Probabilistic Flood Extent Estimates from Social Media Flood Observations

Tom Brouwer[1,2], Dirk Eilander[1], Arnejan van Loenen[1], Martijn J. Booij[2], Kathelijne M. Wijnberg[2], Jan S. Verkade[1], Jurjen Wagemaker[3]

[1]Deltares, Delft, Boussinesqweg 1, 2629 HV, The Netherlands
[2]Dept. of Water Engineering and Management, University of Twente, Enschede, Drienerlolaan 5, 7522NB, The Netherlands
[3]FloodTags, The Hague, Binckhorstlaan 36, 2511 BE, The Netherlands

Correspondence to: Dirk Eilander (Dirk.Eilander@deltares.nl)

Abstract. The increasing number and severity of floods, driven by phenomena such as urbanization, deforestation, subsidence and climate change, creates a growing need for accurate and timely flood maps. In this paper we present and evaluate a method to create deterministic and probabilistic flood maps from Twitter messages that mention locations of flooding. A deterministic flood map created for the December 2015 flood in the city of York (UK) showed good performance ($F^{(2)}$ = 0.69; a statistic ranging from 0-1, with 1 expressing a perfect fit with validation data). The probabilistic flood maps we created showed that, in the York case study, the uncertainty in flood extent was mainly induced by errors in the precise locations of flood observations as derived from Twitter data. Errors in the terrain elevation data or in the parameters of the applied algorithm contributed less to flood extent uncertainty. Although these maps tended to overestimate the actual probability of flooding, they gave a reasonable representation of flood extent uncertainty in the area. This study illustrates that inherently uncertain data from social media can be used to derive information about flooding.

Keywords: User generated content, Social media, Twitter, Flood extent estimation, Uncertainty analysis, Probabilistic flood mapping

1 Introduction

Between 1995 and 2015 2.3 billion people were affected by floods (UN, 2015), which is about one third of the world's population. Worldwide developments such as urbanization, deforestation, subsidence and climate change are expected to increase the occurrence of floods and number of people affected by them. This creates a growing need for timely and accurate information about the locations and severity of flooding. In multiple phases of the disaster management cycle (Carter, 2008), this information is useful. In the mitigation phase, data about previous flood events can be used to evaluate the probability of flooding and prevent urban expansion into flood prone areas. If flood prone areas are already inhabited, information about flood risk can also be used to improve disaster preparedness. In the response phase, information about the current flood situation is useful, for example for rescue workers who want to identify affected areas and assess the accessibility of roads.

Finally, in the recovery phase, flood information can help insurance companies in evaluating flood damages and aid organizations in targeting rebuilding efforts.

Traditionally, flood information in the form of flood maps has been produced using either hydraulic models or remote sensing. Applying these in real-time however, may be problematic. Hydraulic models require a detailed schematization of the study area, knowledge about the cause of a flood and may require considerable computational time. Also forecasts of input data, such as discharge or precipitation may not be readily available. Remotely sensed data may take several hours to become available (Mason et al., 2012) and its temporal resolution is often limited (Schumann et al., 2009).

Data created by users of online platforms such as blogs, wikis and social media, often referred to as 'user generated content', offers an additional source of information about natural disasters. Recent studies focused specifically on using social media content, since platforms such as Twitter, Facebook and Flickr produce large amounts of real-time data. In coarse scale applications for example, these data can be used to detect the occurrence of a natural disaster (Earle et al., 2011). On a more detailed level these data have also been used to assess the geographic extent of a disaster. In the context of flood mapping, some investigations used these data as auxiliary data. Examples include the assessment of the accuracy of remote sensing derived flood maps using Flickr data (Sun et al., 2015) and the selection of the most realistic result of a series of hydraulic model runs based on Twitter data (Smith et al., 2015). Others actually created flood maps directly from the data. Schnebele et al. (2014) used the density of flood related Twitter messages (Tweets) to get an indication of flood extent and in the PetaJakarta project, the number of Tweets in an area is used to indicate flood severity (Holderness & Turpin, 2015). Fohringer et al. (2015) created flood maps by interpolating water levels which were manually derived from photographs on Flickr and Twitter. Eilander et al. (2016) on the other hand used an automatic method to derive water depths and locations from Tweets, and created flood maps using a flood fill algorithm. To our knowledge, no flood-related studies have used data from Facebook until now, which is likely due to Facebook being a more closed network. Flickr and Twitter allow for all public data to be found and extracted using their 'Application Programming Interfaces' (APIs; Interfaces to extract data from online platforms). The Facebook API however, is much more restrictive and cannot be used to retrieve large amounts of public data.

The aforementioned studies all focused on obtaining flood extents from social media content. These flood extents, however, did not contain information about uncertainty, even though uncertainty is an inherent characteristic of information derived from social media content. Locational information of Tweets for example can be uncertain, because geo-tags are available for only a very small number of Tweets and may deviate from the actual location of the observation (Hahmann et al., 2014). McClanahan & Gokhale (2015), who derived locations from the text in Tweets, indicate that the locations they derived from messages in New York City had an average error of 1.72 km. Eilander et al. (2016) were the first to give an estimate of the likelihood of areas being flooded, by harvesting Tweets. This likelihood was based on the number of Tweets found for individual administrative areas, rather than knowledge about the actual errors in the data used.

Information about uncertainties can help in assessing the quality of generated flood maps. In addition, it can also serve as an information source of its own. This was for example the case in the search for Air France flight 447, which disappeared over the Atlantic Ocean in 2009. Probabilistic maps of the location of the wreckage were successfully used to find the wreckage in 2011, while previous attempts, spanning a two-year period, all failed (Stone et al., 2014). More specific to flood mapping, information about uncertainties can be used to direct surveys to areas in which the flood extent is highly uncertain. The information can also be used by rescue workers navigating an affected area, to choose the most optimal route by weighing the length of a route against the probability of it being flooded.

In the present paper we investigate the applicability of using social media content from Twitter to generate probabilistic flood maps. We explicitly address uncertainties in the data and assess the added value of probabilistic maps over deterministic maps. The analyses presented in this paper give insight into the magnitude of errors in flood observations derived from Tweets and improve understanding of how these errors affect the flood extent estimates. Furthermore, we investigate how the uncertainty caused by errors in the Twitter data relates to the uncertainty caused by other sources of error.

This paper starts with a description of the case study (Sect. 2). This is followed by an overview of the methods and data used (Sect. 3). Section 4 subsequently presents the research results and Sect. 5 includes an in depth discussion thereof. Finally, the conclusions of the research are given in Sect. 6.

2 Case study

On the 27th of December 2015 peak water levels on the river Ouse, caused by large amounts of rainfall, led to the flooding of a considerable area within the City of York in the North of England. Up to 120 mm of rain fell in Yorkshire over a 48 hr period between the 25th and 27th of December (Met Office, 2016). These large rainfall amounts resulted in the flooding of York and other places in the north of England. Within York, 453 residences as well as 174 businesses were flooded (Pidd, 2016). Detailed information about damages within York is yet to be published, since one year after the floods a report by the York City Council is still being written.

The flooding of the City of York in December 2015 was selected as a case study, since both high resolution terrain elevation data, as well as recorded flood extents were available for this event. In Fig. 1 an Environment Agency (EA) digital terrain model (DTM) of the city of York is given (EA, 2014). This paper focusses on Central York, delineated by the central administrative areas of York. In most of the study area terrain slopes are moderate, although some higher ridges are found in the south and south-west of the area. The inner-city of York is located to the north of these ridges. At this location there is a confluence of several rivers, of which the River Ouse is the largest.

