offers an additional source of in-situ data that can serve as input for hydraulic models and provides a reference layer for filling spatial and temporal gaps in EO-based products not available during urban floods. The experiments presented demonstrate how such a layer could estimate a flood extent map within the first 24 h of an urban flood. In addition to the benefits mentioned above, the proposed approach uses data that is free of charge except for the geocoding step, therefore having a low economic impact compared to the cost of EO imagery acquisition and analysis.

In the the remainder of this paper, related research work is presented in Section 2, technical details of the data collection, the methodology and the experimental results are presented in Section 3, and finally, a general discussion of the potential of social media analysis for assessing floods extents and future developments is presented in Section 4.

2 RELATED WORKS

The potential of social media for situational awareness during emergencies has been studied by several researchers [5] [18]. Research has also been carried out on how emergency managers could use the information shared by witnesses to plan relief operations [16]. Text and images shared on Twitter have been recognized as containing important information pertinent to humanitarian response [1]. Recent studies have shown encouraging research results related to the use of social media sensors to map flood extent. Brouwer et al. [4] presented a methodology for detecting riverine flood extent using locations derived from Twitter and a normalized digital terrain model. Hydrologically connected tweets are interpolated according to a drainage-normalized representation of the topography. This approach applies to floods driven by an overflow of water from riverbeds, and the focus is on directions of flow in nearby areas rather than on urban floods. Around the same time, Rosser et al. [17] presented a work to estimate flood extent based on a Bayesian model fusing

[7] https://bit.ly/3dMpzkX

[8] EU Member States, the Participating States in the European Civil Protection Mechanism, the Commission's Directorates-General (DGs) and EU Agencies, the European External Action Service (EEAS), as well as international Humanitarian Aid organizations

remote sensing, social media and topographic data sources. The method uses geocoded photographs sourced from social media Flickr, optical remote sensing and high-resolution terrain mapping to estimate the probability of flooding through weights-of-evidence analysis. The results demonstrate that the incorporation of multiple sources of data can aid the prediction of flood extents. Their work does not consider temporal aspects of the data within the modelling process, as the case study involved a prolonged flooding event. Later, Pastor-Escuredo et al. [15] suggested integrating social media data into a framework consisting of authoritative and non- authoritative data for detecting impacts of a disaster. In their work, the function of social sensors regarding mapping the extent is limited to mobile phone use for mobility detection, which is a valuable source of data but makes the methodology hard to scale for cases where such information is not accessible.

Earlier this year Xiaoyan et al. [23] introduced an innovative approach to estimating flood-affected populations, providing high-resolution impact information. The described case study shows that considering mobility patterns during assessment can improve the precision of disaster estimation. Inundation locations and roads blockage are detected by combining flood hazard maps with social media data, thus applying historical statistical data and real-time citizen-generated data. social media data regarding inundations were obtained from the official Weibo account of the Wuhan Traffic Management Bureau.

Heavy rainfall is often the main driver for urban floods, which can happen where there are no rivers, or the flood can be a combination of events such as blockage of the sewage system or coastal floods. Our work aims to map flood extent regardless of the driver of the event. Moreover, our methodology can be applied in near real-time, overcoming the timeliness limitation of post-event analysis. Our work also provides a methodology that can be reproduced in different cities, as it uses open and globally available data.

Several research studies have shown the potential, opportunity and limitations of satellite images and radars for natural disaster analysis. Over the years, increasingly powerful methods and sensors have been launched on satellites, to avoid weather-related signal attenuation. Commonly used technologies include LIght-Detection And Ranging (LIDAR), which can detect the altitude of objects from a long distance, and SAR, which creates two-dimensional images using a signal frequency unaffected by light conditions (day or night) or cloudy weather. Mason at al. [12] studied a method to detect floodwater in urban areas with a SAR simulator in conjunction with LIDAR data. The method allows predicting areas of radar shadow and layover in the image caused by buildings and taller vegetation. The results indicate that flooding can be detected in an urban area with reasonable accuracy. However, the algorithm design assumes that high-resolution LIDAR data are available for the area under analysis. In Mason et al. [11] the same authors use open-access Sentinel-1 SAR data, the World-DEM digital surface model (DSM), and open-access World Settlement Footprint data to identify estimates of flood levels in urban areas locally. Their method searches for increased SAR back-scatter in the post-flood image due to double scattering between water (rather than non flooded ground) and adjacent buildings and reduced SAR back-scatter in areas away from high slopes. The method reports high accuracy in moderately dense housing areas, while the accuracy decreased in dense housing areas when street widths are comparable to the DSM resolution. Yunung et al. [10] employ SAR intensity time-series statistics to create a flood probability map. The resulting extent is selected by applying a global cutoff probability of 0.5. However, smooth surfaces like asphalt roads, SAR shadow, aquatic plants, and soil moisture changes introduce inaccuracies in the prediction. Furthermore, the long time for processing images to build the SAR intensity time-series statistics makes the methodology unsuitable for real-time deployment. The methodology proposed in this research uses social media information fused with digital surface models as sensors for detecting ground truth in near real-time to reduce uncertainty and contribute to solving the issues related to EO technologies.

The so-called "digital divide", and a lack of resources especially in vulnerable area more impacted by global warming [21] [8], have focused the attention of scientists and crisis responders on research tools and methodologies for flood risk management at a global scale. Flood risk assessments for cities produced using Global Digital Elevation Models (GDEMs) are likely to over-predict risks. Past studies found variability in the accuracy of models using different GDEMs, and all substantially estimated higher impacts than the DEM produced from aerial LIDAR [13]. As the world's cities grow, the importance of accurately understanding flood risk has become a high priority. GDEMs enable flood risk assessments to be undertaken globally, defining standard methodologies allowing data integration. Uncertainties in flood risk assessment using GDEMs need to be addressed and reduced in near real-time by local and national authorities and communities, to prevent misinformed decision-making.

