Using Social Sensing to Validate Flood Risk Modelling in England

Joshua Joyce[1,2]([✉]), Rudy Arthur[1], Guangtao Fu[1], Alina Bialkowski[2],
and Hywel Williams[1]

[1] University of Exeter, Exeter EX4 4QE, UK
`jj479@exeter.ac.uk`
[2] University of Queensland, St Lucia, Brisbane, QLD 4072, Australia

Abstract. Floods are amongst the most severe natural disasters. Accurate flood risk maps are vital for emergency response operations and long-term flood defence planning. Currently the validation of such maps is often neglected and suffers from a lack of high-quality data. The proliferation of social media usage worldwide in recent years has supplied access to large amounts of data linked to flooding, and the detection of real-world events using such data is termed 'social sensing'. In this paper we investigate the use of social sensing for the validation of flood risk maps. We apply this methodology to 7 years' worth of flood related Tweets in order to perform a comparison to long term planning flood risk maps in England. The results show that there is a low level of correlation between the collection of socially sensed floods and high-risk flood areas as well as highlighting areas with high levels of socially sensed flooding that have low levels of flood risk, showcasing the potential importance of social media data for use in flood risk validation and planning policy.

Keywords: Social media · Flooding · Twitter

1 Introduction

Natural disasters are among the world's greatest challenges and 80,000 people per day are affected with an economic loss of US\$ 1.5 trillion since 2003. Flooding alone, which is the most frequent and wide-reaching weather-related natural hazards in the world [4], has affected 2.3 billion people with an estimated economic losses of US\$ 662 billion from 1995 to 2015, and US\$ 60 billion in 2016 alone [15]. On top of this, impacts of floods are projected to increase in the future due to climate change [13].

The generation of long term flood risk maps is then of extreme importance for planning procedures. Such maps are produced by utilising features such as terrain elevation, land use and meteorological data as parameters within physical models to estimate the flood extent of various simulated levels of rainfall events [7]. The maps output from this process are then used in urban planning for flood mitigation and defence [14].

The validation of such models is a key topic in flood modelling but as noted by Molinari et al. [8] in their review of existing practices in this area, 'Validation is perhaps the least practised activity in current flood risk research and flood risk assessment'. One major problem in the validation process is a lack of high quality data and when validation is performed often crowd-sourced data is used [10,12].

Social sensing is the use of unsolicited crowd-sourced information to observe real world events. This information can be gathered from a variety of different sources including web searches and social media. The major advantage of social media data over other crowd-sourcing avenues is the volume of data; social media platforms such as Twitter, Instagram and TikTok have millions of active users each month and millions of posts per day. This paper focuses on social sensing using Twitter due to the public accessibility of its data.

A variety of studies have been conducted in the topic of social sensing of floods. Arthur et al. [2] used tweet observations to produce flood maps of the UK validated against data of flood events provided by the Flood Forecasting Centre and concluded that social sensing can reproduce the validation data to a high accuracy, even finding flood events that were not contained in the validation dataset, albeit at the cost of false positives. Moore et al. [9] introduced a method for social sensing of coastal floods. They proposed to use a metric of remarkability of a high-tide event as a way to measure impact of coastal flooding where it is felt the most, as apposed to earlier methods which focus on high population areas due to ease of data collection. Individual regression models were built for counties along the east coast of the USA, building a relationship between number of geo-located tweets that day and maximum daily tide height measured at nearby tidal gauges including controlling for daily rainfall. Young et al. [18] utilised Twitter data as well as data from social media site Telegram in order to analyse the impacts of the 2018 floods in Kerala. They were able to analyse not only the extent of flood impacts but also the kind of impacts such as requests for help. The results showed good agreement with government created post flood database of damages. Ansell et al. [1] introduced a statistical approach involving the use of vine copulas, where they combined social media data including Twitter data, Google Trends data as well as average sentiment with environmental variables such as wave height and water level in order to predict inundation. Here they showed that performance of the model was improved by assuming a relationship between the social and environmental factors rather than assuming independence, showcasing that integration of social media data can produce more accurate forecasts.