3 Data and Method

Data from Twitter were used to derive flood information, since these data are openly and freely available. The research was comprised of several phases. First we extracted locations where flooding was observed from flood-related Tweets. This information was subsequently used to create a deterministic flood extent estimate. After this, based on information about the magnitude of errors in the data, probabilistic flood extent estimates were derived. This section starts with a discussion of the datasets used in this study. In the subsequent paragraphs, each of the three phases of the research is discussed separately. We conclude this chapter by explaining the methods we used to evaluate both the deterministic and probabilistic flood extent estimates.

3.1 Data

An overview of the data we used in this study is provided in Table 1. Elevation data was used throughout all phases of the research. We used a 2 m DTM, which is disseminated by the EA and has a vertical accuracy of ±15 cm (EA, 2016). To reduce the computational time required to create the flood maps, but preserve sufficient detail, we resampled this data to 20 m resolution using the average of the underlying data.

Tweets from between the 25th and 30th of December 2015, were collected using the Twitter streaming API. From these data flood observations were extracted. Google Maps and Google StreetView were used to find locations mentioned by the Tweets and the locations of photographs attached to the Tweets respectively (Sect. 3.2). From OpenStreetMap we downloaded the line-data belonging to the street names that Tweets mentioned, which were used to simulate locational errors along the streets (Sect. 3.4).

Recorded flood extents were used to validate the flood maps (Sect 3.5). A draft version of the fluvial flood extents of the City of York was supplied by the EA. These flood extents only identified areas that were directly affected by flooding from the rivers. However, areas separated from the river around Knavemire Road, Water Lane and Shipton Road were also known to be flooded based on news articles. The flood extents around these locations were approximated by using the EA dataset of recorded flood extents (EA, 2015) from between the years 1991 – 2012. These were merged with the recorded fluvial flood extent from 2016 into one validation dataset.

3.2 Twitter data extraction

The process used to create a database of Twitter based flood observations consisted of several steps (Fig. 2). First of all, we collected all Tweets that contained a number of common flood-related keywords such as 'flood' or 'inundation'. To ensure purely Tweets regarding York were found, only Tweets that mentioned 'York' or '#YorkFloods' were included and messages referring to 'New York' or 'York County' (both in the USA) were excluded. As a last selection step, we only kept Tweets that

contained explicit references to locations, such as streets or points of interest (POIs), by looking for common keywords such as 'street', 'lane', 'museum' or 'school'. Some other minor filters were applied to ensure only relevant Tweets were found, for example by excluding Tweets related to flood barriers and flood warnings.

5   We derived locations from the Tweets in the remaining dataset by manually identifying the section of the Tweet that contained a locational reference. Based on this reference, X and Y coordinates were assigned to the Tweets. To illustrate this process, the following Tweet is used as an example:

"Cumberland Street in York - say they're used to flooding here but only 2000 was worse" (By: @jimtaylor1984)

The locational reference in this Tweet is *"Cumberland Street"*. These locational references were used directly to search Google Maps. The message above however did not refer to a point location, which would be the case for a POI, but to a line element. We derived exact spatial coordinates from such Tweets by using the location of the street from Google Maps in combination with the DTM of the EA (EA, 2014). If topographical depressions were found along the street, the deepest depression,

15   identified by filling the sinks in the DTM, was used as the location of the observation. In case no depressions were found, the point of lowest elevation was used, which was the case for the Tweet in Fig. 3.

To review the accuracy of the spatial coordinates derived from the Tweets, we looked at the photographs attached to some of the Tweets and compared them to Google StreetView. If we found the specific location of a photograph, we compared it to the

20   spatial coordinates derived from the Tweets text to determine the locational error (example: Fig. 3).

3.3 Flood extent mapping

Flood extents were derived using the locations derived from the Twitter messages and the DTM. We applied an interpolation method to derive flood maps from the observations, similarly to Fohringer et al. (2015). Before interpolation, two processing steps were applied to guarantee more realistic flood extent maps. Firstly, we derived water levels relative to the nearest drainage

25   channel from each observation. Since none of the Tweets about York mentioned water depths, water levels were derived by assuming the same default water depth (DWD) for all observations. We calculated these water levels relative to the nearest drainage channel by using a height above nearest drainage (HAND) elevation model (Rennó et al., 2008; Norbre et al., 2011). Secondly, observations were grouped based on the local drainage directions (LDDs) to interpolate only hydrologically 'connected' observations. We used inverse distance weighting (IDW) interpolation to determine the flood extent. Figure 4 gives

30   an overview of this process. These steps are further explained in this section.

Norbre et al. (2016) applied the HAND concept to derive inundation extents for fluvial floods. In contrast to a DTM, which contains elevation values relative to one single reference level, such as mean sea level, elevation values in a HAND map are

relative to the nearest drainage channel. This drainage-normalized representation of the topography has a clear advantage for riverine flood extent mapping, as water depths over land can easily be related to water levels in the river. By using a HAND map instead of a DTM, river slopes are filtered from the dataset. This means that HAND values in an area are directly related to river stage and HAND contour lines describe the flood extent at a specific river stage. Since river slopes are filtered, upstream

5 observations are also less likely to cause overestimations of water levels downstream. We constructed a HAND map from the DTM by deriving the LDDs and using these to determine the elevation value of each grid cell in the study area, relative to the nearest drainage channel. Grid cells were identified as being on a drainage channel if they had an upstream area of 6 km^2 or more, which gave the best representation of drainage channels in the study area. Topographical depressions were filtered to derive the LDDs used to construct the HAND map. To also account for pluvial flooding of local topographical depressions,

10 these depressions were reintroduced in the final HAND map. This map was then used to translate the DWD assigned to each observation to a water level with respect to the nearest drainage channel.

We assumed that the water levels of flooded areas that are separated, are independent of each other. Therefore, we grouped observations to identify to which flooded area each observation belonged. The water levels of each group of observations were

15 then interpolated separately. We grouped observations by combining information about the LDDs in the area, which were derived from the DTM, with the locations of observations. The LDDs were used to determine which cells are downstream of an observation. If the location of the observation is flooded, it is assumed its downstream cells are also flooded, since these are located lower than the observation and are directly connected to it. Therefore, all observations that have downstream cells in common, are located within the same continuously flooded area.

The water levels (relative to the nearest drainage channel) of each group of observations were subsequently interpolated using IDW interpolation as given by Eq. (1) and Eq. (2):

$$Z_{x,y} = \frac{\sum_{i=1}^{n} Z_i * W_i}{\sum_{i=1}^{n} W_i}, \tag{1}$$

25

$$W_i = \frac{1}{(d_{x,y,i}+s)^p}, \tag{2}$$

Where $Z_{x,y}$ (m) is the interpolated water level at spatial coordinates x and y, Z_i (m) is the observed water level of observation i, n is the total number of observations, W_i is the interpolation weight of observation i, $d_{x,y,i}$ (m) is the distance to observation

30 i measured along the flow paths downstream of observations, s (m) is the smoothing parameter and p is the power parameter.