International organizations have put in place emergency management services with the aim of providing support to crisis responders who are in need of resources. The European Union's Earth observation programme, called Copernicus, offers information services that draw from both satellite EO and in-situ (non-space) data. As part of Copernicus, the CEMS supports local authorities and communities needing information to develop environmental legislation and policies or to take critical decisions in the event of an emergency, such as a natural disaster or humanitarian crisis. The Early Warning systems and On-Demand mapping components of the CEMS produce flood hazard maps that have been developed using hydrological and hydrodynamic models, driven by the climatological data of the European and Global Flood Awareness Systems (EFAS [19] and GloFAS [3]). All maps are in raster format with a grid resolution of 100 m (European-scale maps) and 30 arcseconds (global-scale maps). These maps can be used to assess the exposure of population and economic assets to river floods, and to perform flood risk assessments.

Our research is aligned with and complementary to the previous research on providing near real-time information during floods, particularly in densely inhabited areas. A key advantage of our contribution lies in the advantage of the real-time aspect of social media data, together with physical model and EO data. Our research can answer the following important topical question: ***Is it possible to leverage real-time information from social media, fusing it with digital surface models derived from earth observation data to provide a fast estimate of a flood extent?***

The proposed workflow uses social media information to find flooded points (latitude and longitude). It then infers the spatial extent of the flooded area operating a vertical data interpolation based on digital surface grid-based information. Assuming that the same amount of rainfall fell on the city, if point A is flooded and point A is in a higher grid-cell than point B, and their grid-cells are close, we assume that both grid-cells are flooded. The tool's accuracy is determined based on the grid-cells estimated as flooded against the flooded grid-cells provided as reference.

3 DATA, METHODS AND EXPERIMENTS

In order to analyze the quality of flood extent mapping based on social media information during urban flooding, we have carried out two experiments, both in the context of the flood that hit Venice in 2019, as was described earlier in Section 1. In this Section, we describe in detail the data collection, the applied methodology, and the results that were obtained from both experiments.

3.1 Data collection

3.1.1 Weather forecasts. On 12th November 2019, a deep cyclonic circulation affected the Mediterranean area, which resulted in severe weather over the Italian peninsula. In Venice, the extreme weather condition produced a high tide whose maximum recorded value was 189 cm at 10:50 p.m.. At the start of the event, on the morning of 12th November, the municipality emergency managers expected a high tide peak of 170 cm at 11:00 p.m. and another peak of 160 cm for the following

morning at 10:30 a.m. Several warnings were issued, schools were closed and travel restricted, and plans for emergency response were activated.

3.1.2 Social Media. One of the key elements of the methodology proposed here is the consideration of social media as valid ground truth information. We searched for messages posted on Twitter from 12th November at 01:00 a.m. to 14th November at 01:00 a.m., either geocoded inside Venice island or mentioning the keywords "Venice" or "AcquaAlta". We collected roughly 75,000 tweets, 14,000 of which contained pictures.

3.1.3 Digital models. Collected Social Media data are then combined with digital models representing the surface of the city. In order to study the quality and the scalability of the methodology proposed we used several DEMs:

(1) SRTM (Shuttle Radar Topography Mission) is a global research endeavor that yielded nearly-global DEM with a 30 m resolution. SRTM data covers the globe and is free of charge (though registration is required).
(2) The Copernicus DEM EEA-10 instance (hereafter referred to as Copernicus DEM), available free of charge from the European Environmental Agency[9], is a Digital Surface Model (DSM) which represents the surface of the Earth including buildings, infrastructure and vegetation with a resolution of 10m. Since the Copernicus DEM includes building, to reduce as much as possible the error in the elevation of the control points, we adjusted the values of height extracted.
Let $PBUILT(lat, long)$ represent the 0 to 1 share of buildings in a 10 m cell at $(lat, long)$ according to the GHSL built-up-area product[10]. $ELEV(lat, long)$ is the original elevation value from the Copernicus DEM model
Then the adjustment formula that provide the corrected elevation value $ELEV'(lat, long)$ can be described as:

$$ELEV'(lat, long) = ELEV(lat, long) - PBUILT(lat, long) \cdot ELEV(lat, long) \quad (1)$$

While the formula works as it is for cities at sea level, to reproduce such adjustment in higher urban areas, we should lower the elevation $ELEV(lat, long)$ values by subtracting the average altitude of the site or the height of the nearest cell with a near-zero share of buildings before applying the formula.
(3) TINITALY [20] is a seamless DEM of the whole Italian territory. This DEM, which was produced starting from separate DEMs of single administrative regions of Italy, is freely available with a 10 m grid-cells. TINITALY is published with a CC BY 4.0 license and can be used freely, even partly, but it must be cited.

3.2 Methodology

The production of a near real-time flood extent map is carried out in four separate steps: collection of tweets; extraction of social media flood points; interpolation of social media flood points; production of flood extent. Figure 1 depicts the four steps while their details are provided in the remainder of the section.

