SAR-based flood mapping in urban environments

T.S. van der Zee

Technische Universiteit Delft

TU Delft Delft University of Technology

Challenge the future

SAR-based flood mapping in urban environments

by

T.S. van der Zee

in partial fulfillment of the requirements for the degree of

Master of Science
in Applied Earth sciences
Department of Geo-science and Remote Sensing

at the Delft University of Technology,

Supervisor:	Dr. ir. F.J. van Leijen	
Thesis committee:	Dr. ir. F.J. van Leijen	TU Delft
	Prof. dr. ir. R.F. Hanssen	TU Delft
	Dr. ir. M.C. ten Veldhuis	TU Delft

TU Delft Delft University of Technology

Abstract

Flooding is a very costly natural disaster especially when it hits urban areas. Yet synthetic aperture radar (SAR) based flood mapping barely works in urban areas. Buildings and man made objects have similar backscatter signatures as still standing water. This makes them hard to distinguish from one another. Structures can block the visibility of the ground surface for side looking SAR satellites making large parts of potentially flooded ground going unseen by SAR satellites. Smooth surfaces and limited ground visibility make it hard to produce accurate flood maps using SAR in urban environments. Here we have shown how the use of a temporal stack can improve the result of urban flood detection with SAR. Traditionally SAR based flood mapping uses a single image or an image pair to classify flooded and non-flooded pixels. This study found these methods unable to detect flooded pixels in an urban setting. By using a temporal stack of SAR images more pixels are correctly classified as flooded while keeping false positive classifications low. However the number off correctly classified pixels remains too low to be useful on its own, by adding ancillary data in the form of a high resolution DEM an accurate flood map for a very specific area is produced. This means that SAR images are not suitable for flood mapping in urban areas as a single source of information. When they are combined with other data they have the potential to produce accurate flood maps useful for First responders when the next flooding disaster hits.

Contents

1

Introduction

This chapter consists of an introduction to floods and their devastating effects. A description of current flood mapped methods is given. The description is followed by an overview of this research along with an explanation of the research question and its sub-questions. The chapter concludes with an outline of this thesis.

1.1. Floods

Flooding is a natural disaster that occurs when water levels rise to extreme extents and normally dry land finds itself under a layer of water coming from the sky, sea or river. The perception of most experts is that with a changing climate the frequency and severity of floods will rise (Taylor *et al.*, 2014). Due to the complex mechanisms that trigger a flood it is difficult to predict the exact increase of the frequency of floods or their extent. The fact remains however that a rising temperature will result in more evaporation and rainfall. Precipitation extremes will increase with the increase of the global temperature and extreme rainfall events can be a trigger of a flood (Allen and Ingram, 2002, Hegerl *et al.*, 2006). This, in combination with rising sea levels due to melting land ice and the thermal expansion of ocean water (Milne *et al.*, 2009), will not only result in more floods but also in larger floods (Garner *et al.*, 2017).

The potential economic and social impact of flooding can be extremely high. Take as an example the river Rhine , 10 million people live in areas prone to extreme flooding of this river and as a consequence the potential damage of floods in this area was estimated at 165 billion euro's (Friesecke, 2014). An other study estimated that in 2004 the economic value of everything located within 500 meters of the European coastline was between 500 and 1000 billion euro's (Doody *et al.*, 2014). While these were theoretical estimates of assets which could be threatened by floods, the European environment agency has estimated that in the period between 1998 and 2009 the actual damage of floods in Europe was approximately 52 billion euros, 1126 lives were lost and over half a million people were displaced or needed to flee their homes (European-Environment-Agency, 2004, 2011). This makes flooding the most costly natural disaster in Europe. An overview of natural disaster impacts for the European Union is given in Table 1.1.

Maps of flood extent can be a valuable object of information for several purposes. First of all, in the event of a disastrous flood, a high resolution map of the flood extent could be used as a source of information for first responders and emergency services. They could use it to identify which houses or streets are flooded, which regions are in particular need of urgent help and possibly how to get there over non-flooded roads. Hydrological modelling is another area where high accuracy flood extent maps would be useful. Flood maps would serve as validation for the models, validation data could lead to improvements of models or assessing the accuracy of models. Lastly, such maps could prove useful when looking for patterns in flood behaviour. For instance, if the maps show that certain locations are flooded on a regular basis, or that certain spots are always the first to flood and the last to dry. The responsible authorities could use the information to identify and prioritise areas where the sewer

system needs upgrading.

Hazard type	Recorded events	Number of fatalities	Overall losses (EUR billion)
Storm	155	729	44.338
Extreme temperature events	101	77 551	9.962
Forest fires	35	191	6.917
Drought	8	0	4.940
Flood	213	1 126	52.173
Snow avalanche	8	130	0.742
Landslide	9	212	0.551
Earthquake	46	18 864	29.205
Volcano	1	0	0.004
Oil spills	9	n/a	No comprehensive data available (a)
Industrial accidents	339	169	No comprehensive data available (b)
Toxic spills	4	n/a	No comprehensive data available (c)
Total	**928**	**98 972**	**148.831**

Note: (a) Estimation is between EUR 500 and EUR 500 000 per tonne of oil spilled.
(b) Costs for major events reported in Table 12.1 aggregately amount to more than EUR 3.7 billion.
(c) Costs for one particular toxic spill amount to EUR 377 million, see Chapter 13.

Source: EM-DAT, 2010; EMSA, 2010; MARS, 2010.

Table 1.1: Contains an overview of damage from natural disasters in the European Union between 1998 and 2009. The overview gives per natural disaster type, the number of recorded occurrences, the number of fatalities and the overall losses expressed in billions of euros.

Flooding occurs both in rural and urban areas, but the impact on our society in terms of economic losses and injuries are much larger when they occur in urban areas (Mason *et al.*, 2010). Urbanisation of flood planes and land conversion for urban development decrease the water permeability of the soil. This increases the potential impact and likelihood of floods in urban areas (Depietri *et al.*, 2012). Due to the increased impact that flooding of urban areas has on society and the increased likelihood of floods occurring in urban areas, there is a large interest in mapping and modelling these floods.

When mapping the extent of floods there are several sources of information available. Sometimes water level sensors are present but these aren't able to give a complete picture but instead they provide information at a fixed number of locations (Demir and Krajewski, 2013). Therefore, they are useful for validation of other data. Information about the flood extent could be derived through crowd sourcing, gathering eye witness reports, videos or photos witnesses made (Cariolet, 2010). Satellite imagery is widely available, covers large areas simultaneously and comes in a variety of forms that could be useful for flood mapping. Optical imagery from satellites is one of them. Unfortunately, flood events are almost always combined with a cloud cover that obstructs the optical images and possibly renders them useless. Optical images are also dependent on the presence of daylight in order to be useful for flood extent mapping (Mason *et al.*, 2013). Synthetic aperture radar (SAR) is another form of satellite imagery that contrary to optical imagery is not influenced by cloud cover or dependent on the presence of sunlight. SAR is an active remote sensing technique, therefor not depending on the sun or other source of signal other than its own signal. The radar waves that are used for SAR penetrate any cloud cover that is present. This is why SAR data imagery is preferred over optical images for flood mapping (Mason *et al.*, 2013).

Current day technologies enable us to observe and map rural flooding and its extent using SAR imagery with good accuracy (Mason *et al.*, 2012).Flooded areas have a low intensity on radar images because a flat water surface will act as a specular reflector as Figure 1.1 describes.Flood extent maps in rural areas derived from SAR data have been successfully used to calibrate 2D inundation models (Mason *et al.*, 2010). These modelled flows take into account the ground topography and vegetation (Mason *et al.*, 2013). However, mapping urban flood extent using SAR proves to be a lot more difficult.

That urban flood mapping proves to be difficult can be seen in Figure 1.2. The image shown is a

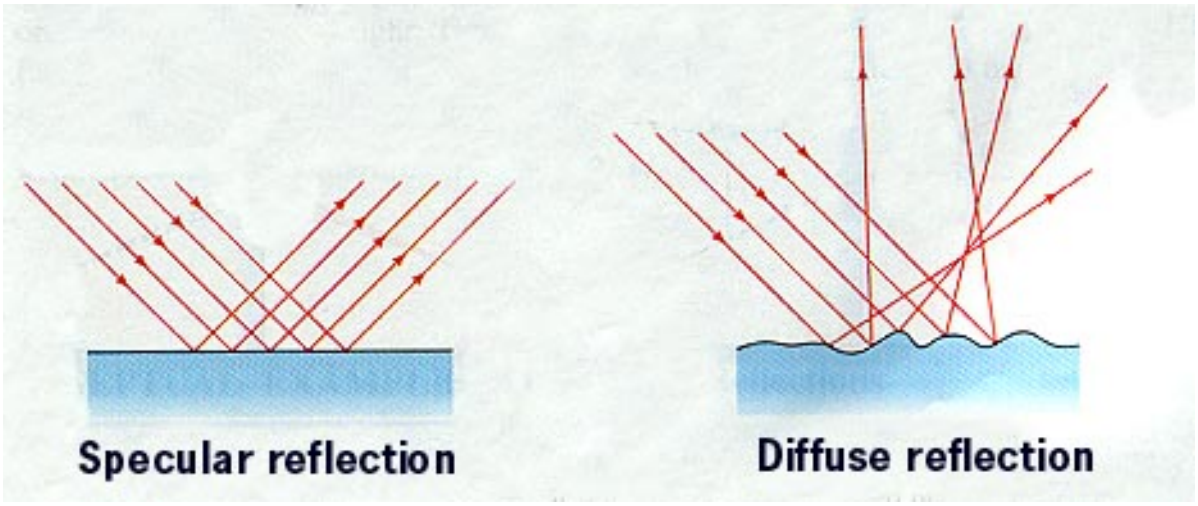


Figure 1.1: Visualisation of the reflection of radar signals on a smooth surface (like water) compared to a reflection of radar signals on a rough surface.

map of the flood extent during hurricane Harvey in the Houston metropolitan area and its surroundings produced by the Copernicus Emergency Management Service. On this flood map large areas around the urban city area of Houston are denoted as flooded, there is however no flooding classified in the city centre of Houston. Figure 1.3 gives a different view, it shows a part of the Copernicus image (a) next to an optical image (b) taken on the same day as the SAR image on which the Copernicus flood map is based. In the optical image there are entire neighbourhoods that are clearly flooded while there is no flooding present in the Copernicus flood map. This is a good example of how the current flood extent mapping methods which can detect rural floods fail to classify urban floods as such.

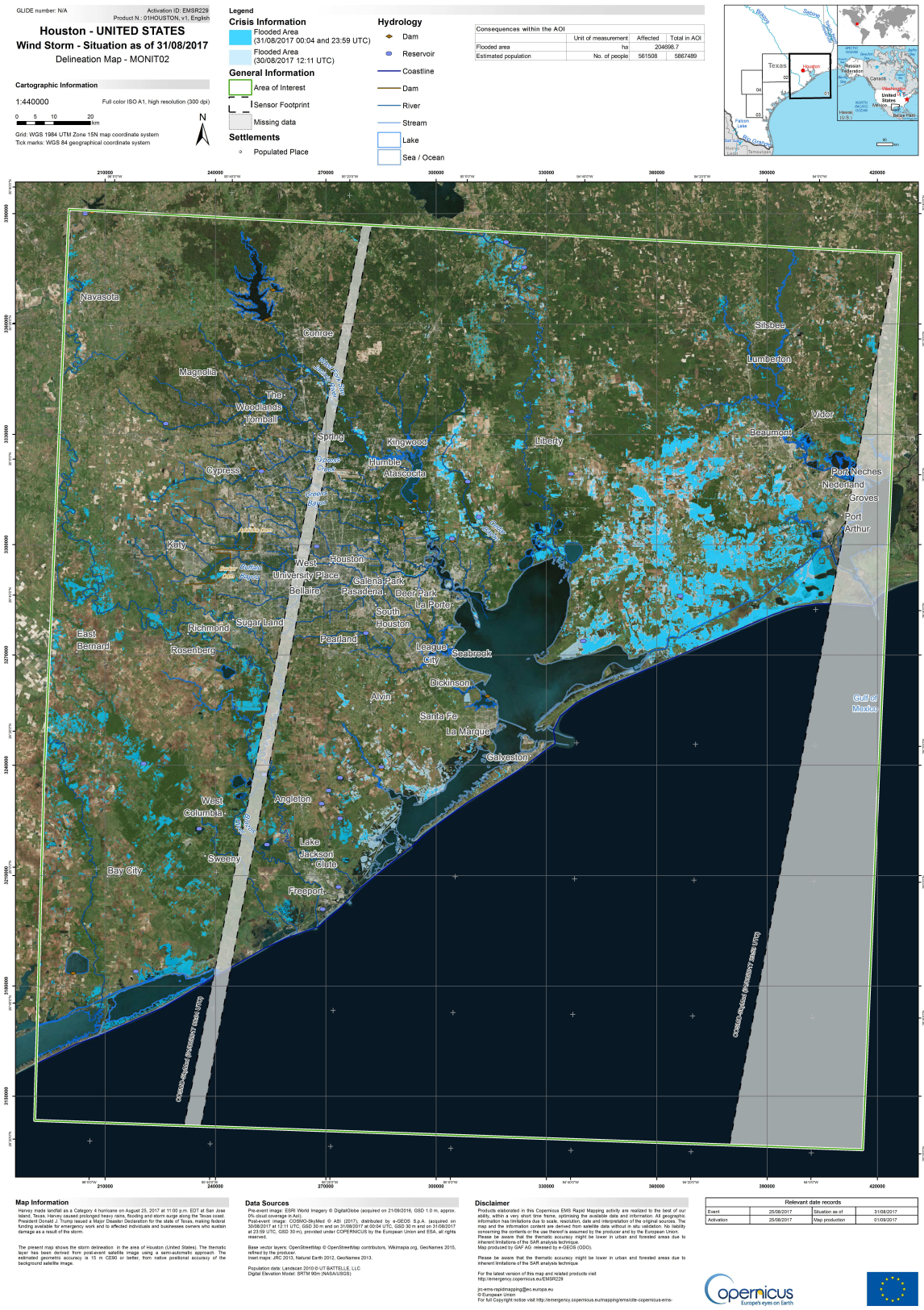


Figure 1.2: Flood map of the metropolitan area of Houston and its surroundings produced by Copernicus Emergency Management Service on the 31st of August 2017. The image shows little to no flooding in the city centre while around the city large regions are classified as flooded.

(a) Zoomed in on an urban section of Figure 1.2.

(b) Part of aerial image made on August 30th.

Figure 1.3: Comparison between flood map made by Copernicus and an aerial image from NOAA. Both images are taken or produced on the 31st of August 2017. The top image contains no flooding outside of the dam while on the bottom image neighbourhoods are clearly flooded.

1.2. Research overview

1.2.1. Objective

The objective of this study is to improve the (semi-)automatic mapping of flood extent in urban areas based on SAR amplitude data. The research aim is to investigate the possibility to create a reliable flood map shortly after an event to be useful for flood crisis management in the circumstance of large floods in urban areas based on SAR data. The study would achieve this objective by creating a functioning (semi-)automatic tool that can process SAR images over areas that are known to be flooded immediately when they become available. The resulting flood map would lead to a better knowledge of the extent and mapping of the flood. The main purpose of this research is to answer the following question:

Can the use of a temporal stack of SAR images improve the mapping of flood extent in urban areas?

Most research on mapping flood extent with SAR has been done on a single image or on a single pair of images. Furthermore most research that processed entire regions was done on rural areas. The research that did focus on urban areas looks at very specific parts of urban streets and develops a method for that specific category of pixels. They do achieve high accuracy results in the range of 75 percent and upwards of correctly classified flooded pixels but see those numbers drop when their method is tested on all pixels. This is why this research will try to use a stack of images instead of a

single or pair of images to see if this leads to improvement of the result.

Sub-questions

With the main question formulated a number of sub-questions can be identified:

1. *How do floods influence the radar reflections measured with SAR?*
   In order to tell apart flooded pixels from non-flooded pixels we need to understand how the flooded pixel values have changed compared to the non-flooded pixels. What happens to the radar reflections when they reflect of a flooded surface compared to when they reflect from a dry surface. When we understand how it changes we can start to quantify it and use it to classify each pixel.

2. *Which methods can be applied to detect flooded pixels from a stack of SAR amplitude images?*
   How can we use a stack of SAR images to detect the difference between the flooded and non-flooded pixels in one SAR image.

3. *Is it feasible to produce a flood map of an urban environment based on SAR data alone or is auxiliary data required to reach a reliable result?*
   Not all streets are visible in urban environments, so will it be enough to only use the streets that are or do we need other data sources to fill in the gaps? Is the change that is caused by floods enough to distinguish flooded pixels from non-flooded pixels?

1.2.2. Method

Starting from a literature review on the topic of SAR flood mapping and SAR flood mapping in urban environments the research will evaluate flood mapping using data from the test case of the Houston floods in 2017. The case uses a stack of Sentinel-1 SAR images taken in VV-polarisation spanning a period of almost three years. The data and test case will be used to test five different methods of which two use a single or double image while the other tree use a stack of images. Based on intermediate results, some methods are pursued more deeply than others in order to reach an optimal flood map for the test case. In order to validate classification results the study uses aerial imagery taken on the same day as the SAR image of the flood.

1.2.3. Thesis outline

A short overview of what to expect in the remainder of this Thesis. Chapter 2 gives a background into all topics relevant for this research. A short introduction of SAR and how water influences these measurements is explained. This is followed by a summary of the research that has already been done on this topic or that is useful for this research. Than comes the description of the test area as well as a background on all data used. Chapter 3 explains the five methods that are tested to produce flood maps in detail and how the methods are evaluated on performance. Chapter 4 goes through the initial results of the five methods. It holds a brief discussion on the results and possible alteration to the methods to improve the results. Chapter 5 hold the discussion of the final results, the shortcomings and strengths of the methods are elaborated on. Chapter 6 draws the conclusion of the thesis and answers the research questions posed in the beginning of the report. It ends with recommendations for future research on the topic of flood mapping in urban areas using SAR.

2

Background

In this chapter the basics of radar, RAR and SAR are explained. The difficulties of flood mapping in urban areas using SAR will be presented alongside an overview of past studies on flood mapping, both in rural and urban specific areas. Furthermore a short explanation on radar technologies is given. The selection of a test case is made and finally the chapter ends with a description of all data used in this research.

2.1. Radar basics

In order to understand the basic concept of SAR a basic understanding of radar and Real Aperture Radar (RAR) is required.

Radar

A radar (Radio Detection And Ranging) system is an active type of remote sensing, meaning that it sends and receives signals. A radar system works in the radio and microwave band of the electromagnetic spectrum. A radar system thus consists of a transmitter that sends out electromagnetic pulses to a target and an antenna that measures received pulses. A radar can use the same antenna for transmitting and receiving so it doesn't require two different antennas although there are radar setups that use different antennas. The radar system measures three things from the pulses it receives: The intensity of the signal, the phase of the signal and how long it takes from sending the pulse and receiving a pulse. The range between the radar system and the target can be derived from the travel time of the pulse. The signal that is measured at the antenna, so called backscatter, is a part of the pulse that reflects or scatters from the target back towards the radar antenna. The intensity of the measured pulse can tell something about the size or surface roughness of the target (Hanssen, 2001).

Radar geometry

Most satellite radar systems are of a side looking nature, meaning the radar doesn't measure straight down to the surface but measures under an angle to the side of the satellite (Bamler and Hartl, 1998). This angle is called the look angle (often denoted by θ) and is visualised in Figure 2.1. When using an air- or space-borne radar angles and positions in the geometry have their own reference system. The coordinate system consists of a range and an azimuth direction. The azimuth direction is parallel to the flight path of the satellite, the range direction is perpendicular to the azimuth direction with the positive side in the direction where the satellite is looking. The point on the earth's surface directly beneath the radar system is called nadir. The nadir line is the line directly beneath the flight path of the system on the Earth's surface. The width of the ground in the range direction that is illuminated by the radar system is called the swath. The swath goes from near-range (closest to the nadir line) to far-range (furthest away from the nadir line). Slant range is the distance between the satellite and the target while ground range is the distance between the nadir and the target over the earths surface. A more complete and detailed glossary on radar, RAR and SAR can be found on the ESA website[1].

[1] http://envisat.esa.int/handbooks/asar/CNTR5-2.html

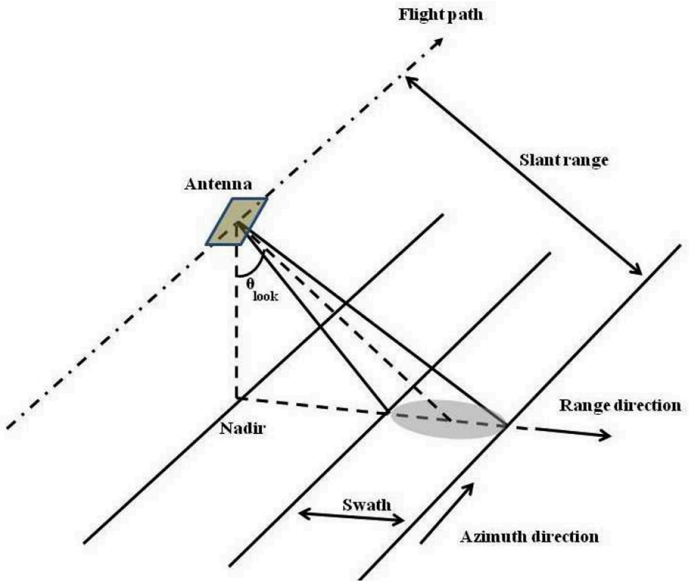


Figure 2.1: Shows a schematic overview of Space/air borne radar imaging geometry and its nomenclature.

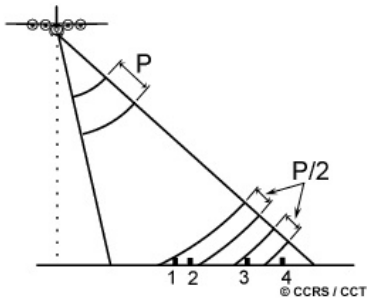
Real Aperture Radar

RAR is a moving radar system that makes images of the ground while it moves. The resolutions of these images are different in the range and azimuth direction. The resolution in the range direction depends on the length of the pulse transmitted by the system, While the resolution in the azimuth direction depends on the width of the pulse beam and the slant range. Both the range and azimuth resolution are illustrated in Figure 2.2. In this figure objects 1 and 2 are too close to be measured separately in the range direction while objects 3 and 4 are measured separately. The opposite is true for the objects in the azimuth direction. The resolution is constant in slant range but changes in the ground range as it is dependent on the incidence angle which changes in the range direction. The azimuth resolution can be changed by changing the angular width of the system, this width is determined by the antenna length. A larger antenna leads to an narrower beam. Because the length of a space born antenna is limited and the slant range is in the order of hundreds of kilometres the azimuth resolution of RAR is limited to a large resolution.

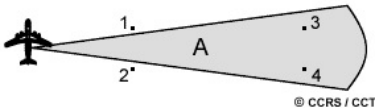
Synthetic Aperture Radar

The difference between RAR and SAR lies in the computation of resolution. The resolution in the range direction is the same as the RAR resolution. But where the resolution in the azimuth direction of RAR is limited by the antenna length SAR can create images with a greater resolution in the azimuth direction. SAR uses a processing technique where it is possible to simulate a very long antenna (synthetic aperture). This is illustrated in figure 2.3. All the radar positions in this figure have object A inside the radar beam. A synthetic antenna can now be created from the radar location it first saw object A to the position at which A was last visible This gives an antenna of length B in figure 2.3. By processing all the radar images object A is in results in a finer and uniform resolution in the azimuth direction. More on RAR and SAR resolution is provided by NRCAN fundamentals of remote sensing[2]. An example of a SAR system is the Sentinel-1 mission. The data that it produces has a resolution that ranges from 2.7x22m to 3.5x22m[3] pixel sizes.

[2]https://www.nrcan.gc.ca/node/9341
[3]https://sentinel.esa.int/web/sentinel/user-guides/sentinel-1-sar/resolutions/level-1-single-look-complex



(a) Image of how pulse length affects range resolution, objects 1 and 2 can not be distinguished separately while objects 3 and 4 can.



