Potential of Twitter derived flood maps: comparing interpolation methods and assessing uncertainties

Author:
Tom Brouwer

Deltares
Enabling Delta Life

UNIVERSITY OF TWENTE.

August 22nd 2016

Potential of Twitter derived flood maps: comparing interpolation methods and assessing uncertainties

This report discusses the main results of a research study performed into using Twitter data to estimate flood extents and the uncertainties therein. This research was performed at Deltares, Delft and served as a master thesis for the programme Civil Engineering and Management, track Water Engineering and Management at the University of Twente, Enschede.

Author:

Tom Brouwer, Bsc.

Under supervision of:

Ir. Dirk Eilander

Ir. Arnejan van Loenen

Dr. Ir. Martijn Booij

Dr. Kathelijne Wijnberg

Status:

Final

Date:

22nd of August, 2016

Summary

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. Therefore 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. Also a digital terrain model at 20 m resolution was constructed for both case studies. 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.

As a first step the sets of Tweets collected for these cases and the locations and water depths derived from them, were investigated. The time variation in the number of Tweets in these datasets was reviewed by comparing it to the time variation in measured water levels. Also 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.

Besides analysing the Tweets gathered for both case studies, different methods of creating flood maps were evaluated. A basic interpolation of water levels, derived from the locations and water depths reported by Tweets, was used as a starting point. Several improvements over this simple method were evaluated. For example flooded areas that were not directly connected to any of the observations were removed from the flood map. Additionally the effect of grouping observations that belonged to the same flooded areas, either based on the vicinity of observations or common cells downstream of the observations, was investigated. A last method focussed on using the cells that lay downstream of observations, called the downstream flow paths of observations, to interpolate water levels along. Also the use of Tweets that did not mention a water depth, by giving them a default water depth, was reviewed. Instead of using a digital terrain model to produce the flood maps, a height above nearest drainage map was used in this research, which reduced the risk of downstream overestimations of water level.

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. Although the analysis of the time variation in the number of Tweets indicated that the severity of flooding was quite accurately reflected in the number of Tweets, there were too few Tweets in the datasets constructed for this research to do a thorough analysis of time variation. For the Jakarta case however, the dataset was intentionally reduced in size, since the Tweets had to be manually analysed.

Although the flood mapping methods used in this research, given their limited computational time, allowed for real-time application, also the manual process of extracting locations and water depths from the Tweets, should be automated to make this possible. Additionally the process of creating uncertainty maps should be further optimized, since these do not accurately reflect the degree of uncertainty caused by locational errors and density of observations. For cases such as the York case, for which only a small amount of relevant Tweets was found, further methods to generate and find more relevant observations should be reviewed. If these issues are addressed however, the real-time flood maps and uncertainty maps created using Tweets have the potential of providing a wealth of information to for example rescue workers or other persons requiring flood information in real-time, where current methods such as hydraulic models and remote sensing are lacking.

Acknowledgements

This report contains the results of the research I did at Deltares over the six month period between February and July 2016. It would however never have been written without Dirk Eilander and Arnejan van Loenen giving me the opportunity to work on this new and exciting subject at Deltares and supporting me both in preparing the research proposal as well as in executing the research itself. Also the help of Patricia Trambauer, who supported me in Dirks absence, was of great importance. Additionally I would like to thank my supervisors from the University of Twente, Kathelijne Wijnberg and Martijn Booij, for supporting me in both phases, and critically reflecting on the results of the research. Finally I would like to thank Jurjen Wagemaker & Ron Boortman from FloodTags, because without their help and data they provided, the research would not have been possible.

Table of Contents

List of abbreviations

The following abbreviations are used in this report:

Abbreviation	Meaning	Page
1D / 2D	1 Dimensional / 2 Dimensional	1
BPBD	Badan Penanggulangan Bencana Daerah (Regional disaster management agency)	21
DKI Jakarta	Special Capitol Region of Jakarta	16
DSM	Digital surface model	4
DTM	Digital terrain model	5
DWD	Default water depth	11
EA	Environment Agency	7
HAND	Height above nearest drainage	5
IDW	Inverse distance weighting (interpolation)	11
LIDAR	Light detection and ranging	5
ME	Mean error	12
POI	Point of interest	5
RMSE	Root mean square error	12
RQ	Research question	2
UK	United Kingdom	1
UN	United Nations	1
UTC	Coordinated universal time	18

1 Introduction

All over the world floods are responsible for a large number of casualties, cause widespread destruction of property and affect daily life by spreading diseases and disrupting businesses. In some places floods can even be a yearly recurring phenomenon. A recent report by the United Nations (UN) indicates that 2.3 billion people, about a third of the world's population, were affected by flooding between the years 1995 and 2015. This makes flooding, besides being the most occurring natural disaster worldwide, also the one affecting the largest number of people. Furthermore the number of floods occurring each year is rising due to phenomena like urbanization and climate change (UN, 2015).

These floods can have a number of causes. For example, if the discharge capacity of rivers is too low, these rivers can flood because of long lasting or short heavy rainfall events. Also the water levels at the outflow point of the river might river might rise, for example during spring tide, which causes water levels further upstream to rise also and can lead to flooding there. These floods originating from rivers are generally called fluvial floods. Rainfall can also cause direct flooding of an area, for example if during heavy rainfall the water cannot infiltrate and also cannot reach the drainage channels. This can be the case at local topographical depressions, and is called pluvial flooding. Additionally floods can occur near the sea. For example, during spring tide a levee might be breached, which causes a large area to flood. These floods are referred to as coastal floods.

Most efforts these days are aimed at preventing all types of floods, for example by installing better flood defences or reducing river runoff. The resources to do so are however lacking in many places. In case of extreme events flood defences might also not be able to cope, which was seen recently during the floods in the United Kingdom (UK) in December 2015. In combination with an ever increasing number of floods this calls for mitigation measures, such as the evacuation of inhabitants of affected areas by rescue workers. For rescue workers to respond effectively to a flood they should have information about which locations are affected by the flood and how severe the flooding at each location is. Such information not only guides the route rescue workers take in a flooded area or the choice at which location to offer flood relief first, but can also help in making the decision to dispatch rescue workers in the first place. Information about flooding can also be used after the flood has passed to help insurance companies in evaluating flood damages, aid organizations in targeting rebuilding efforts or local governments in evaluating flood risk.

1.1 State of the art

Generally flood information is generated in the form of flood maps, indicating which areas are flooded and how deep. These maps are often generated using 1D or 2D hydraulic models, which simulate water levels and flow velocities based on approximate discharges (e.g. Dottori & Todini, 2013; Leandro et al., 2009; Vojinovic & Tutulic, 2009) or using remotely sensed data, for example originating from satellite observations (e.g. Mason et al., 2010; Kussel et al., 2008). Both methods however have considerable drawbacks. Hydraulic models for example generally require large amounts of input data, such as rainfall or runoff data, as well as a detailed schematization of the study area. These models also require comprehensive calibration efforts to accurately simulate floods and take considerable computational time. Especially this last element reduces their utility in real-time applications. Although using remotely sensed data requires less computational time, users are faced with other drawbacks, such as the low temporal resolution of the data. This means the data often does not capture the flood peak and cannot be used to monitor the flood as it progresses. Also some sensors do not work in cloudy conditions and the time between the actual observation being made to the data being released often takes hours to days (Schuman et al., 2009), severely affecting the utility of the data in real-time flood mapping.

These drawbacks, along with the fast growth of several social networks over the last decades, such as Twitter (figure 1), Foursquare, Flickr and Facebook, triggered investigations into a new way of generating flood maps. The rise of these social networks has led to huge amounts of data being available for free and in near real-time. Sun et al. (2015) for example applied this data in a flood mapping context, and used Flickr images to validate remote sensing derived flood maps. Smith et al. (2015) on the other hand proposed to use Twitter data in conjunction with traditional data sources, by using the presence of Twitter messages (Tweets) to invoke hydraulic model runs and selecting the most representative model run based on a comparison with the Tweets. Others studies have investigated using social media observations directly, for example by creating flood maps based on the vicinity of Tweets about flooding (Schnebele et al., 2014) or using location and water depth information extracted from photographs attached to social media messages (Fohringer et al., 2015). Eilander et al. (2016) actually took this one step further and used both water depth and location information that was automatically derived from Tweets to create flood inundation

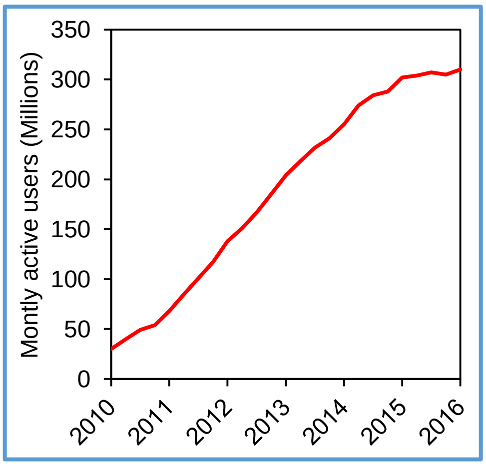

Figure 1: Number of monthly active Twitter users 2010 – 2016 (Statista, 2016)

maps. Validation using photographs in the area indicated that at 69% of locations derived from the photographs the flood map was correct.

1.2 Research gap

The examples mentioned above illustrate that already quite some work has been done on creating flood maps using social media derived data. Nevertheless some key issues are left unresolved. For example, although each study used a different method to create flood maps from social media data, none of them actually compared the results of several methods. Fohringer et al. (2015) already indicate their method contains some flaws and Eilander et al. (2016) indicate that their algorithm is mainly aimed at pluvial flooding but fails when applied to fluvial floods. The lack of a comparison between different methods means insight into the relative advantages and specific characteristics of mapping methods is lacking.

Also Fohringer et al. (2015) indicate that further research should focus on the uncertainties in the maps. Schnebele et al. (2014) point out that the reliability of social media derived observations is likely low, meaning considerable errors in water depth and location can be present, potentially causing large uncertainties in flood maps created using the data. These uncertainties and their impact on flood maps are not discussed in any of the previous research studies, although the assessment of uncertainties is common practice in research studies that involve using 1D or 2D models to map floods. An example of this is the GLUE methodology discussed by Romanowicz & Beven (2003). Since there is no insight into the uncertainty associated with flood maps created using social media data, the actual utility in day to day operation of these maps in unknown.

A last topic that has received little attention is the element of time. Although one of the major advantages of using social media is its real-time availability, most studies only create a map of peak flood extent. However, emergency services in the area, such as rescue workers, benefit from having the latest information at their fingertips. Therefore reviewing the possibility of real-time application of the data is important in assessing the potential of social media data in flood mapping.

1.3 Objective and research questions

The research discussed in this report aims at resolving the three issues described above. For this research data from Twitter was used, since it is freely and openly available. Using this data several flood mapping methods are compared and the uncertainties in these flood maps are reviewed. Also the applicability of the data, meaning the potential of real-time application and the locations for which the data can be used was investigated. The general objective of the research was the following:

"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"

To achieve this objective, the following research questions (RQs) have been answered:

1. How can current methods to create flood inundation maps from Twitter messages be improved upon?
2. How uncertain are the resulting flood extents?
3. In what context can the flood maps be applied?

To investigate at which locations the maps can be applied, case studies of recent floods in both Jakarta (Indonesia) and York (UK) were considered. Where Jakarta was both flooded from the rivers (fluvial) as well as directly by rainfall (pluvial), the flood in York mainly originated from the rivers (fluvial). The differences in Twitter behaviour, area topography and flood origin between these cases gives insight into the applicability of the method.

1.4 Reading guide

Chapter 2 first discusses each case study in more detail. The second paragraph of this chapter discusses the different datasets used in the research and the chapter is concluded by a discussion of the methodology used for answering each of the research questions. The same division in sub-sections found in the methodology paragraph, being the review of the characteristics and uncertainties of the Twitter dataset (used to partially answer RQ2 & 3), the comparison of different flood mapping methods (RQ1) and the uncertainty assessment (RQ2) is also found in the results chapters. Both the chapter discussing the results of the Jakarta Case study (chapter 0) and the York case study (chapter 0) consist of these three paragraphs. The results of the research are further discussed in chapter 5 and the final chapter of this report (chapter 6) discusses the conclusions and recommendations of the research. Specifics of the methodology and results can be found in appendices A and B respectively.

2 Case studies, materials and methods

Two case studies of recent floods in Jakarta and York were investigated. A short description of both is discussed in the first paragraph of this chapter. The second paragraph goes deeper into the datasets used for each of the case studies and the chapter is concluded by discussing the research methodology.

2.1 Case studies

Both the flooding of the city of Jakarta of February 2015 as well as the flooding of the City of York of December 2015 were investigated. Both case studies are shortly discussed below.

2.1.1 Jakarta

The flooding of the city of Jakarta was investigated first. The terrain elevation of Jakarta gives insight into the differences in topography that exist within this area (figure 2). The upstream areas in the south of Jakarta generally have steep slopes, whereas the downstream areas to the north are more flat. Especially these downstream areas subside due to groundwater extractions.

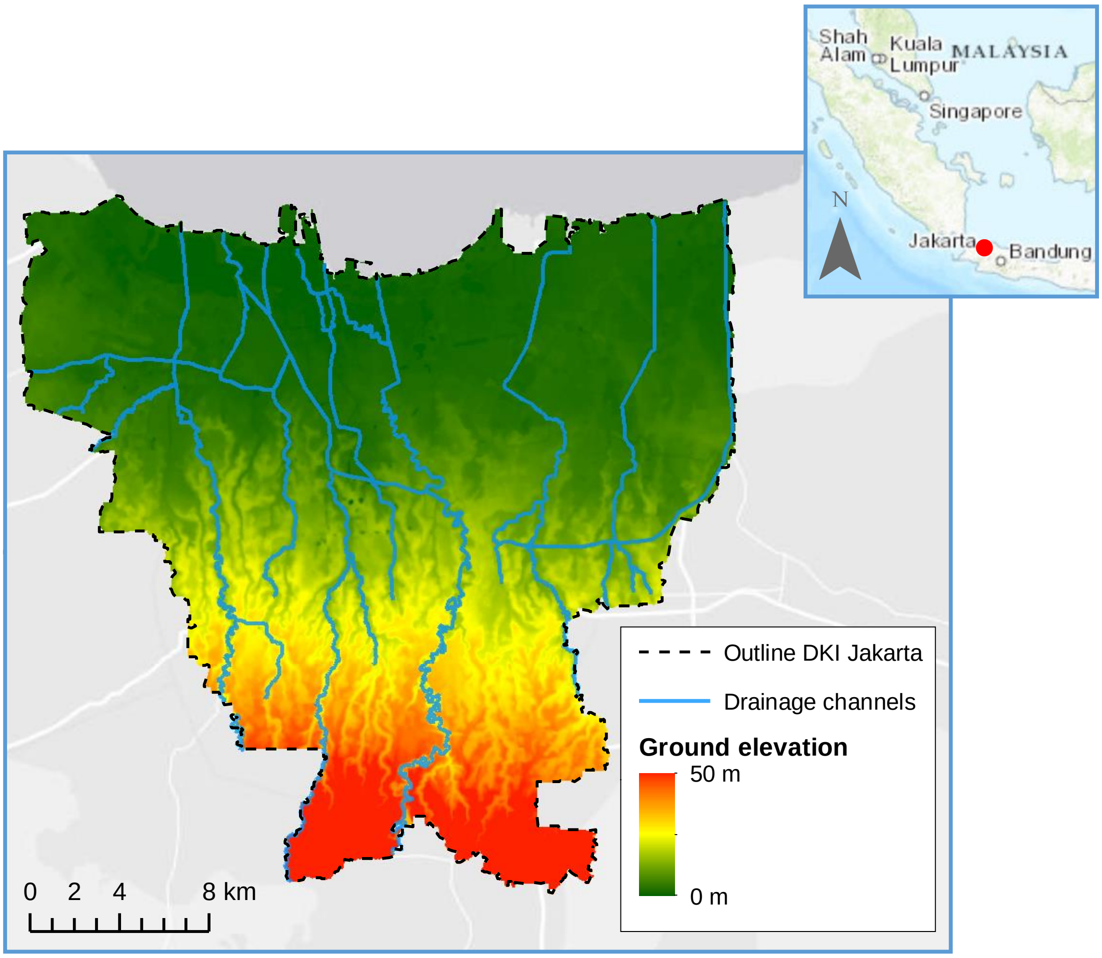

Figure 2: Jakarta - Terrain elevation and main drainage channels. The dashed line gives the boundary of the special capital region of Jakarta, which is repeated in all maps of the Jakarta case study.

This subsidence, in combination with the continued urbanization of the area and the insufficient capacity of the drainage channels causes large scale fluvial flooding in Jakarta on a regular basis (Budiyono et al., 2015). Also, the inefficient drainage system of the city makes that water cannot drain to the rivers, causing pluvial floods (Padawangi & Douglass, 2015). Both types of flooding occurred between the 8th and the 11th of February 2015, when rainfall amounts as high as 310 mm per 24 hours were measured in Jakarta. As a result nearly 5,000 houses were flooded, affecting close to 16,000 people (Davies, 2015). During the four day period over which the floods occurred, close to 29,000 flood related Tweets were generated that either mentioned Jakarta or a neighbourhood within Jakarta.

2.1.2 York

The study area in York was considerably smaller than that of Jakarta. Where the northern parts of the study area have some moderate terrain slopes, there are high ridges in the south and south west of the area (figure 3). The area is crossed by several rivers, of which the river Ouse is the largest. Near the city centre, several smaller rivers converge and eventually discharge into the river Ouse.

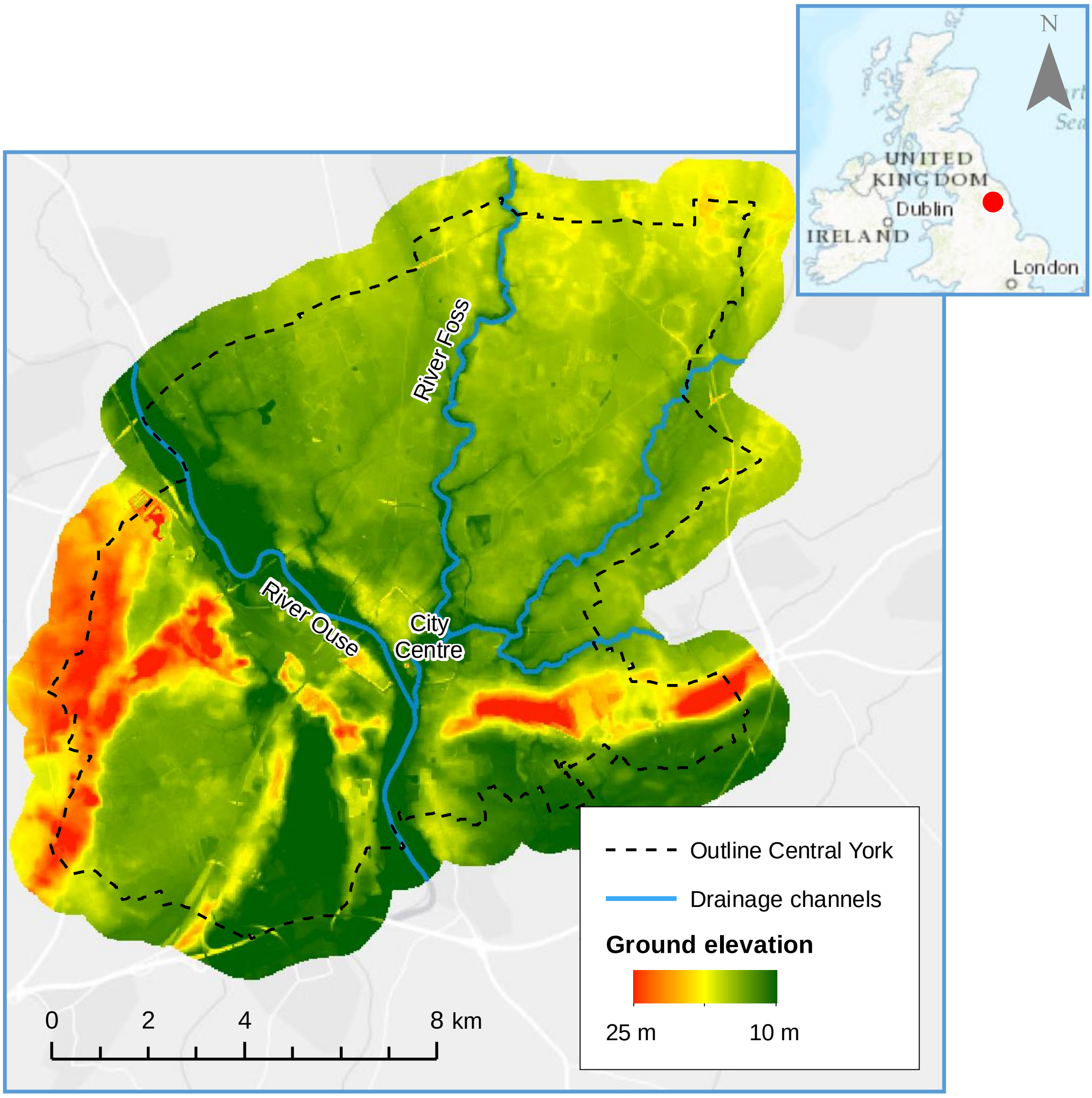

Figure 3: York - Terrain elevation and main drainage channels. The dashed line gives the outline of Central York, which is repeated in all maps of the York case study.

Especially this city centre was badly affected by the floods that occurred between the 25th and the 30th of December 2015. These floods mainly originated from the rivers and were caused by an elongated period of high intensity rainfall in the area. It is estimated that the amount of rainfall that fell over the Yorkshire area in December 2015 had a return period in excess of 100 years.(Parry et al., 2016). This rainfall caused the water level in the River Ouse to rise to 5.2 m. This is only slightly lower than the highest value ever registered, 5.4 m (EA, s.d.), which caused widespread flooding in the year 2000. The 2015 flood also affected a large area within York and submerged almost 600 homes and businesses (Stott, 2016). Over the course of the flood, 38,000 flood related Tweets were produced that either mentioned York specifically, or contained the hash tag ‘Yorkfloods’. Although this is a higher number of Tweets than for Jakarta, they were produced over longer period.

2.2 Materials

To investigate each case study, multiple datasets were used. This paragraph discusses these datasets and the processing steps applied to them.

2.2.1 Elevation data

For both case studies elevation datasets were collected. Although medium resolution datasets from earth observing satellites or spacecraft are widely available, such as 30 metre resolution SRTM[1] or ALOS World 3D[2] data, the quality of these was considered insufficient for flood mapping purposes. Both of these datasets are also digital surface models (DSMs), which contain information about ground, rooftop and canopy elevation, whereas the water depths mentioned in the Tweets are specified relative to ground levels only. Due to the low resolution of these datasets, small areas that purely consist of ground level, for example streets, cannot be distinguished from rooftops. This means that if these medium resolution DSMs are used in flood mapping, ground levels are overestimated, which can lead to large errors in flood extent.

Although high resolution DSMs can be used to distinguish between ground level and rooftop elevation, creating flood inundation maps at very high resolution can be cumbersome, since it requires considerable computational time. Therefore a digital terrain model (DTM) was used. This research used a resolution of 20 m, since it was found that this resulted in reasonable calculation time (about 1 minute) in the Jakarta case study. For this case study a DTM was derived from a 2 m LIDAR DSM (see appendix A). For the York case study a 2 m DTM of the Environment Agency was simply resampled to 20 m resolution. To identify areas susceptible to pluvial flooding, topographical depressions were derived from the DTMs of each case study.

As a last step the DTMs were used to create a height above nearest drainage (HAND) map. Therefore the DTM, in which all depressions were filled, was used to determine the local drainage direction of each cell in the area. These drainage directions were then used to calculate the number of cells draining on each of the cells in the area, termed accumulated flow. The locations of drainage channels in the area were determined using a threshold, since drainage channels generally have a high accumulated flow value. This threshold was adjusted to accurately represent the actual drainage channels in the area. The elevation of all cells on these channels was set to zero. Each of the cells draining on such a drainage channel cell was assigned an elevation value relative to the original elevation of the cell it drained on, which meant each cell contained a HAND value. Since topographical depressions in the landscape can be important in explaining flood patterns, they were again subtracted from the HAND map. The HAND principle is described in detail by Norbre et al. (2015), and the full process used to develop the elevation datasets for both the Jakarta and York case studies can be found in appendix A.

2.2.2 Twitter datasets

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 (figure 4). As a first step the Tweets from the 8th to the 11th of February 2015 were selected for Jakarta and the Tweets from the 25th until the 30th December 2015 for York. These time periods were selected by looking at measured river water levels (see appendix A) and during which time period news reports were generated. The second step was to select only the messages that referred to the Jakarta and York floods respectively. For Jakarta this meant Tweets were selected that either mentioned Jakarta specifically, or mentioned neighbourhoods within Jakarta. For the York case study only messages were selected that mentioned York specifically, or contained the hash tag 'Yorkfloods'. As a last filtering step, only the messages that referred to neighbourhoods, streets or points of interest (POIs), such as market places, schools or railway stations, were included in the dataset. Geo-tags were intentionally not used, since previous investigations have indicated that these often refer to erroneous locations. Often people do not report the floods when actually being in the affected area, but reports are generated from outside the area.

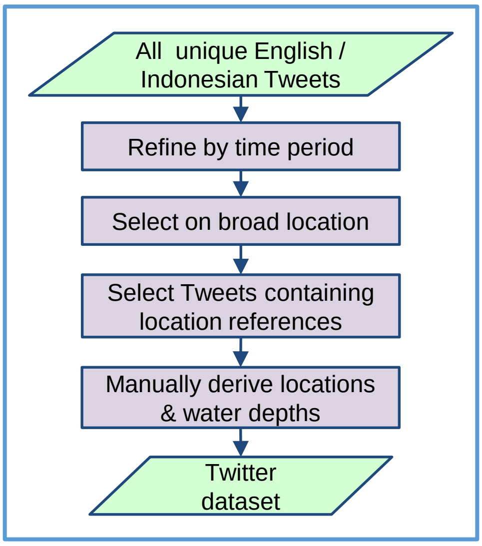

Figure 4: Main steps in creating the Twitter datasets

1 SRTM: Shuttle radar topography mission
2 Elevation dataset disseminated by the Japanese space agency

For the Jakarta case study these filtering steps led to a database containing up to 1000 useful Tweets. Because a manual optimization was used to derive locations and water depths (see paragraph below), this dataset was additionally filtered to obtain a smaller dataset that could be analysed within a reasonable amount of time. The dataset for the York case study was already small enough and was therefore not additionally filtered. Details of the filtering steps applied to both datasets can be found in appendix A.

All filtering steps led to a database of 219 flood related Tweets for the Jakarta case study, of which 117 mentioned a water depth (figure 5). For the York case study only 87 Tweets were found, none of which mentioned a water depth (figure 6). This is an important difference with the Jakarta case study, and is most likely caused by initiatives like PetaJakarta who actively engage with Twitter users in Jakarta, asking for detailed information about the flood. The difference in purely the number of Tweets between both cases however is hardly surprising, since the York study area was considerably smaller, and less densely populated.

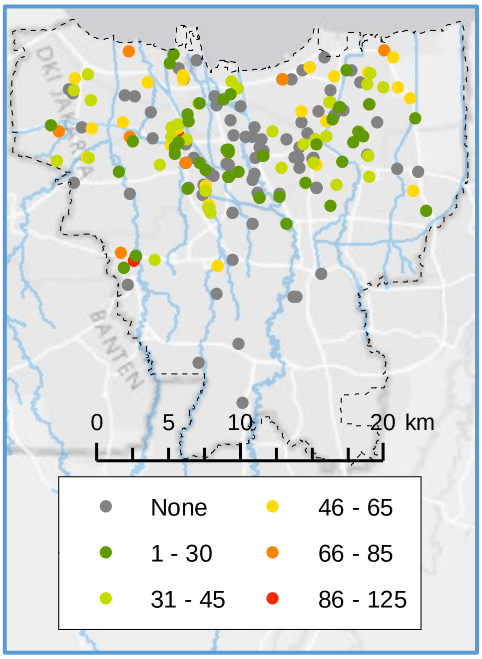

Figure 5: Jakarta – Water depths (cm) of Twitter observations

Manual optimization

To use the collections of Tweets for creating flood maps, both exact locations in the form of geographic coordinates and water depths were manually derived from the Tweets.

Locations were derived by searching for the exact phrases from the Tweets on Google Maps. In case the message mentioned a POI, the exact coordinates of this POI was used. However, if only streets or neighbourhoods were mentioned by the message, additional processing steps were required, since these refer to larger areas and not to point locations.

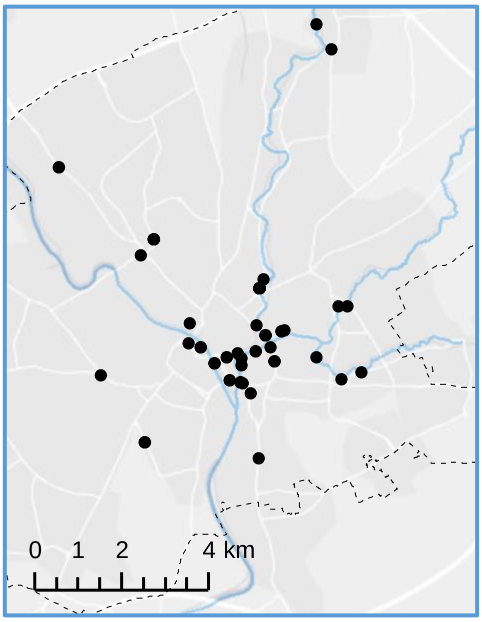

Figure 6: York - Locations of Twitter observations

This was done using the elevation data, from which the locations and depths of topographical depressions in the area were derived. For Tweets that specified entire streets (line elements) or neighbourhoods (polygon elements), the deepest depression on the element was used as the exact location of the Tweet. In case no depressions were found on the element, the location of lowest elevation was used. Since some Tweets mentioned more than one location of flooding, the total number of observations amounted to 233 for the Jakarta case (derived from 219 Tweets) and to 89 for the York case (from 87 Tweets).

Also water depths were derived from the Tweets. In case a Tweet specified a range of water depths (e.g. 20 – 40 cm), the average of this range was used. By performing these steps, databases were created for both cases in which all observations contained exact locations and of which a portion also contained water depth information.

Reference information

Since the quality of both the locational reference and the mentioned water depth were of interest, the actual water depth and location of some of the Tweets were derived by using the photographs attached to these Tweets. When possible the exact location of these photographs was determined using Google Streetview (figure 7). Also water depths were estimated by looking at the flood extents at fixed points in the photograph and comparing these to the Streetview reference, in which no flooding was present.

Figure 7: Example of deriving locational information from the photographs using Google Streetview

2.2.3 Validation data

Additional datasets were used to validate the flood maps. Although a variety of satellite data sources can be used to distinguish between flooded and non-flooded areas, these datasets pose other problems. Some of them for example they do not work in cloudy conditions, which are often present during floods. Since radar data does not have this limitation, the availability of Sentinel-1 radar data, which can be used for this purpose (e.g. Twele et al., 2016), was reviewed. For both case studies however, the temporal availability of data did not coincide with the actual flood peak.

Because of these issues other datasets were used for validation. Flood maps for the Jakarta case study were validated using a collection of photographs attached to Tweets. A set of Tweets completely separated from the input dataset discussed in paragraph 2.2.2, was composed for this purpose. Flood related Tweets having photographs attached were selected over the time period of the Jakarta flood. Using the procedure of adding reference information discussed in paragraph 2.2.2, locations and water depths were added to each of these photographs. In this way a set of 75 validation points at which flooding occurred was constructed.

To validate flood maps for the York case study, a dataset containing the recorded flood extents from the 2015 floods, supplied by the Environment Agency (EA), was used. Still being a draft and only containing information on fluvial flooding, this dataset on its own was not suited for validation purposes. Therefore an additional dataset from the EA containing historic flood extents was used. To also evaluate the flood extent in areas flooded separately from the river, areas that were flooded separately from the river over the period 1991 – 2012 were added to the draft dataset of the recorded flood extents of 2015. The resulting dataset is given in figure 8. The detailed process used to construct this dataset is discussed in appendix A of this report.

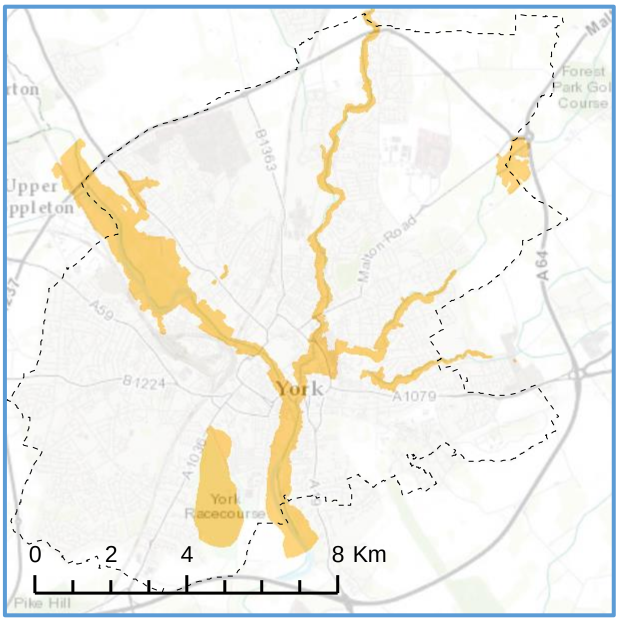

Figure 8: York - Areas identified flooded (orange) in the combined dataset of recorded flood extents by the EA

2.3 Methods

The methods used in the research are closely linked to the research questions from paragraph 1.3. To review the potential of Tweets in real-time flood mapping and assess the uncertainties in these Tweets, the first step of the research was to review the Twitter datasets (RQ3 & 2, paragraph 2.3.1). This analysis was followed by a comparison of a number of different interpolation methods which gave an indication of how flood mapping could be improved (RQ 1, paragraph 2.3.2). The flood mapping method that produced the best results was used to assess the uncertainties in the flood maps (RQ 3), which is discussed in the last paragraph of this chapter (2.3.3).

2.3.1 Dataset characteristics and uncertainties

The dataset of Tweets was investigated to see whether the datasets could be used to create real-time flood maps and to review the uncertainties in the Twitter data. The outline of this process is given in the flow chart in figure 9.

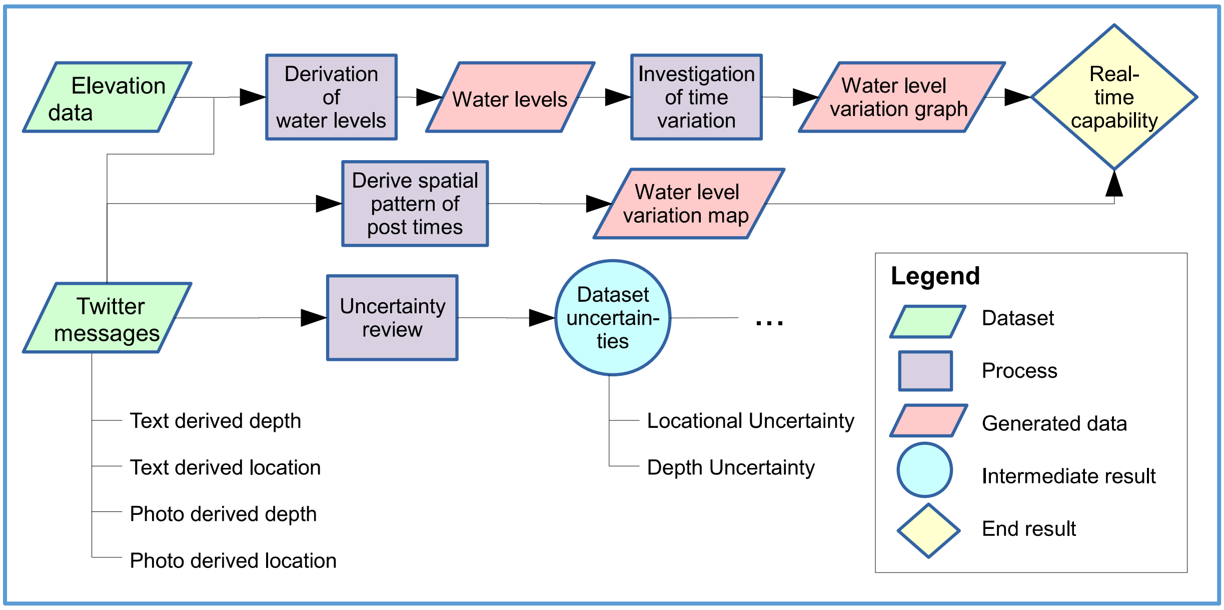

Figure 9: Flowchart for the investigation of Twitter datasets

The main outcomes of the investigation of the Twitter dataset were:

- ✓ An insight into whether the dataset can be used for real-time flood inundation mapping (RQ 3)
- ✓ A quantification of the uncertainties in location and water depth derived from the messages; an intermediate result since it was used as input for the uncertainty analysis (paragraph 2.3.3)

The processes used to produce these results, are shortly discussed in the following paragraphs.

Derivation and analysis of water levels

The analysis started with the derivation of water levels from the Tweets. Water levels were reviewed rather than solely water depths, since ground levels cause the water depths in one continuously flooded area to vary by location. The water depths in the middle of a flooded area for example are generally higher than the water depths at the edge of such an area, though the water level throughout the area is more or less the same. Just like in the work of Fohringer et al. (2015) water levels were derived by adding the water depth of each message to the local ground elevation. In case messages specified not a single, but a range of water depths, the average water depth was added to the local ground elevation.