Firstly, we collected tweets as described in Section 3.1.2 above. For conducting the experiments we relied on the so-called historical search available only to Twitter PowerTrack API users. Nonetheless

[9] https://www.eea.europa.eu/
[10] https://ghsl.jrc.ec.europa.eu/ghs_bu2019.php

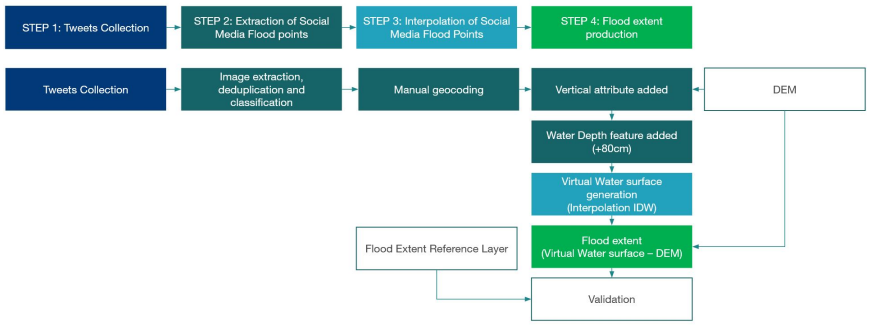

Fig. 1. (best seen in color) The flood extent map production consists of 4 steps: Indigo - Tweets Collection; Teal - Extraction of social media Flood points; Blue - Interpolation of social media Flood Points; Green - Flood extent production. Geocoding is executed manually in this experiment.

the collection of tweets in real-time for replicating the experiment does not require any special data access, since tweets can be filtered from the publicly available Twitter streamer[11].

The work aims to map the flood extent in a city, and therefore it is of the utmost importance that we geocode information as precisely as possible. Although Twitter enables users to post tweets with their current locations (longitude and latitude), only an average rate of 0.85%–3% tweets are being geocoded per day [24] in the USA, where Twitter is most used. Thus, we need to complement the geocoded dataset using other techniques to ensure scalable global products. Past works proved that location mentions were useful to geocode information [9]. However, after a first test using several hundred tweets, we understood that we could not rely solely on automated NLP for a precise geocoding. Figure 2 shows one of the many examples of how a location mentioned in a text does not correctly represent the location of the information. Previous works [2] found that tweet images contained more damage-related information than their corresponding text. Thus, we opted to extract social media images rather than text as they could be better inspected for geocoding within the city and for the extraction of crisis information about the event. During the two days we collected 14,000 images, resulting in 10,000 images after duplicates were removed using a tool for checking and deleting near-duplicate images based on perceptual hash[12].

The second step of the methodology is the identification of social media flood points to be considered as a control point for the flood mapping activity. Once the tweets were collected, we used a Convolutional Neural Network (CNN) model for disaster image classification [18] to classify the images of a flooded location. The model assigns a probability of an image to belong to one of five classes (Flood, Wildfire, Storm, Earthquake, Other). In particular, we set a threshold of flood probability equal to 0.9 for identifying those relevant to the event. We found 2,302 images depicting flooded areas, some examples of which are shown in Figure 3. The vast majority of the relevant images were then geocoded manually. The location of the images was based on the identification of recognizable Points of Interest (POIs) like shops, monuments, street names, bridges, public transport stops, and their comparison with Google Street View. When the image showed a wide area, such as a square photographed from a building, the location of the image was placed in the flooded area, i.e. the center of the square. When, in a flooded street, a shop could be identified, the location was placed in front of it. The annotation was verified by a contractor of CEMS, who found

[11] https://developer.twitter.com/en/docs/tutorials/stream-tweets-in-real-time
[12] https://github.com/knjcode/imgdupes

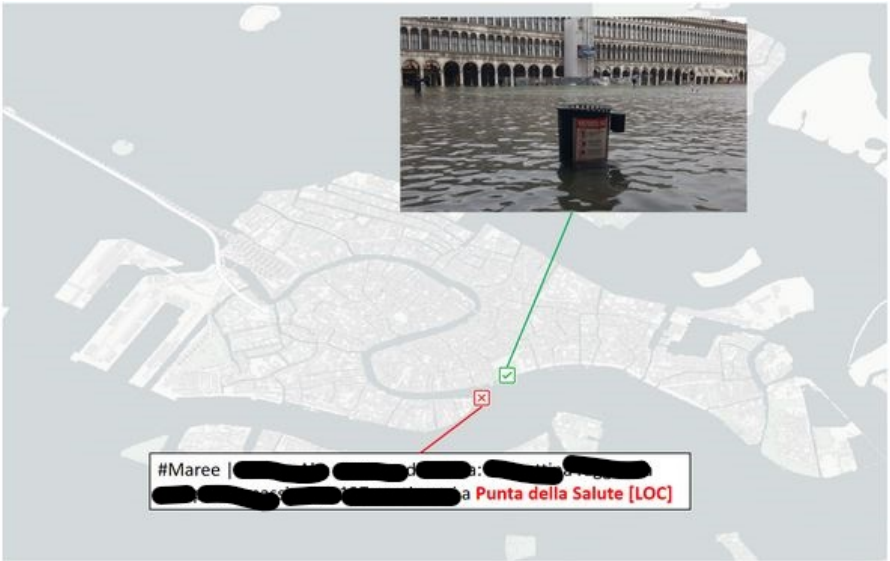

Fig. 2. Example of the wrong facility identified by NLP, as it was mentioned in the tweet text "flood reached the maximum peak cm at Punta della Salute". The correct location is derived from the tweet picture

Fig. 3. Examples of images classified as relevant and manually geocoded.

