

Original citation:

Degrossi, L. C., Albuquerque, João Porto de, Restrepo-Estrada, Camilo, Mobasher, A. and Zipf, A. (2017) Exploring the geographical context for quality assessment of VGI in flood management domain. In: AMCIS 2017 : A Tradition of Innovation, Boston, MA, USA, 10-12 Aug 2017. Published in: Proceedings of the Americas Conference on Information Systems (AMCIS 2017) pp. 1-10.

Permanent WRAP URL:

<http://wrap.warwick.ac.uk/93042>

Copyright and reuse:

The Warwick Research Archive Portal (WRAP) makes this work by researchers of the University of Warwick available open access under the following conditions. Copyright © and all moral rights to the version of the paper presented here belong to the individual author(s) and/or other copyright owners. To the extent reasonable and practicable the material made available in WRAP has been checked for eligibility before being made available.

Copies of full items can be used for personal research or study, educational, or not-for-profit purposes without prior permission or charge. Provided that the authors, title and full bibliographic details are credited, a hyperlink and/or URL is given for the original metadata page and the content is not changed in any way.

A note on versions:

The version presented in WRAP is the published version or, version of record, and may be cited as it appears here.

For more information, please contact the WRAP Team at: wrap@warwick.ac.uk

Exploring the geographical context for quality assessment of VGI in flood management domain

Submission Type: Full Paper

Livia Castro Degrossi

Department of Computer Systems
University of São Paulo
São Carlos, Brazil
degrossi@usp.br

João Porto de Albuquerque

Centre for Interdisciplinary
Methodologies
University of Warwick, Coventry, UK
GIScience Research Group
Heidelberg University, Germany
j.porto@warwick.ac.uk

Camilo E. Restrepo-Estrada

São Carlos School of Engineering
University of São Paulo
São Carlos, Brazil
cerestrepo@usp.br

Amin Mobasher

GIScience Research Group
Heidelberg University
Heidelberg, Germany
a.mobasher@uni-heidelberg.de

Alexander Zipf

GIScience Research Group
Heidelberg University
Heidelberg, Germany
zipf@uni-heidelberg.de

Abstract

Volunteered Geographic Information (VGI) has been used to complement or substitute authoritative data in flood management domain. The main issue regarding the use of volunteered information is to estimate its quality, mainly because it may suffer from heterogeneous quality. Therefore, several methods have been developed in the past few years in order to assess VGI quality. However, existing works lack in assessing VGI quality for the purpose of flood management. To overcome this gap, we propose a method for assessing the quality of VGI for this purpose. This method uses a set of quality metrics that were developed for measuring VGI plausibility. A multiple linear regression was carried out in order to demonstrate the relationship between VGI plausibility and the quality metrics. The results showed that plausibility can be explained by 5 quality metrics. Thus, the proposed method is able to estimate the plausibility of VGI in flood management domain.

Keywords

Volunteered Geographic Information, VGI, Quality Assessment, Flood Management

Introduction

In the past few years, Volunteered Geographic Information (VGI) (Goodchild 2007) has gained special attention in flood management domain due to its potential to underpin the detection of flood events (Longueville, Bertrand; Luraschi, Gianluca; Smits, Paul; Peedell, Stephen; De Groeve 2010; Ludwig et al. 2015), support decision-making (Ahmad and Simonovic 2006; Horita et al. 2015; Simonovic 1999), improve flood forecast (Mazzoleni et al. 2017), and complement authoritative data (Degrossi et al. 2014; Lanfranchi et al. 2014). Potentially, geographic information made available by volunteers can contribute to minimize uncertainty in flood prediction and also improve response to flood events (Ostermann and Spinsanti 2011).

Special attention should be given to information quality when using VGI as source of information in flood management activities (e.g. flood prediction). The volunteers that guarantee a vast amount of information may provide information with varying quality mainly because of the lack of quality standards and systematic documentation during the collection process, but also due to volunteers' knowledge and expertise. Particularly, when predicting a flood event or responding to it, high-quality information is highly desirable. If the information is posted later, this becomes less valuable, for instance, for flood prediction or early warning systems. Furthermore, a slow (disaster) response based on low-quality information could lead to severe consequences such as the loss of lives (Ostermann and Spinsanti 2011). Hence, information quality could limit the use of information. Therefore, verifying the quality of VGI is an important step before its use.