Previous studies where social sensing of flooding has been used have focused mostly on the validation of the method to detect past flood events [2] or in post event analysis [18] and are often short term studies from a temporal standpoint. Only one study was found that utilised social sensing in term of flood risk, Brangbour et al. [5] utilised Twitter data to compute probabilities of rasterised grid cells being flooded during Hurricane Harvey. This was a study done using high quality, highly curated data for a single extremely severe event. On the other hand the main contribution, and unique goal, of this paper is to perform

social sensing of floods using a much greater time period of data, collecting data on all types of flooding over a country wide area in order to compare this analysis with long term flood risk models. As social sensing of floods by its very nature observes the impact of floods on a societal and human level, this report seeks to discover the relationship between areas considered a high risk in flood defence and mitigation planning and areas of high social 'floodiness', and in doing so investigate the potential of social sensing as a useful tool in the key area of flood risk model validation.

The main contributions of this paper are:

– The first long term study of socially sensing of floods via the creation of a dataset consisting of 7 years of geolocated relevant tweets.
– The first study to investigate the relationship between socially sensed flooding and flood risk models on a national scale.

The structure of the paper is as follows. Section 2 describes the data sources used, the methodology for various filtering techniques used to curate the Twitter data as well as the method used to infer locations from tweets. In Sect. 3 the results of the analysis are presented. Section 4 contains a discussion of these results. Finally Sect. 5 presents the conclusions of the paper.

2 Methodology

2.1 Data Collection

Twitter Dataset. Tweets were collected using Twitter's Streaming API and searching for the terms "flood", "flooding" and "flooded" as a basic first filter. It's important to note that this API was accessed using Twitter's Academic track which has significantly increased in price as Twitter have changed their data policies. The API returns tweets in the form of a JSON object which consists of key-value pairs for various metadata such as tweet text, user profile information and user location. In total 160,424,089 tweets were collected between the dates of 22/10/2015 and 11/04/2022. Due to collection issues there are gaps between 28/12/2015 and 04/01/2016 as well as between the dates 26/11/2016 and 02/01/2017 and 17/11/2021 and 10/01/2022.

Flood Maps. Recent flood maps (early 2022) produced by the Environment Agency[1] were used for comparison with the Twitter dataset. These maps are produced using physical modelling methods and separate maps are produced considering different types of flooding. The first of these is called 'Risk of Flooding from Rivers and Sea'. This map consists of 50 m × 50 m gridded areas of England with the likelihood of flooding from rivers and the sea presented in four different categories; namely Very Low, Low, Medium and High, whilst also taking account of flood defences and the condition they are in. High risk flood

---

[1] https://data.gov.uk/.

areas, which refers to a 1 in 30 annual probability of flooding, were chosen for use in this study providing the best comparison to the nearly 7 years worth of Twitter data. The map for the category of high risk can be seen in Fig. 1a.

As well as maps based on river and coastal flooding the Environment Agency also produce maps for surface water flood risk. High spatial resolution maps of this type are unavailable for download and are restricted to tiles, or small grid squares, of England due to the large complexity of these maps. Instead, so called 'Indicative Flood Risk' maps are available where the modelled data is aggregated to 1 km square grids based on 1 in 100 annual probability of flooding as well as minimum thresholds for either area population (200 people per 1 km grid) or critical services (at least one per 1 km square grid) or number of non-residential buildings at risk (at least 20 per 1 km square). The produced grids can be seen in Fig. 1b.
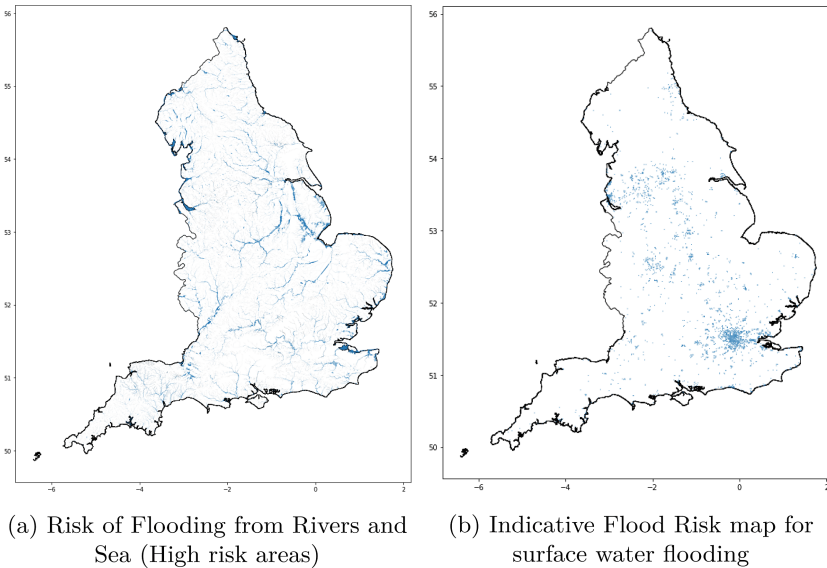