To review the time variation in the Twitter datasets, the changes in reported water levels over time were reviewed. These time variations were compared to measured river water levels, to see whether the variation in water level was reflected in the Twitter messages. Reviewing a cluster of observations rather than all observations in the area is most appropriate, since water levels within the same flooded area are the same, but can be very different at other locations. Although reviewing a cluster of observations, water levels of observations that are made somewhat further upstream can still be different from downstream water levels. To reduce the impact of such differences in the analysis, the water levels were calculated with respect to the HAND map and were consequently relative to the nearest drainage channel.

Spatial pattern of post times

The changes in the spatial pattern of observations over time can also give an insight into the time variations in the Twitter dataset. Such variation might for example be represented if clusters of observations are posted at roughly the same time. Also this variation can be accurately represented if observations in close vicinity of one another have similar water levels at roughly the same time. If observations appear on the map very scattered however and there are large fluctuations in water levels of observations in close vicinity of each other within the same time period, it is less likely the data can be used to create real-time flood maps.

Uncertainty review

The analysis of the dataset was concluded by evaluating the errors in the locations and water depths that were derived from the Twitter messages. Although being uncertain, the timing of the observations was not reviewed, since no reference information was available about the actual time a message was posted.

Reference information was available for both the location and water depths reported by Tweets. These errors were evaluated by comparing the locations and water depths that were derived from the text of the Tweets, with information derived from photographs that were attached to a portion of the Tweets. From these photographs the actual location of the Tweet was derived and a water depth was estimated (see paragraph 2.2.2). The locations and water depths derived from the photograph were compared to the locations and water depths derived from the Twitter message. Thereby the errors in location and water depth were calculated, and information about the magnitude of these errors was used to derive the error probability distributions applied in the uncertainty assessment (see paragraph 2.3.3).

2.3.2 Flood mapping

Several methods of interpolating water levels were compared to see what method performed best. The resulting flood maps were evaluated using the validation data discussed in paragraph 2.2.3. An overview of these steps is given in the flowchart in figure 10.

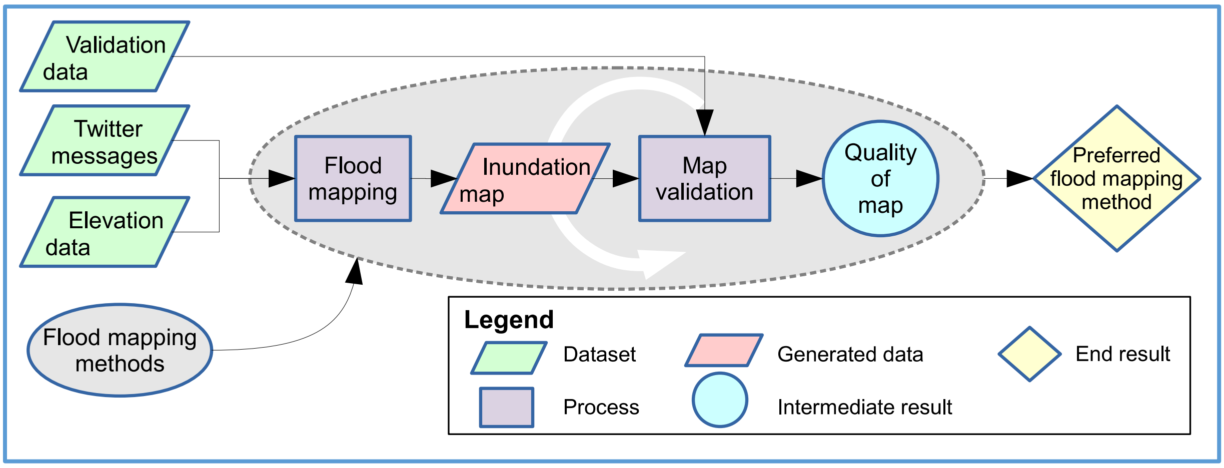

Figure 10: Flowchart of Flood mapping process

The main goal of this comparison of mapping methods was:

- ✓ To define a preferred flood mapping method (RQ 2), which is used in uncertainty propagation

The different methods used for creating flood inundation maps, as well as the methods for validating the created flood maps are discussed in more detail below. The exact implementation of each of these methods can be found in appendix A of this report.

Interpolation methods

Previous investigations that reviewed social media derived flood maps utilized a variety of different mapping techniques (see paragraph 1.2). The method by Schnebele et al. (2014) however only took into account the proximity of observations and did not use elevation data to derive flood extent estimates. The mapping method discussed by Eilander et al. (2016) on the other hand did use elevation data, but assumed only topographic depressions were flood prone and was therefore only applicable to pluvial floods. Only the method discussed by Fohringer et al. (2015), which involved a simple interpolation of water levels, was expected to be applicable to different types of floods and was therefore used as a starting point. Different ways of using interpolation to create flood maps were considered in this research. To limit the number of methods to be reviewed, several requirements were defined with respect to the methods:

- Methods should work using only one single observation. This ensures the method can be applied in data scarce areas.
- Methods should utilize both observations that mention a water depth, as well as ones that do not. This is important since many Tweets omit information about water depths.
- Methods should require limited computational time. Since many runs have to be performed to assess the uncertainties in the maps (see paragraph 2.3.3), this time should be limited to one minute.

- Methods should require a limited number of input parameters. This research is aimed at finding a generally applicable method to create flood maps and having to change large numbers of parameters between cases is undesirable.

The first improvement with respect to the straightforward interpolation method used by Fohringer et al. (2015) is the exclusion of flooded areas that are not directly connected to the observations in the area. By simply interpolating water levels and subtracting ground levels, areas that are separated from observations, for example by a higher lying area, are still identified as being flooded. In reality these areas are physically separated from the flooded area, meaning they will not be flooded. Therefore the methods used for this research identify only flooded areas that are directly linked to an observation. This principle is illustrated in figure 11. In this hypothetical situation the observation (black dot) has a water level of 10 m. If this is simply extrapolated, the area on the other side of the ridge is mapped flooded (figure 11, centre). However, by removing this area, since it is physically separated from the observation, the flood extent estimate is more realistic (figure 11, right).

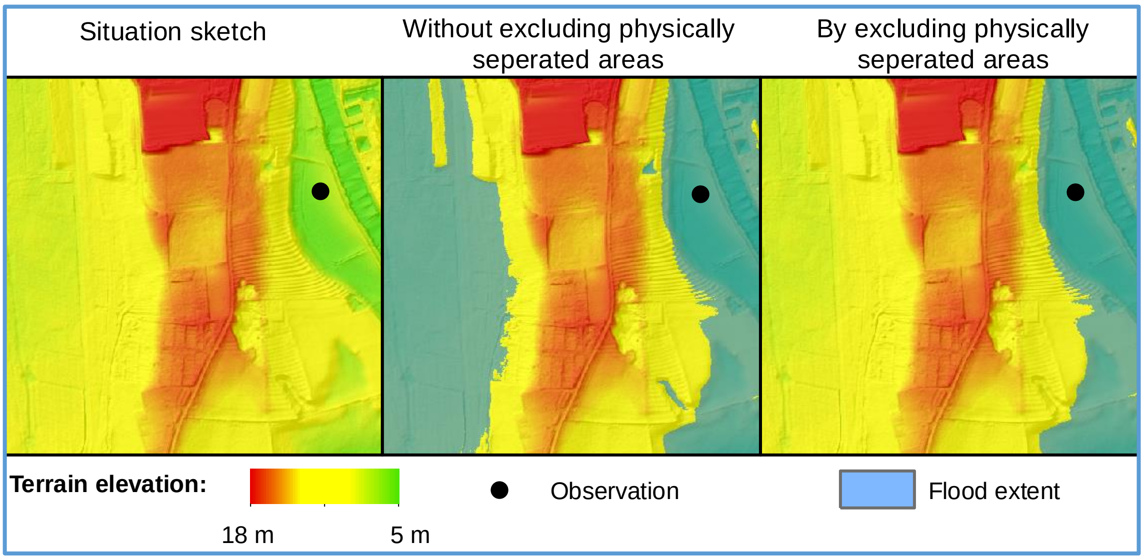

Figure 11: Hypothetical example of excluding flooded areas, with a situation sketch (left), the situation in case physically separated areas were not removed (middle) and the same map in case physically separated areas are removed.

The second possible improvement over a simple interpolation of water levels is the grouping of observations to identify observations that belong to the same flooded area. For example if two areas are separated by a ridge (like in the hypothetical example of figure 11), water levels on one side of the ridge are unlikely to directly influence water levels on the other side of the ridge. By simply interpolation using all observations, observations in one area can influence water levels in another area. Since this is unrealistic, two grouping methods were evaluated in this research:

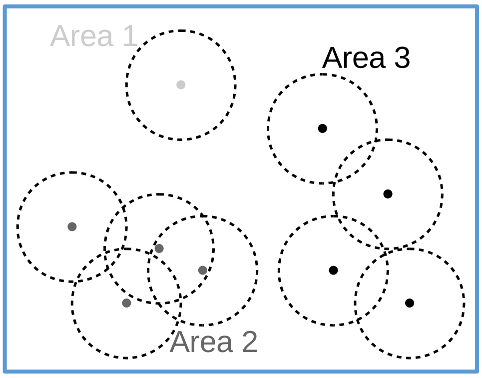

Figure 12: Grouping observations based on proximity

- *Grouping based on proximity*. Observations are clustered based on vicinity, defined by a search radius around each of the observations (figure 12). This method assumes observations close to each other are linked by a continuously flooded area.
- *Grouping based on downstream flow paths*. Observations are clustered based on an intersection of drainage paths downstream of the observations. If a cell is flooded, its downstream neighbours are also flooded, since they are situated lower than the flooded cell and are directly connected to it. Therefore it is likely observations of which the flow path ends up in the same cell, belong to the same continuously flooded area (figure 13)

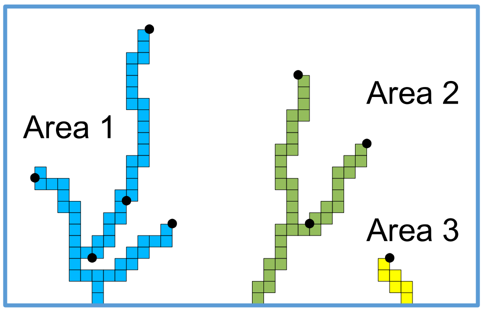

Figure 13: Grouping based on downstream flow paths

The flood maps can also be improved by using observations that do not specify a water depth in addition to the observations that do specify a water depth. As stated at the beginning of this paragraph, the research aimed at using both observations that mention a water depth, as well as observations that do not. Three different ways of using these 'depthless' observations were considered by the research:

- *Assigning all 'depthless' observations, a single default water depth (DWD)*. This is the most straightforward method of including 'depthless' observations in the interpolation procedure. However this causes the DWD of 'depthless' observations that are located close to observations with water depths, to influence the water levels around these observations with water depths. This basically means the precise information which is available at these locations, it deteriorated by the assumed water depth added to the 'depthless' observations. In case for example an observation which specifies a water depth of 120 cm is very close to a 'depthless' observation and a DWD of 30 cm is used, the water levels in this area are artificially drawn down by the DWD specified for the observation without water depth, and the more accurate information of the observation with water depth is ignored.
- *Assigning only 'depthless' observations a DWD if they are close to the flood extent determined using only observations with water depths*. By doing this it is ensured that water levels in an area where observations with water depths are available are hardly affected by the artificial DWD. This means only observations are used that 'add information'. Also less reliable observations which are very distant from the flood extent calculated using observations with a water depth, are excluded from the analysis.
- *Assigning only 'depthless' observations that are clustered a DWD if they are close to the flood extent determined using only observations with water depths*. In this case additionally only clusters of observations are used, since when multiple observations are made in close vicinity, it is more likely the area is really flooded. Again this ensures only observations are used that 'add information'.

Also important for the process of creating flood maps, is the elevation dataset used. Besides water levels being calculated by adding the water depths to local ground elevation values, the dataset is also important in constraining the flood extent. A HAND map was used in this research, instead of a DTM (see paragraph 2.2.1). In this HAND map, river slopes are basically filtered out. Since these are filtered out, water levels in upstream regions are not necessarily higher than in downstream regions, meaning that if a high number of upstream observations are available, a downstream overestimation of water levels is less likely using a HAND map.

In contrast to Fohringer et al. (2015), who used bilinear spline interpolation, inverse distance weighting (IDW) was used to interpolate the water levels derived using the HAND maps. IDW was used since 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. This last feature is especially useful in real-time applications. The smoothing variant of IDW used in this research is given in Equation [1]:

$$\mathrm{Z_{interp}} = \frac{\sum_{1}^{\mathrm{n_{obs}}} \mathrm{Z_{obs}} * \mathrm{W}}{\sum_{1}^{\mathrm{n_{obs}}} \mathrm{W}} \qquad [1]$$

$$\mathrm{W} = \frac{1}{\left(\sqrt{\Delta \mathrm{x}^2 + \Delta \mathrm{y}^2 + \mathrm{s}^2}\right)^{\mathrm{p}}} \qquad [2]$$

With:

Z_{interp}: Interpolated water level (m)
Z_{obs}: Observed water level (m)
n_{obs}: Number of observations (-)
p: Power (-)
Δx: x-distance from observation (m)
Δy: y-distance from observations (m)
s: Smoothing factor

This interpolation method was applied in several ways in this research:

- *Plain interpolation using all observations.* This is the most straight forward interpolation method and uses none of the grouping methods discussed above. Only flooded areas which are not physically connected to any of the observations are removed. This method assumes all water levels in the area are related.
- *Interpolation of groups of observations.* Observations are first divided into groups, based on the flooded areas they belong to. For each of these groups water levels are interpolated in a rectangular extent around the area. Flooded areas which are not physically connected to any of the observations are excluded. Both grouping by vicinity as well as grouping based on downstream flow paths were evaluated for this interpolation method.

- *Interpolation along flow paths.* Observations are first grouped, 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. For this last method, equation [2] is implemented differently, since the distance along the flow path was calculated without calculating the differences in X and Y coordinate used to determine distances as the crow flies (see Appendix A).

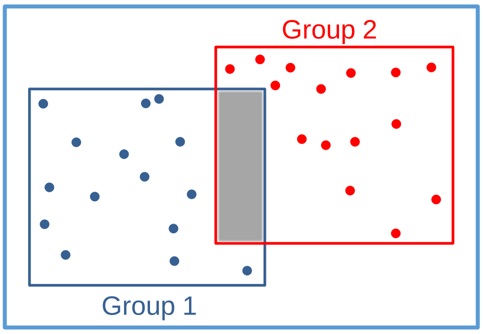

Figure 14: Using interpolation of groups of observations, interpolation extents can overlap (grey area); meaning water levels of the two areas can influence each other.

Both the second and third interpolation methods aim at separating groups of observations. However, with the second method, rectangular extents around the observations are used to interpolate the water level. As illustrated in figure 14 this can still cause overlaps between areas and therefore does not strictly separate the groups. The third method however only assigns the water level of an observation to the cells that are upstream of each cell on the flow path downstream of the observation. Therefore the water levels of an observation can only affect locations within its own subcatchment. This means that the water levels in different areas are strictly separated using the third method.

Each of the three methods given above contains one or more of the improvements discussed in this paragraph. For every method the interpolation parameters (power & smoothing, see equation [1] & [2]) were varied to find an optimal combination. For every method also the different ways of using 'depthless' observations and different settings of the DWD were evaluated. The differences between the two grouping methods were only reviewed for the interpolation of groups of observations. An overview of the applied methods is given in table 1.

Table 1: Overview of mapping methods

Method	**Grouping**	**Interpolation procedure**	**Exclusion of areas**
Plain interpolation	None	IDW, using all observations	Areas not connected to observations
Interpolation of groups	Based on vicinity or intersecting downstream flow paths	IDW, for each group separately, results merged	Areas not connected to observations
Interpolation along flow paths	Based on intersecting downstream flow paths	IDW along flow paths, these values are spread over the subcatchment of each location on the flow path	Areas not connected to downstream flow paths of observations

Map validation

As discussed in paragraph 2.2.3 two very different validation datasets were used for the Jakarta and York case studies. Therefore different validation methods were used for the cases.

For the Jakarta case study actual locations that were flooded and estimates of water depths were derived from photographs attached to Tweets. The flood maps of the Jakarta case study were validated using these points. Just calculating the percentage of points that is within the calculated flood extent is however not sufficient, since overestimations are not assessed this way. Mapping the entire area flooded would yield a score of 100%, although the flood map would be useless. Therefore the number of correctly predicted validation points per km^2 of flooded area was calculated to review the relative performance of the maps. Using the estimated water depths from the pictures also the root mean square error (RMSE) and the mean error (ME) were calculated, to examine whether water depths were correctly represented.

For the York case study actual data was available about flood extents in the area. Therefore a different validation method was used for this case. The flood extent data, supplied as shapefiles, was first converted to raster data at the resolution of the flood map. For each cell in the area, this dataset was compared to the created flood map. The quality of the map was assessed using the F-measure, discussed by Aronica et al. (2002):

$$F = \frac{A_{obs} \cap A_{mod}}{A_{obs} \cup A_{mod}} \tag{3}$$

With:

$A_{obs} \cap A_{mod}$: *Area that is flooded in both in the created flood map and in the validation data (km^2)*
$A_{obs} \cup A_{mod}$: *Area that is flooded either in the created flood map or in the validation data (km^2)*

Using this statistic both the area correctly mapped and the overestimation of flooded area were taken into account. For both case studies the preferred interpolation method and combination of parameters was selected based on the values of these performance measures. This preferred interpolation method was then used in uncertainty propagation.

2.3.3 Uncertainty assessment

The last step in the research was assessing the uncertainties in the flood inundation maps that were created using the preferred mapping method. Therefore the following sources of uncertainty were considered in this research:

- Uncertainty in water depth, for example introduced by inaccuracy of the observer or a wrong interpretation of the text of a Twitter messages (for example interpreting the M of million, when flood damage is reported, as metre).
- Uncertainty in location. This can be due to errors by the observers, ambiguity of the location specified or the failure to specify an exact location (e.g. mentioning a street or neighbourhood).
- Errors in the elevation data, which can be due to errors made by the sensors used to collect the data, or human errors in data collection or processing of the data.

Where the characteristics of the first two sources of error have already been reviewed in the analysis of the Twitter dataset (paragraph 2.3.1), the errors in the elevation models cannot be derived without having accurate measurements of actual ground elevation in the area. Also for both the Jakarta and York elevation datasets no elaborate validation report was available. Therefore the errors in these DEMs were quantified using values from literature. Additional sources of error can be identified, such as the error introduced by misinterpretation of the Twitter messages, coarse resolution of the DTM, errors introduced by the interpolation procedure itself or temporal uncertainty in case real-time maps are created. Additionally the choice of interpolation parameters can seriously affect the results. Although errors introduced by misinterpretation of the Twitter messages are already included in the locational and water depth errors, the remaining error sources can be of importance. Therefore the effects of using different DWDs and grid resolutions were also investigated.

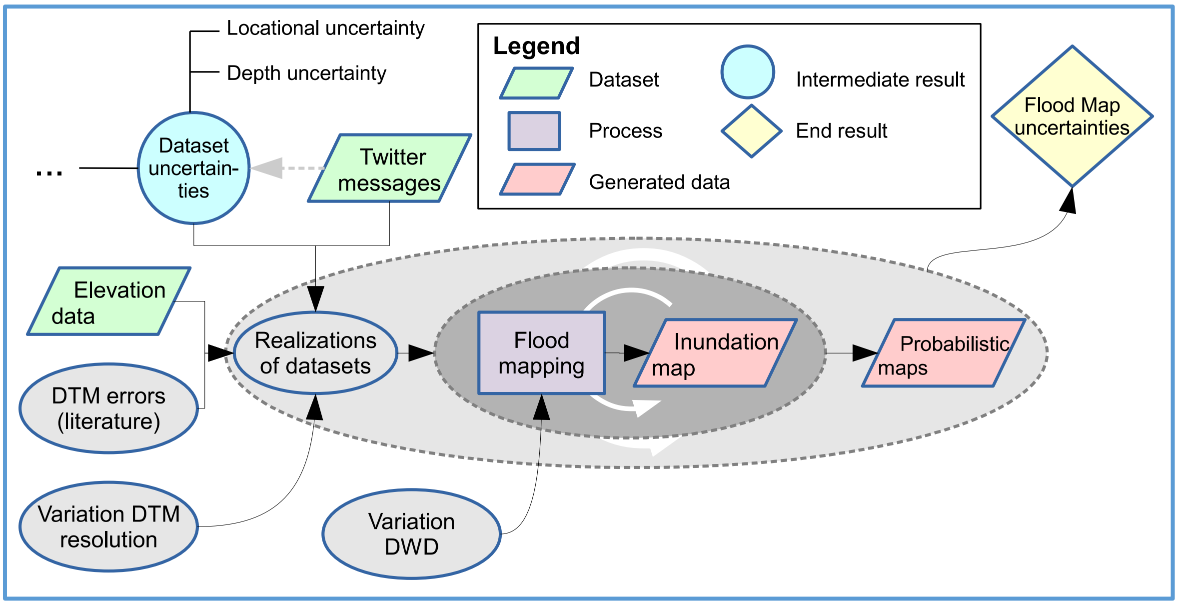

Figure 15: Flow chart of the uncertainty assessment

The flowchart of the uncertainty assessment is given in figure 15. For the uncertainty analysis the three aforementioned main sources of error were used to generate a number of random realizations of the input datasets. Using these random realizations a Monte Carlo analysis was performed to generate a number of flood maps. These individual maps were subsequently used to calculate the probability of flooding of each cell. These probabilistic maps were created using each individual source of uncertainty, as well as a combination of all sources. For the uncertainties caused by the DWD and the resolution of the map, which could not be quantified using a probability distribution, flood inundation maps were generated by varying the DWD and resolution respectively. The main aim of the uncertainty assessment was to:

- ✓ Gain insight into the uncertainty in the created flood extents, and how each individual source of uncertainty affects the total uncertainty in flood extents (RQ 2).

Each step of the uncertainty assessment is described in more detail in the following paragraphs.

Quantification of uncertainty

In order to generate random realizations of the datasets, each of the three sources of uncertainty mentioned above were quantified. The uncertainties in water depth and location were quantified by deriving the mean error and standard deviation in error from the initial analysis of the Twitter dataset (see paragraph 2.3.1).

To specify the errors in the elevation data a similar method was used. Just using the mean and standard deviation of errors for each grid cell however does not accurately reflect the actual errors in the elevation datasets. Both Raaflaub & Collins (2006) and Heuvelink et al. (2007) recognize the importance of the spatial autocorrelation in error for generating random error fields, expressed by the autocorrelation distance of errors.

Since no specific data concerning the quality of the elevation datasets was available, typical error values for LIDAR elevation datasets were derived from literature. Leon et al. (2014) for instance use a value of 0.18 metre standard deviation for their 1 m LIDAR derived DEM. Hodgson et al. (2004) on the other hand found RMSE values in the range of 20-30 cm over different types of land, though they mention errors can sometimes be as high as 150 cm for large scale mapping operations. Since a 2 metre resolution dataset was used, which likely means the errors of multiple LIDAR points are averaged, a standard deviation in error of 20 cm was assumed in this research. It is assumed there is no consistent under or overestimation in the data, meaning that a mean error of 0 cm was used.

Only a few studies also mention the correlation distance of errors in high resolution datasets. Li et al. (2011) for example found values ranging from 173 to 253 metres and Livne & Svoray (2011) found values ranging from 143 to 178 for a 2 m DEM. On the other hand Mudron et al. (2013) found much lower values ranging from 4 metres for a 0.5 metre resolution DEM to 50 metres for a 10 metre resolution DEM. Lastly Leon et al. (2014) found a value of 102 metres for a 1 metre resolution DEM. As a rough average of this, a value of 100 metres was used in this research.

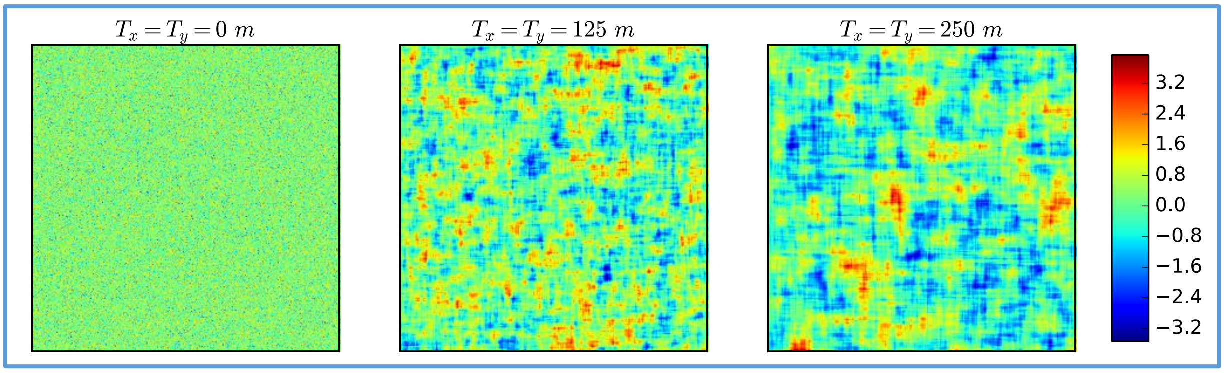

Figure 16: Example of the effects of using different correlation distances in x- and y-directions (T_x and T_y) in generating errors (μ = 0 m, σ = 1 m) for a 10 m resolution 5x5 km grid.

Generating realizations

Random realizations of all input sources were generated to be used for the Monte Carlo simulations of the flood maps. The random realizations of both the errors in water depth as well as the errors in location were created by drawing error values from a normal distribution using the parameters found in the initial analysis of the Twitter data. To simulate the uncertainty in water depth the error was added to either the single value of the water depth specified or the average water depth specified in the message. To simulate the error in location, normally distributed errors were added to the X and Y coordinates of an observation.

Generating the random realizations of the elevation data was more difficult because the spatial autocorrelation in error was taken into account. Random error fields for each realization of the elevation data were generated using the method described by Dullof & Doucette (2014) (implementation discussed in appendix A). Using the mean and standard deviation in error derived above, random error grids were created (example: see figure 16). These were added to the original LIDAR datasets. After adding the errors these datasets were converted to 20 m DTMs.

The number of simulations necessary to get a representative probabilistic flood map was determined using the method described by Raaflaub & Collins (2006), who propose to look at how many realizations are necessary to stabilize the standard deviation in the standard deviations of all realizations. For the random grids used to simulate elevation errors this means the standard deviation of each cell in the grid, using all previous realizations, is calculated. Since a normal distribution is used to generate the grids, it is expected every cell will get the same standard deviation value after a large number of simulations. To see for which number of realizations this happens, the standard deviation of the grid of standard deviations is calculated. A stabilization of this value indicates a full spectrum of values has been drawn from a normal distribution. Adding more realizations would not seriously affect the results anymore.

To determine the realizations necessary to accurately represent the uncertainty in water depth and X/Y location, a similar method was used. Instead of looking at the standard deviation of a grid cell over all previous realizations however, the standard deviation of each observation's water depth or location was calculated over all previous realizations. Again the standard deviation of the standard deviations of all observations was used to indicate how many realizations were necessary.

Probabilistic flood maps

Using the random realizations of the datasets, probabilistic flood maps were generated. This process was first performed for each individual source of uncertainty and then for the combination of all sources. A flood map was created using each realization of the datasets by applying the preferred flood mapping found by the comparison of different methods (paragraph 2.3.2). After maps were generated for all realizations of the input data, the number of simulations a cell is flooded was calculated for each cell in the study area. Thereby the percentage of simulations a cell is flooded was determined, which gives an indication of the probability a cells is inundated.

In case a cell was flooded only a very limited number of times or a very large number of times, it is fairly certain that cell is either not flooded or flooded in reality. However, about cells which are mapped flooded roughly 50% of the simulations, no statement can be made as to whether they are flooded or not. Therefore the degree of uncertainty in the probabilistic flood maps was evaluated by looking at the number of cells in the area, which were flooded in about 25 to 75% of the simulations, since these areas were considered to be most uncertain.

Variation of default water depth and resolution

Additionally the effect of using different DWDs and grid resolutions was evaluated. Since a high resolution DTM was only available for York, the effect of using different grid resolutions was only evaluated for the York case study. For both case studies the DWD was varied from 30 cm below to 30 cm above its original value. For the DTM resolution the effect of using a 6, 10 or 40 m resolution instead of 20 m was evaluated. Additionally the map with total uncertainty, created using the Monte Carlo simulations, was reconstructed using the lowest and highest values of water depth and DTM resolution.

3 Results Jakarta

The first case study discussed is the one of Jakarta, which flooded in February of 2015. This chapter contains the results of this case study and starts with a discussion of the time variation and uncertainties in the dataset of Tweets. This is followed by a comparison of a number of different flood mapping methods. The chapter is concluded by discussion the results of the uncertainty analysis. Each of these three paragraphs is directly linked to the three sections of the methodology (paragraph 2.3).

3.1 Dataset characteristics and uncertainties

The characteristics and uncertainties of the Twitter dataset were analysed to determine if it accurately reflects time variation during the floods (paragraph 3.1.1) as well as to investigate the magnitude of the errors in the dataset (paragraph 3.1.2).

3.1.1 Time variation

To review the time variations, water levels of each observation were calculated by adding the water depth mentioned in the message to the elevation value in the HAND map. Additionally, the spatial pattern of the observations over the course of the flood was examined.

Water levels

Due to the limited number of observations in the dataset, only one large cluster of observations with water depths belonging to the same subcatchment was found in the dataset (see figure 17). Sorting these observations chronologically, creating bins of 3 successive observations and plotting the average water levels and post times of these bins, produced the graph on the left side of figure 18.

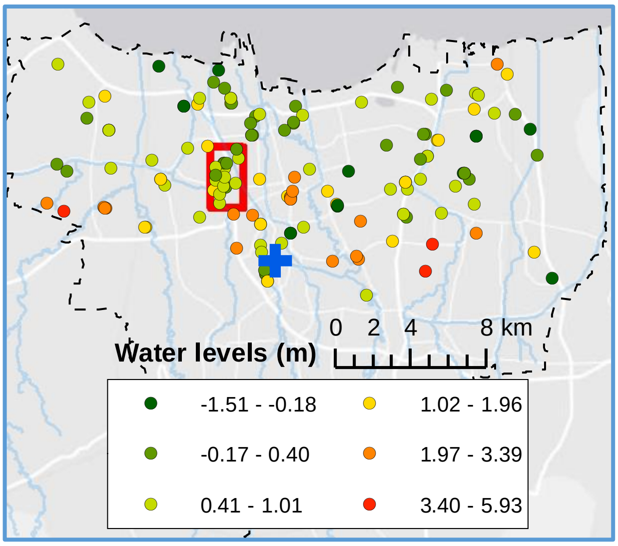

Figure 17: Jakarta - Observations and their calculated water levels. Since sinks were reintroduced in the HAND map (§ 2.2.1), some water levels are negative. The red rectangle and blue cross indicate the cluster of observations and measurement location used for figure 18.

Although a very detailed graph cannot be constructed using the cluster, due to the limited amount of observations (19), still a pattern can be derived from the graph of derived water levels. As a comparison, water levels measurements at a station about 2.5 km from the observations are given on the right side of figure 18.

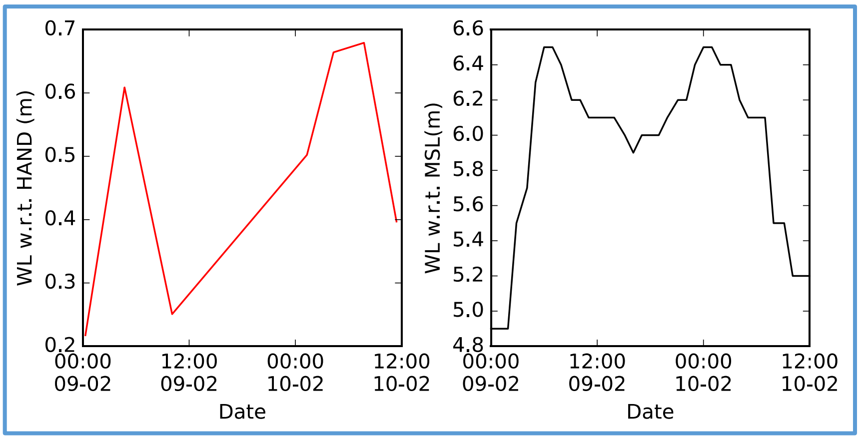

Figure 18: Jakarta - Water levels (w.r.t. HAND) of a cluster of observations averaged over bins of 3 consecutive observations (left, 19 observations total) and water levels measured at the measurement station Karet (right, BPBD DKI Jakarta, s.d.).The cluster of observations and measurement station used are indicated in figure 17 using a red rectange and blue cross respectively.

It can be seen that both graphs contain peaks on both the 9th and 10th of February. These peaks also occur in both graphs at roughly the same time. However, difference between the peaks and troughs of measured water levels is higher than that found in the graph of derived water levels. This can be due to a variety of reasons. First of all the graph on the left contains water levels which are averaged over multiple observations, causing the differences between peaks and troughs to reduce. Secondly it can be seen that already at the start of the 9th of February water levels are reported in the area, whereas the measured water level is significantly lower than the peak water level at that time. This can be the cause of local pluvial flooding in the area due to high intensity rainfall. It is likely the area only later flooded directly from the river, meaning that only close to the peaks, the water levels from the river are relevant. However, to see if time variation is consistently reflected by the dataset, the result should be confirmed by observations from other locations in the area. Unfortunately no further dense clusters of observations with water depth, on which a similar analysis could be performed, were found in the dataset. To make a definite statement about the variation in water levels in the Twitter data, a larger dataset should be reviewed, in which multiple dense clusters of observations with water depths are present.

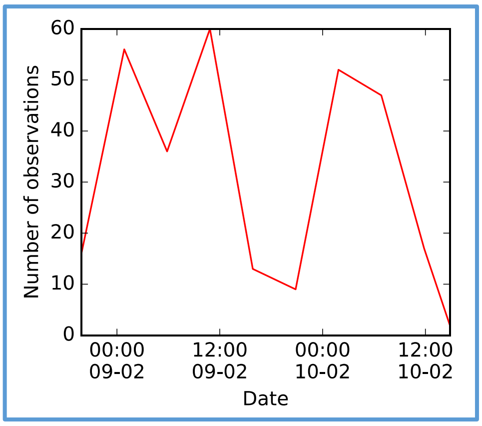

Figure 19: Jakarta - Number of Tweets in time (UTC, using all tweets with locations, assigned to intervals of 5 hours. local time: UTC + 7)

Number of Tweets

Besides the time variation in the water levels reported by the Tweets, the number of Tweets that are posted also varies over time. Figure 19 shows the number of Tweets over 5 hourly intervals. Again the two measured water level peaks (Figure 18 right) are clearly reflected by the dataset. That on the 9th of February actually two peaks are seen, might be the result of people generating less Tweets at night. Just like with the derived water levels however, the number of Tweets about flooding is already rising on the 9th, prior to the moment the measurements indicate the water levels start to rise, which is an indication that direct pluvial flooding affected the area before the flooding from the rivers occurred.

Spatial pattern

Besides time variation in water levels reported by the Tweets or the sheer volume of Tweets, also the spatial pattern of the observations is important in creating flood inundation maps. Therefore the changes in spatial pattern of the observations in time were reviewed. In case clear clusters of observations are visible in certain time steps and water levels of observations in close vicinity are more or less equal within a time step, the dataset might contain real-time variation. Just like for figure 19, Tweets were divided into 5 hourly intervals. The result of mapping all downstream observations that mentioned water depths is given in figure 20.