that 97% of images were properly geocoded[13] Since this work proposes a scalable methodology for detecting a potential flood extent map, manual geocoding has been performed considering the time component. As already described earlier in Section 3.2, only an average rate of 0.85% to 3% of the tweets are originally geocoded. To speed up the geocoding of images, we used an NLP tool[14] to extract mentions of place names from text and leveraged such information to locate the flooded point. However, we have not monitored the time consumed for the single geocoding for the experiments. Thus we cannot estimate how much such automated pre-processing contributed to speeding up the geocoding. Images that could not be geocoded or that were clearly referring to weather conditions in areas outside Venice, were excluded. Finally, after the manual geocoding we could identify almost 800 social media flood points, and 265 unique points. The focus of our research is to support the development of a new product that could be made available within

[13] the report, written by Trabajos Catastrales S.A., is available upon request.
[14] https://github.com/deepmipt/DeepPavlov

the first 24 h after an urban flood happened when neither EO-based nor authoritative maps are available due to technical challenges. The feasibility of such a product is confirmed if we consider that the manual geocoding process, done by non-local personnel using Google Street View, took 6 h (one person). If needed, we can assume that a service provider could allocate more resources to this task during an actual case. Furthermore, manual geocoding can be quickly done during a real crisis using crowdsourcing to leverage the help of local digital volunteers, coordinated by practitioners and emergency-oriented volunteers, such as Virtual Operations Support Teams (VOSTs)[15]

Figure 4 shows a comparison between tweets originally geocoded on a no-flood day and the tweets identified as social media flood points. We divided the city of Venice into a grid of 50x50 m cells and counted the tweets in each cell. We carried out the comparison in order to ensure that the experiment could be scaled to other less-visited cities. The random scattered distribution of geocoded tweets on a no-flood day clearly demonstrates that the image classification step leads to an unbiased distribution. In effect we demonstrated that more pictures taken does not correlate with more flood points.

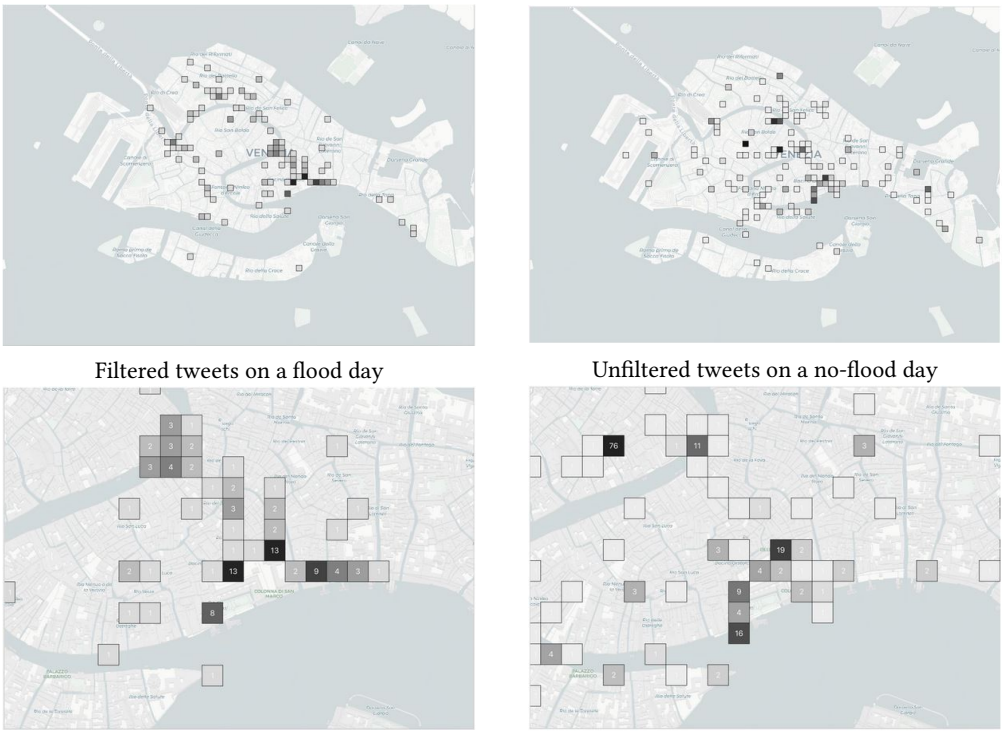

Fig. 4. Geographical distributions of filtered and unfiltered tweets. (Note that the darker the cell, the higher the number of tweets in the cell)

The social media flood points, defined by latitude and longitude coordinates, were given a vertical attribute by sampling the available DEM. We created three datasets of points, one per DEM as defined in Section 3.1. A default water depth (DWD) of 80 cm was added to the ground vertical component based on visual inspection of the images (i.e. water above knees of passengers, doors).

[15] http://vosteurope.org

We considered such approximation done in other works [4] acceptable to contain the time of processing. A better estimate of the water levels in each image would undoubtedly lead to more accurate results, but it would not apply to a semi operational process. This aspect is further discussed in Section 4. In riverine flood hydrology, flood simulation models are used in combination with terrain analysis to detect the flow of water towards lower or unprotected areas. Control points can be used to detect water-levels in statistically generated flood hazard maps. The main driver in the case of urban floods is the amount of rain falling on an area combined with malfunctioning man-made artifacts (i.e. sewage, buildings, roads). Thus, instead of focusing on the water flow, the third step of our methodology is to create a virtual water surface interpolating the social media flood points prepared in the second step to estimate values at other unknown points. We favored an Inverse Distance Weighted (IDW) method, where the sample points are weighted during interpolation such that the influence of one point relative to another declines with distance.

The fourth and last step of our methodology is the identification of the flooded area. This is obtained comparing for each point of our map the DEM and the virtual water surface generated at the previous step. When the water surface is higher we assume the cell is flooded.