Previous studies applied both IDW (Werner, 2001) and bilinear spline interpolation (Fohringer et al., 2015) to calculate flood extents from irregularly spaced flood observations. We used IDW interpolation since it allows for smoothing, which is useful

in averaging the water levels of clusters of *uncertain* flood observations from social media content. In case of *certain* flood observations, which should be followed exactly by the interpolated water surface, bilinear spline interpolation may be more appropriate. An additional advantage of IDW interpolation is that the nominator and denominator of Eq. (1) can be updated with new observations, meaning the additional computational time in real-time applications is limited. We slightly modified the method proposed by Werner (2001) to improve the realism of the interpolated water surface. Firstly, water levels were expressed relative to the elevation of the nearest drain instead of mean sea level. Secondly, observations were interpolated along their downstream flow paths and subsequently projected to the grid cells upstream of these flow paths, to create a grid of water levels. From this grid we subtracted the HAND map, to create an initial grid of water depths in the area. Since the water surface might be extrapolated to areas which were separated from the observations by small barriers, flooded areas that were not connected to any of the observations were removed, similar to the method suggested by Werner (2001). This procedure produced the deterministic flood maps.

3.4 Uncertainty analysis

The uncertainties in the flood extent maps were investigated using a Monte Carlo analysis. We evaluated the uncertainty originating from errors in the locations derived from Tweets, errors in the elevation data, uncertainty in the parameters of the IDW equation and uncertainty in the DWD. The characteristics of the error distributions used to simulate these errors are given in Table 2.

The analysis of the locational errors of Tweets indicated that the locations derived from Tweets that refer to point locations contain less error than those derived from Tweets that refer to streets (see Sect. 4.1). The locational errors of both types of Tweets were therefore simulated differently. We simulated the locational errors of Tweets that referred to point locations by adding random errors to the spatial coordinates of the Tweets. The locational errors of Tweets that referred to streets were simulated along these streets. We did this by extracting the streets to which the Tweets referred from OpenStreetMap. The locational error was modelled using a normal distribution. To generate each realization, the observations were moved a distance along the street, which was drawn from this distribution. As some streets are shorter than six standard deviations, effectively reducing the modelled error, the standard deviation used for modelling these errors was modified so that the resulting errors matched the observed locational errors.

Since no accurate information was available regarding the errors in the EA DTM, these errors were simulated using typical values from literature. Since using independent normally distributed errors does not accurately reflect errors in the elevation data (Heuvelink et al., 2007; Raaflaub & Collins, 2006), spatially auto correlated errors were added using the method described by Dullof & Doucette (2014). Based on typical values of standard deviations and auto correlation distances of errors in LIDAR elevation data found in literature (Leon et al., 2014; Mudron et al., 2013; Li et al., 2011; Livne & Svoray, 2011; Hodgson &

Bresnahan, 2004) a standard deviation of 20 cm and correlation distance of 100 m was used. These errors were added to the original 2 m resolution DTM, before resampling it to 20 m resolution and creating the corresponding HAND map.

The uncertainty caused by the input parameters, being the power and smoothing parameters (Eq. 2) and the DWD, was also evaluated. Based on photographs in news articles about the flooding in York, the water depth in most places was estimated to be between 20 and 80 cm. Therefore, the DWD was varied between 20 and 80 cm. For the smoothing and power parameter, no clear information about the range was available. Errors in these parameters were simulated using the rather conservative ranges of 0-2000m and 2-5 respectively. A uniform distribution was used to simulate errors in the DWD, range- and power-parameters, since there was no specific information available regarding their error distributions.

To determine the number of Monte Carlo simulations required to produce the probabilistic flood maps, multiple maps were created using the same input uncertainties. It was found that using 1000 Monte Carlo simulations, two probabilistic flood maps generated using the same input error distributions were nearly identical.

3.5 Evaluation of results

We evaluated the accuracy of both the deterministic and probabilistic flood maps, by comparing these to the validation data discussed in Sect 3.1. In addition, we reviewed the relative importance of the different error sources for the probabilistic flood maps. The accuracy of the deterministic flood maps was evaluated by calculating the $F^{(2)}$ statistic of Aronica et al. (2002):

$$F^{(2)} = \frac{A_{obs} \cap A_{mod}}{A_{obs} \cup A_{mod}}, \tag{3}$$

Where $A_{obs} \cap A_{mod}$ is the area that is both modelled and observed flooded (true positive area) and $A_{obs} \cup A_{mod}$ is the area that is either modelled or observed flooded (true positive, false positive and false negative area).

We evaluated the accuracy of the probabilistic flood maps using reliability diagrams. These diagrams offer a comparison between the modelled probability (on the horizontal axis) and observed probability (vertical axis) of flooding (Wilks, 2006). To construct the reliability diagrams, we first binned the modelled probabilities in 10% intervals. For each probability bin, the cells on the probabilistic map that fell within that 10% interval were compared to the same cells in the validation data. The observed probability was calculated by dividing the number of selected cells that were flooded in the validation data by the total number of cells in the bin. We assumed that the central value of each 10% interval was the modelled probability, which we plotted along with the calculated observed probability. Note that the first probability bin ranged from 0.01% to 10%, and the 0% probability of flooding value in the diagram included all cells having less than 0.01% probability of flooding.

We assessed the relative importance of the different sources of error on the uncertainties in flood extent by creating three different uncertainty estimates: one by only simulating locational error, one by only simulating errors in the DTM and one by

only simulating errors in the parameters. For every uncertainty estimate, the $F^{(2)}$ statistic of each random simulation was calculated. We used these values to derive three empirical cumulative distributions of the $F^{(2)}$ statistic for the uncertainty estimates generated by simulating the individual sources of error. These were used to review the relative importance of the different types of errors for the accuracy of the maps.

4 Results

During the York floods, 8,000 unique flood related Tweets, posted between the 25th and 30th of December 2015, were harvested. Using the process discussed in Sect. 3.1 a database of 160 Tweets was constructed. Only from 87 of these could a location be derived from the text of the message. 17 Tweets mentioned a point location (an address, intersection or POI) and 70 Tweets mentioned a street name, for which the elevation data was used to derive a point location. Although 56 Tweets from which locations were derived, had photographs attached, we could only match 26 of them to a location on Google StreetView. These were used to assess the quality of the locational references derived from the text of Tweets.

4.1 Locational errors

We compared the locations that were derived from the text in the Tweets and used to create the flood maps to locations derived from attached photographs, in order to evaluate the magnitude of locational errors. Figure 5a gives the result of this comparison. The magnitude of locational errors depends heavily on the type of locational references in the Tweets. If point locations such as POIs or intersections are mentioned in the Tweets, the locational error is limited (Fig. 5c). If streets are mentioned in the Tweets however, locational errors are considerably larger (Fig. 5b). The difference is likely caused by the fact that the locations of point references were directly extracted from Google Maps, whereas an additional procedure was necessary to derive exact spatial coordinates from Tweets referring to streets. The large outlier in Fig. 5b is for example a Tweet that refers to Huntington Road, which is a long street that is located alongside the River Foss (see Fig. 1). The photograph was made at the northern end of Huntington Road, whereas the Tweet was pinpointed to the deepest depression along the road, in the south. There were however photographs attached to other Tweets which indicated that this southern location was also flooded. Without this outlier, the standard deviation in locational errors of Tweets referring to streets reduced to 118 m.

4.2 Flood Extent mapping

We created deterministic flood extent estimates by interpolating the locations and water levels derived from the Tweets. The parameters of the IDW interpolation were calibrated using the $F^{(2)}$ value based on the EA recorded flood extents. A power parameter of 4 in combination with a smoothing of 600 m (Eq. 2) and a DWD of 50 cm gave the best results. Figure 6 gives a comparison between the flood extent generated using information harvested from the Tweets and the validation data. An $F^{(2)}$ value of 0.69 was found, indicating that the modelled area of the flood extent that is correct, makes up 69% of the total flood extent either in the modelled or observed data.

