NHESS Paper Submission

Social sensing of high-impact rainfall events worldwide: A benchmark comparison against manually curated impact observations

Reviewer responses

Reviewer #1:

This paper represents an original contribution on social sensing of high-impact rainfall events, data analysis disseminated by social media against a manually curated impact database created by the Met Office. The methodology, data analysis and discussion are illustrated with scientific rigor, also presenting the limitations of the study at the end.

However, the limitation of the paper is the absence of a section dedicated to reviewing the literature on the subject: are there other studies on the analysis of social media data relating to high-impact rainfall events or other extreme natural events?

there are also no references to risk communication according to the latest UNDRR reports. Does social media have an impact on disasters? do they have a preventive function of them? with the arrival of the pandemic, the use of social media as a communication tool - even institutional - has become increasingly impressive and decisive. Therefore, it is necessary to improve the theoretical structure of the paper; in addition to this, being a study that focuses on the perception of the impact of extreme natural events, it is also necessary to include the vast international literature on the subject of perception of risks associated with extreme natural events, with reference to the type of natural hazard analyzed in this paper.

It would also be interesting to compare social sensing applied to high-impact rainfall events with other experiments on the perception of various extreme natural phenomena such as earthquakes using other similar methodologies (see the crowdsource detections by Bossu, 2020, for example) to further investigate the issue.

We thank the reviewer for their assessment of our work.

To address each comment in turn:

1. **Additional literature review.** We have revised the Introduction section so that it now incorporates a 'Related Work' section. This expands the existing text in the Introduction which discusses other works to provide a more thorough literature review of similar studies which have used social media to analyse the impacts of extreme weather:

Related Work

A number of studies have explored the use of social media as a source of information about the impacts of extreme weather. Social sensing is an approach developed in recent years to analyse unsolicited social media data to detect real-world events of interest.

While social sensing is not specific to natural hazards and can be applied in a variety of contexts (Liu et al., 2015; Wang et al., 2012, 2019), social sensing has demonstrated usefulness for natural hazard events.

Twitter data was used by Sakaki et al. (2010) to detect earthquakes in Japan, with reports arriving in some locations before the shock had been detected by conventional seismography. Many studies have followed, using a number of different approaches to explore the use of social media as an information source during and following natural hazard events. Some studies have focused on the use of social media to better understand risk communication during an extreme natural hazard event. For example, Steward and Wilson (2016) explore the use of social media throughout the crisis lifecycle during Hurricane Sandy in the USA, building the STREMII model to better understand crisis communication during an extreme weather event; Rainear et al. (2018) used Twitter data collected during Hurricane Joaquin to explore the types of information communicated by state emergency management accounts to better understand the flow of risk communication during a crisis; Bossu et al. (2020) explored the use of crowdsourced information, along with Twitter data, in a bespoke application during the 2019 earthquake in Albania, finding that engagement of users with the app provided much more information about the damage caused as a result of the earthquake than was available using conventional methods.

Other studies have explored the use of social media to better understand the impacts of extreme weather events. Many studies focus on individual events. For example Fang et al. (2019) use data from the Chinese social media platform, Sina Weibo, during the 2016 Beijing rainstorm, finding a positive correlation between social media activity and precipitation intensity; Sit et al. (2019) examine Twitter data collected during Hurricane Irma, using geo-located tweets to identify locations with a high density of affected individuals and infrastructure damage; and Han et al. (2019) use data from Sina Weibo during the 2018 Shouguang flood to analyse the changes in sentiment of social media users during the different development stages of the flood. Further examples of other studies examing the impacts of individual weather events at one particular location include: studies relating to specific hurricanes in the United States (Guan and Chen, 2014; Kim and Hastak, 2018; Lachlan et al., 2014; Morss et al., 2017; Niles et al., 2019; Wu and Cui, 2018; Zou et al., 2018) and specific flooding events (Aisha et al., 2015; Brouwer et al., 2017; Cervone et al., 2016; Kankanamge et al., 2020; Li et al., 2018; Rossi et al., 2018).