3.3 Experiments

During the interpolation, we tested different values of the coefficient IDW-P, to create a few different surfaces and adjust this parameter to suit our analysis. A larger coefficient means it takes a larger distance for the values of the surface to become dissimilar from nearby points. A small coefficient means the values of the surface will quickly change as distance increases.

3.3.1 Experiment 1. We interpolated the social media flood points with several weighting values to determine the best accuracy for the maximum extent of the flood. We created an elevation reference layer (altimetry) using the contour lines relative to the elevation of the pavement of the historic centre of Venice with respect to the median sea level, whose accuracy is 1 cm vertical and 2 cm horizontal[16]. According to authoritative sources (as mentioned in Section 3.1), a maximum level of 189 cm was recorded at 10:50 p.m. on 12th November 2019. Thus, by selecting only the points of the elevation reference layer below that value, we were able to define the maximum flood extent to use as reference map. All the layers were transformed to grid-based maps (rasters) during the experiments, and statistics were done on such grid cells. In order to compute Precision and Recall for the experiment, we considered four types of result (True Positives, False Positives, True Negatives, and False Negatives), as described below:

(1) **True Positive** values of the cells that are detected under water from both our methodology and authoritative data (altimetry).
(2) **False Positive** values of the cells that are detected under water from our methodology but not according to authoritative data (altimetry).
(3) **True Negative** values of the cells that are not detected under water from both our methodology and authoritative data (altimetry).
(4) **False Negative** values of the cells that are not detected under water from our methodology but they are, according to authoritative data (altimetry).

Table 1 displays the results of the methodology proposed using several values of the weighting coefficient IDW-P for the interpolation of the social media flood points. We ran the majority of simulations with the Copernicus DEM, because we are convinced that the CEMS could benefit from this work. All the simulations were compared against the maximum extent identified by reference contour lines, as we used the social media flood points collected over the whole period of the event.

[16] http://smu.insula.it/index.php@option=com_content&view=article&id=15&Itemid=111.html

The first row of Table 1 outlines a trivial experiment where we assumed the entire city was flooded. Given the magnitude of the event, it seems that such an assumption brings good results. Thus, for clarity, Table 1 reports also the Matthews Correlation Coefficient (MCC), that ranges in the interval [-1,+1], with extreme values -1 and +1 reached in case of perfect misclassification and perfect classification, respectively, while MCC=0 is the expected value for the coin-tossing classifier. According to [7], this coefficient shows more reliable evaluations versus overall accuracy (OA) and the F1 score, particularly on imbalanced data-sets, such as in our case where the majority of the pixels were flooded. The values in bold represent the best score for each column, maximized in case of true values and minimized in case of false values. We notice that among the runs with the Copernicus DEM, the IDW-P coefficient carrying the best results is 10. Specifically, it detects the highest number of true negative cells, which is valuable information given that almost all the city was flooded. The last two rows of Table 1 represent the methodology's simulation using the best IDW-P (10) but with the other DEMs. The local DEM, TINITALY, gives by far the best results, especially in detecting the true positive values. An interesting feature to emerge was how the SRTM DEM offers an excellent alternative to the Copernicus DEM, although with a lower spatial resolution (30 m against 10 m).

DEM	IDW-P	TN (%)	FN (%)	FP (%)	TP (%)	OA (%)	MCC
AllFlooded	No	0	0	8.91	91.09	91.09	0
COP	2	4.45	38.03	4.46	53.06	57.51	.047
COP	4	5.65	36.85	3.26	54.24	59.90	.132
COP	10	**5.97**	36.58	**2.94**	54.51	60.48	.154
COP	15	5.82	36.14	3.10	54.94	60.76	.147
COP	20	5.80	35.80	3.11	55.28	61.09	.149
COP	25	5.78	35.57	3.13	55.52	61.30	.149
COP	30	5.78	35.53	3.14	55.55	61.33	.149
SMRT	10	5.78	31.04	3.13	60.04	65.82	.182
TINItaly	10	3.37	**7.56**	5.54	**83.53**	**86.89**	**.269**

Table 1. Accuracy comparison between interpolations made with different weighting coefficient IDW-P. (Note: DEM = digital elevation model; IDW-P=weighting parameter; TN = True Negative; FN = False Negative; FP = False Positive; TP = True Positive; OA = Overall Accuracy; MCC = Matthews Correlation Coefficient

The thematic validation was performed by calculating pixel-based confusion matrices from which we can extract the overall accuracy (OA) for the different IDW-P values and DEMs. Figure 5 shows an overview of the validation for IDW-P=10 coupled with the Copernicus DEM.

Areas highlighted in green represent the pixels where there is agreement between the estimated flood and the reference layer (TP or TN). In purple we represented the omission errors (FN) and in orange the commission errors (FP). The dots in yellow represent the control points. It can be seen how the main omissions are in the areas where there are no control points. This can be explained as in these areas, due to the weighting parameter, the interpolation layer does not report the presence of water. Indeed its value, despite being higher than zero, does not reach optimal values like the other coefficients (that range between 0 and 1), demonstrating the difficulty of our method to detect the non-flooded areas (high value of FN) simply because we use control points only where images show an inundation.

