Dr. Arapostathis S.G., Spyrou N., Dr. Drakatos G., Prof. Kalabokidis K., Prof. Lekkas E., Dr. Xanthopoulos G.

Harokopio University, University of Athens, National Observatory of Athens, University of the Aegean, University of Athens, Institute of Mediterranean Forest Ecosystems

sarapos@hua.gr, nspyrou@geol.uoa.gr, g.drakat@noa.gr, kalabokidis@aegean.gr
elekkas@geol.uoa.gr, gxnrtc@fria.gr

Title: Mapping information related to floods, extracted from VGI sources, for effective disaster management within the Greek territory; the floods of West Attica (November 2017 Greece) case study.

Abstract:

While volunteered geographic information (VGI) has emerged as a highly significant aid in emergencies caused by natural disasters, such as floods, related research in Greece is very limited so far. This article aspires to cover this gap by introducing some initial methods for extracting and mapping related information from VGI sources, contributing thus, to the management of disaster events related to floods in Greece. As a case study, a recent flood that affected West Attica (Greece) and especially the city of Mandra, is used. The flood resulted in the death of 23 people and had also major negative consequences in the natural and urban environment of the region.

As a VGI source, Twitter has been chosen, for various reasons, including its micro-blogging nature, its popularity, at least within the area of the case study, and the strong geographical dimension that exists in its semantic structure.

As a first step, all the tweets that were published within 120 hours of the flooding event and contained relevant information, were collected. Next, they were classified to various categories, including: i) information regarding flood, ii) consequences of the flood iii) simple identification of the flood, and iv) simple rain or storm identification. The consequences were associated to certain values for more effective mapping. The final steps of our methodology included the geo-referencing of the classified information and the use of various visualization techniques for creating the final maps.

Keywords: VGI, Social Media, Floods, Disaster Management, Neo-Geography

1. Introduction

Disaster management is a very important scientific field through which human lives can be saved. A plethora of researchers from various fields are constantly trying to improve the response in crisis events that are caused from natural disasters. Various new technological tools equip the scientists and the authorities for applying more effective responses. A very important tool is the phenomenon of Volunteered Geographic Information (VGI) which is the act of having users as reporters (sensors) of geographical information through an Internet Service. The term was, initially, used by Goodchild (2007). According to literature, one of the most important characteristics of this phenomenon is the spatial heterogeneity of the produced data resulting in varying quality standards in each area. This characteristic must always be considered before applying various methodologies that are based on VGI sources. Other characteristics of VGI is the non-structural production of the information (Brando and Bucher 2010) along with the really high rhythm of production (Sui 2008).

While VGI has emerged during the latest years, as a highly significant aid in emergencies caused by natural disasters (Smith et al. 2017), related research in Greece (especially regarding floods) is very limited so far. This article aims to cover this gap by introducing some initial methods and analysis for extracting and mapping related information from VGI sources, contributing thus, on the crisis management of events related to floods in Greece. As a VGI source, Twitter has been chosen for various reasons, including its micro-blogging nature, its popularity at least within the area of the case study, and the strong geographical dimension that exists in its semantic structure. For all these reasons Twitter, is considered as a suitable VGI source for obtaining high volume of information published in different geographical regions and in short-time periods.

2. Published research

A lot of research regarding VGI and disaster management is published in the international literature. Indicatively, McDougall (2011) investigated the effectiveness of social media for identifying events that happened during the Queensland floods in Australia that affected about 30 cities, towns and rural communities over southern and western Queensland, while various inundations in the agricultural environment were also reported. The floods had a major economic impact of five billion dollars. Facebook and Twitter were used as sources and through them various related events were identified through photos and videos. The production volume was huge as about 14,000 to 16,000 tweets were published per hour. The most significant advantages according to this research included the immediacy and depth of information.