(a) Risk of Flooding from Rivers and Sea (High risk areas)

(b) Indicative Flood Risk map for surface water flooding

Fig. 1. Environment Agency flood maps

2.2   Twitter Data Pre-processing

As previously mentioned an initial filter was applied to tweets as they were collected. Several different filters were then applied post collection to remove irrelevant data as follows:

Retweet and Quote Tweet Filtering. One feature of Twitter is the ability to retweet and quote tweets. This is done to promote the tweet and to increase the likelihood of the tweet being seen by other users. As these are not original

and independent flood observations by users all retweets and quote tweets are removed. Sometimes people type 'RT' at the beginning of a tweet to indicate that they are reposting someone else's content instead of using the retweet function and these were also removed.

Bot Filtering. A number of accounts on Twitter are automated bot accounts and in the context of floods a large number of tweets that pass the initial top level filter are from weather stations which tweet out a large amount of flood related information. As we are interested in socially sensed flood events these tweets add a large amount of noise to the data and are removed. In total around 100 accounts are identified as bots and are removed from the dataset, including accounts such as @RiverLevelsUK @ukfloodtweets and @ShropshirePulse. Weather stations in general are removed by searching for a large number of keywords within tweets such as "north", "south", "rain" and "wind" and also units such as "mm" and "m/s". If the number of keyword matches is greater than a threshold then the tweet is removed.

Language Filtering. As the focus of this study is England and for use with geographical databases used in future steps all non English tweets are removed by using the "lang" key within the tweet JSON.

Relevance Filtering. Even after the previous filtering steps there remain a large number of tweets containing the top-level keyword terms in irrelevant contexts. Examples of this include phrases such as 'flooded with' or 'flood of'. A number of manually curated terms such as these were created and tweets containing these terms were filtered out.

Next, tweets were manually tagged as relevant or not relevant where relevant in this case refers to a tweet about an immediate flood situation such as "flooding in Exeter right now" as opposed to tweets about historic flood events or flood warnings which were tagged as not relevant. In total 4524 tweets were tagged with 1733 tagged as relevant and 2791 tagged as not relevant.

Using these tagged tweets as training data, a Multinomial Naive Bayes classifier was built and the tagged dataset of tweets was split into training and validation sets. 75% of the data was used to train the models and 25% was used for validation. Tweets were cleaned to remove stop words, URLs and punctuation. Tweets were tokenized, stemmed and lemmatized. A Bag of Words technique was used and the data was vectorised by counting the number of single word and two word occurrences in the corpus. Overall the Naive Bayes model achieved an F1-score of 0.84 with a Precision of 0.79 and a Recall of 0.83 indicating good overall classification.

Location Inference. Only a very small amount of tweets contain GPS data (less than 1%) [6]. For the purpose of creating accurate flood maps based on this data, it is important to be able to accurately infer locations from tweet metadata.

If a tweet contains an exact GPS tag then that latitude/longitude pair is used to map the tweet. For every other tweet a location inference heuristic is applied which is based on [11].

Table 1. Tweets remaining after each processing step

Filtering stage	Tweets
Top-level	160,424,089
Relevance	2,193,999
Location inference	408,172
Large polygon removal	165,663

