TIER 3: Probabilistic Flood Extent Estimates

Introduction

Globally, floods cause large damages and a huge number of casualties each year. During floods there are a number of parties that benefit from having high quality flood information at their expense. For example rescue workers can use information about the severity of flooding to choose which areas to target and which routes to take in an area. After floods there is also a need for flood information, which can for example help insurance companies in evaluating flood damages, aid organizations in targeting rebuilding efforts or local governments in evaluating flood risk. As a result of the increasing number of floods, caused by phenomena such as urbanization and climate change, there is an increasing demand for accurate and timely flood information. Traditionally this information is produced in the form of flood maps either generated using hydraulic models or remote sensing. However hydraulic models often need detailed schematizations of the study area, require large amounts of input data and can take considerable computational time. For remote sensing data the time it takes from an observations being made, to the release of the data is considerable and this data often has a low temporal resolution. These drawbacks, which particularly affect the potential of these methods for real-time applications, in combination with the rise of social networks over the last decade, have triggered the search for a new way of creating flood maps, using these social media.

The growth of social networks over the last decade has led to huge amounts of data, potentially containing valuable information about flooding, being available almost instantly. Several studies already looked into using this data, which either used it as auxiliary data for other methods to create flood maps or used the data to create flood maps directly. None of these studies however focussed on comparing different methods to create flood maps from social media data, or assessed the uncertainties in maps created using social media data. Therefore the objective of this research was to establish a preferred method of estimating flood extents from Twitter data and assess the uncertainties and applicability of the maps created using this method. Specifically Twitter data was used since it is openly and freely available. To achieve this objective, the research included a comparison of different ways of applying interpolation to create flood maps from the Twitter data, an assessment of the uncertainty in flood extent and a variety of analyses to investigate in what context the Twitter derived flood maps can be applied.

Methods

Filtering Tweets

The Twitter datasets used in this research were supplied by FloodTags. They analyse all Twitter messages in real-time and select messages which refer to flooding. A number of processing steps were applied on this data, to produce a database of Tweets that could be used in flood mapping, see figure with workflow below.

Flood mapping

A basic interpolation of water levels with reference to the local drainage level, derived from the locations and water depths reported by Tweets, was used as a starting point. The inverse distance weighting (IDW) interpolaion method was used here as it requires only limited computational time, smoothing can be used to filter the effects of outliers and the result does not have to be completely recalculated every time an observation is added. Several improvements over this simple method were evaluated. The most optimal method interpolates groups of observations based on their downstream flow paths. For each of these groups, water levels are interpolated along these flow paths. This assigns a water level to all cells belonging to a flow path downstream of an observation. The water levels throughout the study area are then determined by giving the water level value of each grid cell on the flow path, to all cells upstream of it (its subcatchment). Flooded areas which are not physically connected to any of these downstream flow paths are removed.
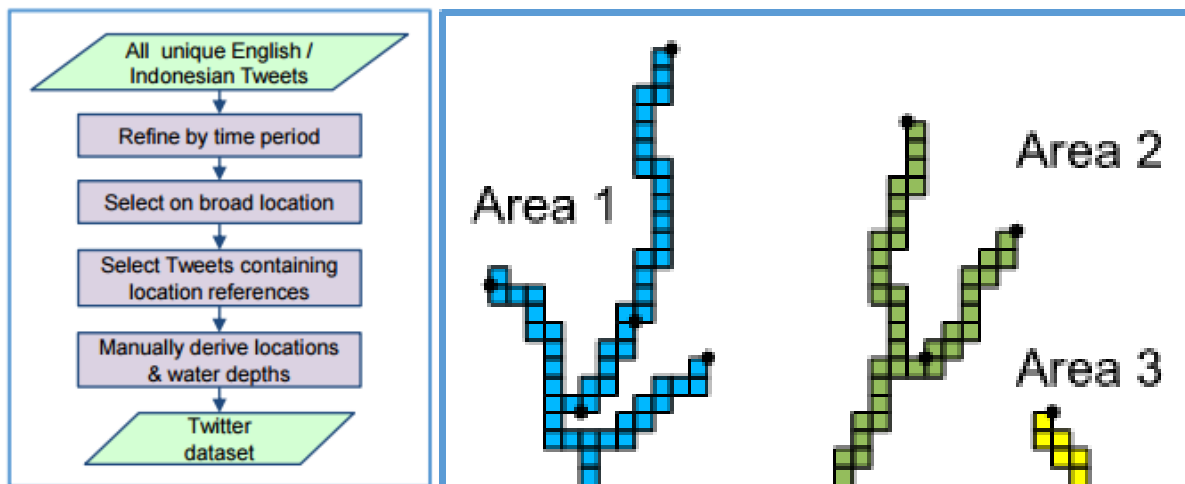
Left: Workflow for filtering tweets; Right: Interpolation along downstream flowpaths

Case studies

York floods

Jakarta floods

Two case studies of recent floods in Jakarta (Indonesia) and York (United Kingdom) were evaluated. For both case studies a dataset of Tweets was constructed, from which both locations and water depths were derived. As a first step the sets of Tweets collected for these cases and the locations and water depths derived from them, were investigated. The magnitude of errors in location and water depth derived from the Tweets was investigated by comparing them to locations and water depths derived from photographs attached to some of the Tweets. Iin Jakarta, the majority of the Tweet's locations are pinpointed within 200 m of their actual location, however, outliers of up to 3.5 km still exist in the dataset.The characteristics of both datasets are given in the table below.

