Determining flooded areas using crowd sensing data and weather radar precipitation: a case study in Brazil

Flávio E. A. Horita*
Center for Mathematics, Computation and Cognition,
Federal University of ABC
Santo André, Brazil
Research & Development Lab, Brazilian Meteorological Agency, CLIMATEMPO
São José dos Campos, Brazil
flavio.horita@ufabc.edu.br

Ricardo B. Vilela
Research & Development Lab, Brazilian Meteorological Agency, CLIMATEMPO
São José dos Campos, Brazil
ricardo.vilela@climatempo.com.br

Renata G. Martins
Research & Development Lab, Brazilian Meteorological Agency, CLIMATEMPO
São José dos Campos, Brazil
renata.martins@climatempo.com.br

Danielle A. Bressiani
Research & Development Lab, Brazilian Meteorological Agency, CLIMATEMPO
São José dos Campos, Brazil
danielle.bressiani@climatempo.com.br

Gilca Palma
Research & Development Lab, Brazilian Meteorological Agency, CLIMATEMPO
São José dos Campos, Brazil
gilca@climatempo.com.br

João Porto de Albuquerque
Centre for Interdisciplinary Methodologies,
University of Warwick
Coventry, United Kingdom
j.porto@warwick.ac.uk

ABSTRACT

Crowd sensing data (also known as crowdsourcing) are of great significance to support flood risk management. With the growing volume of available data in the past few years, researchers have used in situ sensor data to filter and prioritize volunteers' information. Nevertheless, stationary, in situ sensors are only capable of monitoring a limited region, and this could hamper proper decision-making. This study investigates the use of weather radar precipitation to support the processing of crowd sensing data with the goal of improving situation awareness in a disaster and early warnings (e.g., floods). Results from a case study carried out in the city of São Paulo, Brazil, demonstrate that weather radar data are able to validate flooded areas identified from clusters of crowd sensing data. In this manner, crowd sensing and weather radar data together can not only help engage citizens, but also generate high-quality data at finer spatial and temporal resolutions to improve the decision-making related to weather-related disaster events.

Keywords

Crowd sensing data, Weather radar precipitation, Kernel density estimation, Flood management, Collaborative platforms.

*corresponding author

INTRODUCTION

Catastrophic impacts caused by floods around the world have called for the adoption and preparation of measures, which could increase the resilience of vulnerable communities. This is even more important since the occurrence of disaster tends to grow in the next few years due to climate changes. In this manner, flood management requires a proper understanding of flooded areas that are important and fundamental not only to take response actions and conduct relief tasks, but also to trace vulnerable locations and thus training and preparing communities in case of an event. An important component of flood management is the implementation and improvement of flood forecasting and warning systems. Regional and national institutions have been deploying several data collection systems for supporting flood management; these include rainfall gauges, hydrological stations, humidity sensors, and weather data. Among them, a weather radar is a remote sensing-based equipment that measures the volume of rainfall analyzing the backscattered energy of a microwave pulse emitted to the atmosphere (Dixon and Wiener 1993). These are particularly valuable for monitoring the rainfall over an entire area instead of a single point like done by rainfall gauges.

In parallel, the power of crowds in estimating current situation at affected areas and in supporting data analysis for disaster management has been demonstrated in several research works (Horita, Degrossi, et al. 2013; Klonner et al. 2016; Granell and Ostermann 2016). These data comprise a collection of volunteered data produced by individuals and informal institutions, i.e., by ordinary citizens using appropriate tools to gather and disseminate their views and knowledge on the web (Goodchild 2007). For example, Wang et al. (2016) conducted a social media analysis from spatial, temporal, and content perspectives for wildfire hazards. In the particular case of floods, Kusumo et al. (2017) examined volunteered information for planning evacuation shelters in case of floods and Horita, Albuquerque, Degrossi, et al. (2015) presented a decision support system that integrates volunteered information and sensor data for flood management.

Although volunteered information has shown its potential to support flood risk management, it also increases the amount of available and useful data, which can lead to information being overlooked or even misused by decision-makers. Both practice and academia have been developing geospatial approaches to process such data; among them, impressive results were achieved when processing volunteered information using data provided by stationary sensors (e.g., hydrological stations, or rainfall gauges) (Albuquerque et al. 2016; Assis, Albuquerque, et al. 2016). At the same time that these approaches open further opportunities, they also raise new challenges, which the main are twofold: 1) if a sensor fails, or is broken, data processing is not reliable and sound; and 2) stationary sensors are able to monitor a limited region, in which it is located. Here, weather radar data show up as an interesting alternative, because they are able to provide measurements of environmental variables (e.g., rainfall) over an entire area instead of a single point through scans.

On the basis of these challenges, this ongoing work investigates weather radar systems as alternative data sources that could support crowd sensing data assessment and improve decision-making in disaster management (e.g., floods). For doing so, data were collected in the period from December 2017 to January 2018 with weather radar systems and a crowd sensing platform, which in turn were employed in a case study conducted in the city of São Paulo, Brazil for identifying flooded areas. This is a relevant case study mainly because the city has high population density (i.e., 12 million inhabitants, or 20.4 million for the whole metropolitan region) and a recurrent problem with flooding, which have incurred financial losses close to US$ 80 million in 2008 (i.e., 42% of Brazilian losses with disasters). Moreover, in contrast to existing works focused on analyzing the use of crowdsourcing data to determine or monitor weather conditions (Demirbas et al. 2010; Chatfield and Brajawidagda 2014; Muller et al. 2015), this work is the first study in the literature centered on utilizing rainfall data provided by weather radar systems to assess crowd sensing data.

This work is structured as follows. Section "Background" introduces the theoretical background of this study. Section "Research Method" outlines the used research method, while Section "Case Study" details the case study conducted for empirically analyzing the method. Section "Results" presents the preliminary results. Eventually, Section "Discussions and Conclusions" summarizes the findings and also describes future lines of work.