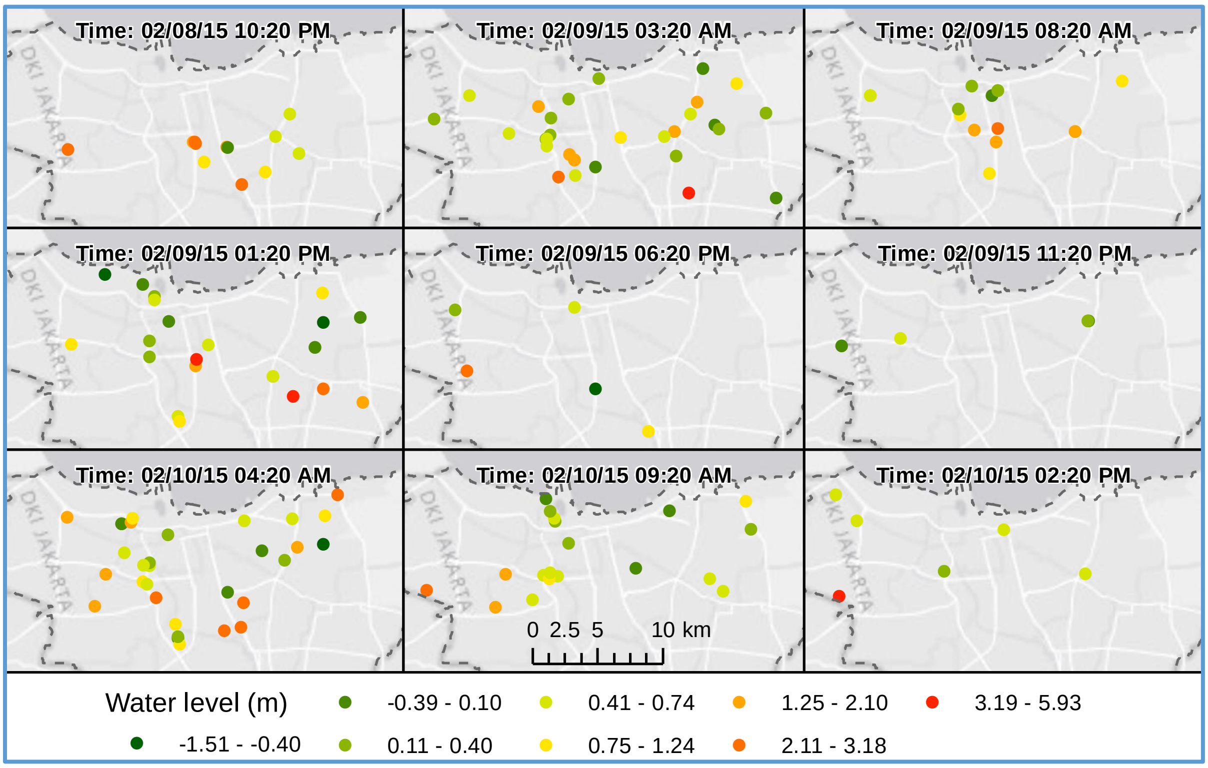

Figure 20: Jakarta - Water levels of observations, binned over 5 hour intervals (time in UTC given is the end of the interval, local time: UTC+7). Since sinks were reintroduced in the HAND map (§ 2.2.1), some water levels are negative.

In most time-steps no real clusters of observations are found. Although some observations are made in close vicinity, often more observations are made at the same location in other time steps. This however can also be an indication that an area was flooded over a longer time period. Also observations that are made in close vicinity of each other often have similar water levels in the same time steps. However some deviations can be seen, especially in the second and seventh time step (02/09/15 03:20 AM and 02/10/15 04:20). This is an indication that at times the water level changes significantly, the interval of 5 hours might be too long, meaning that water level differences within this time interval are large. For the graph of river water levels (figure 18, right) it can be seen that in the early morning of the 9^{th} and the 10^{th} river water levels rose and dropped respectively. Since only about half of the observations for the Jakarta case mentioned a water depth, a review of solely the spatial pattern was also done using all observations in the dataset and smaller time intervals. This analysis however did not generate any new insights (see appendix B).

3.1.2 Uncertainties

The errors in the dataset were reviewed in order to assess the uncertainties in the flood maps. Only Tweets were considered of which a location and/or water depth could be derived both from the text of the message as well as from an attached photograph. Comparing both gave some indication of the accuracy of the text derived properties.

Location

Deviations between text and photo derived locations were quite large, as is illustrated by a review of both the distances between the points as well as the errors in X/Y coordinates (figure 21).

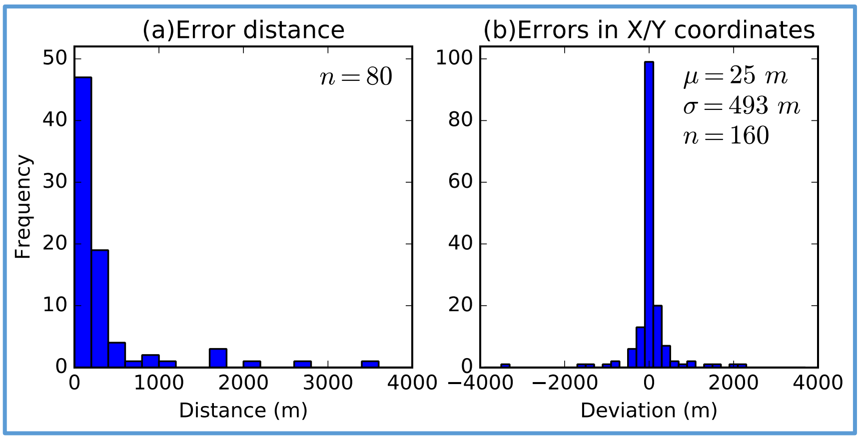

Figure 21: Jakarta - Locational errors of Tweets (Complete dataset), bin size: 200 m

Although the majority of the Tweet's locations were pinpointed (using the text) within 200 m of their actual (photo derived) location, large outliers of up to 3.5 km were also found in the dataset. Specifically looking at these outliers indicated that these observations generally referred to streets or neighbourhoods. To review this further, the dataset was split up based on the locational references that were found in the Tweets.

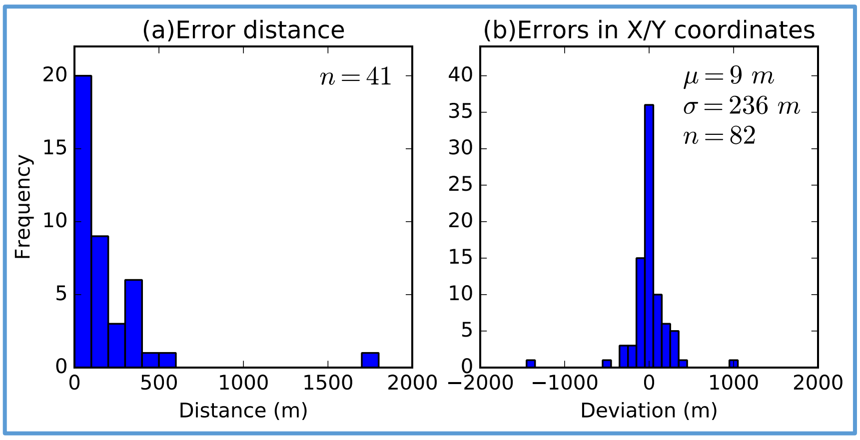

Figure 22: Jakarta - Locational errors of Tweets mentioning POIs, bin size: 100 m

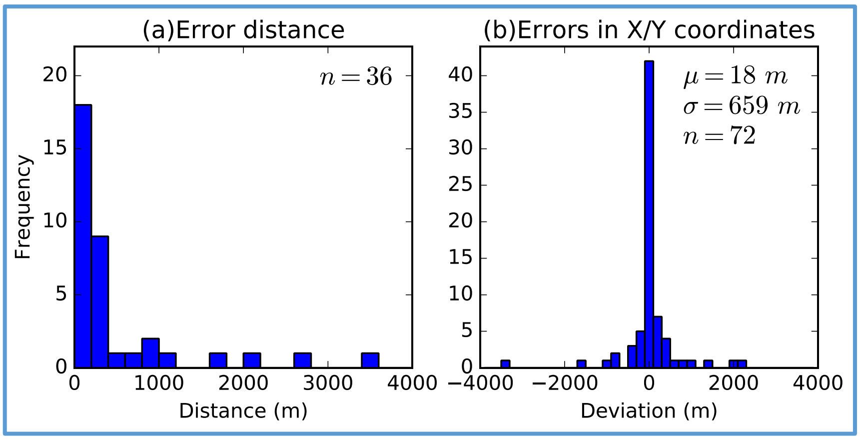

Figure 23: Jakarta - Locational errors of Tweets mentioning streets, bin size: 200 m

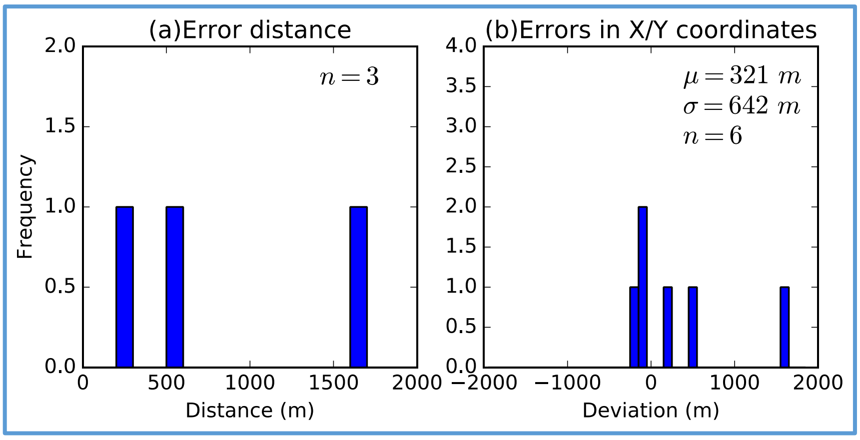

Figure 24: Jakarta - Locational errors of Tweets mentioning neighbourhoods, bin size: 100 m

Figures 22 to 24 illustrate that locational errors remain limited for Tweets that mention POIs, with only one Tweet having a deviation larger than 600 m. The locations of Tweets referring to streets or neighbourhoods were generally more uncertain. Although the number of Tweets referring to neighbourhoods was fairly limited, F-tests (95% confidence limit, see appendix B) still confirmed that the error variances in X/Y coordinates of both Tweets referring to neighbourhoods as well as Tweets referring to streets were significantly higher than that of Tweets referring to POIs.

These differences are likely a result of the process used to derive an exact location from Tweets referring to neighbourhoods and streets. For these Tweets further assumptions were necessary to derive a location, whereas POIs already referred to exact point locations. In this research it was assumed that the exact location of a Tweet referring to a street or neighbourhood lay either in the deepest topographical depression or at the point of lowest elevation along the street or in the neighbourhood. Since especially along long streets or in large neighbourhoods there can be a considerable number of depressions, this can lead to large errors. Also errors can occur when the lowest depression is used to pinpoint a Tweet, whereas the actual flood originated from a river, meaning it is not restricted to topographical depressions. Additionally multiple locations on a street can be flooded, meaning that if a photograph refers to one location, that does not mean the location derived using the text is not flooded also.

Water depth

Uncertainties in the water depth were analysed by comparing either the single water depth mentioned by the Tweet or the average of the range of water depths mentioned, to the water depth which was estimated based on the attached photograph.

In figure 25 it can be seen that water depths mentioned by the message are generally higher than the water depths derived from the photographs. Apart from 3 outliers however, errors are lower than 55 cm. Reviewing only the range mentioned by some of the Tweets and looking whether this covered the water depth derived from the photograph, showed that this was only the case for 14 out of 37 observations. For only 1 Tweet the water depth derived from the photograph was higher than the range specified in the Tweet. The majority of the Tweets, 22 of them, mentioned a range that was higher than the water depth derived from the photograph. Although at first sight these results seem to indicate water depths are almost consistently overestimated in the Tweets, there are a variety of factors that can cause this apparent overestimation.

This bias can for example be introduced by the fact that estimating a water depth from a photograph is quite a subjective process. The photographs attached to the Tweets can also be taken at different locations than the text is referring to. If the flood caused this area to be inaccessible, the photograph likely originates from a more accessible location, which has a lower water depth. Nevertheless it was tested whether the mean water depth error significantly differed from zero. Therefore a t-test was used (95% confidence limit), for which a log transformation was applied to the data (see appendix B). Using this test it could not be proven the mean deviation was any different from zero. Although it cannot be definitively concluded that water depths are overestimated in the Tweets, the results in figure 25 indicate that water depth errors are likely limited to tens of centimetres for the majority of the observations.

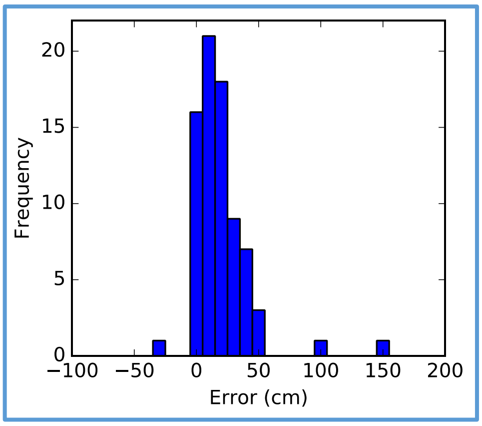

Figure 25: Jakarta - Absolute errors in water depth. Positive values mean the text derived water depth is higher than the water depth estimated from the photograph.

3.2 Flood mapping

Three different methods for creating flood inundation maps (see § 2.3.2) were applied to the Jakarta case study. The optimal set of parameters for each method was determined by comparing the created flood maps with validation data. For the Jakarta case this validation data consisted of 75 locations of flooding and the water depths at these locations, which were derived from photographs attached to Tweets (paragraph 2.2.3). Although the local disaster management agency (BPBD) of Jakarta also released a map of neighbourhoods that were affected by flooding during February of 2015, the quality of this map was found to be quite poor.

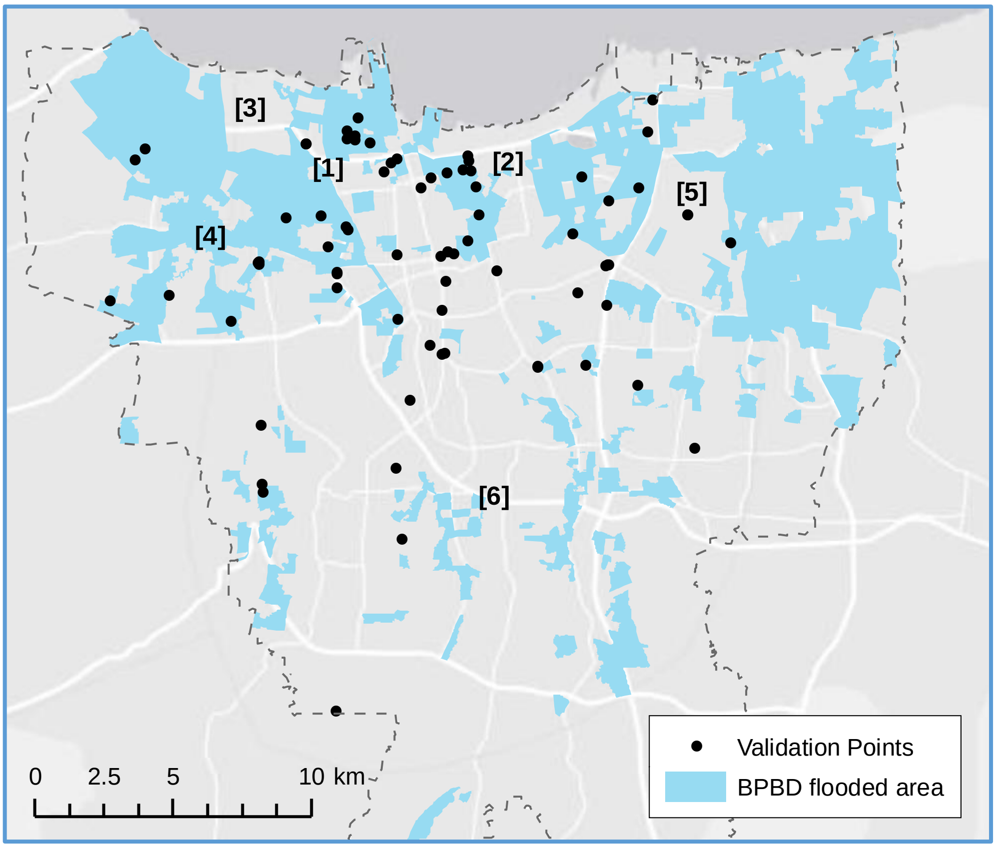

Figure 26: Jakarta – Potential validation datasets for the Jakarta case study. Numbers represent locations being referenced in this paragraph.

This is illustrated by comparing both datasets, which are both displayed in figure 26. It can be seen that a considerable amount of locations derived from the photographs attached to Tweets, are not marked as flooded in the map of the BPBD. In fact, this map only covered 40% of the validation points. This not only indicates that the BPBD map cannot be used in validating the flood maps created in this research, but also illustrates the need for better flood inundation maps of the area. Since by using the point dataset in validation however, overestimations of flood extents are not reviewed, the number of validation points that were within the created flood extent, was divided by the total flooded area of these map. Calculating this value for the BPBD flood map, having a flooded area of 169 km^2, gives a value of 0.18 points correctly mapped flooded per km^2. Since this is the information about the floods that is currently published, it is the benchmark for the flood maps created using the Twitter data.

3.2.1 Plain interpolation

The plain interpolation method utilized all observations at once, and involved performing a basic interpolation of water levels derived from these observations. All flooded areas not directly connected to any of the observations were removed from the maps. The flood maps created using this simple method were already vastly superior compared to the BPBD flood map. The best result was obtained by using both observations with and without water depths, and assigning all 'depthless' observations a DWD of 10 cm.

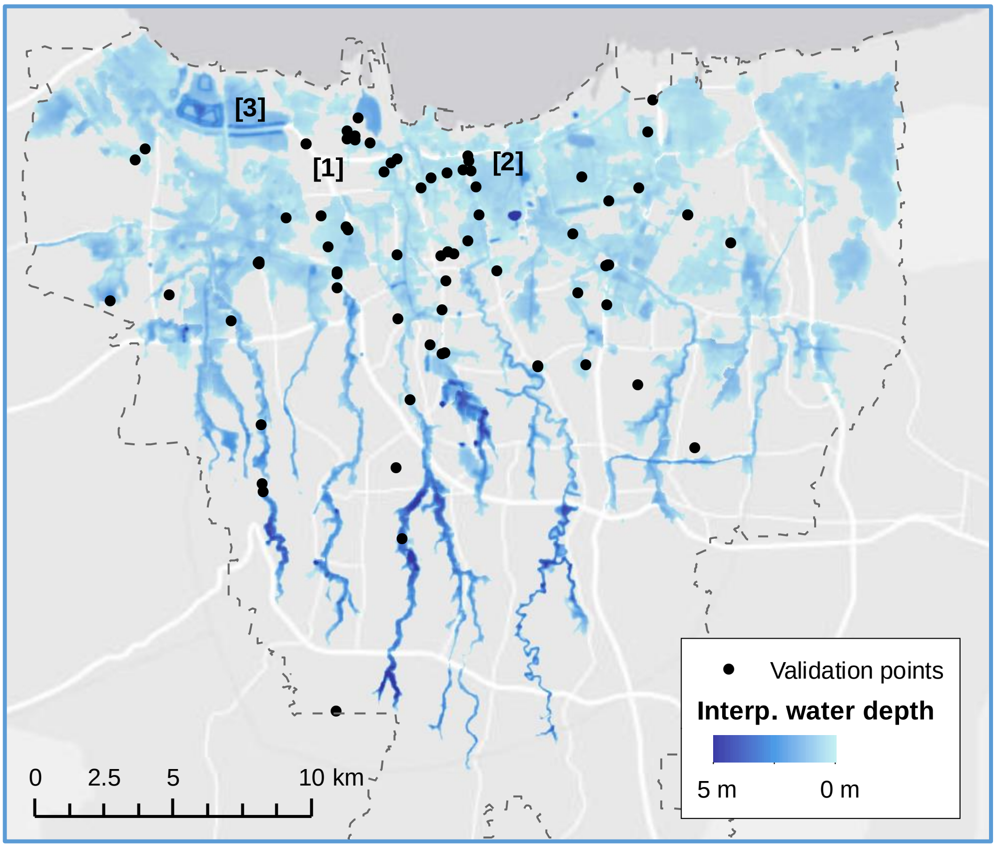

Figure 27: Jakarta - Flood map created using plain interpolation (power = 2, smoothing = 100 m) and a DWD of 10 cm. Locations [1] to [3] are discussed in text.

The resulting flood map is given in figure 27. Since a HAND map was used, especially the flood extents upstream are nicely constrained to the rivers. Downstream however, terrain slopes are only minimal, meaning small overestimations of water level can lead to quite large overestimations of flood extent. Location [2] for example contains no observations in the input dataset, nor photographs in the validation dataset, making it likely this area is erroneously mapped flooded. The same goes for location [3] which, if it was flooded with the depth indicated by the map, would likely have generated a considerable number of Tweets. Also, although the BPBD map was found to omit a large amount of flood locations, it is unlikely that if area [3] indeed had a very high water depth, it would not have been included in the BPBD registration of flooded areas. The main difference with location [2] however is that at location [3] actually input observations were pinpointed. This makes the flooding of this area not a result of the interpolation method, but of the errors in locational reference added to the Tweets. The southern part of the area around location [1], which is not mapped flooded, is better represented in the map, since there are neither validation points nor input observations within this area. A full comparison to these validation points shows that the map covers 77% of them, meaning 0.29 validation points are correctly mapped flooded per km^2 of flooded area. This is considerably higher than the 0.18 that was found for the BPBD flood map, although the BPBD map was also negatively affected by the fact that it was created at neighbourhood scale.

The ME and RMSE in water depth at the validation points were +8 cm and 43 cm respectively. Although the ME seems low, this includes the underestimation of water depths at 23% of the points where no water depth is mapped. Especially in upstream regions, high water depths of up to 5 m are mapped. That the ME remains limited nonetheless indicates that there is no large bias in water depths in the downstream regions.

The overestimations of water levels in these upstream areas were due to the way the HAND was created. For the Jakarta case study the elevation model was clipped to the special capital region of Jakarta. Since the number of upstream cells was used to identify the drainage channels, these drainage channels did not extend all the way to the edge of the map. Therefore the nearest drainage at the edge of the map was further away and was the HAND higher at these locations.

To review the impact of this phenomenon on water levels an additional HAND map was created by using ALOS World 3D – 30 m data around the edges of the original DTM to identify drainage channels at the edge of the map. The resulting water levels assigned to each observation, for both the original and new HAND map are displayed in figure 28.

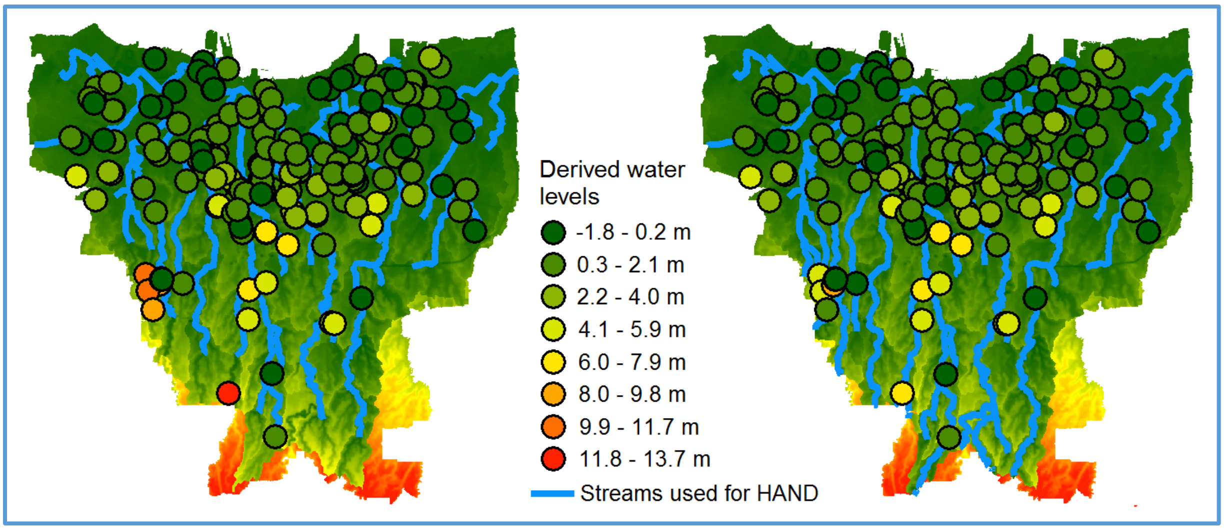

Figure 28: Jakarta - Overestimation of water levels (left) due to drainage channels not extending to the edges of the map. Extending these to the border reduces the water levels upstream (right).

This figure shows that especially in the western parts of upstream Jakarta, the effect of the drainage channels not extending to the edges of the DTM is large. Using a HAND map for which the drainage channels are extended to the edges drops the water levels of some of the observations by several meters (figure 28, right). The flood maps in this paragraph were all created using the old HAND map, meaning overestimations upstream are included.

3.2.2 Interpolation of groups

As an improvement over the simple interpolation method discussed in the last paragraph, observations were grouped prior to interpolation to determine which observations belonged to the same flooded area. The best results were obtained by interpolating the water levels in a 2.5 km extent around the groups of observations. Prior to interpolation, the results of two grouping methods were evaluated: grouping based on proximity and grouping based on downstream flood paths. The latter method was used, since it produced vastly superior results (see appendix B). The best results were produced by additionally using ‘depthless’ observations that were located close to the flood extent that was created by using only observations with water depths. These ‘depthless’ observations were given a DWD of 60 cm. The resulting flood map is given in figure 29.

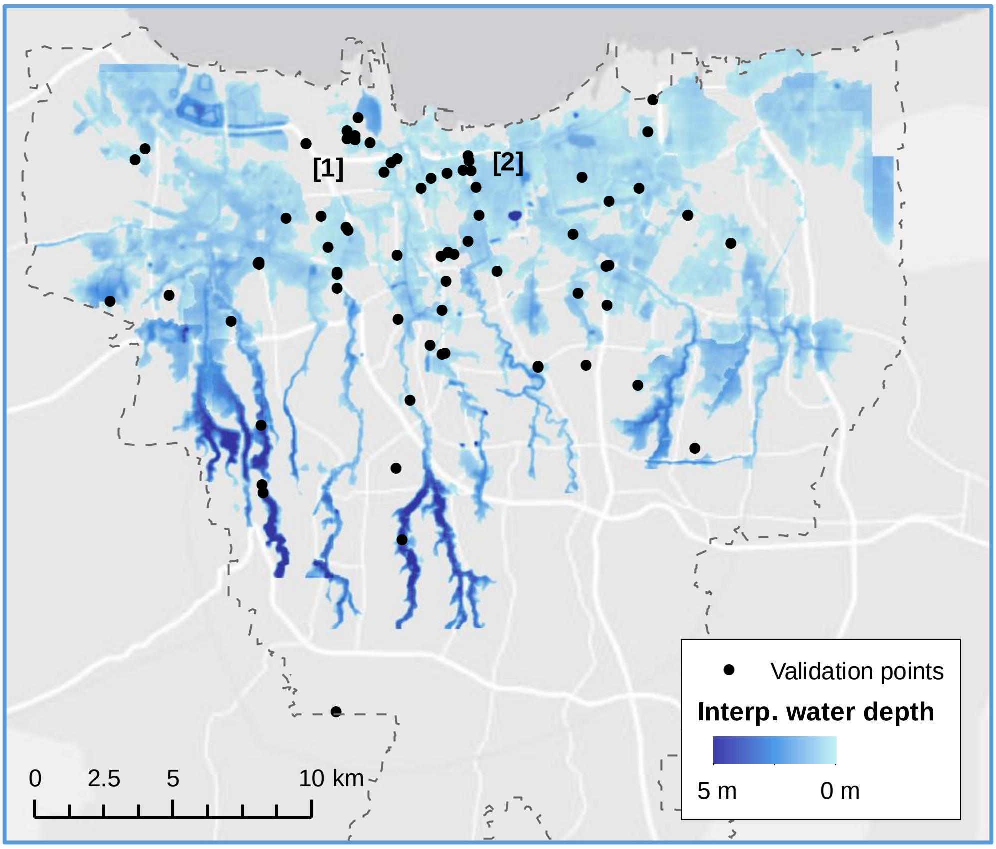

Figure 29: Jakarta - Flood map created using interpolation of groups (power = 2, smoothing = 100 m) and assigning nearby observations a default water depth of 60 cm. Locations [1] and [2] are discussed in text.

For downstream Jakarta this result was nearly identical to that of the plain interpolation method. Again the flood pattern at location [1] is well reproduced and location [2] is erroneously mapped as flooded. The effect of the grouping on the map is however clearly visible. In the north east of the area for example, the rectangular extent used in interpolation causes artefacts to be present. Also, interpolating the water levels of each group separately meant that the overestimations of upstream water levels have a larger effect locally. Downstream water depths on the other hand are in a more credible range of 0 – 1.5 m.

The overestimation upstream is reflected in a high ME of +0.24 m and overall the errors in water depth are also larger with a high RMSE of 0.85 m. Although the total flooded area is similar to the one of the previous method, the number of validation points covered by this area increased to 85%, meaning 0.32 validation points were correctly mapped flooded per km^2 of flooded area. However, because the flood extent upstream is purely limited by the rectangular interpolation extents used, the method is not considered to perform substantially better than the plain interpolation method, especially since flood extents downstream are nearly identical.

3.2.3 Interpolation along flow paths

For the last interpolation method evaluated by this research, observations were grouped using their downstream flow paths. After grouping the water levels were interpolated along these flow paths. After all the water levels on these flow paths were determined, the water level of each cell on the flow path was also given to the cells upstream of it. From these water levels the HAND was subtracted to calculate the water depths. Flooded areas which were not physically connected to any of the flow paths were removed from the flood maps.

This method produced the best results for the Jakarta case. Only using observations which specified a water depth (power = 1, smoothing = 0 m) already gave quite good results, with 67% of the validation points being mapped flooded, meaning 0.31 validation points were correctly mapped flooded per km^2 of flooded area. The best result was however obtained by also including all ‘depthless’ observations and giving them a water depth of 30 cm (power = 2, smoothing = 200 m). The resulting flood map is given in figure 30.

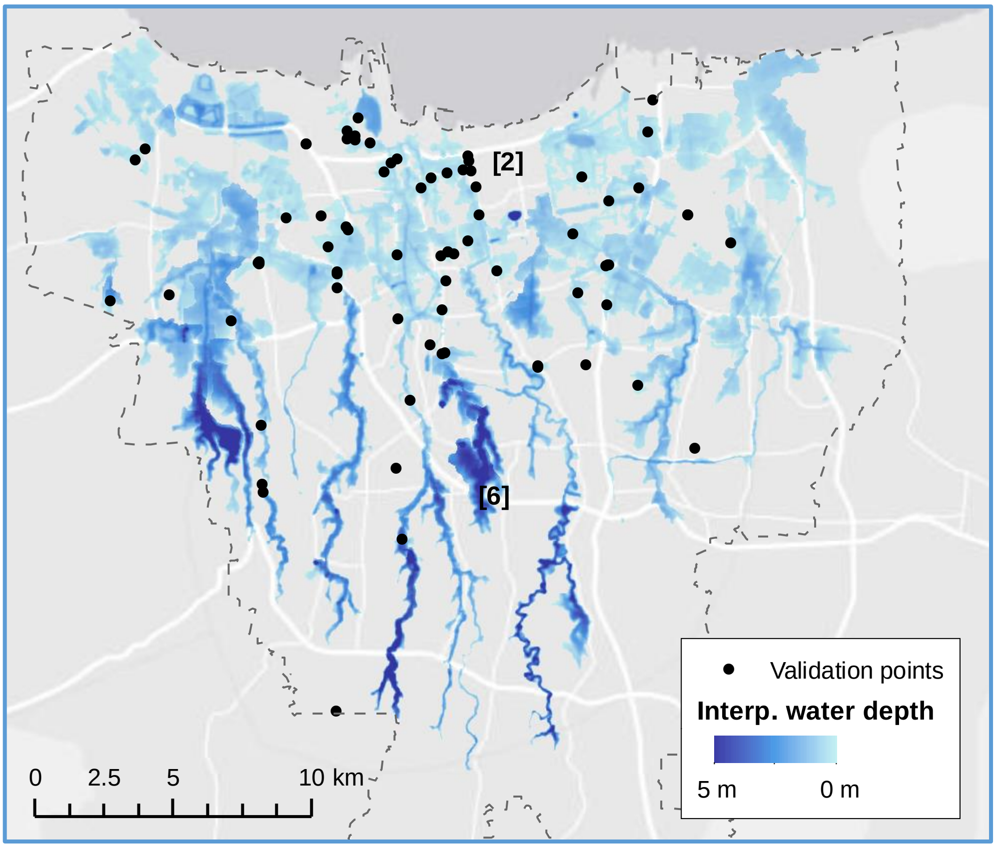

Figure 30: Jakarta – Flood map created using interpolation along flow paths (power = 2, smoothing = 200 m) and assigning all 'depthless' observations a default water depth of 30 cm. Locations [2] and [6] are discussed in text.

Most striking is the large reduction in downstream flood extent. The area around location [2] for example, which was erroneously mapped flooded by both plain interpolation and interpolation of groups, is not mapped flooded anymore. The effects of grouping are also clearly visible, causing the upstream water levels to have a large impact on the results locally. Also the overestimation of water depths at location [6] stands out, which is due to two 'depthless' observations being wrongly assigned a location with a relatively high HAND, causing corresponding water levels to be high. This effect was less severe in the result of the plain interpolation method, since observations were not grouped here.

If specifically the downstream area of Jakarta is reviewed (figure 31) it can be seen that a considerable amount of the validation points in this area are covered by the map. As discussed above, area [3] is mapped flooded due to observations being erroneously pinpointed to this location. Looking at both the presence of validation points and input observations however indicates that such overestimations of flood extent are likely limited. Only at area [4] a limited number of validation points is present, meaning flood extent might be somewhat overestimated in this area.

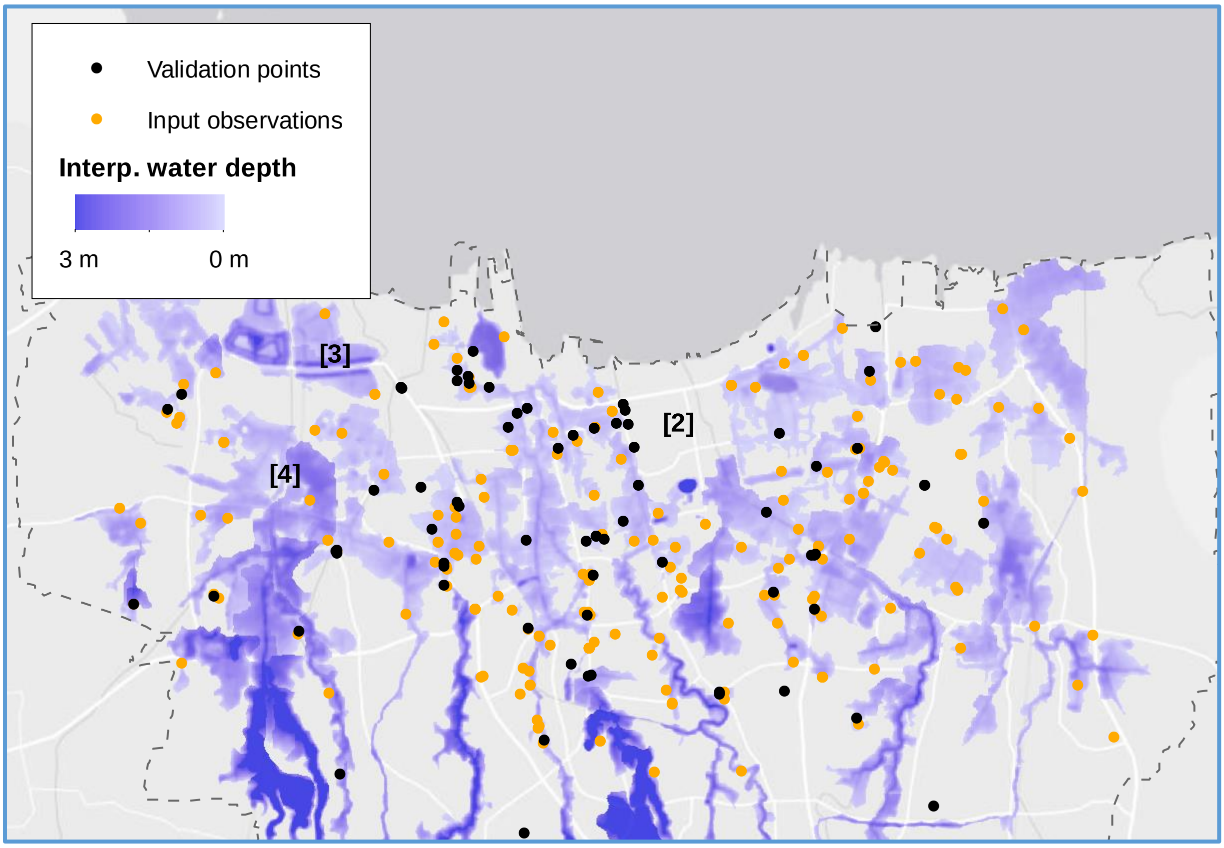

Figure 31: Jakarta – Flood map created using interpolation along flow paths (power = 2, smoothing = 200 m) and assigning depthless observations a default water depth of 30 cm. Zoomed to downstream Jakarta. Locations [2] – [4] are discussed in text

Although the flood map created using interpolation along flow paths covered only 75% of the validation points, the fact that the flood extent was considerably reduced by this method, caused the number of validation points per km^2 of flooded area to be higher (0.33). Although this value is only marginally higher than the one found for grouped interpolation, the fact that in grouped interpolation flood extents were constrained by limiting the interpolation extent, makes that this method is considered to be significantly better than the other methods. Also the ME in water depth (+0.01 m) is the lowest of all methods reviewed and the RMSE (0.44 m) though being slightly worse than that of plain interpolation (0.43 m) is considerably better than the one found for interpolation of groups.