Charasteristics filtered twitter data

	Jakarta	York
date	8/2-11/2 2015	25/12-30/12 2015
total number of filtered tweets	219	87
tweets with a water depth observation	123	0
average location error streets	659 m	287 m
average location error POI	236 m	52 m

Besides analysing the Tweets gathered for both case studies, different methods of creating flood maps were evaluated. The flood extents created for the Jakarta case study were validated using information derived from photographs and the flood extents created for the York case study were validated using actual recorded flood extents. Using the method that created the most accurate flood maps, the uncertainty in flood extent, resulting from locational errors of Tweets, errors in water depths mentioned by Tweets and errors in the elevation data, was evaluated using Monte Carlo simulations. Also the effects of choosing a different default water depth value and using different resolutions on the uncertainty in flood extent was investigated. The comparison of the flood extents generated using the different flood mapping methods with validation datasets showed that for both the Jakarta and York case studies, the best results were obtained by interpolating water levels along the flow paths downstream of observations.

The flood extent calculated for Jakarta covered 75% of the validation points and a comparison of the flood extents calculated for York with recorded flood extents showed that the area of the flood extent that was correct, made up 69% of the total flood extent gotten by merging the created and recorded flood extents. Although two different validation methods were used, making it hard to compare both case studies, the quality of the flood extents varied between the cases. It was seen that in more flat areas, such as downstream Jakarta, flood extents were less precise than in areas with more slopes, such as the inner-city of York. These differences in topography also affected the degree to which errors in the datasets caused uncertainties in flood extent. These uncertainties were especially high in flatter areas, which were mainly affected by locational errors of Tweets and errors in the elevation data. In areas with steeper slopes, these errors caused considerably less uncertainty and at all locations the uncertainty caused by errors in the water depth specified by the messages, or default water depth used for messages that did not mention one, was only minor. Given these large differences in uncertainties, the scale at which maps could be produced varied from fine for the inner-city of York for which flood extents were delineated to within 50 m of their actual location, to more coarse in flatter areas such as downstream Jakarta, where deviations of up to 500m were not uncommon.
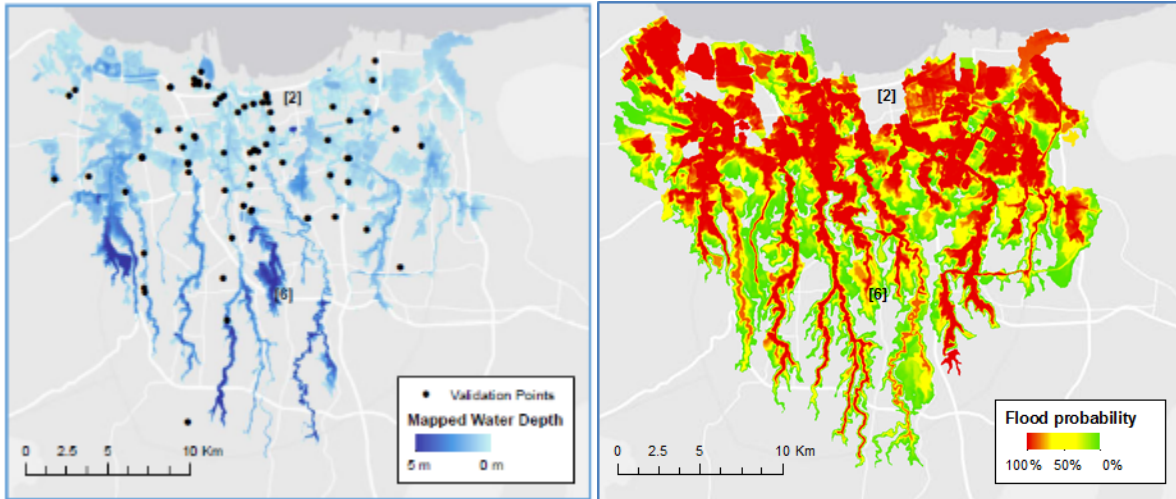
Jakarta modeled flood water depth (left) and flood probability (right)
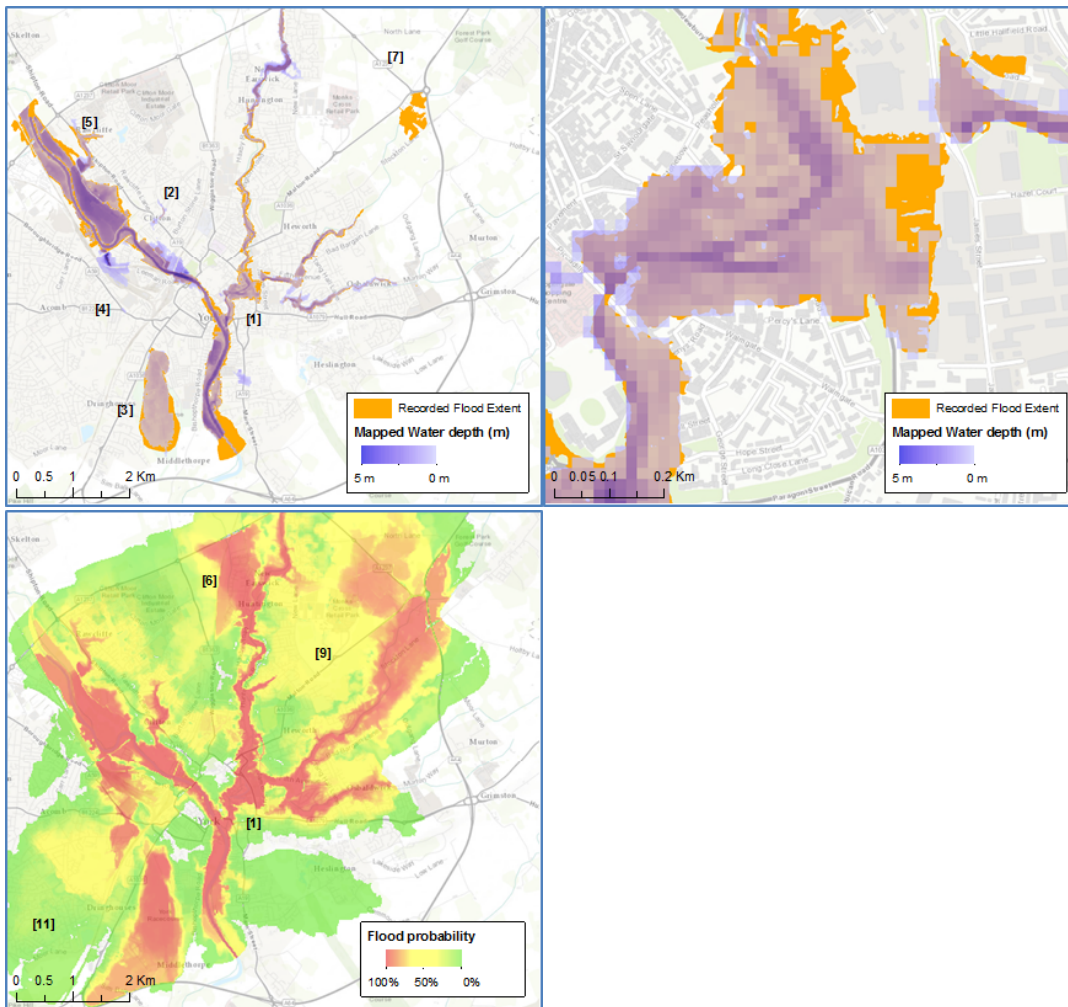
York modeled flood water depth (left) zoomed in to the city center (center) and flood probability (right)