(b) Image of how pulse length affects azimuth resolution, objects 3 and 4 can not be distinguished separately while objects 1 and 2 can.

Figure 2.2: Visualisation of RAR resolution and how it differs in range and azimuth direction.

Speckle

Next to the resolution and geometry a phenomenon that influences the result is called speckle. Speckle is the reason that two neighbouring pixels that contain the same surface can have different values, this leads to the grainy texture that is characteristic of radar images. Speckle occurs because inside each resolution cell there are several scattering points that create positive or negative interference. This is different for each cell and can result in two neighbouring pixels that both contain asphalt having a different intensity value on the radar image.

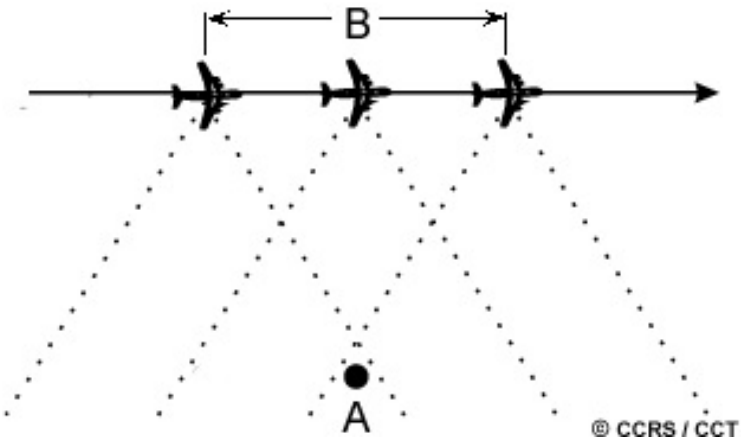


Figure 2.3: Visualisation of SAR resolution and how it uses several images to reduce its resolution compared to RAR.

Data format

The SAR images come in a line-pixel coordinate system with each pixel containing a complex number

$$a + bi \tag{2.1}$$

where a and b are real numbers.

This complex number represents amplitude (A) and phase (ϕ), Figure 2.4 visualises this. The amplitude of each cell can be extracted using

$$A = \sqrt{a^2 + b^2}. \tag{2.2}$$

For this project the amplitude is converted to a decibel scale using

$$A_{dB} = 10 * \log_{10}(A). \tag{2.3}$$

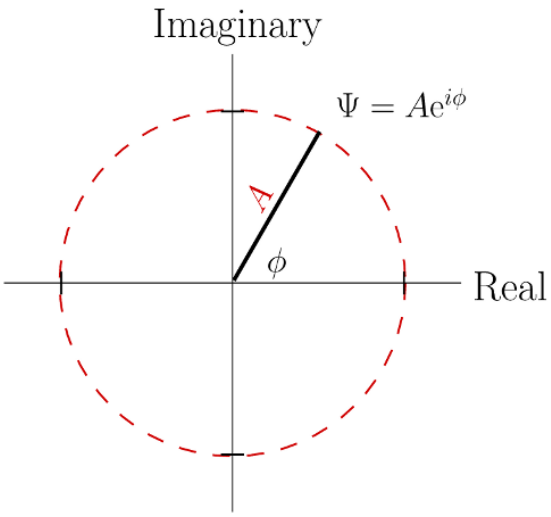


Figure 2.4: Visualisation of how phase and amplitude are stored in a complex number. A is the amplitude and ϕ is the phase of a signal.

2.2. Urban flood mapping with SAR

An urban landscape can be divided into 4 main different categories concerning SAR and flooding:
1.) The first category contains rooftops and other surfaces that lie higher than the ground.
2.) The second category is part of the street that can't be seen by the satellite due to buildings, vegetation, terrain or other man-made structures. This phenomenon is called shadow and is depicted by line CD in Figure 2.5.
3.) The third category is the part of the street that can be directly observed by the radar on the satellite, it is depicted by line BC in Figure 2.5.
4.) The fourth and last category is depicted by line AB in Figure 2.5 and is called layover.

For this research there is no further interest in category 1, as this category will not contain information about flooding on the street level. There is also no further interest in category 2 as there are simply no measurements from this part of the street. Category 3 seems the most straightforward category to work with as the street level is directly measured by the satellite. How large the proportion of a street is that falls into the third category depends on the height of the surrounding buildings, the street width, satellite viewing angle, and the satellite track. Category 4 is often present and harder to interpret than category 3. There are two reasons for this. The first is that the reflection from this part of the street isn't directly measured but is registered after it is reflected again from the side of a building or other object. This phenomenon is called double bouncing and results in a strong return at the radar. An example is displayed as beam M in Figure 2.5. The second is caused by point O and point B in Figure 2.5 being at the same distance from the radar. This means that the points fall into the same range bin and aren't distinguishable from one another.

The first problem when mapping the flood extent using SAR is that part of the flood will not be detected as it isn't seen by the satellite. This concerns all flooded streets that fall in category 2.

The second difficulty that urban flood mapping brings relative to rural flood mapping is the interaction of radar waves with man-made surfaces like concrete or tarmac roads. In SAR images of rural regions water bodies (with low wind and flow conditions) act like specular reflectors as explained in figure 1.1 resulting in a low intensity measurement. The surrounding regions of grass, sand, trees, gravel and bedrock will act more like diffuse reflectors prompting a higher intensity on the SAR image. This is why water bodies tend to clearly stand out in SAR images. This is illustrated by Figure 2.7 where an intensity SAR image is displayed of the Houston metropolitan area, the San Jacinto River, Trinity Bay and part of the Gulf of Mexico. Unfortunately roads tend to have a flat surface that acts as a specular reflector of radar waves, resulting in approximately similar intensity measurements as water. This is illustrated by the histogram in Figure 2.6. In this figure the intensity values of a highway section can be compared to the values of a part of a lake in a SAR image of Houston. These pixels were chosen as

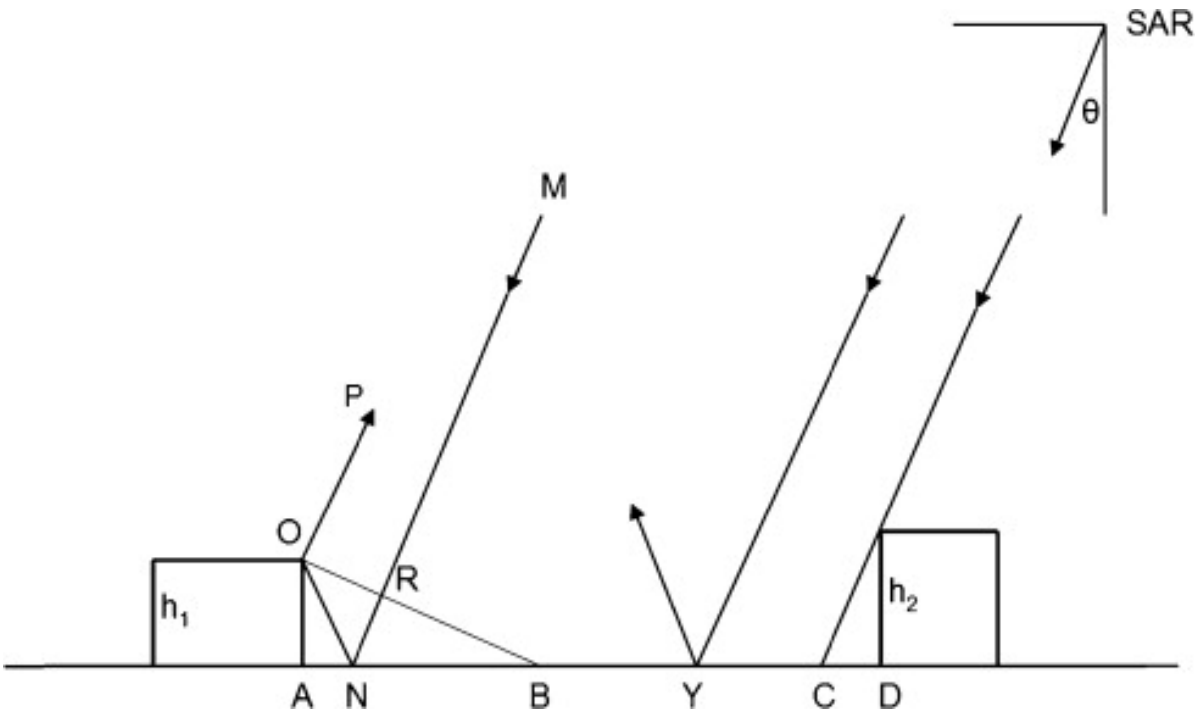


Figure 2.5: Shows different radar beam paths in urban area, detailing how buildings affect different parts of a street in different ways. The image is from (Mason *et al.*, 2013).

an example from a random part of a large water body and a section of interstate 10. Simply classifying low reflections as water in an urban SAR image will result in a lot of road pixels being classified as water hence being seen as flooded while this isn't the case. In a rural setting a small fraction of the total area is covered by roads or concrete so there is a small impact on the classification. In urban areas however the fraction of surface covered with man-made smooth surfaces is considerably larger making it an issue for urban flood mapping.

The final problem of flood mapping in urban areas compared to rural areas comes from reflections in category 4. The multiple reflections in this part of a street result in a higher intensity measurement at the satellite. Flooding may even increase the intensity as more of the signal is reflected towards the satellite. In rural areas flooding leads to a lower intensity measurement but due to category 4 reflections some parts of an urban area flooding can lead to a higher intensity.

The combined effect of the layover and shadow makes it harder to measure and map floods in urban areas with SAR than it is in rural areas. In the city of Karlsruhe (Germany) a study found that two thirds of the streets are unseen by direct radar measurements (Soergel *et al.*, 2003). How much of a particular street falls into each category doesn't only depend on the street width or the height of adjacent buildings, it also depends on the street orientation relative to the satellite direction of travel. A street perpendicular to the path of a satellite will be almost entirely visible for the radar. A street parallel to the orbit of a satellite will contain a maximum amount of shadow and layover and might not have any part measurable by the radar directly. The complicated nature of urban areas makes it more difficult to detect flood water in urban areas than in rural areas in rural areas than in urban areas (Mason *et al.*, 2010).

The difficulty of mapping urban floods in practice can be illustrated with the flood map produced by the Copernicus program of the European Commission. Their flood map is displayed in Figure 1.2. In this image light blue indicates areas flooded on August 30th and blue indicates areas flooded on August 31st. On the flood map most flooding appears on the right side of the image near the coast, these lands consist mainly of farming fields. On the left side of the image there are fewer flooded areas and they are situated along the sides of rivers. In the metropolitan area of Houston there are barely any areas marked as flooded. This is remarkable because images of flooded streets, highways and

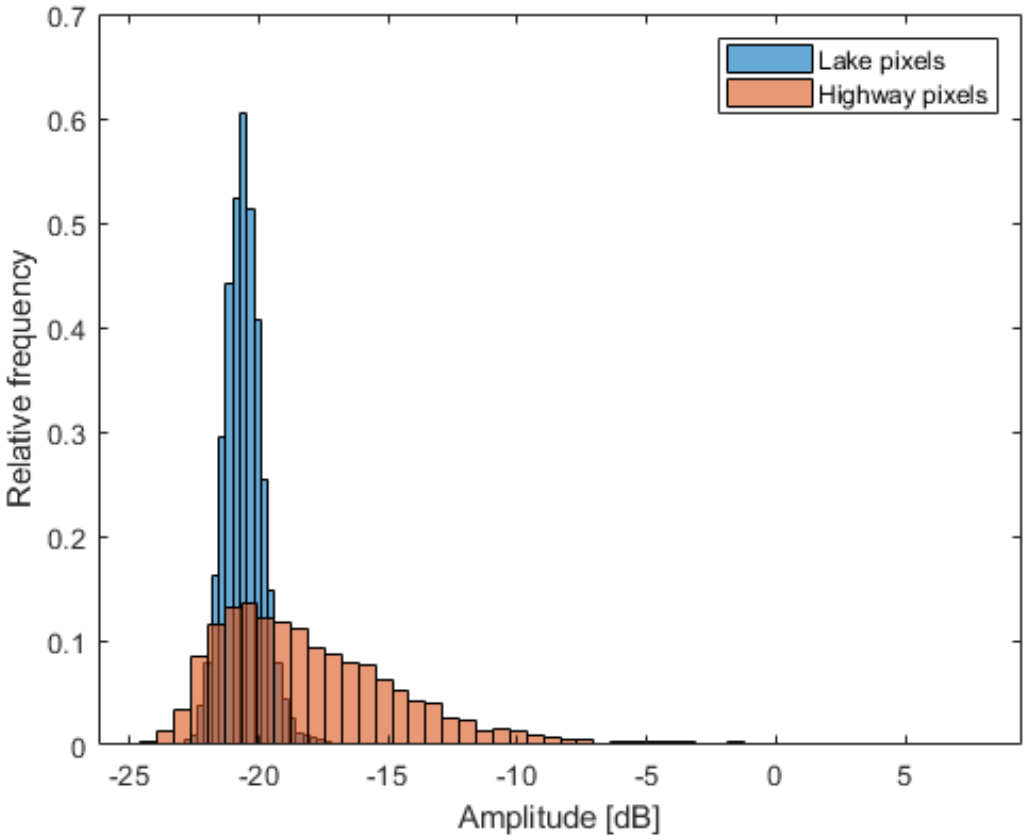


Figure 2.6: Normalised histogram containing pixels of a part of highway and of a part of lake. The histogram demonstrates the similarities in pixel values for roads and water pixels.

entire neighbourhoods were all over the news and can still be found with a Google search of ”Houston flood 2017”. In Figure 1.3 part of the flood map is shown next to an aerial image, the flood map shows floods of 30th and 31st of August and the aerial image was taken on the 30th of August. In this Figure the flood behind the dike in the top left is visible on both images. This is all the flooding indicated by the Copernicus image while on the aerial image it is clear that large parts of the bottom right part of the image are flooded. The Copernicus program is thus able to detect flooded open fields but doesn't classify flooded urban neighbourhoods as such.

In short the difficulties with flood mapping based on SAR images in an urban environment: Man made surfaces such as buildings and roads have smooth surfaces that can act as a secular reflection just as water. This leads to a smaller difference between water surfaces and the ground surfaces. Complication for SAR measurements are the sudden height differences that are caused by buildings or other objects. These differences can lead to parts of streets not being visible for the radar system while other parts will have a higher intensity measurement compared to when no building or height difference would be present. To visualise the influence of these height differences there are several scenarios shown in Figure 2.8. From this image we can expect that floods in urban areas result in the same intensity, a higher intensity and a lower intensity compared to normal conditions.

2.3. Past studies on Flood mapping with SAR

Several research projects have made an effort to map flood extent in both urban and rural areas using SAR images.

A study from D.C. Mason in 2010 focused on classifying pixels from category 3 (pixels in direct view of the satellite) in a city, which he selected using LiDAR data, based on a single SAR image using a semi-automatic algorithm (Mason *et al.*, 2010). Their method used a region growing algorithm to

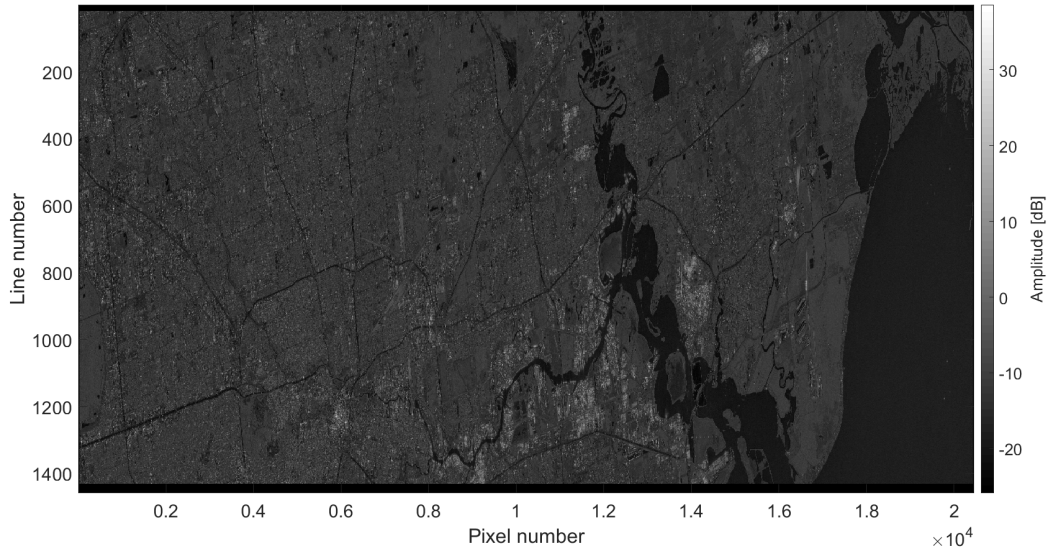


Figure 2.7: Mean intensity image of Houston (part of Sentinel-1, swath 1), the mean is taken from 89 images over the period from 2015-07-18 till 2017-08-24. The SAR images taken during the flood are not included in the mean. On the image it is visible how the water bodies on the right side of the image stand out as dark areas of low intensity compared to the rest of the image. The surface covered in the image is approximately 21km by 86km.

identify regions of homogeneous speckle statistics. They then continued to classify these regions using a supervised classification method. With the validation of aerial photographs the method correctly identified 76 % of water pixels from a flood in Tewkesbury (U.K.) and had a false positive rate of 25 %. These numbers were 58 % and 19 % when all pixels classified as flooded were considered and not only those of category 3.

Mason *et al.* (2013) conducted a similar study, but they only focused on pixels that fall into category 4 (Lay-over areas) (Mason *et al.*, 2013). The study used a SAR simulator based on a LiDAR elevation map to identify regions of layover. The observed strength of double scattering was compared to predicted return strength based on an electromagnetic scattering model to determine whether or not a pixel was flooded. The observed strength was also compared to the observed strength of the same pixel in an earlier recorded SAR image to classify it as flooded or not flooded. For both the single image as well as the change detection method only a very small set of pixels was tested but 100 % of the flooded pixels was correctly identified and 91 % of non-flooded pixels was identified in both cases. This method seems promising due to its highly accurate results but was only applied on a very limited number of pixels, on only one flood event and required the availability of high resolution SAR data and a high accuracy DEM.

In general most studies and algorithms are based on a single image or a single pair of images (Schlaffer *et al.*, 2014). In the single image case a distinction between flooded and non-flooded is made based on pixel values where-as the pair of images case uses the difference between the two images for the classification process. There are a few that use multiple images for flood detection like a study performed by Nico *et al.* (2000). They estimated flood extent using the difference between three SAR images and compared it to the result derived from InSAR coherence images. Both InSAR and SAR were able to produce maps of the flood extent but combining the data resulted in a smoother result. This study was performed on an image with a resolution of 150 meter and in rural areas.
S. Schlaffer et al. published a paper in 2014 that looked at a longer time scale to characterise a seasonal variable back-scatter under non-flooded conditions (Schlaffer *et al.*, 2014). When a pixel value had a larger deviation than a certain threshold from the backscatter expected during a certain season it would be classified as flooded. The study reached accuracy's over 80% for non-urban areas at a pixel resolution of 150m.

All of these studies were either performed on rural areas or on specific parts of an urban area. When the methods are performed on an entire image with urban areas the accuracy of classifying flooded pixels drops. The problem therefore remains of detecting a flood in a wide urban area with a single method. This is exactly what we see in the flood map of Copernicus displayed in Figure 1.2 and 1.3, where rural areas are classified as flooded but none of the flooding in the city is detected.

2.4. Floods and available SAR data

For the application of flood extent mapping with SAR data, the timely availability of the SAR data is of importance. Satellite radar systems fly in orbits which means that they don't observe the same area continuously but return to a particular area every few days. The so-called revisit time is different for each satellite and depends on its orbit and velocity. This means that a satellite will not record every flood as the flood could occur and disappear before the satellite revisits the area. The SAR missions that are currently operational are shown in Table 2.1. From these missions some measure continuously while others only scan certain areas; some of the data is freely available while other data comes from commercial companies and require payment.

SAR System	Launch	Frequency/Polarization	Country
TerraSAR-X/TanDEM-X	2007	X (quad)	Germany
RADARSAT-2	2007	C (quad)	Canada
COSMO-SkyMed 1–4	2007–2010	X (dual)	Italy
RISAT-1	2012	C (quad)	India
HJ-1C	2012	S (VV)	China
Kompsat-5	2013	X (dual)	Korea
PAZ	2013	X (quad)	Spain
Sentinel-1	2013	C (quad)	ESA (European consortium)
ALOS-2	2013	L (quad)	Japan
RADARSAT Constellation Mission 1–3	2017/2018	C (quad)	Canada
SAOCOM-1/2	2014/2015	L (quad)	Argentina

Table 2.1: Overview of currently operational SAR missions, their launch dates, frequency, polarisation and the country that operates the mission. Source (Brisco, 2015)

2.5. Houston flooding test case

The selection of the flooding test case for the study was based upon two criteria. Firstly there needed to be a SAR image taken during the flood and a stack of SAR images from the months leading up to the flood. Secondly validation data of the flood extent from an other source had to be available to assess the accuracy and quantify the result of the proposed flood mapping method.

TerraSAR-X images that have a resolution of 3m were available for large parts of the Netherlands. Unfortunately for this research there were no floods that that occurred when this satellite passed over the Netherlands. Partly because the satellite system has a revisit cycle of 11 days (Roth *et al.*, 2004) and partly because floods in the Netherlands are quite rare, let alone a major flooding that lasts for multiple days. With no overlap in floods and satellite images from the TerraSAR-X satellites, other satellite systems with lower resolution were considered. RSAT-2 and Sentinel-1 data were available but had the same problem as TerraSAR-X of a mismatch between days of floods and days of images taken. An overview of floods considered and SAR data available is given in appendix B. This meant flood events outside the Netherlands needed to be considered. A case of which SAR images taken during a flood were available was the flooding of Houston in 2017 caused by hurricane Harvey in Texas (U.S.A). There were SAR images available from August 2015 till the flood in August 2017 with intervals of 12 days. There was one SAR image taken during the flood and on the same day as an aerial photograph survey was executed. These photos could serve as the validation data. In Figure 2.9 one part of the aerial survey is displayed and shows that the flood extent can be determined from the picture.

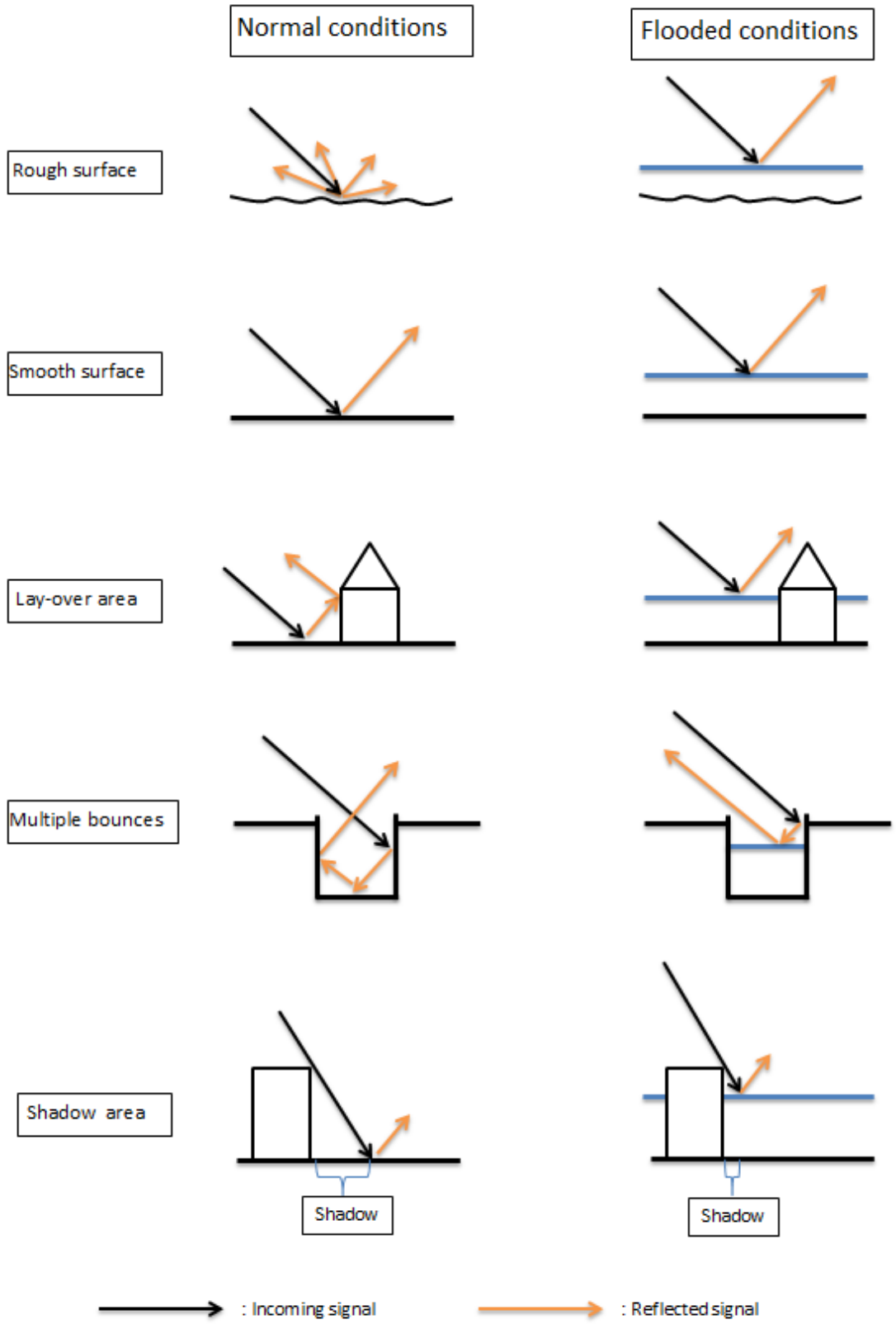


Figure 2.8: Schematic and simplified overview of how different urban surface layouts can affect different radar beam paths under normal conditions and exaggerated flood conditions.