In order to validate the performance of the location inference heuristic comparisons were made to a subset of the filtered tweets with exact GPS coordinates. In total, there were 71,688 tweets with GPS coordinates. Location polygons were inferred using the heuristic method with GPS metadata specifically ignored. Using this, a parameter grid search was performed using a range of different gazetteer database weightings as well as indicator weightings. The displacement between inferred locations and actual locations was calculated in kilometres and a tweet was considered correctly classified if the displacement was lower than 10 km. The best performing set of parameters was shown to be all indicator weightings set to 1 except tweet text which was set to 2. The displacement in kilometres between the inferred location and the actual location for this parameter is shown in Fig. 2. It can be seen that the method performs very well albeit with some large displacements shown for a number of tweets. The total tweets retained after each processing step can be seen in Table 1. Overall, just over 165,000 tweets were retained which were then used to create the socially sensed flood maps.
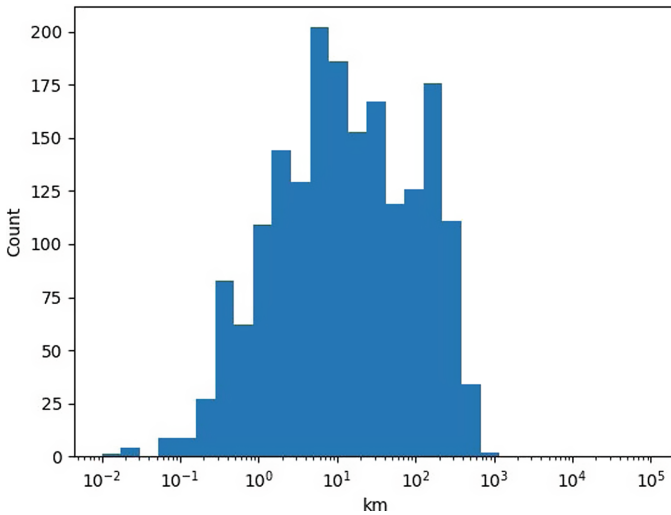

Fig. 2. Displacement of inferred location polygons from true geotagged coordinates

2.3   Flood Map Development

Socially Sensed Flooding Maps. In order to produce maps containing all remaining tweets, we first start with a bounding box of England and discretise it. Based on the results seen in Fig. 2 it was decided to create grid squares of 10 km by 10 km, as this achieves a good balance between accuracy and granularity. Each grid square starts with a weight of 0, $g_W = 0$, and is incremented for each tweet that falls within the grid square:

$$g_W = g_W + \frac{Area_{g \cap p}}{Area_p},\tag{1}$$

where g is the grid square and p is the tweet polygon. When the tweet has a precise location, p is a point and a score of 1 is added to the weight of the grid square, otherwise the proportion of overlap is added. This enables tweets with precise locations to be more influential for the detection of flooding.

The next step is to account for population density as large cities will have vastly more tweets associated with them than small towns. To this end, population data is taken from Lower Layer Super Output Areas (LSOAs), a geographic hierarchy for small area statistics. LSOAs are population areas of at least 1000 people and are designed to be consistent in population size. As a result, LSOAs within cities are much smaller than their counterparts in the countryside. The proportion of overlap between grid squares and their intersecting LSOAs is calculated and the corresponding proportion of the LSOA population is added to the grid square. This population data is taken from the 2011 Census so is somewhat out of date and taking proportions makes the assumption of uniform population across the LSOA which is not necessarily true but as the LSOAs are small enough it provides a reasonable estimate.

We then rescale the grid weights as follows,

$$g_W = \frac{g_W}{g_P{}^\alpha}\tag{2}$$

where g_P is the calculated population for the grid square and α is a scaling factor between 0 and 1. The factor α is necessary as it has been found that there is an imbalance between the number of twitter users in cities and rural areas [3]. A larger value of α will result in the population of the grid square having a larger effect on the weighting, meaning flooding detected in less populated rural areas will be more pronounced. For the purpose of this study α is set to 0.4 as this was found to have the best balancing effect.

Flood Risk Maps. In order to perform direct comparisons between our produced socially sensed flood maps and the flood risk maps produced by the Environment Agency it is necessary to have each type of map at the same spatial resolution. To this end, the Environment Agency flood risk maps are aggregated up to the same grid system as the socially sensed flood maps. In order to do this, for each grid square its intersecting flood risk polygons are obtained. The area of intersection of each polygon with the grid square is then calculated and

summed. This is then divided by the area of the grid square in order to obtain the proportion of the grid square that is consider under risk of flooding.

3   Results

Maps produced from the aggregated Environment Agency flood risk maps can be seen in Fig. 3b for river and seas flooding and in Fig. 3c for surface water flooding. Figure 3a shows the population weighted flood maps based on the fully filtered tweet dataset.

For statistical comparison between the aggregated flood risk maps and the produced socially sensed flood maps, correlation between grid squares was calculated as well as the use of simple linear models. The correlation between grid squares was calculated using Spearman's rank correlation coefficient and is shown in Table 2. Overall correlation between river and coastal flood risk and social floodiness was $r = 0.27$ with a p-value of $1e^{-26}$ and for surface water flood risk was $r = 0.41$ with a p-value of $7e^{-61}$. Scatter plots of tweet weighting against aggregated flood risk can be seen in Fig. 4. The equations for the lines fitted are $y = 4.75x + 1.3$ and $y = 11.4x + 1.2$ for river and coastal flood risk and surface water flood risk respectively. R^2 values were 0.01 and 0.1 respectively showing no linear relationships. Overall the results show a moderate level of correlation between the socially sensed flood map and Environment Agency produced maps, particularly with regards to surface water flooding.