Prior studies have addressed the issue of VGI quality in flood management by developing different methods (e.g., using authoritative data and geographic context) to assess its quality (Hung et al. 2016; Poser and Dransch 2010). Although the relationship between VGI quality and the characteristics of the geographic context was already discussed in previous work (Hung et al. 2016), the combination of those characteristics and the event's characteristics has not been addressed yet. This is relevant because floods are context-dependent and dynamic events, i.e. floods usually occur near to water resources areas and their characteristics change from one event to another, respectively.

In this work, we examine the characteristics of the geographic context and the event in order to assess VGI plausibility. Thus, the overall research question of this paper is "*how can the plausibility of VGI be estimated based on the geographic context and the event's characteristics in flood management domain?*". To answer this, we propose a set of metrics and a method to assess the plausibility of VGI for flood management.

With the development of this method, we aim at supporting the use of VGI in flood management by minimizing the uncertainty regarding its quality. Particularly, we aim at increasing the use of VGI in decision-making process, flood prediction, etc. We argue that having more high-quality information regarding potential flood-affected areas could help decision makers make better decisions, as shown by Horita et al. (2015). Furthermore, we also argue that VGI could be used for reducing the uncertainty regarding flood prediction.

The remainder of this paper is structured as follows: in Section 2, we discuss the related works on quality assessment of VGI. In Section 3, we describe the methodology employed to develop this work. In Section 4, we present the results of our work. In Section 5, we discuss the main findings. Finally, concluding remarks are made in Section 6.

Related Works

The quality assessment is an important step to verify if the information is good enough for the purpose for which it will be used. This step is even more important when VGI is taken into account. An important aspect of the quality assessment is the application domain since the definition of quality strongly depends on it (Bordogna et al. 2016). Hence, several methods have been proposed to assess the quality of VGI in different application domains.

Considering flood management domain, Poser and Dransch (2010) have used authoritative data for estimating the quality of VGI. The authors compared the inundation depth provided by volunteers with the one measured by an authoritative source. Similarly, Moreira et al. (2015) and Degrossi et al. (2014) have

estimated the quality of water level observations by comparing them with real-time sensor data. This method is, however, often limited by the availability of authoritative data (Hung et al. 2016) and costs and licensing restrictions (Mooney et al. 2010). An alternative solution is to analyze the geographic context. Hung et al. (2016) have analyzed geolocation factors in order to assess the credibility of VGI in a flood response scenario.

Although those studies provide interesting findings, little is known about how the characteristics of the geographic context and the flood event could explain the plausibility of VGI.

Quality Metrics

The definition of (information) quality depends on the domain which the information will be used. Considering flood management domain, we define VGI quality based on the concept of plausibility. In general, plausibility can be understood as the likelihood of information being true or how much information can be believed. Here, we define plausibility as the likelihood of rainfall- or flood-related VGI being true.

Based on this definition, we formulate the following hypothesis: “*The plausibility increases with the spatial proximity to water resource areas and/or flood-prone areas and the temporal proximity of the reported event*”. We argue that the closer the information is of, for example, water resources areas, higher is the likelihood of being true (e.g. information about a flood event originating from an area with water resources nearby is more likely of being true than an information from an area with no water resources nearby). Moreover, the closer the VGI location is of the reported event, higher is the likelihood of information being true (e.g. information about a flood event provided near to the event is more likely of being true than information provided far away). It has been argued that volunteers closer to the event are more likely to provide high-quality information (Goodchild and Glennon 2010).

For measuring plausibility, we propose a set of quality metrics (Table 1). The metrics were derived from the works by Craglia et al. (2012), Fava (2015), Friberg et al. (2011) and Sabbata and Reichenbacher (2012) and are based on the characteristics of the geographic context (e.g. water resources areas, land cover, etc.) and volunteers.

| Metric | Description | Measure |
|--|---|---|
| Geographic context (Spatial proximity) | Characteristics of the location near or around VGI; geographic location relative to other information, or known event | (x_1) distance to water resource areas (x_2) distance to flood prone areas |
| Source type | Public authority or common citizen, expert or non-expert | (x_3) source type, e.g. public authority, expert volunteer, or non-expert volunteer |
| Temporal proximity | Information publishing date | (x_4) temporal difference to a known event |
| Verification | Event is also reported by other information source | (x_5) detection in another information source |

Table 1. Metrics for measuring VGI plausibility

Methodology

In the following subsection, we describe the materials and methods used for the development of this work.