The flood extent estimate is correct for a large part of the inner-city (location [1]). Even at smaller flooded areas, such as the ones north-west and south-east of location [2] a good estimate of flood extent is generated. The added value of separating groups of observations that are not in the same flooded area is seen at area [3]. Without separating observations, the underestimation of flood extent in this area would be considerable, whereas separating the observations results in a much better flood extent estimate for this area. Although at some locations minor underestimations of flood extent are seen, there is only one large area missing at location [4]. However, no observations of flooding close to, or in this area were found in the Twitter dataset. This underestimation is therefore a result of the lack of data rather than an error of the interpolation method.

4.3 Uncertainty Analysis

We created probabilistic flood extent estimates by varying the input parameters as well as simulating the locational errors and errors in the DTM in a Monte Carlo analysis. Based on the results in Sect. 4.1 the error distance along streets was modelled using a normal distribution with a standard deviation of 200 m. This modelled error effectively translates into a standard deviation in spatial coordinates of 100 m as some streets were too short to reproduce the full error distribution. Given the results from Fig. 5c errors in point locations were simulated using a normal distribution with a standard deviation of 50 m. The uncertainty resulting from simulating these locational errors along with errors in the DTM and parameters is given in figure 7a.

The uncertainty in flood extent is considerable (i.e.: the flood probability is around 50%). However, near the inner-city, at location [1], the uncertainty is limited. This is only partly caused by the high density of observations in this area and is mostly a result of the fact that the inner-city of York is situated lower than its surroundings, effectively limiting flood extents. For the areas within York that are more flat, the uncertainty in flood extent was generally larger. The density of observations is not well represented in the uncertainty estimates. Generally speaking, one would expect an area that has a high probability of flooding to have multiple observations in it, since a single observation can be placed there due to the Tweet being misinterpreted. At location [3] however, there is a large area with a high probability of flooding, even though only one observation is pinpointed to it. Location [3], as well as location [2] had high probabilities of flooding, although they were not flooded in reality.

We assessed the performance of the probabilistic flood extent map in Fig. 7a by comparing it to the validation data and constructing a reliability diagram (Fig. 8a). Probabilities between roughly 15 to 85% are mostly overestimated, although the most important probabilities close to 0 and 100% are accurately represented in the map. Comparing this map of all uncertainties to a map created by simulating only the errors in the elevation data and the parameters (Fig. 7b) indicates that locational errors are likely responsible for a considerable amount of uncertainty in the flood extent estimates. This is further confirmed by the results in Fig. 9, which shows the empirical cumulative distribution functions of the $F^{(2)}$ measure of accuracy. This was

calculated by using the result of each random simulation of the Monte Carlo Analyses of different types of errors separately. It can be clearly seen that locational errors cause most variation in the accuracy of the maps.

However, the reliability diagram that was constructed using the map generated without simulating errors in location (Fig. 8b), shows that by omitting the simulation of locational error, the uncertainty calculated using the Monte Carlo analysis more accurately describes the real uncertainty in flood extent. This indicates that either the probability distributions used to simulate these errors or the way these errors are propagated cause the flood probability to be overestimated.

5 Discussion and recommendations

This study shows the potential of using inherently uncertain social media content to create deterministic and probabilistic flood maps (Sect. 5.1), although the methods used in this study still contain some limitations (Sect. 5.2). Therefore, recommendations for future research are presented in the last section of this chapter (Sect. 5.3).

5.1 Potential

We showed that a deterministic flood map can be created from social media content. However, large uncertainties, mainly related to the locations derived from the content, still remain. Therefore, the probabilistic maps proved to be a useful addition to the deterministic map. Firstly, they are a source of information in itself. For example, where the deterministic map contained an underestimation of flooded area at location [4] (Fig. 5), the uncertainty estimate showed that flooding was highly uncertain at this location. This information can be used to send staff into the area, to verify if the area is actually flooded and thereby reduce the uncertainty at this location. Similarly, the probabilistic map confirmed the accuracy of the deterministic map near the inner city of York. Furthermore, the probabilistic maps provide information about the flood extent without the need for prior calibration of the model parameters. Therefore, these maps can potentially provide real-time flood extent information, without having to calibrate the method to that particular event or location first. However, to understand how the modelled uncertainty relates to the observed flood extent for a particular area or event, some validation might still be required.

A comparison to the work of Giustarini et al. (2016), who produced probabilistic flood maps from synthetic aperture radar (SAR) data and used the same validation technique, indicates that results are similar. It illustrates that probabilistic flood maps from SAR data provide a degree of accuracy comparable to the ones in our study, with probability-error values up 0.38. Although their reliability diagrams differed among case studies, none of them had a consistent overestimation of flood probability in all bins of the reliability diagram, like the ones from social media content. This indicates that the method presented in this paper still has some limitations.

5.2 Limitations

A possible reason for the overestimation of flood probability may be the fact that photographs were used to evaluate locational errors. A photograph can be taken at a location different from the one in text, for example because that location was too severely flooded, causing the locational error to be overestimated. In addition, the method used to derive locations from Tweets that referred to streets could only identify a single location of flooding along a street, causing others to be omitted. The outlier, mentioned in Sect. 4.1, illustrates this. Although it was pinpointed to a location that actually flooded, a large error was calculated, because the photograph was of a second flooded location along the same street. Since this is an error of omission, rather than a locational error, the exclusion of the outlier is believed to have given a better estimate of the standard deviation in locational errors.

The probability distributions used to simulate locational errors might also have contributed to the overestimation of flood probability. The normal distribution that was used does not reflect the sharp peaks seen at 0m in the graphs of Fig. 6. Using a conventional error distribution also does not give a correct representation of the actual errors in location. In reality, it is more likely that an observation originates from a lower location or a topographical depression, whereas purely using random errors can place observations on top of hills, which are unlikely to be flooded.

Furthermore, the results of the analysis could have been affected by the quality of the maps used for validation. The data for validating the river flood extents was created from a combination of ground observations and aerial photography. In places flooded separately from the river however, recorded historic flood extents were used, which might have been in accurate. However, actual observed flood extents for 2015 were used for the majority of the area. Therefore, we have no reason to believe that there are large uncertainties in the validation data.

We expect that an overestimation of either the errors in the DTM or the parameters is one of the main reasons for the overestimation of flood probabilities. Both the quantification of these errors as well as the methods used to simulate them could have caused this. It is likely that the quantification of parameter errors contributed most, since these were quantified conservatively in absence of accurate information about their error distribution.

Another important reason for the overestimation may lie in the 20 m resolution used for the maps. This resolution was chosen as a compromise between accuracy and computational time, though the results indicate that some barriers in the area were not accurately represented at this resolution. This caused some areas to erroneously be assigned a high probability of flooding.

The probabilistic maps generated in this study also did not consider the density of observations. Although all errors were drawn from the same error distributions in the Monte Carlo simulation, observations that belong to large clusters are more certain

than observations that are completely isolated. Because we did not consider this, the maps contained areas with a high probability of flooding even though these areas contained very few observations.

5.3 Recommendations

Besides resolving the issues related to the quantification and simulation of errors discussed above, a way to include important barriers in the coarse resolution DTM should be investigated. Although using higher resolution data can provide a similar improvement, it will seriously affect computational time and therefore affect the potential of real-time application of the maps. Additionally the inclusion of observation density in the uncertainty analysis should be reviewed.

