MSc thesis

Detection of multi-transitional abrupt changes in SAR time series

B. Scherpenisse

TU Delft

Detection of multi-transitional abrupt changes in SAR time series

by

Brendan Scherpenisse

to obtain the degree of Master of Science
at the Delft University of Technology,
to be defended publicly on Wednesday September 26, 2018 at 13:30.

Student number:	4237390	
Project duration:	December 14, 2017 – September 26, 2018	
Thesis committee:	Prof. dr. ir. R. F. (Ramon) Hanssen,	TU Delft, Geoscience and Remote Sensing
	Dr. ir. F. J. (Freek) van Leijen,	TU Delft, Geoscience and Remote Sensing
	Dr. S. L. M. (Stef) Lhermitte,	TU Delft, Geoscience and Remote Sensing
	Dr. ir. M-. C. (Marie-Claire) ten Veldhuis,	TU Delft, Water Management

This thesis is confidential and cannot be made public until September 20, 2018.

TU Delft

Preface

This project came into existence following a discussion on using SAR time series to detect flooded regions in urban areas in Houston, caused by Hurricane Harvey in August of 2017. It gradually evolved; focussing more on change detection in SAR times series in general and less on the specific application. This also gave way to having more emphasis on the data processing and hands-on high-level programming that came with this, something I really enjoy. The MSc thesis was conducted at the department of Geoscience and Remote Sensing, which is part of the faculty Civil Engineering and Geosciences from the Delft University of Technology, and it presents the findings of a research spanning a little under a year.

I'd like to take this moment to acknowledge the people who helped me during this time, namely my supervisors Ramon Hanssen, Freek van Leijen, Stef Lhermitte and Marie-Claire ten Veldhuis. First, thanks to Marie-Claire for offering a different and fresh perspective in a group consisting of all but one people specialized in satellite Earth observation. Secondly, I'd like to thank Stef for his more application-oriented remarks and his help with the time series analysis. Moreover, I'd like to thank Ramon for his critical remarks and insightful ideas on how to add value to the project. Lastly, thanks to Freek for his help with the data processing and a special thanks for taking time out of his busy schedule to provide weekly feedback and support. In all honesty, without this strong support group the project would've most likely not reached its current state in the given timeframe.

Brendan Scherpenisse
Delft, September 20, 2018

Contents

Summary

Repeat-pass acquisitions with coherent Synthetic Aperture Radar (SAR) systems, preserving both phase and amplitude, are more readily available than ever (Bruzzone, 2016). Phase measurements from SAR systems have seen widespread use in the Interferometric Synthetic Aperture (InSAR) technique to measure deformations and elevations since the late 1980's (Hanssen, 2002). Since the late 1990's an increase in radar-based change detection is observed, mainly relying on amplitude measurements (Ajadi et al., 2016; Dekker, 1998). The unpredictive multiplicative noise-like speckle, inherent to coherent SAR, makes change detection in SAR imagery difficult (Bamler, 2015). However, the advantages in the all-weather mapping capabilities and object penetrating properties of SAR make it a suitable remote sensing technique for certain applications, such as natural disaster damage assessment (Bruzzone and Prieto, 2000).

Broadly speaking, change detection in SAR-based images usually consists of applying an operator on two spatially filtered SAR images to create a difference image (DI), which is then analysed for change points by thresholding and/or clustering (e.g. Alphonse and Biju, 2015). However, such an approach completely neglects the long-term stability of a pixel. When taking the temporal evolution of a pixel into account, the steep increase in data volume puts an emphasis on finding an optimal ('best practice') approach to the multi-temporal change point detection problem.

Here it is shown that change point detection methods that properly take the temporal evolution of a pixel into account can provide good segmentation results in multi-temporal SAR data stacks, even in unfiltered stacks that preserve the complete spatial resolution and without considering the spatial context in a pixel's direct neighbourhood. Moreover, it is found that more sophisticated change point detection algorithms don't necessarily yield superior segmentation results for various discontinuity functions. This means algorithm selection has to be application driven. The results demonstrate that the suitability of algebraic methods in heterogeneous areas is limited, whereas proper time series analysis yields fairly consistent results over different land covers within the same image.

Often, little effort is spend on finding an optimal approach; neglecting data selection and storage, a sensitivity analysis and/or the post-processing analysis procedure, all of which are shown or known to increase the success rate, efficiency and understanding of the segmentation results. It is anticipated that change point detection in SAR imagery will shift away from the classical bi-temporal methodology and multi-temporal approaches will become the norm, be it by decomposing multi-temporal stacks or time series analysis.

Nomenclature

List of acronyms

1D	One-Dimensional
2D	Two-Dimensional
3D	Three-Dimensional
C-SAR	C-Band Synthetic Aperture Radar
CADS	Calibration Annotation Data Set
CLT	Central Limit Theorem
CPD	Change Point Detection
DEM	Digital Elevation Model
DI	Difference Image
DN	Digital Number
EW	Extra-Wide swath
FM	Frequency Modulation
FM	Frequency Modulation
FN	False negative
FP	False positive
GRD	Ground Range Detected
GTC	Geocoded-Terrain-Corrected
HH	Horizontal transmit and Horizontal receive (polarization)
HS	High Resolution Spotlight
HS300	High Resolution Spotlight 300 MHz
HV	Horizontal transmit and Vertical receive (polarization)
I/Q	In-Phase (real part) / Quadrature (imaginary part)
IA	Incidence Angle
IW	Interferometric Wide swath
JC	Jaccard Coefficient
LEO	Low Earth Orbit
LIA	Local Incidence Angle
LOS	Line of Sight
LUT	Look-Up Table

PDF	Probability Density Function
PR	Precision-Recall (curve)
PRF	Pulse Repetition Frequency
PRI	Pulse Repetition Interval
PS	Persistent Scatterer
RA	Ranging Altimeter
RADAR	Radio Detection and Ranging
RAR	Real Aperture Radar
RCS	Radar Cross Section
RNN	Relative Radiometric Normalization
ROC	Receiving Operating Characteristics (curve)
S1	Sentinel-1
SAR	Synthetic Aperture Radar
SC	ScanSAR
SL	SpotLight
SLC	Single-Look Complex
SM	StripMap
SNR	Signal-to-Noise Ratio
SSO	Sun-Synchronous Orbit
ST	Staring Spotlight
TN	True negative
TP	True positive
TSX	TerraSAR-X
VH	Vertical transmit and Horizontal receive (polarization)
VV	Vertical transmit and Vertical receive (polarization)
WGS84	World Geodetic System 1984
WS	Wide ScanSAR
WV	Wave
X-SAR	X-Band Synthetic Aperture Radar
YC	Yule Coefficient
s/c	Spacecraft

List of symbols

ψ	Phase
λ	Wavelength [m]

σ	Radar cross section [m^2]; standard deviation
Φ	Incidence angle [^{o}]
θ	Local incidence angle [^{o}]
τ	Pulse width [s]; threshold
β	Beam width [rad]
ν	Drift
δ	Decision rule
Λ	Test statistic
μ	Mean
σ^0	Backscatter per unit area in ground range [-]
β^0	Backscatter per unit area in slant range [-]
γ^0	Backscatter per unit area perpendicular to slant range [-]
$\Delta_a^{\mathbf{RAR}}$	RAR azimuth resolution
$\Delta_a^{\mathbf{SAR}}$	SAR azimuth resolution
$\Delta_a^{\mathbf{SAR,(unfocused)}}$	Unfocused SAR azimuth resolution
τ_i	Change point location
$\Delta_r^{\mathbf{RAR}}$	RAR range resolution
$\Delta_r^{\mathbf{SAR}}$	SAR range resolution
σ_s	Surface scattering
σ_t	Dihedral return
σ_v	Volume scattering
A	Amplitude; scattering area [m^2]
$A_{\mathbf{DN}}$	Digital number normalization area [m^2]
A_β	Beta nought normalization area [m^2]
A_σ	Sigma nought normalization area [m^2]; standard deviation
A_γ	Gamma nought normalization area [m^2]
A_r	Effective aperture area of the receiving antenna [m^2]
A_t	Effective aperture area of the transmitting antenna [m^2]
D	Antenna length [m]; guard
F	F value
G_r	Directivity gain of receiving antenna [-]
G_t	Directivity gain of transmitting antenna [-]
I	Intensity
K	Sentinel-1 calibration constant; number of groups

L	Pulse length [m]; looks; lines
M	Pixels in azimuth
N	Pixels in range; epochs
P	Phasor; pixels
P_r	Received power [W]
P_t	Transmitted power [W]
R	Distance antenna-target [m]
R_s	Slant range [m]
c	Speed of light [m/s]
$g^{+/-}$	Test statistic CUSUM
i	First iterator
j	Second iterator
k_s	TerraSAR-X calibration constant
m	Number of change points
n	Number of data points; input size
q	Quantile
s	Speckle
t	Travel time [s]; texture
v_s	Velocity on surface
$v_{s/c}$	Velocity of spacecraft
y	Time series/ sequence
$\mathbf{A}$	Design matrix
$\hat{\mathbf{x}}$	Estimator of unknowns
$\mathbf{y}$	Vector of observations

List of operators and annotations

$\{.\}^\intercal$	Transpose
$\{.\}^{-1}$	Inverse
$\{.,.\}$	Set
$[.,.]$	Closed interval
$(.,.)$	Open interval
$\|.\|$	Absolute value
$\|\|.\|\|^2$	Norm
$\sum(.)$	Summation operator
$\mathcal{O}(.)$	Computational cost

$C(.)$	Cost function
$E\{.\}$	Expectation operator
$f(.)$	Probability density function
$Im\{.\}$	Imaginary part
$Re\{.\}$	Real part
$\beta f(.)$	Penalty term
$\exp(.)$	Exponential function
$\max(.)$	Maximum operator
$\min(.)$	Minimum operator
$\arg\max(.)$	Arguments of the maxima operator
$\mod(.)$	Modulo operator
$\log(.)$	Log base ten operator

1 Introduction

1.1. Background

With humanity's ever growing desire for knowledge and its increasing technical capabilities, it is not surprising that some 600 Earth observation satellites orbit Earth as of 2018 (UCS, 2017). A classification for these satellites exists in the sense that it is either an imaging or a non-imaging satellite, referring to the capability of the instrument to generate multi-dimensional arrays of measurements. Furthermore, the sensor can be either of a passive nature, meaning it detects naturally reflected or emitted energy, or active, which detects reflected responses from self-emitted signals (Toan, 2007).

RADAR, the abbreviation for RAdio Detection And Ranging, is one of the techniques used to remotely monitor Earth. Over the years RADAR has lost all capitalization and is now just referred to as the noun *radar* (Philpot, 2015). The term radar is also used for the instrument used in these Earth observation satellites. As radar uses self-emitted electromagnetic pulses in the radio- and microwave frequency bands, it is classified as an active sensor. Both non-imaging and imaging sensors using the radar technique for Earth observation exist, e.g. Radar Altimeters (RA) and Syntheric Aperture Radar (SAR), respectively (Toan, 2007).

A coherent SAR system preserves both amplitude and phase measurements in a repetitive acquisition setting. Repeat-pass phase measurements have been used to measure sub-centimetre deformations and elevation since the late 1980's, with the well-known Interferometric Synthetic Aperture Radar (InSAR) technique (Hanssen, 2002; Van Leijen, 2014). Some 5-10 years prior, change detection in repeat-pass remotely sensed images of the same geographical area garnered attention (Singh, 1989). Since the late 1990's a rapidly growing interest has been observed in this particular topic, due to the increased number of satellites and the new policies for data distribution of old and new satellite missions (Bruzzone, 2016). Especially in recent years radar-based imagery specifically gained increasing importance in change detection applications (Ajadi et al., 2016). This surge in popularity can likely be traced back to the late 1990's when Dekker (1998) introduced the log-ratio operator, as previously no filter type was suited for SAR-based change detection. This operator became the foundation for many SAR change detection algorithms over the years and is still widely used.

Change point detection is a well researched topic in various fields, such as finance (Fernandez, 2004) and medical monitoring (Yang et al., 2006). Likewise, the detection of changes on Earth using space-borne imaging data is becoming indispensable. Applications such as environmental monitoring (Quegan et al., 2000), observation of natural disasters (Brunner et al., 2010), agricultural surveys (Bruzzone and Serpico, 1997) and studying human induced change, such as climate change (Bovolo et al., 2011; Reeves et al., 2007), all play a vital role in today's society.

Multiple remote sensing techniques have been successfully used in detecting changes, from optical to hyperspectral images. Bruzzone (2016) gives a good overview of how these different techniques are used in the field of change detection. A clear advantage of SAR over other techniques, such as optical sensors, is that SAR sensors are not hindered by atmospheric conditions or a lack of sunlight (Alphonse and Biju, 2015; Bruzzone and Prieto, 2000). Besides the all-weather mapping capability, the object penetrating properties of SAR can be extended to forest canopies for example (Mihai et al., 2015). This gives it an edge over other remote sensing techniques in certain applications, despite some of its drawbacks in mainly the highly unpredictive speckle. Some useful applications are listed in Table 1.1.

Table 1.1: Some applications especially suited for SAR time series change detection. For each application a general description and the expected temporal change behaviour is given in terms of familiar discontinuity functions.

Application	Description	Signal type
Natural disaster damage assessment	Certain natural disasters can cause massive structural devastation instantaneously. The damage caused by the likes of landslides, earthquakes, volcanic eruptions, hurricanes and so forth can be assessed to aid rescue operations. These types of events are likely to be accompanied by visual impairing atmospheric conditions.	Heaviside function
Flood mapping	Useful for aiding rescue operations, insurance and retrospective flood hazard zone identification. Again, flood events are likely combined with bad weather conditions, be it due to high spring tide or natural hazards induces, making optical sensors less viable for example.	Dirac function
Military and policing application	Using canopy penetrating radar (e.g. L-SAR) to automatically detect concealed objects in forested areas, such as military vehicles or drugs labs.	Rectangular function
Environmental change	Land use and abuse can be monitored. Topics such as natural deforestation or defoliation due to desertification or human induced changes, such as illegal logging, are becoming more important than ever.	Heaviside function

Most of the methods used for SAR-based change detection have one thing in common, namely the fact that the change detection is based on just two images (*bi-temporal*), one before the event and one after the event. Little research has been done in the field of *multi-temporal* SAR based change detection. Jianya et al. (2008) noted that "almost all classifications for change detection algorithms are based on bi-temporal change detection and little care for temporal trajectory analysis". More recently Dogan and Perissin (2010) mentioned that "there is still little effort about the application of the state-of-the-art techniques to the multi-temporal time series of SAR images within the change detection perspective".

1.2. Research objective

Increasing availability of SAR data offers the opportunity to study the temporal evolution of pixel values. A time series of SAR data holds valuable information on the long-term stability of pixel's backscatter response. Therefore, the objective of this research is to answer the main research question:

> *How can a SAR amplitude data cube be optimally used for automatic abrupt change detection in urban areas, in both an efficient and accurate manner?*

The main research question consists of five key components: SAR amplitude data cube, optimally, urban areas, efficient and accurate. The SAR amplitude data cube is the starting point of the change detection routine, and is used in this context as a stack of co-registered uncalibrated stack of SAR amplitude images. A focus is applied on urban areas, as these are often more troublesome for change detection. Not only is the range of backscatter values generally wider, but various radar inherent distortions, such as radar shadow and side lobes, are generally more prevalent in urban areas than rural areas. Lastly, optimality, efficiency and accuracy have a close relation to one another in the research question. Optimality is a multifaceted and fairly ambiguous term, hence the emphasis on efficiency and accuracy. An effort will be made to not only create reliable and accurate change maps, but also do this in an efficient way. This entails both computational and storage efficiency, as these become of greater importance with the increase in data volume. Descriptors such as computational complexity, underlying methodology of the change point detection method and various quality metrics are considered.

First, four boundary conditions are defined to give a clear direction to the research. The research will be limited to

1. only *pixel-level* methods, feature- and object-level will not be considered;
2. the change detection methods must only use *incoherent* (amplitude based) data, since the phase component doesn't hold that much information;
3. a change point detection method should be able to handle *multi-transitional* time series, that is to say the SAR time series have an unknown number of abrupt changes; and
4. a focus on a change in *mean* backscatter values is applied to dictate the change point detection, where the change points segment the data in sections of different ground states.

By taking the temporal component into account, some extra difficulties are introduced. In fact, it imposes three additional challenges besides dealing with the effects of speckle that are inherent to SAR data. How to deal with the much larger size of the data set? How to include new data, once available? And how to detect an unknown number of change points in a time series?

Looking at, and reflecting upon, different stages and aspects of the change detection procedure will help in answering the research question and how to effectively deal with the aforementioned challenges. Hence, five sub-questions are derived:

1. What are the statistical features of temporal SAR amplitude signals for different land covers?
2. What data handling and processing steps are important to increase the change detection accuracy and computational efficiency in a SAR data cube?
3. What are the current state of the art change detection techniques?
4. What is the performance of various methods for change detection in SAR time series?
5. How much do speckle and the spatial resolution influence the change point detection success rate?

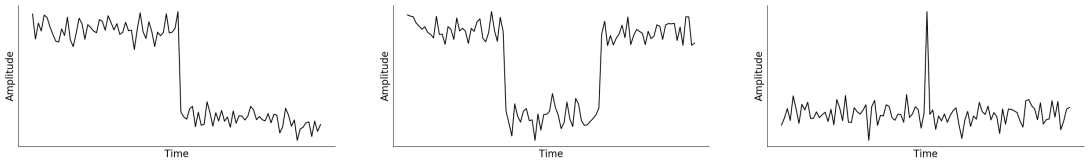
1.3. Methodology

The majority of the research will be focused towards creating software capable of efficiently processing and analysing a large data cube of SAR amplitude data for change points, with the addition of new data in mind. Coherent change point detection methods will not be considered. The software will consist of various elements, but is ultimately designed to detect *abrupt change behaviour in the time series*. When change detection is mentioned a definition by Bruzzone (2016) will be adopted: "[Change detection is] the process that analyses multi-temporal remote sensing images acquired on the same geographical area for identifying changes that occurred between the considered acquisition dates" (Bruzzone, 2016). More specifically, abrupt changes are linked to events such as rain and floods or more permanent changes such as deforestation and urban construction; in general events that cause step-behaviour in time series.

During the research three data sets will be considered, namely simulated data, a Sentinel-1 stack of Houston (the *Houston stack*) and a TerraSAR-X stack of Rotterdam (the *Rotterdam stack*). The Houston data stack will be the main focus of the research, for three reasons. First, TerraSAR-X data has already been extensively used in the change detection context, because of it's superior resolution over Sentinel-1 data. Secondly, Sentinel-1 data can be downloaded free of charge, whereas TerraSAR-X data has to be purchased. Lastly, because of the flood event caused by Hurricane Harvey on the 27th of August, 2017, which provides an interesting test case. Comparing the results for the Houston stack and Rotterdam stack gives an indication of the importance of resolution when using SAR data for change detection.

Both real data sets are available as single-look complex (SLC) data. To compare images taken at different acquisition dates, it is important to note that some pre-processing steps are needed. Every image has to be co-registered (i.e. align pixels to a reference image) and needs to be radiometrically corrected (i.e. reduce backscatter differences not due to target characteristics). These steps are critical in generating reliable change maps. The data is already co-registered using the DORIS InSAR processor, but still has to be radiometrically calibrated. An absolute calibration is performed in order to mitigate the effects of varying satellite geometry for example.

The aforementioned abrupt events are linked to discontinuities in the time series of the multi-temporal data, best described by a Heaviside step function in case of permanent change, see Figure 1.1a for a simulated example time series. A Rectangular function is linked to temporary change, where the mean of the initial



(a) Heaviside response. (b) Rectangular response. (c) Dirac response.

Figure 1.1: Three simulated discontinuity functions linked to change detection in SAR amplitude time series. A Heaviside response models permanent change, a Rectangular function best describes temporary change and a Dirac pulse is associated with a single change event.

group of values is the same as the mean of a third group, as shown in Figure 1.1b. However, this behaviour can also be described as the result of two equal and opposite Heaviside responses over time. Lastly, events resulting in short-term temporary change that are best modelled with a Dirac function are considered, such as floods (Figure 1.1c).

The time series of each pixel will be analysed for change points via a change point detection (CPD) algorithm. As the problem at hand has an unknown number of change points, the algorithm should be able to deal with this. A total of four CPD algorithms will be tested and evaluated with respect to computational efficiency and accuracy. The different change detection methods will be chosen based on comparative studies that focus on classifying the various methods, but the choice will be limited by four constraints mentioned in Section 1.2. This is to say that a change in mean for SAR amplitude time series dictates the detection of an unknown number of change points, while no spatial correlation is taken into account. Furthermore, the statistical features of the time series will be evaluated and used in order to determine which methods are best suited to the data. At this point in the processing chain the need for a temporal and/or spatial filter will be assessed as well. This will be done with some basic averaging filters, as actual filter selection and tuning is outside of the scope of this research.

Subsequently, the time series of each pixel will be segmented in different sections based on the change points detected by the CPD algorithms. The results of the different algorithms in various settings, e.g. temporary change versus permanent change will be compared and reflected upon. The change point solution can be stored and visualized in various ways. For each consecutive image pair a change map can now be created, based on the identified change points. To make planar partitions of changed and unchanged areas a clustering algorithm could be used. This will make sure closely distributed pixels that show similar behaviour are grouped together. Alternatively, a method is proposed that looks at the change point pixel density over time, in order to identify often changing regions.

To validate the results, two tests will be performed. Firstly, the algorithm will be tested on the previously mentioned simulated data stack which has a known truth. These simulated images will have varying and controlled levels of speckle noise added to them, as well as pre-determined change points with a temporal signature comparable to the subfigures in Figure 1.1. This will allow for the quantification of true positives, false positives, true negatives and false negatives. These can be used to define various quality metrics, that allow algorithm comparison in a more controlled environment. Secondly, the change detection maps of the Houston and Rotterdam stack will be interpreted by visual analysis and compared against optical images. To aid this an interactive tool will be developed that allows quick analysis of the change maps over time.

A large emphasis is placed on the temporal evolution of the SAR data. This imposes the practical consideration of how to handle data and change point storage. Hence, the incorporation of a structured file format in the processing chain is evaluated.

1.4. Outline

The research will include different components, starting with a comprehensive literature review. The first part of the review, Chapter 2, gives a brief overview of the RAdio Detection and Ranging (RADAR) technique. Here, space-borne SAR is the main focus. The overview of radar and SAR will focus on the elements that are important to understanding (multi-temporal) SAR based change detection, such as the statistical models used for SAR amplitude data. Moreover, the calibration of SLC data, data storage and the SAR data products are discussed. Subjects such as image co-registration will not be considered.

Chapter 3 focuses on change point detection. An overview of existing change detection methodologies and techniques is given. Both change detection methods applied on SAR data and methods used in other

fields will be considered, to offer a broad perspective on the subject. Some quality metrics will be introduced in order to help compare the different change point algorithms in a later stage. The findings in this chapter will guide the selection of change detection algorithms and quality assessment methodology.

Chapter 4 outlines the implementation of the change point detection algorithms. The entire processing cycle is discussed, starting with the selection of the test sites, as well as the methodology of the pre-processing steps that are performed. Four algorithms are implemented and compared, (simple) Differencing, CUSUM, ANOVA and PELT. The implementation of each method is described, alongside the reasoning behind the selection of these algorithms. Moreover, the threshold selection and methodology of evaluation will be explained.

The results of the application of the algorithms on the various data sets are shown and discussed in Chapter 5. Here, the strengths and shortcomings of the different change detection methods are discussed. Both pixel-level results and spatial solutions are shown. This will eventually result in an assessment of each algorithm, in terms of detection accuracy.