Of particular note, the scatter plots show a number of outlier areas which have little to no modelled flood risk that have a high weighting for socially sensed flooding especially in the case of surface water flooding. Indeed by re-scaling the socially sensed grid weights between 0 and 1 by dividing through by the max of all grid square weights, we can calculate the difference between flood risk and social floodiness.

Figure 5 shows these differences with values between (0,1) indicating higher flood risk than social floodiness and values between (−1,0) indicating lower flood risk than social floodiness. Figure 5a compares the socially sensed response to flood risk from coastal and river flooding. In the north east we see that we have higher flood risk associated with rivers flowing into the Wash and Humber estuaries, as well as for the coast of East Anglia than we would predict from observing floods on Twitter. In contrast we see a much higher Twitter signal than the corresponding risk would predict in the north west (Cumbria) and far south west (Plymouth), likely due to major flood events which occurred during the data collection. Figure 5b which compares against surface flooding shows similar under-estimation in the north- and south-west with over-estimation in Greater London (south east). In general outside of these outliers it can be seen in both maps that flood risk is slightly underestimated against socially sensed flooding.
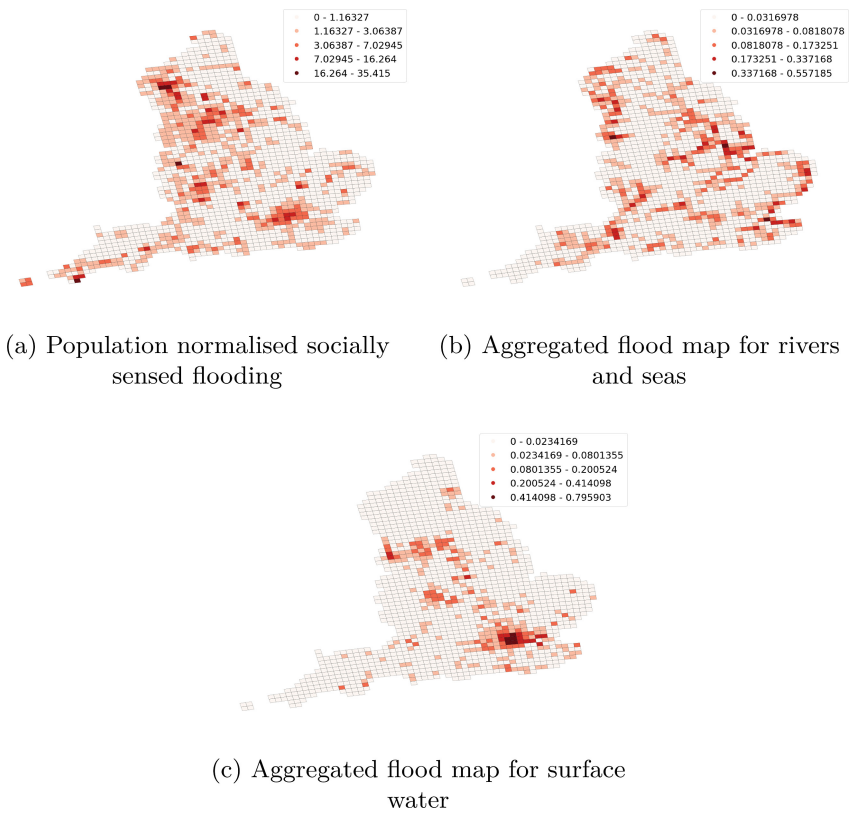
(a) Population normalised socially
sensed flooding