Figure 2.9: Aerial photo of a Houston suburb acquired on the 30th of August 2017. The image clearly shows a large urban area that is flooded. Source: NOAA Hurricane Harvey imagery site.

Houston is a city located on the east coast of the United States of America as shown in Figure 2.10. With 2.3 million people it's the largest city of the state of Texas and the fourth largest city of the U.S.A in terms of population (U.S. Census Bureau, 2017). The city is situated in a wetland coastal area that naturally holds a lot of water and is in the drainage basin for water coming from further inland. The rapid urbanisation of the area has not only led to many houses being built in flood prone areas but also in the reduction of the soil permeability (Davis *et al.*, 2017). This reduces the water that can be absorbed by the ground.

(a) Location of Houston in the U.S.A.

(b) Aerial image of Houston.

Figure 2.10: Location of Houston in the U.S.A and an aerial image of the city. The Aerial image shows how the city is situated at the coast and to a large bay.

The flooding of the Houston metropolitan area in August of 2017 was a result of hurricane Harvey passing over. Harvey was a category 4 hurricane, the storm stalled when it made landfall on the coast of Texas and remained there for 4 days. Harvey brought the largest amount of rainfall ever recorded in the United States, peaking at 1538 mm of accumulated rainwater from the 25th of August till the 1st of September near Nederland, Texas (Blake and Zelinsky, 2017). The Houston metropolitan area received rain varying from 914 mm up to 1219 mm of accumulated water in the period from the 25th of August till the 1st of September (Blake and Zelinsky, 2017). To put that in perspective, the total rainfall in the year 2015 for the Netherlands was 880mm (Rijksoverheid, 2016). It was not only the

extreme rainfall (mm/m^2) that caused the severe flooding but the extent of the heavy rainfall (km^2) that was truly overwhelming.

On top of the incredible precipitation Harvey produced a storm surge that, combined with incoming tides, produced a water level height of 1.83 m to 2.74 m above ground level in the Trinity and Galveston bay's (Blake and Zelinsky, 2017). This might not be nearly as high as the storm surge from Katrina in 2005 but it did reduce the natural outflow of water from land to sea. This caused water to flow up to Houston from both inland as well as from the sea (Fischetti, 2017).

The final contribution to the floods was the decision of the army corps of engineers to release water from the two reservoirs of Houston. This was done to prevent the failing of the dams or an uncontrolled release of water over the top of the dam (Randy Cephus, 2017). The result was an extra input in water flow through the neighbourhoods downstream from the reservoirs.

2.6. Sentinel-1

The SAR images used for this research are acquired by the Sentinel-1 mission. The Sentinel-1 mission is the first of several Sentinel missions currently operational or under development by The European Space Agency (ESA) (Showstack, 2014). The Sentinel missions are part of the Copernicus program of the European Union run by the European Commission in partnership with ESA, the EU Member States and EU Agencies (Showstack, 2014). The Sentinel-1 mission is designed for continuous radar mapping of the Earth's surface. This improves coverage and revisit frequency of available satellite radar data all around the world. The mission constellation consists of two satellites, Sentinel-1A and 1B, the first being launched on April 3rd 2014 and the second one on April 25th 2016 (Torresand *et al.*, 2012). The satellites fly in a near-polar sun-synchronous, 12 day repeat cycle at an altitude of 693km (Attema *et al.*, 2007). The constellation uses a C-band radar that transmits at 5.405 GHz which translates to a wavelength of 5.54 cm. It operates continuously in all weather and during day and night. The instrument is able to operate in dual polarisation (HH+HV or VV+VH), it does so by transmitting either horizontally or vertically and receiving horizontally and vertically simultaneously (Torres *et al.*, 2012). The lifetime of both satellites is 7 years but they carry fuel for 12 years (Attema *et al.*, 2007). The images made by the Sentinel-1 mission and all the other Copernicus data services have a worldwide coverage and are available free of charge to all users.
The Sentinel-1 data is available in two formats:
Level-1 Single Look Complex (SLC) products consist of focused SAR data geo-referenced using orbit and altitude data from the satellite and provided in zero-Doppler slant-range geometry. The products include a single look in each dimension using the full transmit signal bandwidth and consist of complex samples preserving the phase information.

Level-1 Ground Range Detected (GRD) products consist of focused SAR data that has been detected, multi-looked and projected to ground range using an Earth ellipsoid model. Phase information is lost. The resulting product has approximately square spatial resolution pixels and square pixel spacing with reduced speckle at the cost of worse spatial resolution.

2.7. Data description

This research utilises several data sets and each will be described in this section.

2.7.1. SAR images

The main data is the stack of SAR images. These are produced by the Sentinel-1 mission and as described in section 2.6 are freely available online. The extent of the SAR images is displayed in figure 2.11. The stack that was used consisted of 93 images, the first image was acquired on the 18th of July 2015 and the last on the 17th of September 2017. This gives us 89 images before the flood, one image during the flood on the 30th of august 2017 and three images after the flood. All images were taken in descending mode, VV polarisation, in swath 1.

2.7.2. Digital Elevation Models (DEMs)

For this research two DEMs are used.
The first-one is a DEM produced by the Shuttle Radar Topography Mission (SRTM) of NASA. This mission had the goal to produce a Global Digital Elevation Map (GDEM) using space borne radar (Nikolakopoulos

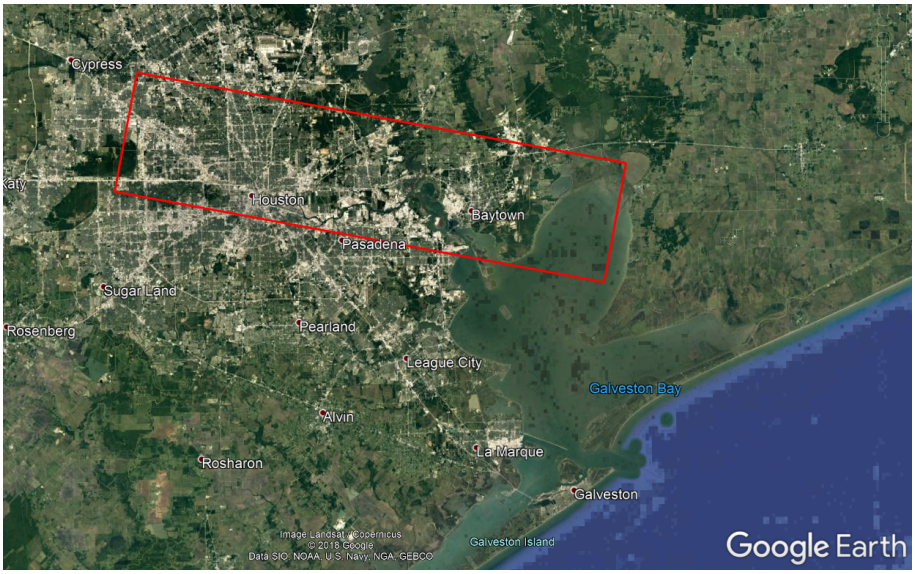


Figure 2.11: Outline of the area covered by the stack of SAR images that is used (in red) on top of a Google earth map, the red shape is approximately 21km by 86km.

et al., 2013). The result was a DEM covering 80% of the earths land surface, in a resolution of 1 arc second (approximately 30m) (Czubski *et al.*, 2013). The second is part of the National Elevation Data set (NED). The NED provides the best available public domain raster elevation data of the United States. The NED is derived from diverse source data, processed to a common coordinate system and unit of vertical measure. The resolution of the NED is 1/3 arc second (approximately 10 meters) and in limited areas it is 1/9 arc second (approximately 3 meters) (Gesch, 2007).

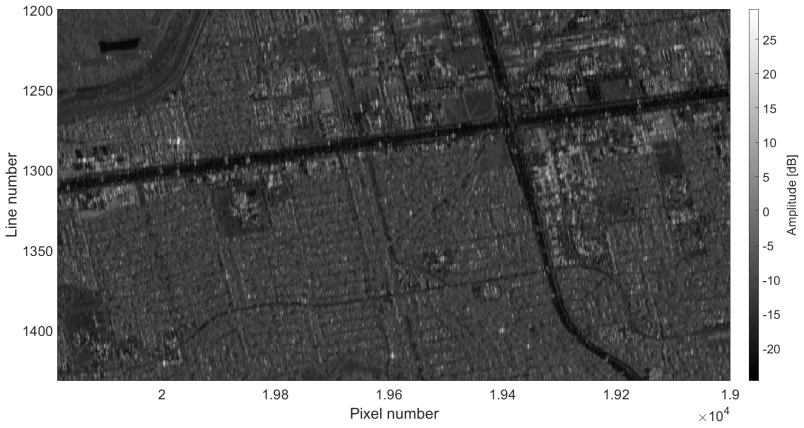
2.7.3. Optical images

In order to validate the results optical airborne imagery is used. This data was acquired by the National Oceanic and Atmospheric Administration (NOAA) remote sensing division. They gathered airborne images every day from August 27th till the 3rd of September to support the NOAA, homeland security and emergency response units. The resolution is approximately 50cm[4].

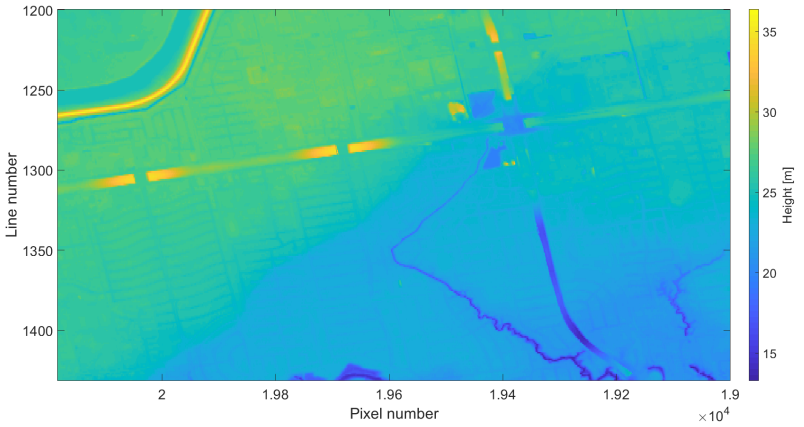
2.8. Study Area

In figure 2.11 the outline of the SAR data extent is given. Because this area is rather large and testing would take a long computational time the development of the method focuses on a smaller subset of the SAR image. In figure 2.12(a) the subset is displayed in the form of a SAR image. This area is chosen because it contains a river, large open highways, a park, roads under dense tree cover and of course flooded areas. In the extent of the image several floods of roads and grasslands can be spotted in the aerial image (Figure 2.12(c)). The outline of these flooded neighbourhoods and parks are displayed on top of the radar-coded DEM in figure 2.12(b). In this DEM we can distinguish some key features, there is an overall gradient present that is lowest in the bottom right corner and highest in the upper left corner, a large highway runs through the image left to right and intersects a smaller but main road running from top to bottom with a small bend at the bottom. Furthermore there is a river running along the lower right side of the image and there is a levee/dike in the upper left corner.

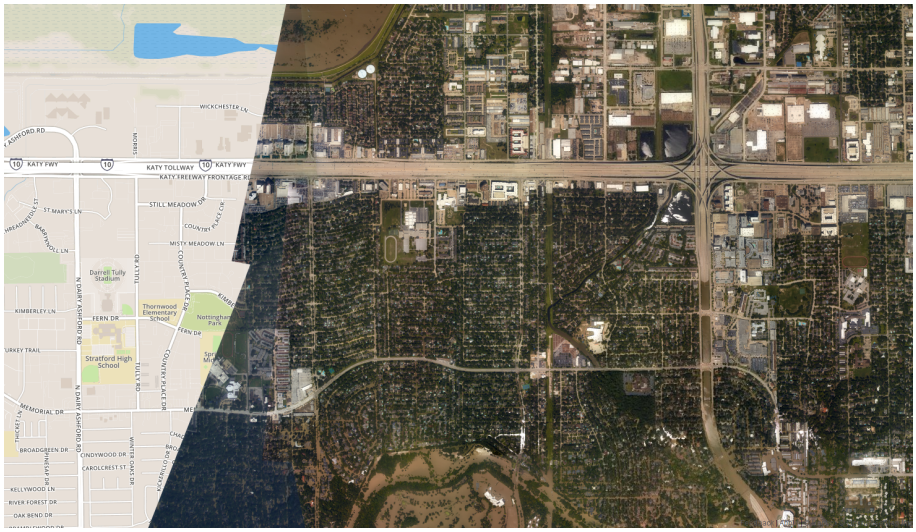
[4]https://storms.ngs.noaa.gov/storms/harvey/index.html



(a) Mean of 89 SAR images over the period of 2015-07-18 till 2017-08-2. It shows the region on which the methods for flood mapping are tested and further developed. The region is approximately 5km by 3km in size.



(b) High resolution DEM freely available for the testing region.



(c) Aerial image of the same region acquired on the 30th of August 2017. Image comes from NOAA Hurricane Harvey imagery site

Figure 2.12: Tree different images made of 3 different data sets all showing the region on which testing for the research was performed.

3

Methodology

In Section 2.3 a brief overview is given on past studies concerning flood mapping using SAR. From these past studies a few methods are selected and in some cases expanded for this research to implement and evaluate. This led to a list of five methods. The methods are selected because they are based on SAR images and process the entire image at once instead of small portions of the image. The methods will be tested on accuracy in an urban setting in Chapter 4.
In this chapter the methodology of the five flood mapping methods used in this research are described. An extensive overview of the processing steps for each method will be given. The five methods are:

1. Single image threshold
2. Image pair threshold
3. Stack of images outliers
4. Stack of images
5. Stack of images plus a DEM

The first two methods come directly from a literature study on flood mapping. The Single image threshold method can be used to, in most cases correctly, classify flooded and non-flooded pixels and used to detect an inundation area (Martinis *et al.*, 2009). This study will use the method with threshold values from literature that were used on rural areas and apply them on an urban setting. Next to the single image an image pair is what most flood mapping studies are generally based on (Schlaffer *et al.*, 2014). The Image pair threshold method used in this study looks at the difference between two consecutive images and applies a threshold on these differences for its classification. The last tree methods use a stack of images in an effort to reduce speckle and create an ”expected” value for each individual pixel. This idea comes from the study performed by Mason *et al.* (2013). Their study used a simulation model to calculate an expected intensity measurement using a high resolution DEM. To bypass the need of this simulated values the last tree methods will use a stack of images to calculate an ”expected value” for each pixel. The Stack of images outliers method will use the maximal and minimal values recorded in the stack under non-flooded conditions to classify pixels as flooded or non-flooded. The Stack of images method will use the mean of the stack to asses whether or not a pixel is flooded. The final method is the Stack of images plus DEM method. It continuous on the Stack of images method with the use of a DEM in order to perform a region growing process to classify pixels as flooded and non-flooded.
These methods, with the exception of the last method, were selected for the fact that they are based solely on SAR amplitude data, are relatively computationally inexpensive and do not require models or other data. This is why a flood map can be computed with these methods the instant a new SAR image is available.
This chapter is limited to only explaining the methodologies. Every method will be applied on the same test case and the results will be displayed in Chapter 4. In Chapter 4 the methods will be further developed based the first results. Chapter 4 will also compare the results from each method with one

another. In order to be able to compare the results the testing area is divided into neighbourhoods that are not flooded and neighbourhoods that are flooded using validation data in the form of airborne optical images. In an effort to quantify the results besides the images a performance metric is introduced.

3.1. Single image threshold

Using a single image that was taken during the flood, a flood map could be created when flooded pixels have a signature on a SAR backscatter measurement image distinguishable from other non-flooded pixels. This is done with two thresholds, the threshold marks the interval of values that are characteristic for flooded pixels. Selecting a threshold is a process on which much time is often spent when making a classification. This is why optimised threshold have been determined in a study performed by Manjusree *et al.* (2012). They found that for VV polarisation the optimal threshold for flood water was between -6 to -15 dB when using higher incidence angle SAR images (20 to 49 degrees). Sentinel-1 incidence angles fall inside this range (Torres *et al.*, 2012) but was not the satellite used in their research, they used data from the RADARSAT-2 satellite. The conclusion of Manjusree *et al.* (2012) was a general one and not limited to RADARSAT-2, that is why the Single image threshold method will start with the threshold from their paper. For the Single image threshold method every pixel that contains a value between -6 and -15 dB will be classified as flooded. When in the first results the thresholds do not seem to be adequate they can be adjusted using first results and histograms from these results. The Single image threshold method is summarised by

$$P = \begin{cases} \text{Flooded}, & \text{if } P \geq \text{-6 dB and} \leq \text{-15 dB} \\ \text{Non-flooded}, & \text{otherwise} \end{cases} \tag{3.1}$$

in which the pixel values in the SAR image a re P. The process thus consists of a single step:

1. Apply thresholds on the SAR image taken during the flood, this creates a binary flood map consisting of flooded and non-flooded pixels.

The advantage of this method is that it only requires one image and that its suitable for rapid flood mapping due to being computationally relatively inexpensive (Martinis *et al.*, 2009). This method can be expanded by adding an image from before the flood and instead of classifying based on the flooded pixels signatures classifying based on the change in individual pixels, as is explained in the following method.

3.2. Image pair threshold

The Image pair threshold method compares an image acquired before the flood with the image acquired during the flood. Flooded pixels are expected to show a difference in amplitude measurement value than when they were not flooded, resulting (in theory) in a larger difference than between a pixel that is not flooded in either image.
The image taken under normal conditions (P_{n}) is subtracted from the image taken during the flood (P_{f}) and gives a difference image (P_{d}), written out this gives

$$P_{\text{d}} = P_{\text{f}} - P_{\text{n}}. \tag{3.2}$$

On the resulting difference image (P_d) a threshold will be applied classifying all pixels according to the threshold into a flooded and not flooded category. The expectation is that flooded pixels have a larger value in the difference image than non-flooded image, these values can however be both positive and negative as illustrated in Figure 2.8. In this figure it is illustrated that a smooth water surface acts as a specular reflector leading to a lower backscatter measurement at the satellite compared to rough ground surface, but it also shows that when the water surface is located next to buildings double bounces can occur leading to a higher measurement. This leads to a threshold implication

$$P_d = \begin{cases} \text{Flooded}, & \text{if } P_{\text{d}} \geq T_{\text{U}} \\ \text{Flooded}, & \text{if } P_{\text{d}} \leq T_{\text{L}} \\ \text{Non-flooded}, & \text{otherwise} \end{cases} \tag{3.3}$$

in which T_L and T_U are the thresholds.

These thresholds will be determined by accessing the image and histogram of P_d. Based on validation data an optimum for T_U and T_L could be chosen so that false positive classifications are minimised while maximising the flooded pixels classified as such. These values for T_U and T_L can than be applied on areas without validation data. The process consists of two steps:

1. Compute the difference image by subtracting the SAR image taken before the flood from the SAR image taken during the flood.
2. Apply thresholds on the difference image, this creates a binary flood map consisting of flooded and non-flooded pixels.

Both the Single image threshold as well as the Image pair threshold methods are limited in the number of data points that they use, using a larger set of SAR data acquired before the flood event could prove to increase accuracy of the flood mapping. The following methods use a stack of images for their classification.

3.3. Stack of images outliers

The Stack of images outliers method is based on the assumption that a flood will change the reflective properties of the pixels it occurs on. As seen in Figure 2.8, the signal is expected to increase or decrease with respect to a signal under normal conditions. This method classifies each pixel individually. The pixel value on the flooded image (P) will be compared to the maximum and minimum value recorded for that pixel in a stack of images (S) (of non-flooded images). When it exceeds one of these values it is classified as flooded, otherwise is it denoted as non-flooded. The classification is expressed as

$$P = \begin{cases} \text{Flooded}, & \text{if } P \geq \max(S) \\ \text{Flooded}, & \text{if } P \leq \min(S) \\ \text{Non-flooded}, & \text{otherwise.} \end{cases} \tag{3.4}$$

This method is different from the others because it works with different thresholds for each pixel. This can be an advantage because of all the different scattering influences there are, as can be seen in Figure 2.8. A general threshold for the entire image has the trade of of missing flooded pixels or having false positives based on the threshold. Taking a different threshold for each pixel might mitigate this effect.

Eq. 3.4 consist of two step. The first step is determining the largest and smallest value for each pixel in the stack. The next step is comparing the pixel values of the image taken on the day of the flood to the largest and smallest values for that pixel, classifying it as flooded when it lies outside of these boundaries and non-flooded if it lies outside of the boundaries. The processing steps of the method are given below:

1. Define a stack of images composed of images with no flood present. A larger stack can represents a wider range of values that can be considered normal for a single pixel but a stack over a long time period can be influenced by new buildings constructed or old buildings removed.
2. Apply Eq. 3.4 on the stack and the flooded image. This creates a binary flood map consisting of flooded and non-flooded pixels

Speckle in the stack and the image of the flood could potentially lead to flooded pixels not being detected by this method. This is because the thresholds sets can be influenced a lot by one measurement where the speckle leads to a very large or a very small value. When there is a very large or small value in the stack the pixel in the flooded image might not exceed the value even when its scattering properties are changed due to the flood, leading to a misclassification of the pixel. Both the Single image threshold as well as the Image pair threshold methods are influenced by speckle. In an effort to reduce the effects of speckle on the classification process the mean of a stack of images could be used. The mean value over several images could be used to represent pixel values before the flood. This reduces the effect of speckle in the value representing the normal state of pixels although speckle is still present in the flooded image. The following methods exploit this theory.

3.4. Stack of images

This method uses a stack of images before the flood under non-flooded conditions to compute an expected measurement value for each pixel. This makes it possible to quantify how much a pixel in an image taken during a flood has changed compared to their expected measured value. The expectation is that by using more than one image before the flood the speckle can be reduced in the value for the normal conditions. The method is based on the research done by Mason *et al.* (2013). Their method uses a model to predict backscatter strengths and compares the predicted values with the measured ones to distinguish flooded and non-flooded pixels. The method proposed in this section replaces the modelled values with the mean of previous images taken. By doing so there is no need for a computationally expensive model. Furthermore, using a stack of images makes it possible to account for the variability that each pixel has under normal conditions. The variability can be expressed by the standard deviation of a pixel under normal conditions in the stack of images. The variability is visualised in Figure 4.11.