Many of these observations are further discussed and compared in Chapter 6. Here, conclusions regarding the optimal change detection approach in multi-temporal SAR imagery are offered. All this combined allows for the formulation of recommendations for further research following the drawn conclusions.

2

Principles of space-borne imaging radar

In this chapter space-borne radar imagery is discussed. A strong focus is on the parts that are vital to proper change point detection in SAR based time series. Topics such as radar geometry, spatial resolution, speckle and the statistical models for SAR backscatter are the focal point. However, the basics and required knowledge in terminology are covered as well. Lastly, pre-processing and storage in structured file arrays of SAR data is also discussed.

2.1. Basics of SAR imaging radar

2.1.1. Radar geometry

Space-borne imaging radar systems often have a side-looking viewing geometry, which is different from some framing and scanning systems in optical remote sensing (NRCAN, 2016). With radar, there are two important axes and a point that make up a sort of 'coordinate system', being the azimuth direction, range direction and nadir point, as seen in Figure 2.1. Azimuth refers to the line parallel to the flight direction of the satellite, range to the across-track dimension and nadir is defined as the point that intersects with the surface directly beneath the spacecraft. For the range dimension, a differentiation between slant range and ground range exists; the former refers to the radial line of sight distance between the radar and some observation point and the latter to these measurements in slant range projected on the Earth's surface. Furthermore, the point in a certain radar swath that is closest to the nadir track of the satellite is referred to as near range and the opposite holds for far range. This means that the angle between the incoming radar beam and the normal to the Earth's surface isn't constant over the swath. This observation is described by the look angle, which increases a couple degrees from near to far range. However, often the *local* incidence angle (LIA) is used, as to account for curvature in the observed surface area (ESA, 2014).

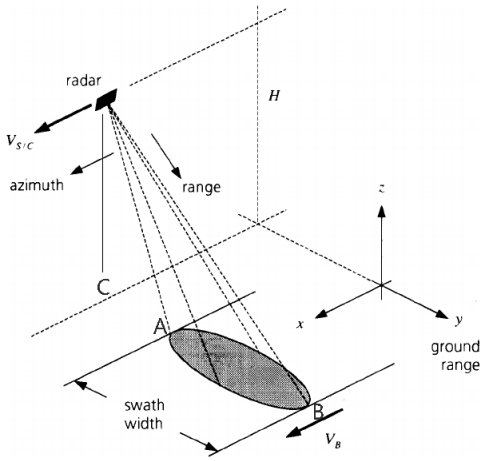


Figure 2.1: A schematic overview of radar geometry showing azimuth and range direction, among other geometry related concepts (Bamler, 2015) [A: near range, B: far range, C: nadir, H: height, $v_{S/C}$: velocity of spacecraft, v_S: velocity on surface].

Many different satellite orbits can be employed, depending on the application of the satellite. Imaging radar satellites are in a *sun-synchronous Low Earth Orbit* (SS-LEO), which means the satellites are at an altitude less than 2000 km above the Earth's surface in a near polar orbit (see Figure 2.2). The Sentinel-1 satellites from the Copernicus program were deployed at an altitude of 693 km (EO, 2002a) and the TerraSAR-X at 514 km (EO, 2002b), for example. This is done to ensure the most constant sunlight conditions possible, as radar imaging satellites are active systems. Figure 2.2 also shows the *ascending* and *descending* passes, referring to the northwards and southwards travel paths respectively. Lastly, each satellite has a certain *repeat (orbit) cycle*, defined as the time it takes for the nadir point of the satellite to pass over the same point on the Earth's surface. This is 12 (EO, 2002a) and 11 (EO, 2002b) days for Sentinel-1 and TerraSAR-x respectively. However, both Sentinel-1 and TerraSAR-X are a two-satellite constellation. Sentinel-1 consists of Sentinel-1A and Sentinel-1B, with a 180^o oribital phasing difference, effectively reducing the repeat cycle to 6 days. The TerraSAR-X and TanDEM-X are only a few hundred metres apart to allow for the generation of a Digital Elevation Model (DEM), hence not decreasing its repeat cycle. The repeat cycle shouldn't be confused with the *revisit time*; the elapsed time between observations of the same area, which is also dependent on e.g. swath widths (EO, 2002a).

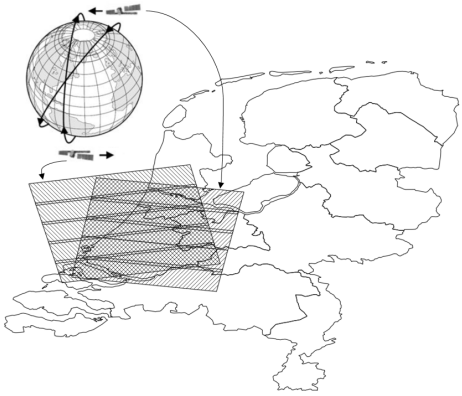


Figure 2.2: Satellite orbit geometry [Top satellite: descending (west looking), Bottom satellite: ascending (east looking), Ground swaths: two S1 sub-swaths of a couple of bursts (see Section 2.3.2)].

The side-looking geometry causes distortions in the radar images, as seen in Figure 2.3. Three objects are shown, A, B and C. The first being nearly flat, the second tall and somewhat symmetrical in its cross section and the last skewed to the left. At object B the radar beam will first hit point 4 and hits point 3 shortly after that. With both measurements projected to the slant range domain, point 4 is mapped before point 3 in the image. In other words the tall object tilts towards the radar (Mather and Brandt, 2001). This phenomenon is referred to as layover and often appears as bright spots on a radar image. Object A is subject to foreshortening. Because the angle between the incoming radar beam and the slope of points 1 and 2 is less than 90 degrees, the projected points 1 and 2 are compressed in the slant range dimension (Mather and Brandt, 2001). In other words; foreshortening. Furthermore, foreshortening also occurs when the base of a tall object is reached before the radar beam reaches the top (NRCAN, 2015). Foreshortened slopes often appear as bright spots in a radar image. The effects of layover and foreshortening are most prominent near range, where the incidence angle is small. Lastly, the phenomenon of radar shadow in object C. Shadow is simply the inability of a radar beam to reach certain parts of the surface due to objects blocking the radars line of sight, which is the case between points 6 and 7. Unlike foreshortening and layover, radar shadow is most prominent in far range situations and appears as very dark pixels in a radar images (NRCAN, 2015).

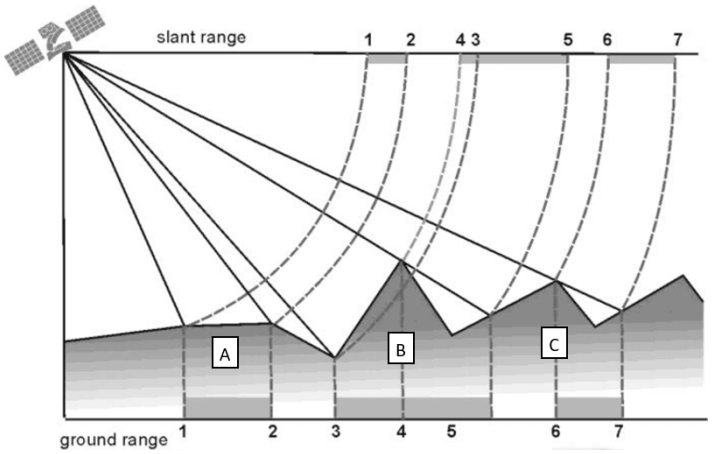


Figure 2.3: The layover (B), foreshortening (A) and shadow (C) effects in radar imagery. These distortions are caused by the side-looking geometry of radar satellites (Toth et al., 2014).

2.1.2. Synthetic Aperture Radar

To understand the SAR imaging radar, it is best to start the Real Aperture Radar (RAR). A RAR is a *monostatic pulsed* radar that emits narrow beams of energy perpendicular to the azimuth direction of the satellite. During a short time interval a pulse with a certain frequency is emitted by the transmitter of the radar. This wave of energy will reach Earth and interact with the surface. Part of the energy is scattered back to the radar ('backscatter'), which is now in receiver mode to record the echo (Bamler, 2015). For the received signal of an imaging radar three types of scattering mechanism are of importance; surface scattering (σ_s), also known as the 'single bounce', the dihedral return (σ_t), or 'double bounce', and volume scattering (σ_v) (Halounova, 2009). The relation between the transmitted and received power, for a specific target and antenna type is given by the *radar equation* (Hanssen, 2017)

$$P_r = \frac{A_r G_t P_t}{(4\pi)^2 R^4} \sigma^0 A. \tag{2.1}$$

Here P_r is the received power [W], P_t the transmitted power [W], A_r the effective aperture area of the receiver antenna [m^2], G_t the directivity gain of the transmitting antenna [-], R the distance between the antenna and the target [m], σ^0 the normalized radar cross section (RCS) [-] and A the scattering area illuminated by a single pulse [m^2]. In Equation (2.1) the range, gain and normalized RCS are defined as

$$R = \frac{ct}{2}, \tag{2.2}$$

$$G_t = \frac{4\pi A_t}{\lambda^2} \tag{2.3}$$

and

$$\sigma^0 = \frac{\sigma}{A}, \tag{2.4}$$

where c is the speed of light [m/s], t the travel time [s], A_t the effective aperture area of the transmitting antenna [m^2], λ is the radar wavelength [m] and σ the RCS of the target [m^2]. In the case of a monostatic radar $A_t = A_r$ and $G_t = G_r = G$ (Philpot, 2015).

The side-looking geometry of the RAR provides sensitivity in the range direction and avoids ambiguous reflections that would occur at a down-looking oriented setup (Hanssen, 2002). The reflected relative intensity from a near range point A to a far range point B (Figure 2.1) is then recorded to create a strip of measurements in range direction (ESA, c). This sequence is repeated at time intervals determined by the pulse repetition frequency (PRF), with the satellite continuously flying in azimuth direction, resulting in a 2D image of the sensed area. Figure 2.4 shows how the PRF and pulse width are related to the signal emitted by pulsed radars.

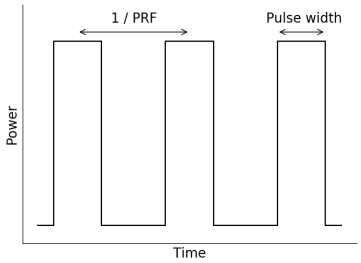


Figure 2.4: A pulsed radar signal showing the pulse repetition frequency and pulse width with respect to the power over time.

The ability of an imaging radar to separate two distinct objects at some minimal distance apart is given by its *spatial resolution*. The spatial resolution, or resolution for short, of a RAR in range and azimuth direction are inversely proportional to half the pulse length

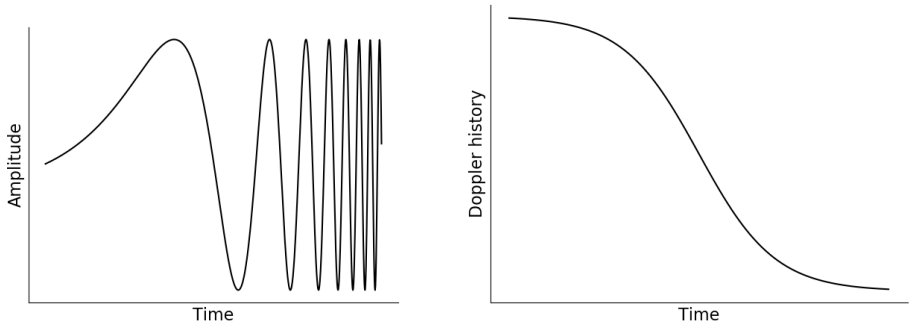
$$\Delta_r^{\text{RAR}} \approx \frac{c\tau}{2} = \frac{L}{2} \tag{2.5}$$

and dependent on the beam width, which is a function of wavelength and antenna dimension, and slant range

$$\Delta_a^{\text{RAR}} \approx \beta R_s = \frac{\lambda R_s}{D} \tag{2.6}$$

respectively (ESA, c; Hanssen, 2017). In Equations (2.5) and (2.6) c is again the speed of light [m/s], τ the pulse width [s], L the pulse length [m], β the beam width [rad], R_s the slant range [m], λ the wavelength [m] and D the antenna length [m]. From Equation (2.6) it becomes clear that the azimuthal resolution degrades with range, as it has a slant range dependency. Both the range and azimuth resolution for a RAR are in the order of 10^3 m, clearly not ideal for proper change detection.

In order to achieve an acceptable range resolution in the order of metres, one would need to decrease the pulse width of current radar systems, which is in the order of 10 µs, with a factor of three. This would mean



(a) Frequency modulated signal (chirp). (b) Doppler frequency history.

Figure 2.5: Visualization of a chirp (Figure 2.5a) and the Doppler effect (Figure 2.5b). The former shows an increase in frequency over time and the latter shows the decrease in frequency over time by changing from approaching to receding from an object.

that the same power has to emitted in a pulse of a couple of nanoseconds, which imposes a physical barrier (Philpot, 2015). Furthermore, a decrease in the pulse width results in an increase in system bandwidth, as these are inversely proportional. With the noise power of a radar system being dependent on the bandwidth, a decrease in pulse width results in a decrease in Signal-to-Noise ratio (SNR) (Hanssen, 2017).

Hence, frequency modulation (FM) is used to achieve *synthetic* shortening of the pulse length in the form of a chirp. The chirp has an increasing frequency over the pulse interval, as can be seen in Figure 2.5a. As this pulse compressed signal has a distinct signature, it can still be discriminated from other incoming signals via a matched filter (Hanssen, 2017; Philpot, 2015). The synthetic pulse now has a width of a couple of nanoseconds, resulting in a range resolution in the order of metres. The use of FM signals to improve the range resolution is called range compression (Johannessen and F., 2013). It has to be noted that the resolution in range direction has a range dependency as well, governed by the incidence angle Φ (Johannessen and F., 2013)

$$\Delta_r^{\text{SAR}} \approx \frac{L}{2\sin(\Phi)}. \tag{2.7}$$

Similarly, an improvement in azimuthal resolution is achieved by creating a large *synthetic* antenna via the principle of *synthetic aperture*, hence the name Synthetic Aperture Radar. As the wavelength for SAR imaging systems are in the range of centimetres and the slant range in the upper hundreds of kilometres; an antenna size of some ten kilometre would be required for a resolution in the range of metres. Obviously, this is not possible. In coherent radars, which receive and preserve amplitude and phase of the return signal (Hanssen, 2002), it is possible to create artificially long antennas by exploiting the Doppler effect. Due to the moving nature of the satellite, a signal returned from a point ahead of the satellite will have a higher frequency than the actual emitted signal and vice versa (Figure 2.5b). Knowing this, one can compensate for the Doppler shifts of a single scatterer over the course of multiple acquisition moments for that particular scatterer. This will result in a measurement of a single point consisting of multiple looks from different angles, sorted by frequency (Philpot, 2015). This process is often referred to as azimuth compression (Johannessen and F., 2013). Elaborating the equation for the Doppler frequency shift, the azimuthal resolution can be obtained as

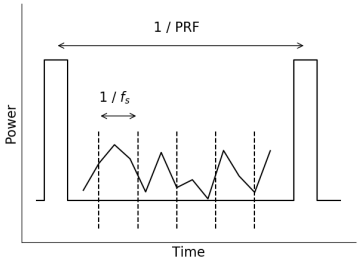


Figure 2.6: The pulse repetition interval and range sampling frequency in relation to the pulsed radar signal first shown in Figure 2.4.

$$\Delta_a^{\text{SAR}} \approx \frac{D}{2}. \tag{2.8}$$

Following Equations (2.7) and (2.8) it can be concluded that both the range and azimuth resolution of a SAR are not influenced by the height of the spacecraft (Hanssen, 2017), which is unique to SAR (Johannessen and F., 2013). Moreover, the spatial resolution has been improved by a factor of three. It has to be noted, however, that the azimuth compression is only effective if the radar has a sufficiently small beam width, that meets the small angle approximation. With the small angle approximation, the radar beam can be approximated as a straight line. In the so-called *unfocused* SAR mode, the beam does not meet this requirement and the

processing is limited to a subset of the azimuth phase history as not to introduce errors, resulting in a coarser azimuth resolution that is still dependent on range (Philpot, 2015)

$$\Delta_a^{\text{SAR,(unfocused)}} \approx \sqrt{\frac{\lambda R}{2}}. \tag{2.9}$$

The limits of the unfocused SAR are governed by the range walk. A range walk occurs when the range difference at the antenna between the range at the normal of the antenna and some other point exceeds the resolution cell size, resulting in an unfocused image (Chen, 2005). In *focused* SAR the range walks are effectively removed by accounting for the curvature in the wave front (Philpot, 2015), which allows a SAR to reach its full potential (Equation (2.8)). In other words; the received data is phase-compensated (Chen, 2005).

A property closely related to resolution, pixel size, is often mistaken for resolution. In the context of SAR, a pixel refers to a measurement in a 2D image stored in an infinitesimally small point (Toan, 2007), (Hanssen, 2002). Hence, pixels don't have a physical size. A pixel 'size' is an arbitrarily chosen sampling interval in the final SAR image (Bamler, 2015). Therefore, the pixel spacing is often used as a sort of measure for spatial distance. This refers to the distance between two pixel-nodes and is also called 'posting' (Hanssen, 2017). The pixel spacing is determined by the pulse repetition interval (PRI) and by the radar sampling frequency (f_s) for the azimuth and range direction respectively (Airbus, 2014b), as shown in Figure 2.6.

2.1.3. Radar signal characteristics

A SAR image is the result of the travel time to the Earth's surface and the strength of the return signals. There are various parameters that influence the incoming signal, with the satellite range and the roughness and dielectric properties of the surface being the major contributors (Hanssen, 2002). However, the local incidence angle, scatterer density and 3D distribution of the scatters are also of importance (Bamler, 2015). The strength of the return signal is often represented by the *amplitude* (A) or intensity (I), with $I = A^2$. The time delay gives a measure of the fractional *phase* (ψ) of the received signal. The amplitude and phase are the two components for the complex *phasor* (P); the radar measurement per pixel (Van Leijen, 2014). A phasor P has a real (Re) and imaginary (Im) part and is formulated as

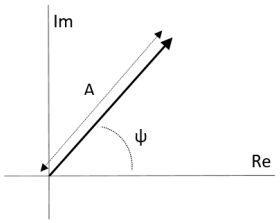


Figure 2.7: The relation between the phasor, amplitude and phase in the complex plane.

$$P = A\exp(i\psi), \tag{2.10}$$

with the amplitude and phase defined as

$$A = \sqrt{Re\{P\}^2 + Im\{P\}^2} \tag{2.11}$$

and

$$\psi = \arctan(\frac{Im\{P\}}{Re\{P\}}) \tag{2.12}$$

such that,

$$Re\{P\} = A\cos(\psi) \tag{2.13}$$

and

$$Im\{P\} = A\sin(\psi). \tag{2.14}$$

The relation between P, A and ψ in the complex plane is shown in Figure 2.7.

A multitude of SAR satellites orbit Earth in a sun-synchronous LEO. The predominant wavelengths used in SAR are in the L, C and X frequency bands (Table 2.1), often abbreviated as e.g. C-SAR. Because these frequency bands reside in the radio spectrum (L-band) and microwave frequency range (C-band and X-band), a SAR imaging radar is able to acquire data acquisitions independent of cloud coverage and time of day. As can be seen in Figure 2.8, the L-, C- and X-band are all located in an atmospheric transmission window.

Table 2.1: The predominant frequency bands of SAR. L-SAR, C-SAR and X-SAR with their respective frequency and wavelength range.

Band	Frequency [GHz]	Wavelength [cm]
L	1-2	15-30
C	4-8	3.75-7.5
X	8-12	2.5-3.75

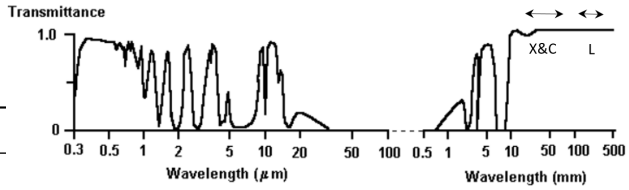


Figure 2.8: The atmospheric transmittance windows (Liew, 2001), with the X-, C- and L-band highlighted.

2.1.4. Coherence and speckle

A very crude dichotomy of reflected signals can be made, based on the amount of scatterers, or objects, contributing to the measurement. On the one hand there is the *point scatterer* and on the other hand *distributed scattering* (also referred to as Rayleigh or Gaussian scattering) (Hanssen, 2002), as seen in Figure 2.9. If the reflection is dominated by a single strong reflector, it is referred to as a point scatterer. Contrariwise, distributed scattering occurs when many small scattering objects contribute to the radar measurement (Van Leijen, 2014). Very strong point scatterers may introduce significant *side lobes*. These side lobes manifest itself in the radar amplitude image as a contamination in neighbouring resolution cells in both range and azimuth direction, as well as contaminate the phase observations (Hanssen, 2002). Side lobes are caused by the spatial signature of the point scatterer; a sinc function (Van Leijen, 2014).

All the contributions of the scatterers in a resolution cell will add up coherently to a final pixel value (Bamler, 2015). If a single pixel, be it a point scatterer or distributed scatterer, is observed with multiple acquisitions over time, a second subdivision can be made. If the response of the sensed objects within a resolution cell doesn't change significantly over time, it is labelled as a coherent scatterer. Contrary to coherent scatterers, incoherent scatterers display a significant phase change in measured radar backscatter over time. This (in)stability of a single pixel over time is described by its coherence.

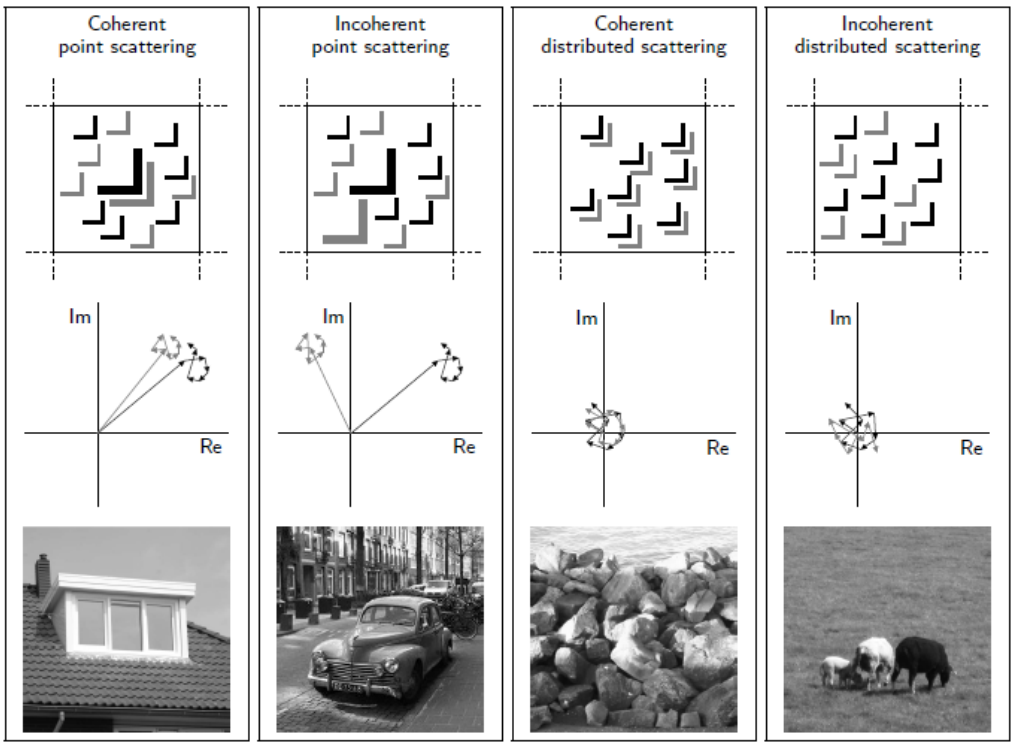
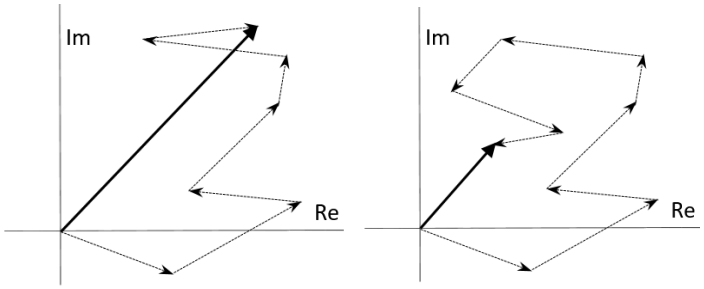


Figure 2.9: The differences between point and distributed scattering in a coherent and incoherent manner (Van Leijen, 2014) [Top row: individual scatterers in a single resolution cell at two different epochs, Middle row: the phasor sum of all scatterers, Bottom row: example images].