BACKGROUND

Crowdsourcing for flood management

The emergence of Web 2.0 and evolution of mobile devices have become the basis for the emergence of a new paradigm, where users in general (i.e., citizens) become established as producers of data and information (Niko et al. 2011). Interestingly, all these generated data and information in many cases, are more detailed and of a higher quality than those provided by official organizations (Goodchild 2007). In this context, Heipke (2010) proposed the

term "crowdsourcing" for this phenomenon, which involves content production being carried out by a third party that is assigned to intelligence and knowledge. It is based on the experience of volunteers, who are independent in the way they use their free time and are located in remote and diverse areas.

In a similar context, but more closely linked to geographical issues, Goodchild (2007) coined the term Volunteered Geographic Information (VGI) to name this phenomenon which was defined as a collection of digital spatial data produced by individuals and non-formal institutions, i.e., by ordinary citizens using appropriate tools to gather and disseminate their views and geographical knowledge on the web. As a result, these volunteered data have a high potential to expand and qualify the amount of information available about the events and experiences of the community members to perform their activities (Coleman et al. 2009). This volunteered information is classified into the following types (Albuquerque et al. 2016): 1) Social media, i.e., information shared using social media platforms (e.g., Twitter); 2) Collaborative mapping, i.e., spatial information generated by volunteers using mapping platforms (e.g., OpenStreetMap); and 3) Crowd sensing, e.g., applications and platforms (e.g., Ushahidi) focused on gathering pre-determined information about a phenomenon (e.g., water level of river).

Different from social media platforms in which information is shared for general purposes and its meaning is only obtained through machine learning techniques, collaborative mapping and crowd sensing platforms demand an active involvement of users to provide structured information. Since these new platforms provide a great volume of useful data, they should be pre-processed before reaching decision-makers. Several filtering and data processing approaches thus have been developed to overcome this issue; for example, analyzing the geographic relations between official stationary sensors and volunteered information (Albuquerque et al. 2016; Assis, Albuquerque, et al. 2016). However, these approaches often are spatially limited to the location of sensors, i.e., a rainfall gauge provides the precipitation information for its location, it is not able to provide data from a rain occurring 5 km far from it. In this manner, weather radar systems are a relevant and important alternative that can be employed to improve situation awareness, which may also improve the quality and use of information generated by volunteers.

Weather radar system for flood management

The weather radar (RAdio Detection And Ranging) is a remote sensing based equipment that measures the rainfall backscattered energy of an emitted microwave pulse to the atmosphere. Different from rain gauges, the weather radar is able to estimates rain over an entire area through scans instead of a single point. The energy measured by the radar is called radar reflectivity factor (Z). To obtain the quantitative rainfall rate (R) in mmh^{-1} a Z-R relationship is applied. Thereby, it is possible to make a real-time rainfall evaluation over urban areas, and provide support for flood warning and management.

Since it is a great source of information for disaster (and flood) management, these weather radar systems have been employed in both practice and research. Within these works, Borga et al. (2014) used rainfall maps using weather radar and rain gauge data for reports of debris flows and flash floods and showed that rain-gauges networks are generally not dense enough to monitor precipitating systems with high temporal and spatial variability. Furthermore, Yang et al. (2004) utilized weather radar measurements to apply a coupled distributed hydrological model for flood forecasting and control. Another application of weather radar for flood monitoring can be seen on Flash Project (Gourley et al. 2017) that uses weather radar data to estimate water balance to improve the techniques of flash flood monitoring and prediction. Likewise, Ehret et al. (2008) described an approach that combines rainfall observations provided by weather radar systems and rainfall gauges for flood forecasting.

There is also an emerging line of works that are focused on combing the crowdsourcing data with remote sensing data for supporting flood management; for instance, Rosser et al. (2017) presented a method that combines data from different data sources (i.e., remote sensing, social media, and topographic maps) with the aim of estimating flood inundation extent. However, existing studies are still very few and further research works are required in order to provide a better understanding of how remote sensing could be employed with crowdsourcing data.

RESEARCH METHOD

Overview

The objective of this paper is to investigate the use of weather radar systems as an alternative source of information to support crowd sensing data processing. This is particularly important because current and existing approaches have used stationary sensor data, which are able to provide data only from a limited region. In contrast, weather radar systems are capable to provide rainfall information over an entire area. Therefore, this work addresses the following research question: *how can weather radar data validate flooded areas identified by crowd sensing data?*

In order to address the proposed question, we propose a method that comprises three components: 1) weather radar data processing, 2) crowd sensing data processing, and 3) data analysis. Figure 1 depicts the method with its components.