The change in every pixel (P_c) will be calculated using the mean amplitude of that pixel over the stack in normal conditions ($\overline{P}$), the variability during that time expressed by the standard deviation (σ_P) and the amplitude on the flooded image (P_f). The calculation of the change in every pixel (P_c) is expressed by

$$P_c = \frac{\overline{P} - P_f}{\sigma_P}. \tag{3.5}$$

In Eq. 3.5 σ_P represents the standard deviation of a pixel or in other words the reliability of the "normal" value of each pixel in the time stack. When the standard deviation is small the pixel is consistent trough time where-as a large standard deviation indicates that the pixel varies a lot under normal conditions and that the "normal" is less reliable. This enables us to compute the difference between the image taken during the flood and the "normal" value for each pixel. The larger the difference the more likely a pixel in the image of the flood has changed. The resulting image of change in every pixel (P_c) will then undergo a classification based on threshold values T_U and T_L. This results in equation

$$P_c = \begin{cases} \text{Flooded}, & \text{if } P_c \geq x \\ \text{Flooded}, & \text{if } P_c \leq y \\ \text{Non-flooded}, & \text{otherwise.} \end{cases} \tag{3.6}$$

The threshold values x and y will be determined using the validation data and the histogram of the image P_c, the values of x and y will be chosen in such a way that a maximum amount of flooded pixels will be classified as flooded while minimising the amount of non-flooded pixels classified as flooded. The selected thresholds will then be tested on other regions to check if they give the desired result and can be implemented on entire images. The processing steps are given below in a step by step workflow:

1. Define a stack, this step consist of setting up the variables, The number of images is defined with the last image being taken during the flood, a longer time period has the risk of being influenced by a changing environment on the ground but a short period filters out less speckle in the mean value.

2. Compute change, this step is the implementation of Equation 3.5 on the stack of images and results in an image where large values (positive and negative) represent pixels that are likely to have changed and could be designated as flooded.

3. Apply thresholds on the change image, this creates a binary flood map consisting of flooded and non-flooded pixels.

Because there is still speckle present in the image taken during the flood, the pixels in the flooded image can deviate from their expected value which might lead to non-flooded pixels being classified as flooded. The next method uses only the highest values in P_c to minimise false positives caused by speckle and uses a DEM to classify the rest of the image.

3.5. Stack of images plus a DEM

The Stack of images plus a DEM method uses seed points derived from a SAR image to determine larger flooded areas on a DEM. The method uses the image of P_c that is computed with the Stack of images method to select pixels which are most likely to be flooded. These pixels are the seed points on a DEM that classify pixels neighbouring them and having a lower elevation as flooded. The process of selecting neighbouring pixels from a seed point is called region growing. Using region growing overcomes the problem of some flooded streets going undetected due to their orientation with respect to the satellite track. This is because a pixel does not need to have a large change to be classified as flooded, if it is next to a pixel that has a high probability to be flooded and is situated on a lower elevation it can still be classified as flooded.
The first step of this method is to compute the change of every pixel (P_c) as defined by Eq. 3.5. While the Stack of images method then uses thresholds to classify every pixel, this method will implement thresholds only to select pixels with the highest probability to be flooded and use regions growing to classify the remaining pixels. The second step is thus to apply a threshold to select pixels most likely to be flooded or in other words the pixels with the highest and lowest values. By doings so the false positive flooded pixels are minimised or even non-existent. Normally, the drawback of implementing very strict thresholds is that a large part of flooded pixels are not detected. This method does not suffer from that drawback as the classification step does not use a threshold. From every pixel that is almost certainly flooded the coordinates are stored. These coordinates are used as seed points on a DEM. Every seed point has an elevation value and 4 neighbouring pixels, with the exception of the 4 corner pixels which only have 2 neighbouring pixels. The elevation value of each neighbour is compared to the elevation of the seed point. When it is a lower elevation the pixel gets classified as flooded, when it is larger it will be classified as non-flooded. When a pixel is classified as flooded its neighbouring pixels get tested in the same way using the elevation value of the original seed point. This process continuous until every pixel that is classified as flooded around the seed point are enclosed by pixels with a higher elevation value than the elevation of the original seed point. The processing steps are given again below in a step by step workflow:

1. Define a stack, this step consist of setting up the variables, The number of images is defined with the last image being taken during the flood, a longer time period has the risk of being influenced by a changing environment on the ground but a short period filters out less speckle in the mean value.
2. Compute change, this step is the implementation of Equation 3.5 on the stack of images and results in an image where large values (positive and negative) represent pixels that are likely to have changed and could be designated as flooded.
3. Filtering, in order to minimise false positives a filtering step can be applied, the step is optional. This filter computes the mean of surrounding pixels for each pixel. This filter ensures that pixels classified as flooded are in the centre of several flooded pixels and therefore have a higher probability to be flooded than a pixel that has a height value but all their neighbouring pixels have a small value. The drawback is the loss of resolution and that single flooded pixels could be missed. (side note, flooded pixels can have positive and negative values, averaging them could lead to a small value and missing them as flooded)
4. Pixel selection, a threshold is applied on the filtered image selecting all pixels above a certain value and below a certain value. From these pixels the coordinates are stored.
5. Region growing, the region growing step uses the selected pixels as seed points for the region growing process. Every pixels that neighbours a selected pixel and has a lower elevation than the selected pixel will be added to the region.
6. Region summation, this step simply consists of adding all the regions from the different seed points to one image

7. Classification, as the name suggests this is the classification step, to start of every pixel that belongs to a region could be classified as flooded, if this results in a large number of false positive classification a threshold could be applied. Implying that a pixel will only be flooded when it is part of a set number of regions.

For a comprehension of how the method computed the pseudo code is given next:

Inputs

-P, the stack of images.
-P_f, the last image of the stack being that of the flood
-$\overline{P}$, the mean of the stack without the flooded image
-σ_P, the std of the stack without the flooded image.
-DEM, the digital elevation map with the same extent as the SAR images in the stack.
-x and y, thresholds for the pixel selection step

Procedure

Compute P_c with $P_c = \frac{\overline{P} - P_f}{\sigma_P}$
Apply filter on P_c, every pixel is mean of surrounding pixels
Pixel selection according to $P_c \geq x$ and $P_c \leq y$
For loop over every selected pixel;
    Identify neighbouring pixels
        for loop over every neighbouring pixel
            if height pixel ≤height seed point
                add to region
                Identify neighbouring pixels
            else
                do not add to region
Add all regions together
If pixel is inside region
    flooded
else
    non-flooded

3.6. Performance metric

In Chapter 4 all methods are tested and their results briefly discussed, in order to compare the results of of the methods to one another a performance metric will need to be defined. A confusion matrix is a typical tool to qualify the results of a classification, the confusion matrix is build up of rows representing the instances in a predicted class and of columns represents the instances in an actual class. For this research the confusion matrix of each method would consist out of two rows and two columns representing the classes flooded and non-flooded, the confusion matrix requires knowledge about the true class of pixels for this research pixels are only known to be inside a flooded neighbourhood or outside. Pixels inside those neighbourhoods do not necessarily mean that they are flooded as the areas contain houses and other non-flooded objects. The performance metric for the purpose of this research will be based on a confusion matrix with the alteration that the true class is not known but only the true region a pixel lies in. This means that pixels in non-flooded regions are all non-flooded but not all pixels inside the flooded regions are necessarily flooded. An example of the confusion matrix used as a performance metric for this research is given as Table 3.1.

	Flooded neighbourhood	Dry neighbourhood
Classified flooded	Correct	False positive
Classified dry	False negative	Correct

Table 3.1: Example of confusion matrix for this research. The columns represent the true group pixels belong to, the rows represent the group pixels are classified as.

4

Results

In this chapter the results of the five flood classification methods are presented by implementing them on the same region of interest. The region on which are used for the tests is shown as an aerial image, a SAR image and a DEM in Figure 2.12. Besides the first results, there is a short discussion on the result and if the initial result could be improved, the results of these potential improvements will be shown as well. The resulting flood map will use the same layout for all methods: blue pixels are classified as flooded while yellow indicates non-flooded. On top of all flood maps red polygon outlines are displayed, these shapes are the outline of the flooded neighbourhoods in the extent of the image. They are derived from aerial images taken by NOAA on the 30th of August, the same date on which the SAR image was taken during the flood. When a classification would be perfect not all pixels inside the polygons would be classified as flooded, the polygons enclose buildings as well as streets that were not flooded or not completely flooded. A perfect flood map would however not contain any pixels classified as flooded outside of the polygons.

4.1. Single image threshold

The method that uses a single SAR image taking during the flood to make a flood map containing flooded and non-flooded pixels. The SAR image of the testing region taken during the flood is displayed in Figure 4.1 This method can be directly implemented as described in Section 3.1. The result is displayed in Figure 4.2.

	Flooded neighbourhood	Dry neighbourhood
Classified flooded	30548 (51 %)	111097 (52 %)
Classified dry	29592 (49 %)	103451 (48 %)

Table 4.1: Confusion matrix for single image classification result using the threshold values of -15 dB and -6 dB. The first number represents the number of pixels and in the brackets the percentile of the total pixels in that neighbourhood. Each column thus adds up to a 100 percent.

The resulting image consists of 48 percent pixels that are classified as flooded. Only 14 percent of the pixels in the entire image lie in flooded neighbourhoods and only a fraction of that 14 percent are streets that are actually flooded. That means that this classification method yields more false positives than positive classifications. The confusion matrix is shown in Table 4.1. In this table it is again shown that there are more pixels classified as flooded outside of the flooded neighbourhoods than inside the neighbourhoods. What does stand out in this image are the main highways. They run through the image as a yellow lines of pixels classified as non-flooded where the rest of the image is an inconsistent mixture of flooded and non-flooded pixels. This is due to their smooth and consistent surface properties that result in a low backscatter that is similar for all pixels in the highway. The rest of the image is made up of a mixture of flooded and non-flooded pixels. There seems to be little difference between

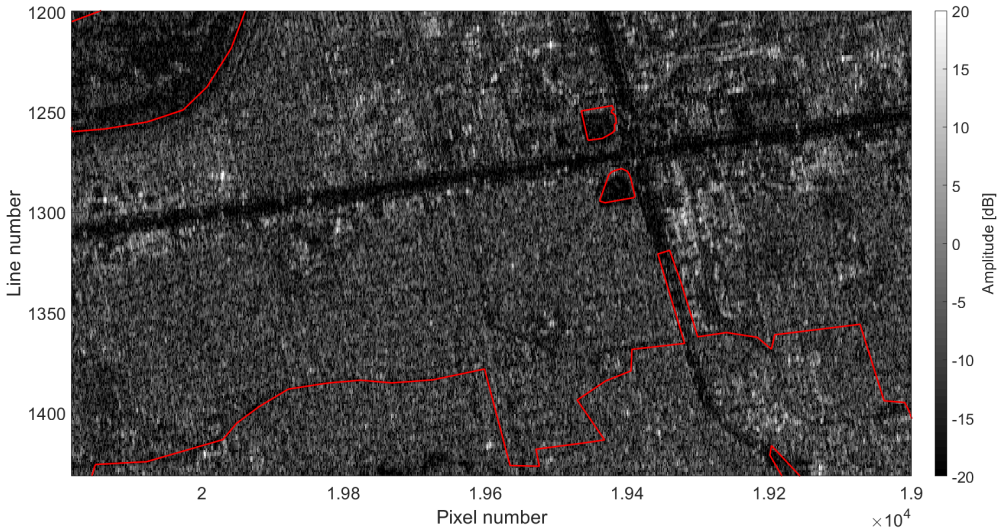


Figure 4.1: Showing a section of the Sentinel-1 SAR image taken on 30-08-2019 that covers the region on which testing is performed. When this image was acquired flooding was present in the area. The red polygons encircle the flooded neighbourhood, outside of these polygons there is no flood water present on the aerial images.

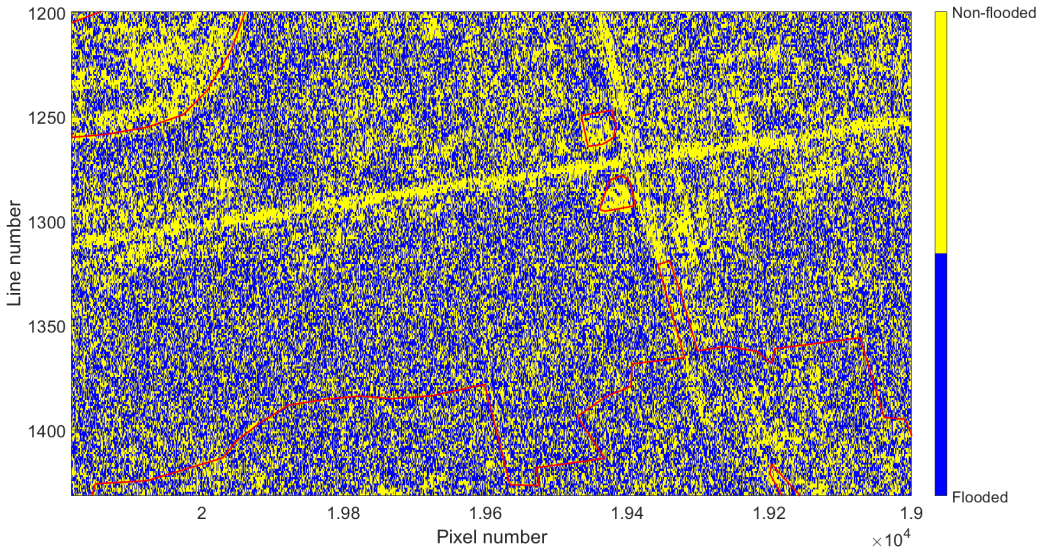


Figure 4.2: Result of the Single image threshold classification method on a SAR backscatter image taken during the flood. All pixels between -15 dB and -6 dB are classified as flooded. The red polygons encircle the flooded neighbourhood, outside of these polygons there is no flood water present on the aerial images.

the inside and outside of the polygons.

Figure 4.3 gives an insight in why this classification does not work, it displays the histogram of the SAR image on which the classification is performed as well as the histogram of the pixels inside the polygons. From the histogram it is clear that there is no threshold that would not lead to more false positive than positive classifications. The flood classification as described in the study done by Manjusree *et al.* (2012), does not yield a useful result in an urban environment when directly implemented.

An option to improve the result is the application of a mean filter window. A mean filter window can reduce speckle noise in SAR images (S *et al.*, 2013), the noise reduction or smoothing of the image can

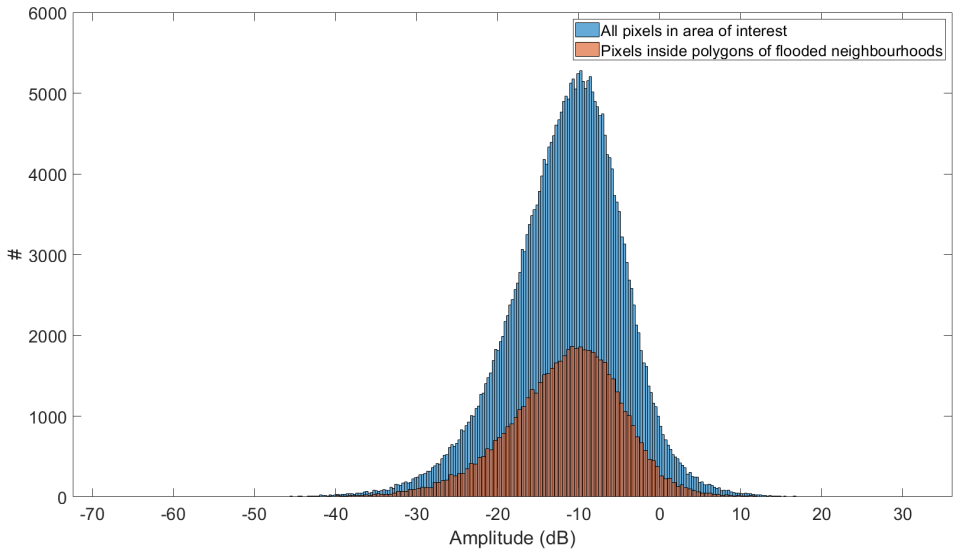


Figure 4.3: Histogram of pixel values in the Sentinel-1 SAR image shown in Figure 4.1. The blue histogram is of all pixels in the image while the orange histogram is only of the pixels that fall inside of the polygons encircling the flooded regions.

improve the quality possibly leading to a better distinction between flooded and non-flooded pixels. The filter is a linear one and is expressed by

$$h_{[i,j]} = \frac{1}{M} \sum_{(k,l)\in N} f_{[i,j]}. \tag{4.1}$$

In Eq. 4.1 M is the number of surrounding pixels N, $h_{[i,j]}$ and $f_{[i,j]}$ are the new and old pixel values. This research uses a mean 3x3 and 5x5 filter window to compute the mean value over a pixel and the eight or twenty-four pixels surrounding it. By doing so larger patches of flood could be better detectable while single outliers will not result in a false positive, drawback of the method is the loss of resolution and potentially missing floods in narrow streets that are only 1 pixel wide. The histogram of the filtered image is shown in Figure 4.4.

From this histogram it seems that the lowest values are mainly located inside of the flooded neighbourhoods, using a threshold of -22 to classify everything lower as flooded could lead to a better classification than in Figure 4.2. The result is shown in Figure 4.5.

	Flooded neighbourhood	Dry neighbourhood
Classified flooded	938 (2%)	1918 (1 %)
Classified dry	57795 (98%)	211209 (99 %)

Table 4.2: Confusion matrix for the classification result shown in Figure 4.5. In that image all pixels in the filtered SAR image lower than -22 dB are classified as flooded. The first number represents the number of pixels and in the brackets the percentile of the total pixels in that neighbourhood. Each column thus adds up to a 100 percent.

This classification lies more in line with expectations, it correctly selects pixels inside of the three polygons in the upper half of the image. These are wide open areas that are flooded, the result even selected some pixels in the lower polygon, water acts as an specular reflector here and results in a low measurement. In the image the main highway, which runs as a cross through the area, also contains a lot of pixels classified as flooded. This underlines the difficulty of flood mapping in urban areas using SAR imagery as the man made concrete results in a similar low backscatter measurement as still water does. Because the highway composes a relatively large part of the image the confusion matrix in Table 4.2 contains more false positive classifications than positive ones but knowing where these lie the result remains an improvement from the non-filtered image. Beside the highway pixels being classified as

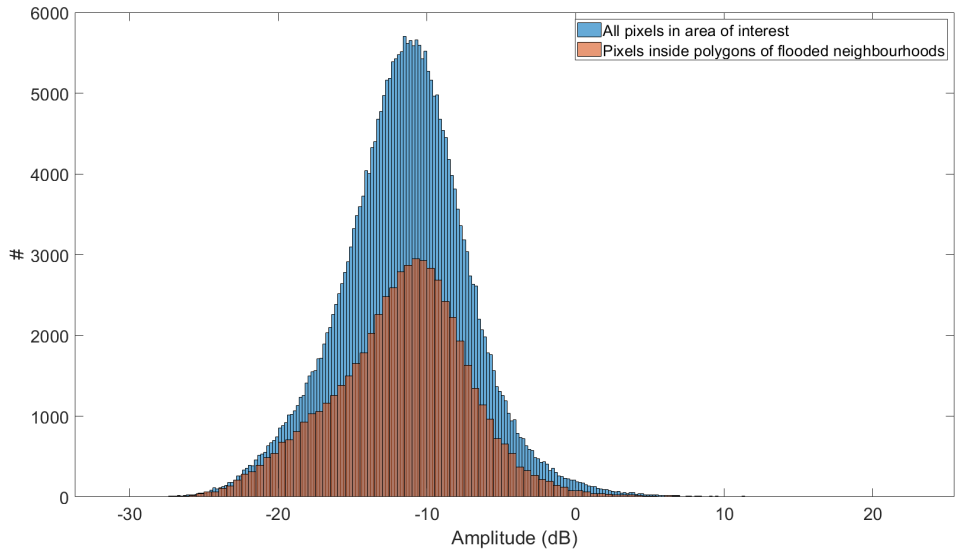


Figure 4.4: Histogram of pixel values in the Sentinel-1 SAR image of the flood after a 3x3 mean filter window was applied. The blue histogram is of all pixels in the image the orange histogram is only of the pixels that fall inside of the polygons encircling the flooded regions.

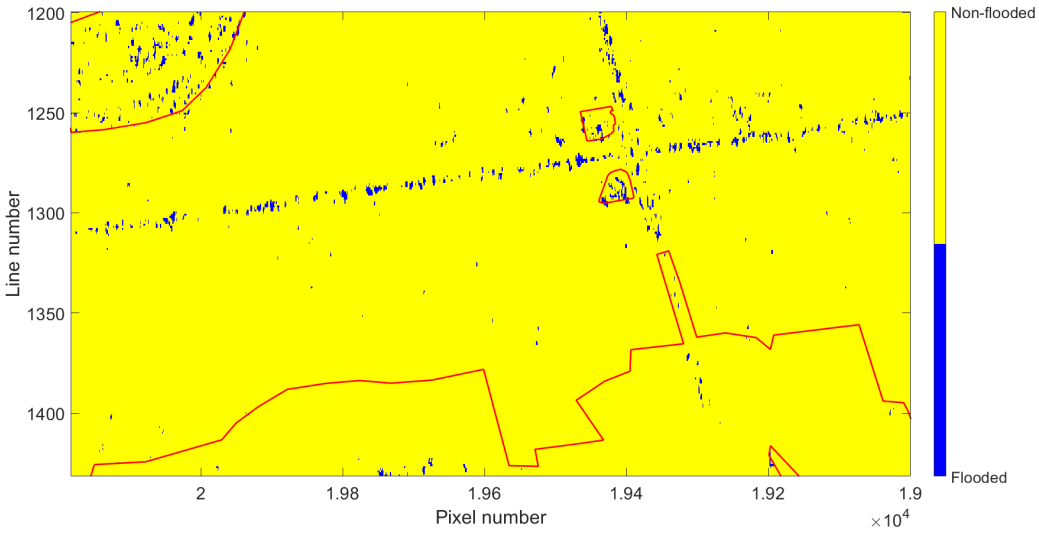


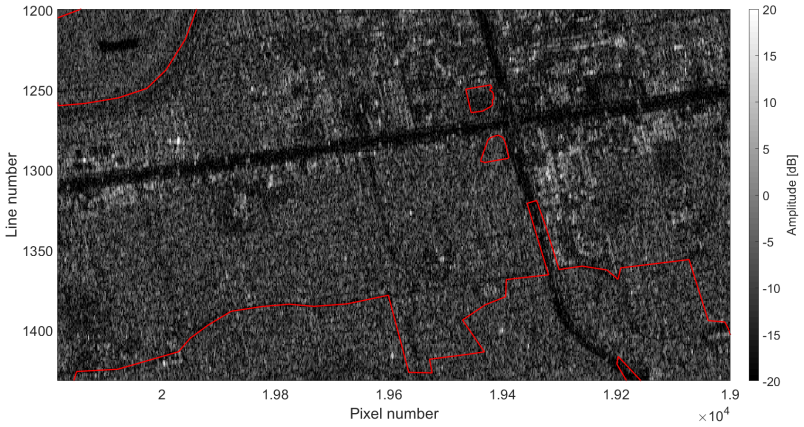
Figure 4.5: Result of the Single image threshold method classification applied on the Sentinel-1 SAR shown in Figure 4.1 after a 3x3 mean filter window was applied. All pixels bellow -22 dB are classified as flooded. The red polygons encircle the flooded neighbourhood, outside of these polygons there is no flood water present on the aerial images.

flooded the result approaches the result of the Copernicus emergency service which only classified the flooded field as flooded. The problem of the highway pixels being seen as flooded might be solved by using an Image pair and using the difference between two images rather than the values of an image itself.