Study Area

The city of São Paulo was selected as the study area (Figure 1). This large city is the capital of the state of São Paulo, with approximately 12,038,175 inhabitants. Furthermore, it has a tropical weather that is characterized by dry winter and rainy summer. Due to the type of weather, flood events are more probable from December to March, summer in the southern hemisphere, when several cases of heavy rain are

registered. Floods in São Paulo are mainly characterized by flash floods, which are caused by excessive rainfall in a short period of time, usually less than 6 (six) hours (National Weather Service, 2016). Flood events are a major concern for local government and the population as they affect countless people.

Datasets

Authoritative data

For measuring the defined metrics, we have used data provided by four authoritative sources, i.e. National Water Agency¹ (ANA), GeoSampa², Emergency Management Center (CGE) and National Center for Monitoring and Early Warning of Natural Disasters³ (Cemaden). Firstly, we obtained data on water resources areas from GeoSampa, a platform that is maintained by the city hall of the city of São Paulo. In the platform, it is possible to obtain hydrological data, topographic data, etc. Secondly, we retrieved information regarding flood-prone areas from ANA, which operates in the management of water resources in Brazil. The agency provides, among other things, a map of the flood-prone watercourses together with their degree of impact and vulnerability. We used the data provided by both sources for measuring the metrics *distance to water resource areas* and *distance to flood-prone areas*, respectively. The distance on both metrics was measured based on the Euclidian distance that provides the distance between two points. After, we selected the water resource and flood-prone watercourse that were the closest to VGI location.

Thirdly, we retrieved information regarding rainfall- and flood-related events from Cemaden and CGE, respectively. Cemaden is a national center that is responsible for monitoring natural disasters, e.g. floods and droughts, in Brazil. Hence, rainfall data is collected through an automatic rainfall gauge network (Figure 1). Each rainfall gauge registers the amount of rain at 10-minute intervals. Here, it is interesting to highlight that not all rainfall gauge work properly. Therefore, we only consider the ones that were working properly in our analysis. In a similar way, CGE is responsible for monitoring flood events in the city of São Paulo. The agency provides information regarding flood-affected areas in São Paulo on a daily basis.

The data from both sources were used for the detection of the reported event in other information source and also for measuring the distance and temporal difference to the known event. First, we verified if the event reported by the volunteer was also detected by CGE. Hence, we checked if the agency has detected any event within an interval of two hours, one before and one after VGI publishing date. We specified this interval because the volunteer could send information after the event has occurred or before it has been detected by another source. If more than one event has been detected, we selected the event that was the closest to VGI location. On the other hand, if any event has been detected by CGE, we checked if the event was detected by Cemaden. For this, we employed the same approach that we employed with CGE. However, instead of searching for flood events, we verified if any rainfall gauge has registered a rainfall event. In this particular case, we restricted our search into a radius of 5 km from the VGI location since it could have considerable amount of variation in rainfall data collected by rainfall gauges far away.

If a flood- or rainfall-related event has been detected by CGE or Cemaden, we set the value of the metric detection of the event in other information source to 1. Otherwise, we set it to 0. Moreover, if the event has been detected by CGE or Cemaden, the distance from the detected event to VGI location was used as a measure for the metric distance to known event. The information regarding the detected event was also used to measure the quality metric temporal difference to known event. For this, we measured the difference from the VGI publishing date to the start or end of the event. However, if VGI has been provided when the event was occurring, this difference is zero.

Finally, for our analysis, we set the value of VGI plausibility according to the distance to the known event (Table 2). We argue that VGI closer to an event is more likely of being true than the one that is far away. However, if any event was detected within the time interval, the plausibility was set to 0.