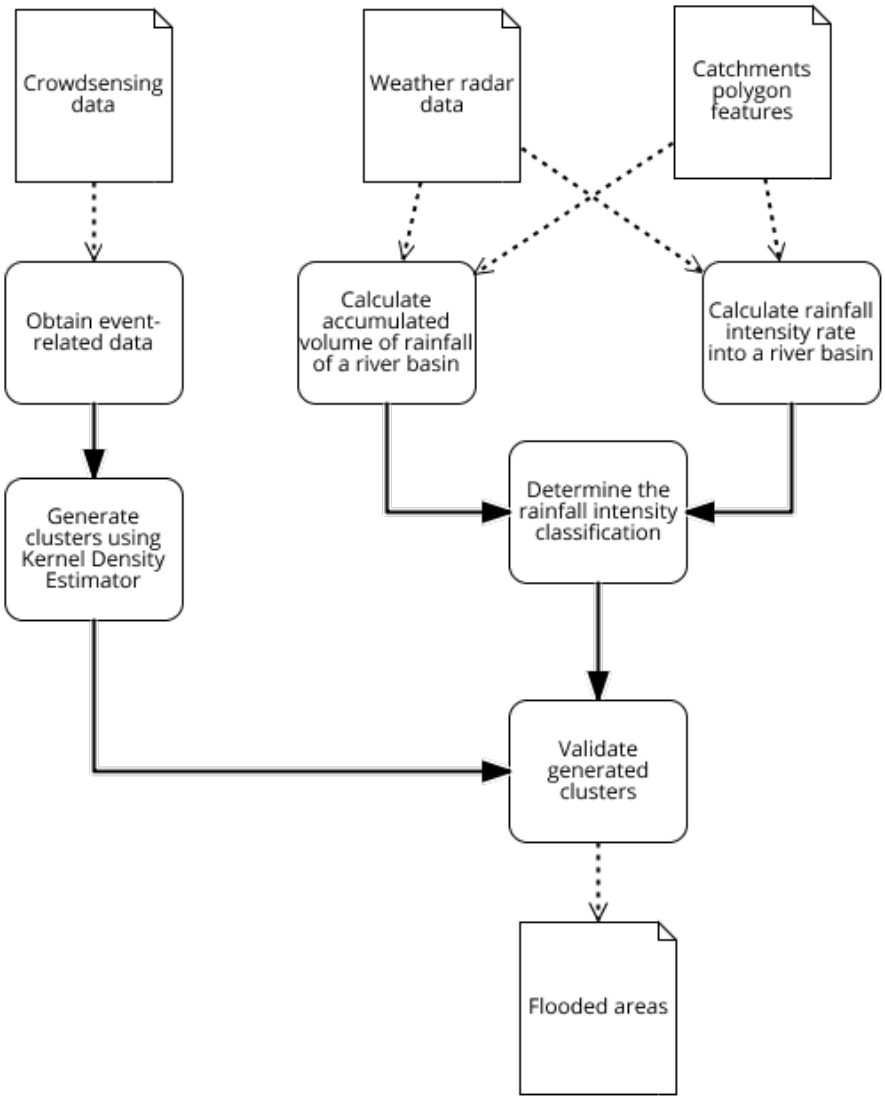

Figure 1. Research method

Weather radar data processing

The weather radar data processing gathers the rainfall data from the weather radar system to calculate the rainfall intensity rate and the accumulated volume of rainfall in a river basin. The rainfall was considered over the area of the river basins, since it is the area in which the precipitation collects and drains into one outlet. Therefore a flood event can happen in a region where there was no rainfall, but is downstream of an area where there was intense precipitation. The observation field of weather radar systems used on this study has a spatial resolution of 100 meters with a time step of 5-minutes per measurements, while the rainfall estimation is calculated using Calheiros and Zawadzki (1987)'s method. This method is based on a probability adjustment between the reflectivity factor and rainfall gauges.

With the rainfall estimation, the intensity and accumulated volume of precipitation within each river basin are calculated in a 5-minutes time span. The volume of precipitation within the river basin was calculated by the spatial mean volume of rainfall inside the basin, considering the largest value, between the last hour and the antecedent hour prior to a flood report (0 to 60 minutes, and 60.01 to 120 minutes prior to an event). Two hours were considered

since the concentration time of the river basins were close to two hours. The concentration time of a river basin can be defined as the time needed for the water to flow from the most remote point in a river basin to its outlet.

The rainfall intensity was calculated using the maximum instantaneous intensity of the mean spatial value for the river basin in the last hour. Therefore, based on the determined volume and intensity of the precipitation within the river basin, a classification of severity of rainfall was developed: This classification comprises four main classes: "0" indicates that there was no rainfall over the location; "1" indicates a weak precipitation with volume that ranges from 1 to 5 mm and intensity of 1 to 10 mmh^{-1}; "2" represents a moderate rainfall with volume from 5 to 10 mm and intensity of 10 to 20 mmh^{-1}; and lastly "3" indicates heavy precipitation with volumes and intensity higher than 10 mm and 20 mmh^{-1}, respectively. Similar categorization is suggested by Stull(2000)

Crowd sensing data processing

In parallel to weather radar data processing, the crowd sensing data processing aims at collecting and processing data provided by a collaborative platform. Having gathered raw data, it focuses on extracting only those data that are related to the disaster event of interest (e.g., floods). Since collaborative platforms, in particular, those focused on crowd sensing data are able to provide more structured data, disaster events often turn to be categories at the platform; therefore, the filtering process selects the data associated only with the category of interest (floods).

Once the event-related data is obtained, a cluster analysis is conducted to identify spatial patterns of the crowd sensing information and thus to create a map with flooded areas. The cluster analysis was based on Kernel Density Estimation (KDE) that "is a non-parametric method using local information defined by windows (also called kernels) to estimate densities of specified features at given locations" (Shi 2010). It is an important and widespread method for mapping spatial patterns of point events.

KDE requires one input for execution, the bandwidth, i.e., the distance threshold between elements to create a kernel. Two other parameters were supplemented to generate the clusters: a) number of cluster elements, i.e., the minimum number of data elements that can be combined to create a cluster and b) the lag time, i.e., the interval of time between data elements and the cluster creation. This was primordial because all the parameters are important and aim to maintain the quality of data; for example, identifying a real-time flooded area using a group of elements from the last 30 minutes may provide more accurate results than another one using data from the last 240 minutes. Based on this, the configuration settings of KDE have been selected based on testing experiments that examined bandwidths of: 100, 150 and 200 meters; the number of elements needed to create a cluster: 3, 4, and 5; and lag time of 15, 30, 45, and 60 minutes. Due to the limit of words however the description of these such experiments are beyond the scope of this work. The best configuration setting resulting from the comparison tests performed was with: a bandwidth of 200 meters; over three elements to create a cluster; and, a lag time of 30 minutes.