3.3.2 Experiment 2. We performed a second experiment simulating a real-time scenario where only forecasts but no authoritative data about the high-tide were available. Assuming our methodology

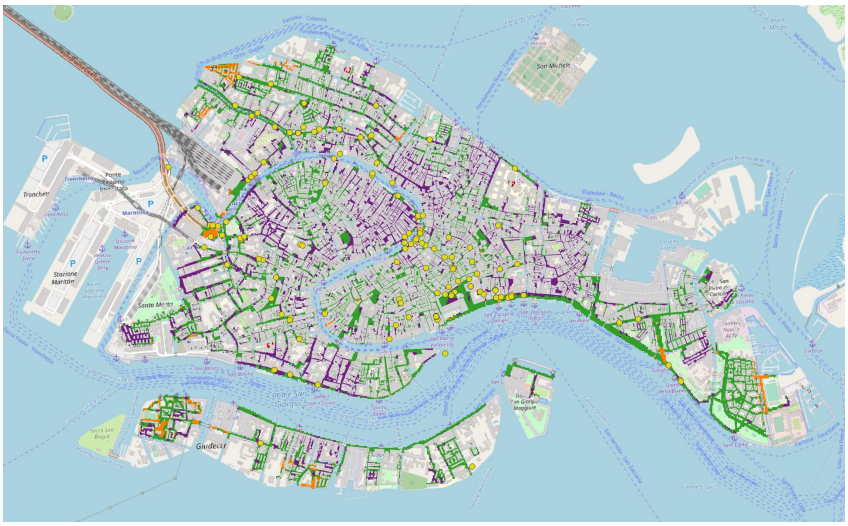

Fig. 5. Overview of the validation for IDW-P=10 and Copernicus DEM (best seen in color).Areas in green show agreement between estimated flood and reference layer (TP or TN). Areas in purple show the omissions (FN) and in orange commissions (FP). Dots in yellow represent the control points.

DEM Copernicus 10m DEM SRTM 30m DEM TINITALY 10m

Fig. 6. Overview of the for IDW-P=10 and Copernicus, SRMT, and TINITALY DEMs

with an IDW-P value of 10 gives the best approximation, we produced a flooded surface map using only social media available the first day until 13th November at 01:00 a.m., and compared it with the reference altimetry below 140 cm forecasted by meteorologists for 12th November. We computed the true and false values comparing simulations with two different contour lines. The first with flood expected for altimetry data below 140 cm and the second with the reported (after the event) value of 189 cm. Table 2 shows the best results in terms of true values, overall accuracy and MCC.

DEM	CL	IDW-P	TN (%)	FN (%)	FP (%)	TP (%)	TN+TP	OA (%)	MCC
COP	189	10	5.42	17.16	3.49	73.92	**79.35**	**.793**	**.286**
COP	140	10	11.90	10.69	23.54	53.87	65.77	.658	.194

Table 2. Accuracy comparison between interpolations made with best IDW-P. with forecasts and 24 h social media for the day November 12 2019. DEM = DEM; IDW-P=weighting parameter; TN = True Neg; FN = False Neg; FP = False Pos; TP = True Pos; OA = Overall Accuracy; MCC = Matthews Correlation Coefficient

We can see from Table 2, based on the values obtained from the social media flood points after 24 h, we could already assume that the forecasted values of 140 cm were exceeded by far. The flood extent map produced after 24 h was closer to the one created by the municipality with a reference water level of 189 cm as reported afterwards.

4 CONCLUSIONS AND FUTURE WORK

The accuracy we obtain when determining whether a cell is flooded or not, using maps that are freely available for the entire world, is 61.3% at the 10 m resolution and 65.8% at the 30 m resolution. Using a more detailed, country-specific map, we arrive to 86.9% resolution (Table 1). We can say that our experiment answers part of our research question ("Is it possible to leverage real-time information from social media, fusing it with digital surface models derived from earth observation data to provide a fast estimate of a flood extent?") in the affirmative, using both local DEM data and the freely available Copernicus DEM. The experiments demonstrate that it is possible to estimate quickly urban flooding extent using freely available resources at a fraction of the cost usually needed for satellite image processing. Such a methodology can help to resolve the issues presented in Section 2 of this paper, particularly regarding the problem of EO-based products in an urban context and the difficulty of capturing the development of the flood events. The target accuracy for a mapping service like Copernicus Rapid Mapping is higher than 80%[17]. Accuracy in Urban areas appears to be usually lower, therefore most of the times urban floods are not analyzed. Furthermore, during a recent CEMS workshop[18], practitioners expressed interest in having rapid exposure assessment while waiting for the first RM product. It appears that while with a high-resolution DEM our tests show a high level of accuracy, even the worst-case accuracy achieved during our experiments could be valuable information for situational awareness in the event's immediate aftermath. For instance, local emergency responders might use the flood extent map as a starting product and refine the map by gathering complementary in-situ data based on their expertise and knowledge of the distribution of the city's critical infrastructures.

While the accuracy may vary depending on the resolution of the DEM, we are aware of the challenges presented by the geographical scalability of the methodology. Ranging in (i) the availability of tweets, (ii) the availability of local expert volunteers, (iii) the availability of technological tools for geocoding of the social media control points, affect the feasibility of the mapping product. Our future work will aim at defining classes of cases and experimenting the timeliness and applicability of the methodology. The classes should range between the two ends:

- Flood extent map feasible: high number of tweets, high presence of local volunteers, google street view, or other similar tools available (i.e. mapillary[19]), and a high-resolution DEM. The time needed for the products is less than 24h and the accuracy is high.
- Flood extent map impossible: few messages, low presence of local volunteers, no digital images to support geocoding.