¹ <http://www2.ana.gov.br>

² <http://geosampa.prefeitura.sp.gov.br>

³ <http://www.cemaden.gov.br/>

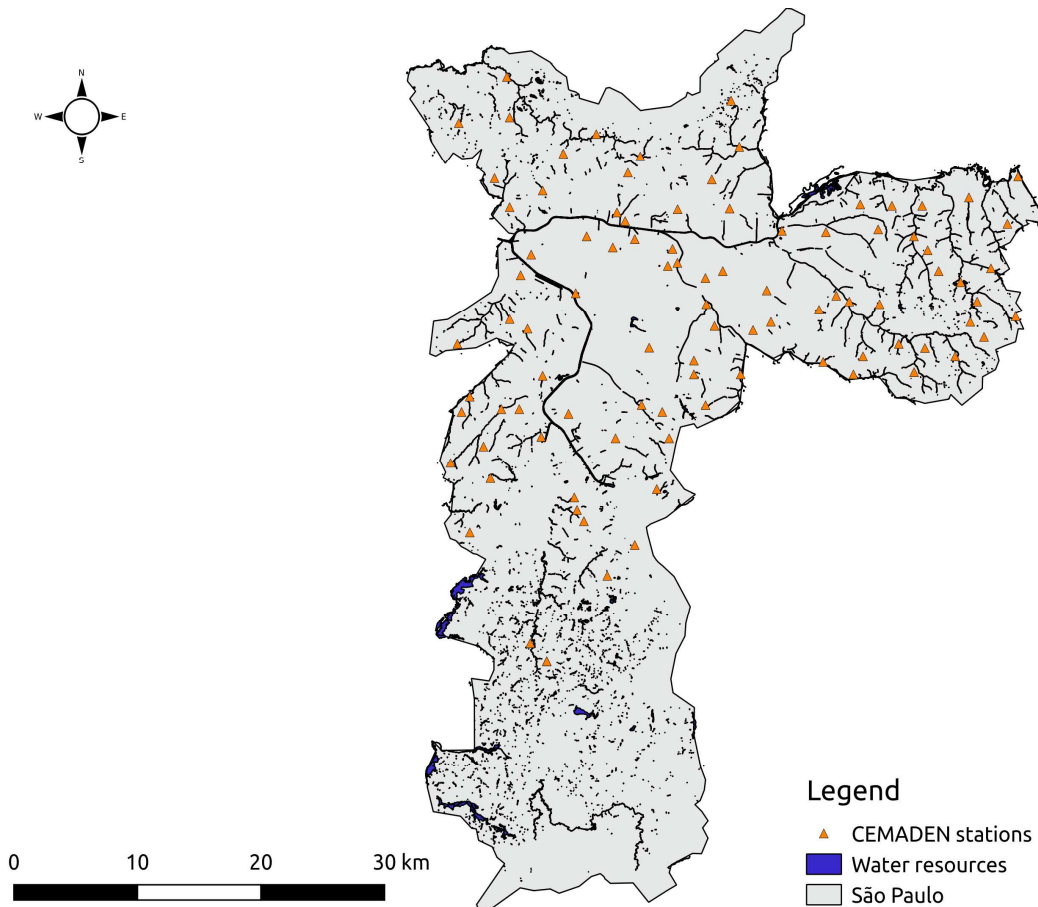


Figure 1. Cemaden stations in the city of São Paulo

Volunteered data

Twitter messages, i.e. tweets, were used to explore the potential of the proposed method which aims at measuring the plausibility of VGI. For this analysis, we retrieved georeferenced tweets from the city of São Paulo. From this set, we selected only the tweets published between 01/01/2016 and 02/20/2016. This period was chosen because it corresponds to the rainy season in South America, between the months of December and March. Moreover, according to the Emergency Management Center⁴, there were 26 records of rainfall- and flood-related events just in this period. The tweets were further filtered according to a set of keywords. The keywords are in Brazilian-Portuguese and are related to rainfall and flood events. The keywords are “chuv*” (chuva, chuveiro, chuvarada, etc.), “garoa*” (gareando, etc.), “temp*” (temporal, tempestade, tempo ruim, etc.), “alag*” (alagamento, alagado, etc.), “inund*” (inundação, inundado, etc.), “enchente” and “enxurrada”. After these steps, the initial dataset contained 5,961 tweets.