The significant reduction in flood extents found by interpolating along flow paths, is mainly caused by the fact that observations can only affect water levels within their own subcatchment using this method. This is illustrated using the example in figure 32. The hypothetical area in (a) shows two drainage channels with a ridge in between. In case the water level on channel 1 is higher than the ridge, the ridge would overflow at some point, as illustrated by the blue line in (b). In case a HAND map is constructed of this area, as is illustrated in (c), there is an offset in elevation on the dividing line between the subcatchments of drainage channel 1 and drainage channel 2, because the higher bed level of channel 2 causes the HAND in this sub-catchment to be lower. If an observation of a water level close to drainage channel 1 is used with plain interpolation, the water level would simply be extended over the entire area, as is illustrated in (d) and (e). However since interpolation of flow paths only spreads the water level of the observation over upstream cells, the flood extent of this method is restricted to the subcatchment of drainage channel 1 (f). Therefore the flood extent is smaller using interpolation along flow paths.

Comparing both the results of plain interpolation and interpolation along flow paths for this hypothetical example to the potentially real water levels given in (b), learns that interpolation along flow paths gives the most accurate description of the flood extent. The overflow of water over the ridge is likely only limited to a small number of locations along the ridge, meaning overland flow is only present in a small part of the subcatchment of drainage channel 2. Where interpolation along flow paths will indicate the sub-catchment of drainage channel 2 is not inundated, using plain interpolation will lead to the conclusion that the sub-catchment of drainage channel 2 has very high inundation depths. Although interpolation along flow paths will omit some locations that actually experience flooding, using plain interpolation leads to a very large overestimation of flood extent. Errors in omission (underestimation) are considered better than errors in commission (overestimations), since in case the maps are for example used by first responders, they want to know which areas are certainly flooded and not which areas can be flooded. Also, providing users with maps that frequently overestimate flood extents reduces the trust people have in these maps.

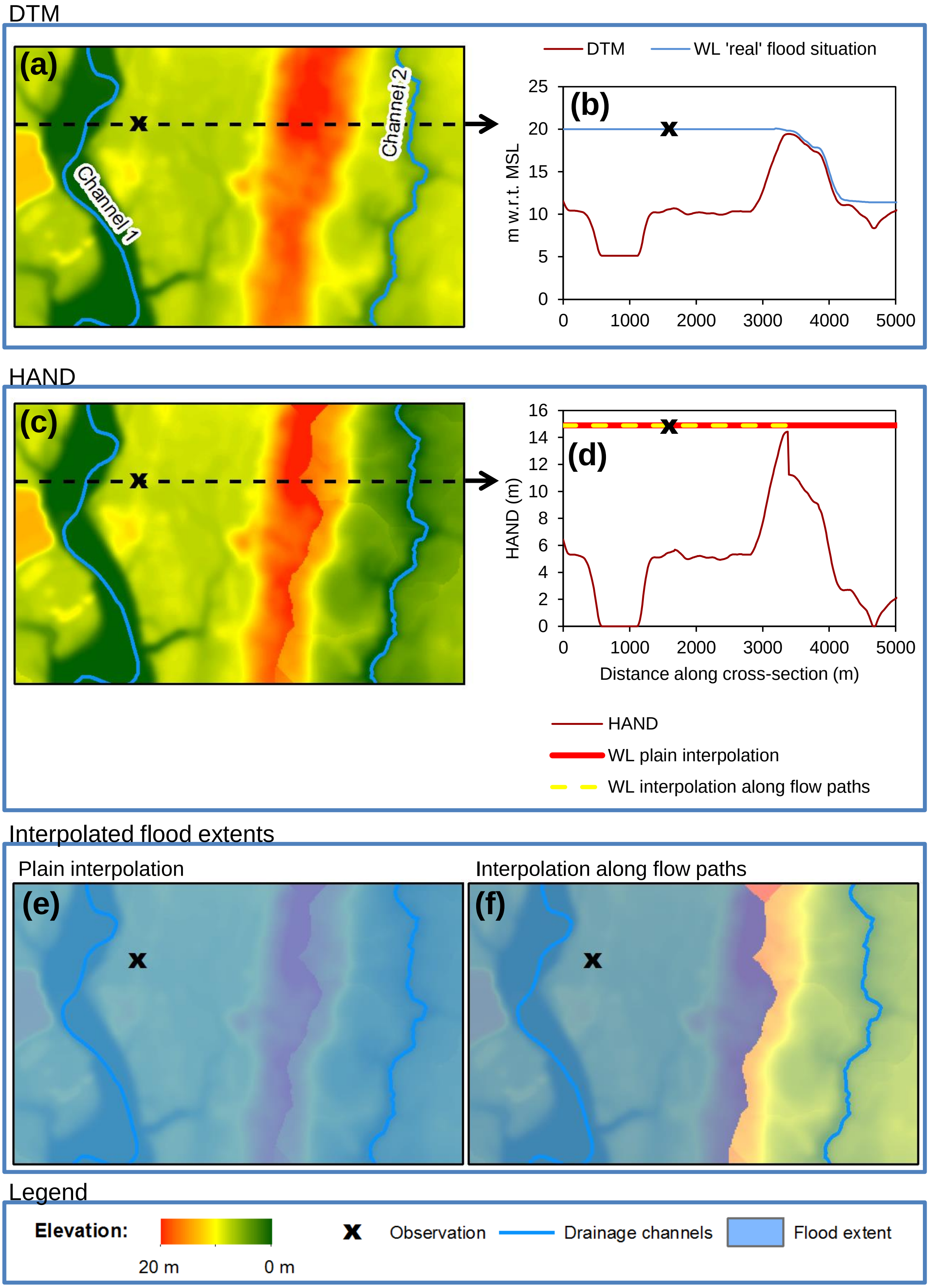

Figure 32: (a) gives the terrain elevation and drainage channels of a hypothetical study area. A cross section of this area is displayed in (b), in which also a potentially real water level is given, in case the water level of drainage channel 1 is 20 m. (c) gives the HAND map of the area, of which a cross-section is given in (d). The offset in elevation here is due to drainage channel 1 being lower than drainage channel 2. Also the interpolated water levels are given in case plain interpolation or interpolation along flow paths is used. The flood extents created using these methods are given in (e) and (f) respectively.

3.3 Uncertainty assessment

The uncertainty in flood extent caused by locational and water depth errors in the Tweets as well as errors in the elevation data was investigated using Monte Carlo analysis. Also the effect of varying the DWD, which was added to the 'depthless' observations, was reviewed.

3.3.1 Quantification of uncertainties

The magnitude of the errors in the datasets was quantified prior to evaluating the uncertainties. The analysis of the Twitter datasets for the Jakarta case study showed that the standard deviation in locational errors of the Twitter messages was about 500 m (§ 3.1.2) and the mean error in X/Y coordinates was close to zero. These errors were simulated by drawing from a normal distribution. The errors in water depths derived from the messages were also quantified using the analysis results. Since a statistical test could not confirm the mean error in water depth was any different from zero (§ 3.1.2), these errors were simulated by drawing from a normal distribution with a mean of zero and a standard deviation of 20 cm. Since no reference information with respect to ground level was available, errors in the elevation model were quantified using literature. As discussed in paragraph 2.3.3 these errors were introduced with a mean error of 0 cm, a standard deviation in error of 20 cm and a correlation distance of 100 m.

Using the quantification of all error sources, the number of random simulations necessary to assess the uncertainties in the flood maps was determined. This was done by evaluating the standard deviation of the standard deviations of errors for each individual grid cell or each individual observation. At the number of realizations this standard deviation stabilizes or approaches zero, the standard deviations of individual grid cells or observations are more or less equal. This means each observation/grid cell has roughly the same error distribution, and adding additional realizations will not improve the resulting uncertainty maps. The results of this analysis are given in figure 33.

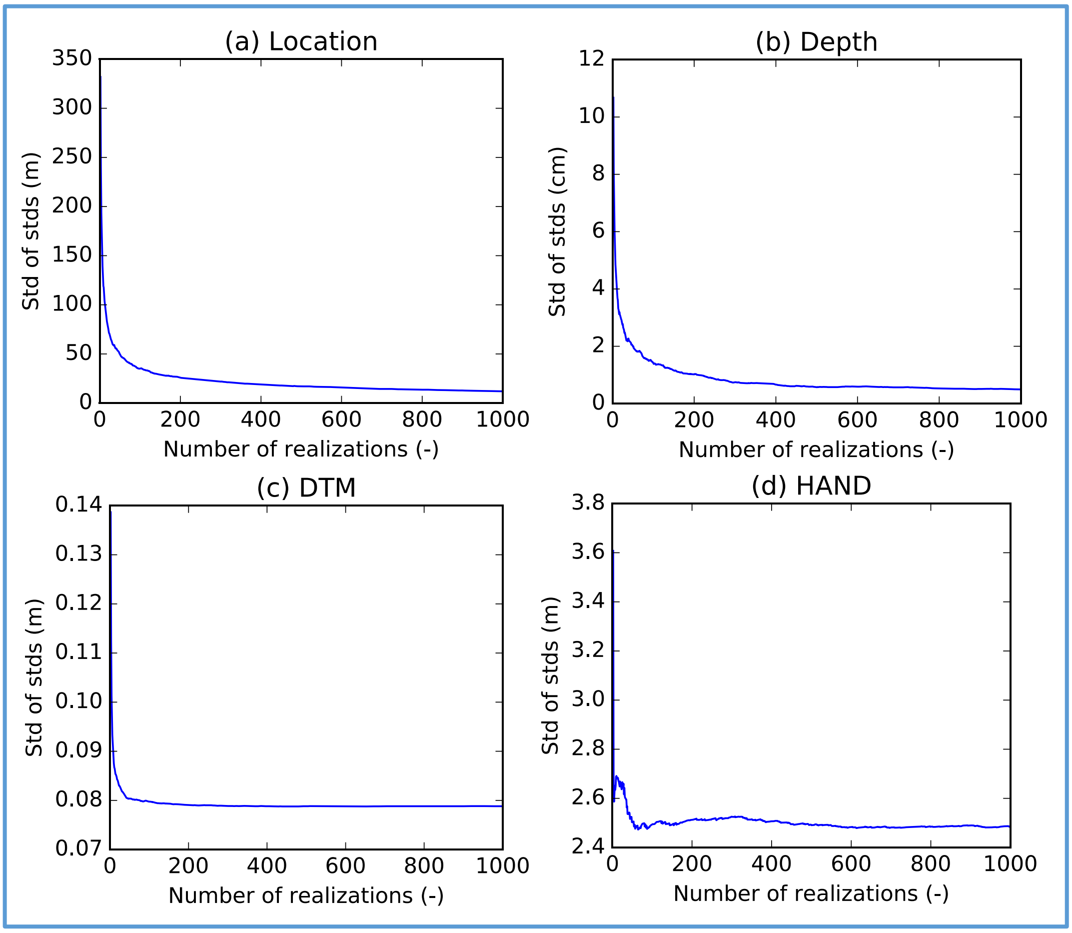

Figure 33: Jakarta - Standard deviation of standard deviations of individual Observations/grid cells

Since all graphs stabilize at about 600 realizations, 600 random simulations were used to assess the impact each individual error source has on the uncertainty in the flood maps. However, it can be seen that the graphs for the DTM (c) and HAND (d) are very different. This is due to the fact that in creating the HAND, elevation values are calculated relative to the elevation values on the drainage channels. Thereby the errors in single elevation values along drainage channels affect entire areas, causing the similarity in error distribution between all cells in the grid to be less. Nevertheless both graphs are more or less stable after 600 realizations. Although a multitude of that number would have been necessary to simulate the combination of error sources, to ensure all combinations of uncertainties were represented in the simulations, 1000 simulations were used, due to time and computational constraints. It was found that the results of doing multiple runs of 1000 simulations did not produce significantly different results.

In paragraph 3.2 an overestimation of upstream water levels was identified. Although this can be mitigated by extending the DTM by using other sources of data, this causes difficulties in simulating the errors in the DTM, since this additional data will have very different error characteristics. Therefore the observations for which water levels were overestimated were excluded, by excluding all observations with latitudes lower than -6.23° from the analysis. Thereby 18 of the total of 233 observations were excluded, leaving a dataset of 215 observations, with 117 mentioning a water depth. Using this reduced set of observations, the uncertainties in the flood maps were evaluated.

3.3.2 Location uncertainty

The uncertainty caused by locational errors was assessed by doing a Monte Carlo simulation using the preferred mapping method from paragraph 3.2. The result of this analysis is given in figure 34.

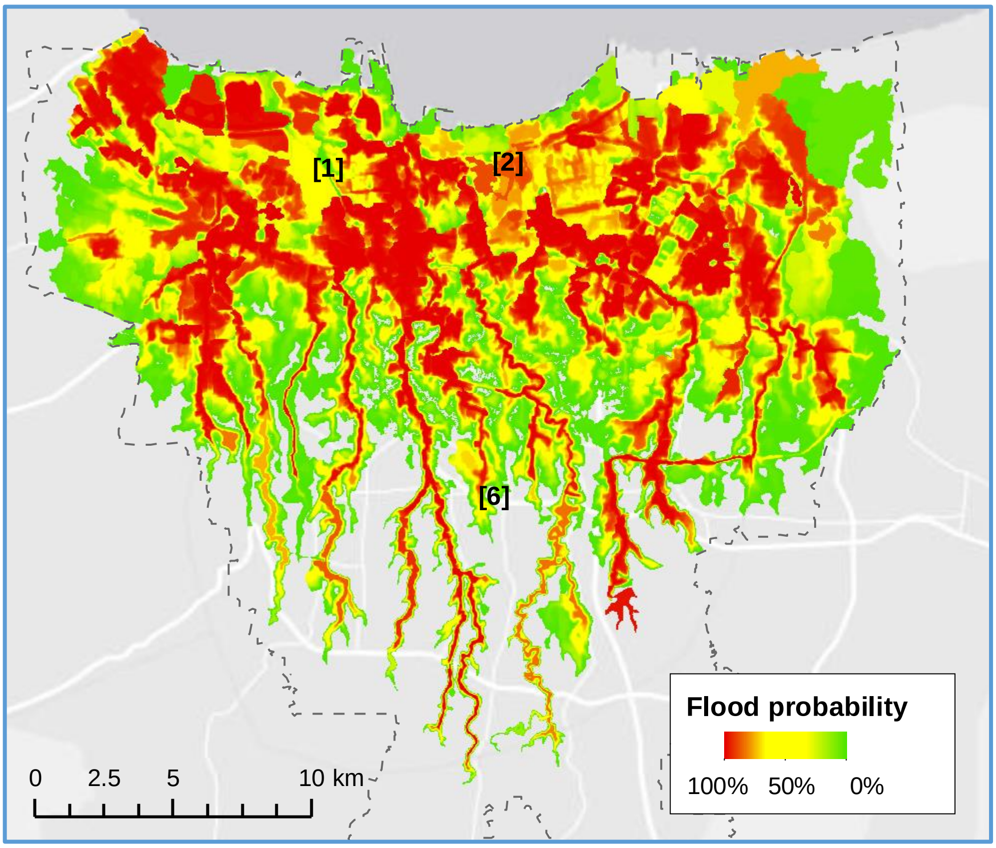

Figure 34: Jakarta – Probability of flooding determined by reviewing the impact of locational errors using a Monte Carlo analysis. Location [1], [2] and [6] are discussed in text.

Locational errors mainly cause uncertainty in the northern downstream parts of Jakarta. These areas are generally more flat and many of them have a probability of flooding around 50% (yellow). Cells that have a probability of flooding closer to 0 or 100% are generally less uncertain, since it is fairly certain these cells are not flooded and flooded respectively. About cells which have a probability of flooding of 50% however no definite statement can be made regarding their likelihood of flooding. That specifically flat areas are sensitive to locational errors is because of the fact that in these areas relatively small changes in water levels can cause large differences in flood extent. Therefore changes in water levels, caused by observations being pinpointed to locations with different ground elevation values, have most effect in the downstream regions of Jakarta. This is for example illustrated by locations [1] and [2], which were not flooded in the flood map in paragraph 3.2.3, but have a considerable probability of flooding in the results of the Monte Carlo analysis. Also it can be seen that the uncertainty in flood extent is large at location [6], which in the flood map in paragraph 3.2.3 had a considerable overestimation of water depths.

3.3.3 Depth uncertainty

Also the uncertainty caused by errors in water depths mentioned by the Tweets, was evaluated using Monte Carlo analysis. The uncertainty resulting from errors in water depths is displayed in figure 35.

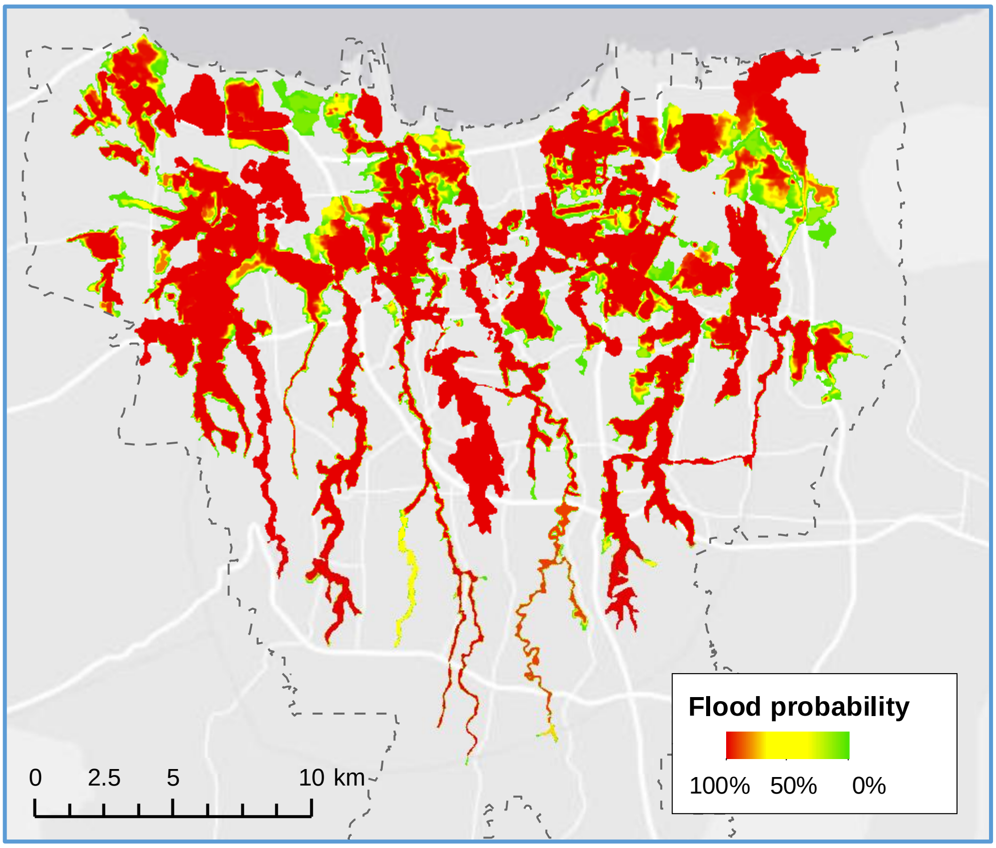

Figure 35: Jakarta – Probability of flooding determined by reviewing the impact of errors in water depth using a Monte Carlo analysis

This uncertainty is considerably lower than the uncertainty caused by locational errors. Only a few downstream locations show some limited uncertainty due to errors in water depths. The uncertainty caused by these errors is mainly in the downstream areas, since these are relatively flat. This causes changes in water level (vertical) to have a relatively large effect on flood extent (horizontal). Water depth errors cause hardly any uncertainty more upstream, due to the steep slopes in these areas. Another factor keeping the uncertainties to a minimum is the interpolation method, due to which the water levels of observations can only affect their own subcatchments.

3.3.4 DTM uncertainty

The last source of uncertainty examined, was the uncertainty caused by errors in the DTM. Again Monte Carlo simulations were run to assess the impact of these errors. The result is given in figure 36.

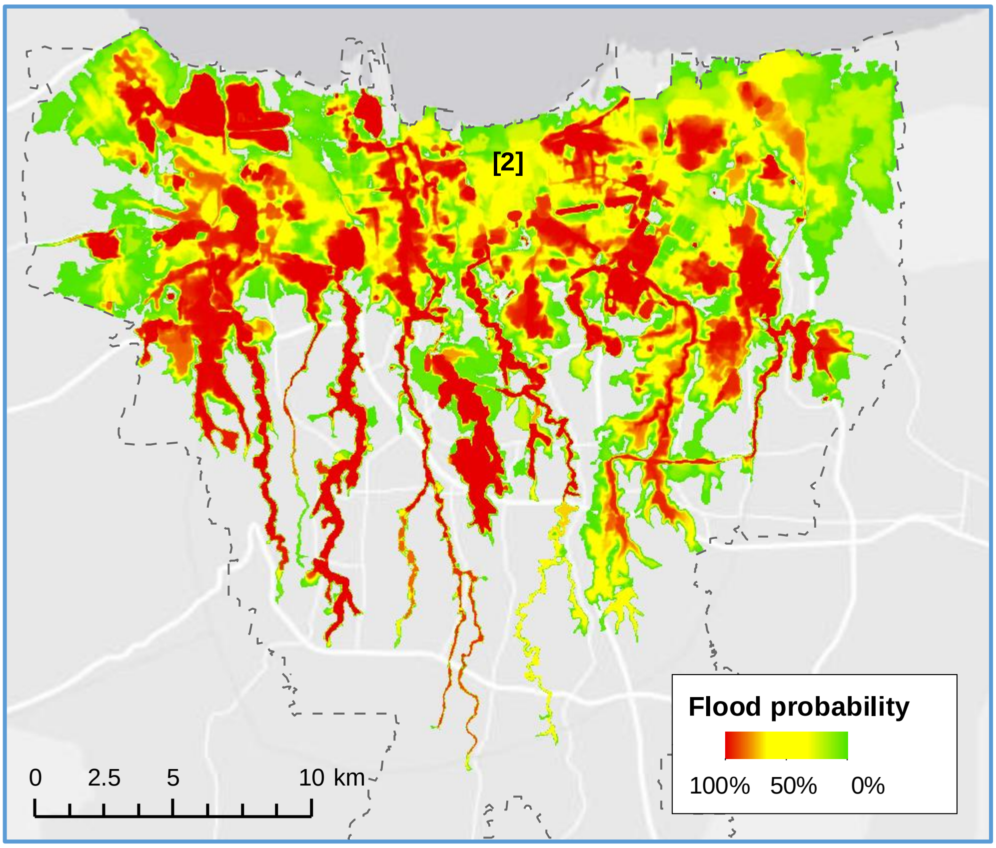

Figure 36: Jakarta – Probability of flooding determined by reviewing the impact of errors in the elevation data using a Monte Carlo analysis. Location [2] is discussed in text.

Especially downstream the uncertainty caused by errors in the elevation data is considerable. Since terrain slopes are low in this area, even minor errors in elevation can have a large influence on the flood extent. Also the subcatchments and downstream flow paths of observations are changed by errors in elevation data, increasing the uncertainty even more. That elevation errors are the most important source of errors in the downstream areas, is illustrated for example illustrated by area [2]. Besides showing that errors in the elevation data can seriously affect the flood extents in flat areas, the analysis also illustrates the importance of using spatially auto correlated errors. If uncorrelated errors were used in simulating the uncertainties, the resampling from the original 2 m DSM to the 20 m DTM, would have filtered all errors, since basically the average of a lot of cells would have been used. Using the auto correlated errors however, the errors are still in the elevation data after resampling. Nevertheless the flood probabilities in areas which are known not to be flooded, such as location [2] are lower than those of the locational errors, meaning that although errors in the elevation data cause considerable uncertainty downstream, the uncertainty caused by locational errors at those locations is even larger.

3.3.5 Combination of uncertainties

To evaluate the combined uncertainty in the flood maps, a Monte Carlo analysis was performed in which the errors in location and water depth of the Tweets as well as the errors in the elevation data were evaluated by performing 1000 random simulations. The result of this analysis is given in figure 37.

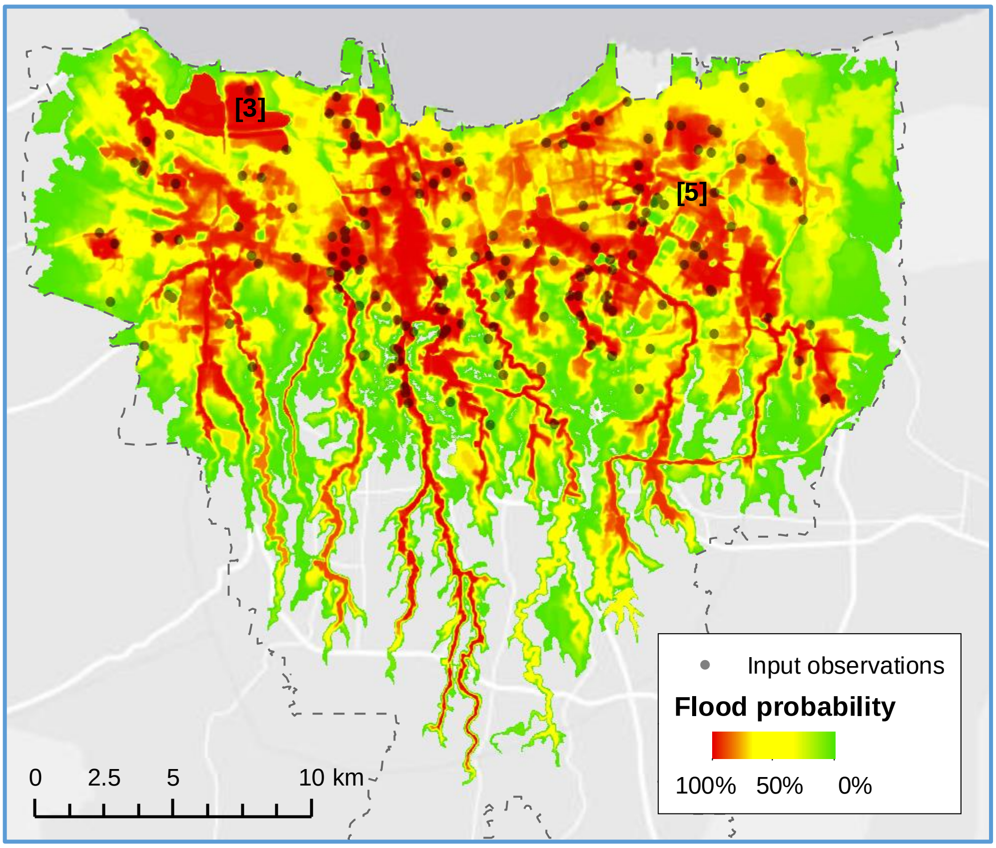

Figure 37: Jakarta – Probability of flooding determined by reviewing the impact of locational errors, errors in water depth and errors in the elevation data using a Monte Carlo analysis. Locations [3] and [5] are discussed in text.

As expected based on the results of reviewing the uncertainty caused by the individual error sources, the flood extents in the downstream area are most uncertain. The uncertainty caused by both errors in the elevation data as well as locational errors is large here. However, given the uncertainties downstream are very similar to those of simulating purely locational errors, these errors are most important in the downstream area. However, the uncertainty calculated using the Monte Carlo analysis, does not necessary reflect the uncertainty of individual messages. For example in area [3] only a very limited number of observations is present, although the results of the Monte Carlo analysis seem to indicate that this area is almost certainly flooded. The opposite is true for the area directly to the west of location [5]. Here four observations are present in very close vicinity, although the results of the Monte Carlo analysis indicate that this area has a very low probability of flooding. This indicates that solely reviewing the results of the Monte Carlo analysis does not give a complete picture of the uncertainty in the flood extent.

3.3.6 Varying default water depth

In addition to the Monte Carlo analysis, which was used to evaluate the impact of all quantifiable sources of error on the uncertainty in the flood maps, the effect of varying the DWD, used for 'depthless' observations, was also examined. The optimal DWD found in paragraph 3.2 was 30 cm. The effect of using water depths of up to 30 cm below and above this value was further examined (figure 38).

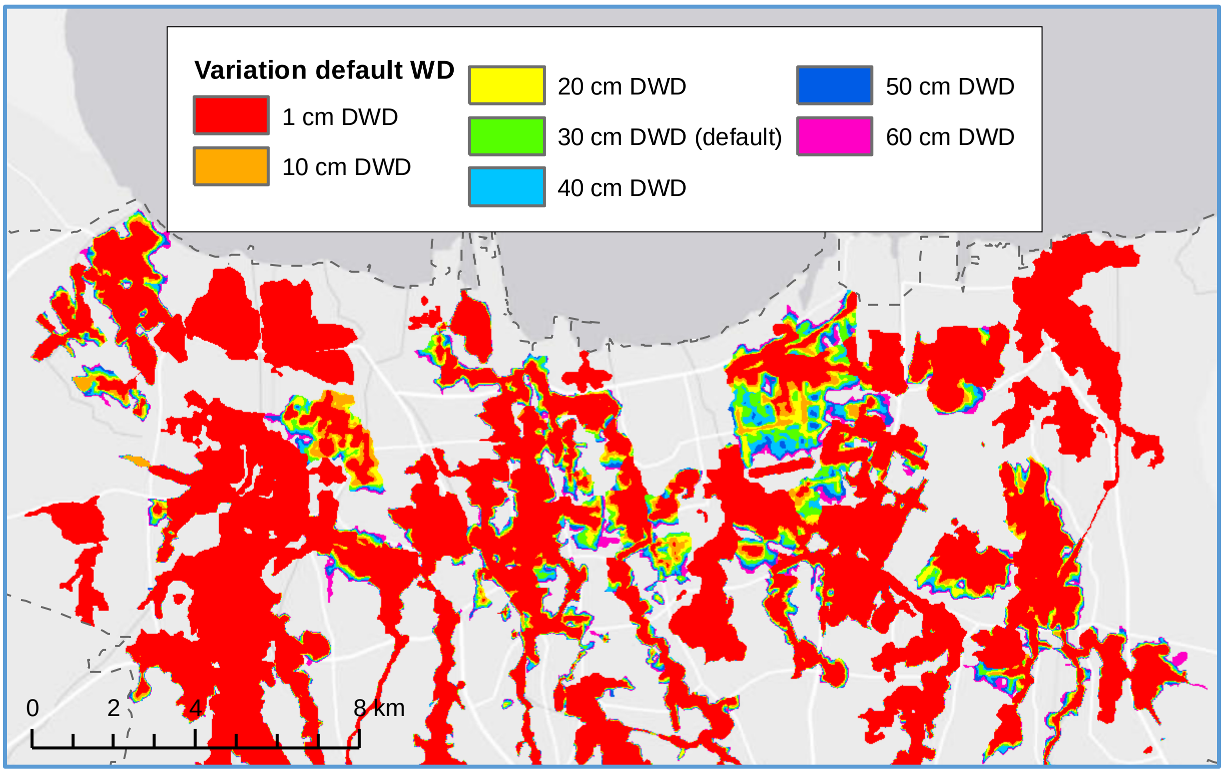

Figure 38: Jakarta - Variation in flood extent as a result of varying the DWD. The red area gives the flood extent in case a DWD of 1 cm is used, and the other colours show the additions in flood extent due to increasing the DWD.

Just like with the uncertainty caused by errors in water depth, the variation of the DWD mainly leads to changes in flood extent in the more flat downstream areas. Therefore a zoom-in of this area is given. Although the downstream area is mainly affected, the impact is still very limited. Upstream hardly any variation in flood extent is seen. Nevertheless the effect of using the lowest (1 cm) and highest (60 cm) DWD on the map of total uncertainty (figure 37) was reviewed. The results are shown in figures 39 and 40 respectively.

Although the differences between the maps with the lower a higher DWD are evident, these changes are mainly visible around the areas which were found to be sensitive to changes in DWD in figure 38. In these areas flood probabilities become higher, if a higher DWD is used. The spatial pattern of uncertainty is also not affected by the differences in DWD. Additionally the differences between both maps, and the original map of total uncertainty created using a DWD of 30 cm (Figure 37), are only minor, especially further upstream.

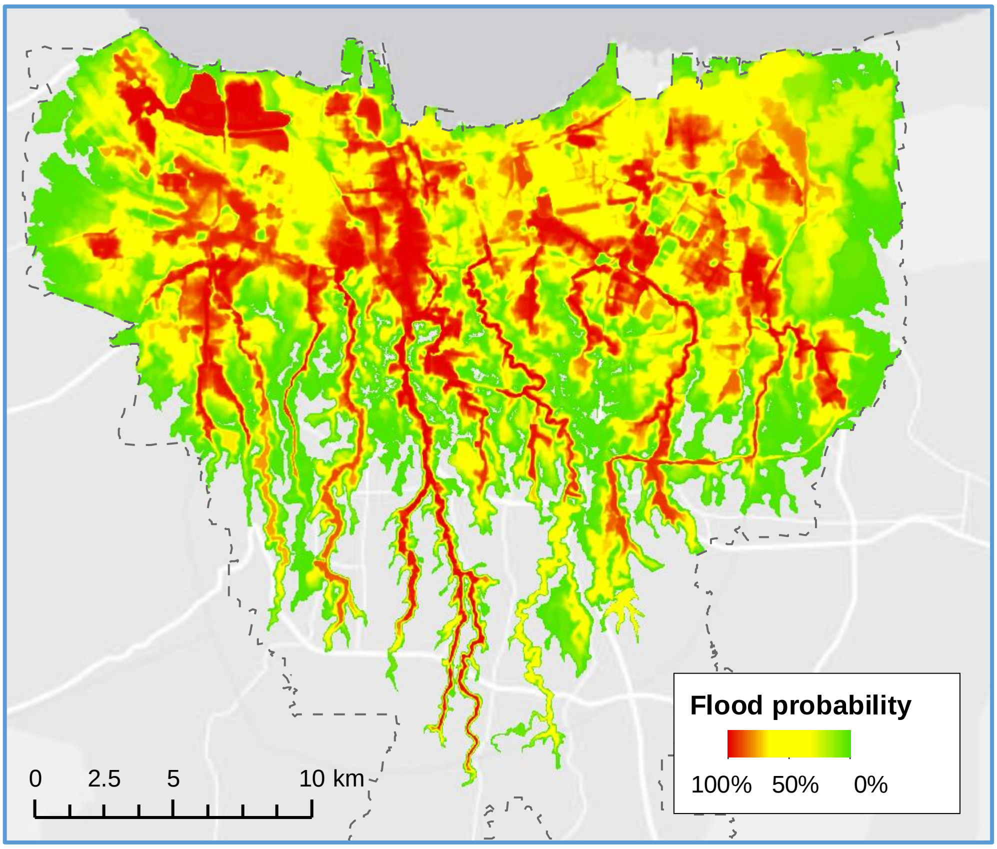

Figure 39: Jakarta – Probability of flooding determined by reviewing the impact of locational errors, errors in water depth and errors in the elevation data using a Monte Carlo analysis in combination with a DWD of 1cm.

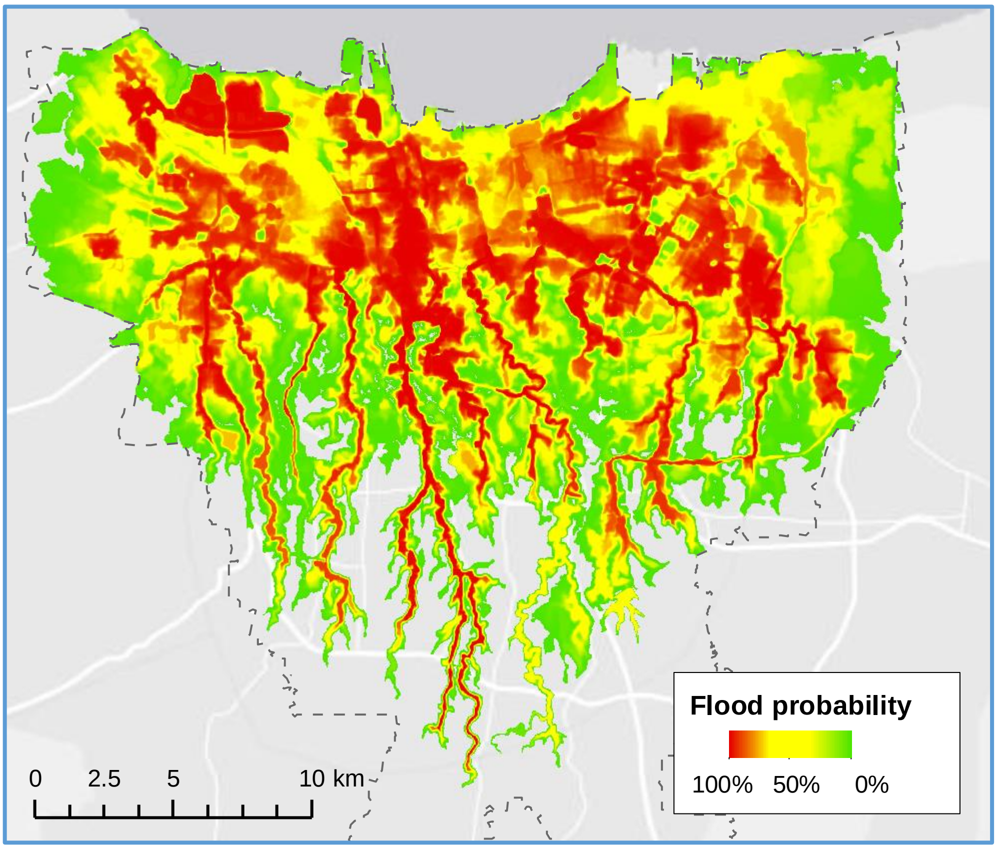

Figure 40: Jakarta – Probability of flooding determined by reviewing the impact of locational errors, errors in water depth and errors in the elevation data using a Monte Carlo analysis in combination with a DWD of 60 cm.