Due to the finite SAR resolution, multiple scatterers with different phases will be present in a single resolution cell (Toan, 2007). As stated before, a SAR is a coherent imager, meaning it images the summation of all the individual scatter responses (see middle row of Figure 2.9) in that particular resolution cell. The scattered

signals from all objects within the resolution cell add coherently based on the relative phase of each scatterer's waveform in either a constructive (Figure 2.10a) or destructive (Figure 2.10b) way (Toan, 2007). This highly unpredictable interference of multiple scatterers, results in a granular noise-like effect, called *speckle* (Hanssen, 2002). This phenomenon is most pronounced in distributed scatterers and inherent to coherent imaging methods (Bamler, 2015). It will manifest itself in radar images as a grainy 'salt and pepper' texture.



(a) Constructive interference. (b) Destructive interference.

Figure 2.10: The unpredictable interference of multiple scatterers, causing a salt and pepper texture in SAR images referred to as speckle. Figure 2.10a shows constructive interference and Figure 2.10b destructive interference.

2.1.5. Speckle reduction

All radar images will be subject to speckle to a certain degree. The random nature of this phenomenon makes interpreting and analysing radar images difficult, hence it is desirable to reduce the effects of speckle as much as possible. This is done by *multi-look processing*, or 'multi-looking', and *filtering*. Multi-look processing will be briefly touched upon and filtering will be discussed in a bit more detail.

Everything discussed so far was in the context of single-look SAR imagery. Multi-look processing is a form of non-coherent averaging, by averaging over range and/or azimuth resolution cells (Halounova, 2009). Rather than processing the single-look of a radar beam, it is divided into several sub-beams (Figure 2.11), each processed separately. Each of these *looks* (L) provide an echo of the same ground area, but will still appear with some degree of speckle. However, by summing and averaging the sub-images to a final image, the speckle will be reduced

$$A = \frac{1}{L}\sum_{i=1}^{L} A^i. \tag{2.15}$$

On the downside, the geometric azimuthal resolution is reduced as the effective azimuthal spectrum beam width is smaller (Philpot, 2015). It differs from filtering, as it is an in-process method, rather than a post image formation method (Gagnon and Jouan, 1999). Lastly, with an increasing number of looks the PDF for the amplitude become more narrow (Toan, 2007). For more on the subject, see e.g. Cantalloube and Nahum (2000).

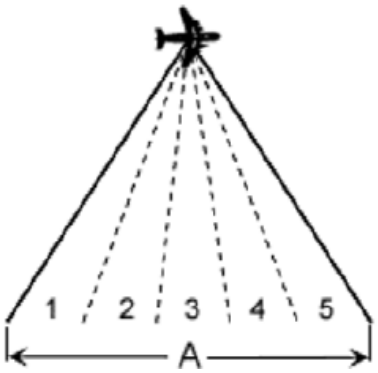


Figure 2.11: A visualization of multi-look processing, a form of non-coherent averaging, with azimuthal bandwidth division in L=5 looks (Philpot, 2015) [A: azimuth].

Filtering on the other hand is done post image formation. A wide variety of filters have been developed with varying degrees of computational complexity and adaptively. Gagnon and Jouan (1999) provides a great overview and comparison of many of the standard (spatial) and more complex-wavelet-based (frequency) filters used in SAR imagery. One of the most standard spatial filters uses a moving window of a couple of pixels in range and azimuth and applies some calculation using those pixels. For example a 3x3 (NxM) moving window that calculates the average of the windowed pixels (Figure 2.12a)

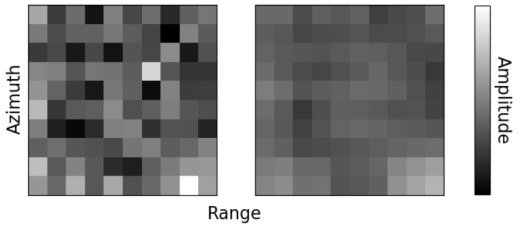
$$A = \frac{1}{NM}\sum_{i=1}^{NM} A^i. \tag{2.16}$$

In Equation (2.16) the central pixel will take the value of the calculation under the window. By doing this for the whole image, the speckle is (visually) reduced by this smoothing operation.

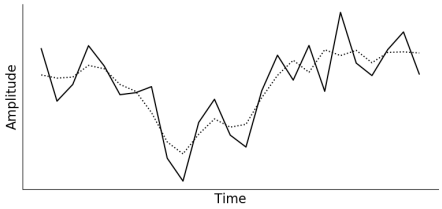
When multiple acquisitions over time are collected it is possible to filter in the time domain. Analog to Equation (2.16), one could create a temporal filter over N=3 consecutive acquisitions that calculates the smoothened value for the central acquisition (Figure 2.12b)

$$A = \frac{1}{N}\sum_{i=1}^{N} A^i. \tag{2.17}$$

These types of filters are referred to as a *moving average* or *rolling mean* and can be expanded by adding weights to pixel values for example. Albeit being very basic filters, they are very quick to compute and require no knowledge of the data set, while reducing the speckle to some degree.



(a) Spatially filtered [Left: original image, Right: filtered image].



(b) Temporally filtered [Solid line: original time series, Dotted line: filtered time series].

Figure 2.12: A moving average filter in the spatial (Figure 2.12a) and temporal (Figure 2.12b) domain. These kind of filters are applied to limit the effects of speckle [Left: window of NxM=3x3, Right: window of N=3].

2.1.6. Decibel notation

Working with decibels (dB) is very common in radar imagery. Besides some equations that require linear values, one often finds characteristics of the intensity of a return signal expressed in dB (Hanssen, 2017). The accuracy and noise levels for various SAR satellite missions, as well as the backscatter characteristics of corner reflectors and transponders, are usually expressed in dB for example. This logarithmic scale allows us to describe very small and big numbers with shorter notation. This relation is given as

$$P_{\mathrm{dB}} = 10\log_{10}(P), \tag{2.18}$$

or inverse as

$$P = 10^{\left(\frac{P_{\mathrm{dB}}}{10}\right)}. \tag{2.19}$$

Figure 2.13 shows the benefits of writing the linear response to the logarithmic domain, as it becomes clear right away that the dB amplitude values have a better effective dynamic range

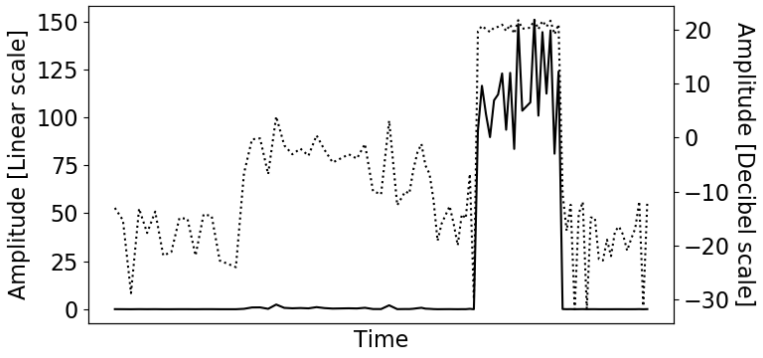


Figure 2.13: A σ^0 calibrated Sentinel-1 time series on a linear and decibel scale. The decibel notation shows a much better effective dynamic range [Solid line: amplitude values on a linear scale, Dotted line: amplitude values on a decibel scale].

2.2. Statistical models for SAR backscatter

To model the statistics of the complex backscattered signal, a number of assumptions of the sensed area have to be made. Ultimately, the characterizations of the phenomena occurring in the area dictate the model. The *speckle model* proposed by Arsenault and April (1976) is the most common statistical model for SAR backscatter and assumes that (Gao, 2010; Hanssen, 2002; Madsen, 1986; Moser et al., 2006):

- Many scatterers contribute to the radar measurement of a particular cell. With a single resolution cell being at least one order of magnitude larger than the wavelength of the radar, this is a fair assumption.
- No dominant scatterer is present. This assumption generally holds for many natural scatterers such as forests and deserts.
- The amplitude and phase of the scatterers are statistically independent random variables. A valid assumption, because the phase delay is independent of the magnitude or the return signal.
- The phase of each individual scatterer is uniformly distributed in [-π,π) and uncorrelated with other scatterers. The correlated scatterers come together in several uncorrelated scatter centers, so this is a reasonable assumption.

When these assumptions hold, it is possible to apply the Central Limit Theorem (CLT), which tells us that the real and imaginary part of the complex backscatter both follow a *Gaussian probability density function* (PDF) (Lopez-Martinez, 2013) and the observations are complex circular Gaussian random variables (Hanssen, 2002). Furthermore, the real and imaginary components are identically Gaussian distributed with a zero mean and equal variances, such that (Bhattacharya; Moser et al., 2006)

$$f(Re\{P\}) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{Re\{P\}^2}{2\sigma^2}\right) \tag{2.20}$$

and

$$f(Im\{P\}) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{Im\{P\}^2}{2\sigma^2}\right), \tag{2.21}$$

which gives the PDF of the (complex circular Gaussian) observation P, or in other words, the joint PDF of Equations (2.20) and (2.21), as

$$f(P) = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{Re\{P\}^2 + Im\{P\}^2}{2\sigma^2}\right). \tag{2.22}$$

For SAR backscatter $2\sigma^2$ is often replaced by E{I}, the expectation of the intensity value of the resolution cell (Hanssen, 2002). A theoretical distription of E{I} can be found in Madsen (1986).

Following Equation (2.22) and Equations (2.13) and (2.14) from Section 2.1.2, the joint PDF can be written as a function of (positive) amplitude and phase ([-π,π))

$$f(A;\psi) = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{A^2}{2\sigma^2}\right), \tag{2.23}$$

which results in the PDF's for the amplitude and phase, by integrating out ψ, between -π and π, and A, between 0 and ∞, respectively. Therefore, the PDF for the amplitude can be written as

$$f(A) = \frac{2A}{\sigma^2} \exp\left(-\frac{A^2}{\sigma^2}\right) \tag{2.24}$$

and for the phase as

$$f(\psi) = \frac{1}{2\pi}. \tag{2.25}$$

Hence, the amplitude can be modelled by the *Rayleigh probability distribution* and the phase follows a *uniform probability distribution*, which contains no information (Lopez-Martinez, 2013). The intensity, under the previously mentioned hypotheses, follows a *negative exponential probability distribution*

$$f(I) = \frac{1}{\sigma^2} \exp\left(-\frac{I}{\sigma^2}\right). \tag{2.26}$$

These distributions represent the *fully developed speckle* case, where the number of scatterers is very large. Modelling for *partially developed speckle* is much more difficult, to the point where even numerical solutions are troublesome (Lopez-Martinez, 2013).

In reality though, SAR amplitude data often follows non-Rayleigh empirical distributions. One of the assumptions noted the absence of point scatterers, which is not a valid assumption for urban areas for example. The *product model*, first proposed by Ward (1981), is a generalization of the speckle model that is applicable to cases where the RCS is not constant. It relates the RCS or 'texture' (t), the noise-less amplitude value, and uncorrelated speckle component (s) to the intensity as

$$I = t \cdot s \tag{2.27}$$

and shows the multiplicative nature of speckle. The product model is often used to develop other statistical models for SAR imagery that take RCS fluctuations in consideration (Gao, 2010).

Research in the field of non-Rayleigh SAR amplitude statistics has been comprehensively studied and many alternative distributions applicable to a particular set of rules have been proposed, such as the Rice/ Rician- and K-distribution (Eltoft, 2013; Lopez-Martinez, 2013). The former is used in the case of single dominant point scatterer (deterministic scattering) and the latter when the number of scatterers is seen as a stochastic quantity (Lopez-Martinez, 2013). However, it is argued that the complex nature of these distributions limit their practical applicability (Eltoft, 2013). In particular Moser et al. (2006) elaborates a method that models the statistics of a wide variety of SAR amplitude distributions in a numerically feasible and analytically consistent way, based on a Generalized Gaussian model. A comprehensive survey of the statistical models used in SAR imagery for different applications is carried out by Gao (2010).

2.3. SAR data products

2.3.1. Polarimetry

Since a SAR uses electromagnetic radiation to illuminate a surface, it can exploit polarization signatures of the imaged surface (Bamler, 2015). Both co-polarization (identical transmit and receive polarizations) and cross-polarization (transmitting in one polarization and receiving in a orthogonal polarization state) are used with space-borne SAR. The polarization is often denoted as a two-letter abbreviation with the transmit polarization first, followed by the receive polarization (e.g. VV or VH). The particular polarization is defined by the *polarization mode* of the satellite. SAR satellites operate in a single-pol, dual-pol or quad-pol polarization system, referring to the number of transmit-receive combinations the satellite is able to manage. A dual-pol system could transmit a horizontally polarized waveform and receive both the vertically and horizontally polarized return signals, resulting in HH and HV imagery (JPL).

Different polarizations are sensitive to different cases of scattering mechanisms, for example cross- polarization (e.g. HV) is sensitive to volume scattering (Bamler, 2015). On the other hand, co-polarization is more useful to distinguish flooded regions from non-flooded areas. A HH polarized signal known to be useful for mapping flooded areas in between vegetation, as it has good canopy penetrating properties and a lower backscatter in rough open water bodies than VV, making it less sensitive to wind conditions (Balanos et al., 2016). For detecting partially submerged features, VV polarized signals are very useful (Manjusree et al., 2012). It has to be noted that there is a clear distinction in the amount of mean backscatter for different types of water bodies. For example, flood water returns a much stronger signal than water from a river (Manjusree et al., 2012). This variable multi-polarization signature can be exploited in the change detection field as well, as shown in Manjusree et al. (2012) and Balanos et al. (2016) for example.

2.3.2. Core products for Sentinel-1 and TerraSAR-X

So far, Section 2.1.1 and 2.1.2 were discussed in the context of the 'conventional' strip mapping mode, where an image is formed over the entire width of the swath (see Figure 2.1). However, a SAR often has multiple *acquisition modes* or *imaging modes*. The imaging mode for Sentinel-1 and TerraSAR-X SAR data used in this paper will be discussed briefly. For Sentinel-1 the StripMap, Interferometric Wide swath (IW), Extra-Wide swath (EW) and Wave (WV) imaging modes are operative (ESA, a). Meanwhile, the TerraSAR-X products are acquired in either Staring Spotlight (ST), High Resolution Spotlight 300 MHz (HS300), High Resolution Spotlight (HS), SpotLight (SL), StripMap (SM), ScanSAR (SC) or Wide ScanSAR (WS) imaging mode (Airbus, 2014b). For Sentinel-1, the most used mode over land is IW. This imaging mode is able to operate in single-pol and dual-pol. IW acquires three 80-90 km wide (sub-)swaths in the range direction, resulting in a swath 251.8 km wide. Each of these swaths is divided by a series of bursts along the entire width of the swath, which are

all processed separately (ESA, a). The SM TerraSAR-X imaging mode also provides single-pol and dual-pol products, with a slightly lower spatial resolution for the dual-pol product (Airbus, 2014b).

The products for SAR-based imagery are provided as a level-0, level-1 or level-2 product. The level-0 product contains the raw compressed and unfocused SAR images and form the basis for the other two products. On the other hand there is the level-2 products, which are geo-located geophysical products derived from level-1 products and provide surface wind measurements for example. The most versatile product is the level-1 product, which has been subject to range and azimuth compression, as well as some other (post-)processing algorithms. This now focused SAR data is available in slant and ground range geometry; Single-Look Complex (SLC) and Ground Range Detected (GRD) respectively. The GRD product is projected to ground range using some terrain corrected ellipsoidal model (e.g. a terrain corrected WGS84 model). This product is also multi-looked, which is briefly discussed in Section 2.1.5. For observations in the natural radar range dimension in a Zero-Doppler orientation, the SLC product is provided. Each row of pixels corresponds with a line orthogonal to the sub-satellite track (ESA, b). The spatial resolution for Sentinel-1 IW SLC data is ~5x20 m and ~3x3 m for SM TerraSAR-X (Airbus, 2014b; ESA, a).

2.4. Calibration of SLC data

2.4.1. Absolute and relative radiometric calibration

To compare images at different acquisition dates, it is important to make sure these images are radiometrically calibrated. Calibration is needed for various reasons, such as a potential offset between twin satellites that generate SAR products (e.g. Sentinel-1A and Sentinel-1B or TerraSAR-X and TanDEM-X), comparing data with a different geometry (e.g. due to ascending-descending tracks, opposite look directions or different incidence angles) and comparing SAR measurements between different satellite missions (Airbus, 2014a). In general, radiometric calibration attempts to diminish the effects of contributions to the radiometric values that aren't due to the target characteristics. Effectively this means that radiometric calibration can be seen as an image normalization (Ajadi et al., 2016).

To produce radiometrically calibrated images, one can choose to *relatively calibrate* or *absolutely calibrate* radar data. For the relative radiometric calibration some image in a data stack is chosen as the point of reference. The remaining images are then normalized to the reference image selected by the user, hence not requiring any measurements of atmospheric and sensor conditions. The relative calibration is also referred to as relative radiometric normalization (RRN) (Yang and Lo, 2000). Yang and Lo (2000) divides the RNN methods for satellite images in three groups:

- linear *statistical adjustments* of images that attempt to create comparable statistics, e.g. mean, standard deviation and dynamic range;
- *histogram matching*, an often used non-linear matching of histograms of images in order to get similar distributions of backscatter intensity; and
- the *linear regression normalization*, that is based on the basic premise that the signal reaching the satellite is a linear function of reflectivity, which means that atmospheric and sensor differences between acquisitions are linearly related.

For SAR imagery in particular *persistent scatterers* (PSs), points that are coherent over long time intervals and for variations in look-angles, are often exploited for RNN. The selected PSs are then used in a histogram matching method for example (Van Leijen, 2014).

Absolute calibration, on the other hand, is based on some reference target with known backscatter characteristics. Often transponders or corner reflectors are used in this regard (Rudolf et al., 2015). An absolutely calibrated radar image has the advantage over relatively calibrated images in the sense that the results relate to the physical properties of the backscatter (Van Leijen, 2014). Three different absolute calibrations exist, where nought is sometimes written as naught (e.g. Eineder et al., 2008)

- *beta nought* (β^0),
- *sigma nought* (σ^0), and
- *gamma nought* (γ^0).

Beta nought is a measure of the backscatter per unit area in slant range (Miranda and Meadows, 2015). The radar brightness, or beta nought, doesn't take the local incidence angle (θ) into account and is therefore only a normalization of the radar backscatter via a slant range based reference area (Small, 2011). Moreover, the local incidence angle is the angle between the normal to the illuminated ground area and the radar beam direction and sometimes also referred to as the local scope (Airbus, 2014a). Because no correction for the local incidence angle is applied, beta nought will contain geometric impact on the backscatter, such as bright pixels in foreshortened areas (Brantly, 2016). In Small (2011) it is argued that sigma nought and gamma nought estimates, who do take the local incidence angle into account, are still inaccurate in foreshortened and layover regions.

Sigma nought is often referred to as the radar backscatter coefficient or the normalized radar cross section. It corrects for the local incidence angle, which makes it is a measure of backscatter from a unit area on the ground with minimal topographic contribution to the backscatter. To compute sigma nought, accurate acquisition geometry, together with a DEM of the imaged area, is needed. The local incident angle has to be computed, which is done using an ellipsoidal Earth model (Small, 2011). Generally the sigma nought value is computed by using an ellipsoidal Earth model that simplifies the normalization factor to a simple sinus correction for the local incidence angle (Miranda and Meadows, 2015).

Gamma nought is a measure of the backscatter per unit area of the incident wavefront, perpendicular to the slant range (Miranda and Meadows, 2015; Small, 2011). Figure 2.14 shows the relation between the radar geometry and the normalization areas used to compute the calibration coefficients. The sigma nought backscatter coefficient can have a significant error, if topographic variation isn't properly accounted for in the sensor model. In order to mitigate this error, gamma nought uses a terrain-flattened normalization by integration of terrain variations. This also provides an improved interpretability of the gamma nought calibrated SAR image (Small, 2011). Similar to the sigma nought computation, the gamma nought value is often computed by simplifying the area normalization factor such that it reduces to the tangent of the local incidence angle. Therefore, both sigma nought and gamma nought can be computed from beta nought, knowing the local incidence angle (Miranda and Meadows, 2015). Additionally, with the use of a DEM both sigma nought and gamma nought can be terrain-geocoded. This will result in a geocoded-terrain-corrected (GTC) product, that corrects the position of the still ellipsoidal-based radiometry (Small, 2011).

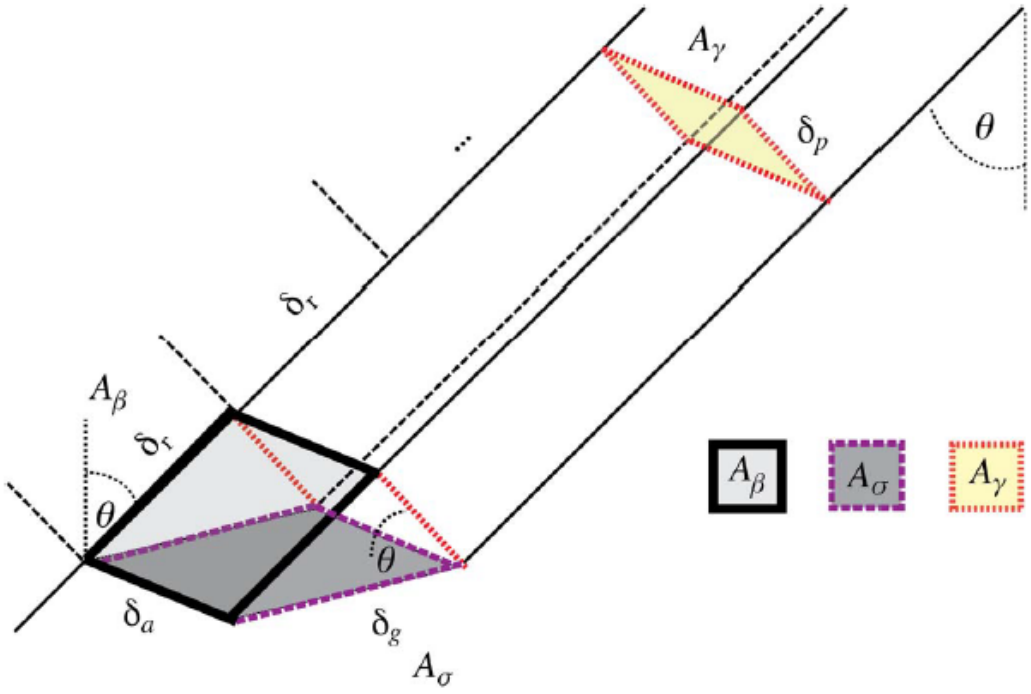


Figure 2.14: The relation between the normalization areas for SAR backscatter in radar geometry. Both sigma nought (f(A_σ)) and gamma nought (f(A_γ)) correct for the local incidence angle, whereas beta nought (f(A_β)) doesn't (Small, 2011) [A_β,A_σ,A_γ: normalization areas, θ: (local) incidence angle, δ_a,δ_r,δ_g: resolution in azimuth, slant range and ground range].