Moreover, Yin et al. (2012) recognized the contribution of twitter in terms of rapid and effective communication regarding disaster management. He claimed that by using this information, the relevant authorities can enhance their emergency-situation awareness, achieving better decisions in planning related operations and giving aid. They also presented a system for leveraging this kind of information for “increasing the perception level of situation awareness” (Yin et al. 2012). The basic principles of this system included data capturing, data processing (classification, clustering, geo-reference) and data visualization, as a research project supported by the official Australian authorities responsible for disaster management.

In addition, Dashti et al. (2013) emphasized the importance of correctly tracking the consequences of a natural disaster event, thus limiting the negative impact in similar future events. Under this framework they designed a system for capturing this kind of information through twitter while as a case study they used the Colorado floods in US that took place in 2013. As they mentioned, while various scientific teams moved to the area in which the floods occurred, important information was lost, as there was a time gap from the event occurrence to the moment that the scientific teams arrived at the place. This gap was covered from data extracted from social media. Specifically, they collected tweets within 9 days of the flood event occurrence by using certain keywords and classifying the tweets as: 1. geo-tagged tweets 2. tweets containing obvious URLs to photos and videos 3. tweets containing place names, 4. tweets containing structural terms as determined by the engineering team. In this way, they extracted significant information. About 1% of the total tweets were geo-located, a similar percentage to that found by other research (Morstatter et al. 2013). Within the findings of the analysis there was significant information regarding various consequences (such as bridge damage, broken sewage, flooded path etc.). Moreover, they combined various sources of information, such as tweets and satellite imagery, to end up with the most effective results. They concluded that along with the conventional methods of field verification, the extracted info could really contribute to the field of disaster reconnaissance.

The phenomenon of Volunteered Geographic Information is not strictly associated with data sources coming from social media. An interesting approach, recently presented by Bello et al. (2018), refers to the development of a community that has been able to reduce the risk of a flood event in Mexico. The community, after appropriate training, was able to perform improvements in the urban environment of a city, thus reducing the potential of having negative consequences after a heavy rain.

Finally, Smith et al. 2012 presented a hybrid approach through which flood-related data extracted from social media are validated with data gathered from a graphics processing unit (GPU) accelerated hydrodynamic modelling. They tested their approach by using as a case study the Tyne and Wear floods of June and August 2012 in the United Kingdom. More than 1,800 tweets were identified as related to the flood events.

3. Case study

In this case study, a recent flood that affected West Attica (Greece) and especially the city of Mandra, was used. The flood event was a sub event of a meteorological storm phenomenon named “Eurydice” that took place in Greece on November 15, 2017. The flood of West Attica resulted in the death of 23 people and had also major negative consequences in the natural and urban environment of the region.

4. Data used

The dataset acquired from the data provider Sifter (https://sifter.texifter.com/), consisted of about 74,000 published tweets containing at least one of the following keywords in Greek: floods, storm, damages, catastrophes, creek, rain, flood, inundation, Mandra. The city name Mandra was used also in Latin. The tweets were published from 15-11-2017 at 00:00:01 GMT until 20-11-2017 at 23:59:59 GMT. The cost for obtaining them was 210 US dollars.

5. Methodology used

In the dataset an initial analysis was performed ending up to the development of initial methodologies. The first step of the analysis consisted on performing all the necessary analytic processes for filtering the data. Specifically, according to the time stamp information of each tweet, all of the tweets that were published within 120 hours were selected. From those, all the tweets published in Greek, English and Spanish were also collected. As a next step, they were classified into certain groups of information related to i) the flood event of the city of Mandra, ii) tracking of the consequences, iii) identification of flood events in general, iv) identification of rain events in general, v) identification of other events not related to floods, vi) identification of reasons and vii) information related to crisis management. Moreover, a certain score was assigned in each tweet related to consequences. The score range was from I to V. The minimum score (I) was assigned to tweets that were describing simple rain or storm identification while the value V was associated exclusively to deaths. In Table 1, there is a brief description for each value and associated tweets.