4 Results York

The second case study discussed by this report is the flooding of the city of York, which occurred in December 2015. The chapter starts with a discussion of the time variation and uncertainties in the Twitter dataset, which is followed by a comparison of a number of different methods to map floods. The chapter is concluded by an analysis of how errors in the input dataset lead to uncertainty in the flood maps. The methods used to produce the results in each of these sections are found in paragraph 2.3.

4.1 Dataset characteristics and uncertainties

The first step in investigating the York case study was an analysis of the Twitter dataset. This section discusses both the time variation as well as the uncertainties in this dataset. In contrast to the Jakarta case study, no water depths were mentioned in the York Tweets, meaning that only errors in positioning were investigated.

4.1.1 Time Variation

Both the time variation in number of Tweets and spatial pattern of Tweets was investigated. These time variations were compared to measured water levels on the River Ouse.

Number

To review the variation in the number of observations, the number of observations over 10 hourly intervals was determined. The result, alongside a graph of measured river water levels, is given in figure 41.

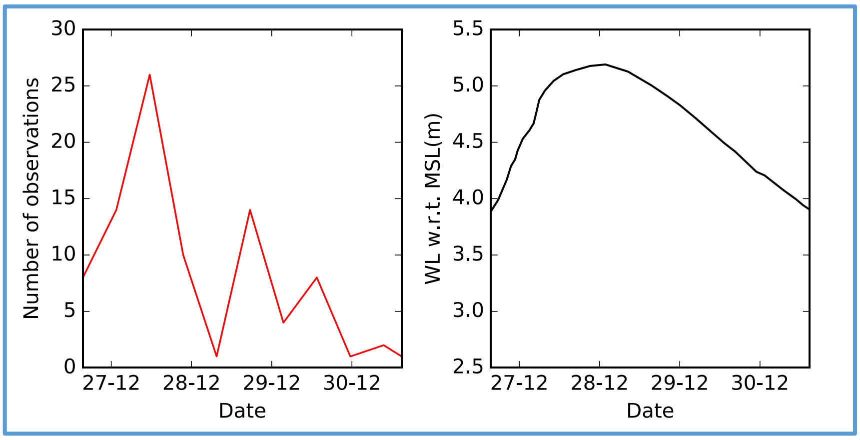

Figure 41: York - Number of Tweets in dataset over 10 hourly intervals (left) and the measured water level on the River Ouse (right) (EA, s.d.).

A comparison of both graphs gives some indication of the extent to which the actual time variation is reflected in the number of Tweets. First of all the largest peak in number of Tweets, on the 27th of December, roughly coincides with the peak in water level on the River Ouse. Also the peaks on later days seem to get lower day by day, just like the measured water levels. The most noticeable differences however, are the drops in the number of Tweets which occur in the early hours of every day. These drops are most likely caused by people generating fewer Tweets during the night. Overall however the number of Tweets in the datasets seems to reflect the severity of flooding in the area.

Spatial pattern

In assessing the utility of the dataset in real-time flood mapping, the spatial pattern of the observations over time is also of interest (figure 42). In the first four time steps most observations are posted scattered throughout the area. In later time-steps however, observations are mainly made around the city centre of York. Also it can be seen that in the eastern part of the inner city, at the confluence of rivers north of location [1], most observations are generated in the first three time steps. Huntington road however has observations in nearly all time-steps, which is consistent with news reports of flooding at this location.

All in all the Tweet dataset for York seems to capture some of the time-variation during the floods. However, only the graph of the number of observations could be compared to actual river water level measurements in the area. Although the review of the spatial pattern of observation seems to indicate that some for some areas observations are reported over a different time span than for other areas, there was no reference information available to compare the changes in spatial pattern with. Only for the Huntingdon Road area it was found to be likely that it was flooded over a longer period of time, since news report indicated this. Overall however, the dataset used for the York case study cannot be used to create real-time flood maps, since the number of observations is very low, meaning that maps over for example half an hour periods would contain very few observations.

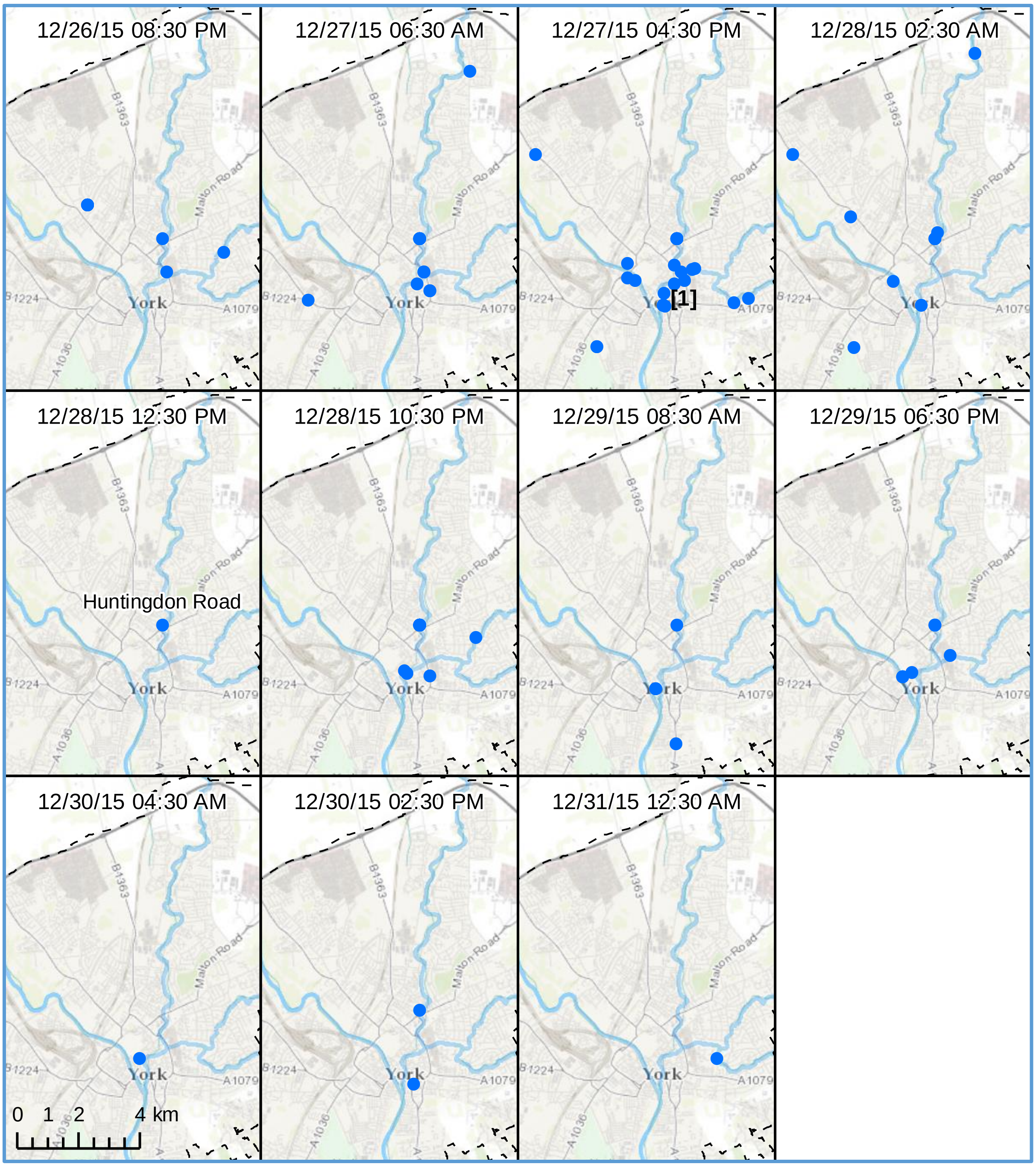

Figure 42: York - Post times of observations, recorded over 13 hour intervals (ahead of the time specified at the individual maps)

4.1.2 Uncertainties

The uncertainties in the Twitter datasets were reviewed in order to investigate the uncertainties in the flood inundation maps. Since no water depths were specified in the Tweets for the York case study, only text-derived locations were compared to locations derived from the photographs. The results for all Tweets for which a location could be derived from an attached photograph are given in figure 43.

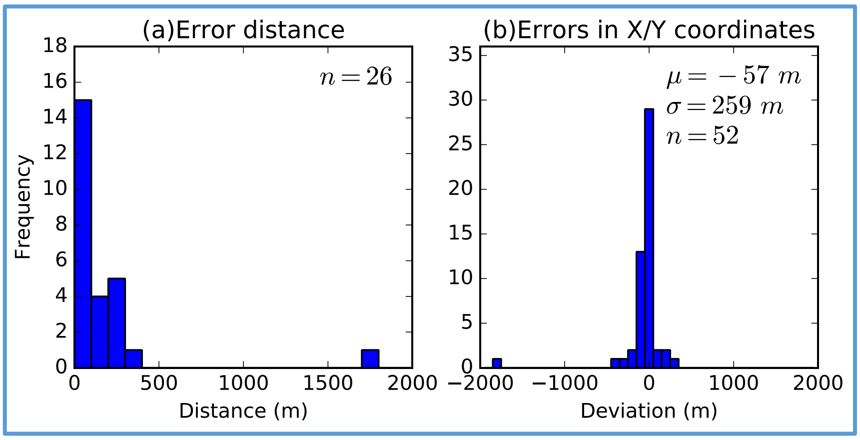

Figure 43: York - Locational errors of Tweets (Complete dataset), bin size = 100 m

Locations could only be derived from attached photographs for a limited number of Tweets. Nevertheless it can be seen that the locational references in the York case study contain a considerable amount of error. However, only one large outlier was found, of which the actual location was over 1.5 km from the location derived from the text. This particular observation refers to Huntington Road being flooded, which is a road stretching over multiple kilometres, causing locational errors of Tweets referring to this road to be large. To review the effect of the type of locational reference on the locational uncertainty, the dataset was split up based on these references (figures 44 and 45). Neighbourhoods were not referenced in the York dataset[3].

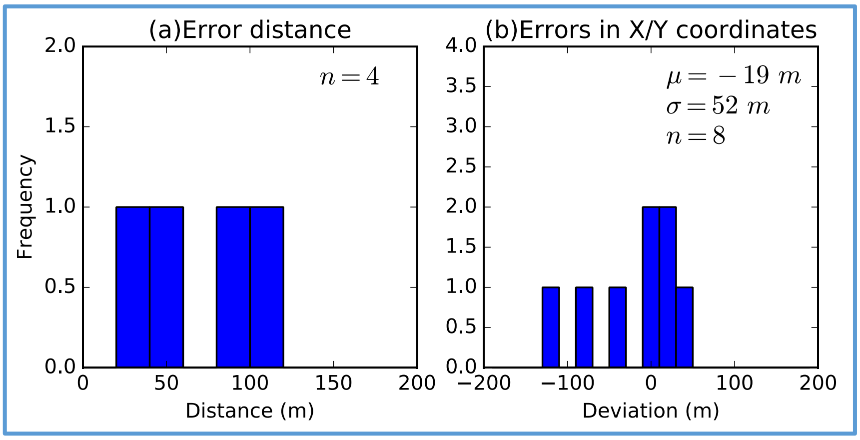

Figure 44: York - Locational errors of Tweets (Mentioning POIs), bin size = 20 m

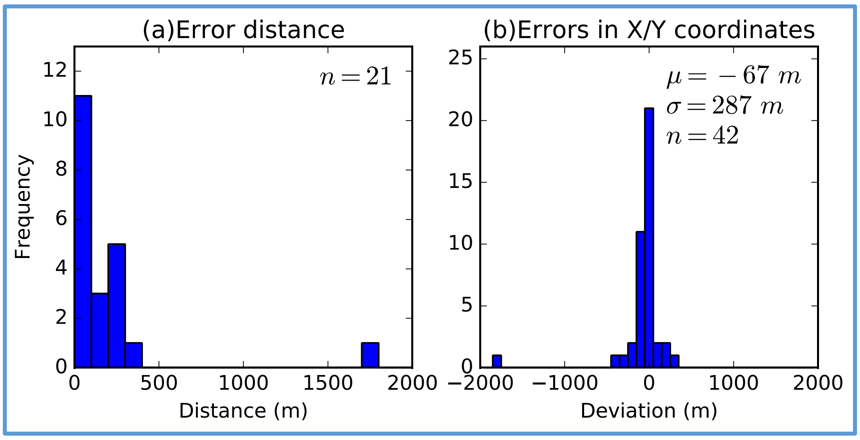

Figure 45: York - Locational errors of Tweets (Mentioning Streets), bin size = 100 m

Although the dataset was quite small, the locational errors of Tweets mentioning POIs were considerably lower than of those mentioning streets. This was confirmed by comparing the X/Y variance of both sets using an F-test (95% confidence limit, see appendix B). This is likely a result of the different methods used to derive locations from the Tweets, as already extensively discussed for the Jakarta case in paragraph 3.1.2.

[3] One observation with photograph however referred to an intersection of roads

A difference between the cases however, is in the magnitude of the errors. Where an overall standard deviation in locational error of about 500 m was found for the Jakarta case study, it is limited to 260 m for the York case study. The variances of both the Tweets referring to POIs as well as Tweets referring to streets were found to be significantly lower for the York case study compared to the Jakarta case study (F-tests,95% confidence limit, see appendix B). This difference between the cases is likely a result of the difference in size between both cities. It was found that the streets Tweets in the York dataset referred to were generally shorter than the streets referenced in the Jakarta case. This causes the locational errors for Tweets containing these street references to drop considerably.

4.2 Flood mapping

Three different methods to create the flood maps were compared for the York case study (see § 2.3.2). The optimal set of parameters for each method to create the maps was determined by comparing the resulting flood maps to the recorded flood extents of the EA (see § 2.2.3). Since the results close to the edges of the maps were neither of interest, nor reliable, only cells within the outline of Central York (figure 3) were used in evaluating the performance of the maps.

4.2.1 Plain interpolation

The interpolation method considered uses all observations within the study area at once in interpolating the water levels. After this interpolation the flooded areas that are not directly connected to any of the observations are removed from the maps. The best map for the York case study was created by using a power parameter of 3 in combination with a smoothing of 200 m. Because none of the observations for the York case study mentioned a water depth, all observations got a DWD assigned. Using a DWD of 20 cm gave the best result (figure 46).

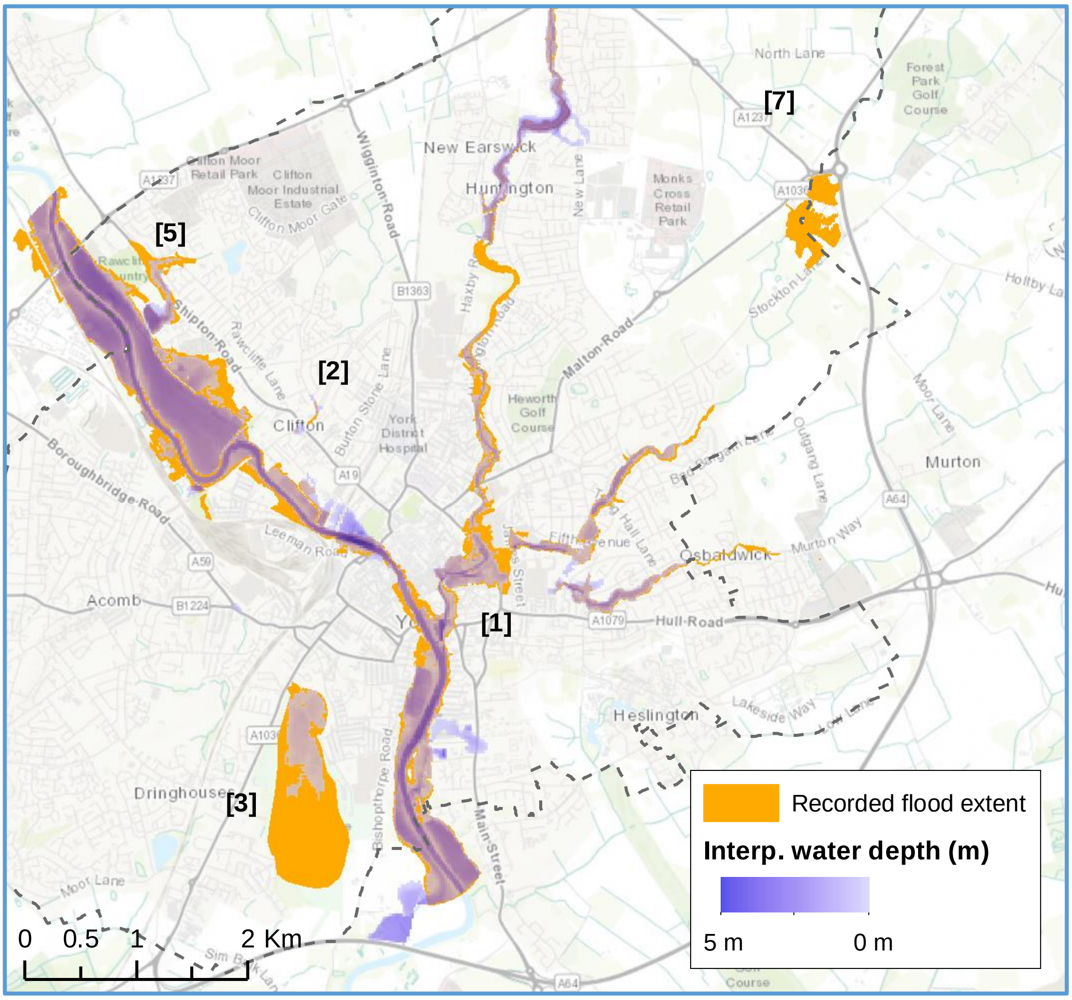

Figure 46: York – Flood map created using plain interpolation (power = 3, smoothing = 200 m) and a DWD of 20 cm.

. At most locations the calculated flood extent is within the flood extent recorded by the EA, meaning flood extents are generally underestimated using this method. This is especially the case at location [3], which is flooded separately from the river. Because the plain interpolation method also uses the observations close to the rivers to determine the water levels in this area, water levels and therefore flood extent are severely underestimated at this location. Also a part of the river Foss (north-east of location [2]) is not mapped flooded at all. Around the inner city (location [1]) the results are better, where especially the western part corresponds well with the recorded flood extents. This recorded flood extent is also well approximated downstream of the inner city and at locations [2] and [5]. The area south of location [7] however is not mapped flooded, since there were no observations in this area

All in all though, results are quite good and large deviations only arise in areas that should be interpolated separately, such as area [3]. The consistent underestimation of flood extent however affects the value of the F-statistic, which is only 0.58. This means that the cells that are indicated as being flooded in both the created flood map and the recorded flood extent make up only 58% of the total number of cells that is mapped flooded in either the flood map or the recorded flood extents. Although choosing a higher DWD reduces the underestimation in some of the areas, this was not used since it would also cause large areas to be erroneously mapped flooded.

4.2.2 Grouped Interpolation

The results obtained using interpolation of groups were nearly identical to those of the plain interpolation method. Although grouping the observations based on their downstream flow paths produced better results than grouping the observations based on proximity, the interpolation extents of the two groups fully overlapped, meaning the areas were not separated in interpolation. Again the most optimal combination of parameters was found using a power parameter of 3 and a smoothing of 200 m, yielding an F-value of 0.58. Because this map is identical to figure 46, it is not included in the main text of this report (see appendix B).

4.2.3 Interpolation along flow paths

For the last interpolation method evaluated in the research, observations were also grouped based on their downstream flow paths. After grouping, water levels were interpolated along these flow paths, and the water levels of each cell on the flow path was also given to the cells upstream of it. It was found that using a power parameter of 4, a smoothing of 600m and a DWD of 50 cm gave the best result (figure 47).

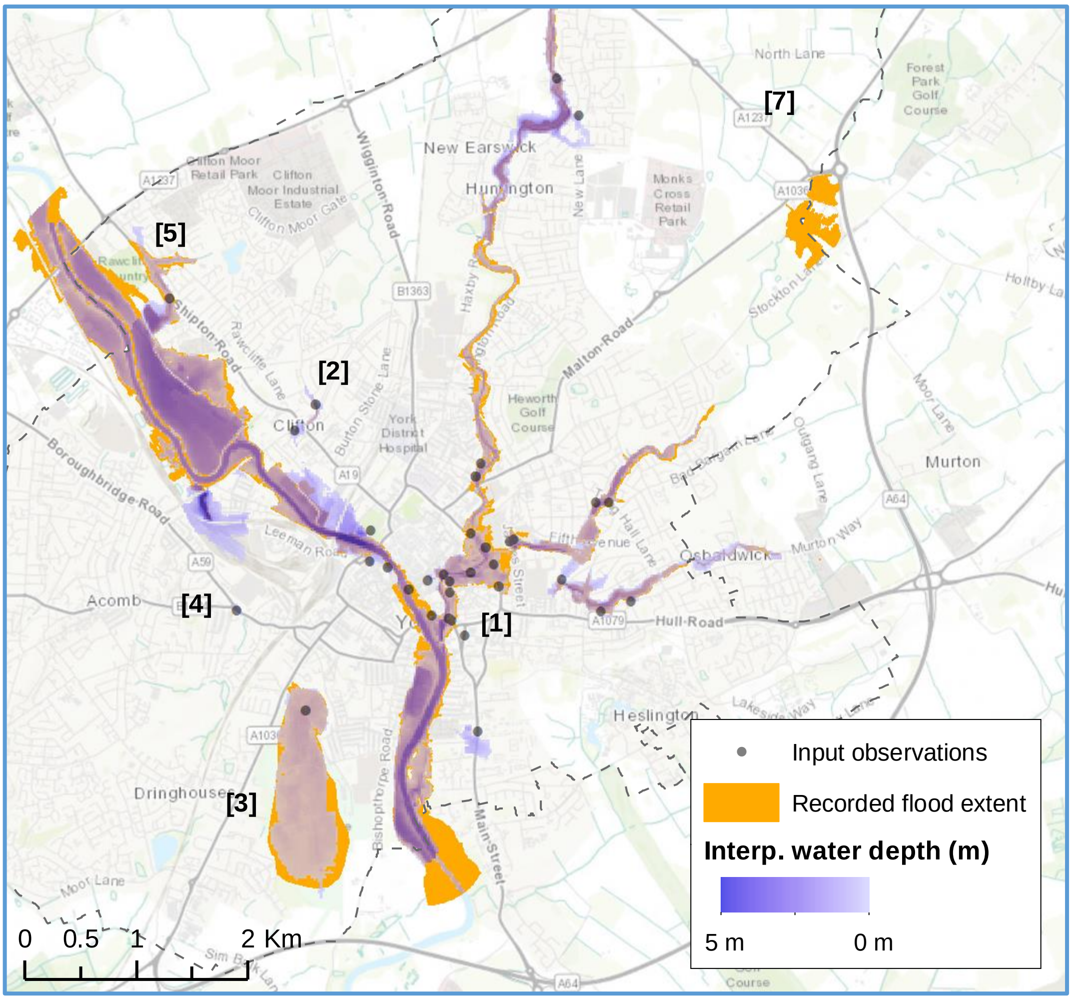

Figure 47: York - Flood map created using interpolation along flow paths (power = 4, smoothing = 600 m) and a DWD of 50 cm.

The most significant result of strictly separating observations that belong to separately flooded areas, is seen at area [3], where the flood extent is reproduced better. North of location [4] however, the flood extent is considerably overestimated. This is partly due to the grouping procedure, which groups the observation to the right of location [4] with the other observations around the river. The flood extent in the remainder of the area however corresponds quite nicely to the recorded flood extent. Still the flood patterns at locations [2] and [5] are well reproduced, and the result at the eastern part of the inner city (location [1]) has improved with respect to using plain interpolation. Where the

Table 2: York - Contingency table containing areas mapped flooded in the flood map and the observed flood extent in km^2

	Obs. wet	Obs. dry
Mapped wet	4.00	0.59
Mapped dry	1.19	63.32

improvement around area [3] was mainly due to the strict grouping of observations, the flood extent in the rest of the area was improved by increasing the DWD, which reduced the underestimations that were present in the map created using plain interpolation. The contingency table (table 2) however indicates that the flood extent is still considerably underestimated. This is largely because of the underestmation of flood extent to the south of location [7], due to the absence of observations near this location.

4.3 Uncertainty assessment

The uncertainties in the flood maps resulting from locational errors and errors in the elevation model were assessed using Monte Carlo simulation. Additionally the effect of changes in DWD and map resolution on the flood maps was reviewed. Errors in the water depth mentioned by the Tweets were not evaluated, since none of the Tweets for the York case study mentioned a water depth.

4.3.1 Quantification of uncertainties

The information about the magnitude of locational errors from paragraph 4.1.2 was used to derive a probability distribution to simulate locational errors. Information about the magnitude of elevation data errors however was derived from literature, as discussed in paragraph 2.3.3. These errors were simulated using a mean error of 0 m, a standard deviation in error of 20 cm and a correlation distance of 100 m.

Based on the results of paragraph 4.1.2, locational errors were simulated using a mean of 0 m and a standard deviation in error of 300 m. This value is different from the one used for the Jakarta case study, since it is expected the difference between the cases were caused by the differences in the length of streets between the two case studies.

The number of simulations necessary to evaluate the impact of each source of uncertainty was reviewed by determining the standard deviation of the standard deviations of individual grid cells or observations. This standard deviation was recalculated after adding each realization and at the point the line stabilizes or reaches zero, adding new observations does not significantly add to the probability distribution of error, meaning enough realizations were generated. The results of this analysis are given in figure 48.

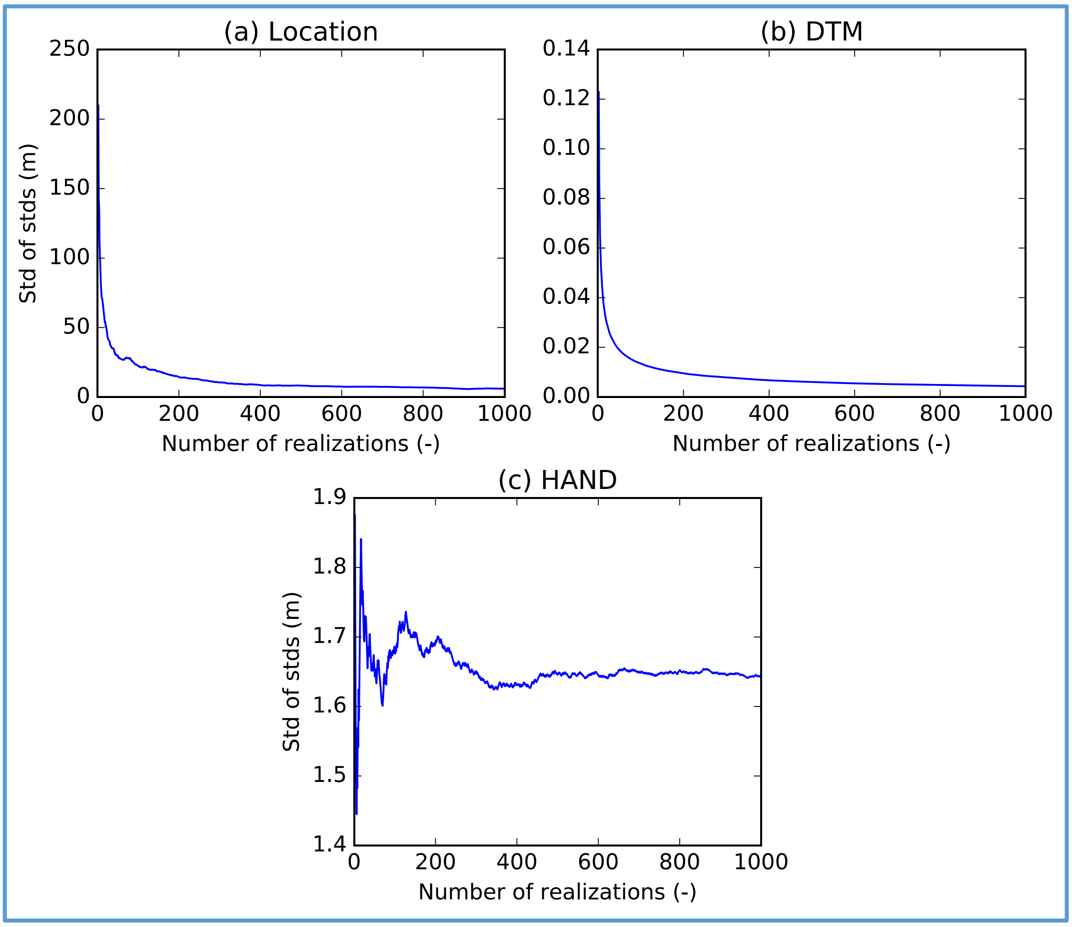

Figure 48: York - Standard deviation of standard deviations of individual Observations/grid cells

Like for the Jakarta case study the graphs of the DTM (b) and the HAND (c) were very different, due to the errors of a single drainage cell affecting its entire upstream area in the HAND map. Nevertheless all graphs stabilized at about 600 realizations and therefore 600 random simulations were used to assess the impact of each individual source of uncertainty. To simulate the effect of the combination of both error sources on the uncertainty in the flood maps, a significantly higher number of simulations is necessary to ensure all possible combinations of errors are evaluated. However, due to time and computational constraints, the number of simulations run to assess the impact of the combination of error sources was limited to 1000. By creating multiple uncertainty maps using the same settings and 1000 simulations, the differences between the maps were found to be negligible.

4.3.2 Locational uncertainty

Using Monte Carlo simulations, the uncertainty caused by locational errors was reviewed. The result of these Monte Carlo simulations is given in figure 49.

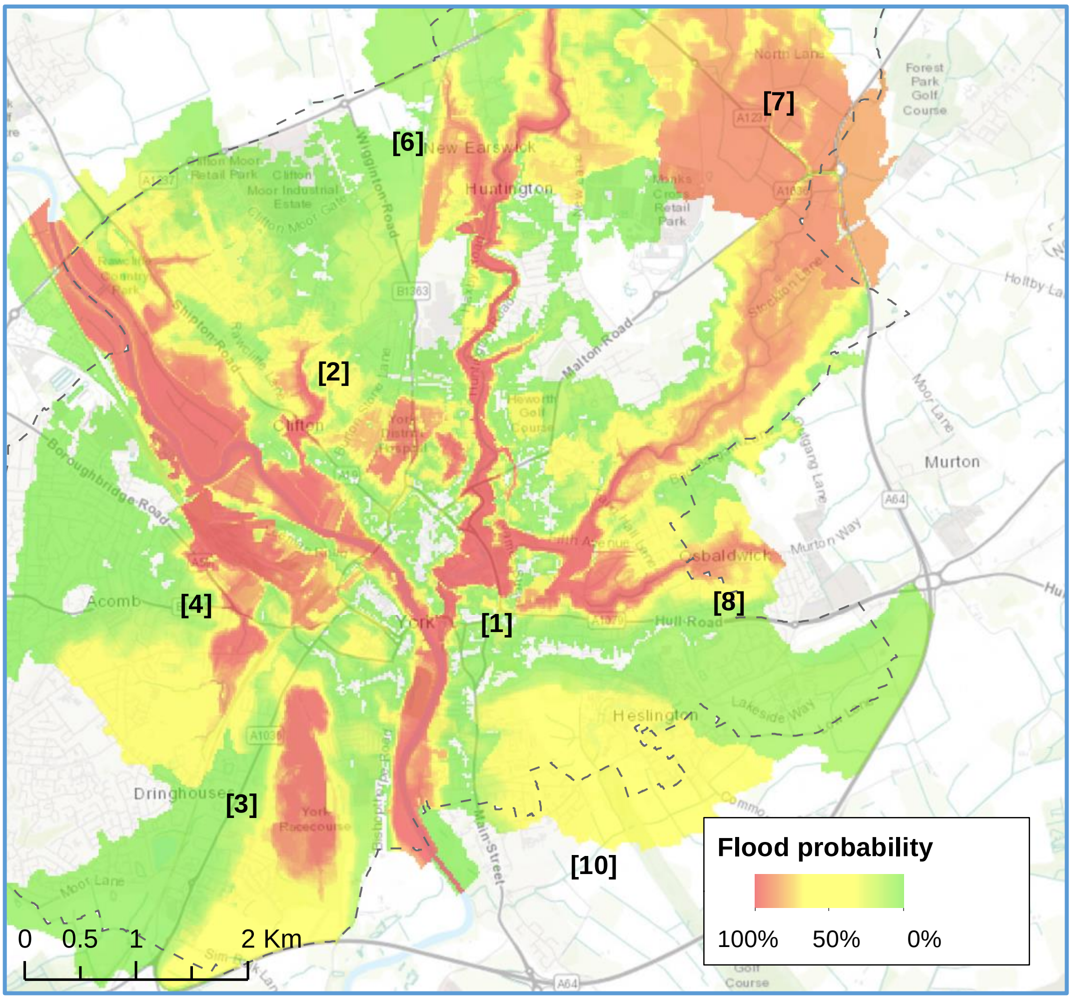

Figure 49: York – Probability of flooding determined by reviewing the impact of locational errors using a Monte Carlo analysis. Location [1] – [4], [6] – [8] and [10] are discussed in text.

This result illustrates that locational errors cause considerable uncertainty in the flood maps. Although for locations having a flood probability close to 100 or 0% it is quite certain that the area is either flooded or not flooded respectively, the uncertainty in flood extent is largest in areas which are flooded in close to 50% of the simulations (given in yellow). As a result of locational errors, there are multiple locations were the uncertainty is high, for example at location [8], south-east of location [2], south of location [4] and at location [10]. What these areas have in common is that they are generally flat and are often located just next to an area with higher elevation (see figure 50). Therefore locational errors can move observations to positions with considerably higher ground elevations, leading to large changes in water level. This change in water level has a large influence on the flood extent in flat areas. The same effect can be seen to the south of area [3]. At location [6] and [7] however, it can be seen that large areas are assigned a high probability of flooding, although they were not flooded in reality. Both of these areas have a sill in elevation in front of the area. A change in water level caused by a locational error of an observation can therefore cause the entire area to flood. At the inner-city of York however (location [1]), uncertainties remain small due to the moderate slope in the area, and the fact that the area around the river is lower than its surroundings, which effectively restricts flood extent.

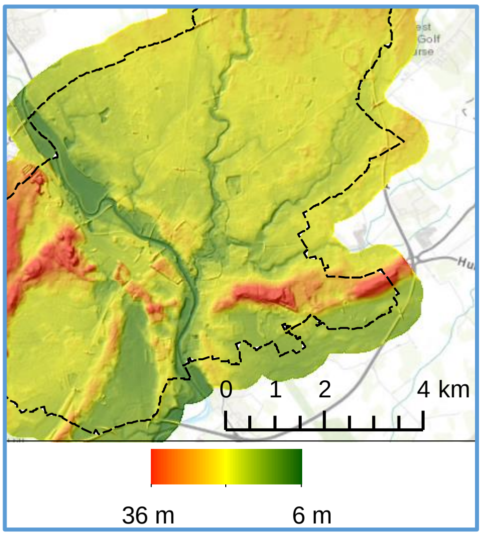

Figure 50: York - DTM of the study area

4.3.3 DTM uncertainty

Also the uncertainty caused by errors in the elevation data was evaluated using Monte Carlo simulations. The result of running 600 of them is given in figure 51.

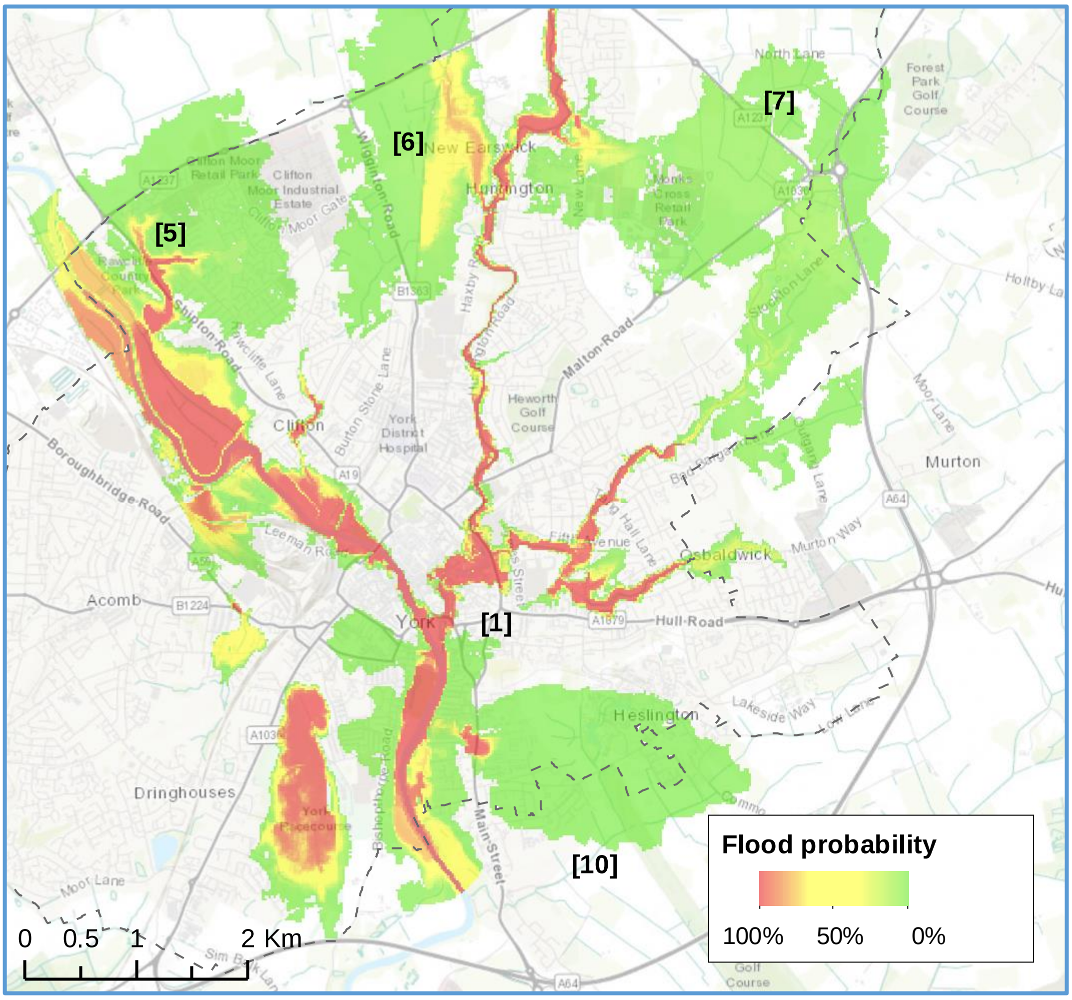

Figure 51: York – Probability of flooding determined by reviewing the impact of elevation data errors using a Monte Carlo analysis. Location [1], [5] – [7] and [10] are discussed in text.