2.4.2. Sentinel-1 radiometric calibration

The radiometric calibration of Sentinel-1 level-1 products is described in Miranda and Meadows (2015). Sentinel-1 provides the normalization areas (A_β, A_σ, A_γ), as shown in Figure 2.14, in look-up tables (LUTs). Furthermore, the Calibration Annotation Data Set (CADS), which includes the LUTs for the different normalization areas, uses an ellipsoidal Earth model with an integrated height contribution per area. Therefore, beta nought, sigma nought and gamma nought can be calculated as

$$\beta^0 = \frac{\mathrm{DN}^2}{A_\beta^2}, \tag{2.28a}$$

with

$$A_\beta = \sqrt{A_{\mathrm{DN}}^2 K}, \tag{2.28b}$$

$$\sigma^0 = \frac{\mathrm{DN}^2}{A_\sigma^2} \tag{2.29a}$$

with

$$A_\sigma = \sqrt{\frac{A_{\mathrm{DN}}^2 K}{\sin(\theta)}} \tag{2.29b}$$

and

$$\gamma^0 = \frac{\mathrm{DN}^2}{A_\gamma^2} \tag{2.30a}$$

with

$$A_\gamma = \sqrt{\frac{A_{\mathrm{DN}}^2 K}{\tan(\theta)}}. \tag{2.30b}$$

In Equations (2.28a), (2.29a) and (2.30a) the Digital Number (DN), in the case of SLC data, is calculated as

$$\mathrm{DN} = \sqrt{I^2 + Q^2}, \tag{2.31}$$

where I is the *in-phase* component and Q the *quadrature* component of the complex signal, analogous to the real and imaginary part in the complex plane. Therefore, DN^2 denotes the product intensity.

It is important to note that the 2-way elevation antenna pattern (EAP) correction and the range spreading loss (RSL) correction are applied by default in Sentinel-1 products. The Sentinel-1 calibration constant (K) is already integrated in the LUTs and is only provided for completeness. Similarly, A_{DN}, which defines the final scaling from the internal SLC (coded in floats) to the final SLC or GRD product (coded in integers), is also incorporated in the normalization areas. This means that Equations (2.28a), (2.29a) and (2.30a) can be directly computed with their respective normalization areas from the LUTs.

2.4.3. TerraSAR-X radiometric calibration

The radiometric calibration of TerraSAR-X images is very similar to that of Sentinel-1 data. The instructions for the absolute radiometric calibration of TerraSAR-X data can be found in Airbus (2014a). In the case of single look slant range complex data, equation 2.31 is used to calculate the DN of the backscattered complex signal. Beta nought, sigma nought and gamma nought from TerraSAR-X SLC data can be computed as

$$\beta^0 = k_s |\mathrm{DN}|^2, \tag{2.32}$$

$$\sigma^0 = (k_s |\mathrm{DN}|^2 - \mathrm{NEBN}) \sin(\theta) \tag{2.33}$$

and

$$\gamma^0 = (k_s |\mathrm{DN}|^2 - \mathrm{NEBN}) \tan(\theta). \tag{2.34}$$

The calibration constant and processor scaling factor (Eineder et al., 2008) for TerraSAR-X data (k_s) can be directly used to calculate beta nought, evident from Equation (2.32). According to Eineder et al. (2008) all the TerraSAR-X products are provided in radar brightness beta nought, including the SLC data. However, a study with corner reflectors performed by the TU Delft shows that this is not the case (Van Leijen and Lopez Dekker, 2018).

To calculate the sigma nought and gamma nought coefficient the LIAs and the noise equivalent beta nought (NEBN) are needed. However, the level-1 calibration files don't provide a complete overview of the local incidence per normalized ground area, rather only the four corner incidence angles and the centre incidence angle. To compute the precise local incidence angles, the geocoded incidence angle mask (GIM) from the enhanced ellipsoidal corrected (EEC) product is needed (Airbus, 2014b), which makes it possible to back-calculate the LIA

$$\theta = \frac{\text{GIM} - (\text{mod}\left(\frac{\text{GIM}}{10}\right))}{100}. \tag{2.35}$$

Hence, the LIA has to be either estimated from the five known LIA points or determined from a DEM of the area in the absence of the EEC product.

It has to be noted that gamma nought isn't explicitly specified in radiometric calibration guidelines for TerraSAR-X data, but can be calculated as shown in Equation (2.34) nonetheless. Furthermore, the TerraSAR-X processor already compensates for the elevation and azimuth antenna pattern, the effects of different range and azimuth bandwidth and sensor settings (e.g. transmit power) (Eineder et al., 2008).

2.4.4. Radiometric accuracy and noise levels

The noise levels and radiometric accuracy of the absolute calibration of S1 and TSX data is given in decibels. The noise equivalent sigma naught (NESN), also referred to as noise equivalent sigma zero (NESZ), is a measure of system noise, or sensitivity, expressed in sigma naught. It depends on e.g. the antenna pattern, pulse power, bandwidth and receiver noise (Eineder et al., 2008). These values are estimations verified by actual measurements. For TerraSAR-X the NESN is between -19 dB and -26 dB, whereas Sentinel-1 NESN values are verified to be -22 dB or better (Airbus, 2014a; ESA, d).

The absolute radiometric accuracy for Sentinel-1 and TerraSAR-X data is within 1 dB and 0.6 dB respectively. The absolute radiometric accuracy is the root mean square (RMS) error between he measured and true radar cross-section in space and time. The stability, or relative radiometric accuracy, is the standard deviation of a known point within one data take (less than 220 s over range), in terms of radiometry. The standard deviation of TerraSAR-X data is 0.3 dB and 0.5 dB for Sentinel-1 data (Eineder et al., 2008; ESA, d).

2.5. Data storage

2.5.1. Multi-temporal I/Q data

So far the term radar image has not been fully elaborated. An image is often used in the context of a 2D matrix of real-valued pixels, such as a picture. With SAR being a coherent imaging method, hence collecting I/Q data, the 'image' often describes the amplitude part of the complex signal. The phase can also be plotted as an image, however a single phase image doesn't show any information whatsoever (as discussed in Section 2.2). This is partly due to the ambiguity in the integer number of cycles of the signal, a well known problem in GNSS as well (e.g. Kampes and Hanssen, 2004). Phase measurements are still extremely valuable and form the basis for radar interferometry (InSAR), a topic extensively discussed in Hanssen (2002).

A SLC SAR image is often stored in binary format, in a *major row order pixel-interleaved* way (Figure 2.15) (DEOS, 2008). Figure 2.15 has row values stored in *lines* (L) and column values stored as *pixels* (P). Especially the term 'pixels' can be confusing, as it is also used to indicate a single measurement in a 2D image (see Section 2.1.2). In this case, the lines correspond with measurements in azimuth direction, whereas pixels are in range geometry. The measurement at (L,P) = (1,1) corresponds with the first measurement of a swath at near range, independent of ascending or descending flight pass geometry.

Naturally, the amount of measurements of the same ground target will increase over time. In order to compare these different acquisitions, SAR image co-registration algorithms are needed. Although not elaborated here, it has to be noted that proper co-registration is vital for change detection in SAR amplitude data. Li and Bethel (2008) provides an overview of image co-registration, as well as a comparison of some of the most used algorithms. Having properly co-registered a set of images, one can create a 3D array of values. This multi-dimensional *data cube,* as visualized in Figure 2.16, allows us to see the behaviour of a single pixel over time.

	Pixel 1	Pixel 2	...	Pixel P
Line 1	Re,Im	Re,Im	...	Re,Im
Line 2	Re,Im	Re,Im	...	Re,Im
...	...	...	...	...
Line L	Re,Im	Re,Im	...	Re,Im

	Pixel 1	Pixel 2	...	Pixel P
Line 1				
Line 2				
...				
Line L				

Figure 2.15: Major row order pixel-interleaved storage of complex SAR data [Top: Pixel-interleaved; Bottom: Major row ordered].

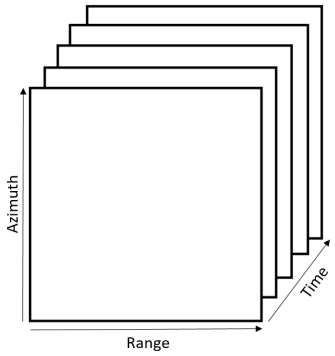


Figure 2.16: A visualization of a data cube of 2D SAR images. Each square represents a SAR image in time.

2.5.2. Storage of array-oriented scientific data

Simply storing SLC data as a binary file in some directory structure, or even in a single directory, is fine. However, for versatile use of the data it is recommended to use some *structured file format*. Two structured file formats will be briefly discussed in Section 2.5.3, alongside the data models that they are used in, namely

- Hierarchical Data Format (*HDF*), and
- Network Common Data Format (*NetCDF*).

When working with a spatial-temporal SLC data cube, two very common access patterns are acquiring data along the time axis or as a spatial cross-section at a certain epoch. In other words, reading a *slice* of the dataset. Moreover, one might even require a subset of a spatial or temporal slice. Such an operation can be easily done with structured file formats, without reading all the data into memory or using memory maps (e.g. memmap in Python). For large data arrays some of the main advantages of these structured files over 'plain' binary files are (Carter, 2015; Hartnett, 2011; Kington, 2014; Koziol and Breitenfeld, 2015)

- *direct access* to subsets of the data, i.e. efficient slicing of complex datasets (as mentioned);
- the data is *appendable* without redefining the structure of the data or copying the whole dataset;
- datasets are *self-describing*, which means that information about the data in the files is present without the need for additional log files;
- it is *portable* between different computational platforms and programming languages;
- multiple users can access the data at once and even write to the file; using parallel input-output (I/O) interfaces, as it is *sharable*;
- it has an *organizational* structure similar to a directory structure; and
- it enables long-term *preservation* of the data.

The access speed of certain access patterns is strongly linked to the way the data is stored. This means that retrieving a full slice along the range, azimuth or time axis will take an unequal amount of time. This performance bias is quite severe, as shown by e.g. Rew (2013) and Kington (2014), and can be some four orders of magnitude slower or faster between access patterns. As mentioned in Section 2.5.1, SLC data is stored in a row-major order (C-contiguous order or C-order for short). Therefore, the range axis is the fastest dimension to receive data from and the time axis by far the slowest. If the data was column-major ordered (Fortran-contiguous order or F-order for short), the azimuth axis would be the main beneficiary. This bias makes the analysis and visualization of large data sets rather unpractical.

Dependent on the application of the data, one can choose the optimal ordering for that particular use. Even creating multiple files of the same data sets using different index orderings could be considered. However, for more flexible use, for example using both spatial slices and time series at the same time, it is better

to find a middle ground. This is achieved by *chunking* the data. Chunking speeds up the access of the slowest dimension(s) at the cost of the speed of the fastest dimension(s) (Rew, 2013). This is achieved by storing multi-dimensional data sets as multi-dimensional rectangular chunks of a predefined size, rather than choosing a row or column ordering. For 2D data, this process is also referred to as *tiling* (Koziol and Breitenfeld, 2015).

2.5.3. HDF5 and NetCDF

The *data model* defines the building blocks for data organization and specification, whereas the *data format* handles the bit-level organization of the files (Hartnett, 2011; Koziol and Breitenfeld, 2015). Only HDF5 and NetCDF-4/HDF5 (sometimes NetCDF-4 for short), the most recent and relevant file format version of HDF and NetCDF will be considered. Only these two file formates are considered, because prior to the upgrade to the NetCDF-4/HDF5 file format, the HDF5 file format performed significantly better in various benchmarks, as shown by e.g. group (2002). The data formats for HDF and NetCDF won't be discussed in more detail here, but more on the topic can be found in HDF (2006) and NetCDF (2018a).

NetCDF has two data models, the classic model and the enhanced model (Hartnett, 2011). The classic model consists of *variables*, *dimensions* and *attributes*. These are used to store the bulk of the data (variables), express the physical dimension of the data such as time (dimensions) and provide information about the data (attributes). The enhanced model, or NetCDF-4 data model, is an extension of the classical model that allows for more expressive power for complex data sets, by allowing the use of *groups* and more data types, for example user-defined data types. One of the main advantages of using NetCDF-4 over HDF5 is the many tools and software packages that handle NetCDF data. However, most of the existing NetCDF software only handles the files in the classic NetCDF data model, to the point where "users are advised to use the classic data model whenever it works well" (Hartnett, 2011).

The HDF5 *abstract data model* consists of *groups*, *datasets* and *attributes*. Datasets are similar to the variables for NetCDF and HDF5 attributes also store metadata. The groups however, allow for the *hierarchical* structure of the files. These groups can make up a directory-like structure and contain zero or more HDF5 datasets and/or attributes, i.e. as group/.../dataset (Koziol and Breitenfeld, 2015). The hierarchical structure makes it really intuitive to work with and reorganize the data in the file. Both the NetCDF-4/HDF5 file format and the enhanced data model have a dependency on HDF5 (NetCDF, 2018b), which enables the group feature in the enhanced model and chunked storage for the data for example. To generalize it: NetCDF-4/HDF5 is a layer on top of HDF5, in order to get similar performance with a simpler application programming interface (API), by using HDF5 storage and the NetCDF API. However, with high-level programming languages, such as Python, the API for HDF5 is very easy to use as well.

3

Survey of change point detection methods

Change in a time series can be looked at in different ways. This chapter starts off by differentiating between some different points of view, as these will ultimately dictate the decision rules used for segmentation. Secondly, multiple classification frameworks used in literature are discussed to get the full picture of what kind of detection methods are used and how they compare with each other. Lastly, some CPD algorithms and evaluation metrics are reviewed, both used in- and outside of SAR-based imagery, to better understand the general flow of CPD algorithms and the assessment of the resulting change maps.

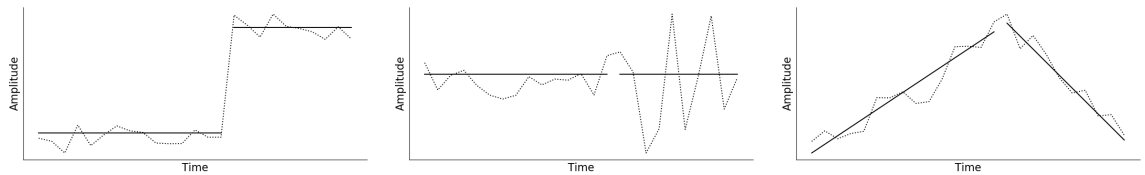
3.1. Categorization of CPD methods

3.1.1. Signal types and detection methods

In the context of change detection in time series many terms are adopted over a multitude of professions, such as step detection, segmentation, edge detection, regime switching, structural breaks, break points, change point detection or detection of transition instants. All of these terms are very similar in concept (e.g. Aminikhanghahi and Cook, 2017). Similar to the multi-temporal change detection terminology, the classification of change detection algorithms is equally dispersed. However, it is generally agreed that no approach is applicable and optimal to any particular signal (Jianya et al., 2008).

In this paper the term *change point detection* (CPD) is used, referring to the detection of abrupt variations in *time series* data when the (statistical) properties of that series change (Aminikhanghahi and Cook, 2017; Killick et al., 2012). Therefore, the *change points* represent transitions between different states in a time series. More formally, a time series of n data points $y_{1:n} = (y_1, ..., y_n)$ has m (unknown) change points. Each change point $\tau_{1:m}$ is an integer value ($1 \leq m \leq n-1$) at a certain time index, or position, τ in y. Furthermore, we set $\tau_0 = 0$ and $\tau_{m+1} = n$. Consequently, this means that the m change points will split the data in $m+1$ segments. An extension of CPD is change point estimation (CPE), where known changes are modelled (or described) and interpreted, rather than detected (Aminikhanghahi and Cook, 2017).

When looking at the statistical properties of a time series, a change points occurs at time τ when the distributions of two windows $(y_1, ..., y_\tau)$ and $(y_{\tau+1}, ..., y_n)$ differ in at least one criterion, e.g. the mean (μ), standard deviation (σ) (or variance (σ^2)) or regression structure (f(α,β)) (Rohrbeck, 2013). This is visualized in Figures 3.1a, 3.1b and 3.1c respectively.



(a) Change in mean. (b) Change in variance. (c) Change in regression slope.

Figure 3.1: Changes at τ=15 for a n=25 time series with different statistical decision criterions. Figure 3.1a, 3.1b and 3.1c show a change in mean, variance and regression slope respectively [Solid line: segments $(y_1, ..., y_\tau)$ and $(y_{\tau+1}, ..., y_n)$, Dotted line: original simulated signal].

This observation can be summarized by the problem of hypothesis testing between two hypotheses; the null hypothesis H_0 and the alternative hypothesis H_A

$$\delta = \begin{cases} H_0, & \text{if} \leq \tau \\ H_A, & \text{if} > \tau, \end{cases} \tag{3.1}$$

where δ denotes the decision rule, H_0 the hypothesis that no change has occurred, H_A the alternative that a change did occur and τ representing some threshold value. For the case of a changing mean the hypotheses would be $H_0 : \mu_1 = ... = \mu_n$ and $H_A : \mu_1 \neq ... \neq \mu_n$. In the case of some data cube this will eventually result in a *change mask* (B), a binary image of zeros and ones that identifies regions of change in some image $I_{i+1}(x, y)$ compared to $I_i(x, y)$ (Radke et al., 2005)

$$B(x, y) = \begin{cases} 0, & \text{if } H_0 \\ 1, & \text{if } H_1. \end{cases} \tag{3.2}$$

A very coarse subdivision for CPD algorithms exists in the sense that an algorithm is either classified *offline* (posterior) or *online* (sequential) and of a *supervised* or *unsupervised* nature. When an offline algorithm is mentioned, the entire time series in considered in one batch and reflected upon. Online algorithms process each new point as it becomes available, which allows it to detect a change in (near) real-time. Furthermore, the difference between supervised and unsupervised is the fact that unsupervised learning algorithms discover patterns in unlabelled data, often based on statistical features, whereas supervised learning algorithms are dependent on input data (Aminikhanghahi and Cook, 2017).

The time series produced by SAR remote sensing satellites are becoming larger by the day. Hence, the computational efficiency will play a vital role in effectively detecting changes in these datasets. To determine the computational cost of certain algorithms, the distinction between *parametric* and *non-parametric* algorithms can be made. A non-parametric approach, or memory based method, does not need to make any assumptions about the underlying function or model of the dataset in contrast to parametric approaches, which makes it scale much better with an increasing data availability at lower computational costs (Raykar, 2007). However, this comes at the expense of needing the entire dataset at every run, making the non-parametric approaches not suited for near-real time analysis. This consideration leaves us with the observation that a successful algorithm strikes the correct balance between decision quality and computational cost (Aminikhanghahi and Cook, 2017).

3.1.2. Classification frameworks

The classification of supervised and unsupervised algorithms, as introduced in Section 3.1.1, allows us to classify different algorithms in this particular framework. Aminikhanghahi and Cook (2017) presents a good overview of some of the different CPD supervised and unsupervised techniques and how they differ from each other. An overview of supervised and unsupervised CPD approaches is given in Figures 3.2 and 3.3.

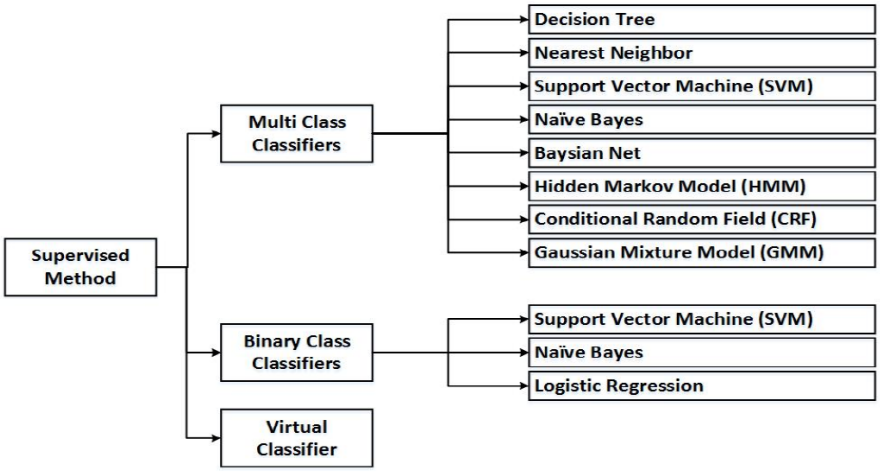


Figure 3.2: The supervised change point detection classification framework by Aminikhanghahi and Cook (2017). The listed algorithms are all dependent on input data.

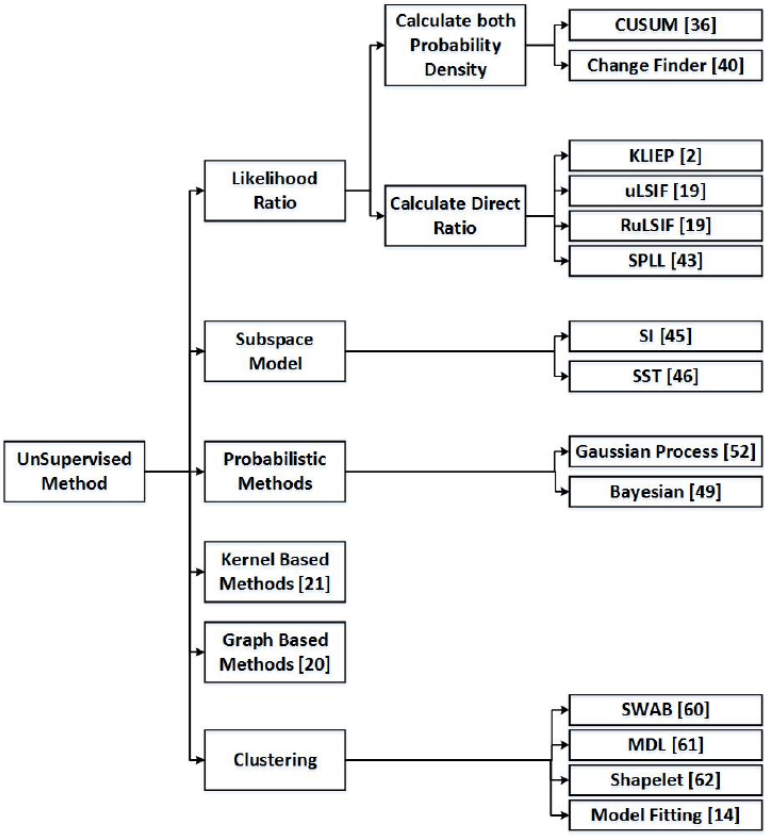


Figure 3.3: The unsupervised change point detection classification framework by Aminikhanghahi and Cook (2017). Each of the listed algorithms is able to discover patterns in unlabelled data.

For the supervised methods, three classifiers are defined: the *multi class* classifier, the *binary class* classifier and the *virtual class* classifier. The binary classifier coincides with the binary change mask in Equation (3.2). The multi class classifier allows for more than two states, but is more often used for change detection in the feature space or spatial domain, e.g. by using a non-parametric (k-)nearest neighbour algorithm that clusters pixels in various groups. The virtual classifier detects and interprets change points, by utilizing supervised decision tree learning into unsupervised change analysis (Hido et al., 2008). This means the virtual classifier is a CDE method, rather than pure CPD.