In the following, we manually analyzed the tweets individually to verify if it is indeed related with rainfall or flood events. This analysis was necessary because words as “garoa” or “chuva” may have different meanings. The keyword “garoa”, for instance, is used as a reference to the city of São Paulo, which is known as the land of drizzle. This analysis consisted of verifying the (information) content and classifying the tweet as related or unrelated (Table 3). Related tweets are all those that through text, photo or video report a rainfall or flood event. Unrelated tweets are all those that do not fit the previous definition. To help us in this classification, we also analyzed, when available, the additional information provided by volunteers through a link. At the end of this analysis, there were 785 related tweets.

⁴ <http://www.cgesp.org/>

| Distance to known event | Plausibility |
|-----------------------------|--------------|
| distance \leq 1 km | 1,0 |
| 1 km < distance \leq 2 km | 0,8 |
| 2 km < distance \leq 5 km | 0,5 |
| distance > 5 km | 0 |

Table 2. Plausibility values based on the distance to known event

| Date/time | Tweet (Portuguese) | Tweet (English) | Status |
|------------------------|--|--|-----------|
| 2016-01-14 17:45:25 | Uma chuva fina, mas intensa cobre a cidade. #Sampa #Chovendo @São Paulo, Brazil https://t.co/7GEtwCqD6M | A fine but intense rain covers the city #Sampa #Chovendo @São Paulo, Brazil https://t.co/7GEtwCqD6M | Related |
| 2016-01-27 17:37:39 | “Chuva chuva e mais chuva.... https://t.co/vzC9w01qQ8 ” | “Rain rain and more rain https://t.co/vzC9w01qQ8 ” | Related |
| 2016-01-11 14:23:04 | Depois de muita chuva, fui conhecer o espaço literário Casa das... https://t.co/zM8rVFhaBH | After a lot of rain, I have gone to known the literary space House of... https://t.co/zM8rVFhaBH | Unrelated |

Table 3. Examples of related and unrelated tweets

Multiple Linear Regression

Linear regression is an approach for modeling the relationship between a dependent variable and an independent variable. In a linear regression model, it is assumed that the relationship between both variables is linear. When the model has one independent variable, it is called simple linear regression. On the other hand, when the model has two or more variables, it is called multiple linear regression. In the latter, the relationship is modeled using several predictors or independent variables. The multiple linear relationship between the variables can be described by the following regression equation:

$$y = \alpha + \beta_1x_1 + \beta_2x_2 + \beta_3x_3 + \beta_4x_4 + \dots + \beta_kx_k + \varepsilon \quad (1)$$

where y is the dependent variable, x_i is the i th independent variable, β_i is the i th regression coefficient, α is the intercept and ε (epsilon) is a random error component. The coefficient β represents the change in the dependent variable y associated with a one-unit increase in the independent variable x . The value of y can increase if the value of β is positive or decrease, if the value of β is negative. Thus, the independent variables provide an explanation or prediction for the dependent variable.

We particularly choose this model because it is simple and often provides an adequate and interpretable description of how the independent variables affect the dependent variable (Hastie et al. 2009). Moreover, we selected it because we did not have any knowledge regarding the functional relationship between the variables.

Results

Firstly, we analyzed the relationship between the quality metrics (Table 1) and the (dependent) variable plausibility. On the one hand, we found out that the metric *source type* is not significant to the model since we classified all volunteers as non-expert. We did this because we had no information regarding volunteers' expertise. Hence, we removed this variable of our model. On the other hand, the other metrics are statistically significant to the model with a confidence level of 99%, as shown in Table 4. The metrics

distance to water resource areas and *distance to flood prone areas* have a weak linear relationship with the dependent variable, i.e. both metrics can explain approximately 4% of the variability of the dependent variable. Therefore, we also removed both metrics of our model. In contrast, the metrics *temporal difference to a known event* and *detection in another information source* have a strong linear relationship with the dependent variable. Together, they can explain 68% (Adjusted R-squared) of the variability of the dependent variable. Thus, VGI plausibility can be measured based on the *temporal difference to a known event* and *detection in another information source*.

| | | Main model | Alternative model 1 | Alternative model 2 |
|---------------------------|---|-----------------------------|--------------------------|---------------------------|
| Independent variables | (x ₄) temporal difference to known event | 0.0017361 (1.22e-14 ***) | - | 0.0129574 (<2e-16 ***) |
| | (x ₅) detection in another information source | 0.4485786 (< 2e-16 ***) | 0.491960 (<2e-16 ***) | - |
| Adjusted R-squared | | 0.6836 | 0.6805 | 0.4715 |
| AIC | | -9819.1 | -9761.5 | -6761.5 |