Data analysis

When a cluster is identified, it is validated through a geospatial data analysis with the precipitation severity classification. The steps of analysis are manifold. It first determines the geometric center (centroid) of the cluster in terms of its longitude and latitude. Then, these coordinates are used to identify the precipitation severity classification that was determined by the weather radar data processing. As mentioned before, a class can assume the following categories: 0 (no rain), 1 (weak rain), 2 (moderate rain), and 3 (heavy rain).

Classified clusters are then filtered and only those clusters at rainy areas (i.e., those classified in categories 1, 2, and 3) measured by the weather radar system remain. As a result, these represent the validated flooded areas identified by the volunteered information.

CASE STUDY

Study area

São Paulo city has more than 12 million inhabitants, an area of 1.5 million km², and a population density of 7,400 inhabitants per km^2. Located in Southeast Brazil (Figure 2), it is the country´s largest city (with 5.9% of its population) and the core of São Paulo Metropolitan Region (SPMR). Which, in turn, is the fifth largest megacity in the world, with around 21.3 million people in 2016 (United Nations 2016). SPMR is Brazil's main economic center, with important industrial, commercial and financial complexes. SPMR contributed with 17.63% of Brazilian's GDP and 54.48% of São Paulo's State GDP, in 2015 (Empresa Paulista de Planejamento Metropolitano (Emplasa) 2018).

Brazil passed through a fast urbanization process, going from an urban population of 36% in 1950 to 86% in 2015. The same occurred with the SPMR, which went from less than 2 to more 20 million inhabitants, from 1940 to

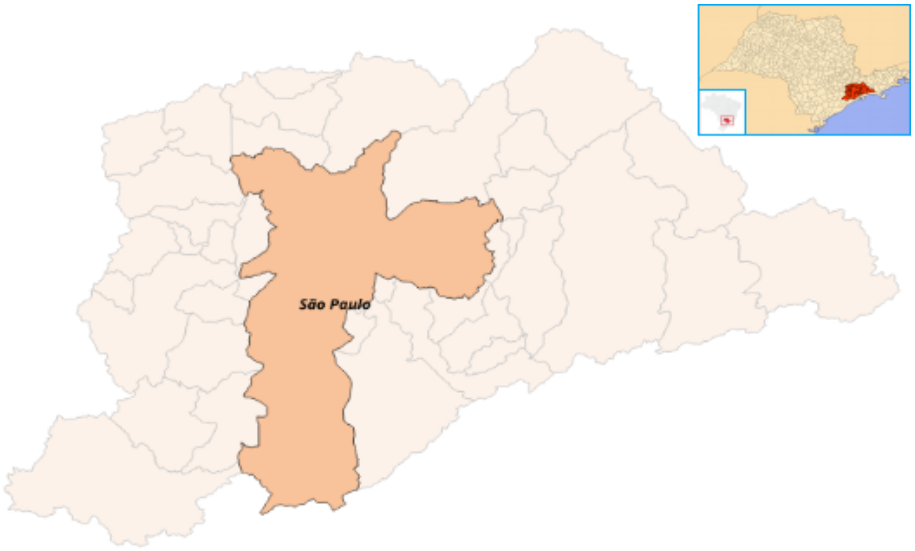

Figure 2. Study area: São Paulo city, Metropolitan Region and State in Brazil

2017 (Silva Dias et al. 2013). The fast urbanization was not accompanied by adequate planning and infrastructure implementation, establishing several vulnerabilities.

São Paulo is a region with heavy rainfall episodes. Large amounts of rainfall occur, due to the South Atlantic convergence zone (SACZ), in the rainfall season (from October until March), as well as heavy rainfall episodes that occur during the dry season, due to cold fronts from southern Chile and Argentina (M. S. Teixeira and Satyamurty 2011). There has been a significant positive trend on daily rainfall extremes in São Paulo city from 1933 to 2010, probably associated to the effects of climate variability and change, as well as urban growth effects, as from heat island and pollution (Silva Dias et al. 2013).

Therefore, these severe weather events associated with environmental and urban vulnerabilities cause frequent losses and disruptions due to flooding. Floods affect citizens, public and private services and activities. Floods in São Paulo city in 2008 alone reflected in losses of up to R$248.55 million to the city, and up to R$564.17 million to the country (Haddad and E. Teixeira 2015). Thus São Paulo city was defined as the study area for its complexity, population size (crowdsource data), and its flooding issues and impacts.

Data

In the period from December 2017 to January 2018, heavy rain damaged several regions of the city of São Paulo in Brazil and most of them have led to impacts to the community routines. Precipitation volumes of these months were significant (i.e., 141mm in December and 161mm in January), although they did no reach high historical records (i.e., 193mm in December and 218mm in January, from 1995). In December 2017, the highest amounts of precipitation were obtained in "Penha" gauge station with 200.9mm (this overwhelmed the historical volume for the month), “Vila Mariana” gauge with 190.0mm, and "Itaquera" gauge with 181.1mm. While, in January, the gauge in the city center (“ Praça da Sé ” gauge) and again ‘Itaquera ” gauge reached the highest amounts of rainfall volume with 185.6mm and 166.4mm, respectively.

Furthermore, 66 flooded locations were issued by the São Paulo Emergency Management Center (CGE, in Portuguese)[1]. Since 1999 when it was created, this center monitors meteorological conditions and issues weather forecasts and warnings of flood events in the city. A team conducts the monitoring task in continuous shifts of 24/7 using satellite imagery, forecasting models, and hydro-meteorological data from stationary stations. This team comprises operators from interdisciplinary backgrounds, i.e., meteorologists, engineers, technicians, and journalists. Members of the County Traffic Engineering Company (CET, in Portuguese), Civil Defense, and Fire Department support activities of CGE by providing updated and reliable information about current occurrences of events in the city.