An operational service could then estimate the applicability of the methodology within the first 24h and decide whether to add it or not to the data available to the crisis responders. If we consider the case of the CEMS activations for urban floods in European cities, with the support of EU-wide local experts (either volunteers such as VOST EU or contractors that provide the services for a fee), where google Streetview is almost overall available, we suggest the systematic use of the

[17] https://etendering.ted.europa.eu/document/document-old-versions.html?docId=44850
[18] https://emergency.copernicus.eu/mapping/ems/cems-week-2021-conclusions-community-insights-and-service-evolutions
[19] https://www.mapillary.com/

methodology as a product to complement disaster risk management services such as CEMS. At the same time, the tool needs further analysis for scalability and future applicability to more cases.

The accuracy of CEMS maps is routinely verified and validated; a contractor of the CEMS repeated this experiment with a more accurate elevation layer and used just a subset of 77 (out of 265) images. They achieved 76% accuracy, higher than our experiments, suggesting that as long as the elevation layer is accurate the number of required labeled images does not need to be very large.[20]. Therefore, we can safely assume that it is possible to provide an estimation of the flood extent map since Twitter streamer filtering can collect tens of tweets within the first hours after an event. A possible solution for scalability could be that a set of maps is automatically produced every 6 h with data available. The geocoding could be contracted to a service provider to allocate resources case by case. CEMS Service Providers could inspect such products before being released, as currently done for other EO-based products.

Although much social media information is textual, it is challenging to geocode precisely (within a few meters) information, using the locations mentioned in the text. These often refer generically to a road or a place such as a square or a large facility. The use of relevant images from social media is crucial as it offers a better possibility of placing social media flood points. The methodology proposed relies on the automated classification of images to facilitate the identification of informational data. Future research should focus on improving the automation of data filtering, and reducing the overall time and resources used for geocoding data, specifically combining textual and visual information where possible to support the manual geocoding of information.

One limitation which we are already planning to address, is the possible combined effects of rain-driven and riverine floods in cities with a mix of built-up areas and river catchments. In this specific context, the methodology described could integrate an additional parameter for the interpolation of control points, namely the Height-Above-Nearest-Drainage (HAND) [14] terrain model, that takes into account groundwater dynamics.

The experiments described here for the case of Venice show very high precision, because almost all of the city was flooded. For this reason we also investigated other measures, and we aimed to derive results as real positive and real negative numbers. We also analyzed results using the mapping validation technique. Our methodology searches for and utilizes social media flood points, thus maximising agreement in terms of true positive values, rather than minimising omission errors, as specified in Section 3.3. Future work could also focus on optimizing the search for non-flooded areas through the inspection of images.

Finally, it is worth mentioning that the social media flood points presented in this paper can also be evaluated as a potential input layer for hydraulic models to reduce uncertainty introduced by weather forecasts.

All the data used in this paper are available in a public repository[21].

ACKNOWLEDGMENTS

The authors would like to thank Niall Mccormick for proofreading this manuscript and for the helpful suggestions. This work has been partially supported by: "la Caixa" Foundation (ID 100010434), under the agreement LCF/PR/PR16/51110009; the Ministry of Science and Innovation of Spain with project "COMCRISIS", reference code PID2019-109064GB-I00; and the EU-funded "SoBigData++" project, under Grant Agreement 871042.

[20]This report, written by Trabajos Catastrales S.A., is available upon request.
[21]https://zenodo.org/record/7022704

REFERENCES

[1] Firoj Alam, Tanvirul Alam, Ferda Ofli, and Muhammad Imran. 2021. Social Media Images Classification Models for Real-time Disaster Response. arXiv:2104.04184 [cs.CV]

[2] Firoj Alam, Ferda Ofli, and Muhammad Imran. 2018. CrisisMMD: Multimodal Twitter Datasets from Natural Disasters. *CoRR* abs/1805.00713 (2018). arXiv:1805.00713 http://arxiv.org/abs/1805.00713

[3] Lorenzo Alfieri, Valerio Lorini, Feyera A. Hirpa, Shaun Harrigan, Ervin Zsoter, Christel Prudhomme, and Peter Salamon. 2020. A global streamflow reanalysis for 1980–2018. *Journal of Hydrology X* 6 (2020), 100049. https://doi.org/10.1016/j.hydroa.2019.100049

[4] T. Brouwer, D. Eilander, A. van Loenen, M. J. Booij, K. M. Wijnberg, J. S. Verkade, and J. Wagemaker. 2017. Probabilistic flood extent estimates from social media flood observations. *Natural Hazards and Earth System Sciences* 17, 5 (2017), 735–747. https://doi.org/10.5194/nhess-17-735-2017

[5] Carlos Castillo. 2016. *Big Crisis Data: Social Media in Disasters and Time-Critical Situations.* Cambridge University Press. https://doi.org/10.1017/CBO9781316476840

[6] L. Cavaleri, M. Bajo, F. Barbariol, M. Bastianini, A. Benetazzo, L. Bertotti, J. Chiggiato, C. Ferrarin, F. Trincardi, and G. Umgiesser. 2020. The 2019 Flooding of Venice and Its Implications for Future Predictions. *Oceanography* 1 (March 2020). https://doi.org/10.5670/oceanog.2020.105

[7] Davide Chicco and Giuseppe Jurman. 2020. The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. *BMC Genomics* 21 (2020). https://doi.org/10.1186/s12864-019-6413-7

[8] Francesco Dottori, Wojciech Szewczyk, Juan-Carlos Ciscar, Fang Zhao, Lorenzo Alfieri, Yukiko Hirabayashi, Alessandra Bianchi, Ignazio Mongelli, Katja Frieler, Richard A. Betts, and Luc Feyen. 2018. Increased human and economic losses from river flooding with anthropogenic warming. *Nature Climate Change* 8, 9 (01 Sep 2018), 781–786. https://doi.org/10.1038/s41558-018-0257-z