To guide further improvements, it should be investigated whether it is useful to invest in optimizing the simulation of the different types of errors, or whether large improvements can be made by post-processing the results. Investigating more case studies can show whether flood probability is consistently underestimated, or if the reliability diagram differs by case. Reviewing more case studies can also show the effect of area topography on the resulting maps. We expect that uncertainties in flood extent are less for hilly areas than for flatter areas. By testing the method on multiple floods at the same location, as well as floods at different locations, also the (in)dependency of model parameters can be further investigated.

Where current methods for flood extent mapping such as hydraulic models and remote sensing have shortcomings in real-time application, this is where the real value of using social media content lies. The methods used in this report can potentially be applied in real-time. Random simulations for the York case were generated at a pace of about 100 simulations per minute, and the fact that calculations for single observations can just be added to the nominator and denominator of Eq. 1, ensures that adding new observations does not call for a complete recalculation of the results. To further improve computational time, alternative sampling techniques should be reviewed, since this can reduce the number of Monte Carlo simulations necessary.

Besides optimizing computational time, a further look into the gathering of observations is required. For real-time applications, it is vital to collect a high number of observations to ensure an accurate and up to date map can be produced at any point in time. The selection techniques used in this paper were only able to find a small number of tweets. It should be reviewed whether using different search techniques, additional sources of data, or techniques such as crowd interaction can increase the number of observations available for creating the maps.

6 Conclusions

This study illustrates that social media content has real potential in generating flood extent estimates. Although errors in locations derived from the Tweets were considerable, the deterministic flood extent map presented in this paper showed good

agreement with validation data. The deterministic flood maps can therefore be used to gain insight into the current situation of flooding.

Using information about the errors in the Tweets, DTM and parameter settings, we constructed a probabilistic flood extent
5   map. The uncertainty in flood extent mainly originated from the locational errors of Tweets, whereas DTM and parameter errors contributed less to flood extent uncertainty. A comparison of the probabilistic map to validation data showed that by simulating errors in the Tweets, DTM and parameters, a reasonable estimate of flood extent uncertainty is generated, which provides users with additional information on top of the deterministic flood map.

10   These results illustrate that social media content can be used to derive information about floods, even more so when exploiting the uncertainties in this data source. If further improvements are made, so that the methods used in this report can be applied in real-time, these maps have the potential of filling in the gap where hydraulic models and remote sensing are lacking.

Code Availability

The analyses in this paper were performed using Python 2 scripts. The code used for the different analyses (Brouwer, 2016) is
15   publically available on GitHub and published in the Zenodo research data repository (doi: 10.5281/zenodo.165818).

Data Availability

In the study data from downloaded from the Twitter API as well as data from the Environment Agency was used. The filtered subset of Tweets used in the research, the information about streets extracted from OpenStreetMap as well as the 20m resolution DTM and HAND map can be found in the aforementioned GitHub project (doi: 10.5281/zenodo.165818). Also the
20   data used to create the plots and maps are available at this location.

Competing Interests

The authors declare that they have no conflict of interest.

References

Aronica, G., Bates, P. D. and Horrit, M. S.: Assessing the uncertainty in distributed model predictions using observed binary
25   pattern information within GLUE, Hydrol. Process., 16, 2001-2016, doi:10.1002/hyp.398, 2002
Brouwer, T.: Twitter Flood Mapping Scripts: First Release [Data set], doi: 10.5281/zenodo.165818, 2016

Carter, W.N.: Disaster Management: A Disaster Manager's Handbook, Asian Development Bank, Mandaluyong City, Philippines, 2008

Dullof, J. and Doucette, P.: The Sequential Generation of Gaussian Random Fields for Applications in the Geospatial Sciences. Int. J. Geo-Inf., 3, 817-852. doi:10.3390/ijgi3020817, 2014

5   Environment Agency, LIDAR Composite DTM – 2 m: https://data.gov.uk/dataset/lidar-composite-dtm-2m1, 2014

Environment Agency, Recorded Flood Outlines: https://data.gov.uk/dataset/recorded-flood-outlines1, 2015

Environment Agency, Environment Agency LIDAR data Technical Note: http://www.geostore.com/environment-agency/docs/Environment_Agency_LIDAR_Open_Data_FAQ_v5.pdf, 2016

Earle, P. S., Bowden, D. C. and Guy, M.: Twitter earthquake detection: earthquake monitoring in a social world, Ann.
10   Geophys-Italy., 54(6), 708-715, doi:10.4401/ag-5364, 2011

Eilander D., Trambauer P., Wagemaker J. and Van Loenen A.: Havesting Social Media for Generation of Near Real-time Flood Maps, Procedia Engineering, 154, 176-183, doi: 10.1016/j.proeng.2016.07.441, 2016

Fohringer, J., Dransch, D., Kreibich, H. and Schröter, K.: Social media as an information source for rapid flood inundation mapping. Natural Hazards and Earth System Sciences, 15, 2725-2738, doi:10.5194/nhess-15-2725-2015, 2015

15   Giustarini, L., Hostache, R., Kavetski, D., Chini, M., Corato, G., Schlaffer, S., Matgen, P.: Probabilistic Flood Mapping Using Synthetic Aperture Radar Data. Ieee T. Geosci. Remote, 54(12), 6958-6969, doi:10.1109/TGRS.2016.2592951, 2016

Hahmann, S., Purves, R. S. and Burghardt, D.: Twitter location (sometimes) matters: Exploring the relationship between georeferenced tweet content and nearby feature classes. Journal of Spatial Information Science, 9, 1-36, doi: 10.5311/JOSIS.2014.9.185, 2014

20   Heuvelink, G. B. M., Brown, J. D. and Van Loon, E. E.: A probabilistic framework for representing and simulating uncertain environmental variables. Int. J. Geogr. Inf. Sci., 21(5), 497-513, doi:10.1080/13658810601063951, 2007

Hodgson, M. E. and Bresnahan, P.: Accuracy of Airborne Lidar-Derived Elevation: Empirical Assessment and Error Budget. Photogramm. Eng. Rem. S., 70(3), 331-339, doi:10.14358/PERS.70.3.331, 2004

Holderness, T. and Turpin, E.: From Social Media to GeoSocial Intelligence: Crowdsourcing Civic Co-management for Flood
25   Response in Jakarta, Indonesia, in: Social Media for Government Services, Nepal S., Paris C. and Georgakopoulos D., Springer International Publishing, Basel, Switzerland, 115-133, 2015

Leon, X. J., Heuvelink, G. B. M. and Phinn, S. R.: Incorporating DEM Uncertainty in Coastal Inundation Mapping, PLOS ONE, 9(9), doi:10.1371/journal.pone.0108727, 2014

Li, S., MacMillan, R. A., Lobb, D. A., McConkey, B. G., Moulin, A. and Fraser, W. R.: Lidar DEM error analyses and
30   topographic depression identification in a hummocky landscape in the prairie region of Canada. Geomorphology, 129, 263-275, doi:10.1016/j.geomorph.2011.02.020, 2011

Livne, E. and Svoray, T.: Components of uncertainty in primary production model: the study of DEM, classification and location error. Int. J. Geogr. Inf. Sci., 25(3), 473-488. doi:10.1080/13658816.2010.517752, 2011

Mason, D. C., Davenport, I. J., Neal, J. C., Schumann, G. J.-P., & Bates, P. D.: Near Real-Time Flood Detection in Urban and Rural Areas Using High-Resolution Synthetic Aperture Radar Images, IEEE T. Geosci. Remote, 50(8), 3041-3052, doi:10.1109/TGRS.2011.2178030, 2012