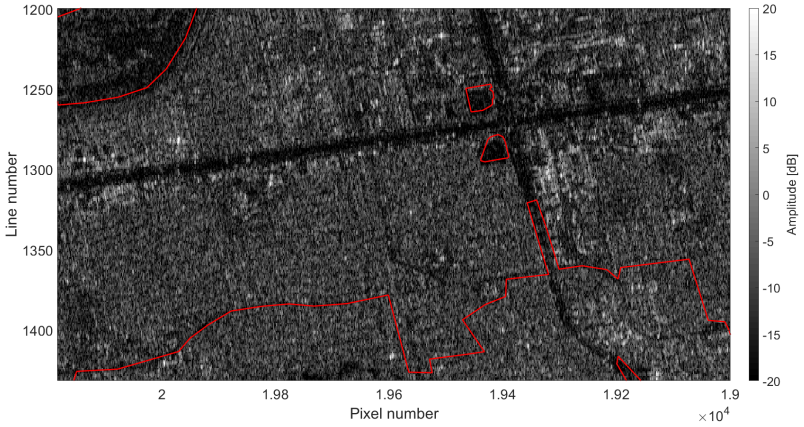
4.2. Image pair threshold

The method that uses the difference between a SAR image before the flood and one taken during the flood to produce a flood map containing flooded and non-flooded pixels. Computing the difference image is the first step of this method. The difference image is computed using Eq. 3.2. The images used are displayed in Figure 4.6 and the resulting image is displayed in Figure 4.7. In this image the

upper left corner stands out as it appears generally darker than the rest of the image, there is also a dark spot on the right side of the figure. These two darker areas are flooded. They are however also open fields. The bottom right appears generally lighter than the rest of the image. This is the area that contains flooded streets.



(a) Sentinel-1 SAR image taken on 24-08-2017 (before the flood).



(b) Sentinel-1 SAR image taken on 30-08-2017 (during the flood).

Figure 4.6: Sentinel-1 SAR images that are used in the Image pair threshold method. The red polygons encircle the flooded neighbourhood, outside of these polygons there is no flood water present on the aerial images.

The lighter shades are not the result of being the pixels that are the lightest or the only light pixels in the image. This becomes clear when looking at the histogram of the pixel values of the entire image and of the pixels inside the polygons in Figure 4.8. From this histogram no threshold seems to lead to more positive than false positive classifications. That is, any threshold to classify the image pixels as flooded would not work as it would classify more pixels outside of the polygon as flooded as well. The streets that are flooded are thus not directly distinguishable from dry streets in this figure based on pixel values. Because some patches do stand out in the difference image (Figure 4.7) a 3x3 mean filter window can be applied in an effort to isolate them. Just as the Single image threshold method used mean filter window to improve the result.

The 3x3 mean filter window takes takes the mean of the eight surrounding pixels as a value for that pixel. The filtering process leads to the histogram shown in Figure 4.9.

From this histogram it initially seems that any threshold applied on the image would result in a large percentages of the pixels classified as flooded would fall outside of the flooded neighbourhoods. But when looking at the highest percentiles on both sides of the histogram a threshold might work. From looking at the higher percentiles the values chosen for the classification are -15 and 15 dB difference.

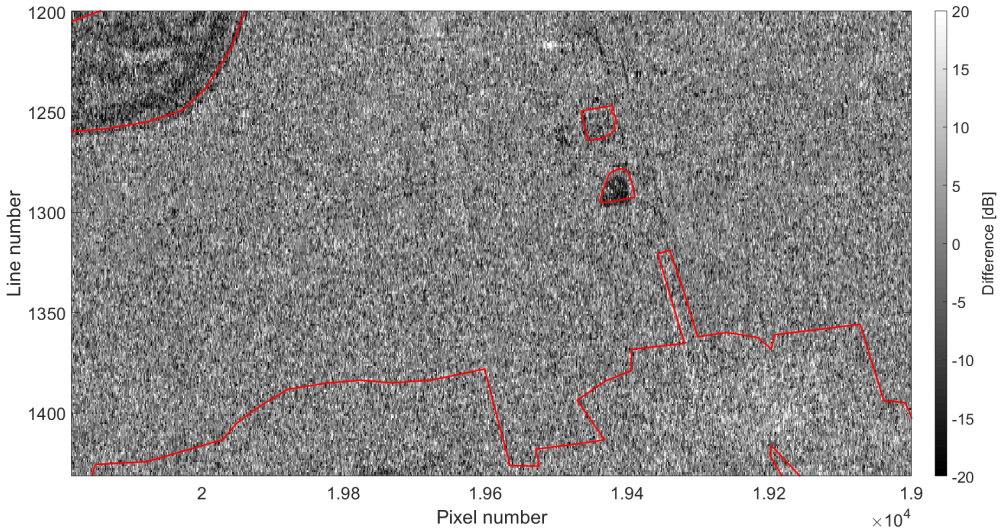


Figure 4.7: The difference image that is the result of subtracting Figures 4.6 (a) and (b) from one another. The red polygons encircle the flooded neighbourhood, outside of these polygons there is no flood water present on the aerial images.

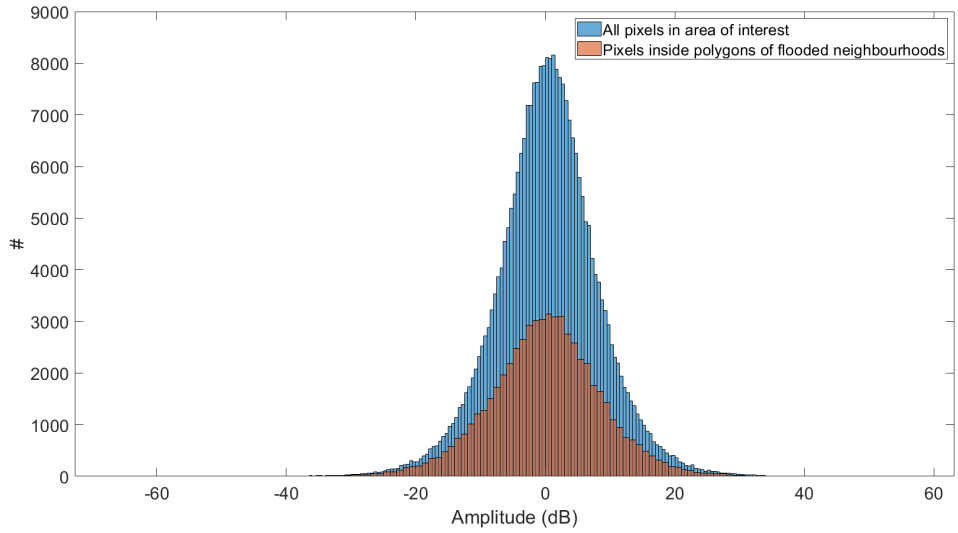


Figure 4.8: Histogram of pixel values in the difference image shown in Figure 4.7. The blue histogram is of all pixels in the image while the orange histogram is only of the pixels that fall inside of the polygons encircling the flooded regions.

The result of the classification attempt is shown in Figure 4.10.

	Flooded neighbourhood	Dry neighbourhood
Classified flooded	160 (0 %)	53 (0 %)
Classified dry	58573 (100 %)	213074 (100 %)

Table 4.3: Confusion matrix for image pair lower than -15 or higher than 15 (after the 3x3 mean filter window is applied). The first number represents the number of pixels and in the brackets the percentile of the total pixels in that neighbourhood. Each column thus adds up to a 100 percent.

In Figure 4.10 only a small number of pixels are classified as flooded, but most of them fall inside of the polygons on the top half of the image. In Table 4.3 it is visible that 75% of pixels classified

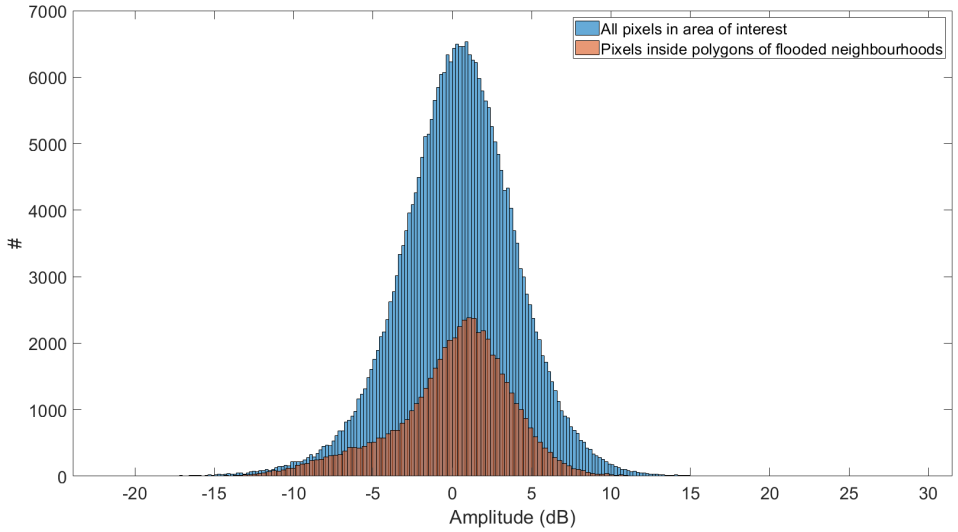


Figure 4.9: Histogram of pixel values in the difference image shown in Figure 4.7 after a 3x3 mean filter window has been applied. The blue histogram is of all pixels in the image while the orange histogram is only of the pixels that fall inside of the polygons encircling the flooded regions.

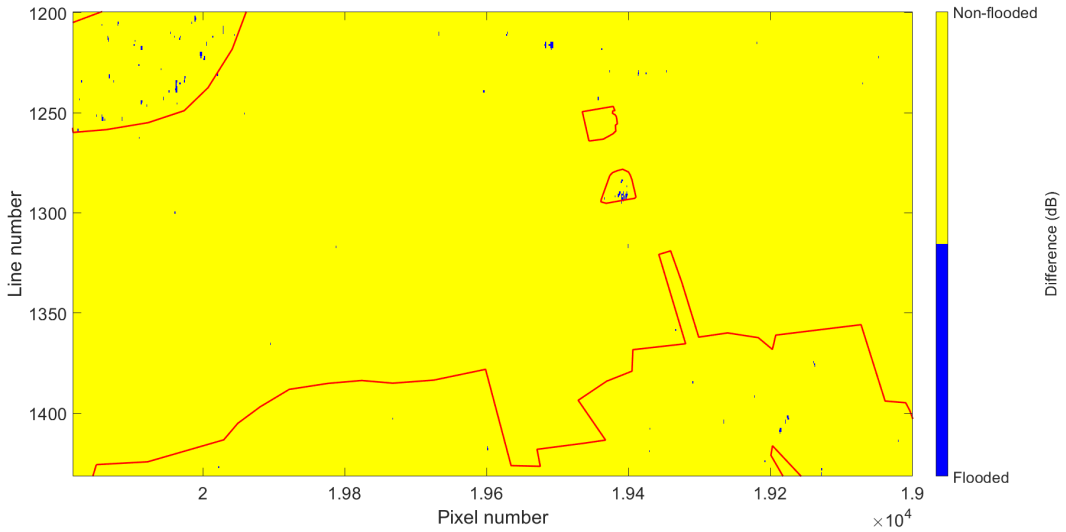


Figure 4.10: Result of the Image pair threshold classification method on the difference image shown in Figure 4.7 after a 3x3 mean filter window has been applied. All pixels smaller than -15 dB and larger than 15 dB are classified as flooded. The red polygons encircle the flooded neighbourhood, outside of these polygons there is no flood water present on the aerial images.

as flooded lie in a flooded neighbourhood. This is an improvement in terms of percentage compared to the result of the single image threshold 4.2, where the majority of pixels classified as flooded fell outside of the flooded areas. Although the number of mis-classifications seems to be low, so is the number of classifications of flooded pixels inside the polygon in the bottom of the image, where all the flooded streets are.

This result is an overall improvement over the result of the Single image threshold method (Figure 4.5) as the flooded pixels (albeit a lower number) are found in the top polygons of the image but more importantly the highway is no longer classified as flooded as it was with the previous method. To sum up, the Image pair threshold method (partially) spots open fields that are flooded but does not seem to be able to identify a flooded neighbourhood. This is nearly the same result the Copernicus emergency services produced for this area.

The problem thus still lies in the bottom polygon where the flooded streets lie. In order to understand why the approach of a difference image does not classify this part of the city correctly and misses some pixels in the flooded fields it is best to look at the behaviour of an individual pixel over time. In Figure 4.11 a pixel value over time is displayed. This pixel in taken from a highway that was not flooded on August the 30th 2017. The pixel value ranges from -14 dB to -40 dB seemingly at random.

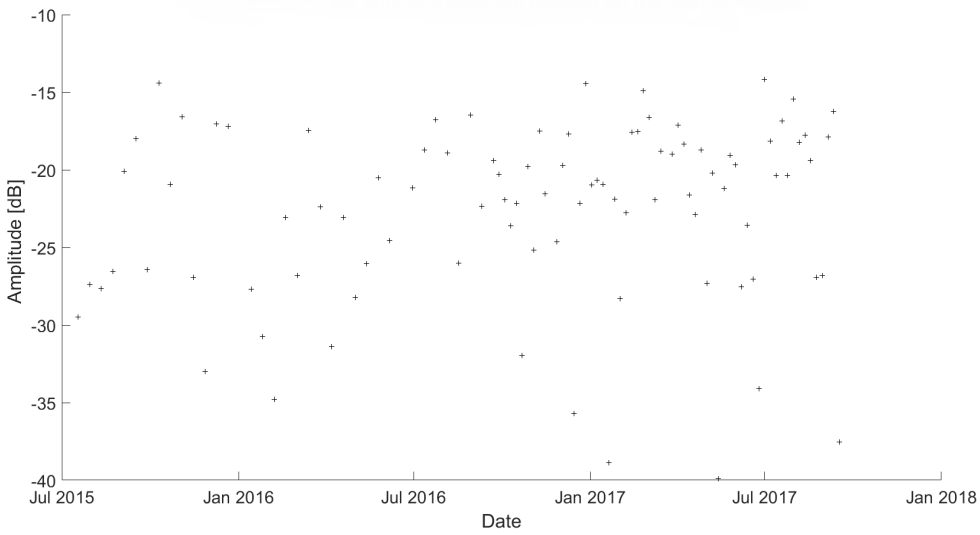


Figure 4.11: Amplitude backscatter behaviour of a single pixel that was not flooded on the 30th of August 2017 over time.

When two consecutive images are taken from this time series a difference of 15 dB would not be uncommon. In Figure 4.12 the pixel values over time are displayed from a pixel that was flooded on the 30th of August. The value on the image that was flooded is indicated by a red '*'. This value stands out as it is the highest value recorded in two and a half year.

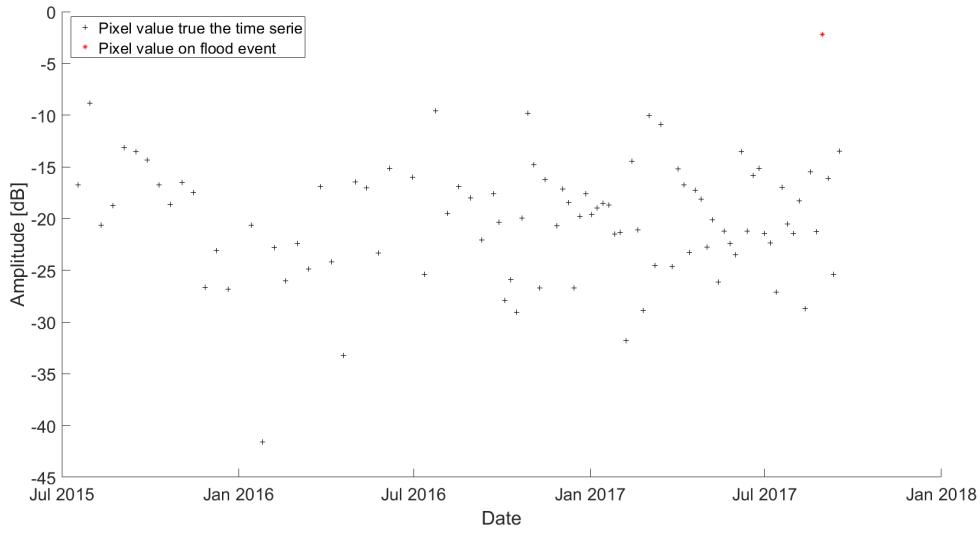


Figure 4.12: Amplitude backscatter behaviour of a single pixel that was flooded on the 30th of August 2017 over time.

It is however not the largest difference between two consecutive images. Because the pixel value has such a large range of values in non-flooded conditions the difference between two images can be large and small regardless of a flood occurring. From this single pixel it would again seem to be easy to classify flooded or not flooded based on a threshold of in this case -5 dB but with a look at Figure 4.3 it is again clear why this would not work. Taking all of this into account, using a difference image

will not yield a result in which dry streets can be distinguished from flooded streets. The method on this data can spot large open fields that are flooded but gives no satisfying result in urban areas.

4.3. Stack of images outliers

The method uses a stack of images to determine when a pixel value falls outside of the range of values that are expected of a pixel in a SAR image when it is not flooded. With the pixels that fall outside of the range a flood map is produced, the change is attributed to the flood and therefore pixels are clasified as flooded when they fall outside of their range of normal behaviour. Eq. 3.4 can be directly implemented on the data when the image on which a flooding could be present and the stack of images before the flooded image are defined. The resulting flood map is split up in two flood maps and shown in Figure 4.13. The values that were higher than ever recorded before are shown in (a). They are clearly present throughout the image but there is a cluster in the bottom right corner. In the flooded open fields there seems to be an absence of pixels classified as flooded. The smallest values ever recorded are shown part (b). The pixels classified as flooded in this image seem to be uniformly spread trough-out the image both in flooded neighbourhoods and not flooded ones.

At a first glance the minimum values seem to be useless as there is no difference between flooded and non-flooded neighbourhoods, the confusion matrix is shown in Table 4.5. The confusion matrix for the largest values recorded image is shown in Table 4.4, from this table the method does not seem use-full as the number of pixels classified as flooded is larger outside the flooded regions than inside. The cluster of pixels in the flooded neighbourhood is however the result that is desired, an effort to isolate this cluster by improving the method is consequently made. A possible solution is the implementation of a buffer.

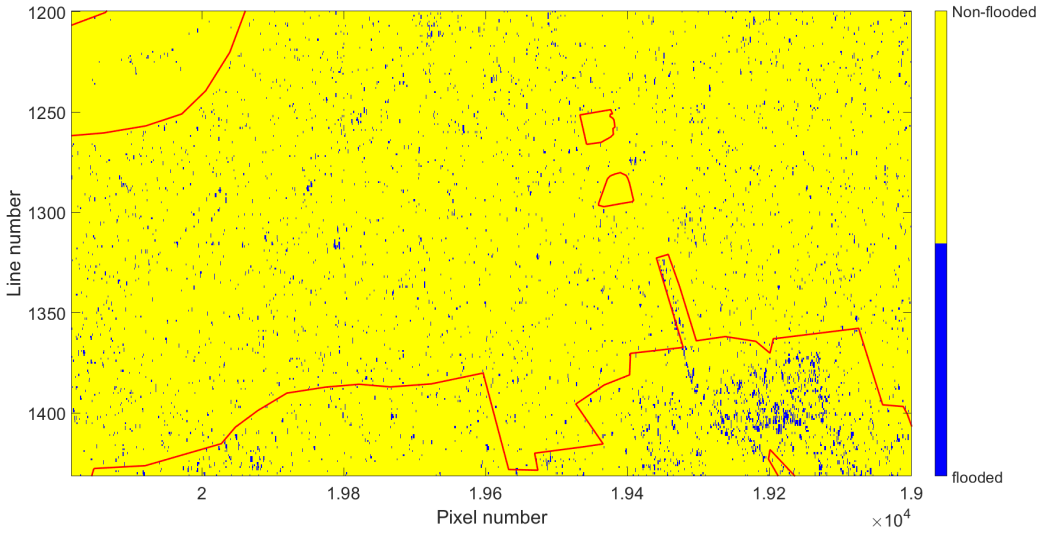
	Flooded neighbourhood	Dry neighbourhood
Classified flooded	2340 (4 %)	3563 (2 %)
Classified dry	57800 (96 %)	210985 (98 %)

Table 4.4: Confusion matrix for stack of images outliers larger than in the previous 2 years. The first number represents the number of pixels and in the brackets the percentile of the total pixels in that neighbourhood. Each column thus adds up to a 100 percent.

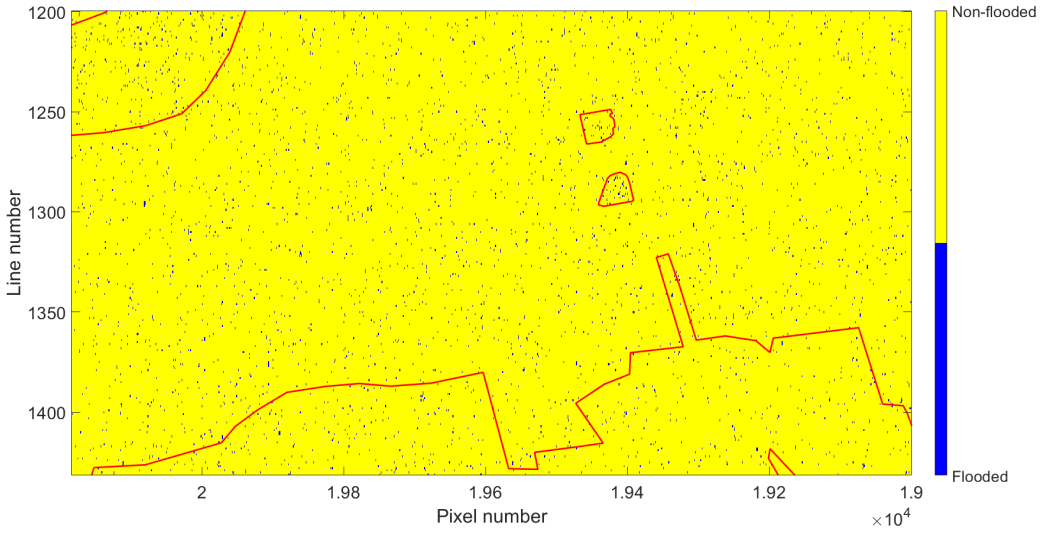
	Flooded neighbourhood	Dry neighbourhood
Classified flooded	859 (0 %)	2543 (1 %)
Classified dry	59281 (100 %)	212005 (99 %)

Table 4.5: Confusion matrix for stack of images outliers smaller than in the previous 2 years. The first number represents the number of pixels and in the brackets the percentile of the total pixels in that neighbourhood. Each column thus adds up to a 100 percent.

In Figure 4.12 the variability of a pixel under normal conditions is shown, along with the value of the pixel when flooded. This value stands out as the largest recorded in that time span. In this case classifying the largest value as flooded would be correct. There will however always be a value that is the largest of the stack, even for pixels that are not flooded. For the non-flooded pixels the largest value is in a random image in the stack which could accidentally be the image of the flood. When this value is present in an image taken during the flood the pixel will be classified as flooded, regardless whether or not it is flooded. This is why there seems to be a uniform spread of maximum and minimum values in Figure 4.13 in the non-flooded neighbourhoods. Its in the flooded neighbourhoods in the maximum values that patterns emerge. The flooded fields in the top halve contain very little of these pixels and are almost uniformly classified as non-flooded. This could be because they are used as reservoirs during the year and might be under water several times a year and therefore in several of the SAR images in the stack. The pixel values thus do not exceed the pixel values in the stack and are not detected. The cluster in the bottom also stands out. The fact that it only stands out on the



(a) Values on flooded image higher than in the stack of images.



(b) Values on flooded image lower than in the stack of images.

Figure 4.13: Result of the Stack of images outliers classification method on the stack of images consisting of 90 images spanning two years. All pixels on the SAR image taken during the flood that are larger or smaller than in the entire stack are classified as flooded. The top image are all pixels higher that in the stack, the bottom image shows all pixels lower than in the stack. The red polygons encircle the flooded neighbourhood, outside of these polygons there is no flood water present on the aerial images.

maximum image in Figure 4.13 suggest that the signal has increased strength due to a double bounce. This is the first method that makes this area of flooded urban streets stand out. In order an effort to remove the false positive classifications but keep this section classified as flooded an extra buffer could be implemented. The buffer means that not all the largest and smallest values are classified as flooded but only the values that are the largest and smallest by a margin are. The implementation of the margin translates Eq. 3.4 to

$$P = \begin{cases} \text{Flooded}, & \text{if } P \geq \max(S) + T_{\text{up}} \\ \text{Flooded}, & \text{if } P \leq \min(S) - T_{\text{low}} \\ \text{Non-flooded}, & \text{otherwise.} \end{cases} \tag{4.2}$$

By increasing T_{up} in Eq. 4.2 fewer pixels outside the polygons are marked as flooded while pixels

inside the polygons continue to be classified as flooded. This leads to a better result with the drawback that only a small amount of pixels are classified as flooded, leading to an increase in false negative classifications. The result of the using $T_{up} = 10$dB in Eq. 4.2 is shown in Figure 4.14 where only the maximum values with a buffer of 10 dB are classified as flooded. The result holds no flooded pixels in the open fields and only a few flooded pixels in the flooded neighbourhood, it does however also hold no flooded classifications outside flooded neighbourhoods. This results in the confusion matrix in Table 4.6, the best in ratio flooded inside polygon versus flooded outside polygon thus far. This means that the method is the first one to correctly classify flooded pixels inside the flooded neighbourhood without classifying non-flooded pixels as flooded. The image correctly shows some flooding where the Copernicus emergency services map did not.