Table 1: Description of values of consequence score

Consequence score	**Description**
I	Simple identification of rain or storm
II	Torrential storm, human fear, terror
III	Damages, problems in the traffic network, minor human injuries, flooded streets
IV	Huge damages, missing people, serious danger to human life, emergency situation
V	Loss of human life

The third step of the methodology was related to the geo-referencing part; x, y coordinates have been added to each tweet that contained geographic reference within the text. The publication positions (geo-located tweets) were not considered as the position in which a tweet was published, following Huiji and Barbier (2011) who found that these positions do not necessary reflect the location that is related to the information. In case of having more than one geographic references within a tweet, the tweet was duplicated and each one (the original and the duplicate) received the geographic coordinates of each mentioned area respectively. Moreover, each geographic reference received a certain precision score according to the size of each mentioned area. Cities at a size of a municipality received a score of I; areas of the size of a hamlet or a street of a certain place or a sub-area within a municipality or a certain geographic entity outside a municipality, received a score of II; finally, areas larger than a city (general area, prefecture etc.) were assigned a score of III.

Next, various validations to confirm the correct classifications were applied. The validations were mostly related to text queries that were combined through regular expressions. As a next step and after adding the x, y coordinates, the data were inserted as layers in a Geographic Information System (GIS) and various geo-processing techniques were applied. The techniques included automatic counting of published tweets per geographic area. Moreover, the tweets were spatially randomized within the polygon that defines the boundaries of the geographic area in which a tweet is associated to. This technique was applied in similar previous research of the authors (Arapostathis et al. 2016, 2018). The last step of the methodology included creation of various maps based on the classified content.

6. Results – Maps

As already stated, the dataset consisted of about 74,000 tweets (Figure 1). From those, 9,188 contained information related to floods or other events and were classified into the categories described earlier. As mentioned, many tweets had references of more than one geographic areas and were duplicated. Thus, the final dataset of classified information consists of 12,829 point-observations coming from tweets. 1,169 of them were related to simple rain event identification while 11,731 were describing various consequences of flood events. 6,112 of them were identifying flood events (Figure 1). Finally, 10,126 of them were in Greek, while 2,173 were in English and 370 in Spanish (Figure 2). About 160 tweets written in different languages were also recognized and classified.

Figure 1: Number of tweets per classification category

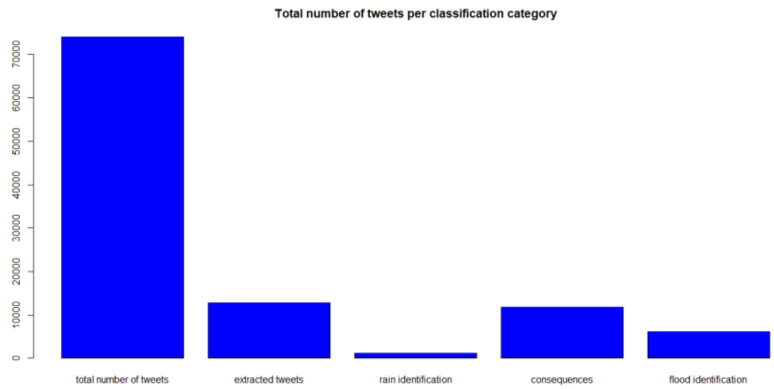
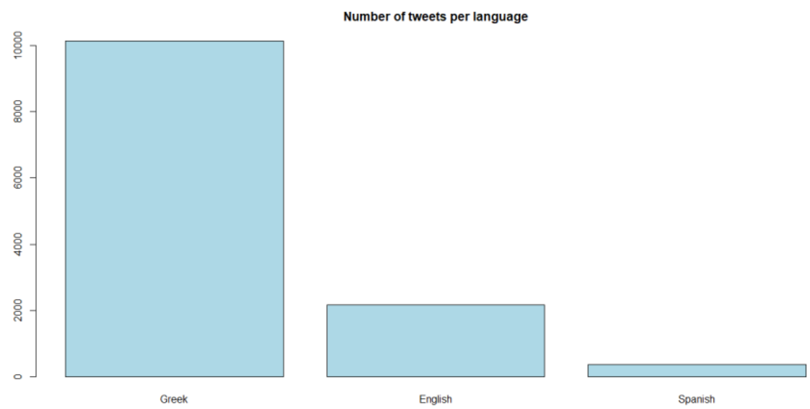