The unsupervised methods consist of *likehood ratio* methods, *subspace models*, *probabilistic* methods, *Kernel based* methods, *graph based* methods and *clustering* methods. Likelihood ratio methods can be either parametric or non-parametric. The former are based on the calculation of the PDFs of consecutive intervals defined by the change point search windows. Change points are defined when these distributions exhibit significant differences, by evaluating the logarithm of the density-ratio. Non-parametric likelihood ratio methods don't rely on pre-designed density model estimations of the consecutive windows, but rather estimate the density-ratio directly by a non-parametric Gaussian kernel model (Aminikhanghahi and Cook, 2017).

The subspace model algorithms are based on the principle that the subspace spanned by the columns of an extended observability matrix, which represents the subspace of the entire time series, is similar to the ones spanned by the sub sequences of the time series. The observability matrix is constructed by the system matrices of a pre-designed state-space model, that models the time series (Kawahara et al., 2007). The measure of change is given by the dissimilarity; the distance between the subspace of the time series and sub sequences of the time series data (Liu et al., 2013). Both parametric and non-parametric methods exist.

Probabilistic methods use a single sliding window for change point detection. The underlying observation for these methods is that a time series can be divided in some non-overlapping product partitions, each of which has its own PDF. A change point is found when adjacent segments are modelled by a PDF with different characteristics, e.g. a Gaussian PDF with different mean and variance (Adams and MacKay, 2007).

Kernel based methods exist as supervised and unsupervised, but are always used in the non-parametric context. Windowed segments are compared based on a test statistic from the kernel Fisher discriminant

ratio (KFDR). This ratio can be interpreted as a measure of homogeneity between windows and is often compared to some self-defined threshold value. To find the most likely change point the KDFR is calculated for all consecutive window combinations, after which the maximum ratio value is compared to the threshold (Aminikhanghahi and Cook, 2017).

The non-parametric approach of graph based methods uses tests based on graphs with the observations from the time series as nodes. Numerous ways to construct the graph exist, such as minimum distance paring or a nearest neighbour graph. In general the graph is derived from a dissimilarity or distance on the sample space, representing the similarity between observations. The edges of the graph connect observations that have a short distance to one another (Chen and Zhang, 2014).

Lastly, clustering methods. Clustering is often used in the context of the spatial domain, in order to yield some planar partition of different classes, by grouping data points. This approach can be translated to the temporal domain. The bottom-up approach for example, treats all observation points as separate segments. It then groups observations until some stopping criteria is met. The goal being to create segments with similar points, whereas different segments are as dissimilar as possible (Aminikhanghahi and Cook, 2017).

Deer (1999) first proposed a very straightforward classification of change detection methods in three main categories: *pixel-based*, *feature-based* and *object-based* change detection, where

- the *pixel* level uses isolated brightness values within an image;
- the *feature* level uses some transformation on the spatial or spectral properties of the image, before further analysis; and
- the *object* level refers to comparison of subsets of an image that show homogeneity in the spatial and spectral domain, e.g. texture.

Over time this classification framework by Deer (1999) has evolved. Jianya et al. (2008) proposes a classification framework which begins by determining how the signal in question is approached, rather than which type of algorithm is used. In Figure 3.4 this classification concept is shown.

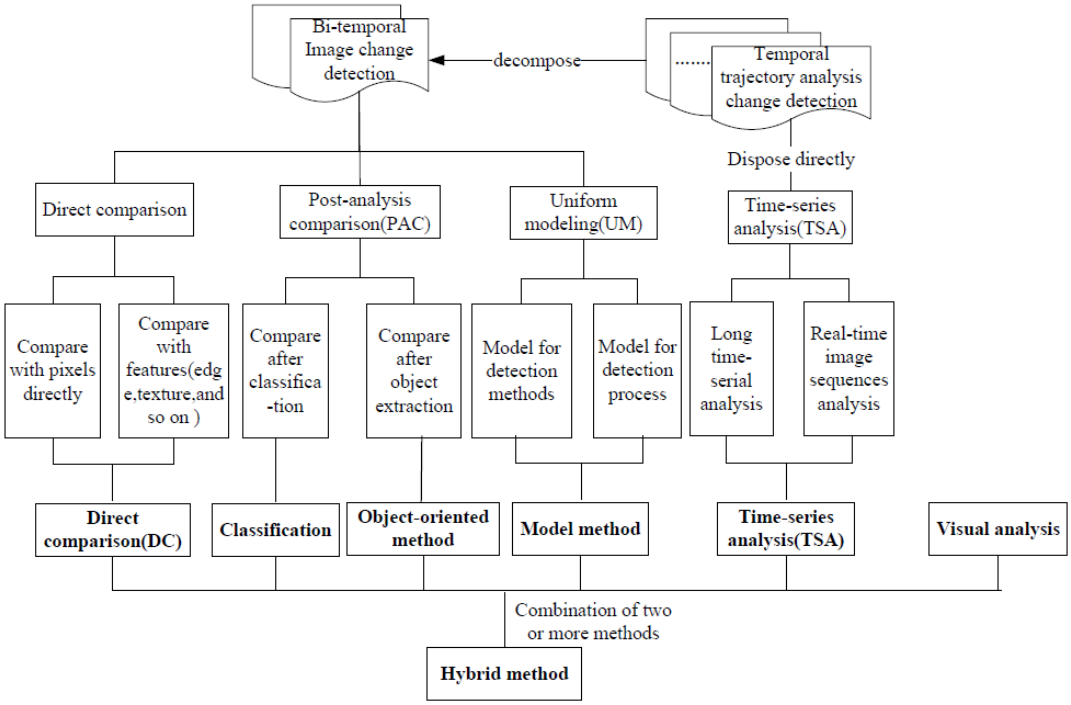


Figure 3.4: The CPD classification concept by Jianya et al. (2008). A division between decomposing a data cube for bi-temporal change detection or analysing the temporal trajectory of pixels with the help of time series analysis is applied.

Starting from multi-temporal data, a subdivision is made between bi-temporal image change detection by decomposing the multi-temporal data and temporal trajectory analysis in the form of time series analysis. This

division ultimately results in seven CPD methods: *direct-comparison*, *classification*, *object-oriented method*, *model method*, *time-series analysis*, *visual analysis* and a *hybrid method* (a combination of the aforementioned methods) (Jianya et al., 2008).

Direct comparison methods compare pixels or features, either directly or after some basic transformation. Classification and object-oriented methods on the other hand, rely on comparing extracted information from the images. The most important post-analysis comparison methods are dependent on object extraction from images. These methods are often based on feature extraction prior to the analysis, hence 'post-analysis' comparison. Following some image segmentation, these different segments are then analysed for change regions. Object-oriented methods are mostly used in image-to-map change detection, such as GIS integrated methods. This is due to the proneness of errors in feature and object extraction. The uniform modelling approach, has a subdivision in the sense that the model is made for the detection methods (approach-based) or the detection process (process-based). Modelling methods propose a big challenge in the creation of suitable models (Jianya et al., 2008).

For multi-temporal data, a time series analysis can often makes sense. At the expense of spatial details, the temporal trajectory allows us to exploit the temporal consistency of pixels at different times. Jianya et al. (2008) subdivides the temporal trajectory analysis in two approaches: (non-real-time) long-time serial analysis and real-time image sequences analysis. These approaches are analogue to the offline/ posterior and online/ sequential analysis respectively, as introduced in Section 3.1.1.

Similar classification approaches to Jianya et al. (2008) for bi- and multi-temporal change detection algorithms are presented in Lu et al. (2004) and Radke et al. (2005). In Lu et al. (2004), a total of six CPD categories are presented, as well as one category which is compiled of CPD techniques that don't quite fit in the six proposed categories and aren't used extensively in the field of change detection. These categories are *algebraic* operations, *transformations*, *classification*, *advanced models*, *GIS* integration and *visual analysis*. Radke et al. (2005) groups techniques in the broader scope that is image analysis as a whole, in a total of six groups: *simple differencing*, *significance and hypothesis tests*, *predictive models*, based on a *shading model*, *background modeling* and *change mask consistency*. Here temporal and spatial models are subgroups of the predictive models group. Of course, many more change detection classification frameworks have been proposed. These include, but are not limited to, Singh (1989) (subdivision in classification and direct comparison), Coppin and Bauer (1996) (similar to Radke et al. (2005), but remote sensing only) and Li et al. (2002) (classification on image registration and data sources).

3.2. Change detection methods and algorithms

3.2.1. CPD in multi-temporal data sets

A great many CPD algorithms exist for multi-temporal data, which are not typically used for remotely sensed imagery. Three such algorithms will be discussed, namely *Binary Segementation* (BS) (Sccott and Knott, 1974), *Optimal Partitioning* (OP) (Jackson et al., 2005) and *Pruned Exact Linear Time* (PELT) (Killick et al., 2012). These algorithms are highlighted because of their widespread use, computational efficiency and close relation to one another.

Starting from a sequence $y_{1:n} = (y_1, ..., y_n)$, as described in Section 3.1.1, each change point in y has a position τ, such that $\tau_{1:m} = (\tau_1, ..., \tau_m)$ is the set of change points. A common approach to identify multiple transitions within a time series is by minimising

$$\sum_{i=1}^{m+1} [C(y_{(\tau_{i-1}+1):\tau_i})] + \beta f(m), \tag{3.3}$$

where each ith segment in y is made up of $y_{(\tau_{i-1}+1):\tau_i}$, C is some cost function for a segment and the $\beta f(m)$ term denotes a penalty term to avoid over fitting. Equation (3.3) will be referred to as the *penalised case*. A very commonly used cost function is twice the negative log likelihood, however other cost functions such as the cumulative sum and quadratic loss are also used (Killick et al., 2012). In this notation the penalty term is not linear, but rather a function of the number of segments. However, this guard is also often considered to be linear in m , such that $\beta f(m) = \beta m$ (Rohrbeck, 2013).

Binary segmentation is probably one of the most widely used CPD algorithms to the point it can be considered a 'standard' method (Killick et al., 2012; Rohrbeck, 2013). In order to detect multiple change points, it iteratively uses the same approach to segment a sequence, by applying the same methodology to segmented subsets of that sequence. In relation to Equation (3.3), it tests whether some integer τ in y exists that satisfies

$$C(y_{1:\tau}) + C(y_{(\tau+1):n}) + \beta < C(y_{1:n}). \tag{3.4}$$

The point τ that yields the lowest value is compared against the cost of the entire time series. If Equation (3.4) holds, τ is identified as a change point and the time series is split up in two segments, split at τ. This change point method is repeated iteratively for all sub sequences until Equation (3.4) is not satisfied anymore, at which point the algorithm will break. BS does not minimise the penalised minimalization problem automatically, but rather approximates it (Rohrbeck, 2013). This means that BS isn't an exact method. This non-exactness is best explained by an analogy of BS to the problem of finding the shortest path between two points. A shortest path between points a and b connected by a series of parallel and serial nodes does not necessarily start by connecting the node that is closest to a with a.

On the other hand, Optimal Partitioning is an exact method, as it does satisfy Equation (3.3). The basis for OP is dynamic programming, an efficient recursive procedure that solves combinatorial optimisation problems, such as the penalised case (Maidstone et al., 2014; Rohrbeck, 2013). OP is based on the idea that the minimalized penalised cost over segmentations can be split in a minimisation over the position of the last change point and all the earlier change points. Dynamic programming can be used to solve the minimisation problem because the *principle of optimality* holds, which tells us that any sub-partition of an optimal partition is optimal (Rohrbeck, 2013).

Consider some sequence $y_{1:t} = (1, ..., t)$, a subset from $y_{1:n} = (1, ..., n)$. After conditioning on the last point of change, OP determines the optimal partition of the first $t+1$ observations, by using the previously obtained optimal partitions for the first $1, ..., t$ data points. For each iteration step $t+1$ all the possible starting locations j are considered, with $1 \leq j \leq t$. In other words, it considers all previous data points as *candidates* for the previous change point to the current data point (Jackson et al., 2005). The partition that minimises the cost is stored after the iteration, with previous iteration steps having stored the cost of the optimal partition prior to j. The set of possible change points for the first t data points is stored as

$$\mathcal{T}_t = \{\tau : 0 = \tau_0 < \tau_1 < ... < \tau_m < \tau_{m+1} = t\} \tag{3.5}$$

with τ_m the last point of change (Wambui et al., 2015). Furthermore, the cost of the optimal partitioning of the data until t is denoted as $F(t)$. The cost can be evaluated, following Equation (3.3) and using the principle of optimality, with the recursion

$$F(t) = \min_{0 \leq \tau_m \leq t} \{F(\tau_m) + C(y_{\tau_m+1:t}) + \beta\}. \tag{3.6}$$

It starts by calculating $F(1)$, recursively up until $F(n)$, storing the optimal segmentation up until τ_{m+1} at every step (Wambui et al., 2015). When the end of the sequence is reached at $t+1 = n$, the algorithm breaks, having recorded the location of the change points (Rohrbeck, 2013).

The Pruned Exact Linear Time algorithm by Killick et al. (2012), a fast and exact algorithm, is widely used in time series change analysis. It is a continuation of Optimal Partitioning, with an added inequality based pruning step to increase the efficiency of OP based change detection. This works by limiting the set of change point candidates under the condition that a constant κ exists for some data point s, such that for all $t < s < T$

$$C(y_{t+1:s}) + C(y_{(s+1):T}) + \kappa \leq C(y_{t+1:T}), \tag{3.7}$$

which translates in

$$F(t) + C(y_{t+1:s}) + \kappa \geq F(s). \tag{3.8}$$

This means if Equation (3.8) holds then for any $T > s$ the best segmentation with the most recent change point prior to T being at s is better than any change point at t (not optimal), which avoids a search over $t < s$. In practice, it removes values of τ that can never be minima following the penalised case. Moreover, this also means that at each minimisation iteration step fewer values are considered, hence increasing performance. Almost all commonly used cost functions satisfy the assumption as stated in Equation (3.7) (Killick et al., 2012).

Lastly, two CPD methods that won't be discussed here, but deserve a mention nonetheless are Segment Neighbourhoods (SN) (Auger and Lawrence, 1989) and Bayesian Online Changepoint Detection (BOCD) (Adams and MacKay, 2007). The SN solves the *constrained case* of the minimalization problem in an exact way, similarly to OP. For the constrained case, Equation (3.3) is now simplified as it does not include the $\beta f(m)$ penalty

term this time and the number of segments $m+1$ is assumed to be known. This means it has a search space of a maximum $m+1$ change points, pre-defined by the user. Moreover, SN is known to be much slower than the likes of PELT for example, with SN having a cubic complexity and a linear computational complexity for PELT (Killick et al., 2012).

Probably the most well known online CPD algorithm is the BOCD by Adams and MacKay (2007). An algorithm that focusses on predictive filtering and generation of an accurate distribution of the unseen data point (Adams and MacKay, 2007). Instead of retrospective segmentation (offline), it uses past measurements to generate an accurate distribution of the next unseen point (online). Taking into account the highly unpredictive speckle, it is unlikely that this kind of approach is suitable for SAR change detection. Moreover, it can be argued that an online segmentation approach is not beneficial at all with consecutive measurements for Sentinel-1 and TerraSAR-X being multiple days apart.

3.2.2. Classical CPD in remotely-sensed imagery

A lot of research has been done in the field of *bi-temporal change detection* using SAR amplitude data. The 'classical' change detection using SAR images is often based on the widely used *difference image* (DI), which is created by taking two SAR images at different acquisition times and applying a pixel-by-pixel rationing operator or subtraction/ difference operator. Moreover, image correlation, image regression and change vector analysis (CVA) are also used to create SAR based DIs (Tan and Hao, 2017). CVA describes the change of magnitude and direction of a vector through time (Bruzzone, 2016). However, the most widely used comparison operator for SAR imagery is the *log-ratio operator* (LR), as it reduces the effect of the multiplicative speckle and it produces a more statistical symmetrical distribution of the changed and unchanged classes (Bruzzone, 2016; Zhang et al., 2014). It is defined as

$$A_{\mathrm{LR}} = \left|\log\left(\frac{A_2}{A_1}\right)\right| = \left|\log(A_2) - \log(A_1)\right|, \tag{3.9}$$

with A_1 and A_2 the SAR amplitude values of consecutive images. After the DI generation, the DI is analysed, which includes a segmentation step and/ or an iterative clustering step. Moreover, a spatial filter is nearly always applied, either before or after the DI generation. These filters range from basic moving-average filter to complex-wavelet-based filters. A comparative study between complex-wavelet-based filters and standard filters by Gagnon and Jouan (1999) showed a similar performance for both types of filters in case of low-level speckle, whereas the complex-wavelet-based filters slightly outperform the standard filters in the case of high-levels of speckle.

In the DI analysis step a wide variety of techniques have been deployed to generate a binary change map, either supervised or unsupervised. The segmentation of the DI in a changed and unchanged part is either performed by a *threshold method* (Bazi et al., 2006, 2005; Bruzzone and Prieto, 2000; Jin, 2005; Zhang et al., 2014), a *clustering* method (Ajadi et al., 2016; Alphonse and Biju, 2015; Benedek et al., 2015; Kasetkasem and Varsney, 2002) or a combination of the two (Tan and Hao, 2017). Other techniques include the addition of a transformation on the input image before the segmentation step, using contourlets (Varghese and Jose, 2014), curvelets (Schmitt et al., 2010) or wavelets (Celik and Ma, 2011). Moreover, fractals (Aghababaee et al., 2013), machine learning (Gao et al., 2016; Wieland and Yamazaki, 2016) and singular value decompositions (SVDs) (Gao et al., 2017) have all been successfully used with regard to change detection in SAR based data.

Image fusion is also a common technique used in SAR based change detection. Image fusion is the process of combining information from various images into a single image (Wang et al., 2017). One such approach is combining the information of multiple DIs resulting from different operators, such as the mean ratio operator and log ratio operator in Varghese and Jose (2014). More recently, Tan and Hao (2017) used image fusion to fuse two change detection maps, one based on a threshold method and one on a clustering method, thereby attempting to combine the best of both worlds. Furthermore, image fusion can also be used to enrich single sensor remotely sensed data and add complementary features of e.g. multispectral imagery (Wang et al., 2017).

Rather than visually inspecting a difference image, trial-and-error thresholding or using an empirically derived threshold approach such as $A_i > \mu + \tau\sigma$ (Fung and LeDrew, 1988), with A_i the pixel value, τ some constant threshold and μ and σ the mean and standard deviation of the DI; more efficient and accurate methods have been developed over the years. A widely used method for pixel level based automatic thresholding is based on *Bayes decision theory*, first introduced by Bruzzone and Prieto (2000). By applying Bayes rule for minimum error, it enables automatic threshold selection for the DI, by minimizing the overall error probability in the change detection process. This approach assumes that all the pixels are spatially indepen-

dent and that the DI is composed of two Gaussian PDFs; one for the changed class and one for the unchanged class such that

$$p(DI) = p(DI|A_{\text{changed}})P(A_{\text{changed}}) + p(DI|A_{\text{unchanged}})P(A_{\text{unchanged}}), \tag{3.10}$$

with the Gaussian distributions for the changed and unchanged classes given as

$$p(DI|A_j) = \frac{1}{\sqrt{2\pi}\sigma_i} \exp\left(-\frac{DI-\mu_i{}^2}{2\sigma_i^2}\right). \tag{3.11}$$

Here $p(DI)$, $p(DI|A_{\text{changed}})$ and $p(DI|A_{\text{unchanged}})$ are the PDFs of the DI, changed and unchanged pixels respectively. Moreover, $P(A_{\text{changed}})$ and $P(A_{\text{unchanged}})$ are the a priori probabilities and i an iterator over the classes, with $i \in (changed, unchanged)$. Thresholding based on Bayesian inferencing is often combined with the Expectation-Maximization (EM) algorithm to determine a priori statistical parameters for the aforementioned Gaussian distributions. EM is an iterative method to find the maximum likelihood estimates for latent variables in a statistical model, by alternating between expectation and maximization steps (Bruzzone and Prieto, 2000), (Celik, 2010), (Bazi et al., 2006).

The Bayesian framework can also be extended to the spatial-contextual domain. This is often done in combination with a Markov Random Field (MRF) model. MRFs can serve as a powerful method to exploit interpixel correlation with neighbouring pixels, i.e. a pixel of class i is likely to be surrounded by pixels all belonging to that same class. MRFs are incorporated in a two-threshold methodology, where a low threshold is used to create an initial (absolute) unchanged class and a high threshold for the (absolute) changed class. The pixels not assigned to a class are then further processed with the help of MRF (Bruzzone and Prieto, 2000), (Jin, 2005), (Kasetkasem and Varsney, 2002).

Clustering techniques can be divided in two main categories; *hard clustering* schemes and *fuzzy clustering* schemes (Alphonse and Biju, 2015). The Fuzzy C-Means (FCM) algorithm is by far the most popular choice, as it allows for partial membership of a pixel to both the changed and unchanged class; hence retaining more image information than hard clustering schemes in certain instances (Subbarao, 2014), (Varghese and Jose, 2014). FCM will obtain the membership probability ($u_{x_i\,j} \in [0,1]$) of all the pixels x_i to each cluster $j = 1,...,c$ by minimizing the FCM objective function (Tan and Hao, 2017)

$$J_m = \sum_{i=1}^{n}\sum_{j=1}^{c}\left(u_{x_i\,j}\right)^m ||\,x_i - v_j\,||^2, \tag{3.12}$$

where the Euclidean distance is given by $||\,x_i - v_j\,||^2$, v_j denotes the prototype of the centre of the cluster j and m a weight component. The standard FCM algorithm only works well with de-noised images, due to the absence of spatial information (Alphonse and Biju, 2015), (Tan and Hao, 2017).

In order to overcome this sensitivity to noise, more robust methods based on FCM have been proposed that incorporate spatial information. Reformulated Fuzzy Local Information C-Means Clustering (RFLICM) (e.g. Varghese and Jose, 2014) and Markov Random Field Fuzzy C-Means (MRFFCM) (e.g. Subbarao, 2014) are two such methods. The latter uses MRF to decide whether a pixel is located in a hetero- or homogeneous area, which is used to improve the FCM membership assignment. The 'reformulated factor' in RFLICM algorithm adds a local similarity measure to the classical FCM algorithm, introducing a trade-off between noise level and image detail.

In general, it is argued that threshold methods have less false negatives and more false positives compared to clustering algorithms that take into account spatial information (Tan and Hao, 2017). Using multi-sensor data, with the help of image fusion, often results in more accurate change detection maps at the expense of difficult image processing and difficulty in the choice of suitable CPD technique (Lu et al., 2004).

3.2.3. Multi-temporal SAR-based CPD

The analysis of the temporal trajectory of co-registered SAR data is getting more common. CPD algorithms used in SAR amplitude time series analysis are often based on statistical features of the time series. It is good to note that besides amplitude-based *incoherent* change detection, phase information can be used as well. *Coherent* approaches have been developed, that study the coherence map, such as in Preiss and Stacy (2006). These won't be discussed here however.

A CPD method based on the ANalysis Of VAriance (ANOVO) is first presented in Li et al. (2011). It allows the detection of changes along the time domain that show *statistical significant stepwise* changes. The variation in amplitude over time is still subject to speckle, which can introduce many false positives if not taken

into account. The ANOVA approach in Li et al. (2011) is applied on 47 TerraSAR-X images, segmenting each pixel's time series with a 2-step (long-term changes) or 3-step model (short-term changes), i.e. identifying a maximum of 1 or 2 change points respectively. ANOVA measures the variation between two or more groups by analysing the variance of those groups, via F statistics. This allows us to compare the sample means of these groups. It has to be noted that ANOVA assumes that the data in each group fits the Normal distribution.