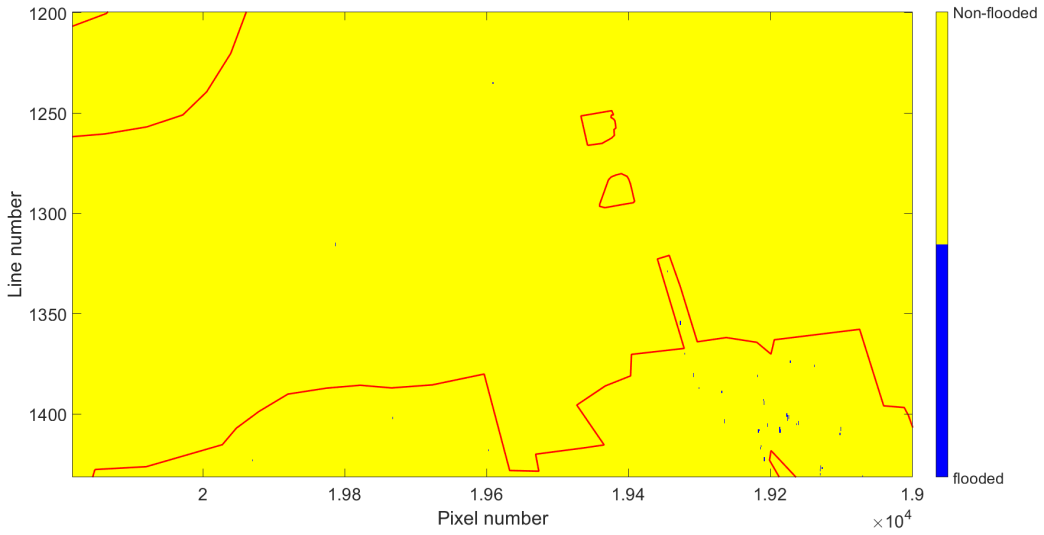


Figure 4.14: Result of the Stack of images outliers classification method (only the largest values) with a value of 10 dB for T_{up}. All pixels on the SAR image taken during the flood that are larger by the margin of at least 10 dB than in the entire stack are classified as flooded. The red polygons encircle the flooded neighbourhood, outside of these polygons there is no flood water present on the aerial images.

	Flooded neighbourhood	Dry neighbourhood
Classified flooded	55	3
Classified dry	60085	214545

Table 4.6: Confusion matrix for stack of images outliers larger with a minimum of 10 dB than in the previous 2 years. The first number represents the number of pixels and in the brackets the percentile of the total pixels in that neighbourhood. Each column thus adds up to a 100 percent.

The minimum values behave differently, increasing T_{low} leads to fewer pixels classified as flooded but eventually almost no pixels inside the polygons are classified as flooded while there are still pixels classified as flooded outside of the polygons. An example of this is when T_{low} is 15, the result is displayed in Figure 4.15. The result in underlined by the confusion matrix in Table 4.7 where the number of pixels classified as flooded outside the flooded neighbourhoods is a multiple of the number of pixels classified as flooded inside flooded neighbourhoods. These result go against the expectation that floodwater acts like a secular reflector leading to smaller back-scatter measurements, as the smallest pixel values do not correspond to the flooded pixels but the highest pixel values due.

By using each image individually speckle has an influence on each image and could therefore influence the result of the Stack of images outliers method. By using the mean of several images taken under normal condition the speckle can be reduced. This is what the Stack of images method tries to accomplish.

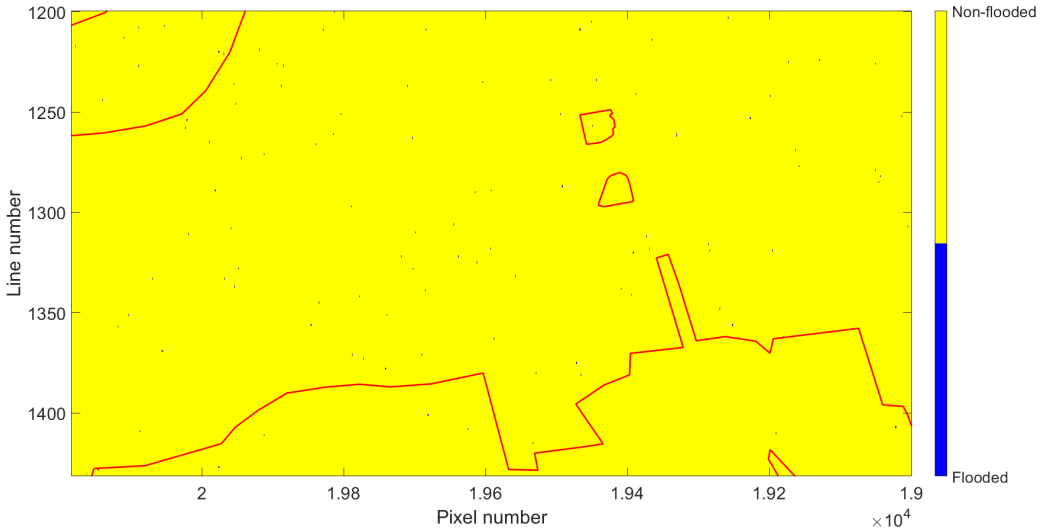


Figure 4.15: Result of the Stack of images outliers classification method (only the largest values) with a value of 15 dB for T_{low}. All pixels on the SAR image taken during the flood that are lower by the margin of at least 10 dB than in the entire stack are classified as flooded. The red polygons encircle the flooded neighbourhood, outside of these polygons there is no flood water present on the aerial images.

	Flooded neighbourhood	Dry neighbourhood
Classified flooded	17 (0 %)	90 (0 %)
Classified dry	60123 (100 %)	214458 (100 %)

Table 4.7: Confusion matrix for stack of images outliers lower with a minimum of 15 dB than in the previous 2 years. The first number represents the number of pixels and in the brackets the percentile of the total pixels in that neighbourhood. Each column thus adds up to a 100 percent.

4.4. Stack of images

The method uses a mean of a stack of images as a value for each pixel under normal conditions. The normal is than compared with the pixel value during the flood to determine if a pixel is flooded or not. When the stack of images is defined Eq. 3.5 can be implemented. The resulting image with pixel values P_c is shown in Figure 4.16. P_c is the result of $P_c = \frac{\overline{P} - P_f}{\sigma_p}$, it is unitless and can be described as a normalised difference value for pixels. The higher the absolute value of P_c the higher the likelihood that the pixel value in the flooded image has changed compared to the normal pixel value. This change is attributed to flood water.

Figure 4.16 contains the values of P_c. In this image the flooded polygons in the top half of the image seem to contain pixels with high values where-as the flooded neighbourhood in the bottom half of the image contains mainly low pixel values. This is in line with the expectations, the polygons in the top half are open fields where the flood water acts like a specular reflector resulting in a lower value than usual. In the lower half of the image the values of the pixels on the flooded image seem to be higher than usual resulting in low values when using Eq. 3.5. The water still acts as an specular reflector but due to buildings or other objects multiple reflections occur resulting in a high intensity backscatter measurement at the radar system.

The regions that stand out are roughly the same as the regions that stand out using the Image pair threshold method in Figure 4.7. Applying a threshold on the Image pair threshold method did not lead to a result in which all flooded streets were classified as such but a few pixels from flooded streets were classified correctly. In order to determine a threshold to apply on the image shown in Figure 4.16 the histogram of that image is used. The histogram is displayed in Figure 4.17.

On the positive side of the histogram it seems that a threshold will lead to more false positive classifications than positive classifications. On the negative side the histograms of all pixels and pixels

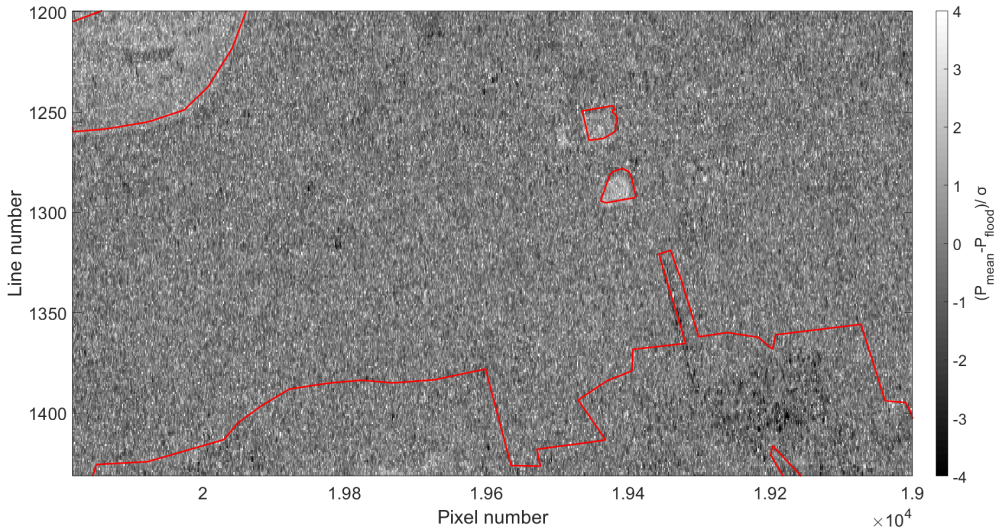


Figure 4.16: Result of using Eq. 3.5 on a stack of 90 Sentinel-1 SAR images taken during a 2 year period.The red polygons encircle the flooded neighbourhood, outside of these polygons there is no flood water present on the aerial images.

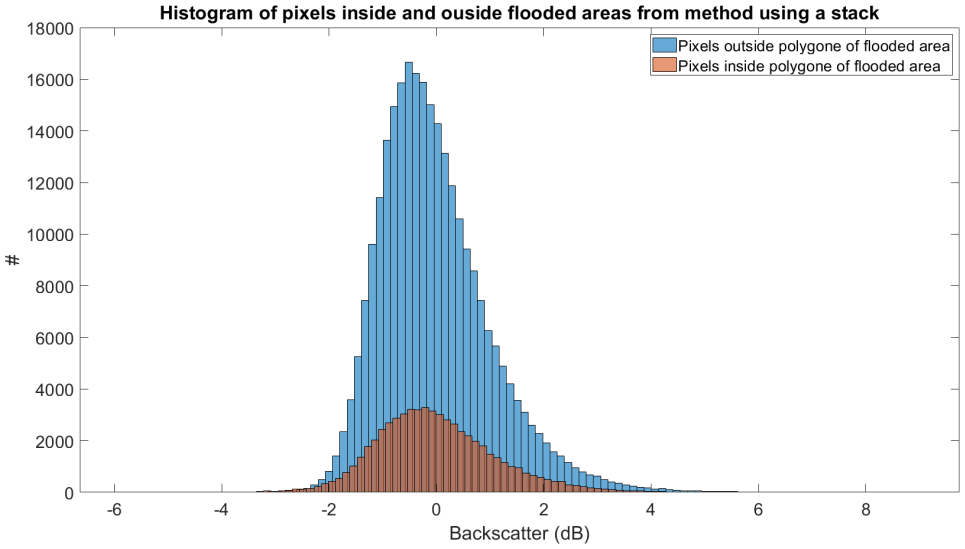


Figure 4.17: Histogram of pixel values in the image shown in Figure 4.16. The blue histogram is of all pixels in the image while the orange histogram is only of the pixels that fall inside of the polygons encircling the flooded regions.

.

inside the flooded neighbourhoods come much closer to one-another. Applying a threshold on this side could lead to a desirable result. In Figure 4.18 the results is shown of implementing a threshold of -3 to the image of Figure 4.16. Although there are some some false positives still present the majority of flooded pixel lie inside of polygon of the flooded neighbourhood. Lowering the threshold eliminates most of the false positives but also reduces the number of pixels classified as flooded in the flooded neighbourhood to a third. The confusion matrix for a threshold of -3 is shown in Table 4.8.

In order to decrease the number of false positive classification while retaining the pixels correctly classified as flooded a 3x3 mean filter window can be applied to the image with pixels P_c. This filter computes the mean for each pixel of its eight surrounding pixels. The histogram of this filtered P_c image is shown in Figure 4.19

The histogram shows that the highest values lie inside of the polygons enclosing the flooded neighbourhoods, it becomes easier to extract thresholds that lead to fewer false positives than in the non

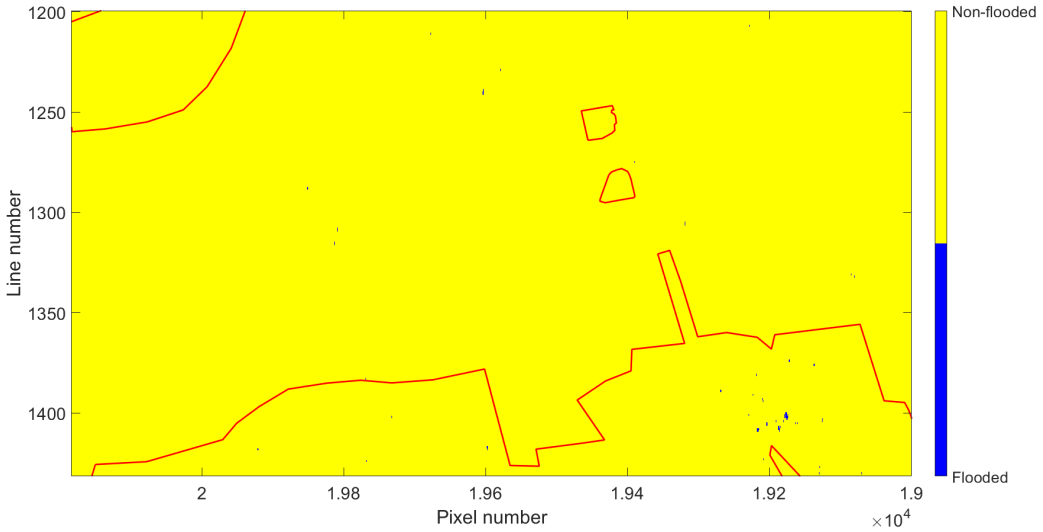


Figure 4.18: Result of the Stack of images method classification applied on the Sentinel-1 stack of images. All pixels in Figure 4.16 bellow -3 are classified as flooded. The red polygons encircle the flooded neighbourhood, outside of these polygons there is no flood water present on the aerial images.

.

	Flooded neighbourhood	Dry neighbourhood
Classified flooded	169 (0 %)	73 (0 %)
Classified dry	59971 (100 %)	214475 (100 %)

Table 4.8: Confusion matrix for stack of images with a threshold of -3. The first number represents the number of pixels and in the brackets the percentile of the total pixels in that neighbourhood. Each column thus adds up to a 100 percent.

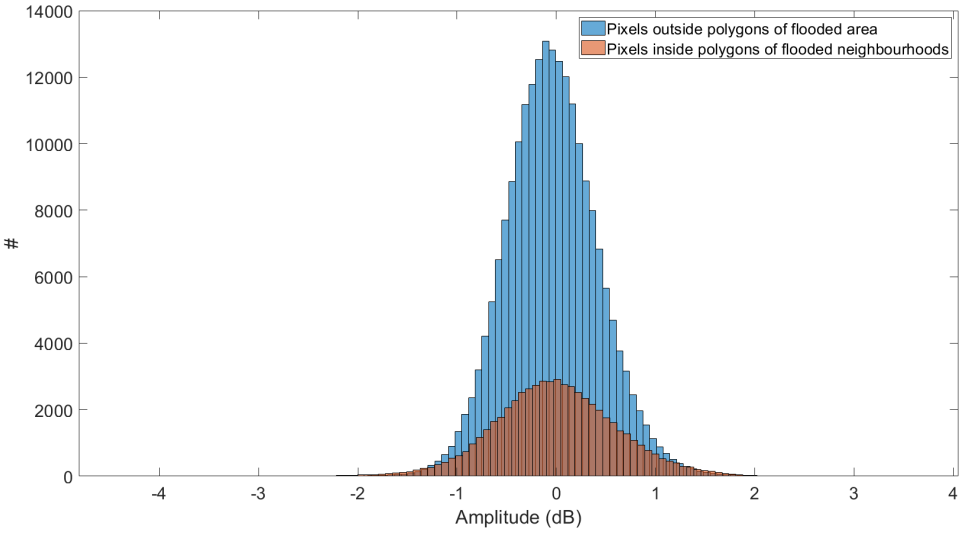


Figure 4.19: Histogram of image P_c (Figure 4.16) after mean 3x3 filter window is applied. The blue histogram is of all pixels in the image the orange histogram is only of the pixels that fall inside of the polygons encircling the flooded regions.

filtered image. This becomes even more apparent when a 5x5 mean filter window is applied that computes the mean of the 24 neighbouring pixels for each pixel. The histogram of the image that results from such a filter is shown in Figure 4.20. In this histogram the largest values from inside the flooded regions stand out even further than from the dry regions.

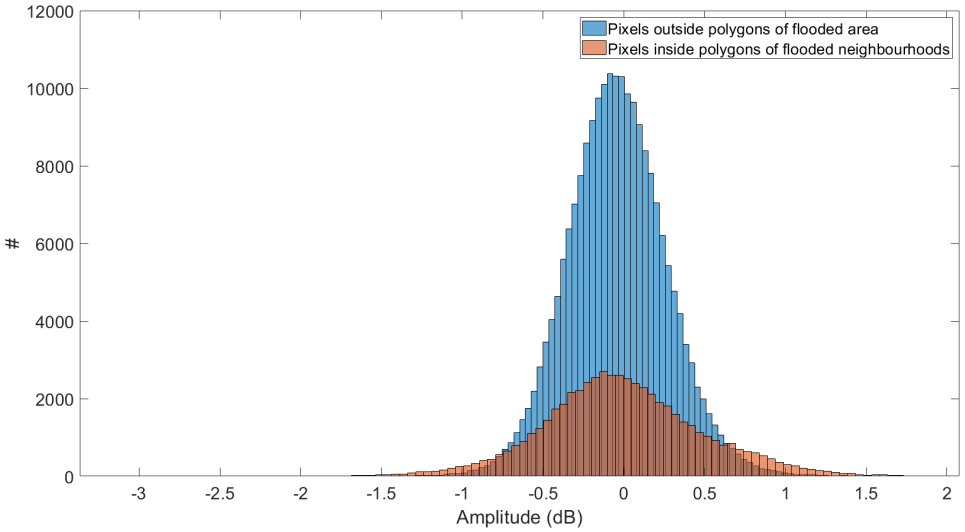


Figure 4.20: Histogram of image P_c (Figure 4.16) after mean 5x5 filter window is applied. The blue histogram is of all pixels in the image the orange histogram is only of the pixels that fall inside of the polygons encircling the flooded regions.

Using the histogram of Figure 4.19, a threshold of -2 is selected, everything below that value is classified as flooded. The resulting image is shown in Figure 4.21.

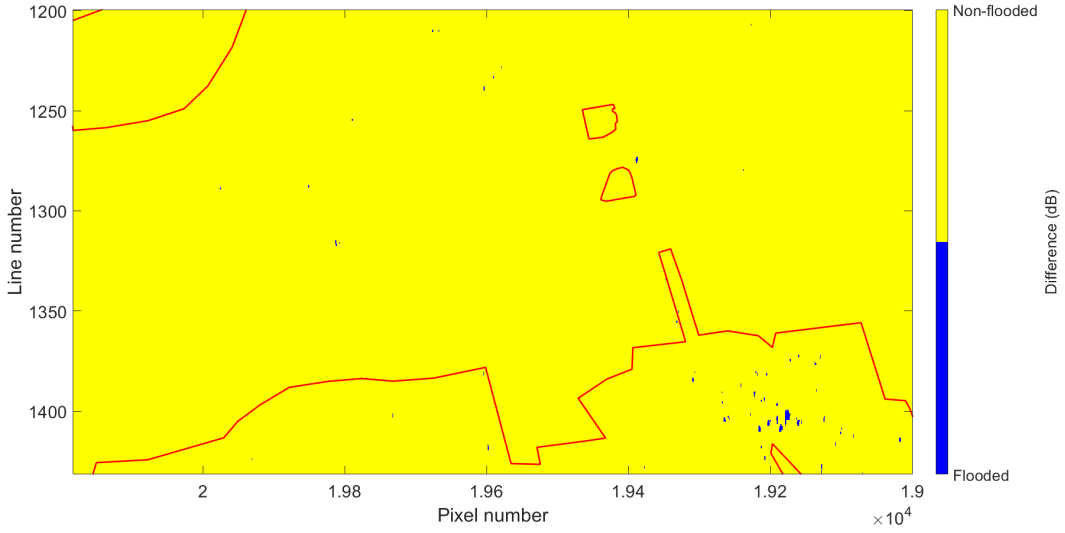


Figure 4.21: Result of the Stack of images method classification applied on the Sentinel-1 stack of images. All pixels in Figure 4.16 after a 3x3 mean filter window is applied that are bellow -2 are classified as flooded. The red polygons encircle the flooded neighbourhood, outside of these polygons there is no flood water present on the aerial images.

	Flooded neighbourhood	Dry neighbourhood
Classified flooded	169 (0 %)	32 (0 %)
Classified dry	58564 (100 %)	213095 (100 %)

Table 4.9: Confusion matrix for stack of images with a threshold of -2 after a 3x3 mean filter window is applied. The first number represents the number of pixels and in the brackets the percentile of the total pixels in that neighbourhood. Each column thus adds up to a 100 percent.

When comparing Figure 4.18 and 4.21 most of the false positives are eliminated in the second image, while most of the pixels inside of the polygons that were first classified as flooded still are. This is apparent when comparing the two confusion matrices in Tables 4.8 and 4.9, The number of flooded pixels inside the flooded neighbourhoods is the same while the number of pixels classified as flooded outside of the flooded areas is drastically lower in the filtered image. The drawback of filtering remains that single flooded pixels can be missed as their neighbouring pixels are not flooded. That the number of correctly classified flooded pixels is exactly the same is coincidence and does not mean that the same pixels are classified as flooded in both images.

The Stack of images method does not seem capable to detect the flooded fields. This is probably due to the fact that they are used as reservoirs and can be flooded several times in the stack of images. This results in a large value for σ_P and a low value for P_c. The method does however identify pixels in the flooded neighbourhood in the bottom half of the image, something the other methods did not manage to accomplish. The result does not nearly classify all or most of the flooded streets, it only seems to be able to detect the larger flooded streets. This is logical considering a mean filter window is used that reduces the resolution but is does miss the goal of this research to classify most if not all flooded streets in an urban environment. In order to fill in the gaps an additional data source could be used. The stack of images plus a DEM method will try to accomplish this.

4.5. Stack of images plus a DEM

The method continues on the image of P_c displayed in Figure 4.16. Instead of applying a threshold, as the Stack of images method does, the largest values are selected as these are most likely flooded and from these locations the flooded areas are grown using a DEM. The region growing result is shown in Figure 4.22. In this figure the colour represents the number of seed points every pixel is connected to. A higher number means that there are more seed points with a higher elevation connected to a pixel giving it a higher probability to also be flooded (as the seed points are flooded).

