

Nat. Hazards Earth Syst. Sci. Discuss., referee comment RC2
<https://doi.org/10.5194/nhess-2022-109-RC2>, 2022
© Author(s) 2022. This work is distributed under
the Creative Commons Attribution 4.0 License.

Comment on nhess-2022-109

Zhang Hongping (Referee)

Referee comment on "Exploring the utility of social media data for urban flood impact assessment in data scarce cities" by Kaihua Guo et al., Nat. Hazards Earth Syst. Sci. Discuss., <https://doi.org/10.5194/nhess-2022-109-RC2>, 2022

Urban flooding is a hot issue in the world. There are still many challenges in urban flood emergency response, one of which is collecting the timely flooding information. This research demonstrates a technological workflow to get flooding information from social media and deeply mine valuable information from them. It's doubtlessly a good try, though there is still a long way to put it in practical use. Here are some specific comments and suggestions as following.

- About the introduction section:

The introduction section seems a little logically weak and needs a better organization. For example:

(1)The content in the first paragraph seems a little far away from the research topic. I would like to introduce the topic directly from urban floods in the world, and then explain why we need to get data from social media for urban flood emergency management and what the difficulties are.

(2)At the beginning of the second paragraph, it would be better to state the urban flood situation in a worldwide view, not only from China, because urban flooding is a worldwide hot issue. China can be used as one of examples.

(3)The traditional observation methods are mentioned as "costly, delay and time-consuming" methods in Line 57. But discussion about the traditional observation methods seems insufficient. It would be better to give more discussion on what kinds of traditional

observation methods there are and how costly, delay and time-consuming they are before Line 57.

(4) Line 45 seems to indicate that the existing researches use the unified format data. If there are unified format data, why do we need to use the unformatted data from Weibo and Tiktok? Please give more explanations about it. It would be better to give more details about the existing researches on utilizing the social media data for flood emergency management. What have they done and what are their limits?

(5) From Line 57, two research questions are listed. There might be one more question: how are the data from social media validated before they are used? Or what are the uncertainties when they are used? Please give a little more statements about it.

- About Table 2:

(1) Table 2 seems a little confusing. Usually the first row is the header which defines the content of each column, and each column has contents matching with the header. In this table, row 5 seems like a new header, and row 7 doesn't seem to match either headers. It would be better to reorganize this table.

(2) More explanation about time lag in Table 2 would be better. Does it mean: the Pearson correlation coefficient between the hourly precipitation and the hourly tweet number with a certain time lag? If it does, please specifically clarify the hours of precipitation and hours of tweets analyzed during calculating their Pearson correlation coefficient.

(3) Line 212 describes Storm 1 filtering as by excluding relevant discussions. Please give more explanations on how relevant discussions are identified in a huge number of tweets.

- About Table 4.

It might be more figurative to transpose this table by placing the check-in data and flood-point data for one district in one row in order to give a better comparison.

- About Section 4

Some contents in Section 4 seem a weak relation with the results of this research. For example:

(1) In Section 4.1 the author seems to discuss the value or the role of social media for various stakeholders. It's a fact, but not a result of this research. It might be better to use this content to develop the necessity of this research in the introduction section.

(2) Section 4.3 is titled as "comparison with other platforms and countries". I would like to make comparison just between researches, weakening the idea of countries. Furthermore, the content in this section seems like the progress of other similar research. Is it possible to put it in the introduction section as existing research progress?

(3) Section 4.2 gives a good example of deeper mining of the value of social media data. It might be better to focus on this example and give more details about it in Section 4.

- Others

Some other minor questions or suggestions are seen in the manuscript.

Please also note the supplement to this comment:

<https://nhess.copernicus.org/preprints/nhess-2022-109/nhess-2022-109-RC2-supplement.pdf>