Rainfall data provided by two weather radar in the city of São Paulo were used in this work. These are X-band Weather Radar Systems that generate spatially distributed horizontal data, on the temporal evolution of precipitation

[1] http://www.cgesp.org

around a determined area, which can have a radius coverage of up to 60 km (in our case, 21.6 km) and with a time lag between scans of 5 minutes. Through analyzing data from the period of December 2017 to January 2018, two events of heavy rain were determined and later adopted as main case studies. Table 1 shows the regions where the highest values of the accumulated amount of precipitation were observed by the weather radar systems within the indicated day (i.e., January 16, 2018 and January 21, 2018).

Table 1. Accumulated volume of rainfall in São Paulo, Brazil.

Date	Region	Accumulated volume of rainfall (mm)
2018/01/16	Campo Limpo	21,5
2018/01/21	São Miguel	63,4

Together, these two events reached values close to 19.47% of the historical maximum precipitation of January (i.e., 218mm), which also provided concrete evidence of the passage of intense meteorological systems over the city of São Paulo. Therefore, we gathered rainfall data provided by the weather radars system in a 5-minutes resolution for the studied period; these data are generated as a NetCDF file[2] that associates metadata to spatial positions.

The crowd sensing data used in this study were shared by citizens utilizing a mobile app. When creating a report through the app, a user first indicates the type associated with the report (e.g., hazard), then its subtype (e.g., flood or fog). Geospatial location of the report is automatically gathered using the GPS of the smartphone. It is also possible to include a photo or a comment. In the period of study, 735 geo-referenced reports in the area of São Paulo were obtained from this platform. These reports consist of a set of data: type, subtype, date in milliseconds, a unique ID, information about the user, and the coordinates of the report. Metadata is shared in a JSON format. Table 2 outlines the crowd sensing dataset separated among the existing subtypes

Table 2. Crowd sensing dataset per subtype.

	Subtype			
Date	Flood	Weather	Fog	Hail
2018/01/16	230 (46.09%)	6 (8.45%)	14 (15.55%)	21 (28.00%)
2018/01/21	269 (53.90%)	65 (91.54%)	76 (84.44%)	54 (72.00%)
	499	71	90	75

The majority of data is related to reports that indicate flooded areas - 499 reports (close to 67.89%). While, reports of Fog represent the second subtype with - 90 reports (12.24% of all data), which is followed by remaining subtypes: Hail (10.20%) and Weather (9.67%).

RESULTS

A high concentration of crowd sensing data could be identified at the city center and southwest region of São Paulo on January 16^{th}, 2018, while the condition on January 21^{st}, 2018, was restricted only at the central region of the city. Figures 3 and 4 display the condition on these two cases studied; with crowd sensing data as green points, weather radar systems locations as red triangles, the area covered by these systems as circles, and the geographical area of São Paulo delimited.

After employing the research method in the case study, it was identified a total of 97 clusters associated with flooded areas. Interestingly, 4.12% of these clusters were attributed to the "no rain" severity class, while the others were assigned to "heavy rain" class and "moderate rain", 54.64% and 41.24%, respectively. Table 3 shows the distribution of the clusters according to the rain severity classification.

The clusters attributed to the rain severity classification "0" were removed and only the remaining ones were adopted as validated clusters of rainfall response. Since the aim of this paper was to investigate the use of weather radars system precipitation to support the processing of crowd sensing data and to improve information quality and situation awareness in a disaster. We compared the location of generated clusters with historical hot spots of flooded areas reported by CGE; the main idea here is that cluster polygons are expected to be located in known areas of

[2]NetCDF (Network Common Data Form) is a set of software libraries and self-describing, machine-independent data formats that support the creation, access, and sharing of array-oriented scientific data.

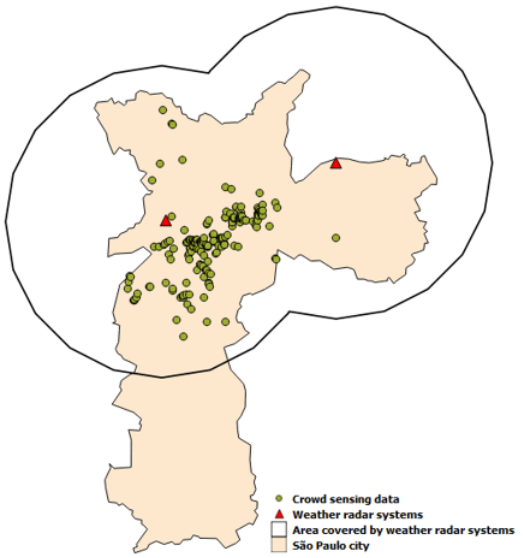

Figure 3. Crowd sensing data on January 16th, 2018.

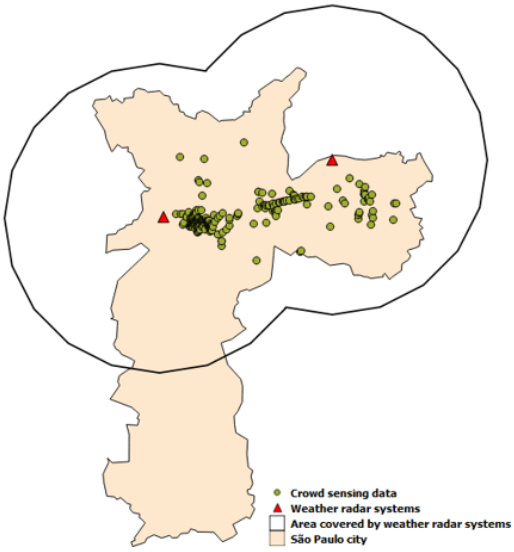

Figure 4. Crowd sensing data on January 21st, 2018.

flooding (i.e., historical data of flood events). This historical data covered a period from January 2007 to December 2017, which consists of 9,667 reports of flooded areas in the city of São Paulo, i.e., almost 1,000 of events per year. Areas with frequent events can be verified at the city center, close to the Municipal Market; further areas can also be verified alongside Tietê Expressway and Pinheiros Expressway. Other verified locations with a high concentration of reports are close to Ibirapuera Park, Mario Pimenta Park, and São Paulo Golf Club. These areas concentrated over 60% of reports published by CGE on its website. Figure 5 displays the validated clusters as red polygons, weather radars locations as red triangles, and historical data of reported flooded areas by CGE as a heat map.