(b) Aggregated flood map for rivers
and seas

(c) Aggregated flood map for surface
water

Fig. 3. Aggregated flood maps

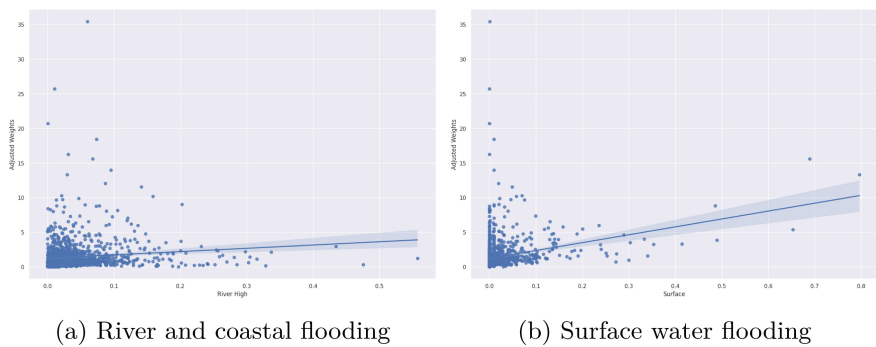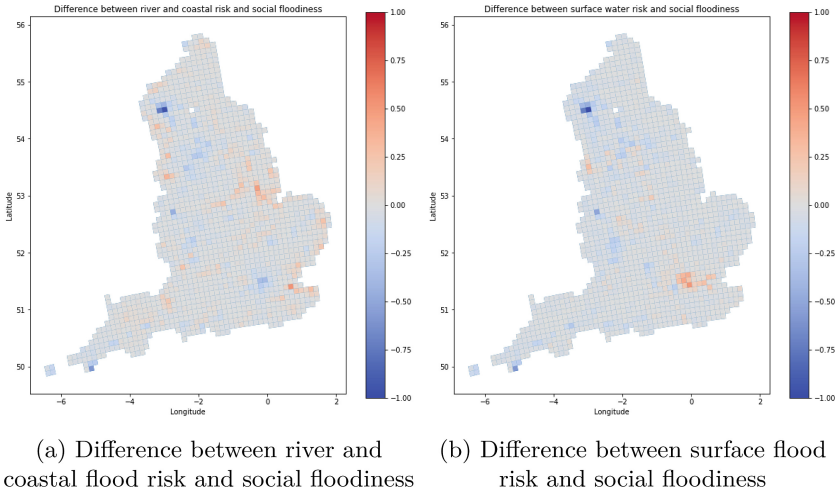(a) River and coastal flooding

(b) Surface water flooding

Fig. 4. Scatter plots of each flood map against socially sensed flooding

Table 2. Spearman's rank correlation of the socially sensed flood map against specified maps

Flood Map	Spearman's rank correlation
Rivers and coastal	0.27
Surface water	0.41

(a) Difference between river and coastal flood risk and social floodiness

(b) Difference between surface flood risk and social floodiness

Fig. 5. Maps indicating the difference between flood risk and social floodiness for each type of flood risk, red indicates grids with higher risk than social floodiness, blue indicates grids with lower risk than social floodiness. (Color figure online)

4   Discussion

From the results obtained we have found that there is a low to moderate correlation between socially sensed flooding and the high risk flood map for rivers and seas and that there is a moderate correlation between socially sensed flooding and surface water flooding showing that overall there is a modest level of agreement between socially sensed flooding and flood risk maps.

While 7 years is quite a long period for social media, it is not necessarily a long period for flood risk, which commonly predicts 1 in 10 to 1 in 100 year events. However, the fact that we observe significant over and under-predictions is notable. For example, historic flood events in the north west (Cumbria and Lancashire) during the data collection period produce a Twitter signal far in excess of the predicted risk. This occurs because the social response to a flood is highly non-linear [2,3], doubling the size of the flood generates much more than 2-fold increase in tweets. If tweets are taken as a rough proxy for impact, this implies that risk models which aim to predict not just the probability of occurrence of a flood, but also its potential impact, should incorporate a non-linear scaling of impact with event size.

There are a large amount of grid squares which have extremely low tweet weighting which could indicate that these areas do not suffer from flooding or that there is a demographic bias in some areas due to the fact that Twitter users are unrepresentative of the population as a whole. Further work could be done to explore this hypothesis. Related to this is the idea of the 'remarkability' of a flood. For example, an area which floods often may not produce tweets as this is considered a normality. This could be checked using historical data for severity and amount of flooding over a long term period, techniques like this were utilised in [9,16] to great effect and could be used in further work to improve the social floodiness response.

The most important step in the process of producing socially sensed flood maps is the initial filtering stage. Social media data is inherently noisy and improved relevance filtering would lead to a much less noisy dataset and a more accurate representation of flood events. The model used in this report is a classic classifier method used in this area but recent transformer models such as BERT as well as LSTM based models have been shown to perform well [17].