Some authors have begun to explore the use of Twitter for more wide-scale specific weather event detection, Arthur et al. (2018) use Twitter data to detect and locate flood *events in the UK to produce maps of flood activity. De Brujin et al. (2019) compare Twitter activity relating to flooding and hydrological information with flood events in the NatCatSERVICE disaster database, finding a good comparison between these data sources. Boulton et al. (2016) use Twitter data collected during several time periods to detect and locate wildfires in the USA. Cowie et al. (2018) find that user reports on Twitter during the year can help to locate peaks in hayfever symptoms as a result of pollen levels in the UK. Furthermore, Spruce et al. (2020) examine Twitter data relating to named storms, wind and precipitation in the UK finding that it is possible to identify tweets which can be used to assess the impact of storms both temporally and spatially.*

The following references will also be added as follows:

Bossu, R., Fallou, L., Landès, M., Roussel, F., Julien-Laferrière, S., Roch, J. and Steed, R.: Rapid Public Information and Situational Awareness After the November 26, 2019, Albania Earthquake: Lessons Learned From the LastQuake System, Front. Earth Sci., 8, doi:10.3389/feart.2020.00235, 2020.

Fang, J., Hu, J., Shi, X. and Zhao, L.: Assessing disaster impacts and response using social media data in China: A case study of 2016 Wuhan rainstorm, Int. J. Disaster Risk Reduct., 34, 275–282, doi:10.1016/j.ijdrr.2018.11.027, 2019.

Han, X. and Wang, J.: Using social media to mine and analyze public sentiment during a disaster: A case study of the 2018 Shouguang city flood in china, ISPRS Int. J. Geo-Information [online] Available from: https://www.mdpi.com/2220-9964/8/4/185, 2019.

Rainear, A. M., Lachlan, K. A., Oeldorf-Hirsch, A. and DeVoss, C. L.: Examining twitter content of state emergency management during Hurricane Joaquin, Commun. Res. Reports, 35(4), 325–334, doi:10.1080/08824096.2018.1503945, 2018a.

Sit, M. A., Koylu, C. and Demir, I.: Identifying disaster-related tweets and their semantic, spatial and temporal context using deep learning, natural language processing and spatial analysis: a case study of Hurricane Irma, Int. J. Digit. Earth, 12(11), 1205–1229, doi:10.1080/17538947.2018.1563219, 2019.

Stewart, M. C. and Gail Wilson, B.: The dynamic role of social media during Hurricane #Sandy: An introduction of the STREMII model to weather the storm of the crisis lifecycle, Comput. Human Behav., 54, 639–646, doi:10.1016/J.CHB.2015.07.009, 2016.

2. **Coverage of risk communication.** The reviewer raises two interesting questions of how social media is used for risk communication, and whether social media might have a

preventive impact on extreme weather disasters. While these are important and worthy of study, they lie beyond the scope of this paper, which focuses on the utility of social media as a means of *observing the impacts* of weather hazards. Given the already lengthy nature of the paper, we feel any additional commentary relating to risk communication would take away from the key focus of the paper - which seeks to explore the use of social media as a tool for impact information curation, rather than an analysis of how social media data is used during a disaster. However we thank the reviewer for their comments and advice on this area and will certainly bear this in mind for future studies.

3. **Analysis of risk perception.** Again, the reviewer has identified an interesting research topic. It is certainly possible that social media could be used to assess the public perception of the risks arising from extreme weather. But as with questions around risk communication and preventive effects, this is beyond the scope of our study. To make a proper assessment of risk perception would require establishment and validation of suitable metrics, collection of data, analysis, interpretation etc. Essentially, a whole other study. Therefore we do not feel able to give this subject the time and space it deserves within our current manuscript, where the focus lies elsewhere.

4. **Comparison with other crowdsourcing experiments on natural disasters.** We thank the reviewer for the suggestion of Bossu 2020 as a relevant citation and have included that work in our literature review. Our expanded literature review considers social sensing experiments for a number of natural phenomena, including floods, wildfires, wind and storms, helping to establish the context for our current work.

Reviewer #2:

GENERAL COMMENTS

The paper entitled "Social sensing of high-impact rainfall events worldwide: A benchmark comparison against manually curated impact" presents a valuable contribution to identify rainfall events generating flooding or landslides with tangible impacts on the population. Specifically, the Authors compared events identified with the proposed procedure based on the use of social media, with events included in an existing database (Met Office Global Hazard Map - GHM), verifying the related accuracy. Generally, the paper meets the scope of NHESS resulting well structured and presented, with a critical analysis of findings provided in the Discussion section. Despite this, I identified some points that the Authors should address with the aim of improving the paper. In my opinion, the several limitations affecting the use of Twitter data do not allow to use the proposed procedure at global scale, as well as many criticisms highlighted in the Discussion section should be limited since they depend on the research design decided by the authors. In addition, more references to similar studies should be included in the Introduction section and some changes should be perform to clarify both the methodology and results sections, as highlighted in the specific comments.

We thank the reviewer for their favourable summary of our work.