The detection of multi-transitional instants using ANOVA is further elaborated in Dogan and Perissin (2010). Where Li et al. (2011) used pre-determined number of groups, Dogan and Perissin (2010) generalizes this problem by iteratively using a 2-step approach in their research. The 70 TerraSAR-X images are analysed for changes along the time axis and segmented at point τ_m for pixel i, creating two new time series. The last segment is then iteratively analysed in a similar fashion, until a segment doesn't show significant statistical variation or only a few time steps remain denoted by a *guard* (D). This approach allows for the detection of an unknown number of change points in a time series. ANOVA will be further elaborated in Section 4.2.4.

For time series analysis the widely used log-ratio operator is not suited, as it only allows the comparison of two images at a time. An alternative to this is the use of the ratio of the *generalised (temporal) means*, also knows as power mean or Hölder mean, of the amplitude component of the SAR time series. In SAR image time series the Pythagorean generalised means are often used; being the geometric (M_0), arithmetic (M_1) and quadratic mean (M_2) (Quin et al., 2013). Staring with a sequence of pixel amplitude values $y_{1:n} = (y_1, ..., y_n)$, the jth order generalised mean of y is denoted as (Bullen, 2003)

$$M_j(y) = \begin{cases} \sqrt[j]{\frac{1}{n}\sum_{i=1}^{n} y_i^j}, & \text{if } j \in \mathbb{Z}^* \\ \sqrt[n]{\prod_{i=1}^{n} y_i}, & \text{if } j = 0. \end{cases} \tag{3.13}$$

The differences between various generalised means are enhanced in areas that are subject to change, whereas unchanged areas show similar values in their generalised means. This observation is the basis for methods developed by Lombardo and Olivier (2001) and Quin et al. (2013).

The former proposes a method based on the generalised maximum likelihood approach for 11 ERS SAR images, to distinguish between agricultural and forest areas. A statistical test is derived as the ratio of the geometric and arithmetic generalised means, such that

$$\Lambda = \frac{\sqrt[n]{\prod_{i=1}^{n} y_i}}{\frac{1}{n}\sum_{i=1}^{n} y_i} < \tau, \tag{3.14}$$

where Λ is the test statistic and τ denotes some threshold value. This test is a realisation of the *generalised likelihood ratio test* (GLRT). According to the generalised likelihood criterion this is the optimal change detection test (Lombardo and Olivier, 2001). The GLRT is combined with either pre-analysis spatial partitioning following a multi-temporal segmentation annealing algorithm (MT-SEGANN) or a simple square moving average of NxN pixels (see Section 2.1.5, Equation (2.16)). Both introduce a loss in spatial resolution. The former is argued to be the most intuitively correct, whereas the latter doesn't require any prior analysis of the images resulting in much lower computational cost.

A Method for generelIsed Means Ordered Series Analysis (MIMOSA) for 7 TerraSAR-X and 2 airborne CARABAS-II is presented by Quin et al. (2013). MIMOSA analyses a stack of images by comparing two generalised temporal means. The geometric and quadratic mean were selected in this instance. Each amplitude value y_i is assumed to be the product of a texture t and varying speckle component s_i, as given by the product model discussed in Section 2.2. In the no-change case the texture is assumed to be constant. The texture and speckle are modelled by an inverse Rayleigh-Nakagami and normalized Rayleigh-Nakagami distribution respectively, yielding a Fisher distribution for the amplitude values.

First an estimation of the amplitude Fisher distribution parameters over the images is made using a log-cumulant-based method, after which the joint PDF between two means m_0 and m_2 is estimated. These means are the result of substituting the product model in Equation (3.13). In the bi-temporal case the two means can then scattered in a 2D graph, such that two *isolines* I_1 (based on the joint PDF $p(m_0, m_2)$) and I_2 (based on the conditional PDF $p(m_2|m_0)$) make up a region of change. An automatic iterative thresholding step, based on a false alarm rate (FAR), is used to select the two thresholds τ_1 and τ_2 that define these isolines with the following hypothesis test

$$\delta = \begin{cases} H_0 : p(m_0, m_2) \geq \tau_1 \vee p(m_2|m_0) \geq \tau_2 \\ H_A : p(m_0, m_2) < \tau_1 \wedge p(m_2|m_0) < \tau_2. \end{cases} \tag{3.15}$$

For the multi-temporal case the conditional PDF can't be obtained, hence two *guides*, G_1 and G_2 are used instead. The former defines where I_1 becomes vertical in the 2D plane and the latter is used to define the point where I_1 becomes horizontal.

In Su et al. (2014) a *likelihood ratio test* based method is conducted on various TerraSAR-X data sets, namely the NORamlized Cut on chAnge criterion MAtrix (NORCAMA). Rather than using the generalised means, it uses a combination of the noisy and de-noised amplitude values. This method follows a three-step approach, starting with a multi-temporal de-noising step. The second step uses both the noisy and de-noised images in a likelihood ratio test to determine the change criteria between images. Lastly, the classification of types of change is done by a clustering and recognizing method on the change criterion matrix (CCM). The CCM presents the temporal behaviour of the time series of each pixel. Two likelihood ratio tests are proposed to determine the change criterion, a GLRT

$$\Lambda^{GLRT} = \frac{p(y_1, \hat{x}_1|x_{12}, H_0)p(y_2, \hat{x}_2|x_{12}, H_0)}{p(y_1, \hat{x}_1|x_2, H_A)p(y_2, \hat{x}_2|x_2, H_A)} < \tau \tag{3.16}$$

and an approximate likelihood ratio test (ALRT)

$$\Lambda^{ALRT} = \frac{p(y_1, y_2|x_{12}, H_0)}{p(y_1, y_2|x_1, x_2, H_A} < \tau. \tag{3.17}$$

The ideal de-noised amplitude values x_i are not available of course, hence estimated $\hat{x}_i$ values are used in the ALRT following a two-step denoising method. Moreover, GLRT uses the maximum likelihood estimations for x_1, x_2 and x_{12}.

Similar to the bi-temporal case, image fusion can be used in a multi-temporal setting. Sofiane and Ferdaous (2012) proposes a combination of a Rayleigh *Kullback-Leibler* (KL) *divergence* measure and a Rayleigh Distribution Ratio (RDR) measure to detect changes in SAR images. This methodology is tested on 14 ENVISAT images, as well as simulated data. Prior to the analysis, the input images are first filtered with a modified version of a Non-local (NL) means spatio-temporal filter that uses weights based on similarity of a pixel in the spatial and temporal domain. As a pixel is assigned a weight for each time step; the pixel in a time series with the highest weight can be thought of as the most representative of the normal state. From all these most stable pixels a *reference image* is created.

Following this reference image and the assumption of Rayleigh distributed (de-noised) data, the symmetric Rayleigh KL divergence measure can defined. As the KL divergence $D_{KL}(P||Q)$ of two distributions P and Q is not symmetrical, a symmetric version is defined as $D_{KL}(P,Q) = D_{KL}(P||Q) + D_{KL}(Q||P)$ (Sofiane and Ferdaous, 2010), such that

$$D_{KL}(P_{A_{ref}}, P_A) = L\frac{\left|\sqrt{\hat{A}_{ref}} - \sqrt{\hat{A}}\right|^2}{\sqrt{\hat{A}_{ref}}\sqrt{\hat{A}}}, \tag{3.18}$$

where $D_{KL}(P_{A_{ref}}|P_A)$ is the divergence from the 'true' PDF $P_{A_{ref}}$ to the observed target distribution P_A, $\hat{A}_{ref}$ and $\hat{A}$ are equivalent to the texture square root of t in the product model and L is the number of looks (Le et al., 2016). For the symmetrical case, KD divergence should actually be referred to as KL distance (KLD) (Inglada, 2003). This measure gives a way to quantify the similarity in PDFs (Sofiane and Ferdaous, 2012). KL-based methods using different PDFs for the SAR data have been developed, such as in Le et al. (2016). The second change criterion, RDR, is actually often used in the filtering process. It is formulated as

$$RDR(A_{ref}, A) = \frac{1}{N}\sum_{(k,l)\in V} \log\left(\frac{A_{ref,i}(k,l)}{A_i(k,l)} + \frac{A_i(k,l)}{A_{ref,i}(k,l)}\right), \tag{3.19}$$

with V the neighbourhood of the pixel i, controlled by a pre-determined window size in the de-speckling step, and N the number of pixels present in V. The resulting change maps following these two decision rules are then fused via the Dezert-Smarandache Theory (DSMT), creating a detector using both local (RDR) and global (KLD) variations.

3.3. Performance evaluation

3.3.1. Quality metrics

It is good to first define *quality metrics* and the *accuracy assessment* procedure, as used in this research. Both are part of the *performance evaluation* of a certain change detection methodology and its consecutive result. The former is used in the context of quantifying a classifier's performance, with tools and metrics that result in a certain comparable unit of measure. Accuracy assessment originates from image classification in remote sensing, for single images (Jianya et al., 2008). Accuracy assessment can be done in a quantitatively way using the quality metrics, but also in different ways, such as a visual comparison with the ground truth or a combination of the two (Radke et al., 2005).

To define different quantifiers, we first define the following quantities (Aminikhanghahi and Cook, 2017; Radke et al., 2005):

- *True positives* (TP); refers to the pixels that are correctly assigned to the change class.
- *True negatives* (TN); the pixels correctly assigned to the no-change class.
- *False positives* (FP) (or false alarms or type I error); pixels assigned to the change class, that should've been assigned to the no-change class.
- *False negatives* (FN) (or misses or type II error); the pixels that should've been assigned to the change class, but are assigned to the no-change class.

These four quantities can be used to construct a *confusion matrix* or error matrix, the most well known visualization of the performance of a classifier algorithm (Aminikhanghahi and Cook, 2017).

Table 3.1: An example confusion matrix of a binary classifier. A confusion matrix can be used to visualize the performance of classifier algorithm by tabulating the FPs, FNs, TPs and TNs.

Truth \ Classified	**Positive**	**Negative**
Positive	TP	FN
Negative	FP	TN

A lot of classifiers have been developed over the years, Rosin and Ioannidis (2002) and Radke et al. (2005) describe three methods to evaluate the performance with the *Percentage Correct Classification* (PCC), *Jaccard coefficient* (JC) and *Yule coefficient* (YC). Aminikhanghahi and Cook (2017) adds some additional classifiers to this list in the *precision*, *sensitivity*, *specificity*, *G-mean*, *F-measure*, *Recieving Operating Characteristics* (ROC) *curve* and the *Precision-Recall* (PR) *curve*. Additionally, Aminikhanghahi and Cook (2017) lists some performance metrics that include the difference in position, or change in time, of the detected change point with respect to the actual change points. These won't be discussed here, however. The different quality metrics, with accompanying explanation and equations can be found in Table 3.2 (Aminikhanghahi and Cook, 2017; Davis and Goadrich, 2006; Radke et al., 2005).

3.3.2. Accuracy assessment

Employing and comparing the quality metrics mentioned in Section 3.3.1 is dependent on the premise of the availability of reliable (ground) truth data. Moreover, it does not only assess the performance of a certain algorithm, but also incorporates pre-processing steps that might have introduced errors pre-analysis, including precise co-registration and calibration of the data (Jianya et al., 2008), (Lu et al., 2004). Even limitations or errors in the data play a vital role in the overall accuracy assessment of a CPD algorithm, such as a limiting image resolution. Shao (2006) divides the accuracy of CPD in 5 distinct classes, which can be summarized as *data errors and limitations*, *pre-processing errors*, *algorithm errors*, *ground truth errors* and *post-processing errors*.

Finding a suitable CPD method for a certain change detection problem can be difficult, even with the numerous quality metrics to compare them. Some problem statements might favour taking into account feature-levels, object-levels or the temporal trajectory of the data set. The first two impose a problem in the sense that the majority of research has been done on pixel-level based change detection, with little effort in the feature- and object-level change detection (Jianya et al., 2008). Taking the temporal trajectory into

account can introduce a different limitation for accurate accuracy assessment. Obtaining reliable ground references can be especially troublesome for long time series of data (Jianya et al., 2008). Lastly, one might want to incorporate higher-level constraints in the accuracy assessment, that reflect the ability of an algorithm to detect changes that are deemed most important (Radke et al., 2005). Pixels could be assigned weights for example, rather then being treated equally.

Table 3.2: Various quality metrics that are often used, provided with corresponding descriptions and equations when applicable. Quality metrics can be used to assess the performance of change detection algorithms, and compare the algorithms against one another.

Quality metric	Description	Equation	
PCC (or accuracy)	Provides a high-level measure of the performance of a classifier	$\frac{TP+TN}{TP+FP+FN+TN}$	(3.20)
Precision (or positive predictive value)	The ratio of correctly identified change points out of all change points. A negative predictive value also exists.	$\frac{TP}{TP+FP}$	(3.21)
Sensitivity (or recall or true positive rate)	The ratio of correctly identified change points out of all points. A true negative rate (or specificity), false negative rate and a false positive rate (or fall-out) can also be defined similarly.	$\frac{TP}{TP+FN}$	(3.22)
G-mean	Combines the ratio of true positive and true negative rate to overcome the small ratio of changes to the total data points.	$\sqrt{\frac{TP}{TP+FP}\frac{TN}{FP+TN}}$	(3.23)
F-measure (or F-score or F1 score)	A measure of overall effectiveness, that combines a weighted positive predictive value and true positive rate.	$\frac{(1+\beta)^2 \frac{TP}{TP+FN}\frac{TP}{TP+FP}}{\beta^2 \frac{TP}{TP+FN}+\frac{TP}{TP+FP}}$	(3.24)
JC	A variation of the accuracy measurement, that ignores large parts of stable background pixels, i.e. TNs.	$\frac{TP}{TP+FP+FN}$	(3.25)
YC	Similar to the JC in use. Useful for a low change point to data points ratio.	$\left\lvert \frac{TP}{TP+FP}\frac{TN}{TN+FN}-1 \right\rvert$	(3.26)
ROC curve	A 2D graph with the true positive rate plotted against the false positive rate for various thresholds. Both the ROC and PR curve can be used to assess the overall effectiveness and performance of alternative algorithms.	N/A	
PR curve	A 2D graph with the positive predictive value plotted against the true positive rate for various thresholds. This curve is preferred over the ROC curve for highly skewed data sets.	N/A	

3.4. Summary

Chapter 3 gives an overview of the current state of the art change detection techniques, in order to determine suitable CPD algorithms for the simulated, Houston and Rotterdam stack. Both bi- and multi-temporal approaches are reviewed, not necessarily previously used with SAR data. Moreover, in order to assess the results of the change detection algorithms, various components of the performance evaluation are analysed.

A change point in a time series occurs when at some point in the series the statistical distribution changes in at least one criterion, such as the mean. A large selection of methods to detect these changes exist, both previously used with SAR time series and beyond. CPD algorithms are classified as offline or online, and of a supervised or unsupervised nature. Multiple classification frameworks have been proposed in order to help make an informed decision on which algorithm is the most applicable to some change detection problem.

CPD with SAR data is mostly done in a bi-temporal way, evident from the lack of multi-temporal approaches in literature. Therefore, a selection of often used CPD techniques using bi-temporal SAR, SAR time series and multi-temporal time series in general are listed. In the bi-temporal case the temporal pattern is neglected, often in favour of including spatial relations via MRF models, clustering methods and so forth. Multi-temporal change detection, in- and outside of SAR-based data, has a much larger emphasis on pixel-level CPD. Applied bi- and multi-temporal approaches share certain common practices, such as spatial filtering to reduce speckle, using smart operators like the log-ratio operator for bi-temporal and generalised temporal means for multi-temporal data, and an emphasis on statistical methods.

Comparing the segmentation results of various methods requires the use of comparable quality metrics. These quality metrics are derived from 4 quantities; true positives, true negatives, false positives and false negatives. To compute these quality metrics, ground truth data is required. Quality metrics can be used for the overall accuracy assessment of a change detection procedure. The assessment procedure should also include reflection on the sources of errors, by considering data errors and limitations, pre-processing errors, algorithm errors, ground truth errors, and post-processing errors.

In the next chapter the survey on change point algorithms, classification frameworks and performance evaluation in this chapter is used to make an informed decision on what change point algorithms to use in this research. This procedure starts by first introducing the data sets and looking at the statistics of these sets. Moreover, the absolute calibration and CPD algorithm implementation methodology is discussed. Lastly, a framework is created in which the performance of the chosen CPD algorithms is evaluated. This is done by assessing specific aspects (micro evaluation) of the algorithms' performance and by looking at the overall performance (macro evaluation).

4

Implementation of methods for change point detection

Chapter 4 is aimed at outlining the methodology of implementing the used change point detection methods, namely Differencing, CUSUM, ANOVA and PELT. ANOVA and PELT were already introduced in Section 3.2. The used assumptions, steps for implementation and rationale behind the chosen algorithms is covered. Pre-processing (e.g. data selection and calibration) and post-processing (evaluation) steps are discussed as well.

4.1. Image pre-processing for change detection

4.1.1. Data selection and test locations

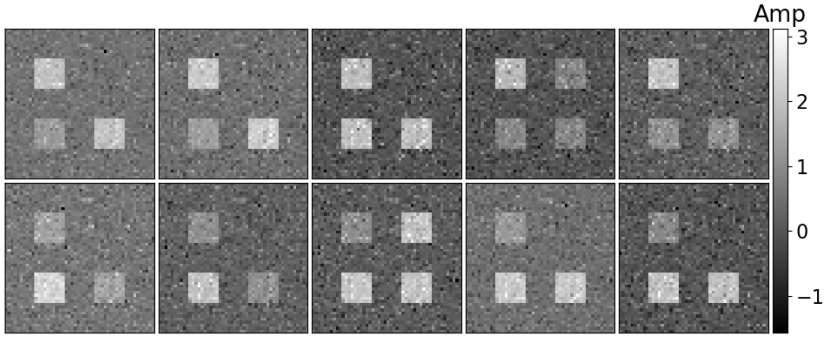
In Section 1.3 the three different types of data sets used throughout this research were briefly touched upon. These three data sets are:

1. simulated data, following the model descriptions as given in literature;
2. the *Houston stack*, a co-registered Sentinel-1 Terrain Observation by Progressive Scans mode (Interferometric Wide Swath) SLC stack in the Houston area from 2015/07/18 to 2017/09/17 containing 94 VV images in a descending orbit geometry, from swath 1; and
3. the *Rotterdam stack*, a co-registered descending TerraSAR-X StripMap mode stack in the Rotterdam area from 2009/04/08 to 2017/08/22 containing 240 HH images.

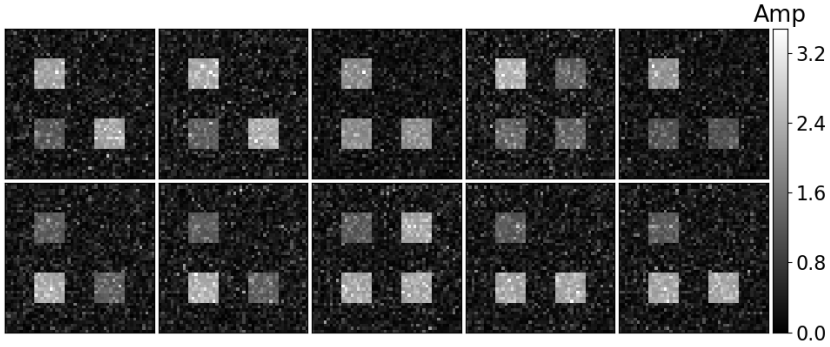
The simulated stacks follow the product model description, introduced in Section 2.2, Equation (2.27). For the texture t, assumed constant over time, random samples are generated from a distribution. Change regions are then imposed on the texture images, where the type of signal and the amount of offset are both chosen variables. These texture images are then subjected to multiplicative speckle s. The speckle is modelled by a normalized Rayleigh distribution. A realization of this modelling approach can be seen in Figure 4.1.

Using a simulated stack of images has a couple of distinct advantages over using real SAR data. Working in a controlled environment means the user gets to decide the boundaries of the simulation. The distribution of the underlying noise-free texture, the shape and magnitude of the spatial change regions and the type of temporal signal response linked to the change (i.e. a Dirac or Heaviside response) for example. Unlike real data, simulated data isn't dependent on ground truth data from external sources. This makes it the ideal data source to compute and assess various quality metrics. Two simulated stacks are considered, one following a texture that is Normally distributed and one with a 'standard' Rayleigh distribution. For convenience, these simulated stacks are referred to as the *Gaussian stack* and *Rayleigh stack*, respectively. Both are subject to normalized Rayleigh distributed multiplicative speckle.

The Houston and Rotterdam stack contain a significant amount of data points, measuring 227 MB and 0.98 GB per image respectively. Therefore a subset is selected for each area of interest. To emphasize the importance of data selection; the Houston stack consists of roughly 3 billion pixels if the entire area spanned by the co-registered images would be considered. Figure 4.2 shows the mean amplitude image of both subsets, by taking the mean amplitude value along the time axis for each pixel. Because both data cubes are in a descending orbit geometry, the images are flipped along the vertical for each visualization of the stacks.



(a) Gaussian stack.



(b) Rayleigh stack.

Figure 4.1: The simulated data stacks, with texture following a Normal distribution. A total of 10 large squares can be seen, each representing a SAR image of 50 pixels by 50 lines at some point in time. On each simulated acquisition up to 4 smaller 10 by 10 squares can be seen. These are superimposed change regions that are subject to significant changes over time.

(a) Houston sensed with S1 (VV).

(b) Rotterdam sensed with TSX (HH).

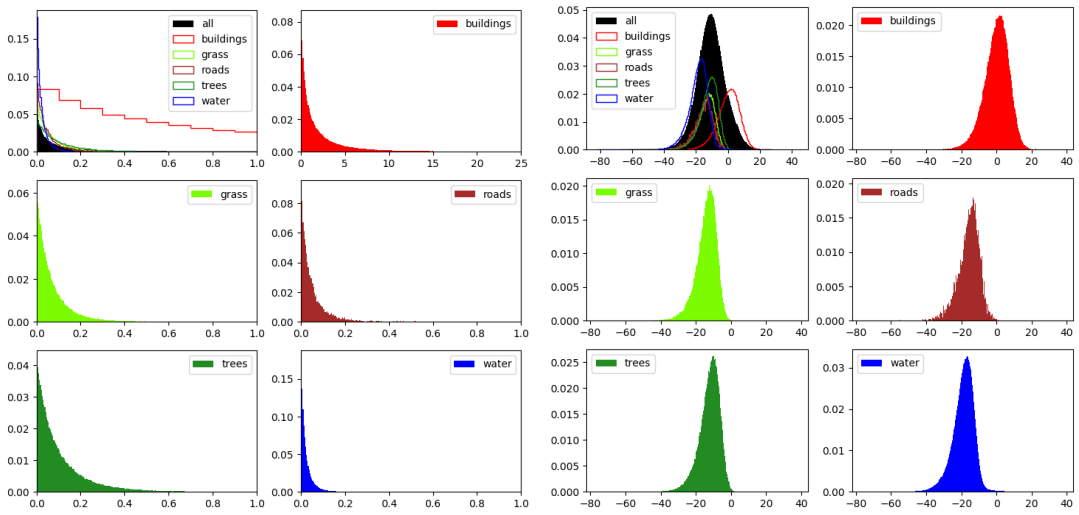
Figure 4.2: Mean amplitude images for the Houston and Rotterdam subsets. Both stacks show similar ground features, including a canal and similar type of buildings. However, the Rotterdam stack consists of more images, 240 against 94, and has a superior resolution [Left: flipped Houston stack with an aspect ratio of 3 (vertical:horizontal) to improve visibility, Right: flipped Rotterdam stack].

It can be seen that both areas of interest have comparable ground features, including a canal, some high rise buildings, low rise building and shipping containers. However, the Rotterdam stack looks much sharper due to it's superior resolution. An overview of the characteristics of the Houston and Rotterdam stack can be found in Table 4.1. If the statistical distributions of various landcovers of the sigma nought calibrated Sentinel-1 subset are analysed, the sigma nought values show a negative exponential distribution (Figure 4.3a). Recall that sigma nought is a measure of intensity, as DN^2 is used in the equations. When the calibrated data is transformed to dB, following Equation (2.18), an approximate normal distribution is observed (Figure 4.3b). These distributions do all show a slight negative skew. Furthermore, the general shape of the histograms

and the standard deviation of each land covers is comparable. The most discriminating factor is the mean backscatter response.