[9] Ullah Irfan, Khan Sharifullah, Imran Muhammad, and Lee Young-Koo. 2021. RweetMiner: Automatic identification and categorization of help requests on twitter during disasters. *Expert Systems with Applications* 176 (2021), 114787. https://doi.org/10.1016/j.eswa.2021.114787

[10] Yunung Nina Lin, Sang-Ho Yun, Alok Bhardwaj, and Emma M. Hill. 2019. Urban Flood Detection with Sentinel-1 Multi-Temporal Synthetic Aperture Radar (SAR) Observations in a Bayesian Framework: A Case Study for Hurricane Matthew. *Remote Sensing* 11, 15 (2019). https://doi.org/10.3390/rs11151778

[11] David C. Mason, Sarah L. Dance, and Hannah L. Cloke. 2021. Floodwater detection in urban areas using Sentinel-1 and WorldDEM data. *Journal of Applied Remote Sensing* 15, 3 (2021), 1 – 22. https://doi.org/10.1117/1.JRS.15.032003

[12] David c. Mason, Sarah L. Dance, Sanita Vetra-Carvalho, and Hannah L. Cloke. 2018. Robust algorithm for detecting floodwater in urban areas using synthetic aperture radar images. *Journal of Applied Remote Sensing* 12, 4 (2018), 1 – 20. https://doi.org/10.1117/1.JRS.12.045011

[13] Fergus McClean, Richard Dawson, and Chris Kilsby. 2020. Implications of Using Global Digital Elevation Models for Flood Risk Analysis in Cities. *Water Resources Research* 56, 10 (2020), e2020WR028241. https://doi.org/10.1029/2020WR028241 arXiv:https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2020WR028241 e2020WR028241 10.1029/2020WR028241.

[14] A.D. Nobre, L.A. Cuartas, M. Hodnett, C.D. Rennó, G. Rodrigues, A. Silveira, M. Waterloo, and S. Saleska. 2011. Height Above the Nearest Drainage – a hydrologically relevant new terrain model. *Journal of Hydrology* 404, 1 (2011), 13–29. https://doi.org/10.1016/j.jhydrol.2011.03.051

[15] David Pastor-Escuredo, Y. Torres, Maria Martinez, and Pedro Zufiria. 2018. Floods impact dynamics quantified from big data sources. (04 2018).

[16] Hemant Purohit, Andrew Hampton, Shreyansh Bhatt, Valerie L. Shalin, Amit P. Sheth, and John M. Flach. 2014. Identifying Seekers and Suppliers in Social Media Communities to Support Crisis Coordination. *Computer Supported Cooperative Work (CSCW)* 23, 4 (01 Dec 2014), 513–545. https://doi.org/10.1007/s10606-014-9209-y

[17] J. F. Rosser, D. G. Leibovici, and M. J. Jackson. 2017. Rapid flood inundation mapping using social media, remote sensing and topographic data. *Natural Hazards: Journal of the International Society for the Prevention and Mitigation of Natural Hazards* 87, 1 (May 2017), 103–120. https://doi.org/10.1007/s11069-017-2755-0

[18] Paola Rufolo, Davide Muraro, and Valerio Lorini. 2021. Social Media Image Analysis in the Immediate Aftermath of the 2020 Beirut Blast, 2021. 50 (2021). https://doi.org/10.2760/944555

[19] P.J. Smith, F. Pappenberger, F. Wetterhall, J. Thielen del Pozo, B. Krzeminski, P Salamon, D Muraro, M. Kalas, and C. Baugh. 2016. Chapter 11 - On the Operational Implementation of the European Flood Awareness System (EFAS). In *Flood Forecasting*, Thomas E. Adams and Thomas C. Pagano (Eds.). Academic Press, Boston, 313–348. https://doi.org/10.1016/B978-0-12-801884-2.00011-6

[20] S. Tarquini, I. Isola, M. Favalli, M. Mazzarini, F. Bisson, MT. Pareschi, and E. Boschi. 2007. TINITALY/01: a new Triangular Irregular Network of Italy, 2007. *Ann. Geophys.* 50, 3 (2007). https://doi.org/10.4401/ag-4424

[21] Michalis I. Vousdoukas, Lorenzo Mentaschi, Evangelos Voukouvalas, Alessandra Bianchi, Francesco Dottori, and Luc Feyen. 2018. Climatic and socioeconomic controls of future coastal flood risk in Europe. *Nature Climate Change* 8, 9 (01 Sep 2018), 776–780. https://doi.org/10.1038/s41558-018-0260-4

[22] Annett Wania, Inès Joubert-Boitat, Francesco Dottori, Milan Kalas, and Peter Salamon. 2021. Increasing Timeliness of Satellite-Based Flood Mapping Using Early Warning Systems in the Copernicus Emergency Management Service. *Remote Sensing* 13, 11 (2021). https://doi.org/10.3390/rs13112114

[23] Liu Xiaoyan, Yang Saini, Ye Tao, An Rui, and Chen Cuizhen. 2021. A new approach to estimating flood-affected populations by combining mobility patterns with multi-source data: A case study of Wuhan, China. *International Journal of Disaster Risk Reduction* 55 (02 2021). https://doi.org/10.1016/j.ijdrr.2021.102106

[24] L. Yue, L. Qinghua, and S. Jie. 2017. Discover Patterns and Mobility of Twitter Users—A Study of Four US College Cities. *ISPRS International Journal of Geo-Information* 6, 2 (2017). https://doi.org/10.3390/ijgi6020042

Received July 2021; revised November 2021; accepted April 2022