An analysis of the results presented in Figure 5 show that most of the validated cluster polygons - i.e., flooded areas - were located at regions with a high concentration of historical flooded areas, from CGE in both case studies. Furthermore, it is worthwhile to mention that none of the clusters were located in areas with no previous reports from CGE. This provides evidence that the research method is able to identify valuable flooded areas from volunteered data and weather radars system data.

DISCUSSIONS AND CONCLUSIONS

This ongoing work presented a research method that could be employed to support crowd sensing data processing and thus improve decision-making in disaster management (e.g., by identifying flooded areas). Through analyzing information provided by a crowd sensing platform, the method first generates spatial clusters, which are later analyzed and validated using weather radars system data. From December 2017 to January 2018, a case study has been carried out in the city of São Paulo, Brazil using rainfall data provided by two weather radars located in the city, as well as information shared by volunteers utilizing a crowd sensing platform. Study findings indeed indicated that rainfall data provided by weather radars system are able to validate flooded areas, which were identified by volunteered information. Furthermore, the research method utilized a clustering approach to determine those flooded areas that thus may support more informative decisions in flood management. To the best of our knowledge, this is the first research work that uses weather radar data to process crowdsourcing data.

Results obtained in this research provided preliminary evidence, which shows that rainfall data provided by weather radars system are of great value for validating volunteered information. The results were promising in exploring

Table 3. Distribution of clusters according to the rain severity classification.

Time (min)	Rain Intensity Class				Total
	0	1	2	3	
30	4 (4.12%)	0 (0%)	40 (41.24%)	53 (54.64%)	97

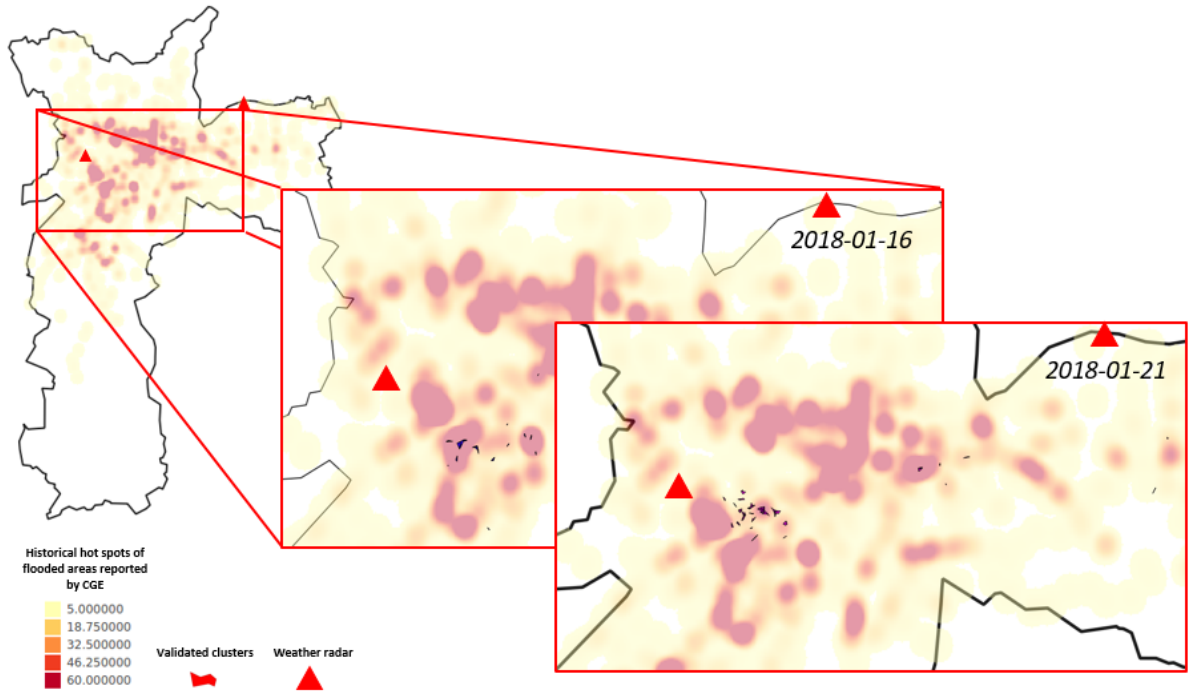

Figure 5. Validated clusters of flooded areas. Validated clusters are displayed as red polygons, weather radars locations as red triangles, and historical flooded areas reported by CGE as heat map.

the potential of the crowd sensing data, by validating and ensuring its quality based on precipitation data from weather radars in the drainage basin. The results showed compatible with historical records from CGE. Public organizations that monitor flooding areas usually have limited capability of reach and resources to be in all places needed, especially when a crisis situation strike. As for example, CGE monitors the entire São Paulo city (i.e. with limited personnel or hydrological gauge stations to cover a large area). Therefore using population input from crowdsourcing data associated with a spatial precipitation information (from weather radars) to ensure data quality can greatly enlarge the reach and coverage of real-time information. These preliminary results as the future works here proposed, can provide significant gains in disaster management, flood and situation awareness, forecasting and empowerment of citizens in relation to disaster (flooding) risks.

Hence, this can be employed as an additional step in existing research works that are focused on data processing; for example, in the approach proposed by Assis, Behnck, et al. (2016), which employed data provided by rainfall gauge stations to prioritize on-topic social media messages. In order to improve the rigorous of spatial analysis (Mohaymany et al. 2013), there is also a need to employ other clustering approaches like Moran's I or DBSCAN, or even understand in which cases those approaches are more suitable (García-Palomares et al. 2015).