The uncertainty caused by errors in the elevation data generally remains limited. Some large areas however are still flooded in a limited number of simulations. This is for example the case at locations [6], [7] and [10], where locational errors remove sills in topography or breach barriers in a very limited number of simulations. At location [5] the only observation in the area is sometimes raised by an error in the elevation data, causing it to flood in a limited number of simulations. Nevertheless the total uncertainty caused by locational errors is very limited. Especially at the inner-city of York, north of location [1], the uncertainty is small.

4.3.4 Combination of uncertainties

The uncertainty caused by both the locational errors as well as the errors in the elevation data was evaluated using Monte Carlo simulation. Running 1000 simulations gave the result in figure 52.

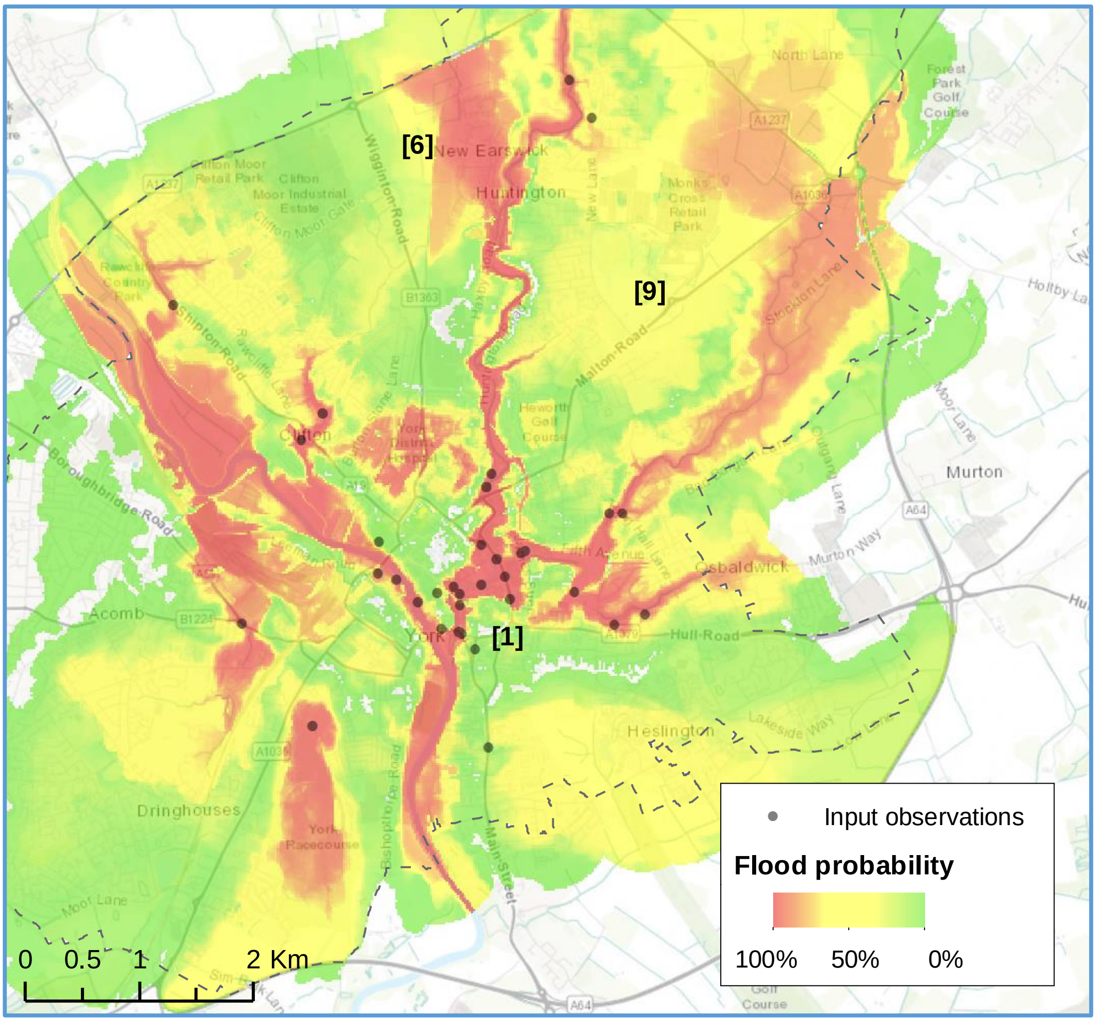

Figure 52: York – Probability of flooding determined by reviewing the impact of both locational errors as well as elevation data errors using a Monte Carlo analysis. Location [1], [6] and [9] are discussed in text.

Overall the result is very similar to simulating purely locational errors. Nevertheless the two sources of error interact in some places. For example location [9] was not flooded in any of the maps created using the individual error sources, but is highly uncertain when reviewing both sources of uncertainty. This additional uncertainty is caused by the fact that by combining the error sources not only the location of the observation changes, but also the downstream flow path of an observation can change significantly.

Also at location [6] the result of the interaction between the two error sources is also clearly visible. Here the area which already flooded in a large number of simulations when reviewing solely locational errors is extended because errors in the elevation model additionally cause the barrier in the back of the area to breach. At most other locations however, the result is very similar to the map of uncertainty created using locational errors only, illustrating that locational errors are an important source of uncertainty for the York case. Just like with the map created using the locational errors, the uncertainty around the inner-city (location [1]) remains limited. However, just like with the Jakarta case study, it can be seen that the uncertainty in flood extent is hardly influenced by the presence of observations. There are large areas having a high probability of being flooded in which only a few observations are located, and also some observations at locations where the probability of flooding is fairly low. This illustrates that solely the results of a Monte Carlo analysis do not give a complete picture of the actual uncertainty the flood maps generated using Tweets.

4.3.5 Varying default water depth

A DWD was added to all of the observations for the York case study, since none of them mentioned a water depth. The effect of choosing a different DWD value was evaluated by reviewing the result of varying the DWD from 30 cm below (20 cm) the most optimal value (50 cm) to 30 cm above (80 cm). Figure 53 gives the resulting flood extents.

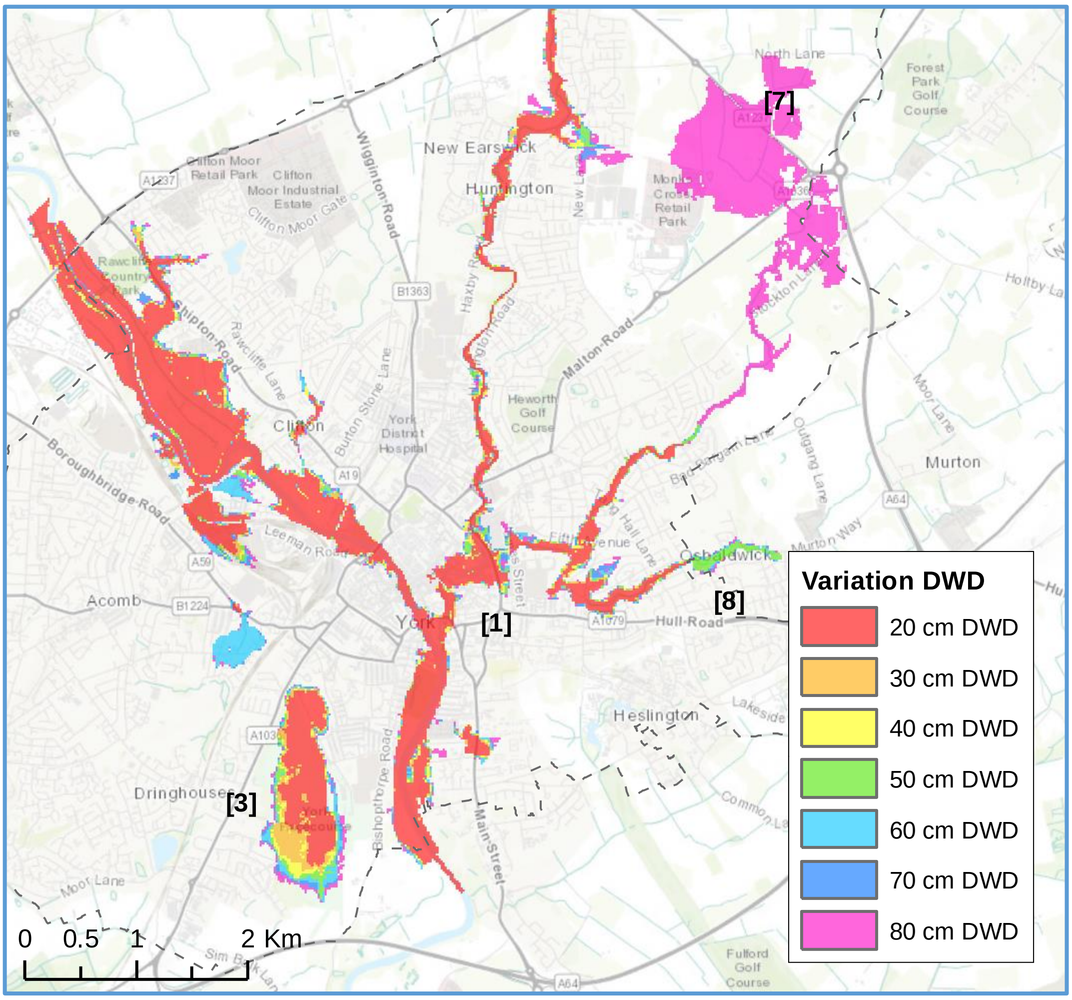

Figure 53: York - variation in flood extent as a result of varying the DWD. The red area gives the flood extent in case a DWD of 20 cm is used, and the other colours show the increases in flood extent due to increasing the DWD. Locations [1], [3], [7] and [8] are discussed in text.

The result indicates that some areas are more sensitive to changes in DWD than others. Areas [7] and [8] for example flood as soon as a certain water depth is exceeded. Also the small terrain slope to the south of area [3] causes this location to be quite sensitive to changes in DWD. At most other locations however, the changes in flood extent are only minor. This is especially true for the inner-city of York (location [1]).

Nevertheless the effect of using DWDs of 20 and 80 cm on the combined uncertainty map was reviewed. The resulting maps are given in figures 54 and 55 respectively. Although for the map created using an 80 cm DWD, the total surface area that is flooded in a majority of the simulations, given in red, increases somewhat, no major differences are seen. Only to the southeast of location [2] there is a considerable change in the probability of flooding between the low and high DWD values. A comparison to the original map of combined uncertainty (figure 52) indicates however that the differences are small. This is especially true near the inner city (location [1]).

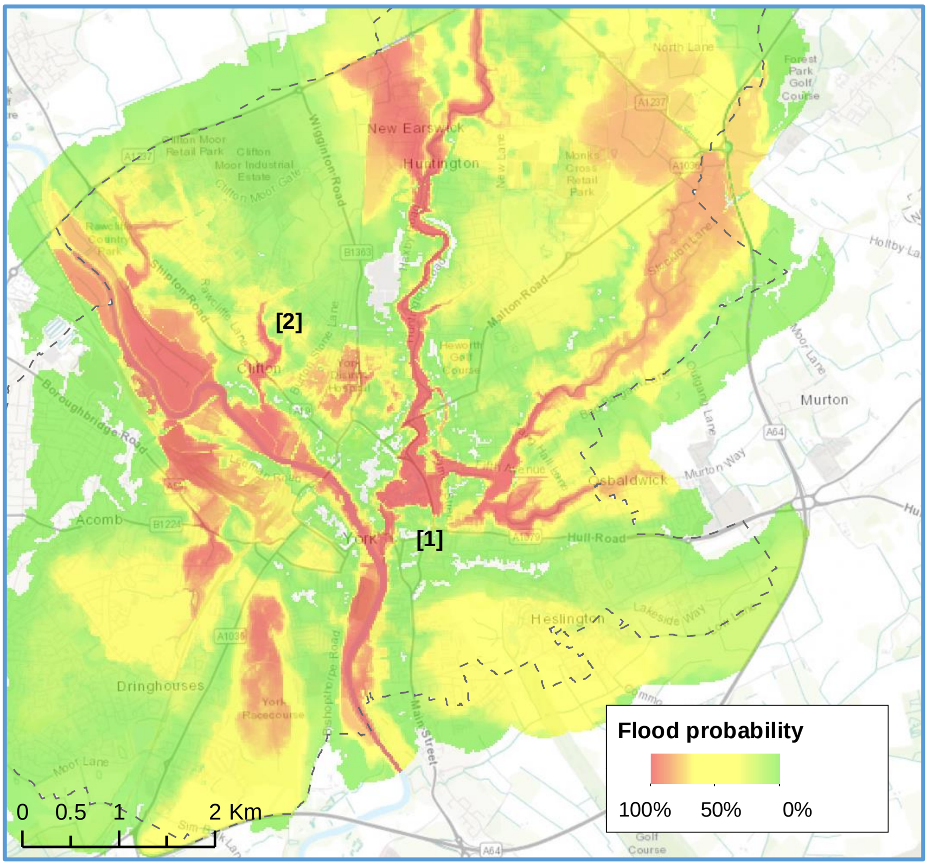

Figure 54: York – Probability of flooding determined by reviewing the impact of both locational errors and elevation data errors using a Monte Carlo analysis (DWD = 20 cm). Location [1] and [2] are discussed in text.

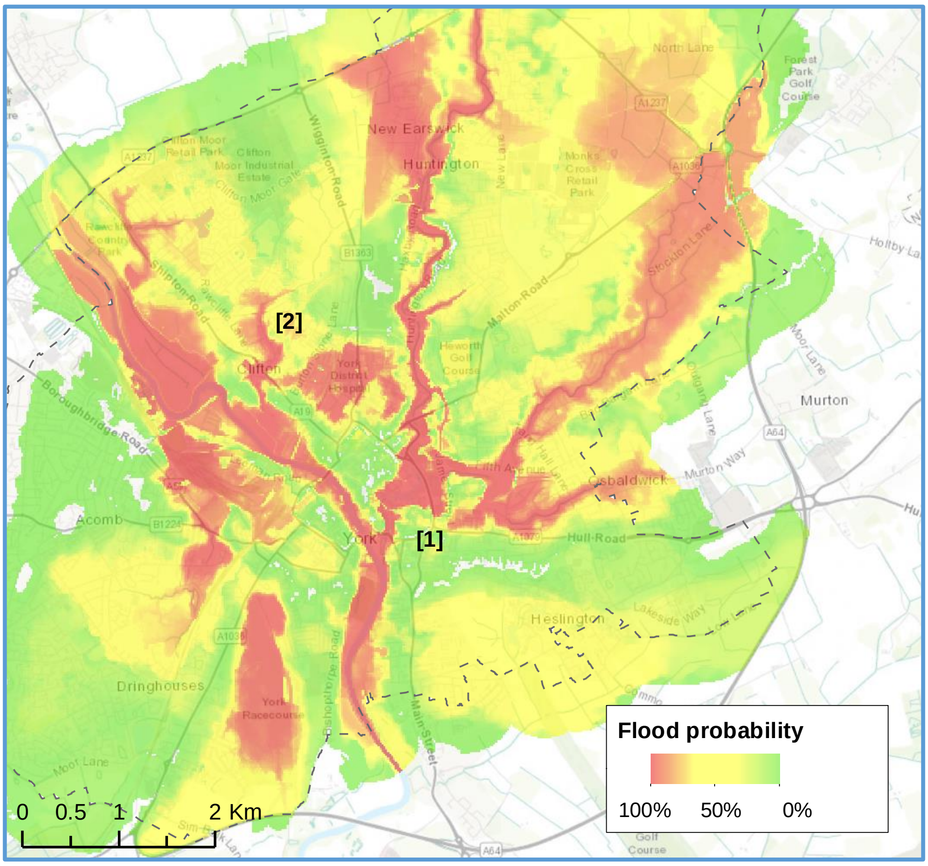

Figure 55: York – Probability of flooding determined by reviewing the impact of both locational errors and elevation data errors using a Monte Carlo analysis (DWD = 80 cm). Location [1] and [2] are discussed in text.

4.3.6 Varying DTM resolution

The impact the choice of resolution (20 m) had on the flood maps was also reviewed. Therefore additional flood maps were created at a 6 m, 10 m and 40 m resolution (figure 56). How these maps were created is described in appendix B.

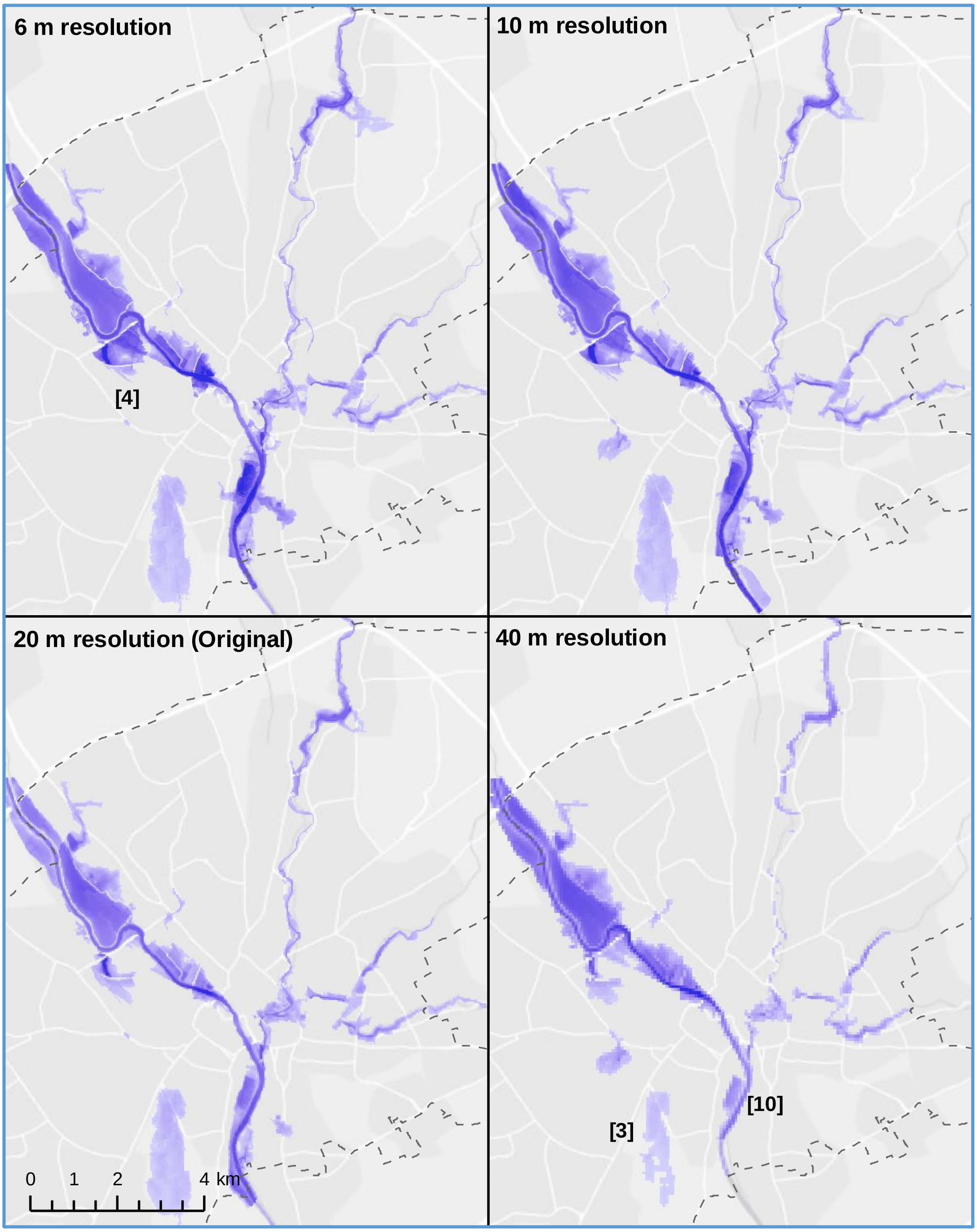

Figure 56: York – Flood maps created using interpolation along flow paths at 6 m, 10 m, 20 m (original) and 40 m resolution

The overall flood pattern is the same at each resolution. Especially the 6 m and 10 m resolution maps are very similar to the 20 m resolution map. The only major difference occurs at location [4], where a somewhat larger area is flooded at higher resolutions. This is because a barrier is breached at these resolutions, due to the resampling from the 2 m DTM. Using a 40 m resolution however seriously affects the quality of the flood map, since at locations [3] and [10] large areas are not flooded anymore. These results indicate that using a resolution higher than 20 m does not significantly affect the mapped flood extent. Secondly, since the 40 m map shows degradation in performance, it confirms the choice of using a 20 m DTM to create the flood maps for the York case study.

The uncertainty caused by the combination of locational errors and errors in the elevation data was also re-evaluated at 6 m and 40 m resolutions. The results are given in figures 57 and 58 respectively. When using a 6m resolution, some barriers in the area are better represented. For example at location [7], the area which has a high probability of flooding is more or less restricted to the area which is given in the validation data. Also the area at location [6] has flood probability that is considerably lower compared to the 20 m resolution map. It can also be seen that the extent of the area with a high probability of flooding around the rivers is significantly reduced. This is especially clear by comparing the map to the 40 m resolution map. At 40 m resolution also the flood probability of area [9] is significantly reduced. Nevertheless the total uncertainty in the complete map remains more or less the same among each resolution. Although in the 40 m resolution map for example, the uncertainty at location [9] is lower, it is higher at again at location [11]. Furthermore the uncertainty at the inner city (location [1]) is limited for the maps at all resolutions.

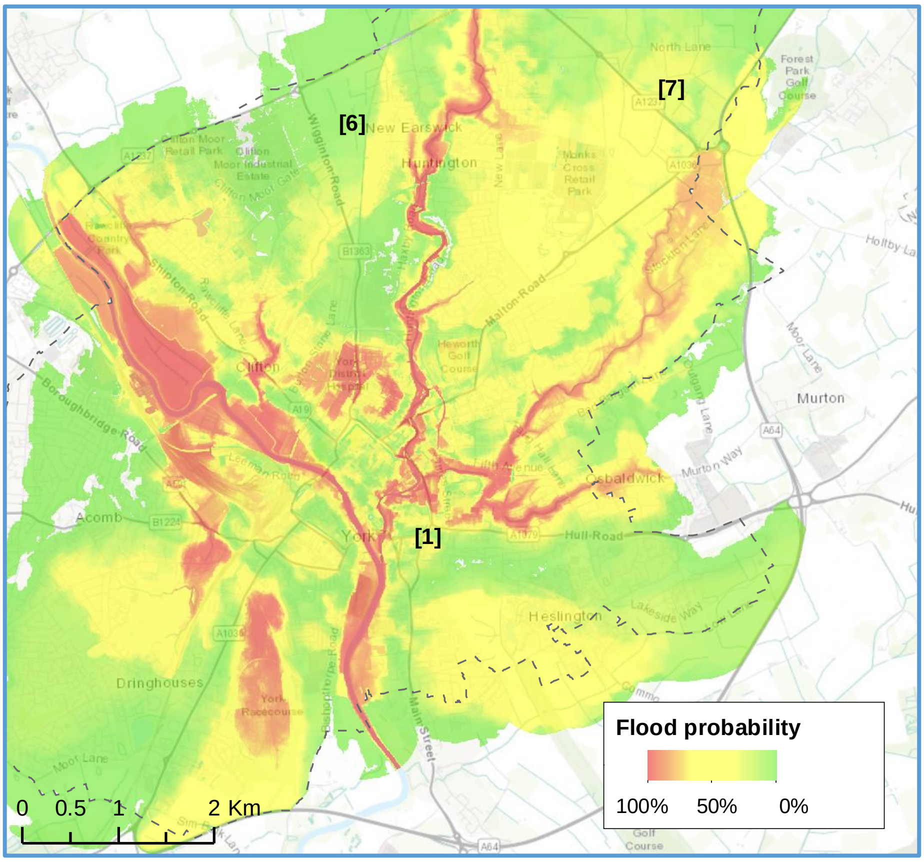

Figure 57: York – Probability of flooding determined by reviewing the impact of both locational errors as well as elevation data errors using a Monte Carlo analysis (6m resolution). Locations [1], [6] and [7] are discussed in text.

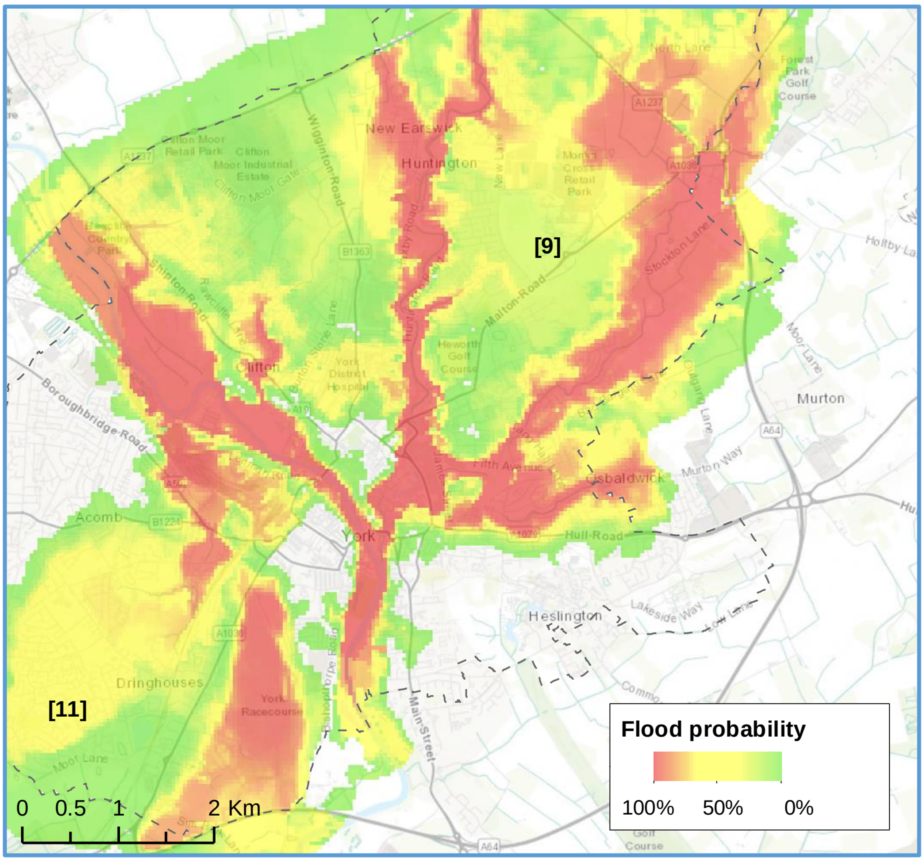

Figure 58: York – Probability of flooding determined by reviewing the impact of both locational errors as well as elevation data errors using a Monte Carlo analysis (40m resolution). Locations [9] and [11] are discussed in text.

5 Discussion

The objective of the 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"*. This research succeeded in formulating a preferred flood mapping method, which improves upon previously employed methods. The uncertainties in the maps originating from a number of error sources were also assessed, although this does not yet accurately reflect the actual uncertainty in flood extents. Using an analysis of the data and the results of the uncertainty analysis also the scale at which the method could be applied and the potential of real-time application were reviewed. The flood mapping methods and their results as well as the uncertainties in the maps and the applicability of the maps are further discussed in this chapter.

5.1 Flood mapping

Several different ways of interpolating water levels were evaluated. However, each of these techniques was based on IDW interpolation. Although IDW is used in some other studies in which water levels are interpolated (e.g. Palasceanu & Pearlstine, 2008; Werner, 2001), also many other interpolation methods can be used, such as spline interpolation (Fohringer et al., 2015) or different types of Kriging (Yan et al., 2014). However, the differences caused by using another interpolation method are likely small in areas where a large number of observations are available. Also the IDW interpolation allowed for smoothing, which was important for creating the best results, and cannot be done with every interpolation method.

IDW interpolation was applied in several different ways. For both case studies, the best results were generated by interpolating along flow paths, although the parameters used to obtain these results differed. For the Jakarta case study (power = 2, smoothing = 200m, DWD = 30 cm) the power parameter, smoothing and DWD were lower than those of the York case study (power = 4, smoothing = 600 m, DWD = 50 cm). It is expected these differences are a result of the differences in topography between the cases. The most densely populated areas within Jakarta are relatively flat, which cause small water depths to lead to large flood extents. This can be the reason that the DWD for the Jakarta case was lower. Since lower water depths already cause considerable flood extents in Jakarta, more Tweets about small flood depths are posted. For York however, large areas are only inundated when water depths are higher. Also the higher power parameter for the York case can be a result of the steeper slopes, because a higher power parameter makes local observations more important, which is important if the differences in water levels are large.

Whether the same parameters could be used for future flood events was however not investigated. Effectively only a calibration of the method was performed, without a real validation of its results. To be successful in creating flood maps, the parameter values should be known prior to the flood event. To further investigate this both floods at different locations as well as multiple flood events at the same location should be reviewed. By both calibrating and validation the method, it can be seen if the interpolation parameters purely depend on the topography of the area and if they also change between flood events. Ideally the optimal parameter set is more or less constant between flood events, since only then it can successfully be applied. Additionally an analysis of flood events af different locations will indicate if the parameters indeed are related to the topography of the area, meaning that the optimal parameter values can be estimated in advance based on this topography.

Despite the same method producing the best results for both cases, the quality of the maps was different. Although the maps could not be objectively compared because different methods and datasets were used to analyse the performance of the maps, it is expected the results for the York case study were superior to those of the Jakarta case study. This is surprising, since for the York the number of Tweets was less and Tweets were less detailed. However, the quality of the maps created using Tweets largely depends on the topography of the study area. Especially the inner city of York, where the best results were found, is located lower than the terrain surrounding it, meaning the flood extent is well constrained by the surrounding terrain. The most densely populated area in Jakarta however, was rather flat, meaning small differences in water levels seriously affected flood extents. Also the fact that for Jakarta a 2 m DSM first had to be converted to a 2 m DTM, could have caused some important barriers to be missing in the DTM used for the Jakarta case, which especially affects flat areas.

Since the flood maps between the cases could not be objectively compared, the results are compared to other studies. Eilander et al. (2016) for example used a validation method similar to the one used for the Jakarta case study. They found that 69% of the flood locations they had derived from photographs were within 500 m from the modelled flood extent. Given 75% of validation points exactly overlapped with the flood maps created in this research, the results of using the methods described in this report are considerably better. However, Sun et al. (2015) who compared their satellite derived flood maps with geo-tagged Flickr images, found that 95% of the images intersected with their mapped flood extent. It is however expected that the best flood maps are always created from remote sensing data, because these are direct observations of the flood extent in the area. Nevertheless, all of the above values only assess the errors made due to underestimations of flood extent, but say nothing about overestimations of flood extent. ME and RMSE values are however also mentioned in literature. Neal et al. (2009) for example compared their hydraulic model simulations for the city of Carlisle with gauge measurements, and found an RMSE of 0.28 m and an ME of 0.06 m. Comparing this with the RMSE and ME values found for Jakarta, 0.44 m and 0.01 m respectively, indicates that the maps created for Jakarta are not considerably worse. Also the performance of the Twitter flood map for Jakarta was found to be a lot better than the BPBD map, which is the flood information that is currently available.

For the York case study an F-statistic of 0.69 was found using the best performing method. Comparing this to a variety of modelling studies shows that this result is not bad. Bates & De Roo (2000) for example reviewed the performance of the 2D LISFLOOD-FP model at 25 m resolution, and found F-values of up to 0.82. Horrit et al. (2007) on the other hand found values ranging from 0.65 to 0.89 evaluating 2-D finite volume and finite element models at roughly 50 m resolution for a number of different discharge values. Where both these studies however predominantly focus on rural areas, Ozdemir et al. (2013) looked at a densely populated urban area and applied the LISFLOOD-FP model at high 1 m resolution. For different points during the flood wave they found the F-statistic to vary between roughly 0.7 and 0.85, depending on the moment in time. They however used model results of the same model at 10 cm as a reference, which has likely led to an overestimation of model performance. Although the value of 0.69 found for the York case is lower than most of the studies above, it is not considerably worse. Furthermore, a higher F-statistic can be found if purely reviewing the inner-city of York. These results are not presented separately here however, since the arbitrary choice of selecting the inner city of York, seriously affects the F-value found.

The performance of the maps created for both case studies was likely influenced by the fact that Tweets over the entire course of the flood were used, whereas the maps were validated using the maximum flood extent. For example in case many water level observations of different times are placed at roughly the same location, meaning that there are both low and high water levels at this location, the current method averages these water levels, which can lead to an underestimation of flood extent. If the observations are further away from each other, this likely only leads to unrealistic slopes in the calculated water level causing unrealistic water depths, but has only a small influence on the actual flood extent. Using observations over a shorter time period, for example in real-time application (also see §5.3), might improve the resulting flood maps. The datasets used for this research however, were too small to split up into smaller subsets, meaning the flood extents calculated in this research are probably more reliable than the actual water depths calculated.

Other improvements, such as using additional sources of data, might also lead to increased performance of the method. For example, using measured river water levels, which in the UK is supplied at a 15 minute interval, can lead to better flood maps. Basically all additional point observations of either water levels or water depths can easily be integrated in the interpolation procedure in order to improve the results.

5.2 Uncertainties

The main focus of the research was on evaluating the uncertainties in the flood maps. The uncertainties in the Tweets were analysed by comparing the location and water depth information derived from the Tweet itself, to information derived from the photograph attached to the Tweet. Tweets having photographs attached however are less likely to give false reports of flooding, whereas Tweets without one can erroneously refer to floods. Therefore the magnitude of errors might have been underestimated by deriving them using the photographs attached to the Tweets.

On the other hand however, the use of photographs might at the same time have led to a considerable overestimation of the magnitude of locational errors. For example, if a Tweet reports the flooding of a seriously affected and therefore inaccessible area, the photograph might originate from a more easily accessible location, whereas the location derived from the Tweet might still refer to a correct location of flooding. Also the way in which locational errors were simulated, might have contributed to an overestimation of the uncertainty caused by them. The analysis of the Twitter dataset indicated that locational errors often arose because a Tweet referring to a street was pinpointed to a wrong location along this street. Nevertheless these errors were simulated by adding random errors to the X/Y coordinates of the observations, whereas it would have been more appropriate to define errors along the street. This can for example be done by using the data about streets and neighbourhoods from OpenStreetMap, and might be a useful addition in future research studies. Overall it is estimated the combination of the methods used to derive errors from the messages and simulate the resulting errors using Monte Carlo simulation used in this research, caused a considerable overestimation of the uncertainty resulting from these locational errors.

Also the errors in the elevation model might have been overestimated as a result of the procedure used to generate the random realizations of the HAND maps. In creating these random realizations, the threshold used to identify the locations of drainage channels was kept constant. In reality, if a DTM with errors is used, the threshold will be adjusted to accurately reflect the main drainage channels in the area. Therefore additional errors were introduced, which would not be there in reality. For example, if the threshold used is too low, drainage channels are identified at additional locations, causing these to flood. Overall however it is expected that the influence of this overestimation of errors on the uncertainty in flood extents was only minor, especially compared to the overestimation of locational errors discussed above.

Although the method used to simulate these locational errors might have led to a considerable overestimation of the uncertainty caused by these errors, it is still expected these errors still have a considerable share in total uncertainty. Just like the errors in the elevation data, these location errors mainly affect more flat areas, and led to limited uncertainty in flood extent for areas with steep slopes. These findings illustrate that the degree to which errors in the input propagate to the flood maps depends heavily upon the topographical characteristics of the area. The upstream area of Jakarta for example has steep slopes, due to which the uncertainty caused by errors in the datasets was low. These topographical characteristics will also affect the choice of resolution. Although a 20 m resolution was found to produce good results for the York case study, it was still seen in the uncertainty analysis that some barriers were omitted at this resolution, causing the flood probabilities at these locations to be overestimated. Although the effect on York was limited, it is expected that especially for the Jakarta case study many small barriers in the downstream area were omitted from the DTM, causing the large uncertainty in flood extent in downstream Jakarta.