I response to the general comments:

- **Use of methods at a global scale**
  The reviewer makes an interesting point regarding the limitations of the study and its validity at a global scale. For the countries examined in the study, including those where English was not a native language, we did find good performance overall, despite a low volume of tweets in some places (Figure 12). For example, the tweet volume in Kenya was low, however we found a high F2 score. Furthermore, the study only focused on a 6 month period of 2017, therefore missing the 'rainy' season for some countries (e.g. India) and therefore this may have impacted performance there. Extending these methods over a longer timeframe should improve performance for these countries and therefore there would be more confidence in the use of the method globally.
  We have ensured these points have been added to the discussion to hopefully clarify this issue.
- **Literature review and related work**
  We have added a 'Related Work' section to the Introduction which provides a more in-depth literature review on similar studies:

  Related Work
  A number of studies have explored the use of social media as a source of information about the impacts of extreme weather. Social sensing is an approach developed in recent years to analyse unsolicited social media data to detect real-world events of interest.
  While social sensing is not specific to natural hazards and can be applied in a variety of contexts (Liu et al., 2015; Wang et al., 2012, 2019), social sensing has demonstrated usefulness for natural hazard events.

  Twitter data was used by Sakaki et al. (2010) to detect earthquakes in Japan, with reports arriving in some locations before the shock had been detected by conventional seismography. Many studies have followed, using a number of different approaches to explore the use of social media as an information source during and following natural hazard events. Some studies have focused on the use of social media to better understand risk communication during an extreme natural hazard event. For example, Steward and Wilson (2016) explore the use of social media throughout the crisis lifecycle during Hurricane Sandy in the USA, building the STREMII model to better understand crisis communication during an extreme weather event; Rainear et al. (2018) used Twitter data collected during Hurricane Joaquin to explore the types of information communicated by state emergency management accounts to better understand the flow of risk communication during a crisis; Bossu et al. (2020) explored the use of crowdsourced information, along with Twitter data, in a bespoke application during the 2019 earthquake in Albania, finding that engagement of users with the app provided much more information about the damage caused as a result of the earthquake than was available using conventional methods.

Other studies have explored the use of social media to better understand the impacts of extreme weather events. Many studies focus on individual events. For example Fang et al. (2019) use data from the Chinese social media platform, Sina Weibo, during the 2016 Beijing rainstorm, finding a positive correlation between social media activity and precipitation intensity; Sit et al. (2019) examine Twitter data collected during Hurricane Irma, using geo-located tweets to identify locations with a high density of affected individuals and infrastructure damage; and Han et al. (2019) use data from Sina Weibo during the 2018 Shouguang flood to analyse the changes in sentiment of social media users during the different development stages of the flood. Further examples of other studies examining the impacts of individual weather events at one particular location include: studies relating to specific hurricanes in the United States (Guan and Chen, 2014; Kim and Hastak, 2018; Lachlan et al., 2014; Morss et al., 2017; Niles et al., 2019; Wu and Cui, 2018; Zou et al., 2018) and specific flooding events (Aisha et al., 2015; Brouwer et al., 2017; Cervone et al., 2016; Kankanamge et al., 2020; Li et al., 2018; Rossi et al., 2018).

Some authors have begun to explore the use of Twitter for more wide-scale specific weather event detection, Arthur et al. (2018) use Twitter data to detect and locate flood events in the UK to produce maps of flood activity. De Brujin et al. (2019) compare Twitter activity relating to flooding and hydrological information with flood events in the NatCatSERVICE disaster database, finding a good comparison between these data sources. Boulton et al. (2016) use Twitter data collected during several time periods to detect and locate wildfires in the USA. Cowie et al. (2018) find that user reports on Twitter during the year can help to locate peaks in hayfever symptoms as a result of pollen levels in the UK. Furthermore, Spruce et al. (2020) examine Twitter data relating to named storms, wind and precipitation in the UK finding that it is possible to identify tweets which can be used to assess the impact of storms both temporally and spatially.

The following references will also be added as follows:

Bossu, R., Fallou, L., Landès, M., Roussel, F., Julien-Laferrière, S., Roch, J. and Steed, R.: Rapid Public Information and Situational Awareness After the November 26, 2019, Albania Earthquake: Lessons Learned From the LastQuake System, Front. Earth Sci., 8, doi:10.3389/feart.2020.00235, 2020.

Fang, J., Hu, J., Shi, X. and Zhao, L.: Assessing disaster impacts and response using social media data in China: A case study of 2016 Wuhan rainstorm, Int. J. Disaster Risk Reduct., 34, 275–282, doi:10.1016/j.ijdrr.2018.11.027, 2019.