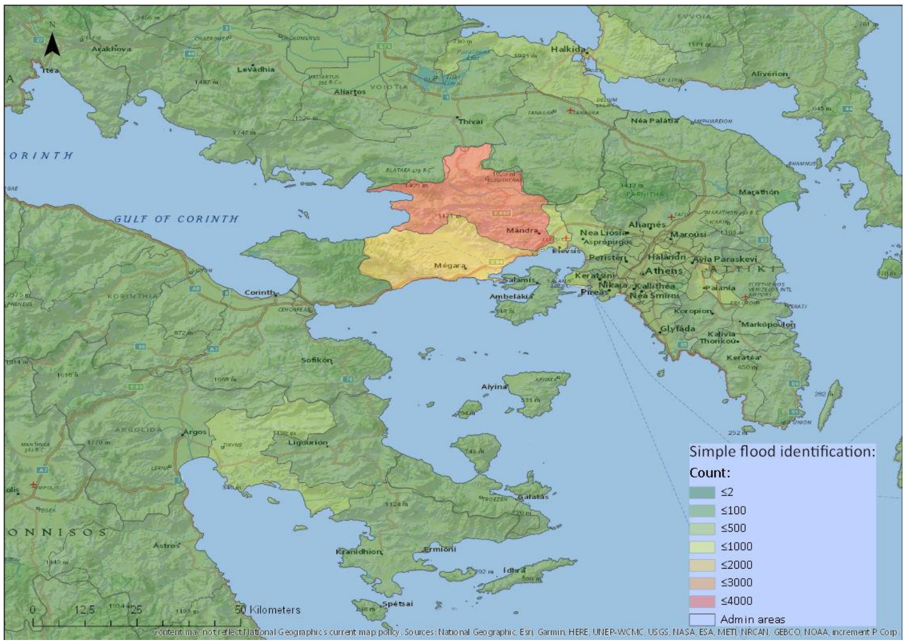
Figure 2: Total Number of tweets per language



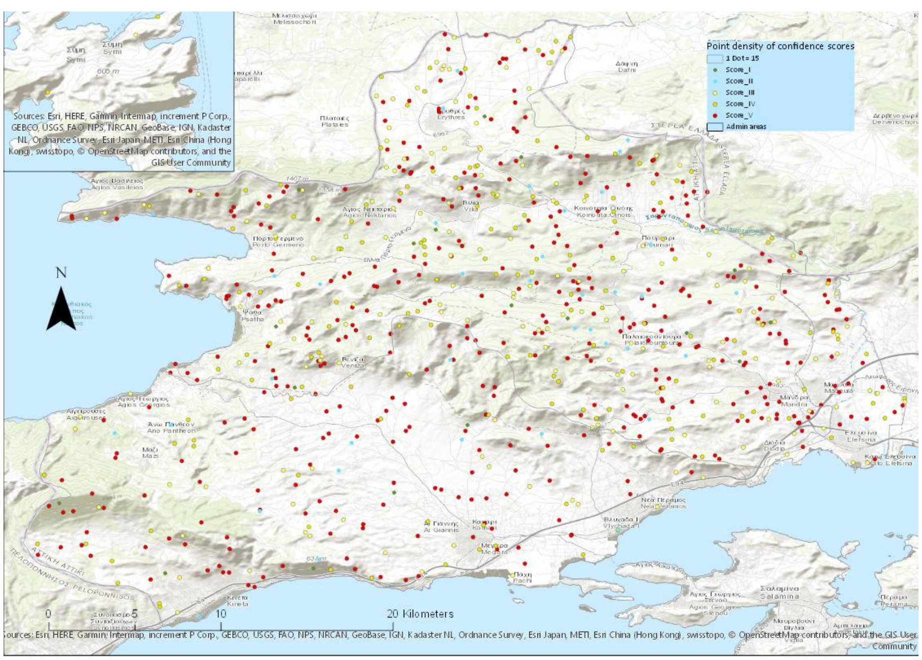
Map 1: Tweets related to rain identification within 120 hours

As it is observed in Map 1, the majority of the rain identification information was located in Attica and in North West Attica in particular. Few other rain events were identified in northwest Greece, and in islands located in the east and west parts of the country, while there were also few identifications in Peloponnese. Regarding floods, as it is displayed in Map 2, there were more than 3,800 tweets that contained information related to flood event identification. The Municipality of Megara, 16.5 km SW of Mandra, also had a high volume of references.