Table 4.1: An overview of the characteristics of the Houston and Rotterdam stack to summarize and highlight the differences and similarities between the stacks [(...): notation, [...]: unit, S/c: spacecraft, Res.: (spatial) resolution, Pol.: polarisation, Orb. geom.: orbit geometry, Acc.: absolute radiometric accuracy, NESN: noise equivalent sigma nought, Std: relative radiometric accuracy, A: azimuth, R: range, L: lines, P: pixels, T: time].

	S/c	Res. (AxR) [m]	Cycle [day]	Sensor	Pol.	Orb. geom. (ASC/DSC)	Mode	Acc. [dB]	NESN [dB]	Std [dB]	Size (LxPxT)
Houston	S1A & S1B	5x20	6	C-SAR	VV	DSC	IW	<1	<-22	0.5	201x501x93
Rotterdam	TSX & TDX	3x3	11	X-SAR	HH	DSC	SM	<0.6	<-19	0.3	501x1001x240



(a) Linear (b) Logarithmic (dB)

Figure 4.3: The histograms for various land covers of the Sentinel-1 stack, with the values on the x-axis and the normalized frequency ([0,1]) on the y-axis. Note that for the individual land cover histograms the y-axis varies. It can be seen that after a transformation from a linear scale to decibel notation the data stacks follow a near-Normal distribution. Moreover, the most significant difference between land covers is the mean value (e.g. 2 dB for buildings and -19 dB for water), whereas the overall histogram shape and standard deviation is comparable. [Left: exponential distribution showing exponential decay, Right: an approximate Gaussian distribution with a small negative skew].

As discussed in Section 2.5, array-oriented data can be stored in a structured file format offering some distinct short-term and long-term advantages over plain binary file storage. The subsets are stored in a HDF5 file format, without any chunking or compression steps. Chunking in particular could speed up the change point detection process significantly, if done properly. Even without chunking a significant increase in speed is observed, mainly when quickly analysing parts of the data. Furthermore, the HDF5 file allows for the storage of multiple data sets of the same area of interest in the same environment, i.e. the radiometric calibrated data.

4.1.2. Radiometric calibration methodology

Section 2.4 introduced the equations needed to calibrate level-1 SLC data generated by Sentinel-1 and TerraSAR-X. The used stacks are pre-processed by the Doris InSAR processor (DEOS, 2008). These pre-processing steps include co-registration of each slave image w.r.t. a chosen master image and resampling of the images, yielding an output file of a data type that is major-row ordered pixel-interleaved with 32 bit floats (.raw extension). These files don't include any header information, as relevant information is stored in separate log files (.res extension). These two files, alongside a calibration file (.xml extension) storing the needed variables, altogether form the required input for the absolute calibration. The steps taken to calibrate the SAR images is best summarized with a flowchart, see Figure 4.4.

Besides the different calibration equations, the actual workflow is slightly different as well for the two sensors. Sentinel-1 requires the selection of the particular swath and burst that the user wants to calibrate. As the calibration file is for the complete swath, it is important to choose the normalization areas from the

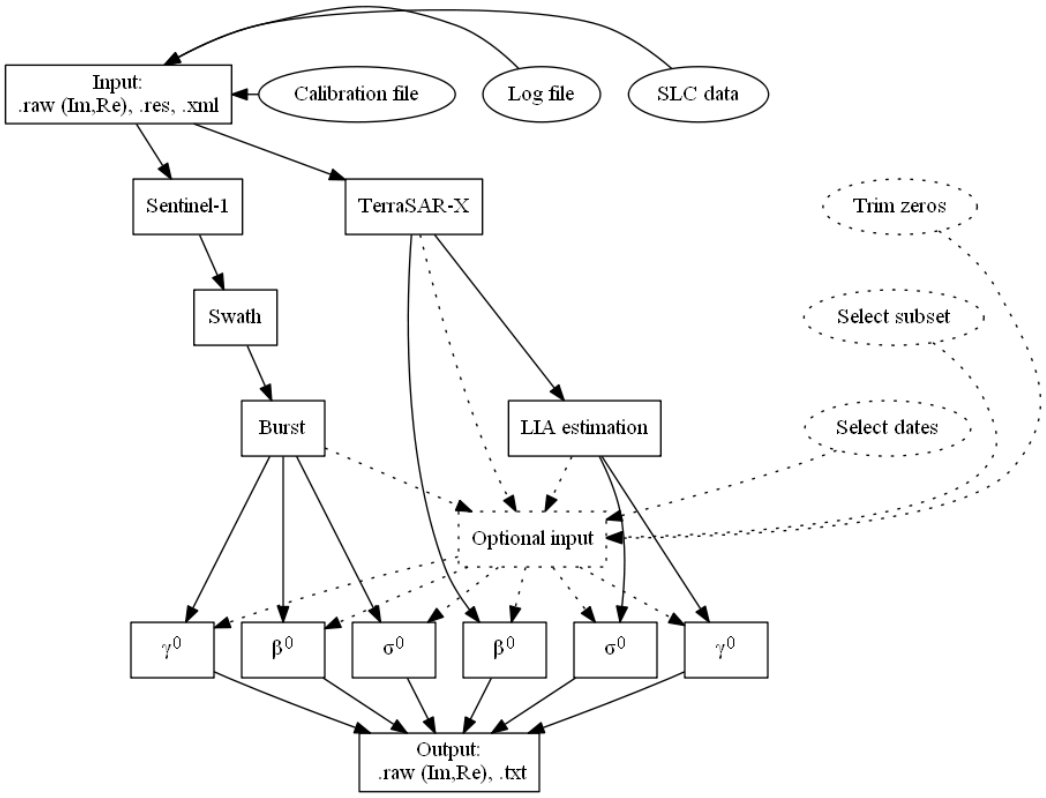


Figure 4.4: A flowchart of the absolute calibration implementation. Sentinel-1 and TerraSAR-X follow a different methodology, as can be seen. Most noticeably, in the absence of the EEC product, the absolute calibration of TerraSAR-X requires an estimation of the local incidence angle (LIA) to calibrate to sigma nought or gamma nought values. After calibration the phase information is preserved as well.

LUTs that correspond with the correct relative range and azimuth lines. Lastly, some optional input can be specified, namely to trim complete zero-valued rows or columns, to select a subset of the image and to specify the precise dates to calibrate.

TerraSAR-X's calibration files only provide five local incidence angles (LIAs), as mentioned in Section 2.4.3. To provide a better estimate of the LIA a plane has been fitted through these five points using linear Least Squares (LS) estimation. A plane can be described as (e.g. Eberly, 1999)

$$\begin{aligned} &ax + by + cz + d = 0 \\ &-z = ax + by + d, \qquad \text{if } c = 1. \end{aligned} \tag{4.1}$$

Here x and y should be interpreted as the lines and pixels, and z the local incidence angle. Equation 4.1 yields a model, which can be solved for a, b and d using linear LS

$$\begin{bmatrix} -z_1 \\ -z_2 \\ \cdots \\ -z_n \end{bmatrix} = \begin{bmatrix} x_1 & y_1 & 1 \\ x_2 & y_2 & 1 \\ \cdots & \cdots & \cdots \\ x_n & y_n & 1 \end{bmatrix} \begin{bmatrix} a \\ b \\ d \end{bmatrix}. \tag{4.2}$$

Equation (4.2) is of the form $\mathbf{y} = \mathbf{A}\mathbf{x}$, with $\mathbf{y}$ the vector of observations, $\mathbf{A}$ the design matrix and $\mathbf{x}$ the vector of unknown variables. To estimate the unknown variables, LS is used, so that

$$\hat{\mathbf{x}} = (\mathbf{A}^\intercal \mathbf{A})^{-1} \mathbf{A}^\intercal \mathbf{y} \tag{4.3}$$

with $\hat{\mathbf{x}}$ the estimator of the unknowns and where $\{.\}^\intercal$ and $\{.\}^{-1}$ denote the transpose and inverse, respectively. By solving Equation (4.3) an estimation is obtained for the LIA over the entire area of interest. Furthermore, the optional input parameters are the same as those of Sentinel-1.

Both sensors yield the same output; the now calibrated data (.raw extension) and a log file (.txt extension) that contains some important information, such as the polarization of the data. During this research the

sigma nought calibrated values are used, as these are directly related to the ground, allowing for better interpretation of the images. Moreover, beta nought doesn't account for the incidence angle. A ~0.6 dB difference between near and far range is observed for the entire beta nought calibrated Sentinel-1 swath (Appendix A, Figure A.1).

4.2. Change detection methods and algorithms implementation

4.2.1. Algorithm selection

Four algorithms are chosen, based on the conducted survey in the CPD research field. These methods combined give a good balance of computational efficiency, complexity in methodology and the level of sophistication in which the temporal evolution is considered. As shown in Section 4.1.1, the sigma nought values, transformed to decibels, of the test sites follow a slightly skewed Normal distribution. This is important to take into account when choosing an CPD algorithm. The algorithms that are chosen are Differencing, CUSUM, ANOVA and PELT. These four algorithms can be placed in the classification frameworks given in Section 3.1.2, to give context to them (Table 4.2). In particular, how the temporal evolution of a pixel is considered with each algorithm.

Table 4.2: Differencing, CUSUM, ANOVA and PELT placed in the classification frameworks given by Aminikhanghahi and Cook (2017), Jianya et al. (2008), Radke et al. (2005) and Lu et al. (2004), introduced in Section 3.1.2. Most important to note is how the temporal evolution of a pixel is considered in each of the algorithms. Differencing uses an iterative bi-temporal approach, whereas ANOVA performs a proper time series analysis, for example.

Method	Aminikhanghahi and Cook (2017)	Jianya et al. (2008)	Radke et al. (2005)	Lu et al. (2004)
Differencing	Unsupervised; N/A	Direct comparison	Simple differencing	Algebraic
CUSUM	Unsupervised; N/A	Hybrid method (Direct comparison and time series analysis)	N/A	Algebraic
ANOVA	Unsupervised; Probabilistic method	Time series analysis	Significance and hypothesis tests	Other
PELT	Unsupervised; Probabilistic method	Time series analysis	N/A	Other

Differencing and CUSUM have some overlap in approach, as both use the relative differences in amplitude of consecutive measurements. The implementations of Differencing and CUSUM are both $\mathcal{O}(n)$, with n the input size. The Differencing approach is really straightforward, fast and easy to implement, and therefore makes it a good baseline method. It does not take the temporal behaviour of the time series into account. This makes it a bi-temporal method extended to the multi-temporal domain. Variations on the Differencing approach, using simple thresholds, are widely used in SAR change detection.

The CUSUM algorithm was selected as it takes into account the temporal evolution of the time series in contrast to the suggested Differencing method, while still being easy to understand and implement. Moreover, CUSUM change detection has never been applied on SAR time series. It incorporates all the information of the samples of a sequence by using the cumulative sum of the deviations of those samples to a target, or threshold, value. Moreover, CUSUM is known to be good at detection small shifts in an observed parameter (Page, 1954). These two methods will give an insight in the reliability and overall applicability of simple algebraic CPD methods in combination with SAR time series, influenced by the inherent speckle.

ANOVA was chosen because it takes uses the data statistics and fully utilizes the temporal evolution of the time series by using the between- and within-group relationships. The significantly long time series in the used data cubes and the methodology of ANOVA should make it fairly robust against speckle. Previous implementation in literature by Li et al. (2011) and Dogan and Perissin (2010) of ANOVA-based change detection on SAR time series show promising results. Where Dogan and Perissin (2010) improved the methodology of Li et al. (2011), this research attempts to improve upon the implementation of the former by inspecting

all created segments for change points in an iterative way. ANOVA is the most computationally expensive algorithm of the four, with $\mathcal{O}(n \log n)$.

Lastly, PELT was chosen, as it has never been used for SAR based change point detection, even though it offers some great benefits, such as linear computational cost ($\mathcal{O}(n)$) under certain conditions and a choice of cost function dependent on the input data. Variations of the underlying search method of this algorithm have seen widespread use in literature.

4.2.2. Differencing implementation

In Radke et al. (2005) and Lu et al. (2004) an approach for *(simple) differencing* is elaborated, by taken the difference of two consecutive images and thresholding the difference image

$$DI_i = I_{i+1} - I_i, \tag{4.4}$$

which defines a binary decision rule based on one simple threshold

$$\delta = \begin{cases} 0, & \text{if } |DI_i| \leq \tau \\ 1, & \text{if } |DI_i| > \tau. \end{cases} \tag{4.5}$$

This approach is purely based on an arithmetic operation and a threshold chosen by the user and therefore very sensitive to the chosen threshold. The actual threshold selection will be discussed in a separate section, namely Section 4.3.

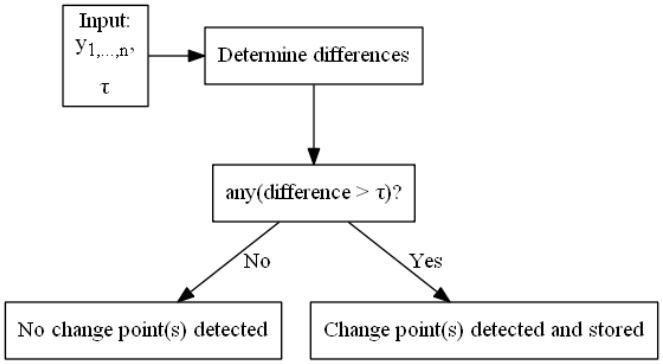


Figure 4.5: The Differencing algorithm flowchart. Differencing is a basic approach to the change detection problem, by simply comparing differences in amplitude of consecutive measurements against a threshold.

When converting the linear values to a decibel scale, an alternative approach is identified, closely related to the log-ratio operator

$$DI_{i,\mathrm{dB}} = 10\log(I_{i+1}) - 10\log(I_i). \tag{4.6}$$

This alternative approach, using the dB scale, is considered superior for the same reasons the log-ratio operator is so widely used (e.g. minimizing the effects of speckle). Therefore, Equation (4.6) is used. In order to make the algorithm less sensitive to a high degree of speckle and to counter overfitting, a guard can be introduced. This approach follows the same methodology as ANOVA, discussed in Section 4.2.4. While a time series has more samples than the defined guard dictates, the series is iteratively segmented at the largest offsets if Equation (4.5) yields a positive outcome.

4.2.3. CUSUM implementation

The CUSUM algorithm has seen many adaptations over the years and can be traced all the way back to Page (1954) who introduced the tabular CUSUM chart. Each approach follows the same basic premise; let θ be a term that defines a parameter in the probability distribution of a continuous random variable, e.g. the mean or variance. The cumulative sums of the derivations of the sample values of such a parameter are then collected over time. This is a measure of (significant) change in a process that can be used for change detection.

Monitoring the mean of a sequence has seen two implementations; the tabular (or algorithmic) approach and the V-mask by Barnard (1959). The tabular approach is the one that prevailed, due to its relative simplicity compared to the V-mask approach (Nenes and Tagaras, 2005). Originally only the upward shifts in mean were observed in the one-sided CUSUM test, however this test can be extended to a two-sided CUSUM test, where the CUSUM tests for positive and negative change are formulated as

$$g_i^+ = g_{i-1}^+ + s_i - \nu, \tag{4.7a}$$

$$g_i^- = g_{i-1}^- - s_i - \nu, \tag{4.7b}$$

where g_i^+ and g_i^- are referred to as the one sided upper and one sided lower respectively (Gustafson, 2000), ν is a drift term used to prevent positive drifts eventually triggering a false alarm and $s_i = y_i - y_{i-1}$. To prevent a slower change detection time, i.e. negative drift, the test statistic g_i is reset when it becomes smaller than 0

$$g_i^+ = \max\left(0, g_{i-1}^+ + s_i - \nu\right), \tag{4.8a}$$

$$g_i^- = \max\left(0, g_{i-1}^- - s_i - \nu\right). \tag{4.8b}$$

If this reset value is not equal to 0, but rather some arbitrary negative value, the algorithm is referred to as the Sequential Probability Ratio Test (SPRT) (Gustafson, 2000). The test statistics can now be compared to some threshold τ, so that a decision rule can be formulated

$$\delta = \begin{cases} 0, & \text{if } g_i^+ \leq \tau \wedge g_i^- \leq \tau \\ 1, & \text{if } g_i^+ > \tau \vee g_i^- > \tau. \end{cases} \tag{4.9}$$

The chart in Figure 4.6 shows the general flow of the algorithm. It has to be noted that the algorithm also returns the index at which the change was detected.

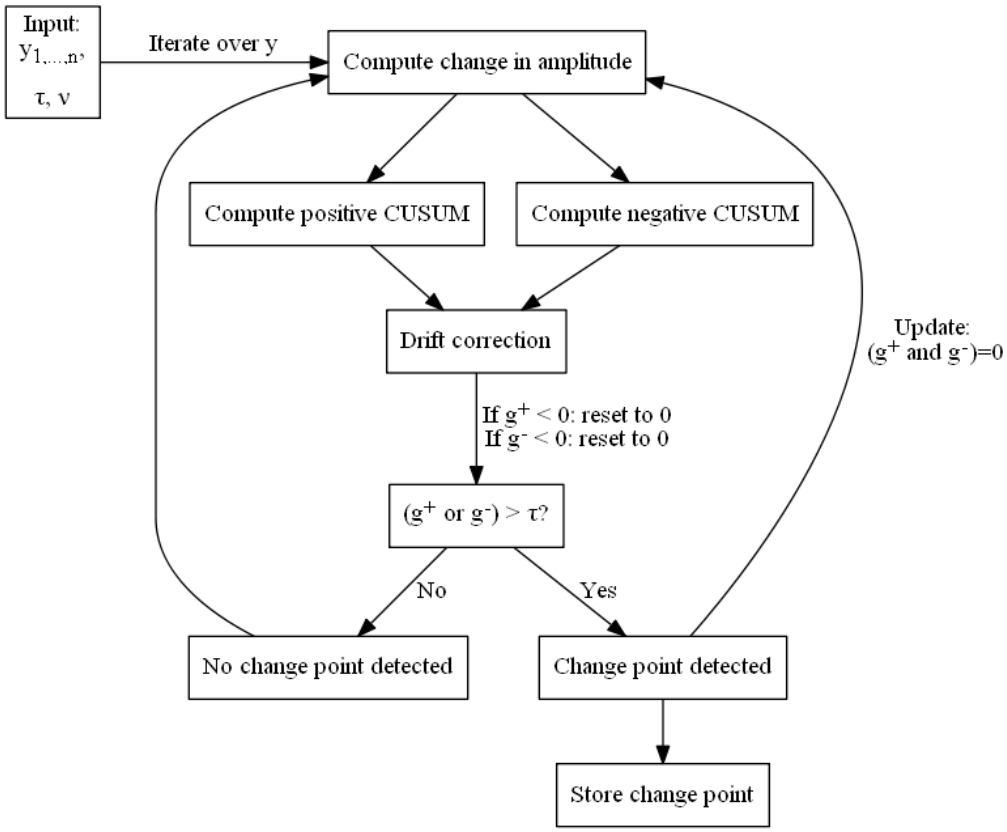


Figure 4.6: The CUSUM algorithm flowchart. CUSUM partly takes the temporal evolution of a pixel into account by considering the cumulative sum of previously determined changes in amplitude of consecutive measurements.

4.2.4. ANOVA implementation

The ANOVA method has been briefly introduced in Section 3.2.3. ANOVA is an F-statistic, which is a test statistic that can be used to estimate the variation between and among groups in order to analyse the differences among group means in a sample. Staring with a sequence $y_{1,\dots,n}$ with N samples, one can split the data in an arbitrary number of groups K and compare those using ANOVA, making it a measure off dissimilarity between groups. However, the used adaptation of the algorithm always uses $K = 2$ groups; the reasoning behind this will be discussed in Section 4.3.

A slightly modified version of the approach suggested by Dogan and Perissin (2010) is proposed. The proposed method is different from Dogan and Perissin (2010) in the sense that rather than segmenting a time series at point τ and only using an iterative approach for the last segment $y_{\tau+1,\dots,n}$; both the newly created segments ($y_{1,\dots,\tau}$, $y_{\tau+1,\dots,n}$) are iteratively analysed for more change points. Let's first look at the general flowchart (Figure 4.7) to get a better understanding of the flow of the algorithm.

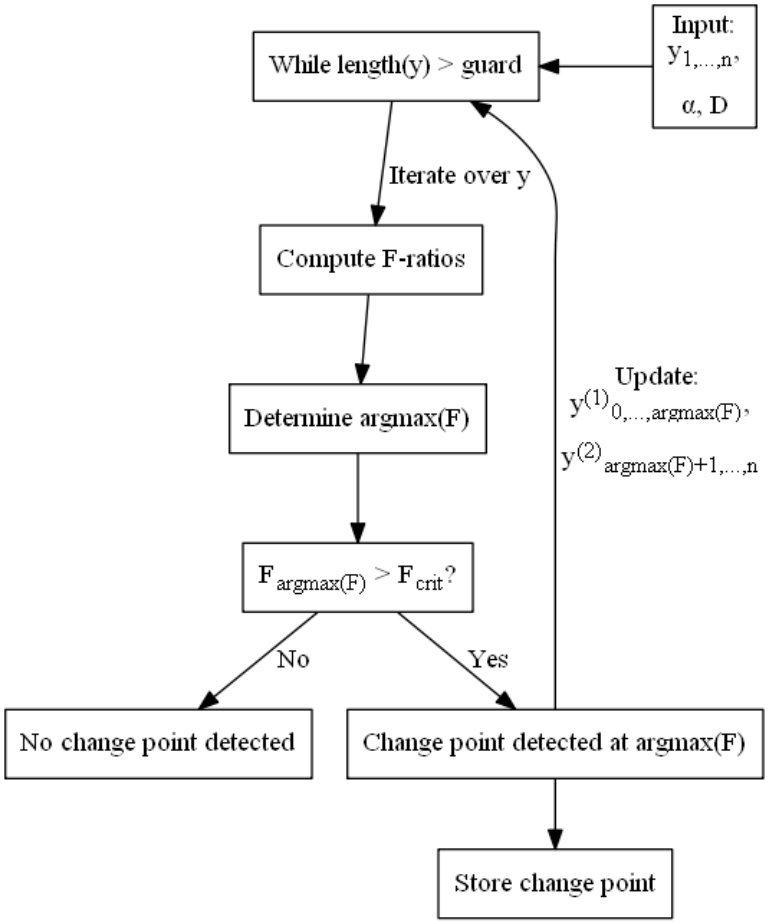


Figure 4.7: The ANOVA algorithm flowchart. ANOVA performs a proper time series analysis and takes all measurements into account before initiating the decision rule, by comparing the within- and between-group variability of all the combinations of measurements in two consecutive groups.

After comparing the length of the sequence to some guard D that ensures sufficiently long sequence, the F ratios of all the possible combinations of two consecutive groups are computed. Under the null hypothesis (see section 3.1.1) for constant means, or the no-change hypothesis, the F values for the single transition case are defined as

$$F = \frac{N-K}{K-1} \frac{\sum_{i=1}^{K} N\left(\mu_i - \mu\right)^2}{\sum_{i=1}^{K}\sum_{j=1}^{N}\left(A_{ij} - \mu_i\right)^2}. \tag{4.10}$$

Here N and K still denote the number of samples and groups respectively, μ is the overall sequence mean, μ_i the group means and A_{ij} the actual pixel values corresponding to a certain group. Equation (4.10) consists of two parts, the within-group (wg) variability and between-group (bg) variability, such that Equation (4.10) can be interpreted as

$$F = \frac{\mathrm{MS}_{\mathrm{bg}}}{\mathrm{MS}_{\mathrm{wg}}} = \frac{\frac{\mathrm{SS}_{\mathrm{bg}}}{\mathrm{DF}_{\mathrm{bg}}}}{\frac{\mathrm{SS}_{\mathrm{wg}}}{\mathrm{DF}_{\mathrm{wg}}}}, \tag{4.11}$$

where DF is the degrees of freedom, SS the sum-of-squares and MS the variability.

