Dear Valerio Lorini,

Thank you very much for giving us the opportunity to improve our manuscript. Upon the insightful and constructive comments given by you, we have conducted a careful and thorough revision of our manuscript. Before the presentation of our point-by-point responses to the comments given by you, we summarize the major revisions of the manuscript as follows:

1. We have added the most recent progress in the text classification using the recurrent neural network.
2. We have revised our manuscript according to your comments and suggestions.

We believe these revisions have significantly improved the quality of our manuscript, and made our method more clearly to users. Once again, we sincerely thank you for the constructive comments.

Sincerely yours,

Hui Liu, Ya Hao, Wenhao Zhang, Hanyue Zhang, Fei Gao, Jinping Tong*

The paper describes how information extracted from micro-blogging platform Weibo can be used to build a model for detecting urban floods. The model was trained using ground truth data merged with social media posts as training dataset and a set of known events have been used as reference for evaluation of the transfer learning potential of the model.

I think the paper is overall clear and well structured. While the work does not present a novelty in terms of technology or methodology, the effort of (i) applying it to a new, according to my knowledge, micro-blog data source with an extensive authoritative ground truth and (ii) to a specific type of flood, namely urban waterlogging makes the paper interesting for community researchers. The literature review covers widely the related works but is missing some of the most recent developments where similar methodology was applied.

Response: Thank you for your constructive suggestion. In our manuscript, we have added the most recent progress in text classification using recurrent neural network.

The methodology applied is correct and the results well clarified and documented with clear figures and tables. The authors made the code and data available, which is a much appreciated and very good practice.

General comments - Since the publisher is the European Geoscience Union (EGU) and the audience is expected to have little experience with Weibo, it would be useful to give some more context to non-users of the platform about how the original posts are structured and how the data was extracted in terms of text and location.

Response: Thank you for your comment. Weibo is run by sina.com Company in China and it is similar to Twitter. A Weibo post often contains the fields such as user ID, user name, microblog content, posted time, etc. Some posts include pictures.
    We used keywords "drowning" or "waterlogging" to download posts using Sina Weibo API, as such posts often describe waterlogging events already occurred in some places. To identify the location of waterlogging, we first marked the names of communities, roads and orientation from the posts using named entity recognition technique. Next, the names were matched to the national geographic names and national community names to extract the accurate geographic location.

- While it is very clear and well explained how the model was trained and the data was prepared, little description of the operational Monitoring of urban waterlogging in real time was provided. The authors could elaborate more on the scalability of their system.

Response: Thank you for your suggestion. We have completed the development of the urban waterlogging monitoring system based on the WeChat, which is a social media software similar to WhatsApp in Europe and American. Our applet can be launched by one click from WeChat. Once the applet is started, it will automatically calculate the distance from the current location to the nearby waterlogging point. When the distance exceeds the predefined threshold, the applet will trigger voice alarm to remind drivers that there is a waterlogging point nearby and drive carefully. With the help of WeChat's powerful web service capability and wide application, it is helpful for a lot of people to monitor flood deposits, especially for taxi and bus drivers.

Minor comments:

L120 It is not clear to me why multiple posts actually located to the same flood deposit were removed. It means posts had same text and locations or different text and same location? in case could you explain better maybe with examples or by stating the relation between positive samples and location. Is it 1 to 1? or many to 1? or many to many?

Response: Thanks for your insightful comments. We deleted duplicate posts that have same text and same location. If their content is different, we will keep all, even if their location is same. For example: (i) On the morning of May 12th, the reporter learned from the Nanning Traffic Police Detachment that because of the rainy day leading to Pingle Avenue surrounding roads have part of water. (ii) At present, there have been water accumulation in some sections of Yinhai Avenue, Ruihejiayuan and Pingle Avenue, please pedestrians and vehicles are requested to pass carefully. We kept both posts. If the content of these two posts was similar or the same, then we would keep one of them.

L120 It would be of great help to have the table 1 partially translated and described further, as it has been done in the following paragraph about the selection of negative samples.

Response: Thank you for your important suggestion. We have translated Table 1 in the revised manuscript.

L210 How the undersampling was achieved? Randomly removing posts or is there a methodology applied?

Response: The undersampling is done by using the third-party library imblearn implemented in Python. The RandomUnderSampler function in imblearn package implements boostrap sampling by setting the parameter replacement to true. The function randomly remove samples in original dataset with multiple classes, to build the balanced subset of the whole samples.

L335 At some point the authors introduce the term 'flood deposit' and it seems to be used as an interchangeable term with waterlogging. Since the term is repeated extensively it would be clearer for reader to get the definition in the introduction of the paper when waterlogging is introduced.

Response: Thank you for your careful review of our manuscript. "Flood deposit" and "waterlogging" are interchangeable in most cases. However, "Flood deposit" is more sense of a geographical location, which is easy to be flooded in the case of rainstorm. "Waterlogging" is more sense that flooding events occurred.

L300 Why not using an example referring to a flood post?

Response: Thank you for your constructive suggestion. In the manuscript, we have modified this part as follows:
For example, a post reads as follows: In the early hours of this morning, there was a heavy rain in Shenzhen. Many roads were flooded, and only the top of cars could be seen. Among them, the water at the bottom of Hezhou Bridge was serious, which caused the road to be interrupted. TP means true positive: the true is positive, and the classifier predicts a positive. For example, Bottom of Hezhou Bridge is flooded, and the classifier accurately reports this. TN means true negative: the truth is negative, and the classifier predicts a negative. For instance, Bottom of Hezhou Bridge is not flooded, and the classifier accurately reports this. FP means false positive: the truth is negative, but the classifier predicts a positive. Such as Bottom of Hezhou Bridge is not flooded, but the classifier inaccurately reports that it is. FN means false negative: the truth is positive, but the classifier predicts a negative. For example, Bottom of Hezhou Bridge is flooded, but the classifier inaccurately reports that is not.

L405 If numbers of events is based on microblogging, could it be that 'new technology adoption' rather than GDP is a leading factor for bigger orange spots? i.e. there are more waterlogging because the population discuss more about it on weibo? in order to clear such doubt it could be useful to report the number of posts per day in the several areas.

Response: Thank you for your insightful comment. We agree with you very much that new technology adoption is a factor related to the size of the orange dot. For example, Nanjing, Jiangsu Province has more than 200 posts every day, while Lhasa, Tibet Autonomous Region has about 40 posts per day. When a flood occurs, people discuss more waterlogging-related content on Weibo, we will collect more posts related to the waterlogging, which will lead to the orange dot bigger. Therefore, the residents of Nanjing area that discussed waterlogging via Weibo may make the orange dot bigger. However, it is worth noting that the new technology adoption itself is correlated to GDP. We will carry out more analysis to

explore the factors related to the number of Weibo posts discussing waterlogging in the future work.