Table 4. Results of multiple linear regression

Secondly, we verified the quality of the model by measuring the AIC (Akaike Information Criterion) value, which indicates the relative quality of statistical models for a given dataset. The preferred model should be the one with minimum AIC value (Sakamoto et al. 1986). The AIC value of our model is -9819.1 (Table 4), indicating the relative quality of the proposed model. Thus, the model for measuring the plausibility of VGI is presented in Equation (2).

$$\text{plausibility} = \beta_1 \text{temporalDifferenceKnownEvent} + \beta_2 \text{detectionOtherSource} \quad (2)$$

Besides the main model, we investigated if one of the metrics could be removed and, even so, keep the relative quality of the model. After some tests, we identified 2 alternative models that can be employed for estimating the plausibility of VGI (Table 4). These models differ in accordance with the type of the independent variable. To determine them, we removed one metric at a time and tested the statistical significance of the remaining metric. Finally, we measured the quality of the model through the AIC value.

First, we removed the metric that is the most difficult to measure in real-time, i.e. the temporal difference to a known event, since determining the exact time that the event has started or ended could be a challenge. According to the results, the relative quality of the alternative model 1 is almost similar to the main model since its AIC value is -9761.5. Moreover, the metric detection in another information source is significant to the model because it can explain 68% of the variability of the dependent variable. In the following, we removed the metric detection in another information source. The relative quality of the alternative model 2 is smaller than the main model, its AIC value is -6761.5. However, we argue that it is also relevant to measure the plausibility of VGI because the metric temporal difference to a known event can explain 47% of the variability of the dependent variable.

Comparing both alternative models, the first model is better than the second because the AIC value of the second model is lower than the AIC value of the first. Therefore, if the temporal difference could not be determined, the plausibility of VGI can be measured by the metric detection in another information source.

Discussion

In this work, we propose a method for assessing the plausibility of VGI in flood management domain. For this, we also propose a set of metrics for measuring the plausibility. These metrics were developed based on the different aspects of the geographical context that could have an influence on VGI plausibility. Two of them were proposed by Hung et al. (2016) and four of them are new metrics proposed by us.

To measure their statistical significance, we carried out a multiple linear regression. By employing this model, we aimed at verifying the relationship between the dependent variable (i.e. plausibility) and independent variables (i.e. quality metrics), i.e. we investigated if the independent variables provide an explanation for the dependent variable indeed. We found out that the metric source type cannot explain the plausibility of VGI, possibly because all volunteers were classified as non-expert. The metrics *distance to water resource areas* and *distance to flood prone areas* are statistically significant however have a weak linear relationship. On the other hand, the metrics *temporal difference to a known event* and *detection in another information source* are statistically significant to our model and have a strong linear relationship. Thus, both metrics can be employed to predict the plausibility of VGI. By validating their significance, we took the first step towards answering our initial research question (Section 1).

Differently from Hung et al. (2016), in this work, we verified that the distance of VGI to water resource areas and flood-prone areas do not have an influence on the plausibility of VGI. We also found out that the detection of the event in another information source and the temporal difference to a known event can explain the plausibility of VGI since they have an influence on it, when considering flood management domain.

As well as in previous works, we also used authoritative data to measure our metrics. A drawback of using them is that it might be out-of-date (e.g. data from ANA or GeoSampa), or it is not provided in real-time (e.g. data from CGE), which could hinder its use. Therefore, further investigation should be carried out in order to search for alternative information sources. Previous works, for instance, showed the potential of Location-based Social Network for event detection (Andrade et al., 2017.; Longueville et al. 2010; Steiger et al. 2015). Moreover, it is interesting to investigate some aspects that were not addressed in this work. Here, we did not consider surface elevation such as Hung et al. (2016). Thus, further investigation is required to verify if and how surface elevation, together with the proposed metrics, could explain the plausibility of VGI. Another aspect not addressed here is the number of users' followers and the number of retweets. We argue that these points have the potential of explaining the plausibility of VGI because individuals usually follow people that they trust and retweet information that they believe it is true.