The topographic characteristics of an area can however not be changed, and can therefore only be used to assess to which extent errors in the input data will lead to uncertainties in the flood maps. The locational errors in the data, which were found to cause considerable uncertainty in the flood maps for both case studies, can however be altered. Considering Tweets which refer to POIs generally contain more accurate locational references than Tweets referring to streets or neighbourhoods, interacting with Twitter users to actively ask for detailed locational references might help in reducing locational errors.

It is also interesting to see if already in place information, such as geo-tags, can be used to improve the positioning of individual observations. For example if geo-tags indicate a location on the street being referenced in the Twitter message, this information can be used to improve the location of the Tweets. Also a more fundamental aspect of the method used to convert locational references to streets and neighbourhoods to a point location should be further reviewed. Primarily street and neighbourhood references were pinpointed to sinks, whereas with fluvial floods only looking at the point of lowest elevation might produce better results. Another way of reducing the impact of uncertain observations referring to streets and neighbourhood is to already use the information about the type of locational reference in flood mapping. For example the likelihood of an observations can be reflected in the interpolation weights, meaning an uncertain observation (referring to a street or neighbourhood) will affect the interpolated water levels less than an more certain observation (referring to a POI). In real-time application, this likelihood and therefore the weight an observation gets can additionally reflect the time at which the observation was posted, giving a higher weight to more recent observations.

A factor that is likely to cause only minor differences in the uncertainty in the maps created by the Monte Carlo analysis, is including more observations. For both cases it was found that the uncertainty maps created using Monte Carlo analysis were quite insensitive to the density of observations, with some areas containing only very few observations but having a large probability of flooding, and others containing multiple observations but having a low probability of flooding. Although it is not reflected in the Monte Carlo analysis results, the presence of clustered observations makes it more probable an area is flooded and combining the water level information in these observations can give a better estimation of the actual water levels. This last element is actually accurately reflected when purely doing a Monte Carlo simulation of water level errors. In case many observations with water level errors are in close vicinity of each other, the interpolation method will average their values, meaning that errors are more or less filtered. However, the locational errors of many observations which are originally in close vicinity will not cancel each other out, since the locational errors added to them cause them to shift. Especially this last characteristic of the Monte Carlo analysis causes the calculated uncertainties to not accurately reflect the density of observations. This should be included however, to accurately represent flood extent uncertainty. This can for example be done by reducing the locational errors of observations which are in close vicinity to other observations. Alternatively, the results of the Monte Carlo analysis can be post-processed to reflect the presence or absence of observations.

5.3 Applicability

For both the Jakarta case study, in which pluvial as well as fluvial flooding occurred, as for the York case study, in which mainly fluvial flooding occurred, the same method was found to give the best results. This indicates that this method can be successfully applied to both pluvial and fluvial floods. Coastal floods however were not reviewed in and should be further investigated before the applicability of the method in these floods can be determined.

Given the results of the uncertainty analysis, the map scale at which conclusions can be drawn, depends on both the errors in the input data as well as the topographical characteristics of the area, which determine to what extent errors propagate to the flood maps. The inner city of York for example was located lower than the surrounding area, and especially for this location flood maps at street scale could be produced. Errors in flood extent at the inner city were limited to 50 m. For the more flat downstream area of Jakarta however, uncertainties were considerably bigger. The errors in flood extent at this location estimated to be limited to 500 m. Although the method discussed in this research can also be used at an even coarser scale, for example to create a map of an entire country, other methods such as the one by Schnebele et al. (2014) will likely give comparable results, but require less computational time.

One of the potential advantages of using Twitter data to create flood maps is that these maps can be generated in real-time. For this to be possible however, the Twitter messages should reflect the time variation during the flood, there should be enough Twitter messages to be able to create maps over short time intervals and the methods to create the flood maps should require limited computational time. The results of multiple studies indicate that time variations during events are reflected in the number of Tweets during the event (e.g. Terpstra et al., 2012; Yin et al., 2014). The evolution of the spatial pattern over time has however only been reviewed at a coarse resolution, for example by Guan & Chen (2014). The analysis of time variation performed in this research indicated that indeed the variations in number of Tweets were similar to the variations in measured water levels. However, the datasets used contained too few observations to accurately review the time variation in the water levels derived from the Tweets and could not be used to create real-time flood maps. Nevertheless, many more observations can be gathered for the Jakarta case study, although the number used in this research was limited, because the Tweets needed to be analysed manually.

Also the flood mapping methods used in this research required only limited computational time. Flood maps were created in less than one minute for Jakarta and for the York case study 1000 Monte Carlo simulations were performed in about 10 minutes. The fact that observations can be added to the inverse distance weighting results, without having to recalculate using all observations in the area can help in reducing the computational time in real-time application even more.

The big drawback in real-time applicability is however the process of deriving locations and water depths from the Tweets. Although it is possible to do this manually in real-time, it might require more than one person if multiple observations are generated per minute. Therefore further research should be performed into automating this procedure. This can potentially be done using data from OpenStreetMap, since its underlying dataset can be accessed. This means that for example the line elements of streets can be used to pinpoint locations using a DTM. For real-time application it is essential that such a method can keep up with the speed at which Tweets are supplied.

OpenStreetMap datasets can also be used the other way around. Instead of searching for the locational reference in a Tweet on OpenStreetMap, it might be possible to use the most important OpenStreetMap features in an area (e.g. roads, neighbourhoods, POIs) to search for Tweets. For example Tweets which only mention a street name, but not specifically mention a city, can then be found also. This might lead to even higher numbers of Tweets being found, since the Twitter datasets in this research were restricted to Tweets that explicitly mentioned a city or neighbourhood.

Automating this process ensures that locations and water depths can be extracted from the messages in real-time. Although the creation of single flood maps can already be applied in real-time using the methods discussed in this report, the uncertainty maps that were created do not yet accurately reflect the real uncertainty in flood extent. Also methods of gathering more observations from Twitter or other data sources, such as including water level measurements, interacting with Twitter users and using an using a different method to search for Tweets with accurate locational references, should be analysed in order to have enough observations available to create maps at a high temporal resolution. If the uncertainty maps are successfully adapted however and enough observations can be obtained, the real-time flood maps and uncertainty maps created using Tweets have the potential of providing a wealth of information to for example rescue workers or other persons requiring flood information in real-time, where current methods such as hydraulic models and remote sensing are lacking.

6 Conclusions and recommendations

6.1 Conclusions

The below paragraphs shortly discuss the answers to the research questions from paragraph 1.3.

1. *How can current methods to create flood inundation maps from Twitter messages be improved upon?*
Interpolating 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.

2. How uncertain are the resulting flood extents?
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.

3. *In what context can the flood maps be applied?*
The 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.

6.2 Recommendations

The most important recommendation made based on the research is to review more case studies of flooding. Both reviewing multiple floods at the same location as well as reviewing floods at different locations are important to fully understand where the values of the parameters depend on. By reviewing multiple floods at the same location, it can be investigated whether the interpolation parameters remain equal between the cases, which is important in real-time application of the methods. Reviewing multiple floods at different locations can help to assess if these parameters can be estimated based on the topography of the study area.

Also possible improvements to the interpolation method and uncertainty analysis should be further reviewed. Examples of this are the inclusion of water level measurements in the interpolation procedure and reflecting the uncertainty of individual observations in the interpolation weights used to calculate the water levels. The method of producing uncertainty maps should be further optimized, since they did not accurately reflect the real uncertainty in flood extent. Reducing the overestimation of uncertainty caused by locational errors, and methods to consider the density of observations should be further investigated.

The real-time application of both the flood maps and uncertainty estimates should also be further reviewed. Therefore a large set of Tweets should be investigated, in order to find to which extent they relay time-variation in flood water levels and affected locations. Also for cases such as York, where only few observations were found, the effect of techniques to gather more observations, for example by using different search techniques or crowd interaction, should be analysed. A final recommendation with respect to the real-time application of the method is to investigate how the process of geocoding the locations from the Tweets can be automated.

The last recommendation of the research relates to the large uncertainties caused by locational errors. Although the uncertainty in flood extent caused by locational errors in the Tweets was likely overestimated, it is expected these errors are still an important source of uncertainty in flood extents, especially in more flat study areas. By minimizing the magnitude of these errors, the uncertainty in the flood maps can be considerably reduced. Different ways to achieve this should be reviewed should be reviewed, such as using different methods to transform the street and neighbourhood references in the Tweets to exact locations, requesting more specific locations using crowd interaction or using geo-tags in combination with street or neighbourhood references.

References

Aronica, G., Bates, P. D., & Horrit, M. S. (2002). Assessing the uncertainty in distributed model predictions using observed binary pattern information within GLUE. *Hydrological Processess, 16*, 2001-2016. doi:10.1002/hyp.398

Bates, P. D., & De Roo, A. P. (2000). A simple raster-based model for flood inundation simulation. *Journal of Hydrology, 236*, 24-77. doi:10.1016/S0022-1694(00)00278-X

BPBD DKI Jakarta. (n.d.). *Info Tinggi Muka Air*. Retrieved July 21, 2016, from BPBD Provinsi DKI Jakarta: http://bpbd.jakarta.go.id/waterlevel/

Budiyono, Y., Aerts, J., Brinkman, J., Marfai, M. A., & Ward, P. (2015). Flood risk assessment for delta mega-cities: a case study of Jakarta. *Natural Hazards, 75*, 389-413. doi:10.1007/s11069-014-1327-9

Davies, R. (2015, February 10). *Jakarta Floods Force 6,000 Evacuations*. Retrieved July 20, 2016, from FloodList: http://floodlist.com/asia/jakarta-floods-force-6000-evacuations

Dottori, F., & Todini, E. (2013). Testing a simple 2D hydraulic model in an urban flood experiment. *Hydrological Processes, 27*, 1301 - 1320. doi:10.1002/hyp.9370

Dullof, J., & Doucette, P. (2014). The Sequential Generation of Gaussian Random Fields for Applications in the Geospatial Sciences. *International Journal of Geo-Information, 3*, 817-852. doi:10.3390/ijgi3020817

EA. (n.d.). *Real time flood monitoring data: Archives*. Retrieved July 28, 2016, from environment.data.gov.uk: http://environment.data.gov.uk/flood-monitoring/archive

Eilander, D., Trambauer, P., Wagemaker, J., & van Loenen, A. (2016). Harvesting social media for generation of near real-time flood maps. *12th International Conference on Hydroinformatics, HIC 2016.* Elsevier. Retrieved from https://www.floodtags.com/wp-content/uploads/2015/09/Near-real-time-flood-maps_Eilander2016.pdf

Fohringer, J., Dransch, D., Kreibich, H., & Schröter, K. (2015). Social media as an information source for rapid flood inundation mapping. *Natural Hazards and Earth System Sciences, 3*, 4231-4264. doi:10.5194/nhess-15-2725-2015

Guan, X., & Chen, C. (2014). Using social media data to understand and assess distasters. *Natural Hazards, 74*(2), 837-850. doi:10.1007/s11069-014-1217-1

Heuvelink, G. B., Brown, J. D., & Van Loon, E. E. (2007). A probabilistic framework for representing and simulating uncertain environmental variables. *International Journal of Geographical Information Science, 21*(5), 497-513. doi:10.1080/13658810601063951

Hodgson, M. E., & Bresnahan, P. (2004). Accuracy of Airborne Lidar-Derived Elevation: Empirical Assessment and Error Budget. *Photogrammetric Engineering & Remote Sensing, 70*(3), 331-339. doi:10.14358/PERS.70.3.331

Horrit, M. S., Di Baldassarre, G., Bates, P. D., & Brath, A. (2007). Comparing the performance of a 2-D finite element and a 2D finite volume model of floodplain inundation using airborne SAR imagery. *Hydrological Processes, 21*, 2745-2759. doi:10.1002/hyp.6486

Kussul, N., Shelestov, A., & Skakun, S. (2008). Grid system for flood extent extraction from satellite images. *Earth Science Informatics, 1*(3), 105-117. doi:10.1007/s12145-008-0014-3

Leandro, J., Chen, A. S., Djordjevic, S., & Savic, D. A. (2009). Comparison of 1D/1D and 1D/2D Coupled (Sewer/Surface) Hydraulic Models for Urban Flood simulation. *Journal of Hydraulic Engineering, 135*(6), 495 - 504. doi:10.1061/(ASCE)HY.1943-7900.0000037

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

Li, S., MacMillan, R. A., Lobb, D. A., McConkey, B. G., Moulin, A., & Fraser, W. R. (2011). Lidar DEM error analyses and 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

Livne, E., & Svoray, T. (2011). Components of uncertainty in primary production model: the study of DEM, classification and location error. *International Journal of Geographical Information Science, 25*(3), 473-488. doi:10.1080/13658816.2010.517752

Mason, D. C., Speck, R., Devereux, B., Schumann, G. J., Neal, J. C., & Bates, P. D. (2010). Flood Detection in Urban Areas Using TerraSAR-X. *Transaction on geoscience and remote sensing, 48*(2), 882-894. doi:10.1109/TGRS.2009.2029236

Mudron, I., Podhoranyi, M., Cirbus, J., Devecka, B., & Bakay, L. (2013). 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

Neal, J. C., Bates, P. D., Fewtrell, T. J., Hunter, N. M., Wilson, M. D., & Horrit, M. S. (2009). Distributed whole city water level measurements from the Carlisle 2005 urban flood event and comparison with hydraulic model simulations. *Journal of Hydrology, 368*, 42-55. doi:10.1016/j.jhydrol.2009.01.026

Norbre, A. D., Cuartas, L. A., Momo, M. R., Severo, D. L., Pinheiro, A., & Norbre, C. A. (2015). HAND contour: a new proxy predictor of inundation extent. *Hydrological Processes*. doi:10.1002/hyp.10581

Ozdemir, H., Sampson, C. C., De Almeida, G. A., & Bates, P. D. (2013). Evaluating scale and roughness effects in urban flood modelling using terrestrial LIDAR data. *Hydrology and Earth System Sciences, 17*, 4015-4030. doi:10.5194/hess-17-4015-2013

Padawangi, R., & Douglass, M. (2015). Water, Water Everywhere: Toward Participatory Solutions to Chronic Urban Flooding in Jakarta. *Pacific Affairs, 88*(3), 517-550. doi:10.5509/2015883517

Palaseanu, M., & Pearlstine, L. (2008). Estimation of water surface elevations for the Everglades, Florida. *Computers & Geosciences, 34*, 815-826. doi:10.1016/j.cageo.2007.08.004

Parry, S., Barker, L., Prosdocimi, I., Lewis, M., Hannaford, J., & Clemas, S. (2016). *2016 Hydrological summary for the United Kingdom: December 2015.* Wallingford, UK: NERC/Centre for Ecology & Hydrology. Retrieved from http://nora.nerc.ac.uk/512654/1/HS_201512%20v2.pdf

Poulter, B., & Halpin, P. N. (2008). Raster modelling of coastal flooding from sea level rise. *International Journal of Geographical Information Science, 22*(2), 167-182. doi:10.1080/13658810701371858

Raaflaub, L. D., & Collins, M. J. (2006). The effect of error in gridded digital elevation models on the estimation of topographic parameters. *Environmental Modelling & Software, 21*, 710-732. doi:10.1016/j.envsoft.2005.02.003

Romanowicz, R., & Beven, K. (2003). Estimation of flood inundation probabilities as conditioned on event inundation maps. *Water Resources Research, 39*(3). doi:10.1029/2001WR001056

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

Schumann, G., Bates, P. D., Horrit, M. S., & Matgen, P. (2009). Progress in integration of remote sensing-derived flood extent and stage data and hydraulic models. *Review of Geophysics, 47*(4). doi:10.1029/2008RG000274

Smith, L., Liang, Q., James, P., & Lin, W. (2015). 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

Statista. (2016). *Twitter: number of montly active users 2016*. Retrieved May 5, 2016, from statista.com: http://www.statista.com/statistics/282087/number-of-monthly-active-twitter-users/

Stott, J. (2016, February 6). *After the floods, York's residents try to piece their lives back together*. Retrieved July 20, 2016, from The Guardian: https://www.theguardian.com/money/2016/feb/06/after-york-floods-residents-piece-lives-back-together

Sun, D., Li, S., Zheng, W., Croitoru, A., Stefanidis, A., & Goldberg, M. (2015). 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

Terpstra, T., De Vries, A., Stronkman, R., & Paradies, G. I. (2012). Towards a realtime Twitter analysis during crises for operational crisis management. *Proceedings of the 9th International ISCRAM Conference*, (pp. 1-9). Vancouver. Retrieved from http://www.iscramlive.org/ISCRAM2012/proceedings/172.pdf

Twele, A., Cao, W., Plank, S., & Martinis, S. (2016). Sentinel-1-based flood mapping: a fully automated processing chain. *Interational journal of remote sensing, 37*(13), 2990-3004. doi:10.1080/01431161.2016.1192304

UN. (2015). *The human cost of weather related distasters 1995 - 2015.* Geneva: United Nations. Retrieved from https://www.unisdr.org/we/inform/publications/46796

Vojinovic, Z., & Tutulic, D. (2009). On the use of 1D and coupled 1D-2D modelling approaches for assessment of flood damage in urban areas. *Urban Water Journal, 6*(3), 183-199. doi:10.1080/15730620802566877

Werner, M. G. (2001). Impact of grid size in GIS based flood extent mapping using a 1D flow model. *Physics and Chemistry of the Earth, Part B: Hydrology, Oceans and Atmosphere, 26*(7-8), 517-511. doi:10.1016/S1464-1909(01)00043-0

Yan, Y., Xiao, F., & Du, Y. (2014). Construction of lake bathymetry from MODIS satellite data and GIS from 2003 to 2011. *Chinese Journal of Oceanology and Limnology, 32*(3), 720-731. doi:10.1007/s00343-014-3185-4

Yin, J., Lampert, A., Cameron, M., Robinson, B., & Power, R. (2012). Using Social Media to Enhance Emergency Situation Awareness. *IEEE Intelligent Systems, 27*(6), 52-59. doi:10.1109/MIS.2012.6

Appendix A: detailed materials and methods

This appendix discusses the details of the datasets and processing steps discussed in chapter 2 of this report. The layout of the appendix is similar to the methodology chapter, starting with a discussion of the datasets which is followed by the discussion of methodology.

Materials

Elevation datasets

For both cases LIDAR datasets were used. The processing used to create the DTM used in flood mapping however differed among cases. Therefore the procedure used for each case is separately discussed.

Jakarta

A 2 m resolution LIDAR DSM covering the special capital city district of Jakarta was used as a starting point. This dataset was processed using a series of commands in SAGA GIS. The processing steps applied are listed below:

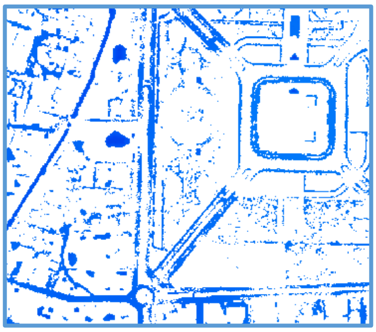

Figure 59: Ground elevation cells extracted from the DSM

1. The LIDAR DSM was resampled to both 20 m and 40 m resolution, to reduce the computational time necessary for each of the steps.
2. Of the 40 m DSM the slope was calculated. A Gaussian filter with a radius of 8 cells and standard deviation of 4 cells was applied to find the general areas with high slopes. Only continuous areas in which the filtered slope was higher than 0.02, consisting of at least 500 cells were kept. The result was resampled to 20 m resolution. This map was created since the upstream areas with the higher slopes had to be treated differently than the downstream areas.
3. For the 20 m DSM the percentile elevation value with respect to the other cells in a radius of 10 cells around it was calculated.
4. This map with percentiles was divided by 1 plus 5 times the value of the slope map. Thereby the values in upstream areas were further attenuated, and the high elevation values often present in the areas with higher slopes, were not considered to be outliers.
5. From the 20 m DSM every cell having a percentile value higher than 80 in the previous map was filtered, and the result was smoothed using a Gaussian filter with a standard deviation of 1 cell and a radius of two. This gave a smoothed surface representation.
6. This smoothed map was subtracted from the original 2 m resolution DSM, and cells having values below -0.15 were considered to represent ground elevation (figure 59). The elevation values of these cells were resampled to 20 m resolution.
7. From this 20 m resolution grid the percentile elevation values of each cell with respect to all cells in a radius of 5 cells were computed. Cells with values below 5 and above 95 were excluded.
8. The remaining cells were transformed to points, and used in IDW interpolation using a power of two and no smoothing.
9. To remove peaks the result is filtered using a Gaussian filter with a standard deviation of 2 and a radius of 4

The resulting DTM was used in the creation of the flood inundation maps and to create the HAND maps.

York

For the York case a 2 m DTM by the Environment agency was used as a start. This was clipped to the city of York with a buffer of 1 km. Afterwards the dataset was resampled to 20 m resolution. This large scale dataset was used to create the HAND. Both the HAND and DTM for the York case were clipped by a shapefile of the wards of the inner city of York, with a buffer of 500 m. The larger area DTM was used to create the HAND because HAND results at the border of the DTM are unreliable, since drainage channels do not extent to the border of the DTM.

Creation of HAND

The HAND was generated from the DTM using PCRaster. By computing the drainage directions of each cell of the DTM (with sinks removed), the accumulated flow (number of upstream cells flowing into a cell) could be derived. Since drainage channels have a considerably higher accumulated flow than cells outside of drainage channels, these could be identified by using a threshold. For the Jakarta case, setting a threshold of 10,000 cells (4 km^2) gave a good representation of the streams in the area. For the York case the streams were best represented by using a threshold of 15,000 cells (6 km^2). By using the subcatchment function of PCRaster, the elevation values of cells with an accumulated flow over 10,000 were distributed over the upstream area of each of the cells. By subtracting this from the original grid, the elevation values of the drainage channels were set to zero, and the remaining cells got an elevation value relative to the nearest drainage channel. Since depressions can be important in the mapping of floods, these were reintroduced, by subtracting the difference between the sink filled DTM and normal DTM from the result.

Validation datasets

The validation dataset for the city of Jakarta was produced by using flood related Tweets that contained photographs. These Tweets were kept strictly separated from the input dataset used to create the flood maps. For each of the photographs an exact location was derived, by using the locational reference already in the Tweet, searching Google Streetview for this location, and seeing if an exact match with the picture could be made. This way the location of each photograph was determined to within a few metres accuracy. Also the water depth was determined from the photograph, by comparing the flooded situation with the not flooded picture on Google Streetview. Using this process a dataset was created of points at which flooding certainly occurred, including the estimated water depths at these locations.

The validation dataset for the city of York was constructed solely using data from the Environment Agency (EA). The most important dataset used was a (draft version of a) dataset with recorded fluvial flood outlines for the actual floods in December 2015. Since this dataset only included the fluvial flood extent, and not the areas that were flooded separately from the river, an additional dataset of historic flood events was additionally used. This dataset of historic flood events was used to determine the flood extents at 3 places which were flooded further away from the river. Historic flood events were added in these places, since photographs were available actually indicating these locations were flooded. The dataset of fluvial flood extents, extended with the historic flood extents at three locations is given in figure 60.

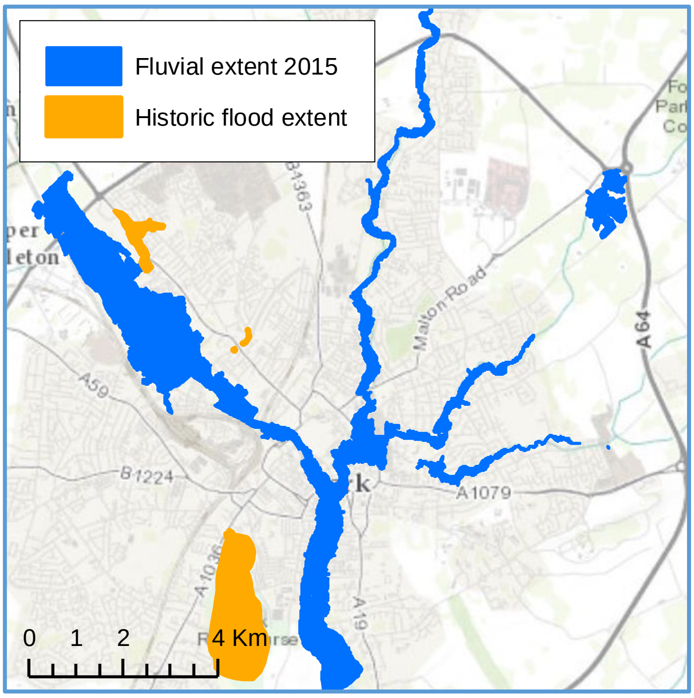

Figure 60: York - Combined validation dataset

Twitter datasets

The datasets with all the flood related Twitter messages over the period of interest for each of the case studies were downloaded using the FloodTags API. To come to the final dataset used to create the actual flood maps, several filtering steps were performed. These are discussed below for each case separately.

Jakarta

The time period to review for the Jakarta case was selected by looking at the news reports of flooding in the area and a chart of river water levels during the flood event (figure 61).

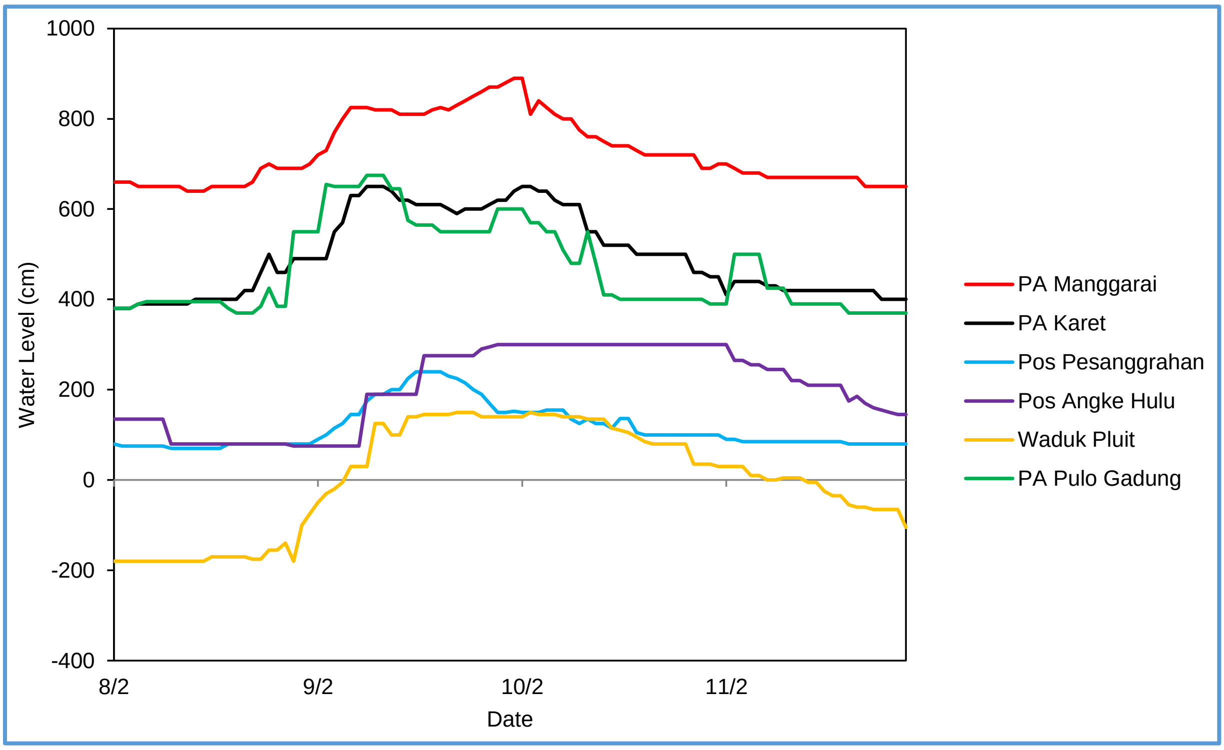

Figure 61:Jakarta - Water levels at several measurement stations during the 2015 floods (BPBD DKI Jakarta, s.d.)

The water levels started to rise on the 8th of February, and all declined on the 11th (UTC). For this time period about 135,000 flood related Tweets were supplied by FloodTags. The following processing steps were applied to the dataset to come to a dataset that could be used in flood mapping:

1. Duplicates or closely matching Tweets were removed from the dataset, by removing the links from the messages, and looking for exact matches in the dataset.
2. Only Tweets from which FloodTags already derived a location within Jakarta, were included. This location, currently added to Indonesian Tweets by FloodTags, at best refers to neighbourhoods.
3. This dataset was further filtered for the presence of locational information. Therefore only Tweets remained in the dataset, which contained certain keywords, such as references to streets, neighbourhoods or POIs. To find these words, sentences were not only split up by using spaces, but also by points, commas, semicolons or colons. Additionally Tweets were included which:
4. Contained certain combinations of letters, in this case ‘cm’, related to water depth, and ‘kel’ related to a reference to a neighbourhood.
5. From the dataset created using the steps above, tweets were excluded which contained a certain combination of letters and symbols, for example related to re-tweets, or questions about flooding.
6. As a last step only messages were included of which FloodTags determined they belonged to the class ‘flood’. This class is assigned to each of the messages by a process called natural language processing, and indicates the topic of a message. Since relevant messages were also found in other classes, such as ‘mixed’ or ‘news’, this step was mainly performed to reduce the size of the dataset and create a manageable dataset to be used in the manual process of assigning locations and water depths to each of the messages.

The specifics of each step are given in table 3. By applying these filters a dataset of 419 Tweets was left. After manual filtering of the dataset, excluding messages from which no accurate location could be derived, a dataset of 219 Tweets was left.

Table 3: Specifics of filtering steps used to construct the Jakarta dataset.

Filter	Specifics
2. Only including messages pinpointed to this geographical extent	North: -6.02, East: 107.2,South: -6.7,West: 106.45
3. Only including Tweets with these complete words or…	'jl.', 'jl', 'jalan', 'universitas', 'pos polisi', 'persimpangan', 'dekat', 'gedung', 'persegi', 'toko', 'depan', 'stasiun', 'pasar', 'ITC', 'kampung', 'perum'
4. … contain these combinations of letters	‘cm’,’kel’
5. Exclude Tweets containing these sequences of letters/symbols	'via @', 'RT @', '?'
6. Only include Tweets belonging to this class	‘flood’

For the Jakarta case also a dataset was constructed for validation purposes. This was constructed by using a combination of datasets, which used a variety of filters. First of all the filtering steps in Table 3 were used, without the filtering for class, and by including only Tweets with photos attached to them, or containing links to photos. A second dataset was constructed by looking for Tweets that mentioned both a water depth and had a photograph or link to a photograph attached to it. These two datasets were combined, and Tweets already in the input dataset were removed from them. This yielded a total of 571 Tweets with photographs. Only from 75 of these an exact location could be derived.

York

The time period over which York was flooded was determined by looking at news reports of the floods and the water levels on the River Ouse (figure 62). To capture the entire flood peak, Tweets from the 25th until the 30th of December were used, yielding a database of about 900,000 Tweets. To this dataset, the following filtering steps were applied:

1. Only Tweets that mentioned ‘York’ or ‘YorkFloods’ were included
2. Of this selection, Tweets that mentioned either ‘New York’ or ‘York County’ were excluded, since these locations are in the US.
3. Excluding re-tweets by removing links from messages, and searching for exact matches.

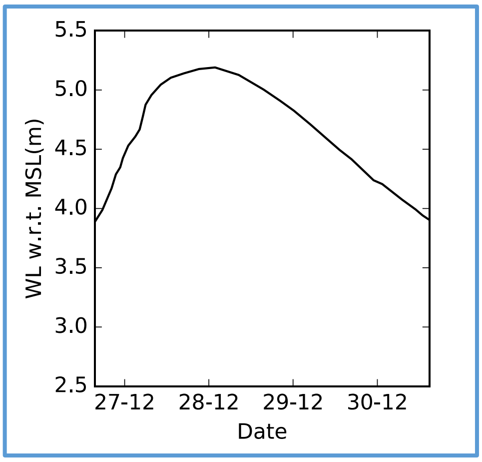

Figure 62: York - Water level on the river Ouse during the 2015 floods. Source: (EA, s.d.)

Using these steps an initial database was created, consisting of about 38,000 Tweets. This initial database was then further filtered by:

1. Including only Tweets which included locational references
2. Excluding tweets which did not refer to flooded locations directly, such as tweets referring to flood barriers, locations where sandbags were placed, the looting of flooded houses and questions about flooding.

The details of these steps are given in table 4.

Table 4: Detailed filtering steps to construct the York Twitter dataset

Filter	Specifics
1. Including Tweets mentioning specific words	'Ln', 'Rd', 'St', 'Lane', 'Road', 'Street', 'Hospital', 'Museum', 'golf', 'ave', 'Avenue', 'Boulevard', 'School', 'library'
2. Excluding Tweets containing the combination of letters/symbols	'flood barrier', 'flood gates', 'sandbag', 'good', 'view', 'robbing', 'well', 'news', '?', 'the road', 'road closures', 'looting', ' OK ', 'flooded street', 'clean', 'Flood warning', 'rescue', 'help', 'fine', 'open', 'flooded road', 'roads', 'major road', 'my road', 'police', 'main road', 'near york', 'some roads', 'a street', 'York Street', 'Are there', 'Flood Alert'

Using these steps yielded a database of 266 Tweets. Of 87 of these an accurate location could be derived.

Manual optimization

Currently FloodTags only adds low resolution positional information (up to neighbourhood level) to Indonesian Tweets, and no positional information to English Tweets. Therefore manual geocoding of the messages was performed. This was done by searching for parts of Tweets using Google Maps, a process that could potentially be automated in the future. First of all the relevant parts of sentences, that contain location information were identified. These are often places after words like 'in' or 'at'. In some cases these can also be identified by looking at the occurrence of the word 'street' in a message, which might indicate the presences of address information. Other words like 'near' of 'in front of' also indicate the presence of locational information. The combination of words after these keywords, were further optimized, before using them in a search engine. For Jakarta for example if a number is given after a street name which often refers to the name of the street itself and not an address on the street. If no exact match can be found for the street name (including the number), the number was transformed to Roman numerals, which are often used in street names in Jakarta. If that didn't yield any exact matches, these numbers were transformed to words, for example from 2 to 'dua', which are also found in street names.

Some messages also contained multiple locational references. These can be separated either by commas or the words 'and' or 'in front of'. For example: "Flooding on main street in front of the supermarket". In these cases two locations were derived from the messages. In case one of the locations was a street and the other one was a POI or address, the location on the street nearest to the POI or address was used. In case more than two locations were mentioned, either POI's or Streets, they were treated separately, and considered a separate observation. Only if the two streets mentioned in the message intersected, the location of the intersection was used.

The combination of words which resembled the location was used as a search query on Google Maps, and the first search result was selected as being the best. If in case of searching for a street name, multiple matching streets were found, the most important road (e.g. main road or highway) was selected. For the Jakarta Case in some cases it helped to add the word Tol after Jalan (street) in order to ensure a highway was found.

When using an entire road however, a precise location of flooding is lacking. To add this precise location in latitude/longitude (WGS84) coordinates, the DTM and information about depressions therein, were used. These were imported in Google Earth Engine. If depressions were present on a road, the deepest depression on the road was assumed to be the location of the report. In case no depressions were found on the road in question, simply the lowest location on the road was used as the location of flooding.

A similar analysis was performed in case only reference was made to a certain neighbourhood. In this case the deepest depression was selected, if this was not at open water, or very near to open water. In case no depression was in the neighbourhood, again the lowest point was taken as the location of flooding.

Besides locations, also water depths were extracted from the messages. These water depths were found by searching for the presence of indicators of water depth, such as 'cm', 'm' or 'metre' after a number. If these letters were present, the numbers in front of them were considered to be the water depth. In many cases actually a range of water depths was given. This was identified by searching for the separator '-' or the word 'to' in between numbers. The low and high values of these ranges were stored separately, since they can be relevant in determining the uncertainty in water depth specified. If multiple locations were found in a Twitter message in the geocoding process and these were not a combination of a POI and a street, multiple water depths were extracted from the Twitter message. The water depth was linked to a specific location by either looking as the water depths that were in the same part of the message as the location (so in between the same commas), or that were closest to the locational reference.

Methods

The details of both the flood inundation mapping process and the uncertainty assessment are discussed in the paragraphs below.

Flood mapping

The different flood mapping methods discussed in paragraph 2.3.2 were implemented in Python scripts. For the calculation of drainage directions of the data and spreading water level values over the subcatchments of flow paths, the PCRaster module for Python 2.7 was used. Some of the specifics of the implementation of the mapping methods are discussed below.

Grouping of observations

The grouping of observations was performed by either clustering observations, or by reviewing which observations had the same downstream flow path.

The clustering of observations was performed by starting at the first observation in the dataset, and looking in a pre-specified radius around this observation, to see whether any observations were close. If this was the case, these observations were added to the area, and for these observations it was also reviewed whether they had any observations in a specified radius around them. This was repeated until no more additional observations were found, and the resulting set of observations was considered to belong to the same continuously flooded area. After this a new observation was taken from the database, which not yet belonged to an area, and the process was repeated. This way all observations were divided into groups.