Although it is possible to move towards the effective identification of flood clusters, an additional stage in the analysis should be also considered. This includes incorporating environmental characteristics, such as the physical parameters related to flood vulnerability as in research conducted by Rimba et al. (2017). In this sense, an overlap considering the physical processes involved with flooding, as well as the precipitation from the weather radars system, can provide a greater understanding to better identify and map, in an area, the likelihood of flooding. A better understanding of decision-making requirements should be also covered when filtering and analyzing the usefulness of these flooded clusters (Horita, Albuquerque, and Marchezini 2018). In this manner, the framework proposed by Horita, Albuquerque, Marchezini, and Mendiondo (2017) could be combined with the method introduced in this work and thus describe how decision-making tasks might be impacted by processed crowd sensing data.

More case studies should be conducted to improve the generalization of the overall work, as well as to gather new insights for improving the research method. This includes performing new studies adopting different empirical settings; for example, using weather radar data for validating volunteered information in the situation of landslides. Furthermore, KDE settings could diverse in three perspectives: spatial, temporal, and minimal number of elements. Different combinations of these perspectives should be also investigated to obtain more accurate clustering results; for example, by determining a smaller area (150m), a longer period (120min), and a greater number of minimal elements (five elements). Furthermore, other collaborative platforms (e.g., Ushahidi) and social media platforms (e.g., Twitter, Flickr, and Instagram) should be investigated.

Finally, future work lines should also employ this volunteered data as alternative data for flood modelling, forecasting and warning through precipitation-streamflow artificial intelligence (neural networks, etc) or hydrological models, using radar rainfall and hydrological gauges as inputs; for example, by using the water balance models developed by Gourley et al. (2017). This approach thus has a great potential to improve monitoring because it considers the soil's characteristics and orography. Therefore, we believe a model that integrates precipitation data (from radars, satellites, gauge stations), together with the characteristics of the river basins (shape, time of concentration, land use, soil's characteristics, slope, hydraulic features, etc) and crowd sensing data can improve the way we monitor, forecast and warn flooding areas and flooding risks, providing better tools for citizens, as well as public and private entities to be aware, collaborate and prevent the impacts of flooding.

ACKNOWLEDGEMENTS

The authors are thankful for the financial support from FAPESP and FINEP (Grant no. 2016/10229-3). RGM is also grateful for the financial and institucional support from FAPESP (Grant no. 2017/17007-9). FEAH thanks the financial support from FAPESP (Grant no. 2017/09889-1). The authors would also like to thank Climatempo, the University of São Paulo (USP) and the Meteorological Remote Sensing of Storms (STORM-T) and Prof. Dr. Carlos Augusto Morales Rodriguez for the radar data.

REFERENCES

Albuquerque, J. P., Eckle, M., Herfort, B., and Zipf, A. (2016). "European Handbook of Crowdsourced Geographic Information". In: ed. by C. Capineri, M. Haklay, H. Huang, V. Antoniou, J. Kettunen, F. Ostermann, and R. Purves. Ubiquity Press. Chap. Crowdsourcing geographic information for disaster management and improving urban resilience: an overview of recent developments and lessons learned, pp. 309–321.

Assis, L. F. F. G., Behnck, L. P., Doering, D., Freitas, E. P., Pereira, C. E., Horita, F. E. A., Ueyama, J., and Albuquerque, J. P. (2016). "Dynamic Sensor Management: Extending Sensor Web for Near Real-Time Mobile Sensor Integration in Dynamic Scenarios". In: *Proceedings of the 30^t h International Conference on Advanced Information Networking and Applications (AINA)*, pp. 303–310.

Assis, L. F. F. G., Albuquerque, J. P., Herfort, B., Steiger, E., and Horita, F. E. A. (2016). "Geographical prioritization of social network messages in near real-time using sensor data streams: an application to floods". In: *Brazilian Journal of Cartography* 68.6, pp. 1231–1240.

Borga, M., Stoffel, M., Marchi, L., Marra, F., and Jakob, M. (2014). "Hydrogeomorphic response to extreme rainfall in headwater systems: flash floods and debris flows". In: *Journal of Hydrology* 518, pp. 194–205.

Calheiros, R. and Zawadzki, I. (1987). "Reflectivity-rain rate relationships for radar hydrology in Brazil". In: *Journal of climate and applied meteorology* 26.1, pp. 118–132.

Chatfield, A. T. and Brajawidagda, U. (2014). "Crowdsourcing Hazardous Weather Reports from Citizens via Twittersphere under the Short Warning Lead Times of EF5 Intensity Tornado Conditions". In: *Proceedings of the 47th Hawaii International Conference on System Sciences (HICSS)*, pp. 2231–2241.

Coleman, D. J., Georgiadou, Y., Labonte, J., Observation, E., and Canada, N. R. (2009). "Volunteered Geographic Information : the nature and motivation of producers". In: *International Journal of Spatial Data Infrastructures*.

Demirbas, M., Bayir, M. A., Akcora, C. G., Yilmaz, Y. S., and Ferhatosmanoglu, H. (2010). "Crowd-sourced sensing and collaboration using twitter". In: *Proceedings of the 2010 IEEE International Symposium on "A World of Wireless, Mobile and Multimedia Networks" (WoWMoM)*, pp. 1–9.

Dixon, M. and Wiener, G. (1993). "TITAN: Thunderstorm identification, tracking, analysis, and nowcasting—A radar-based methodology". In: *Journal of atmospheric and oceanic technology* 10.6, pp. 785–797.

Ehret, U., Götzinger, J., Bárdossy, A., and Pegram, G. G. (2008). "Radar-based flood forecasting in small catchments, exemplified by the Goldersbach catchment, Germany". In: *International Journal of River Basin Management* 6.4, pp. 323–329.