McClanahan, B. and Gokhale, S. S.: Location Inference of Social Media Posts at Hyper-Local Scale, 3rd International Conference on Future Internet of Things and Cloud, Rome, 25-26 August 2015, doi:10.1109/FiCloud.2015.71, 2015

Met Office, Further rainfall and flooding across north of the UK:

http://www.metoffice.gov.uk/climate/uk/interesting/december2015_further , last access: 27 December 2016

Mudron, I., Podhoranyi, M., Cirbus, J., Devecka, B. and Bakay, L.: Modelling The Uncertainty of Slope Estimation from A Lidar-Derived Dem: A Case Study from A Large-Scale Area in The Czech Republic, GeoScience Engineering, 59(2), 25-39, doi:10.2478/gse-2014-0051, 2013

Norbre, A. D., Cuartas, L. A., Hodnett, M., Renno, C. D., Rodrigues, G., Silveira, A., Waterloo M. and Saleska, S.: Height Above the Nearest Drainage – a hydrologically relevant new terrain model,. J. Hydrol., 404, 13-29, doi:10.1016/j.jhydrol.2011.03.051, 2011

Norbre, A. D., Cuartas, L. A., Momo, M. R., Severo, D. L., Pinheiro, A. and Norbre, C. A.: HAND contour: a new proxy predictor of inundation extent. Hydrological Processes, 30(2), doi:10.1002/hyp.10581, 2016

Pidd, H., A year after the deluge, York is still counting the cost: https://www.theguardian.com/uk-news/2016/dec/26/a-year-after-the-deluge-york-is-still-counting-the-cost, last access: 2 February 2017, 2016

Raaflaub, L. D. and Collins, M. J.: The effect of error in gridded digital elevation models on the estimation of topographic parameters. Environ. Modell. Softw., 21, 710-732, doi:10.1016/j.envsoft.2005.02.003, 2006

Rennó, C. D., Nobre A. D.,Cuartas L. A., Soares J. V., Hodnett M. G., Tomasella J. and Waterloo M.J.: HAND, a new terrain descriptor using SRTM-DEM: Mapping terra-firme rainforest environments in Amazonia, Remote Sensing of Environment, 112(9), 3469-3481, doi: 10.1016/j.rse.2008.03.018, 2008

Schnebele, E., Cervone, G., Kumar, S. and Waters, N.: Real Time Estimation of the Calgary Floods Using Limited Remote Sensing Data. Water, 6, 381-398. doi:10.3390/w6020381, 2014

Schumann, G., Bates, P. D., Horrit, M. S., Matgen, P. and Pappenberger, F.: Progress in Integration of Remote Sensing-derived Flood Extent and Stage Data and Hydraulic Models. Rev. Geophys., 47. doi:10.1029/2008RG000274, 2009

Smith, L., Liang, Q., James, P. and Lin, W.: Assessing the utility of social media as a data source for flood risk management using a real-time modelling framework. Journal of Flood Risk Management, doi:10.1111/jfr3.12154, 2015

Stone, L. D., Keller, C. M., Kratzke, T. M. and Strumpfer, J. P.: Search for the Wreckage of Air France Flight AF 447. Stat. Sci., 29(1), 69-80. doi:10.1214/13-STS420, 2014

Sun, D., Li, S., Zheng, W., Croitoru, A., Stefanidis, A. and Goldberg, M.: Mapping floods due to Hurricane Sandy using NPP VIIRS and ATMS data and geotagged Flickr imagery. International Journal of Digital Earth, 1753-8947, doi:10.1080/17538947.2015.1040474, 2015

UN: The human cost of weather related disasters 1995 – 2015, United Nations, Geneva, Switzerland, 30 pp., available at: http://www.unisdr.org/files/46796_cop21weatherdisastersreport2015.pdf, 2015

Werner, M.G.F.: Impact of Grid Size in GIS Based Flood Extent Mapping Using a 1D Flow Model, Phys. Chem. Earth Pt. B, 26(7-8), 517-522, doi:10.1016/S1464-1909(01)00043-0, 2001

5   Wilks, D.S.: Statistical Methods in the Atmospheric Sciences, Elsevier, Oxford, UK, 2006
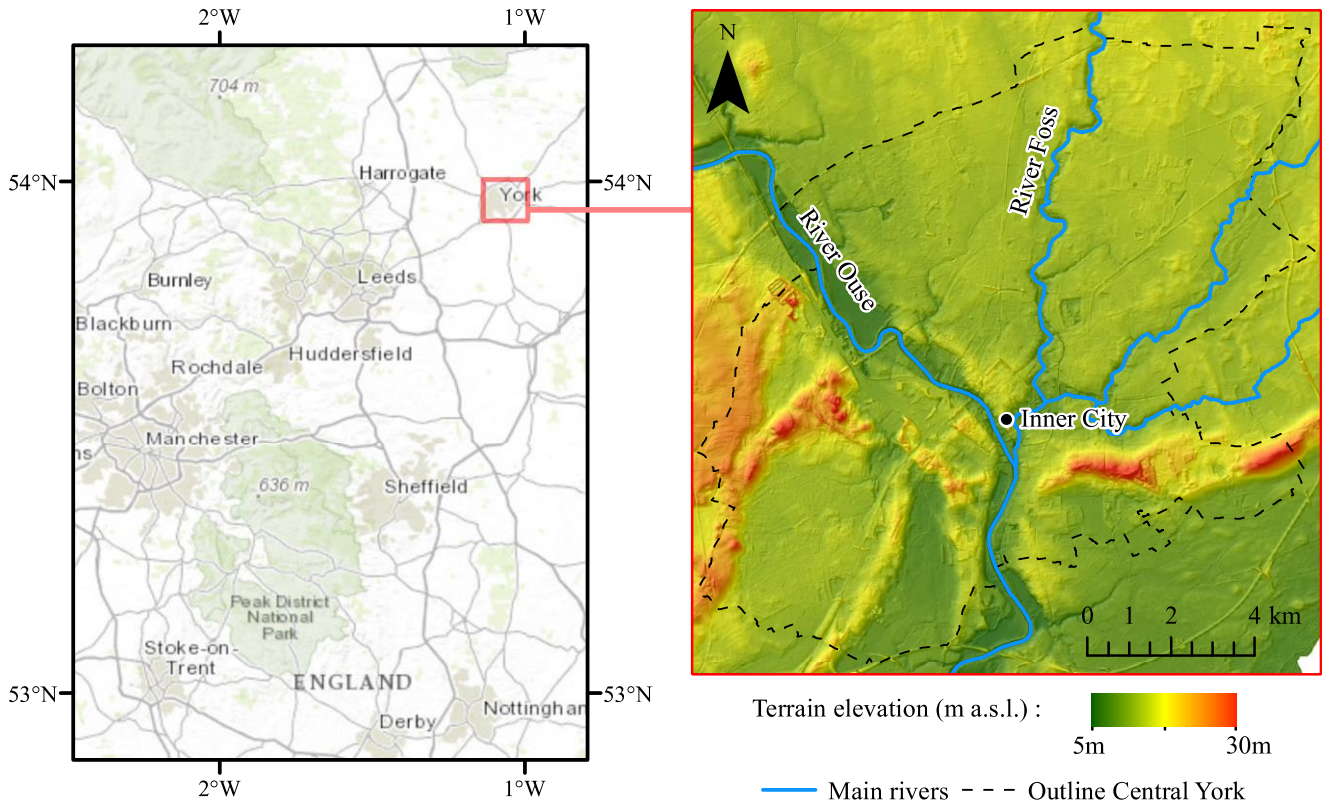
Figure 1. Digital Terrain model of the York study area
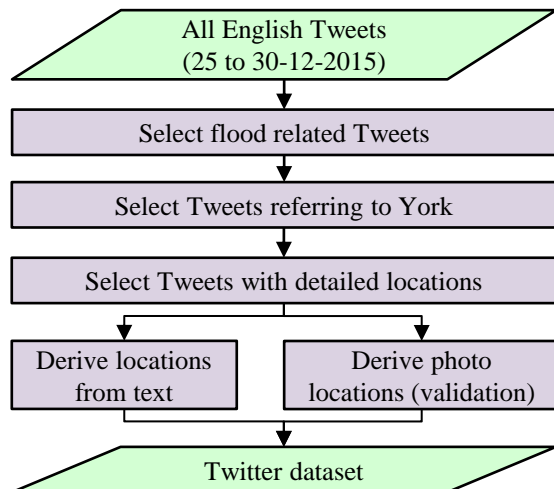
Figure 2. Process of constructing the dataset of Tweets