The second grouping method, based on the intersection of downstream flow paths was executed by first calculating the local drainage directions of the elevation model using PCRaster. After this was done, for each observation in the dataset, the downstream flow path was traced. This way a raster was created of which all the downstream flow paths of observations had a value of 1. Using the 'clump' function of PCRaster, all connected areas were given a unique number. These numbers where then assigned as group numbers to the observations within the areas. To account for sinks, the map with local drainage directions was generated using the 20 m resolution DTM, and only filtering out a limited number of sinks, which were believed to be generated by errors in the DTM, rather than actual features in the area. Sinks that had an area lower than a certain pre-specified threshold, were filtered. Setting this threshold to 800 and 1200 for the Jakarta and York case studies respectively, gave the best representations of the drainage networks and sinks in the area. By leaving these sinks in the dataset to a certain extent, all the observations draining to the same sink were grouped.

Interpolation procedure

Water levels of the observations were interpolated using Inverse distance weighting (Equation [1] & [2]). The implementation of the interpolation along flow paths required a slightly adapted version of this equation. For this method, the distances of each cell the observation along the flow path was calculated using the 'spread' function of PCRaster. This way the distance to each cell (in number of cells) was calculated. Therefore the smoothing parameter could not be added to the squares of the X/Y distance. Due to this the following slightly adapted version of equation [2] was used:

$$W = \frac{1}{(d+s)^p} \qquad [4]$$

With:

W: Weight used in inverse distance weighting
d: Distance calculated using PCRaster 'spread' function (cells)
p: Power (-)
s: Smoothing factor

Using equation [4] in combination with equation [1], water levels on cells belonging to the flow paths were determined. This yielded a grid in which only the flow paths had water level values. Using the PCRaster 'subcatchment' function, these water levels on the flow paths were also given to their upstream cells. This way the water levels for all areas connected to flow paths were calculated. From these water levels then the ground level was subtracted, to derive the water depth. Flooded areas not directly connected to the flow paths were removed using the flood fill procedure discussed below.

Constraining flood extents

For all methods flooded areas that were not directly connected to either the downstream flow path of an observation, or the observation itself, were removed from the maps. This was done by using a flood fill algorithm, and seeding from the cells belonging to the flow path or the cells containing observations respectively. Since relatively coarse resolution elevation data was used in the research, a four-side connectivity rule was used, since an eight side connectivity rule can lead to an overestimation of flood extents (Poulter & Halpin, 2008).

Uncertainty assessment

The uncertainty assessment was performed by first generating input datasets with random errors added to them, and consequently using these datasets to create binary flood maps, meaning they indicated flooding by giving cells a value of 0 or 1. After all flood maps were created, the number of runs a cells was flooded, was divided by the number of runs to calculate the percentage of simulations a cell was flooded. These steps were implemented in Python 2.7. The details of this implementation are discussed in the paragraphs below.

Generation of datasets with random errors

Errors were added to the Twitter datasets by drawing values from a normal distribution with a mean value and standard deviation based on the errors derived in the analysis of the Twitter datasets. For each simulation a dataset of observations with random errors was generated.

The generation of Random Grids was performed using the methodology described by Dullof & Doucette (2014), who also discuss the implementation of the numerical scheme in detail. The following sets of equations (Dullof & Doucette, 2014) were used to implement this method:

$$z(k+1,l+1) = r * z(k+1,l) + s * z(k,l+1) - r * s * z(k,l) + N(0,\sigma_u) \quad [5]$$

$$s = e^{-\delta_y/T_y} \quad [6]$$

$$r = e^{-\delta_x/T_x} \quad [7]$$

$$\sigma_u^2 = (1-s^2)(1-r^2)\sigma_z^2 \quad [8]$$

With:

$z(k,l)$: Error at row k, column l (m)
$\delta_{x/y}$: Grid resolution in x/y direction (m)
$T_{x/y}$: Spatial correlation distance in x/y direction (m)
$N(0,\sigma_u)$: Error drawn from random distribution with mean of zero, and standard deviation σ_u
σ_z: Standard deviation of errors in grid (m)

To compute the cell first row and column equation [9] was used. For computing the remainder of the first row and column equations [10] and [11] were used respectively (Dullof & Doucette, 2014).

$$z(1,1) = N(0,\sigma_z) \quad [9]$$

$$z(1,l+1) = r * z(1,l) + N(0,\sigma_u) \quad [10]$$

$$z(k+1,1) = s * z(k,1) + N(0,\sigma_u) \quad [11]$$

With:

$z(k,l)$: Error at row k, column l (m)
$N(0,\sigma_z)$: Error drawn from random distribution with mean of zero, and standard deviation σ_z
$N(0,\sigma_u)$: Error drawn from random distribution with mean of zero, and standard deviation σ_u
σ_z: Standard deviation of errors in grid (m)
σ_u: (Equation [8])
r: (Equation [7])
s: (Equation [6])

Appendix B: detailed results

The detailed results of both cases are discussed in this appendix.

Jakarta

Several aspects of the results for the Jakarta case are discussed in more detail in the paragraphs below.

Dataset characteristics and uncertainties

Spatial pattern of observations

As an extension to figure 20, a map was created of all observations, including the ones without water depths, to see if this would lead to new insights. The result is given in figure 63.

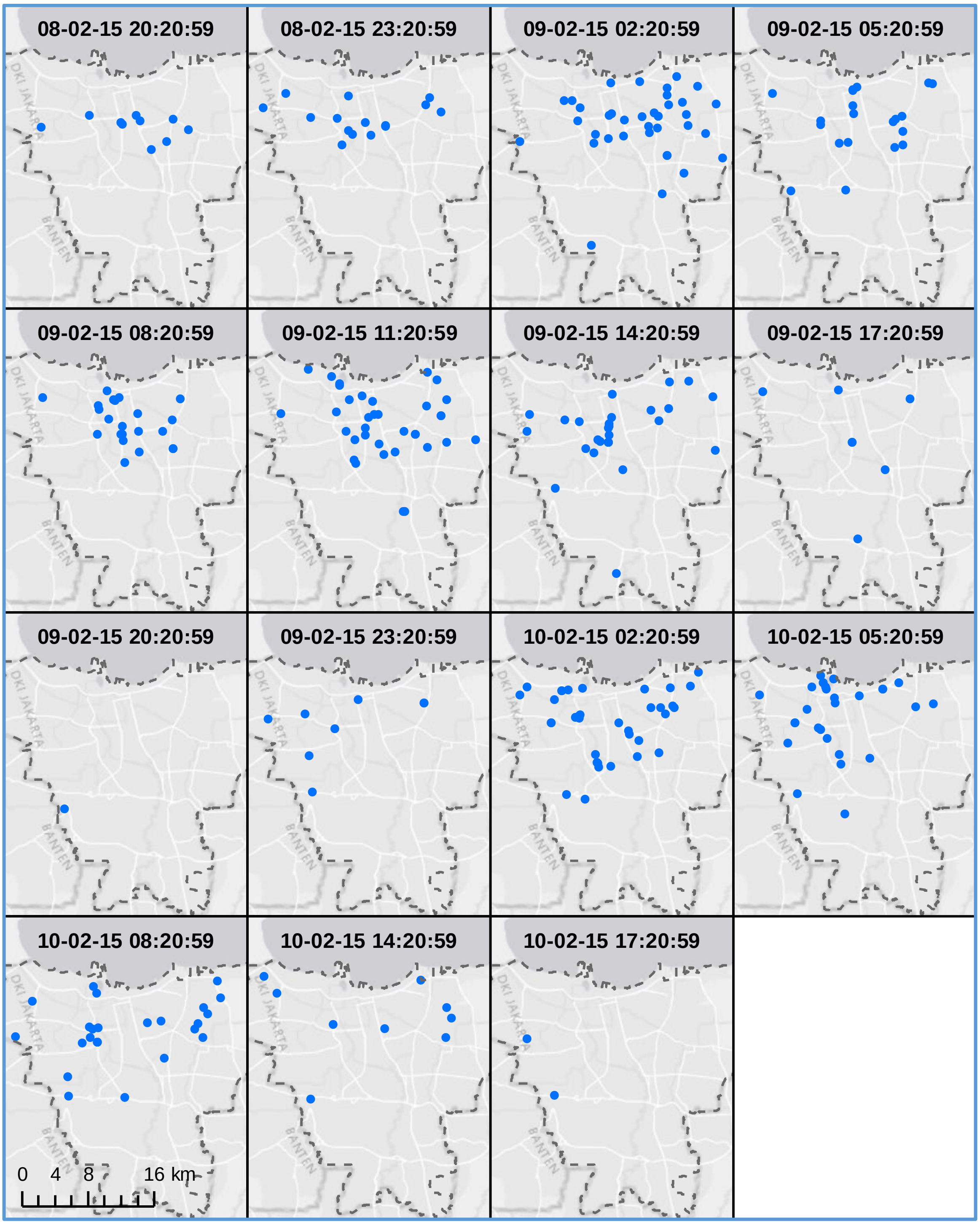

Figure 63: York - Post times of observations, recorded over 5 hour intervals (ahead of the time specified at the individual maps), using all observations with a location reference.

Just like the map in figure 20, clear peaks are visible on the ninth and tenth of February. A clear pattern or clustering of observations within each time step however, seems to be absent. Therefore using all observations with locational references does not lead to any additional insights with respect to the map in figure 20 of this report.

Differences in X/Y error variance

Judging from figures 22 to 24 it is likely significant differences in locational error variance exist between Tweets that mention POIs, and Tweets referring to Streets or Neighbourhoods. Therefore F-tests were applied, to review the differences between Tweets having the different locational references. POI and Street references as well as POI and Neighbourhood references were reviewed. The variance of each subset is given in table 5. The F-Values are:

Table 5: Jakarta, variances in locational errors

Type of references	Variance (m^2)	n
POIs	56,822	82
Streets	441,412	72
Neighbourhoods	496,121	6

$$F_{streets>POIs} = \frac{s_1^2}{s_2^2} = \frac{441{,}412}{56{,}822} = 7.77$$

$$F_{neighbourhoods>POIs} = \frac{496{,}121}{56{,}822} = 8.73$$

Using these results and a 95% confidence interval (one-sided) it is confirmed that both the locational errors of Tweets referring to streets as well as the locational errors of Tweets referring to neighbourhoods have a larger variance than that of the locational errors of Tweets referring to POIs.

Mean deviation of water depth

To test whether water depths significantly overestimated by the Twitter messages, a t-test was used to test whether the mean deviation was significantly different from zero. Therefore a log transformation was applied to the dataset with water depths. The result is given in figure 64. Doing a z-test using this set of transformed water depths only yielded a z-score of 0.61, which is considerably lower than the critical value of 1.64 (One tailed, 95%). Therefore it cannot be proven there is a mean overestimation of water depth.

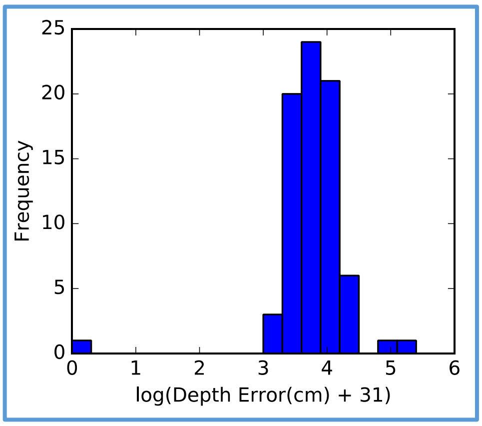

Figure 64: Log transformed water depths (original in Figure 25)

Flood mapping

For each of the flood mapping methods discussed in this report, parameters were varied to see how the quality of the map was affected. This paragraph first discusses some of the constants used in grouping of the observations. This is followed by a paragraph which discusses the results of the variation of other mapping parameters.

Mapping constants

Several parameters were kept constant for all flood maps. These were the search radius parameter used in the grouping of observations based on vicinity and the number of downstream cells to consider as well as the area threshold to filter sinks in the grouping based on downstream flow paths.

For the grouping based on vicinity, a search radius of 34 cells (680 m) was found to give the most accurate representation of the groups in the area. For the grouping of downstream cells, the complete downstream flow paths of observations were traced. This, in combination with filtering sinks with an area below 800 cells (0.32 km^2), yielded the best separation of groups. The results of using the parameters on the grouping of observations are given in figure 65.

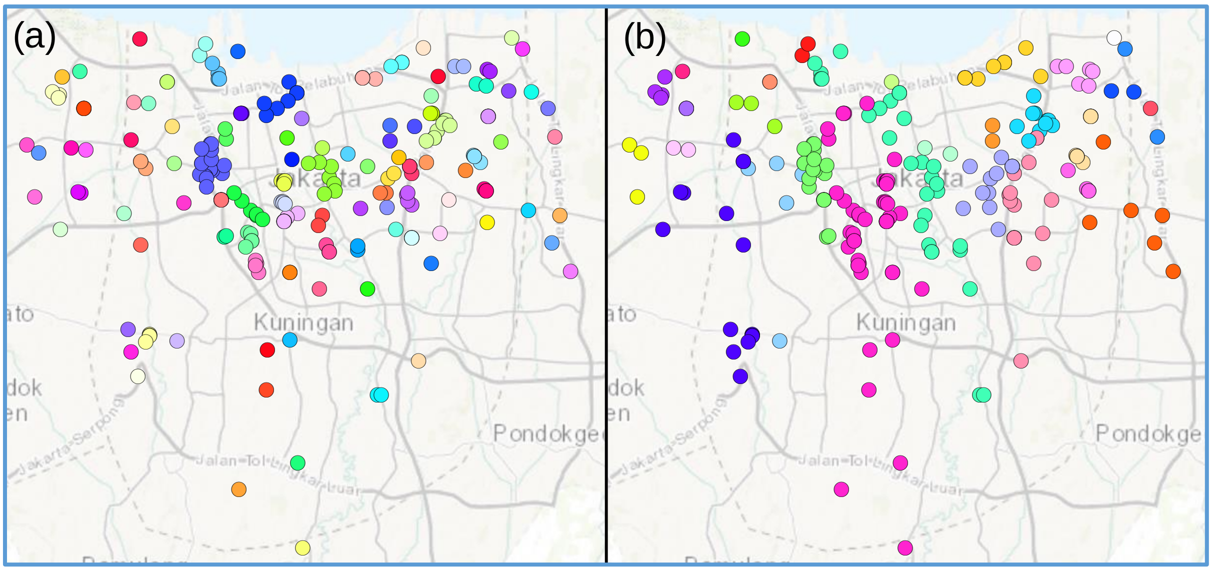

Figure 65: Jakarta - Maps of applying both grouping procedures, with colours indicating unique areas. (a) Grouping based on vicinity and (b) grouping based on intersection of downstream flow paths.

From this figure it can be seen that the grouping of observations by flow paths gives more credible groups, especially for observations close to drainage channels. Since not all sinks are filtered from the DTM, the groups that drain to large sinks are also correctly grouped by the procedure.

For the identification of observations that lay close to the original flood extent mapping using only observations with a water depth, a value of 200 m was used. This meant that observations situated within 200 m from this original flood extent were given a default water depth. Also in case clusters of observations were used in combination with a default water depth a value of 200 m was used to cluster the observations.

Plain interpolation results

To find the optimal set of parameters, at which the mapping methods performed best, a number of parameters were varied. Also the method of including observations without water depths was varied. The Mean error (ME) in water depth, Root mean square error (RMSE) in water depth as well as the percentage of validation points correctly mapped flooded and the amount of these points per km^2 of flooded area were calculated. The combination of these parameters was used in evaluation of the result, in which the amount of points per km^2 of flooded area was of most importance. The results for applying the plain interpolation procedure to the Jakarta case are given in table 6.

Table 6: Jakarta - Results of applying plain interpolation

Power	Smoothing (m)	Default water depth (cm)	Observations with no WD	Search Radii (m)	Percentage correct	ME (m)	RMSE (m)	Flooded Area (Km^2)	Points/area (Km^{-2})
1	0				76	+0.22	0.46	215.70	0.26
2	0				76	+0.14	0.64	201.14	0.28
3	0				75	+0.17	0.87	197.63	0.28
4	0				73	+0.20	1.02	200.15	0.27
5	0				72	+0.22	1.09	201.92	0.27
2	100				76	+0.14	0.62	201.68	0.28
2	300				75	+0.17	0.62	201.22	0.28
2	500				75	+0.19	0.67	202.08	0.28
2	750				75	+0.21	0.70	203.25	0.28
2	1000				75	+0.22	0.69	204.40	0.27
4	500				76	+0.21	0.92	199.94	0.29
1	100				75	+0.23	0.48	215.65	0.26
2	100	1	DWD		72	+0.04	0.43	190.66	0.28
2	100	5	DWD		75	+0.05	0.43	194.09	0.29
2	100	10	DWD		77	+0.08	0.43	203.28	0.29
2	100	20	DWD		80	+0.11	0.45	214.73	0.28
2	100	30	DWD		80	+0.14	0.47	221.23	0.27
2	100	40	DWD		83	+0.17	0.49	226.97	0.27
2	100	50	DWD		85	+0.21	0.51	232.42	0.28
2	100	10	Nearby	200	81	+0.19	0.67	220.09	0.28
2	100	20	Nearby	200	81	+0.19	0.67	221.69	0.28
2	100	30	Nearby	200	81	+0.20	0.68	223.23	0.27
2	100	5	Nearby + Clusters	200 / 200	83	+0.24	0.70	231.70	0.27
2	100	10	Nearby + Clusters	200 / 200	83	+0.24	0.70	232.65	0.27
2	100	20	Nearby + Clusters	200/200	83	+0.25	0.81	234.56	0.26
2	100	20	Nearby	10000	85	+0.29	0.75	249.00	0.26

Most optimal results were found by using a default water depth (DWD) of 10 cm for observations lacking one, in combination with a power parameter of 2 and a smoothing of 100 m. Other combinations were tested, for example by only using observations close to the flood extent mapped using observations with water depths (Nearby), or the use of clusters of observations. For both a search radius of 200 m (10 cells) was used. Using only additional observations that were in clusters is not in the table above, since only a limited number of clusters were identified. Also using all observations outside the original flood extent was tried (Nearby in combination with a radius of 10.000). This did not yield better results.

Grouped interpolation results

For the method that groups observations prior to interpolating also a variety of parameters was varied. The extent over which water levels were interpolated in addition to the highest/lowest x/y values was fixed at 2.5 km, since this setting ensured most of the downstream area of Jakarta got water level values assigned. As grouping methods both the grouping based on vicinity as well as the grouping based on intersection of downstream flow paths were evaluated. As the results in Figure 65 already indicate, grouping using downstream flow paths gave considerably better results. Using this grouping method, several parameters of the mapping method were varied. The results are given in table 7.

Table 7: Jakarta - Results of applying grouped interpolation

Power	Smoothing (m)	Default water depth (cm)	Observations with no water depth	Search radii (m)	Percentage correct	ME (m)	RMSE in water depth (m)	Flooded Area (Km^2)	Points/area (Km^{-2})
1	0				79	+0.24	0.69	189.42	0.31
2	0				73	+0.17	0.82	181.32	0.30
3	0				75	+0.20	0.98	177.65	0.32
4	0				73	+0.22	1.07	177.05	0.31
1	100				77	+0.27	0.72	189.14	0.31
2	100				77	+0.18	0.80	184.33	0.31
2	200				77	+0.20	0.79	184.09	0.32
2	300				75	+0.21	0.81	184.16	0.30
2	100	5	DWD		72	+0.07	0.54	185.71	0.29
2	100	10	DWD		73	+0.08	0.54	192.16	0.29
2	100	20	DWD		77	+0.12	0.55	205.95	0.28
2	100	5	Nearby	200	80	+0.20	0.83	197.49	0.30
2	100	10	Nearby	200	80	+0.21	0.83	198.05	0.30
2	100	20	Nearby	200	83	+0.21	0.83	199.19	0.31
2	100	30	Nearby	200	84	+0.22	0.84	200.18	0.31
2	100	40	Nearby	200	84	+0.23	0.84	201.22	0.31
2	100	50	Nearby	200	84	+0.23	0.84	202.16	0.31
2	100	60	Nearby	200	85	+0.24	0.85	203.06	0.32
2	100	70	Nearby	200	87	+0.25	0.85	204.08	0.32
2	100	5	Clusters	200	79	+0.21	0.82	190.67	0.31
2	100	10	Clusters	200	79	+0.21	0.82	190.73	0.31
2	100	20	Clusters	200	79	+0.21	0.83	190.90	0.31
2	100	30	Clusters	200	79	+0.21	0.83	191.06	0.31
2	100	50	Nearby	10000	89	+0.29	0.87	229.39	0.29
2	100	30	Nearby	10000	87	+0.27	0.86	225.61	0.29
2	100	20	Nearby	10000	85	+0.26	0.86	220.63	0.29
2	100	50	Nearby + Clusters	200 / 200	88	+0.27	0.86	208.53	0.32
2	100	40	Nearby + Clusters	200 / 200	87	+0.26	0.86	207.47	0.31
2	100	30	Nearby + Clusters	200 / 200	87	+0.25	0.86	205.95	0.32

Using this method, assigning observations without water depths a default water depth only when they lay close to the originally mapped flood extent or were part of a cluster, actually gave better results than purely assigning a default water depth to all observations without a water depth. Also using all observations outside the flood extent mapped using observations with a water depth did not give better results.

Interpolation along flow path results
For the interpolation along flow paths also the model parameters were varied to find an optimal combination. The results are given in table 8.

Table 8: Jakarta - results of applying interpolation along flow paths

Power	Smoothing (cells)	Default water depth (cm)	Observations with no water depth	Search radii (m)	Percentage correct	ME (m)	RMSE in water depth (m)	Flooded Area (Km^2)	Points/area (Km^{-2})
1	0				67	+0.07	0.56	162.93	0.31
2	0				63	+0.02	0.51	165.73	0.28
3	0				60	+0.01	0.50	168.05	0.27
4	0				60	+0.01	0.50	170.24	0.26
5	0				60	+0.00	0.51	172.02	0.26
1	10				64	+0.10	0.59	164.04	0.29
1	20				63	+0.13	0.61	164.85	0.29
1	30				64	+0.15	0.63	165.40	0.29
2	5				64	+0.02	0.52	165.40	0.29
2	10				65	+0.02	0.52	165.18	0.30
2	20				67	+0.03	0.53	165.06	0.30
2	30				67	+0.05	0.54	164.98	0.30
2	40				65	+0.06	0.56	164.91	0.30
3	20				60	+0.01	0.52	167.25	0.27
3	40				65	+0.03	0.53	166.58	0.29
3	60				64	+0.04	0.55	166.27	0.29
1	0	5	DWD		63	+0.00	0.43	151.17	0.31
1	0	10	DWD		64	+0.02	0.43	154.74	0.31
1	0	20	DWD		68	+0.05	0.45	161.31	0.32
1	0	30	DWD		71	+0.08	0.46	167.75	0.32
1	0	40	DWD		75	+0.11	0.48	173.49	0.32
1	0	50	DWD		75	+0.14	0.51	178.45	0.32
2	10	5	DWD		67	-0.07	0.42	153.95	0.32
2	10	10	DWD		69	-0.05	0.42	157.57	0.33
2	10	20	DWD		72	-0.02	0.43	164.35	0.33
2	10	30	DWD		75	+0.01	0.44	171.17	0.33
2	10	40	DWD		76	+0.04	0.46	177.31	0.32
2	10	20	'Nearby'	200	69	+0.08	0.60	179.56	0.29
2	10	10	'Nearby'	200	69	+0.07	0.60	177.90	0.29
2	10	30	'Nearby'	200	69	+0.09	0.61	180.98	0.29
2	10	40	'Nearby'	200	71	+0.10	0.62	182.57	0.29
1	0	20	'Nearby'	200	69	+0.13	0.59	178.61	0.29
2	10	10	'Nearby'	10000	76	+0.16	0.70	197.89	0.29
2	10	20	'Nearby'	10000	73	+0.15	0.69	195.83	0.28
2	10	20	'Clusters'	200	65	+0.08	0.59	172.32	0.28
2	10	30	Clusters	200	65	+0.08	0.59	172.62	0.28
2	10	20	'Clusters + Nearby'	200/200	69	+0.12	0.65	185.99	0.28

Although the number of points per km^2 of flooded area was only slightly higher for this method, the deviations in water depth were much lower. The best result was obtained by giving observations without water depths a default water depth of 30 cm, using a power parameter of 2 and a smoothing of 10 cells (200 m).

York

Several aspects of the York case study are discussed in more detail below

Dataset characteristics

Differences in X/Y error variance

A F-test was applied to review the differences in variance between the locational errors of Tweets referring to POIs and Tweets referring to streets. The variances of both subsets are given in Table 9 . Using these numbers the F-value was calculated:

Table 9: York - variances in locational error

Type of references	Variance (m^2)	n
POIs	3,134	8
Streets	84,609	42

$$F = \frac{84{,}609}{3{,}134} = 27.0$$

Using this result, in combination with a 95% confidence interval indicates that the variance in locational errors of Tweets mentioning streets is significantly larger than that of Tweets mentioning POIs.

Also, using the data in Table 5 and Table 9 it was reviewed whether the variance in locational errors of Tweets referring to streets of the Jakarta case study was significantly higher than for of the York case study. The F-statistic was therefore calculated:

$$F = \frac{441{,}412}{84{,}609} = 5.22$$

Using a 95% confidence interval it is confirmed both values are indeed significantly different.

Flood mapping

For each mapping methods, several combinations of parameters were tested. The effects of using different sets of parameters are discussed below. This paragraph starts with a discussion of the results of both grouping methods. This is followed by all the results obtained using different parameters for the interpolation procedure

Mapping constants

Both grouping based on vicinity as well as grouping based on downstream flow paths were tested for the York case. Grouping based on vicinity was found to give the best results using a radius of 52 cells (1040). The grouping based on downstream flow paths was found to perform best by filtering sinks smaller than 1200 cells (0.48 km^2). The results of using the different grouping methods are given in Figure 66.

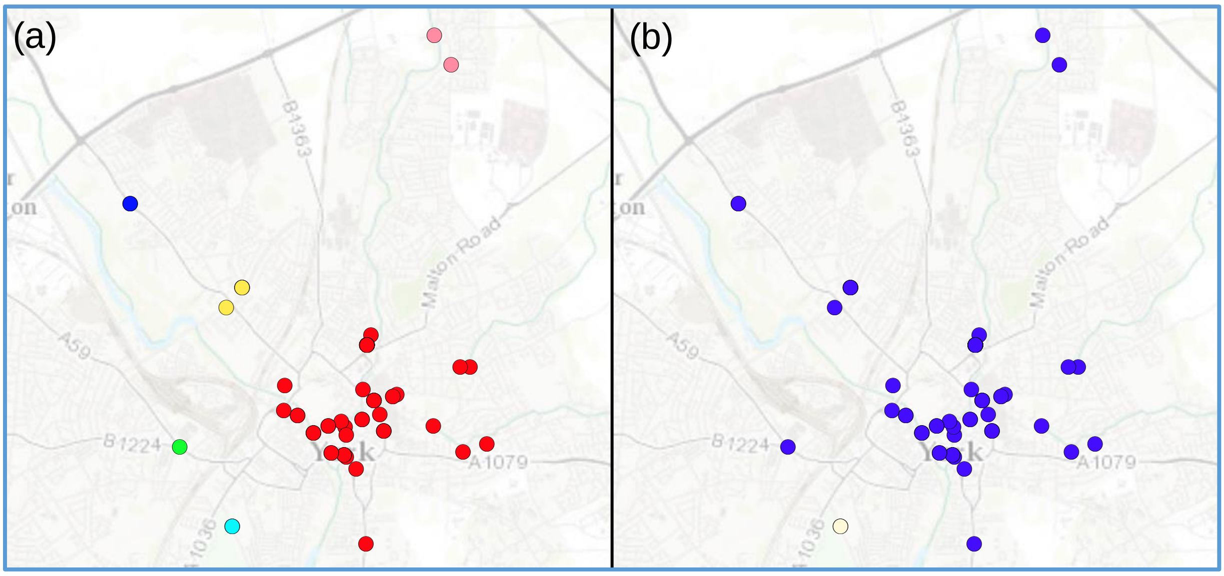

Figure 66: Maps of applying both grouping procedures in York, with colours indicating unique areas. (a) Grouping based on vicinity and (b) grouping based on intersection of downstream flow paths.

From this picture it is obvious none of the grouping methods is perfect. In reality only the green and light blue groups in figure 66 (a) should be separated from the rest of the observations, since these are believed to be not directly connected to the main rivers. It can be seen that grouping by vicinity (figure 66 a) produces to many groups, whereas grouping by downstream flow paths produces only two groups instead of three. By filtering less sinks in the procedure of grouping based on flow paths, other observations are erroneously not being omitted from the group directly connected to the main rivers. Nevertheless the grouping using the downstream flow paths is assumed to perform best for the York case study, and was therefore used for grouping the observations.

Plain interpolation results

The results of using different combinations of parameters for creating the flood inundation map for the York case study are given in table 10. Since no water depths were given in the dataset for York, all observations were assigned a DWD.

Table 10: York - results of applying plain interpolation

Power	Smoothing (m)	DWD (cm)	F-value
1	0	10	0.26
1	0	20	0.23
1	0	30	0.25
1	0	40	0.26
1	0	50	0.26
1	0	60	0.26
1	0	70	0.26
1	0	80	0.27
1	0	90	0.33
1	0	100	0.33
1	50	90	0.33
2	0	20	0.50
2	0	30	0.52
2	0	40	0.51
2	50	30	0.50
2	100	30	0.50
3	0	20	0.48
3	0	30	0.47
3	0	40	0.49
3	0	50	0.50
3	0	60	0.49
3	50	50	0.50
3	100	50	0.50
3	200	20	0.58
3	300	60	0.59
4	0	20	0.48
4	0	30	0.49
4	0	40	0.51
4	0	50	0.52
4	0	60	0.50
4	100	50	0.52
4	300	60	0.51

The best results were produced using a power parameter of 3 in combination with a smoothing of 200 m and a DWD of 20 cm. Although using a slightly higher power parameter in combination with a higher water depth yielded a higher F-value, a quite large area was also erroneously mapped flooded. Therefore it was not considered an improvement.

Grouped interpolation results

The interpolation of groups of observations was applied, and several interpolation parameters were varied. The resulting values of the F-statistic are given in table 11.

Table 11: York - results of applying grouped interpolation

Power	Smoothing (m)	DWD (cm)	F-value
1	0	10	0.25
1	0	20	0.28
1	0	30	0.25
1	0	40	0.26
1	0	50	0.27
1	0	60	0.27
1	0	70	0.27
1	0	80	0.28
1	0	90	0.35
1	0	100	0.35
1	50	90	0.35
2	0	20	0.50
2	0	30	0.52
2	0	40	0.51
2	50	30	0.50
2	100	30	0.49
3	0	20	0.48
3	0	30	0.47
3	0	40	0.49
3	0	50	0.50
3	0	60	0.49
3	50	50	0.50
3	100	50	0.50
3	200	20	0.58
3	300	60	0.59
4	0	20	0.48
4	0	30	0.49
4	0	40	0.51
4	0	50	0.52
4	0	60	0.50
4	100	50	0.52
4	300	60	0.51

Overall these values are almost identical to values found by using plain interpolation. This is large due to the interpolation extents of both groups of observations overlapping. Due to this it is hardly surprising the optimal flood map is created at the same combination of parameters. The flood map resulting from using this parameter combination is given in figure 67.

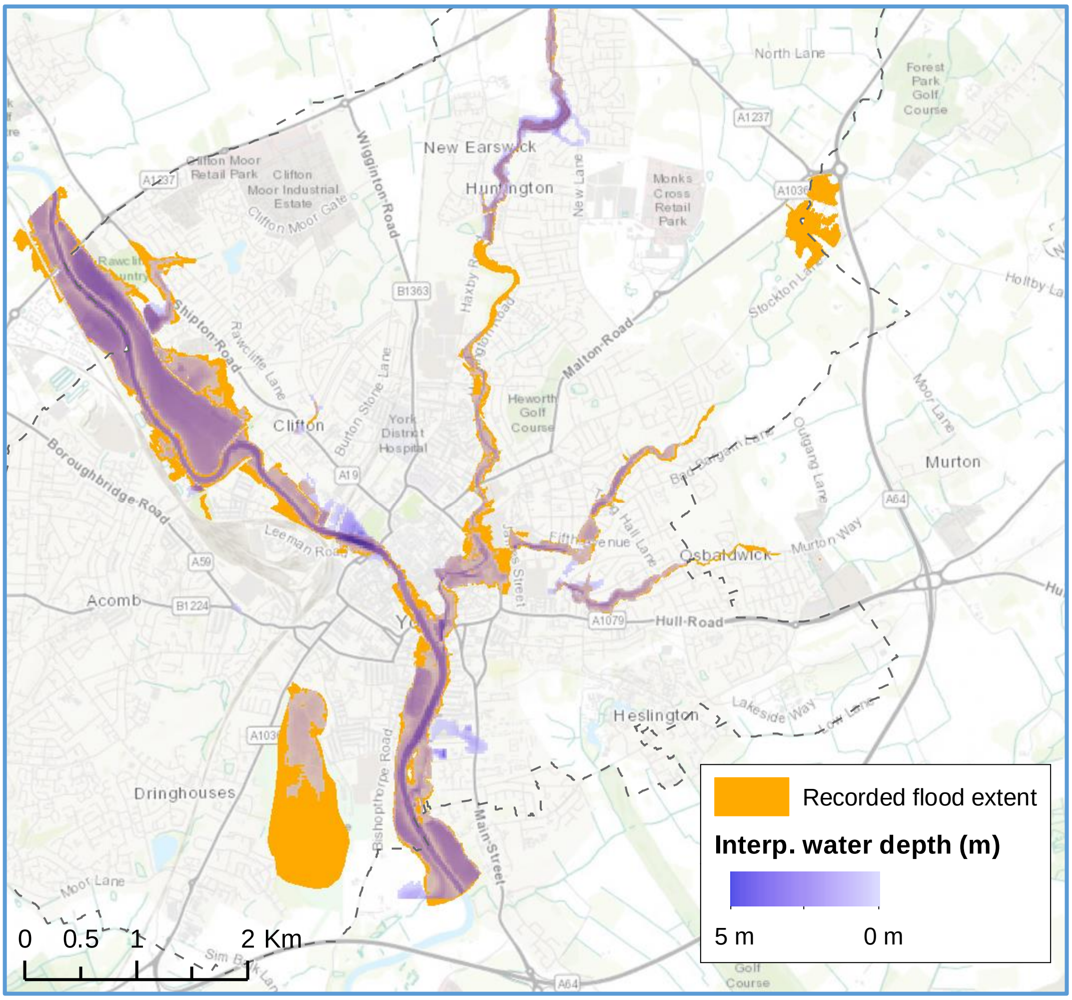

Figure 67: York - Flood map created using grouped interpolation (power=3, smoothing = 200 m) and a DWD of 20 cm.

Interpolation along flow paths results

The last and most promising method applied, was the interpolation of water levels along the downstream flow paths of observations. The results obtained by applying this method are given in table 12.

Table 12: York, results of applying interpolation along flow paths

Power	Smoothing (m)	Default water depth (cm)	F-value
1	0	10	0.44
1	0	20	0.48
1	0	30	0.53
1	0	40	0.56
1	0	50	0.58
1	0	60	0.50
1	0	70	0.50
1	0	80	0.49
1	5	40	0.47
2	0	30	0.58
2	0	40	0.60
2	0	50	0.62
2	0	60	0.63
2	0	70	0.65
2	0	80	0.56
2	10	60	0.62
3	0	30	0.63
3	0	40	0.65
3	0	50	0.66
3	10	40	0.65
3	10	50	0.67
3	10	60	0.68
3	5	50	0.67
3	5	60	0.67
4	0	30	0.63
4	0	40	0.64
4	0	50	0.65
4	10	40	0.66
4	20	50	0.67
4	30	70	0.70
4	40	70	0.70
4	30	40	0.67
4	30	50	0.69
4	30	60	0.69
4	30	70	0.70
4	30	80	0.59
5	0	30	0.61
5	0	40	0.63
5	0	50	0.63
5	20	50	0.68

Using high power parameters between 3 and 5, the results do not change significantly. Although the absolute highest F-values were found using default water depths of 70cm, a water depth of 50 cm is considered most optimal. Since using a water depth of 80 cm directly deteriorates the results severely, using a default water depth of 70 cm would make the results very sensitive for changes in this depth, and the addition of observations. Due to this using a DWD of 50 cm was chosen as an optimal value, in combination with a power parameter of 4 and a smoothing of 30 cells.

Uncertainty assessment

To review the uncertainties caused by the limited resolution of the DTMs used to create the flood maps, flood maps at different resolutions were created. The DTMs at 6 m, 10 m and 40 m resolution were constructed by resampling the original 2 m DTM. Both the threshold (in number of upstream cells) to identify drainage channels in creating the HAND map, as well as the parameter used to indicate sinks up to which surface area should be filtered, were changed in creating the 6 m, 10 m and 40 m Grids.

The threshold for identifying drainage channels was set to a value at which the main rivers in the area were identified. The parameter used to filter sinks up to a certain surface area was adjusted to the point only two groups of observations were identified, being the groups displayed in figure 66 (b). Only for the coarse 40 m resolution no accurate grouping could be applied, meaning three groups of observations were created. The settings used in creating the HAND maps and interpolation are given in table 13.

Table 13: Settings used for creating the HAND maps and flood maps at different resolutions

Resolution (m)	Threshold for HAND (cells)	Minimum allowed sink area (cells)
6	196,000	15,000
10	60,000	6,000
40	4,400	800