Map 2: Tweets related to flood event identification within 120 hours



Map 3: Mapping of consequences published within 120 hours



Various other flood events of smaller impact were identified in various parts of Attica, in Chalkida and in other areas located in the continental south part of Greece. Finally, Map 3 that is a point density map, displays the volume of information of consequences, classified to certain values. The red bullets represent the worst effects (fatalities). Each point represents 15 point observations (tweets).

7. Conclusions and future steps

Concluding, it is assessed that Social Media Sources can be really valuable in identifying, rain and flood events as well as, the consequences of those events in Greece. After this initial analysis, there is a remarkably high level of produced information regarding the consequences of an event. In 5 days more than 12,000 point-observations containing related information and precise geo-location were extracted through twitter. The contribution in disaster management is expected to be significant as a lot of information can be mapped in a way that can help the authorities to understand the information rapidly and perform more effective response actions. A further analysis is considered as necessary though to be able to calibrate and thus display more effectively the consequences of each event.

8. Acknowledgements

The tweet dataset was acquired from data provider Sifter. RStudio along with ArcGIS Pro were used for the creation of Maps and Figures. A layer of the administration structure of Greece at a municipality level was used, from geodata.gov.gr. Expenses were covered by the main author.

9. References

Arapostathis S.G., Lekkas E., Kalabokidis K., Xanthopoulos G., Drakatos G., Spyrou N., Kalogeras I. (2018) Developing seismic intensity maps from twitter data; The case study of Lesvos, Greece 2017 earthquake: Assessments, improvements and enrichments on the methodology. In Proceedings of the Gi4DM Congress, 18 – 21 of March, Istanbul

Arapostathis S, Parcharidis I, Stefanakis E, Drakatos G, Kalogeras I (2016) A method for creating seismic intensity maps from twitter data. Journal of Civil Engineering and Architecture, David Publishing 10:839-852

Bello G. M. C., Cohen M. A., Angelica N., Zaragoza M., Corona A. H. L., Perez J. R. (2018) Flood vulnerability reduction using a Partial Participatory GIS Approach. In Proceedings of the Gi4DM Congress, 18 – 21 of March, Istanbul

Brando C, Bucher B (2010) Quality in User Generated Spatial Content: A matter of specifications. In Proceedings of the 13th AGILE International Conference on Geographic information science, Giomaraes, Portugal.

Dashti S., Palen L., Heris M. P., Anderson K. M., Anderson S., Anderson T. J. (2013) Supporting disaster reconnaissance with social media data: a design-oriented case study of the 2013 Colorado Floods. In Proceedings of the 11th International ISCRAM Conference – University Park, Pennsylvania, USA

Goodchild MF (2007) Citizens as sensors: the world of volunteered geography. Geojournal 69(4):211-221.

McDougall K. (2011) Using Volunteered Information to Map the Queensland Floods In: Proceedings of the Surveying & Spatial Sciences Biennial Conference 2011. 21 – 25 of November 2011, Wellington, New Zealand

Sui DZ (2008) The wicification of GIS and its consequences: Or Angelina Jolie's new tattoo and the future of GIS. Computers, Environment and Urban Systems, 32:1-5.

Smith L., Liang Q., James P., Lin W. (2017) Assessing the utility of social media as a data source for flood risk management using a real-time modelling framework. Journal of Flood Risk Management 10:370-380

Yin J., Lampert A., Cameron M., Robinson B., Power R. (2012) Using Social Media to Enhance Emergency Situation Awareness IEEE Computer Society pp 1541-1672