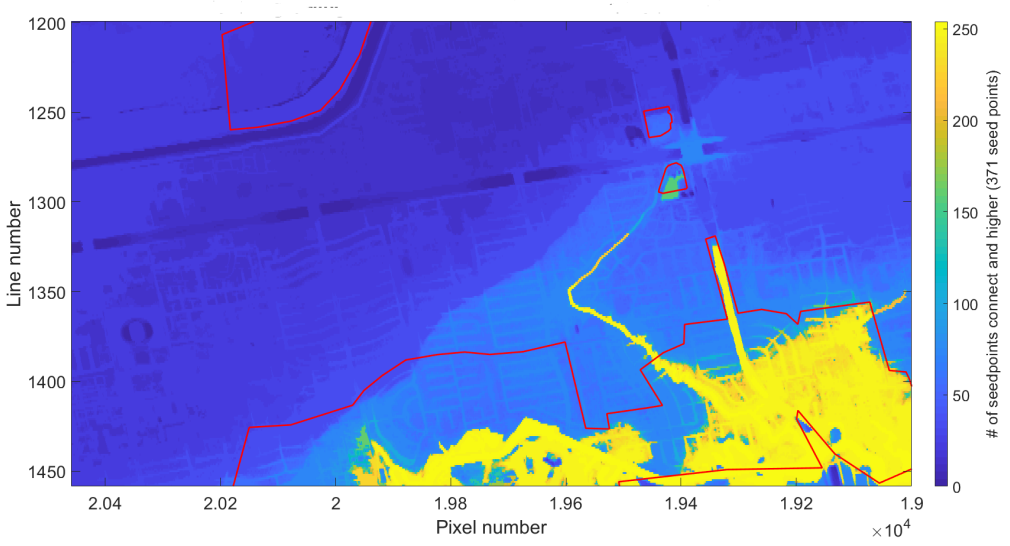


Figure 4.22: Result of region growing process in the Stack of images plus a DEM method. The region growing is performed using seed points taken from Figure 4.16 after a 5x5 mean filter window is applied. The pixels that are lower than -1.5 and higher than 1.5 are selected as seed points. In the image the number of each pixel represents the number of seed points it is connected to by a path going from high to low elevation trough neighbouring pixels. More seed points connected should mean a higher probability to be flooded when the SAR image was acquired.

The image shown in Figure 4.22 is the result of using the thresholds -1.5 and 1.5 to select the seed points. There are 371 seed points that meet this criteria and their coordinates are used for region growing on the DEM. The result is the first one where individual streets can be determined, the higher the value the more likely it is to be flooded. In this case the highest pixel values are all found inside of

the flooded neighbourhood in the lower part of the image.
There are however two problems that arise:
Firstly this map shows the likelihood of pixels to be flooded but is not a classification map, if every pixel that is connected to a seed point is classified as flooded the entire image would be flooded. Therefore a threshold needs to be applied to transform it into a flood map.
This is the second problem, every flooded region has a different number of seed points in it. This number is influenced by the size of the region and the number of flooded areas visible for the satellite. A very large flooded region naturally will contain more seed points than a small region, to identify the smaller region the threshold should be low, but a low threshold means that a few seed points in a non-flooded region will lead to a flooded classification.

The first problem can be dealt with by changing the thresholds to -2 and 2 this leads to an image that is displayed in Figure 4.23. Only 44 seed points meet this criteria and every one of them is located in the bottom polygon, in this image the step of applying a threshold is unnecessary as every pixel that lies in at least a single region determined by the seed points is inside the polygon as well. Classifying every pixel that is inside a region made by a seed point as flooded would lead to a desirable result.

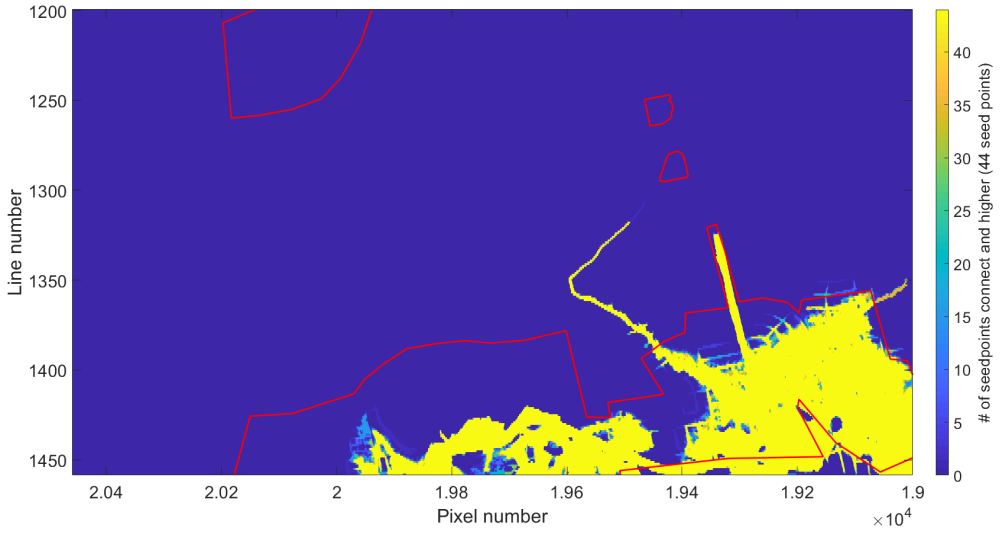


Figure 4.23: Result of region growing process in the Stack of images plus a DEM method. The region growing is performed using seed points taken from Figure 4.16 after a 5x5 mean filter window is applied. The pixels that are lower than -2 and higher than 2 are selected as seed points. In the image the number of each pixel represents the number of seed points it is connected to by a path going from high to low elevation trough neighbouring pixels. More seed points connected should mean a higher probability to be flooded when the SAR image was acquired. The image is created using a high resolution DEM.

	Flooded neighbourhood	Dry neighbourhood
Classified flooded	38852 (44 %)	4421 (2 %)
Classified dry	49306 (56 %)	214077 (98 %)

Table 4.10: Confusion matrix for stack of images plus a DEM with a threshold of -2 and 2 after a 5x5 filter is applied. The first number represents the number of pixels and in the brackets the percentile of the total pixels in that neighbourhood. Each column thus adds up to a 100 percent.

The polygons in the upper half of the image do not get any seed points in this method. This is because the fields of which they consist are storm drains. These basins are filled wit rainwater during periods of heavy rainfalls. Therefor it could be present during storms resulting in the basins being full on several of the radar images in the stack. This leads to a high value for σ_P in Eq. 3.5 and thus in a small value P_c going undetected in this method.

In the confusion matrix shown in Table 4.10 the highest number of correctly classified flooded pixels is shown. The number of flooded pixels outside of flooded areas might seem high but from Figure

Classified	Flooded		Dry		
Actual state	Flooded	Dry	Flooded	Dry	Flooded Flood/Flooded Dry
Single image threshold	938	1918	57795	211209	0.49
Image pair threshold	160	53	58573	213074	3.02
Stack of images outliers	55	3	60085	214545	18.33
Stack of images threshold	169	32	58564	213095	5.28
Stack of images plus a DEM	38852	4421	49306	214077	8.79

Table 4.11: Overview of the results produced by the 5 methods in one table.

4.23 we note that all of them either lie at the edge of a flooded area or in the river that is now easily identifiable in the figure. For this single case and this single image the result is satisfying as it clearly states where the flooded streets are even though it misses the flooded fields.

4.6. Summary

To sum up all of the results of the five methods Table 4.11 is given, in which the confusion matrices of the results are merged into one overview. A final column is added in which the ratio of pixels classified as flooded in flooded areas over pixels classified as flooded inside dry areas is given. This is one way of directly comparing methods as it says something about the accuracy of the method in different areas. The higher the number the better with a side note that the number doesn't tell us anything about the number of false positives or the number of positive flood classifications. For example, the number for the stack of images outliers is the highest in Table4.11 but the method only classifies a relatively small number of pixels as flooded inside the flooded area.

A closer look at the Stack of images plus a DEM method will be given in the next chapter, where the method is examined over a broader area and its potential shortcomings are investigated.

5

Discussion

In Chapter 4 the first results of all five methods are displayed along with a brief analysis and explanation. In this chapter a closer look at the results will be taken along with a discussion on their performances and shortcomings against the expected results. The main part of this chapter will focus on the Stack of images plus DEM method as it gave the best results in Chapter 4. The benefits and shortcomings of this method will be discussed as well as the performance of the method on a larger image.

5.1. Single image and image pair thresholds

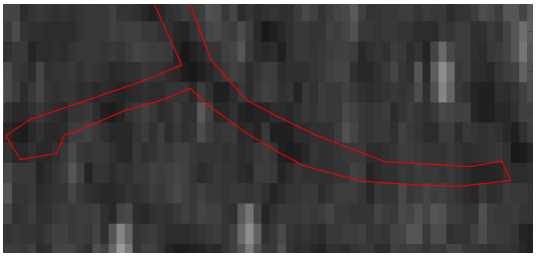
The expectation from still standing water was that is would act as a specular reflector resulting in a low reading of the satellite measurement. From the single image, low values should thus correspond with (flood)water. Using values from past literature on floodwater as thresholds yielded no satisfying results. Selecting the lowest values on the original image brought forth an image that contained pixels classified as flooded on the dry highway and on the flooded open fields. After the application of a 3x3 mean filter window and the selection of the lowest values there were three patterns that emerged.
Firstly, pixels in a large open field that were flooded were classified as such. Not all of them but enough to note that the area is flooded. This confirms that water acts as an specular reflector and can be identified on SAR images.
Secondly, a large part of the highway running through the image is incorrectly classified as flooded. This confirms the expected difficulty of urban flood mapping using SAR, namely that concrete and possibly other man-made structures with a smooth surface act as a specular reflector. This results in low intensity measurements at the radar that are similar to the readings from water surfaces.
Thirdly, the absence of pixels classified as flooded inside of the flooded neighbourhood. The bottom half of the image is full of streets that are flooded but almost no pixels are marked as flooded in this area. This could be the result of complex scattering mechanisms that influence the backscatter of the water before it reaches the satellite resulting in a similar or higher intensity measurement compared to a non-flooded street.

The use of an image pair improved on the Single image threshold method in the sense that the highway is no longer classified as flooded. This is because the backscatter of the highway is similar in both the before image and the image taken during the flood, resulting in a small difference. The Image pair threshold method is able to correctly classify some pixels as flooded in the flooded open sections in the image (after a mean-filter window is applied) while only classifying a low number of pixels as flooded outside of the flooded neighbourhoods. The flooded neighbourhood however is still not classified as such.

The first explanation for missing the flooded streets in a neighbourhood can be found in the pixel size. In Figure 5.1 two images are shown. The first is an aerial images of a section of flooded street where a polygon is drawn over the flooded street, the second image is from the same location but on the Sentinel-1 SAR image and it contains the same polygon. From the figure the size of a road is put into perspective with the size of the SAR pixels. On the areal image the flooded road is clearly

distinguishable from the lawns and houses. In the SAR image however a road often barley covers half a pixel or two half's of two pixels. In the case of Figure 5.1 it covers half a pixel, the flooded street is not the only contributor to the backscatter signal. This means that any change caused by flooding has less influence in the signal than when it would cover the entire pixel.
The second contributor to missing the flooded streets also has to do with size. Besides having less influence in pixel values the flood influence is further decreased when a mean filter window is applied. When a flooded street only covers half of a pixel and a 3x3 mean filter window is used the flood only covers 1/18 off the area over which the mean is taken. This influence is too small to result in a detectable change in the image.

(a) Part of an aerial image with a polygon shape drawn over a flooded street.



(b) Same polygon shape over the SAR image illustrating how the pixel size corresponds to the size of a street.

Figure 5.1: Visualisation of pixel size versus street size.

5.2. Filter

The 3x3 and 5x5 mean filter windows are used in this research to reduce speckle. When it comes to floods and urban areas there is an important downside. Consider a flooded road with buildings on both sides that is three pixels wide and parallel to the path of flight of the satellite. The pixel closest to the satellite track is not in direct sight of the satellite radar, the pixel in the middle is in direct sight and the floodwater scatters the radar signal away from the satellite resulting in a low measurement, the third pixel scatters the signal away but the building at the edge reflects the signal again towards to satellite resulting in a higher than usual signal. Taking the average over these tree pixels results in a signal similar to normal conditions letting the flood go unnoticed.
A street that is only one pixel wide will not benefit from a mean filter window either. The pixels on the road side will be taken into the average when such a filter is applies therefor reducing the influence of the flooded street pixel value P_c. In an open field a larger number of adjacent pixels acts in a similar way thus making the use of a mean filter window appropriate, in streets there will not be a large number of pixels flooded thus reducing the usefulness of the mean filter window. Than why use a Filter in the first place? The filter, especially the 5x5 filter, increases the accuracy of the pixels classified as flooded. When a pixel is marked as flooded after a mean filter window is applied the surrounding pixels are probably also flooded. A single pixel influenced by speckle or other changes not related by floodwater will no longer be classified as flooded as there is no change present in the surrounding pixels, reducing false positive classifications. This is especially useful for the selection of seed points where false positive classifications result in seed points that are not flooded. If the seed points are at a high elevation the region growing process could grow over the entire image.

5.3. Use of Stack

From the Figures 4.11 and 4.12 it is noted how large the range of values of the reflection from a single pixel over time can be in. A stack of several images could be used to identify these ranges and register when a pixel value falls outside of those ranges. The Stack of images outliers method did exactly that. Initially putting the ranges equal to the maximum and minimum value recorded in the stack did not give a result where the flooded areas could be easily identified. This could have been expected as the threshold values already occurred once so a slightly higher value under normal conditions was possible without the need of a flood having occurred. There was a clear difference between the values larger than previously recorded and the values smaller than previously recorded.

The smaller values were evenly distributed throughout the image. When a buffer was applied, meaning the values had to be smaller by a certain margin, there were no pixels identified as flooded in flooded regions and some in non-flooded regions. The smallest values were expected in the open fields, these open fields are however flood basins and possibly contained water on images in the stack. These previously flooded images enlarge the range of values that are considered non-flooded and result in no positive classification in this area. That there are no small values inside of the flooded neighbourhood could be the result of only narrow streets being present that cause double bounces and high reflection measurements at the satellite. When this is the case it should be visible in the image that classifies the pixels with larger values than in the stack as flooded. When a buffer is implemented on the on the larger values classification, meaning that only pixels that are larger by a certain margin are classified as flooded, the pixels that are classified as flooded all fall inside of the flooded neighbourhoods (although they are small in number). The buffer might filter out some values that are flooded but fall inside of the range that speckle can introduce and are thus filtered out.

The introduction of a stack in the Stack of image outliers method used every image in the stack separately and thus keeping the speckle in every image. Taking the mean of a stack as normal conditions can reduce the speckle in the image of ”normal” conditions. The size of the stack is a factor to take under consideration. Starting from a single image, adding images reduces the effect of speckle but with every image added the time span from first to last image increases. Using a longer time span leaves more time for changes in the ground conditions. New buildings arise while others are demolished, trees might be cut down and new parks or roads can be constructed. The surface in urban areas is subject to constant change. This could lead to values in the stack that are larger or smaller than in the present image only due to objects that are no longer present. Flooding might introduce change in these pixels that is not registered because its a smaller change than the building or removal of objects in the past.

For this thesis the stack covers a time span of two years. The initial results in Figure 4.13 show that floodwater does not necessarily lead to lower backscatter measurement at the satellite. In fact it shows no clear difference in the values that were smaller during the flood than in the previous two years inside or outside the flooded regions. The values that are larger than previously recorded do show a pattern, the open fields that are flooded are misclassified and the flooded neighbourhood contains a higher concentration of flooded pixels than outside of this region. The open fields can have been used to hold water in previous images, the stack thus contains low pixel values for these pixels. Containing water during the flood wont be a large deviation from the stack and thus not necessarily classified as flooded. The cluster of pixels classified as flooded in the flooded neighbourhood in Figure 4.13 (a) can suggest that the water leads to double bounces resulting in a higher measurement than before. The effort to keep these pixels classified as flooded while reducing the number of false positives was partly successful as Figure 4.14 suggests. But from this image it is not necessarily apparent that a large flooding has occurred in the area.

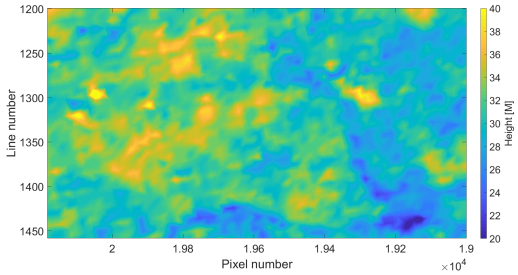
On its own the method can not detect open fields albeit that this is because the field in the extent of the image are often under water and that this is not always the case for all field in urban environments. The method detects and isolates more pixels in the flooded neighbourhood than the Image pair threshold method but not enough to confidently mark the entire neighbour that is flooded as such.

The Stack of images method takes the mean and the standard deviation from the stack instead of individual values. This in an effort to reduce the influence of speckle. The initial result (Figure 4.16) is promising as the number of pixels classified flooded inside the flooded neighbourhood rises compared to the stack of images outliers. The pixels classified as flooded outside of the flooded neighbourhoods unfortunately is also larger but this number can be greatly reduced by applying a mean filter window as shown in Table 4.8 and 4.9. Nevertheless, as a final product Figure 4.21 is not very useful as it is not capable of making a distinction between flooded and non-flooded streets. At this point one could argue that with this specific data set an accurate and detailed flood map purely based on the SAR images is not possible. The final result from the Stack of images method were however able to filter out some pixels in the flooded neighbourhood. The question then arises if it would it be possible to use these correctly classified pixels and use them to in conjunction with other data to produce a useful flood map.

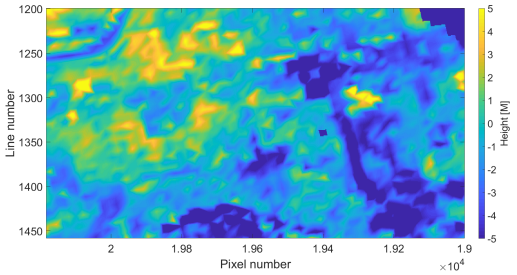
5.4. Region growing and DEM

By adding data in the form of a high resolution DEM a useful flood map is produced. Figure 4.22 shows a map on which the flooded neighbourhood is clearly different from the non-flooded neighbourhoods. Figure 4.23 goes even further as to completely isolate large parts of the flooded neighbourhoods from the non-flooded ones. The pixels that give a false positive are all at the edge of the flooded neighbourhood or in a river. The river always consist of water, therefore it does not necessarily show any difference between flooded and non-flooded images. Because the elevation of the river is lower than the land surrounding it seed points easily connect to it and classify the river as flooded. This is not necessarily negative but these pixels are labelled as non-flooded in the validation data thus resulting in a false positive in the result. The flooded fields are not classified as flooded because they contain water in several images in the stack and therefor do not result in a large absolute value of P_C. All in all the method produces a flood map with a high accuracy for this very specific case and area. There are however some questions that need to be posed before using it on other areas and larger scales.

To start with the DEM, the method highly relies on the DEM. For this case study a high resolution DEM was available freely, but this is not the case worldwide. What happens when a lower resolution DEM is used that is available around the globe? The SRTM DEM and the TerraSAR-X DEM are both freely available everywhere. Both DEMs are shown in Figure 5.2.



(a) SRTM DEM of the testing region in line pixel coordinates.



(b) TerraSAR-X DEM of the testing region in line pixel coordinates.

Figure 5.2: DEMs section of the testing region that are available world wide. The heights should not be compared between the images as they use a different reference point.

The same seed points that are used to create Figure 4.22 can be used on the two lower resolution DEMs. The results of the region growing processes are shown in Figure 5.3 and Figure 5.4. The resulting image based on the SRTM DEM does correctly detects flooding in a small part of the region but does not tell us much more than the result of the Stack of images method that does not use a DEM (Figure 4.21). The TerraSAR-X DEM results in a flood map where a very small region holds a large amount of seed points. The part around this region holds 6 seed points but is much larger. Unfortunately this part lies inside and outside of the flooded region, classifying it as flooded would result in a large amount of false positive classifications. The images show the influence of the DEM on the method. The SRTM DEM leads to an area classified as flooded that is smaller than the high resolution DEM. The TerraSAR-X DEM either leads to an even smaller flooded region or depending on the threshold a much larger area but one that contains a lot of false positive flood classifications. Without the high resolution DEM the method loses much of its value even on this small scale. On a larger scale more problems could arise, small elevation changes might not be included in a courser DEM resulting in a larger region for seed points to grow while in reality a barrier stops the flooding. Even when a high resolution DEM is available it might not contain every ledge or small elevation change that reality has. A simple edge of a sidewalk could redirect water or the curb could keep it from flooding over into the next street. Emergency services can also place temporary dams and dikes in an effort to protect regions against the water. These obstructions wont be present in the high resolution DEM.

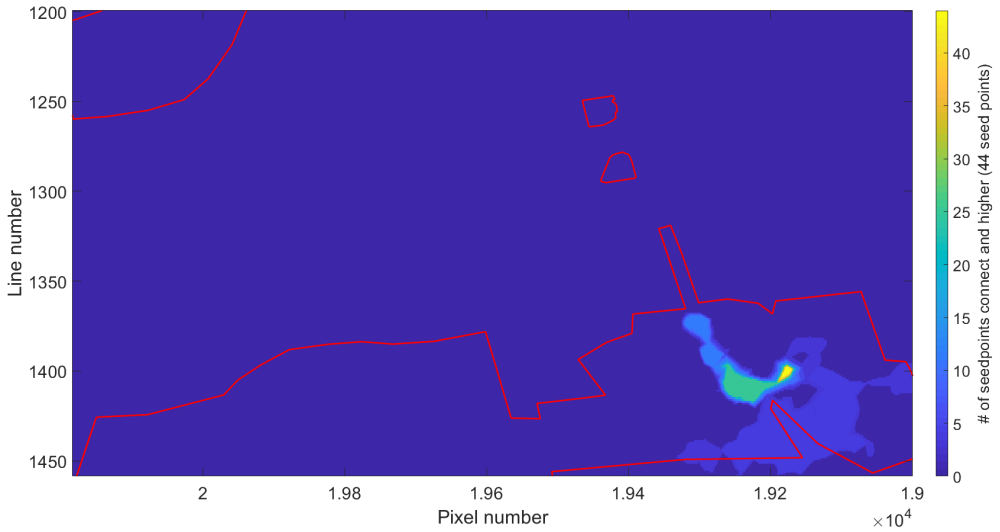


Figure 5.3: Result of region growing process in the Stack of images plus a DEM method. The method uses a SRTM DEM instead of the high resolution DEM used previously. The region growing is performed using seed points taken from Figure 4.16 after a 5x5 mean filter window is applied. The pixels that are lower than -2 and higher than 2 are selected as seed points. In the image the number of each pixel represents the number of seed points it is connected to by a path going from high to low elevation trough neighbouring pixels. More seed points connected should mean a higher probability to be flooded when the SAR image was acquired.

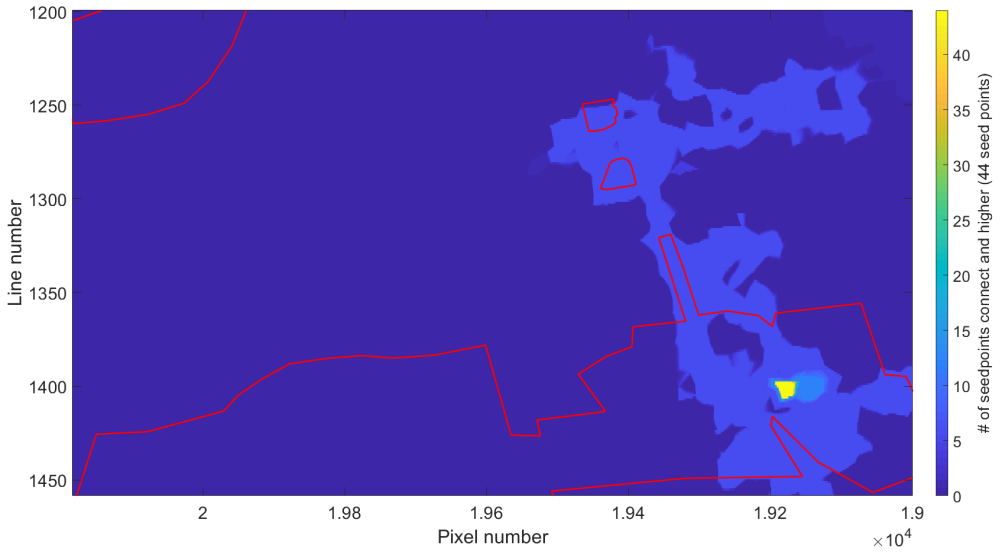


Figure 5.4: Result of region growing process in the Stack of images plus a DEM method. The method uses a TerraSAR-X DEM instead of the high resolution DEM used previously. The region growing is performed using seed points taken from Figure 4.16 after a 5x5 mean filter window is applied. The pixels that are lower than -2 and higher than 2 are selected as seed points. In the image the number of each pixel represents the number of seed points it is connected to by a path going from high to low elevation trough neighbouring pixels. More seed points connected should mean a higher probability to be flooded when the SAR image was acquired.