Conclusions

The resulting flood maps can be applied to both pluvial and fluvial floods. The scale at which the flood maps can be used varies from case to case, depending on the topographical characteristics of the area, which determine to what extent errors in the input datasets propagate to the flood maps. For areas with high terrain slopes maps at fine scale can be produced, delineating flood extents to within 50 m of their actual location, whereas for more flat areas only conclusions can be drawn at more coarse scale, because deviations of 500 m in flood extent are not uncommon. Although the real-time application of the flood maps could not be fully reviewed, since the datasets used in this research contained too few observations and Tweets were not processed automatically, the computational time of the methods used to create the flood extent and uncertainty estimates, allows for application in real-time

Current methods to create flood inundation maps from Twitter messages have been improved. The method that uses interpolating of water levels along flow paths produced the best results. This method improved upon the basic interpolation of water levels by using the flow paths downstream of observations to determine which observations belong to the same continuously flooded area. Further improvements are made by first interpolating the water levels along these flow paths, instead of directly interpolating them throughout the entire area and subsequently excluding flooded areas which are not directly connected to the downstream flow paths of observations.

The degree of uncertainty caused by errors in the input dataset depends largely on the topographical characteristics of the area and can be large for flat areas with low terrain slopes. Mainly locational errors of Tweets and errors in the elevation data affect these locations. Since fluctuations in water levels have less of an effect in areas with steep terrain slopes, the uncertainty in these areas remains relatively limited. Also the uncertainties caused by errors in the water depth mentioned by the Tweets and default water depth added to observations without a water depth was found to have only a minor influence on the uncertainty in flood extent.

Operationalization

Some recommendations for operationalization of the Tier III method are made in this memo.

Furter reading

Brouwer, Tom. *Potential of Twitter derived flood maps: comparing interpolation methods and assesing uncertainties*. MS thesis. University of Twente, 2016.

Brouwer et al. *Probabilistic flood extent estimates from social media flood observations,* Nat. Hazards Earth Syst. Sci., 17, 735-747, doi:10.5194/nhess-17-735-2017, 2017