The development of this method contributes to the research field of quality assessment of VGI by providing a new approach for estimating VGI quality in flood management domain. Moreover, by employing this method, the use of VGI in flood management domain could increase since it minimizes the uncertainty regarding its quality. Thus, high-quality VGI could be used in decision-making, flood prediction, early warning systems, etc.

Conclusion

In this paper, we presented a method for assessing the plausibility of VGI in flood management domain. We also developed a set of metrics that is used by the method for measuring the plausibility. A multiple linear regression was carried out to verify the relationship between the dependent variable (i.e. plausibility) and independent variables (i.e. quality metrics). As a result, we demonstrated that 2 metrics are statistically significant to our model and have a strong linear relationship. The metrics *temporal difference to a known event* and *detection in another information source* are the mainly metrics that can predict the plausibility of VGI since both metrics are statistically significant in the main and alternative models.

Further investigation should be carried out in order to collect more evidence and, thus, accept or reject our initial hypothesis. Therefore, we will employ the method to estimate the plausibility of a different VGI dataset. Moreover, it is still necessary to investigate alternative information sources to measure the quality metrics mainly because some authoritative sources could hamper their use.

Acknowledgements

The authors would like to express thanks for the financial support provided by CAPES. Livia Castro Degrossi is grateful for the financial support from CNPq (Grant no. 201626/2015-2) and CAPES (Grant No. 88887.091744/2014-01). João Porto de Albuquerque acknowledges financial support from CAPES (Grant no. 88887.091744/2014-01), and Heidelberg University (Excellence Initiative II/Action 7). Camilo Restrepo-Estrada is grateful for the financial support from CAPES-PROEX.

REFERENCES

- Ahmad, S., and Simonovic, S. P. 2006. "An Intelligent Decision Support System for Management of Floods," *Water Resources Management* (20:3), Kluwer Academic Publishers, pp. 391–410 (doi: 10.1007/s11269-006-0326-3).
- Andrade, S. C. de, Restrepo-Estrada, C., Delbem, A. C. B., Mendiondo, E. M., and Albuquerque, J. P. de. (n.d.). "Mining rainfall spatio-temporal patterns in Twitter: a temporal approach," in *Proceedings of the 20th AGILE Conference on Geographic Information Science*, Wageningen, The Netherlands.
- Bordogna, G., Carrara, P., Criscuolo, L., Pepe, M., and Rampini, A. 2016. "On predicting and improving the quality of Volunteer Geographic Information projects," *International Journal of Digital Earth* (9:2), Taylor & Francis, pp. 134–155 (doi: 10.1080/17538947.2014.976774).
- Craglia, M., Ostermann, F., and Spinsanti, L. 2012. "Digital Earth from vision to practice: making sense of citizen-generated content," *International Journal of Digital Earth* (5:5), Taylor & Francis Group, pp. 398–416 (doi: 10.1080/17538947.2012.712273).
- Degrossi, L. C., Albuquerque, J. P. de, Fava, M. C., and Mendiondo, E. M. 2014. "Flood Citizen Observatory: a crowdsourcing-based approach for flood risk management in Brazil," in *Proceedings of the 26th International Conference on Software Engineering and Knowledge Engineering*.
- Fava, M. C. 2015. "Modelo de Alerta Hidrológico com Base Participativa usando Sistema de Informações Voluntárias para Previsão de Enchentes."
- Friberg, T., Prödel, S., and Koch, R. 2011. "Information Quality Criteria and their Importance for Experts in Crisis Situations," in *Proceedings of the 8th International ISCRAM Conference*, Lisbon, Portugal, pp. 1–10 (available at <http://www.iscramlive.org/ISCRAM2011/proceedings/papers/150.pdf>).
- Goodchild, M. F. 2007. "Citizens as sensors: the world of volunteered geography," *GeoJournal* (69:4), pp. 211–221 (doi: 10.1007/s10708-007-9111-y).
- Goodchild, M. F., and Glennon, J. A. 2010. "Crowdsourcing geographic information for disaster response: a research frontier," *International Journal of Digital Earth* (3:3), pp. 231–241 (doi: 10.1080/17538941003759255).
- Hastie, T., Tibshirani, R., and Friedman, J. 2009. *The elements of statistical learning*, Springer-Verlag New York (doi: 10.1007/978-0-387-84858-7).
- Horita, F. E. A., Albuquerque, J. P. de, 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," *Computers & Geosciences* (Vol. 80) (doi: 10.1016/j.cageo.2015.04.001).
- Hung, K.-C., Kalantari, M., and Rajabifard, A. 2016. "Methods for assessing the credibility of volunteered geographic information in flood response: A case study in Brisbane, Australia," *Applied Geography* (68), pp. 37–47 (doi: 10.1016/j.apgeog.2016.01.005).
- Lanfranchi, V., Wrigley, S. N., Ireson, N., Ciravegna, F., and Wehn, U. 2014. "Citizens' Observatories for Situation Awareness in Flooding," in *Proceedings of the 11th International ISCRAM Conference*, Pennsylvania, USA, pp. 145–154.
- De Longueville, Bertrand; Luraschi, Gianluca; Smits, Paul; Peedell, Stephen; De Groeve, T. 2010. "Citizens as Sensors for Natural Hazards: A VGI integration workflow," *Geomatica* (64:1), p. 2010.
- Longueville, B. De, Luraschi, G., Smits, P., Peedell, S., and Groeve, T. De. 2010. "Citizens as Sensors for Natural Hazards: A VGI Intergration Workflow," *Geomatica* (64), pp. 41–59.
- Ludwig, T., Reuter, C., and Pipek, V. 2015. "Social Haystack: Dynamic Quality Assessment of Citizen-Generated Content During Emergencies," *ACM Transactions on Computer-Human Interaction* (22:4), p. 17:1–17:27 (doi: 10.1145/2749461).
- Mazzoleni, M., Verlaan, M., Alfonso, L., Monego, M., Norbiato, D., Ferri, M., and Solomatine, D. P. 2017. "Can assimilation of crowdsourced data in hydrological modelling improve flood prediction?"