Empresa Paulista de Planejamento Metropolitano (Emplasa) (2018). *Sobre a RMSP*. Technical Report, GIP/CDI, 2018.

García-Palomares, J. C., Gutiérrez, J., and Mínguez, C. (2015). "Identification of tourist hot spots based on social networks: A comparative analysis of European metropolises using photo-sharing services and GIS". In: *Applied Geography* 63, pp. 408–417.

Goodchild, M. F. (2007). "Citizens as sensors: the world of volunteered geography". In: *GeoJournal* 69.4, pp. 211–221.

Gourley, J. J., Flamig, Z. L., Vergara, H., Kirstetter, P., Clark III, R. A., Argyle, E., Arthur, A., Martinaitis, S., Terti, G., Erlingis, J. M., et al. (2017). "The FLASH Project: improving the tools for flash flood monitoring and prediction across the united states". In: *Bulletin of the American Meteorological Society* 98.2, pp. 361–372.

Granell, C. and Ostermann, F. O. (2016). "Beyond data collection: Objectives and methods of research using VGI and geo-social media for disaster management". In: *Computers, Environment and Urban Systems* 59, pp. 231–243.

Haddad, E. A. and Teixeira, E. (2015). "Economic impacts of natural disasters in megacities: The case of floods in São Paulo, Brazil". In: *Habitat International* 45, pp. 106–113.

Heipke, C. (2010). "Crowdsourcing geospatial data". In: *ISPRS Journal of Photogrammetry and Remote Sensing* 65.6, pp. 550–557.

Horita, F. E. A., Albuquerque, J. P., Marchezini, V., and Mendiondo, E. M. (2017). "Bridging the gap between decision-making and emerging big data sources: An application of a model-based framework to disaster management in Brazil". In: *Decision Support Systems* 97, pp. 12–22.

Horita, F. E. A., Albuquerque, J. P., and Marchezini, V. (2018). "Understanding the decision-making process in disaster risk monitoring and early-warning: A case study within a control room in Brazil". In: *International Journal of Disaster Risk Reduction* 28, pp. 22–31.

Horita, F. E. A., Albuquerque, J. P., Degrossi, L. C., Mendiondo, E. M., and Ueyama, J. (2015). "Development of a spatial decision support system for flood risk management in Brazil that combines volunteered geographic information with wireless sensor networks". In: *Computers & Geosciences* 80, pp. 84–94.

Horita, F. E. A., Degrossi, L. C., Assis, L. F. G., Zipf, A., and Albuquerque, J. P. (2013). "The use of Volunteered Geographic Information (VGI) and Crowdsourcing in Disaster Management: a Systematic Literature Review". In: *Proceedings of the 19th Americas Conference on Information Systems (AMCIS)*. 1-10.

Klonner, C., Marx, S., Usón, T., Albuquerque, J. P., and Höfle, B. (2016). "Volunteered geographic information in natural hazard analysis: a systematic literature review of current approaches with a focus on preparedness and mitigation". In: *ISPRS International Journal of Geo-Information* 5.7, p. 103.

Kusumo, A. N. L., Reckien, D., and Verplanke, J. (2017). "Utilising volunteered geographic information to assess resident's flood evacuation shelters. Case study: Jakarta". In: *Applied Geography* 88, pp. 174–185.

Mohaymany, A. S., Shahri, M., and Mirbagheri, B. (2013). "GIS-based method for detecting high-crash-risk road segments using network kernel density estimation". In: *Geo-spatial Information Science* 16.2, pp. 113–119.

Muller, C., Chapman, L., Johnston, S., Kidd, C., Illingworth, S., Foody, G., Overeem, A., and Leigh, R. (2015). "Crowdsourcing for climate and atmospheric sciences: current status and future potential". In: *International Journal of Climatology* 35.11, pp. 3185–3203.

Niko, D. L., Hwang, H., Lee, Y., and Kim, C. (2011). "Integrating User-generated Content and Spatial Data into Web GIS for Disaster History". In: *Computers,Networks, Systems, and Industrial Engineering 2011*.

Rimba, A. B., Setiawati, M. D., Sambah, A. B., and Miura, F. (2017). "Physical Flood Vulnerability Mapping Applying Geospatial Techniques in Okazaki City, Aichi Prefecture, Japan". In: *Urban Science* 1.1, p. 7.

Rosser, J. F., Leibovici, D. G., and Jackson, M. J. (2017). "Rapid flood inundation mapping using social media, remote sensing and topographic data". In: *Natural Hazards* 87.1, pp. 103–120.

Shi, X. (2010). "Selection of bandwidth type and adjustment side in kernel density estimation over inhomogeneous backgrounds". In: *International Journal of Geographical Information Science* 24.5, pp. 643–660.

Silva Dias, M. A. F., Dias, J., Carvalho, L. M. V., Freitas, E. D., and Silva Dias, P. L. (2013). "Changes in extreme daily rainfall for São Paulo, Brazil". In: *Climatic Change* 116.3, pp. 705–722.

Stull, R. (2000). *Meteorology for scientists and engineers*. Brooks/Cole.

Teixeira, M. S. and Satyamurty, P. (2011). "Trends in the frequency of intense precipitation events in southern and southeastern Brazil during 1960–2004". In: *Journal of Climate* 24.7, pp. 1913–1921.

United Nations (2016). *The World's Cities in 2016*. Technical Report, Data Booklet (ST/ESA/ SER.A/392). Department of Economic and Social Affairs, Population Division.

Wang, Z., Ye, X., and Tsou, M.-H. (2016). "Spatial, temporal, and content analysis of Twitter for wildfire hazards". In: *Natural Hazards* 83.1, pp. 523–540.

Yang, D., Koike, T., and Tanizawa, H. (2004). "Application of a distributed hydrological model and weather radar observations for flood management in the upper Tone River of Japan". In: *Hydrological Processes* 18.16, pp. 3119–3132.