"Cumberland Street in York - say they're used to flooding here but only 2000 was worse"

Attached Photograph   Google StreetView

Figure 3. Example of determining the error in the spatial coordinates derived from the text of a Tweet, based on an attached photograph. The grey dot is the location derived from the text of the Tweet, and the black dot is the location derived from the attached photograph.
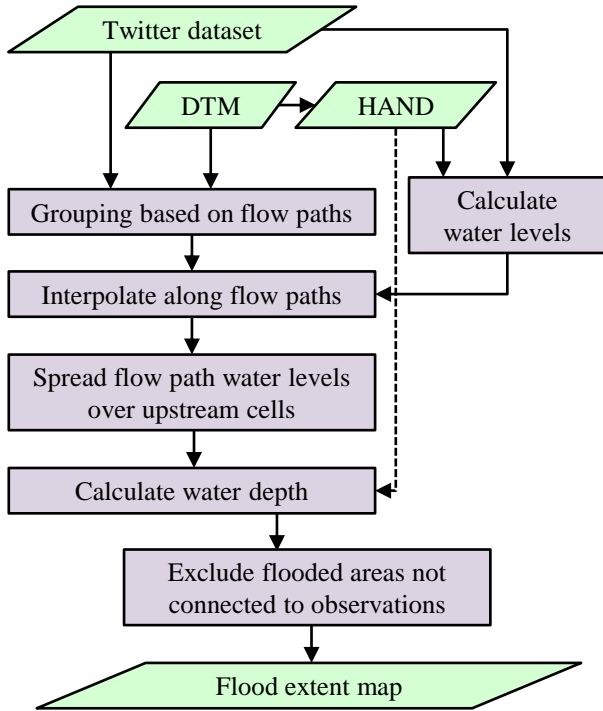
Figure 4. Process of creating flood extent maps
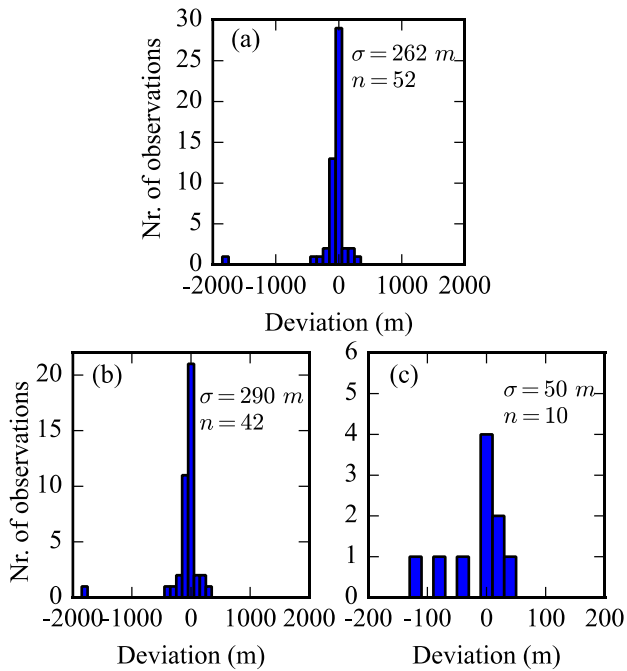
Figure 5. Locational errors in X/Y coordinates of all Tweets (a), only the ones referring to streets (b) and only the ones referring to point locations (c)
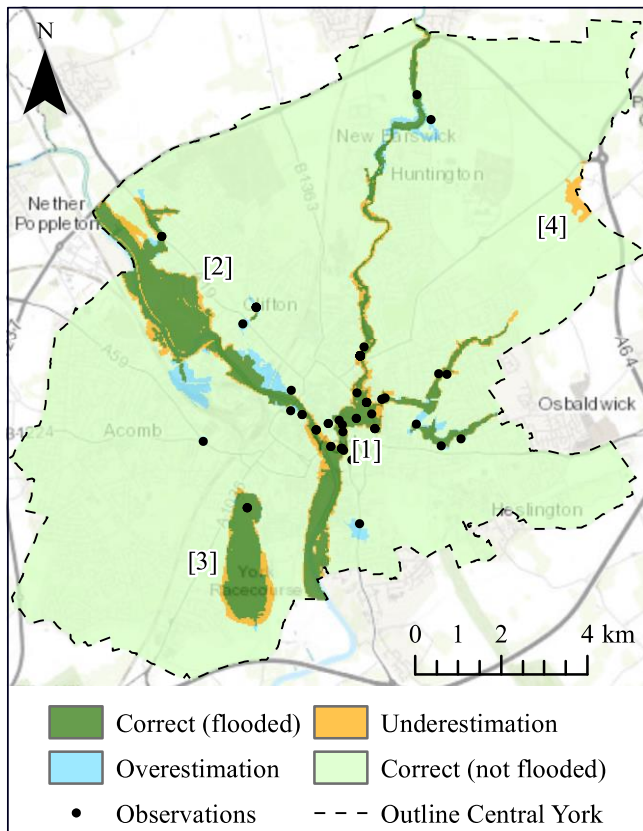
Figure 6. A comparison between the deterministic flood map (modelled) and validation data (observed). The locations denoted by the numbers [1] to [4] are referred to in text.
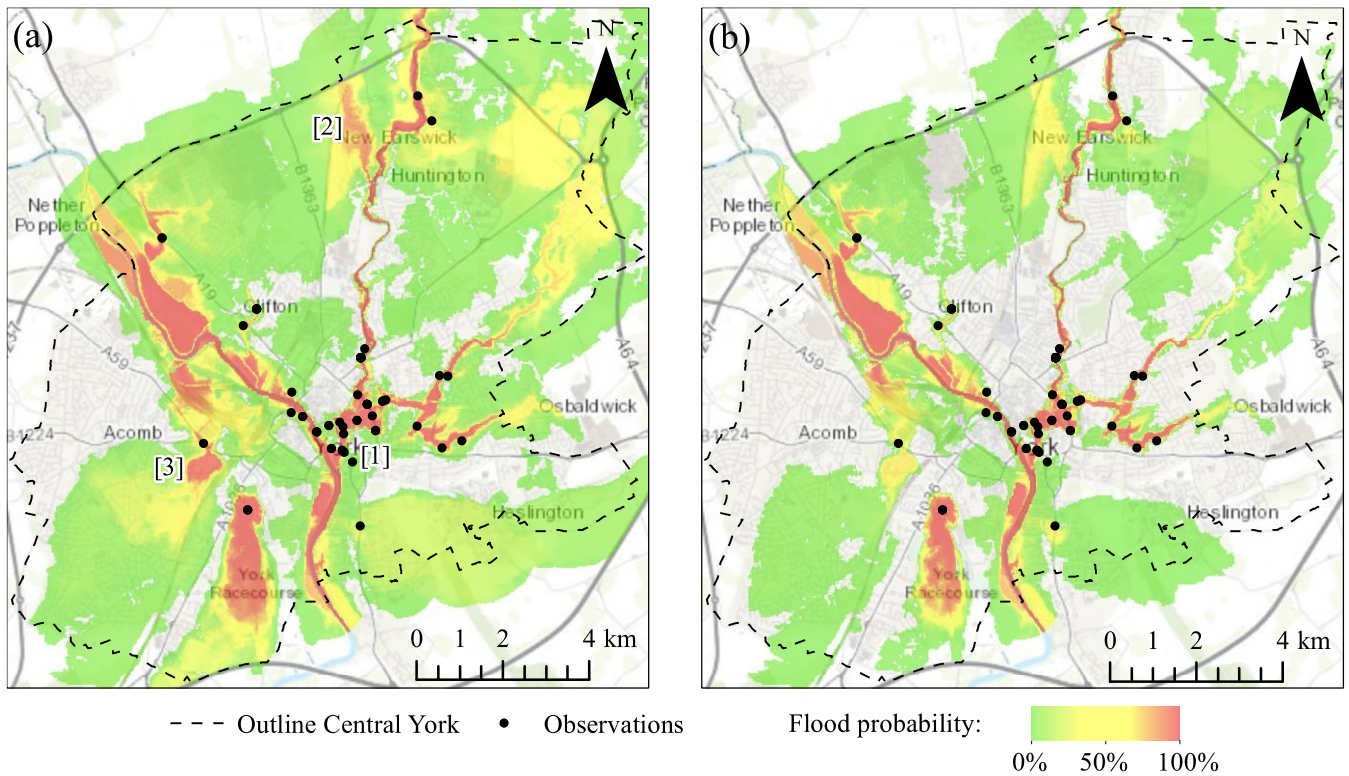
Figure 7. Probabilistic flood map generated by simulating locational errors, errors in elevation data and errors in parameters (a) and a probabilistic map generated simulating errors in the elevation data and parameters only (b). The locations denoted by the numbers [1] to [3] are referred to in text.
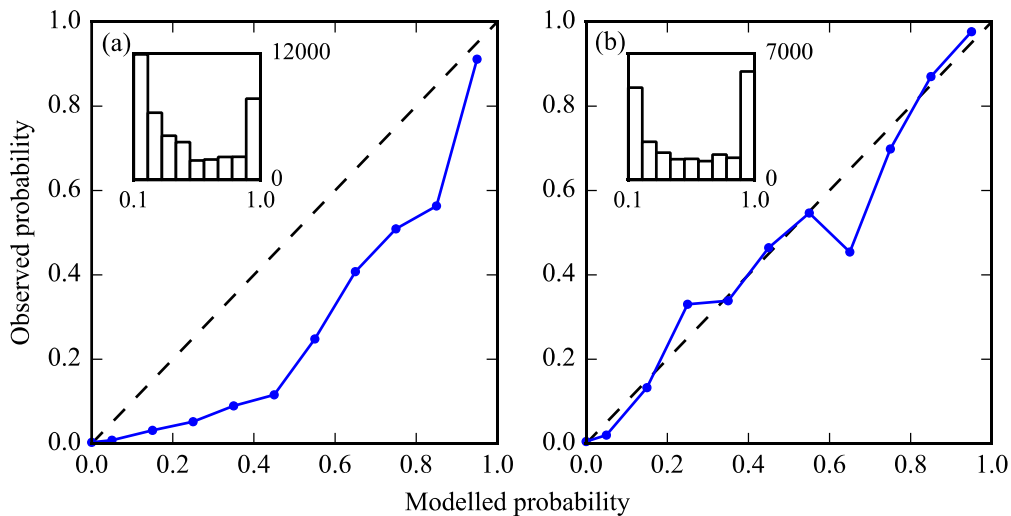
Figure 8. Reliability diagrams constructed from the probabilistic flood map generated by simulating all errors (a) and by simulating only errors in the elevation data and parameters (b). The small histograms give the number of cells within each 10% bin of modelled flood probability.
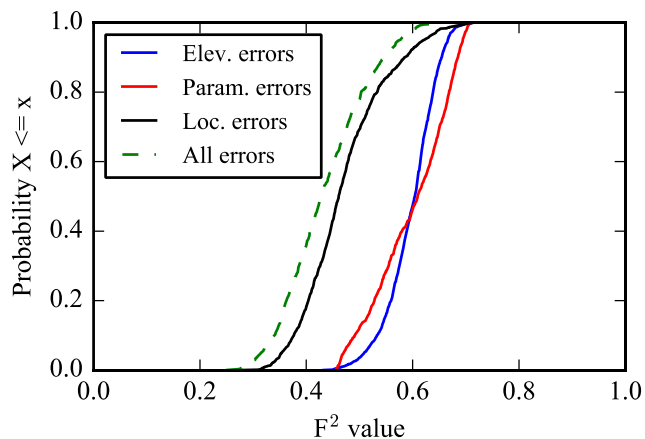
Figure 9. Empirical cumulative distribution functions of the $F^{(2)}$ performance statistics derived by simulating purely errors in the elevation data (blue), errors in the parameters (red), locational errors (black) and the combination of these errors (dashed green).