Han, X. and Wang, J.: Using social media to mine and analyze public sentiment during a disaster: A case study of the 2018 Shouguang city flood in china, ISPRS Int. J. Geo-Information [online] Available from: https://www.mdpi.com/2220-9964/8/4/185, 2019.

Rainear, A. M., Lachlan, K. A., Oeldorf-Hirsch, A. and DeVoss, C. L.: Examining twitter content of state emergency management during Hurricane Joaquin, Commun. Res. Reports, 35(4), 325–334, doi:10.1080/08824096.2018.1503945, 2018a.

Sit, M. A., Koylu, C. and Demir, I.: Identifying disaster-related tweets and their semantic, spatial and temporal context using deep learning, natural language processing and spatial analysis: a case study of Hurricane Irma, Int. J. Digit. Earth, 12(11), 1205–1229, doi:10.1080/17538947.2018.1563219, 2019.

Stewart, M. C. and Gail Wilson, B.: The dynamic role of social media during Hurricane #Sandy: An introduction of the STREMII model to weather the storm of the crisis lifecycle, Comput. Human Behav., 54, 639–646, doi:10.1016/J.CHB.2015.07.009, 2016.

We will now address the specific comments made in turn.

SPECIFIC COMMENTS
INTRODUCTION

1) The use of social sensing for the analysis of impactful rainfall events does not represent an innovation since many studies already focused on this topic in the past years. I suggest to include more references about this, trying to cite also works based on other social platforms.

1. A 'Related Work' section has been added to the Introduction which provides a more in depth literature review with additional references to similar work, as detailed above.

2) Please, provide a brief description of the Met Office and its activities. In addition, I suggest to provide a comparison between the Met Office Global Hazard Map (GHM) and other available global databases that you might have used to test your procedure.

2. We have added the following to the Introduction as follows:
    a. *"The Met Office is the national meteorological service for the UK, providing weather services and contributing to climate science research worldwide (https://www.metoffice.gov.uk/about-us/who)."*
    b. *There are limited options available for other global databases containing weather impacts with which to compare our methodology against. There are databases such as NatCatSERVICE, produced to record insurance loss as a result of natural catastrophes. However we would like to consider impacts of extreme weather (i.e. disruption to daily life) which don't necessarily lead to financial loss which could be missing from this kind of record. ReliefWeb, which is a humanitarian information source on global crises and disasters, is another possible database from which to compare our results, however this is filtered for disaster events which are most relevant to global humanitarian workers and decision-makers, rather than all impactful events. Other available databases rely on citizen input (e.g. the European Severe Weather Database (ESWD)), may be*

limited to certain geographical areas, and are unlikely to contain the same level of rigour as the Community Impacts Database in terms of criteria for inclusion. Considering the options available to us, the Community Impacts Database therefore provides the most comprehensive database for comparing our methodology against.

METHODS

3) Line 88: In my knowledge, there are many countries where the use of Internet and, as consequence, of social media is not so widespread to highlight the occurrence of weather events. I suggest to take into account such consideration and modify this sentence accordingly.

3. Line 88 sentence has been modified as follows:
   "There are still some countries where use of the internet is not as widespread or where social media is limited to certain platforms. Despite this limitation, however, Twitter is still one of the most prevalent social media platforms across the world and therefore likely to be a good source of information for understanding where people are being affected by extreme weather, and how they are being impacted by it. "

4) Line 106: Please, specify the meaning of "unique events"

4. Line 106 sentence has been modified as follows:
   "The dataset used in this study contained 519 entries (135 unique events) in the period January-June 2017. Unique events refers to the fact that a single rainfall event can lead to impacts in multiple locations."

5) Line 122: Before in the manuscript, you stated that only tweets containing one or more keywords were downloaded. Here, you are stating that a filtering step was necessary to extract only those with one or more of the selected keywords. Please, clarify this inconsistency.

5. Line 122 sentence has been modified to explain this apparent inconsistency as follows:
   "The Twitter Streaming API searches the whole of the tweet metadata for the search terms requested in the search including tweet text, urls, and usernames. Therefore collected tweets were filtered to extract only those with one or more of the selected keywords in the tweet text and to remove any duplicate tweet IDs"

6) Generally, the paragraph 2.3.2 seems quite confused. In order to fix this issue, I suggest to list the several steps, trying to clarify their description.