As well as this, location inference is a necessity due to the low level of geo-tagged tweets. Therefore more accurate location inference will lead to more accurate flood detection which is crucial for any validation of this method. Currently toponym recognition is limited as it only searches for proper nouns, and methods could be developed which expand this. Limitations also exist with the use of geo-databases, with toponym resolution being limited to nonexistent at below town level meaning potentially useful fine grained data is discarded. Improvements in this area would also allow future work to be expanded to smaller regional flood risk maps such as cities.

5   Conclusion

We have shown that it is possible to produce long term socially sensed flood maps that can be used to form a comparison - and potential validation tool - to long term flood maps. As socially sensed flood detection by nature detects floods which affect people, the results show not only the potential use of social sensing as a new data collection tool to validate flood risk maps compared to traditional validation methods which take time and may require a large workforce to manually collect and synthesise flood observation data, but also highlights a need for models to better take into account the impact of floods. As social media use continues to grow into the future, the growth of quality observation data that can be obtained from it will only improve its usability in this area.

References

1. Ansell, L., Valle, L.D.: Social media integration of flood data: a vine copula-based approach. arXiv preprint arXiv:2104.01869 (2021)
2. Arthur, R., Boulton, C.A., Shotton, H., Williams, H.T.: Social sensing of floods in the UK. PLoS ONE **13**(1), e0189327 (2018)

3. Arthur, R., Williams, H.T.: Scaling laws in geo-located twitter data. PLoS ONE **14**(7), e0218454 (2019)
4. Berz, G.: Flood disasters: lessons from the past-worries for the future. In: Proceedings of the Institution of Civil Engineers-Water and Maritime Engineering, pp. 3–8. Thomas Telford Ltd (2000)
5. Brangbour, E., et al.: Computing flood probabilities using twitter: application to the houston urban area during harvey. arXiv preprint arXiv:2012.03731 (2020)
6. Cheng, Z., Caverlee, J., Lee, K.: You are where you tweet: a content-based approach to geo-locating twitter users. In: Proceedings of the 19th ACM International Conference on Information and Knowledge Management, pp. 759–768 (2010)
7. Mignot, E., Paquier, A., Haider, S.: Modeling floods in a dense urban area using 2D shallow water equations. J. Hydrol. **327**(1–2), 186–199 (2006)
8. Molinari, D., De Bruijn, K.M., Castillo-Rodríguez, J.T., Aronica, G.T., Bouwer, L.M.: Validation of flood risk models: current practice and possible improvements. Int. J. Disaster Risk Reduction **33**, 441–448 (2019)
9. Moore, F.C., Obradovich, N.: Using remarkability to define coastal flooding thresholds. Nat. Commun. **11**(1), 1–8 (2020)
10. Schnebele, E., Cervone, G., Waters, N.: Road assessment after flood events using non-authoritative data. Nat. Hazards Earth Syst. Scie. **14**(4), 1007–1015 (2014). https://doi.org/10.5194/nhess-14-1007-2014. https://nhess.copernicus.org/articles/14/1007/2014/
11. Schulz, A., Hadjakos, A., Paulheim, H., Nachtwey, J., Mühlhäuser, M.: A multi-indicator approach for geolocalization of tweets. In: Seventh International AAAI Conference on Weblogs and Social Media (2013)
12. See, L.: A review of citizen science and crowdsourcing in applications of pluvial flooding. Front. Earth Sci. **7**, 44 (2019)
13. Tellman, B., et al.: Satellite imaging reveals increased proportion of population exposed to floods. Nature **596**(7870), 80–86 (2021)
14. Tsakiris, G.: Flood risk assessment: concepts, modelling, applications. Nat. Hazards Earth Syst. Sci. **14**(5), 1361–1369 (2014). https://doi.org/10.5194/nhess-14-1361-2014, https://nhess.copernicus.org/articles/14/1361/2014/
15. Unisdr, C., et al.: The human cost of natural disasters: a global perspective (2015)
16. Weaver, I.S., Williams, H.T., Arthur, R.: A social beaufort scale to detect high winds using language in social media posts. Sci. Rep. **11**(1), 3647 (2021)
17. Yaseen, Q., et al.: Spam email detection using deep learning techniques. Procedia Comput. Sci. **184**, 853–858 (2021)
18. Young, J.C., Arthur, R., Spruce, M., Williams, H.T.: Social sensing of flood impacts in India: a case study of Kerala 2018. Int. J. Disaster Risk Reduction **74**, 102908 (2022)