Table 1. Datasets used in this study

Data	Source	Purpose
2 m LIDAR DTM	EA (2014)	To group observations (Sect. 3.3)
		To calculate water levels (Sect 3.3)
		To estimate flood depth & extent (Sect 3.3)
		To pinpoint Tweets referring to streets (Sect 3.2)
Twitter	Twitter streaming API	To extract flood observations (Sect. 3.2)
Google Maps	Used online	To find locations with Tweets (Sect. 3.2)
Google StreetView	Used online	To find exact locations of photographs (Sect. 3.2)
OpenStreetMap	Exported from osm.org	To simulate locational errors along streets (Sect 3.4)
Recorded historic flood outlines	EA (2015)	To evaluate flood extent in areas affected by non-fluvial flooding (Sect 3.5)
Recorded 2015 fluvial flood outline York (draft)	EA (Personal communication)	To evaluate flood extent in areas affected by fluvial flooding (Sect. 3.5)

Table 2. Quantification of error sources. [a]See Sect. 4.3

Error source	Distribution	Parameter	Value
Elevation data	Normal (spatially auto correlated)	μ:	0 m
		σ:	0.2 m
		Corr. distance:	100 m
Tweets (point location)[a]	Normal (X/Y coordinate)	μ:	0 m
		σ:	50 m
Tweets (Street location)[a]	Normal (along street)	μ:	0 m
		σ:	200 m
Power parameter	Uniform (integers only)	Lower bound:	2
		Upper bound:	5
Smoothing parameter	Uniform	Lower bound:	0 m
		Upper bound:	2000 m
DWD	Uniform	Lower bound:	0.2 m
		Upper bound:	0.8 m