- Hydrology and Earth System Sciences* (21:2), Copernicus GmbH, pp. 839–861 (doi: 10.5194/hess-21-839-2017).
- Mooney, P., Corcoran, P., and Winstanley, A. C. 2010. “Towards quality metrics for OpenStreetMap,” in *Proceedings of the 18th SIGSPATIAL International Conference on Advances in Geographic Information Systems*, New York, NY, USA, pp. 514–517 (doi: 10.1145/1869790.1869875).
- Moreira, R. B., Degrossi, L. C., and de Albuquerque, J. P. 2015. “An experimental evaluation of a crowdsourcing-based approach for flood risk management,” in *Proceedings of 12th Workshop on Experimental Software Engineering*, Lima, Peru, pp. 393–403.
- Ostermann, F. O., and Spinsanti, L. 2011. “A Conceptual Workflow For Automatically Assessing The Quality Of Volunteered Geographic Information For Crisis Management,” in *Proceedings of the 14th AGILE International Conference on Geographic Information Science*, Utrecht, Netherlands.
- Poser, K., Dransch, D. 2010. “Volunteered geographic information for disaster management with application to rapid flood damage estimation,” *Geomatica* (64:1), pp. 89–98.
- Poser, K., and Dransch, D. 2010. “Volunteered Geographic Information for Disaster Management with Application to Rapid Flood Damage Estimation,” *Geomatica* (64:1), pp. 89–98.
- Sabbata, S. De, and Reichenbacher, T. 2012. “Criteria of geographic relevance: an experimental study,” *International Journal of Geographical Information Science* (26:8), pp. 1495–1520 (doi: 10.1080/13658816.2011.639303).
- Sakamoto, Y., Ishiguro, M., and Kitagawa, G. 1986. *Akaike information criterion statistics*, Dordrecht, The Netherlands: D. Reidel Publishing Company.
- Service, N. W. 2016. “What is flash flooding?” (available at <http://www.weather.gov/phi/FlashFloodingDefinition>; retrieved January 1, 2016).
- Simonovic, S. P. 1999. “Decision support syst for flood management in the Red River Basin,” *Canadian Water Resources Journal* (24:3), Taylor & Francis Group, pp. 203–223 (doi: 10.4296/cwrj2403203).
- Steiger, E., de Albuquerque, J. P., and Zipf, A. 2015. “An Advanced Systematic Literature Review on Spatiotemporal Analyses of Twitter Data,” *Transactions in GIS* (19:6), pp. 809–834 (doi: 10.1111/tgis.12132).