That brings on the next question, for the current image seed points are not restricted in growth. This means that as long as a lower elevation pixel is present the region grows. Imagine a seed point finding its way into a river, it could include the river and everything downstream until the edge of the image and if the coast is at a lower elevation than the seed point the entire coast line could be included into the flooded region. This could be partly solved by not using the elevation of the seed point as a criteria for all neighbouring pixels but use the value of each individual pixel if it is added to the region

to compare to its neighbours. By using this method even when a seed point finds its way into a river it should not be able to leave that river downstream as the river edges are higher than the river itself. Another problem of the region growing is its reach in urban areas. Not all water flows above ground and follows the elevation profile. Sewer systems are in place that drain the water and prevent it from potentially flooding other parts of the city. A pixel that is flooded and lies on top of a shallow slope would mean that the entire slope is flooded, even when it is 10 km long in the current method. This would only be true if all the water would flow above ground and does not infiltrate the ground or is drained by man made systems. What is the reach of a flooded pixel? A radius of ”infection” could be programmed for each seed point limiting its sphere of influence. Determining this radius is a research project on its own.
When the current method is applied on a larger area it produces a result that is shown in Figure 5.5. The area that was used for testing all five methods is in the bottom corner. It clearly stands out from its surroundings but it contains a lower number of seed points than the entire left side of the area. This is partly the result of a natural slope running through the image. Seed points on the right side of the image are always higher than most of the left side of the image The river jumps out of the image as a bright yellow line. The line gets brighter when it is followed downstream as every seed point that grows into the river will grow into every pixel downstream of the river as well.
Converting the result of the region growing into a flood map containing flooded/non-flooded pixels is another challenge. The classification of image 4.23 would not give any problems. All pixels connected so a seed point are flooded. This could be the criteria to classify flooded and non-flooded pixels (is connected to at least 1 seed point). Classifying Image 4.22 is a little bit harder as there are some seed points outside of the flooded neighbourhood. Because there is only one flooded neighbourhood (not counting the open field that are water reservoirs), using the threshold of being connected to at least 150 seed points would result in a satisfying result similar to Figure 4.23. Classifying gets harder when there are several unconnected flood areas. Take Figure 5.5 as an example. Every threshold of minimal number of seed points a pixel must be connected to will only classify the region that was tested in as flooded if the entire right side of the images is also classified as flooded. When there is one large region and several small flooded regions it is probable that the large region contains a lot more seed points that the smaller regions. In order to correctly classify the smaller regions a threshold would need to be set at a low number of seed points. This low number could mean that a small number of seed points that are outside of flooded regions result in large patches of dry areas being classified as flooded. Restricting the potential infinite growth of the seed points could partially reduce the difference in number of seed points connected to pixels in larger and smaller areas.

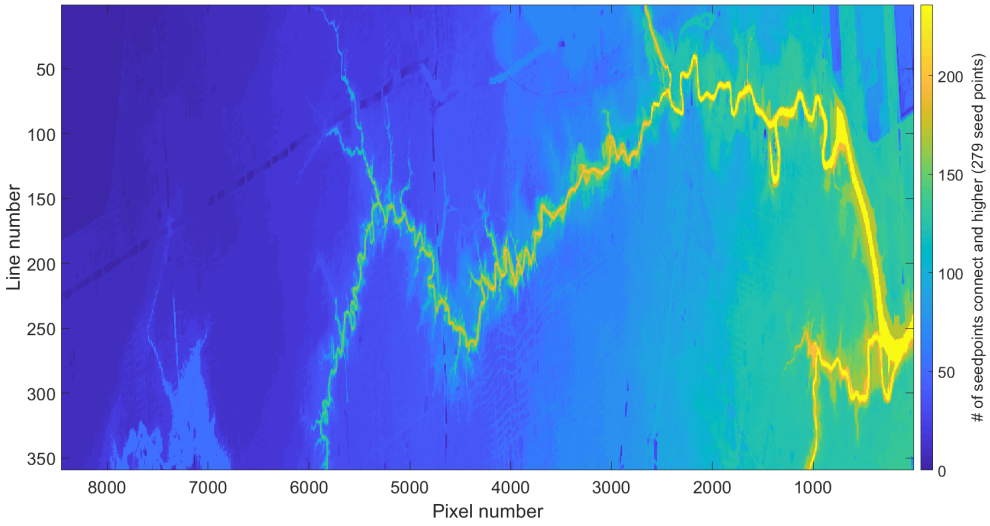


Figure 5.5: Result of region growing process on the Stack of images plus a DEM method. The region growing is performed on a larger area than the the testing area used for all previous images. Seed points are taken after a 5x5 mean filter window is applied. The pixels that are lower than -2 and higher than 2 are selected as seed points. In the image the number of each pixel represents the number of seed points it is connected to by a path going from high to low elevation trough neighbouring pixels. More seed points connected should mean a higher probability to be flooded when the SAR image was acquired. The image is created using a high resolution DEM.

5.5. Real world application

One of the goals of this research was the production of a flood mapping method suited for emergency situations in urban environments. The result from the Stack of images plus a DEM (Figure 4.23) is perfectly suited for that use. It does not matter that not every street in the flooded neighbourhood is identified as flooded. The result clearly shows where flooding has occurred. Based on the flood map help could be send to the lower right area of the map and no help to the top half, resulting in help getting to the area where it is needed. Although the result can be useful, the method is not yet useful, it only works on this small specific region and has far from a desirable result in other or larger areas. Take the result shown in Figure 5.5 as an example, emergency response can not yet be directed according to this result. To conclude, the current method has no practical use yet but if, after refinement, the same results are archived on a large scale than it could be of use. For the time being in case of a flood disaster the course of action is to take aerial images with a plane as the ones used in this research for validation purposes. On these images the flooded regions were easily distinguishable from the non-flood ones. When flying is not an option the availability of satellite imagery should be checked. Cloud cover might render them useless for flood mapping. When that is the case and this method is working on all areas than SAR images could be considered to be used in creating a flood map.

6

Conclusion & Recommendations

6.1. Conclusion

The main research question of this thesis as stated in Chapter 1 was:

Can the use of a temporal stack of SAR images improve the mapping of flood extent in urban areas?

This research shows that indeed the use of a temporal stack of SAR images can improve the mapping of flood extent in urban areas. The use of a stack results in more flooded pixels being classified as such but does not lead to a perfect result yet. Flood mapping with SAR in urban areas is difficult as the image released by the Copernicus Emergency Management Service (Figure 1.2) proves. The image shows little to no flooding in the metropolitan area whereas aerial images show large parts of the city under water. The traditional methods of a single image and an image pair had no difficulty detecting the flooded open areas in the region they were tested on. These open areas were classified as flooded in the Copernicus flood map. Like the Copernicus image however they did not detect flooding in the populated neighbourhoods in the same image.
All tree methods that used a stack of images produced a higher ratio of pixels classified as flooded in a flooded region compared to pixels classified as flooded outside of flooded regions (Table 4.11) than the methods that did not use a stack. In that aspect using a stack of images improves the mapping of flood extent in urban areas. The emphasis lies on improve as the introduction of a stack of images does not yet lead to an accurate flood map, only to more flooded pixels being detected than with the traditional methods.

The best flood map result produced for the specific region used to test all the methods is computed using the method Stack of images plus DEM. It clearly separates flooded from non-flooded streets. However, due to problems discussed in Chapter 5, the method is not yet able to do the same for larger or other regions without making alterations and defining a thresholds for P_c and minimum number of seed points connected to an individual pixel to classify as flooded.

The flood map produced for the region it is tested on correctly identifies most flooded streets while not classifying dry streets as flooded. The method to select seed points with the use of a stack, selects seed points on flooded streets that are not classified as flooded by the Single image and Image pair method.
The downside of the method is the heavy reliance on the DEM. When a low resolution DEM as the SRTM DEM or the TerraSAR-X DEM are implemented the resulting flood map does not give a clear image of flooded and non-flooded streets. The seed points are selected after a 5x5 mean filter window is applied. This increases the accuracy of the selected seed points but reduces the resolution on which floods are detectable. The resolution of the Sentinel-1 SAR images is already low compared to the dimensions of the flooded streets but a mean filter window decreases the resolution even more. A neighbourhood that would only consist of narrow streets will not contain any seed points when a filter is applied. The Stack of images plus DEM method has even more shortcomings when applied on larger regions. The region growing method needs to be revised as unlimited growth for each seed point in

not sustainable over larger areas. The classification step is not suitable for larger images either. When different flooded neighbourhoods are present in one image, especially the combination of larger and smaller regions, one threshold will not work. The smaller regions will naturally contain less seed points than the larger region but implementing a low threshold increases the influence of false positive seed points. Speckle is another problem of the Stack of images plus DEM method. The speckle present in the SAR image taken during the flood can lead to false positives seed points. Again these seed points can have a large influence when a low threshold is used.

Besides the main research question there were three sub-questions, these will be answered below:

1. *How do floods influence the radar reflections used in SAR?*
   For this research it was assumed that flood water acts as a specular reflector. With waves and wind the surface of water can become a diffuse reflector. In the single image and the image pair method it was clear that (in an open field) the flood water gave low amplitude backscatter measurements confirming the assumptions. The Stack outliers method showed a different effect in the area with roads and buildings. When the area was flooded there were a lot of pixels that had higher than ever backscatter measurements. Floodwater still acts as a specular reflector but objects in and around these pixels scatter the signal multiple times, resulting in a high backscatter measurement. This is largely in line with the findings of Mason *et al.* (2013). They concluded that urban flood mapping using SAR could benefit when taking into account the presence of double bounces that result in high intensity measurements caused by flooding instead of ignoring them and only looking for low intensity measurements to detect floodwater.

2. *Which methods can be applied to detect flooded pixels from a stack of SAR amplitude images?*
   All three methods using a stack were able to detect flooded pixels from the SAR amplitude images without false classifying large amounts of pixels as flooded that were dry in reality. It does need to be noted that no method can detect enough flooded pixels to make a flood map or to roughly differentiate flooded from non-flooded neighbourhoods as the correctly detected pixels were often clustered together in one part of the flooded neighbourhood. The thresholds used for this research are specific for the test area and do not necessarily produce the same results when applied on different regions.

3. *Is it feasible to produce a flood map of an urban environment based on SAR data alone or is auxiliary data required to reach a reliable result?*
   When using Sentinel-1 data for the flood of Houston in 2017 it isn't feasible to produce a flood map in the metropolitan area without using auxiliary data. The resolution of the SAR images is too low to distinguish single narrow streets which means a flooded street has too little influence on the backscatter value measure for the pixel it is on. It could be that a higher resolution like the TerraSAR-X images are able to produce a flood map on there own.
   The use of auxiliary data does produce a flood map that is usable. Albeit only for the one specific testing area and not yet for other areas. The addition of a high resolution DEM produces a flood map as in Figure 4.23. This flood map clearly and correctly differentiate flooded streets from dry ones.

To conclude, the research developed the Stack of images plus DEM method that give a desirable result of the region it was tested on. The Single image and Image pair methods are implemented as they are the most often used methods in past studies. These methods are shown to produce no desirable results in urban areas compared to open fields. Lastly, the importance of ancillary data for the Stack of images plus a DEM is stated as well as the shortcomings of the method, therefore there are several recommendations to potentially improve the method.

6.2. Recommendations

In the conclusion the Stack of images plus a DEM method is presented as a working method on a specific area. It is also noted that before application on other regions or on a larger scale is possible, several aspects of the method need to be altered. Those aspects are given in the following section.
The main problem that all five methods dealt with was the resolution of the Sentinel-1 SAR images. In future research all five methods should be tested using higher resolution SAR images to determine if flooded pixels can be better identified. With higher resolution data the mean filter windows could produce a better result. The filters would still decrease speckle and resolution but with higher resolution SAR images the resolution after the filter application would still be higher than the unfiltered Sentinel-1 SAR image. The increased resolution could lead to narrower flooded streets being correctly classified.

The filter itself could also be subject to improvement, instead of a mean filter window a ”smart” filter could be developed. This filter would recognise that if one pixel has a higher that usual value and its neighbouring pixel a lower than usual value both pixels are likely to have flooded. As this is the result of one pixel causing only a single bounce while the neighbouring pixel causes a double bounce reflection from the flood water and urban surroundings.

Further research into the application of a threshold on the region growing image is needed. When processing a large image the region growing could be applied of several parts of the image instead of the entire image at once. This limits the number of seed points a pixel can be connected to as it is dependent on the number of seed points in that part of the image. Another potential solution is to have different thresholds for different part of the image.
Limiting the range on which seed points can expand could also make it easier to apply one threshold as it also limits the number of seed points that can be connected to a single pixel. The range of seed points should be further investigated as it does not only influence the number of seed points that can be connected to pixel but also how far a flood can be extrapolated over the ground surface. Different city’s have different water drainage systems, different soils have different saturation rates. These are all factors of influence on how far floodwater flows over ground. The study should point out what a realistic range for seed point is and if this range can be set the same for every flood case.

The selection of seed points is done with a threshold. This threshold is subject to optimisation. The question should be answered if every flood case can be mapped using the same thresholds for the selection of seed points. There could be floods in urban areas that do not cause double scatterings and therefore will not result in any pixels having a higher than normal amplitude measurement. If a flood event is known to have occurred, instead of a threshold value to select seed points maybe the top percentiles of pixels that have changed to most could be used as seed points. This would lead to a number of seed points that can be chosen depending on the size of the flood and leads to a different threshold value for each individual flood without the need to manually select a value. This is under the assumption that flooded pixels will always display the most change of all pixels in an image compared to the image in dry conditions.

Finally the addition of other data sources is recommended. A map of street boundaries or the location and shape of buildings could be valuable for the analysis and validation of the method. When all the pixels in an image that lie on a road could be selected separately a better histogram could be created. In the current histograms all pixels are included but when only roads are selected a clearer difference could emerge between flooded and non-flooded streets. This could lead to the selection of better thresholds.

A

Abbreviations

DEM	Digital elevation model
EU	European Union
ESA	European Space Agency
GDEM	Global Digital Elevation Map
InSAR	Interferometric synthetic aperture radar
LiDAR	Light detection and ranging
NED	National Elevation Dataset
NOAA	National Oceanic and Atmospheric Administration
SRTM	Shuttle Radar Topography Mission
WGS84	World Geodetic System 1984

B

Flood events versus available TU Delft SAR data

Appendix contains a table showing dates of flood events in Rotterdam and Amsterdam versus the closest recording dates of several SAR satellites over those city's. The satellites are selected because their data was available on the TU Delft servers. The table demonstrates that even though there are several SAR missions ongoing there is no guarantee that a one day flood event is recorded by a SAR satellite.

					Sentinel			
	Flood event	tsx_dsc	tsx_asc	rsat2	t037_dsc	t110_dsc	t088_asc	t161_asc
Rotterdam								
	23-06-16	30-06-16	09-07-17	No data	05-07-16	23-06-16	08-07-16	01-07-16
	30-07-17	31-07-17	09-08-17	No data	No data	No data	No data	No data
	01-09-17	02-09-17	22-09-17	No data	No data	No data	No data	No data
Amsterdam								
	28-07-14		31-07-14	No data	No data	No data	No data	No data
	08-09-15		08-09-15	No data	20-11-15	14-09-15	12-09-15	17-09-15

Table B.1: Table contains the major urban floods in Amsterdam and Rotterdam vs the closest recording dates for the satellites of which data is available at the TU delft.
-tsx stands for TerraSAR-X.
-dsc and asc stand for descending and ascending.
-rsat stands for Radarsat-2.
-The Sentinel satellite has different tracks that pass over the city's.
-No data means that there was no data available around the data of the flood, on the TU Delft server not that there were no measurements at all.

Bibliography

A. L. Taylor, S. Dessai, and W. B. de Bruin, *Public perception of climate risk and adaptation in the uk: A review of the literature,* Climate risk management **4-5**, 1 (2014).

M. R. Allen and W. J. Ingram, *Constraints on future changes in climate and the hydrologic cycle,* Nature **419**, 224 (2002).

G. C. Hegerl, thomas J. Crowleu, M. Allen, W. T. Hyde, H. N. Pollack, J. Smerdon, and E. Zorita, *Detection of human influence on a new, validated 1500-year temperature reconstruction,* Journal of climate **20**, 650 (2006).

G. A. Milne, W. R. Gehrels, C. W. Hughes, and M. E. Tamisiea, *Identifying the causes of sea-level change,* Nature Geoscience **2**, 471 (2009).

A. J. Garner, M. E. Mann, K. A. Emanuel, R. E. Kopp, N. Lin, R. B. alley, B. P. horton, robert M. DeConto, J. P. Donnelly, and D. Pollard, *Impact of climate change on new york city's coastal flood hazard: Increasing flood heights from the preindustrial to 2300 ce,* PNAS **114**, 11861 (2017).

F. Friesecke, *Precautionary and Sustainable Flood Protection in Germany – Strategies and Instruments of Spatial Planning*, Tech. Rep. (FIG Regional Conference, 2014).

P. Doody, M. Ferreira, S. Lombardo, I. Lucius, R. Misdorp, H. Niesing, A. Salman, and M. Smallegange, *Living with coastal erosion in europe, sediment and space for sustainability, results from the erosion study*, Tech. Rep. (European Commission, 2014).

European-Environment-Agency, *Mapping the impacts of recent natural disasters and technological accidents in europe,,* Environ **35** (2004).

European-Environment-Agency, *Disasters in europe: more frequent and causing more damage,* (2011).

D. C. Mason, R. Speck, B. Devereux, G. J.-P. Schuman, J. C. Neal, and P. D. Bates, *Flood detection in urban areas using terrasar-x,* IEEE Transactions on geoscience and remote sensing **48**, 882 (2010).

Y. Depietri, F. G. Renaud, and G. Kallis, *Heat waves and floods in urban areas: a policy-oriented review of ecosystem services,* Sustainability Science **7**, 95 (2012).

I. Demir and W. F. Krajewski, *Towards an integrated flood information system: Centralized data access, analysis, and visualization,* Environmental Modelling & Software **50**, 77 (2013).

J.-M. Cariolet, *Use of high water marks and eyewitness accounts to delineate flooded coastal areas: The case of storm johanna (10 march 2008) in brittany france,* Ocean & Coastal Management **53**, 679 (2010).

D. Mason, L. giustari, J. Garcia-Pintado, and H. Cloke, *Detection of flooded urban areas in high resolution synthetic aperature radar images using double scattering,* Journal of applied earth observation and geoinformation **28**, 150 (2013).

D. Mason, I. J. Davenport, J. C.Neal, G. J.-P. Schumann, and P. D. Bates, *Near real - time flood detection in urban and rural areas using high resolution synthetic aperture radar images,* IEEE Transactions on Geoscience and Remote Sensing **50**, 3041 (2012).

R. F. Hanssen, *Radar Interferometry data interpretation and error analysis*, Ph.D. thesis, TU Delft (2001).

R. Bamler and P. Hartl, *Synthetic aperture radar interferometry,* Inverse problems **14**, R1–R54 (1998).

U. Soergel, U.Thoennessen, and U. Stilla, *Visibility analysis of man-made objects in sar images,* GRSS/ISPRS Joint Workshop on Remote Sensing and Data Fusion over Urban Areas **2** (2003).

S. Schlaffer, P. Matgen, M. Hollaus, and W. Wagner, *Flood detection from multi-temporal sar data using harmonic analysis and change detection,* International Journal of Applied Earth Observation and Geoinformation **38**, 15 (2014).

G. Nico, M. Pappalepore, G. Pasquariello, A. Refice, and S. Samarelli, *Comparison of sar amplitude vs. coherence ood detection methods—a gis application,* International Journal of remote sensing **21**, 161 (2000).

B. Brisco, *Remote sensing of wetlands: Applications and advances,* (2015).

A. Roth, G. Braun, G. Schreier, and R. Werninghaus, *Terrasar-x science plan,* Released by DLR (2004).

U.S. Census Bureau, *Quickfacts houston city, texas,* (2017), https://www.census.gov/quickfacts/fact/table/houstoncitytexas/PST045217, Last accessed on 2019-09-10.

A. Davis, J. Gillum, and A. Tran, *Houston's 'wild west' growth how the city's development may have contributed to devastating flooding,* The Washington post (2017).

E. Blake and D. A. Zelinsky, *Hurricane Harvey*, Tech. Rep. (National hurricane center, 2017).

Rijksoverheid, *Jaarlijkse hoeveelheid neerslag in nederland, 1910-2015,* (2016), www.clo.nl/printpdf/18206, Last accessed on 2019-09-10.

M. Fischetti, *Hurricane harvey: Why is it so extreme?* Scientific American, a division of Nature America, Inc. (2017).

Randy Cephus, *Corps releases at addicks and barker dams to begin,* (2017), http://www.swg.usace.army.mil/Media/News-Releases/Article/1291369/corps-releases-at-addicks-and-barker-dams-to-begin, Last accessed on 2019-09-10.

R. Showstack, *Sentinel satellites initiate new era in earth observation,* Eos **95**, 239 (2014).

R. Torresand, P. Snoeij, D. Geudtner, D. Bibby, M. Davidson, E. Attema, P. Potin, B. Rommen, N. Floury, M. Brown, I. N. Traver, P. Deghaye, B. Duesmann, B. Rosich, N. Miranda, C. Bruno, M. L'Abbate, R. Croci, A. Pietropaolo, M. Huchler, and F. Rostan, *Gmes sentinel-1 mission,* Remote Sensing of Environment **120**, 9 (2012).

E. Attema, P. Bargellini, P. Edwards, G. Levrini, S. Lokas, L. Moeller, B. Rosich-Tell, P. Secchi, R. Torres, M. Davidson, and P. Snoeij, *Sentinel-1, the radar mission for gmes operational land and sea services,* ESA Bulletin **131**, 10 (2007).

R. Torres, P. Snoeij, D. Geudtner, D. Bibby, M. Davidson, E. Attema, P. Potin, B. Rommen, N. Floury, M. Brown, I. N. Traver, P. Deghaye, B. Duesmann, B. Rosich, N. Miranda, C. Bruno, M. L'Abbate, R. Croci, A. Pietropaolo, M. Huchler, and F. Rostan, *Gmes sentinel-1 mission,* Remote Sensing of Environment **120**, 9 (2012).

K. G. Nikolakopoulos, E. K. KAMARATAKIS, and N. CHRYSOULAKIS, *Srtm vs aster elevation products. comparison for two regions in crete, greece,* International journal of geoinformatics (2013).

K. Czubski, J. Kozak, and N. Kolecka, *Accuracy of srtm-x and aster elevation data and its influence on topographical and hydrological modeling: Case study of the pieniny mts. in poland,* International journal of geoinformatics (2013).

D. B. Gesch, *Chapter 4 – the national elevation dataset,* Bethesda, Maryland, American Society for Photogrammetry and Remote Sensing , 99 (2007).

S. Martinis, A. Twele, and S.Voigt, *Towards operational near real-time flood detection using a split-based automatic thresholding procedure on high resolution terrasar-x data,* Natural Hazards and Earth System Sciences **9**, 303 (2009).

P. Manjusree, L. P. Kumar, C. M. Bhatt, G. S. Rao, and V. Bhanumurthy, *Optimization of treshold ranges for rapid flood inundation mapping by evaluationg backscatter profiles of high incidence angle sar images,* Int. J. Disaster Risk Sci. **3(2)**, 113 (2012).

K. S, G. Akpeko, and A. K. E. Isaac, *On the performance of filters for reduction of speckle noise in sar images off the coast of the gulf of guinea,* International Journal of Information Technology, Modeling and Computing (IJITMC) **1**, 41 (2013).