6. Paragraph 2.3.2 has been amended to more of a bullet point format as follows:
   *"**2.3.2 Gazetteer look-up***
   This filter checked the tweet to determine if a discernible place name could be detected from the user location and/or the tweet text using gazetteers including Geonames (Geonames, 2020) and DBPedia (DBpedia, 2020). The following methodology was applied to each tweet which did not contain geo or place coordinates as described in 2.3 above:
   - *Geonames was used as our primary source of gazetted features as it is a geographical database with information about all countries with over eight million*

places, such as cities and points of interest. Where there was no match found in the Geonames database, the DBpedia database was used.

- *Where a match to a place name is found, a set of co-ordinates or bounding boxes from the gazetteer database is returned.*
- *Where locations were found in both the user profile and tweet text, place names in the tweet text are preferred as they are more likely to relate to the subject of the tweet.*
- *In a small number of cases, the user profile location and tweet text locations may differ; in that case, the place determined from the tweet text is given more weight during the location inference process.*
- *Where multiple matches to a place name were found in Geonames (i.e. where a place name exists in more than one country), then if there was no reference to the country elsewhere in the tweet or the country had not already been determined by the country filter described in 2.3.1 above, the place with the largest population (which has been found in previous studies to be the most likely location for the tweet (Schulz et al., 2013, Arthur et al., 2018)) was logged and coordinates returned.*
- *In addition, where multiple place names are determined from a tweet, to infer the most probable location, areas of overlap between the matching location polygons are detected before a final coordinate or bounding box is returned. This assumes that polygon overlaps are the highest likelihood locations.*
- *Since some place names are also commonly used to denote something other than a location (Liu et al., 2011), a database of words which are also places was used to remove apparent locations which were more likely to be a word than a place (e.g. dew, aka, var, etc).”*

7) Lines 227-228: Please, clarify the use of polygons and the overlapping concept

7. Lines 227-228 sentence has been modified to clarify the use of polygons and overlapping concept as follows:
   “In addition, where multiple place names are determined from a tweet, to infer the most probable location, areas of overlap between the matching location polygons are detected before a final coordinate or bounding box is returned. This assumes that polygon overlaps are the highest likelihood locations.”

8) Lines 228-230: I suggest to move this sentence at the end of the paragraph

8. Lines 228-230 - sentence has been moved to end of paragraph, as shown in 6. above.

9) Line 234: The term “likely” is repeated two times. Please, modify.

9. Line 234 - this sentence has been rewritten as described in 10 below.

10) Line 234: Provide a more convincing explanation on why you select the “place with the largest population”. Contents in the brackets are quite trivial.

10. Line 234 has been amended as follows:

“Where multiple matches to a place name were found in Geonames (i.e. where a place name exists in more than one country), then if there was no reference to the country elsewhere in the tweet or the country had not already been determined by the country filter described in 2.3.1 above, the place with the largest population (which has been found in previous studies to be the most likely location for the tweet (Schulz et al., 2013, Arthur et al., 2018)) was logged and coordinates returned.”

RESULTS

11) The results section is structured in too many parts. Given that this may confuse the reader, I suggest to join some sections trying to reduce the used Figures and text. Some Figures, for example, may be unified in multiple panels.

11. On review of the results section we agree that this section could be more streamlined. We suggest merging figures 7 and 8 and the related sections of text. Section 3.4 could also be reduced by removing figure 11 and merging figures 13 and 14; as well as the related text.

DISCUSSION

12) I believe that the core points of the work are the used methodology and validation strategy. In the light of this, it may be useful avoiding a lot of information about the comparison between social sensing and Met Office data, since the latters are affected by numerous biases that you have correctly highlighted. In addition, it is important to note that some limitations are due to the research design chosen by you (e.g. the analysed time frame). Therefore, I think that sentences related to these aspects should be limited.

12. We thank the reviewer for this observation. The Met Office database was laborious and time consuming to collect, however it is very useful because it pulls information from a wide range of sources; includes all events found, regardless of location in the world; and has clear and consistent criteria for events to be included within it. In this study, we have shown that Twitter data is a good source of data for event detection. Furthermore, the type of record that Twitter provides (i.e. eye-witness accounts, individual reports of events taking place), is different in nature to the aggregated sources that the Met Office database and other similar databases use. Therefore Twitter data can be used as a ‘first pass’ event detection tool, largely automating the difficult manual curation task. We have strengthened our point on this in the discussion to clarify this issue.

13) Taking into account the several limitations affecting the proposed procedure, I disagree to propose it as a valid methodology at global scale. Probably, the country scale represents the best one, allowing to limit many drawbacks such as the used language.

13. As detailed in our response to General comments, we have added some further points to the discussion to hopefully strengthen our argument for the method being suitable to be used at a global scale