After computing the F ratios of each combination of $K = 2$ groups, the position of the maximum F value $\hat{i}$ is determined, which is the most likely change point candidate

$$\hat{i} = \underset{1 \leq i \leq N-1}{\arg\max} \left(\frac{N-K}{K-1} \frac{\sum_{i=1}^{K} N\left(\mu_i - \mu\right)^2}{\sum_{i=1}^{K} \sum_{j=1}^{N} \left(A_{ij} - \mu_i\right)^2} \right). \tag{4.12}$$

This candidate change point is then compared to a critical F value F_{crit}, which is taken from a Fisher-Snedecor distribution, or F-distribution for short. This distribution is governed by the degrees of freedom of the within- and between-group variability and the critical F value is ultimately dependent on a users input of significance level α. The F-distribution reads as

$$f(x; d_1, d_2) = \frac{\sqrt{\frac{(d_1 x)^{d_1} d_2^{d_2}}{(d_1 x + d_2)^{d_1 + d_2}}}}{x B\left(\frac{d_1}{2}, \frac{d_2}{2}\right)}, \tag{4.13}$$

where d_1 and d_2 are the (positive integer) degrees of freedom of the within-group and between-group variability and B is the Beta function. The actual selection of this threshold value F_{crit}, based on α, is also discussed in Section 4.3.

This ANOVA approach yields a decision rule for a single transition instance that embodies the null hypothesis of constant mean and an alternative hypothesis of a changed mean at a certain instance i

$$\delta = \begin{cases} 0, & \text{if } F_{max} \leq F_{\mathrm{crit}} \\ 1, & \text{if } F_{max} > F_{\mathrm{crit}}. \end{cases} \tag{4.14}$$

To detect multiple change points, this approach is repeated iteratively. For each of the two newly defined segments the whole procedure is repeated, until either the guard or the threshold condition is not met, at which point the algorithm breaks for that particular segment.

4.2.5. PELT implementation

In Section 3.2.1 the Optimal Partitioning and Pruned Exact Linear Time algorithms were introduced. This mainly focussed on the reasoning behind these algorithms, as well as the decrease in time complexity made possible by the pruning step in PELT over OP. To quickly summarize; PELT (and OP) attempt to minimize Equation (3.3) by using dynamic programming to find the optimal partitioning of a sequence in a recursive manner.

PELT first has to be initialised by choosing a starting value for the penalty term β, which as discussed, is often taken as linear in the number of change points. Multiple examples are given in literature, e.g. Akaike's Information Criterion (AIC) (Akaike, 1974) of the form $\beta = 2k$ and Schwarz (or Bayesian) Information Criterion (BIC/ SIC) (G., 1978) of the form $\beta = k \log(N)$, where k is the number of free parameters to be estimated (G., 1978), (Killick et al., 2012). In this research a penalty function of $\beta = \log(N)$ is used (e.g. Rohrbeck, 2013). An optimal choice of penalty is still an open research question, as noted by Killick (2017). This chosen penalty is subsequently used to initialise the cost of $F(0)$, where $F(0) = \beta$.

To evaluate the cost for segmentation (F), a cost function has to be defined. It is assumed that the time series is described by a Normal distribution, as discussed in Section 4.2.4. For the application of PELT on Normally distributed data, the cost function of a segment of the sequence $y_{1:n}$ has to be picked. Twice the negative log-likelihood is chosen to model the cost for a segment $(y_{s+1}, ..., y_t)$, with $0 \leq s \leq t \leq n$, such that (Rohrbeck, 2013)

$$C(y_{s:t}) = -2l(\mu, \sigma_s, ..., \sigma_t | y_s, ..., y_t) = \sum_{i=s}^{t} \left[\frac{(y_i - \mu)^2}{\sigma_i^2} + \log\left(2\pi\sigma_i^2\right) \right]. \tag{4.15}$$

This likelihood approach is commonly used as a cost function in change point literature (Killick et al., 2012). For a negative log-likelihood, a constant $K = 0$ is used in order to meet the condition to do pruning (Killick et al., 2012). Furthermore, as mentioned in Section 1.3, a focus lies on a change in mean. Therefore, a constant standard deviation σ is assumed. The whole procedure for the PELT algorithm is summarized in a flowchart, see Figure 4.8.

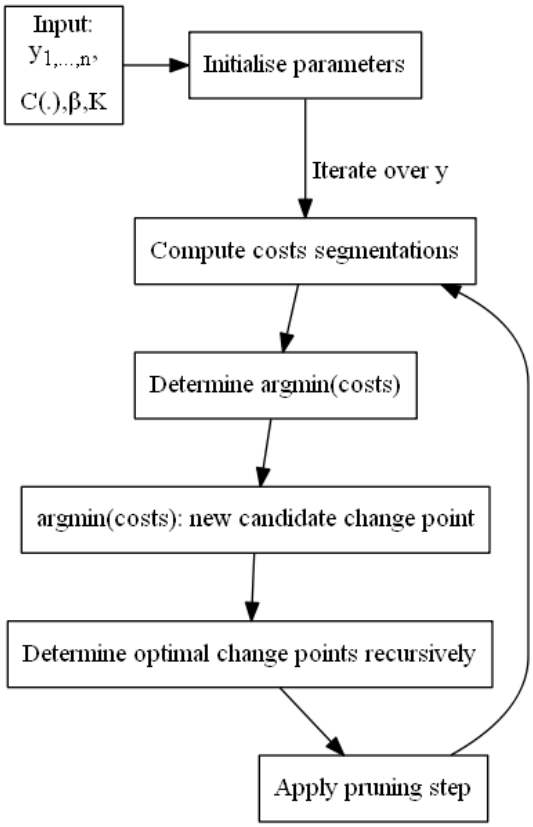


Figure 4.8: The PELT algorithm flowchart. PELT uses dynamic programming to recursively determine the optimal partitioning of a time series, by minimizing the cost of the penalised case.

First the algorithm is initialised, by setting $F(0) = \beta$, $K = 0$ and choosing a cost function (Equation (4.15)). After initialising, the cost of each possible segmentation is computed by iterating over the time series. From all these segmentations up until i, the segmentation yielding the lowest cost is selected and stored as a candidate change point τ_i, alongside its cost. Now the cost of the new optimal partitioning of the data up until i is calculated by recursion using the previously calculated costs, following Equation (3.6). Lastly, a performance increasing pruning step is applied removing previous τ_i's which cannot yield an optimal partitioning, as explained in Section 3.2.1. This is repeated until the end of the time series, resulting in a list of change points that segment the series in an optimal way.

4.3. Threshold selection

Proper tuning of the Differencing, CUSUM, ANOVA and PELT algorithms ultimately determines the success rate in the detection of change points. Therefore, a separate section is dedicated to threshold selection. It has to be noted that PELT doesn't use a threshold per definition, but is rather governed by the choice of penalty and the expected standard deviation of the time series. The choice of constant standard deviation is partly

responsible for the cost in segmentation, whereas the penalty is used to prevent over fitting; similar to the drift term in CUSUM.

The Differencing algorithm is very sensitive to speckle, especially without any filtering steps, therefore making threshold selection a delicate matter. As the data has been calibrated in sigma nought, a quantity of *absolute* calibration, documented backscatter intensities of various objects can be used as a guideline for the tuning of the algorithm. Alternatively, a reference time series can be used. This time series is then segmented using Differencing, yielding a certain threshold τ used image wide. Here, a combination of the two is used.

The CUSUM algorithm is tuned, following Gustafson (2000). A drift is chosen to half the expected change in mean, i.e. $\nu = 0.5\tau$. An increase in drift will result in fewer false alarms, at a slower detection rate. This also results in an automatic dependency of the drift on the chosen threshold value. This threshold value is chosen based on minimizing the false alarm rate, as observed in a reference time series. Moreover, some analytical expression for the drift have also been proposed in literature (e.g. Cheng and Yu, 2014). Lastly, it is good to note that an optional Fast Initial Response (FIR) can be used. This means that the CUSUM test statistic g_i is given a head start by setting $g_0 > 0$, possibly identifying change points more effectively.

ANOVA thresholding is done indirectly, based on an input value in the level of significance α that yields a critical F value F_{crit} from a F-distribution governed by the degrees of freedom, as seen in Figure 5.21. The significance level is related to the quantile q of the F-distribution as $q = 1 - \alpha$. Furthermore, a guard D has to be defined. This guard ensures a minimum number of observations for a sequence, i.e. a guard against over fitting. With a repeat cycle of 12 and 11 for Sentinel-1 and TerraSAR-X the guard is set at $D = 3$, equivalent to at least a temporal span of a month. Please note that Sentinel-1 'effective' repeat cycle can be reduced to 6 days when data from both S1A and S1B are used. This guard is also used for the Differencing approach.

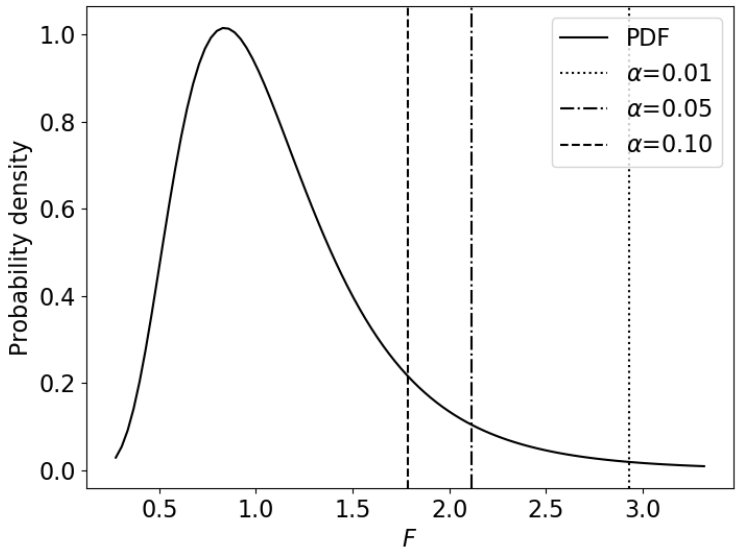


Figure 4.9: Critical F value selection for different significance levels, as used in the ANOVA algorithm. A lower significance level yields a higher threshold.

In general, there is no one-size-fits-all threshold for each method. Threshold selection should be carefully approached and assessed, with e.g. ROC curves. For all the following applications of the algorithms in Section 5 the used parameters will be specified.

4.4. Evaluation

4.4.1. Evaluation criterion

The evaluation of the different algorithms is split into six parts. These six parts combined will allow for a better substantiated conclusion on the strengths, weaknesses and overall applicability of Differencing, CUSUM, ANOVA and PELT on SAR amplitude time series. To generalize; a distinction between micro and macro evaluation is made. The former focusses on specific aspects of the CPD, whilst the latter is aimed towards evalu-

ation of the overall solution. For the micro evaluation, five tests are performed.

1. The sensitivity of each algorithm to each of the three defined signal response types (i.e. Heaviside, Rectangular and Dirac) is assessed by using simulated data. First, a mixture of all three signal types is used to get a general picture, before assessing each signal type separately. For the latter some quality metrics are computed per signal type, namely the accuracy, precision and ROC curve. These were already discussed in Section 3.3.1.

2. To cover the 'grey' area between Dirac pulses and a clear Rectangular response, a separate sensitivity analysis is performed with synthesized data. Here, aspects such as the needed number of subsequent significant offsets and the magnitude of these offset are evaluated. This is only done for ANOVA and PELT, as Differencing and CUSUM are fairly straightforward methods. The exact methodology will explained alongside the test results in Section 5.3.

3. The sensitivity of the algorithms to non-normality is assessed by comparing the time series solution with its quantile normalized counterpart. Again, this is only done for ANOVA and PELT in the separate section on sensitivity analysis. The methodology of quantile normalization is explained in Section 4.4.2.

4. The effects of speckle on the CPD solution is assessed by visually comparing the results of the unfiltered Houston stack, with a spatiotemporal filtered solution. This is done with non-weighted moving average filtering in the spatial (second order neighbourhood) domain, following Equation (2.16) from Section 2.1.5.

5. The impact of (spatial) resolution is evaluated by comparing unfiltered Sentinel-1 and TerraSAR-X CPD solutions. Here, the results in a spatial setting are also used, i.e. on image scale, rather than comparing time series.

Usually, the assessment of the overall change map solution of a single image pair relies on a ground truth map to compare to. However, these are largely absent or hard to come by or create. Not to mention that this is aggravated by taking the temporal component into account. Therefore, the temporal density of change points and the direct neighbourhood of change pixels are taken into account; being the final part of the evaluation procedure.

6. The temporal and spatiotemporal density of the change points is assessed. The creation of the temporal change point density map is a simple summation of the detected change points along the time axis. The spatiotemporal change density map sums the number of changes found in the second order neighbourhood (Figure 4.10) of a pixel i over time, with every pixel in the second order neighbourhood being assigned an equal weight.

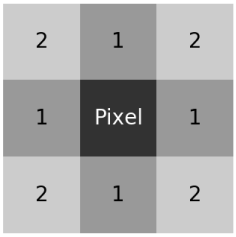


Figure 4.10: The first (1's) and second (1's and 2's) order neighbourhood of a pixel.

These will be compared with optical data, taken from Google Earth, in order to assess the reliability of the maps. Hotspots in these change point density heatmaps should be easily explainable. This sixth evaluation part, combined with the five micro evaluation tests, make up the macro evaluation.

4.4.2. Quantile normalization

The F-test is known to be extremely sensitive to non-normality (Markowski and Markowski, 1990), however the ANOVA F-test uses an F statistic and this is not the same as it being 'the F-test' (Moore and McCabe, 2003). Still, ANOVA works best when the compared segments are described by samples following a Normal distribution. When comparing these segments for inference about the segments means, ANOVA is known to be quite robust to non-normality and unequal variances. This is especially true when the compared segments are composed of a sufficient number of samples (Moore and McCabe, 2003). The implementation of PELT uses a cost function assuming Normally distributed data, with constant standard deviation. It is therefore good to assess the sensitivity of PELT to non-normality, as well. Moore and McCabe (2003) describes some possible steps to deal with severe non-normality in data, in order to make the inference of the group means more reliable; one can try to remove *outliers*, *transform* the data or use some other statistical test, better suited to the distribution that best describes the overall pattern.

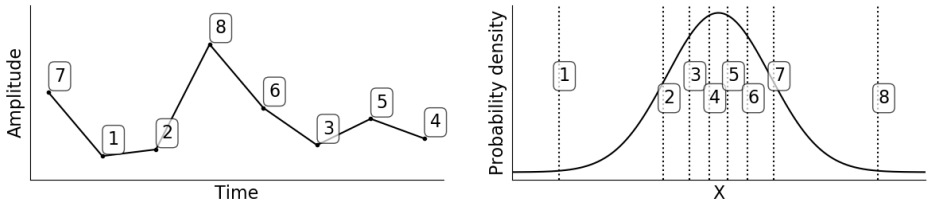
Section 4.1.1, Figure 4.3, showed the that the transformation from a linear scale to a decibel scale yields a near-Normal distribution from a heavily right-tailed distribution. It has to be noted that the observed near-Normal distribution for the Houston stack is for the entire data stack and doesn't necessarily entail normality for the temporal trajectory for a pixel i. To assess the reliability of ANOVA- and PELT-based change detection on the logarithmic transformed data sets, a second transformation is suggested based on *quantile normalization*.

Quantile normalization is usually used to get equal distributions of one or more sequences, all of the same length, by writing these to a reference distribution, creating identical statistical properties (e.g. Bolstad et al., 2003). Because ANOVA is used to compare inter-sequence segments, rather than different realizations of the same observation sequence (with equal length); each time series as a whole is transformed to a Normal distribution via quantile normalization. This ensures that there is at least normality in the entire population of samples. The quantile normalization is performed by

- assigning a rank to all the data points, based on their amplitude values (Figure 4.11a);
- convert each rank to a quantile representation of the ranking, i.e. the first ranked element is assigned the highest quantile value and so forth;
- pull values from a Normal distribution, based on the assigned quantile values, as a replacement representation of that particular data point (Figure 4.11b); and
- normalize these values between the minimum and maximum of the original data to get the quantile normalized data y_q, by computing

$$y_{\mathrm{q}} = (\max(y) - \min(y)) \frac{y - \min(y)}{\max(y) - \min(y)} + \min(y). \tag{4.16}$$

This will yield a time series following a Normal distribution and it still being representative of the original data.



(a) Signal with ranked values.

(b) Quantiles with corresponding ranks.

Figure 4.11: The quantile normalization procedure illustrated. Amplitude values are ranked and assigned a probability quantile. This quantile representation is used to pull a value from a Normal distribution, yielding a normalized time series. The difference in segmentation of the original and quantile normalized time series is a measure of robustness of a CPD algorithm against non-normality.

4.5. Summary

Chapter 4 looks at the data distribution of temporal SAR signals for different land covers in linear and decibel notation of the sigma nought calibrated values. Based on the data distribution and the CPD survey conducted in Chapter 3 four CPD algorithms are selected. Furthermore, the implementation of these algorithms and the methodology of the absolute radiometric calibration for the Rotterdam and Houston stack is discussed. The chapter as a whole highlights the various processing steps, how they are implemented and how the change detection results are assessed in this research.

Simulated and real SAR data is used to assess the accuracy of the CPD algorithms. The simulated data follows the product model description, with Normal or Rayleigh distributed texture and multiplicate normalized Rayleigh distributed speckle; resulting in the Gaussian and Rayleigh stack. The SAR test data sets in Houston (S1) and Rotterdam (TSX) are calibrated in sigma nought. When transformed to a decibel scale, the data follows a near-Normal distribution.

Differencing, CUSUM, ANOVA and PELT are selected based on the data statistics and to give a good balance in computational efficiency, complexity in methodology and level of sophistication in which the temporal evolution is used. Respectively, the implementations of the Differencing, CUSUM and PELT are in a best-case scenario $\mathcal{O}(n)$ and $\mathcal{O}(n \log n)$ for ANOVA. Thresholds are selected based on visual time series segmentation, literature guidelines and documented backscatter intensities of different objects.

The evaluation is split in two parts, the micro evaluation and macro evaluation. The micro evaluation focusses on certain aspects of the algorithm's results, by assessing the sensitivity to different change signal behaviour, speckle and the different resolutions found in Sentinel-1 and TerraSAR-X. The macro evaluation, assessing the overall CPD solution, is done by computing the temporal and spatiotemporal density change maps and comparing these with optical data from Google Earth.

In the next chapter the Differencing, CUSUM, ANOVA and PELT algorithms are tested with the simulated (Gaussian and Rayleigh) and real SAR (Houston and Rotterdam) data. Simulated data with only one specific temporal change behaviour, i.e. a Dirac, Heaviside or Rectangular-like function, is tested with the various algorithms as well. The accuracy, precision and ROC curve quality metrics are computed to get a better understanding of particular strengths and weaknesses of each algorithm with relation to certain change behaviour. For the Houston stack individual time series' segmentation is analysed. Moreover, for both the Houston and Rotterdam stack the temporal and spatially weighted change density maps are computed and analysed. Lastly, a sensitivity analysis for ANOVA and PELT, and an assessment of the change detection performance of all algorithms is conducted.

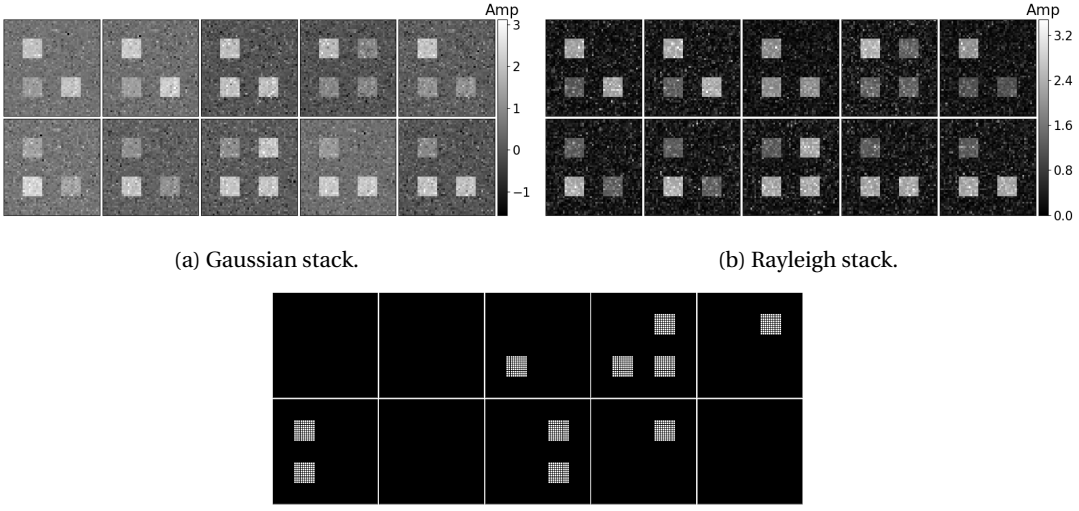
5
Results

Chapter 5 describes the results from the application of Differencing, CUSUM, ANOVA and PELT on simulated and real SAR time series. Moreover, the results from the various outlined sensitivity tests are shown and discussed.

5.1. Application on simulated data

5.1.1. Simulated stacks

The simulated stacks, introduced in Section 4.1.1, Figures 4.1a and 4.1b, consist of 4 square regions of change, that each exhibit a different temporal signal. Figure 5.1 shows the superimposed changes on both the Gaussian stack (Figure 5.1a) and the Rayleigh stack (Figure 5.1b) as a binary change mask (Figure 5.1c).



(a) Gaussian stack.

(b) Rayleigh stack.

(c) Truth map.

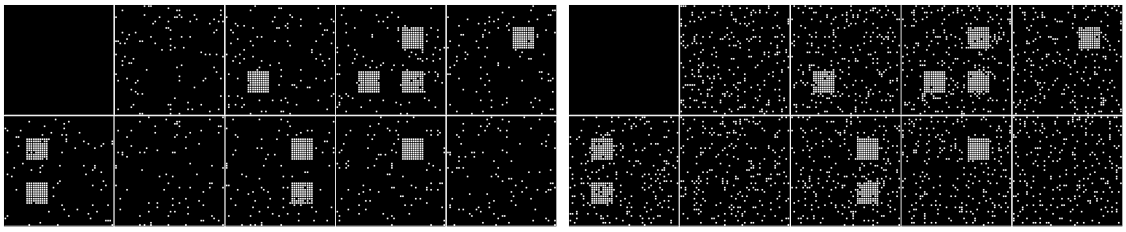
Figure 5.1: The Gaussian stack, Rayleigh stack and a binary truth map showing the true change points. For Figures 5.1a and 5.1b a total of 10 large squares can be seen, each representing a SAR image of 50 pixels by 50 lines at some point in time. On each simulated acquisition up to 4 smaller 10 by 10 squares can be seen. These are superimposed change regions that are subject to significant changes over time. Figure 5.1c shows the binary change maps, where a white dot indicates a different state than the previous image.

For both simulations the following change points were imposed on the images, each in an area of 10 by 10 pixels:

- Top left; a negative Heaviside response half-way through the time series.
- Top right; two positive Dirac responses of different magnitude, with the first being twice as weak as the second.

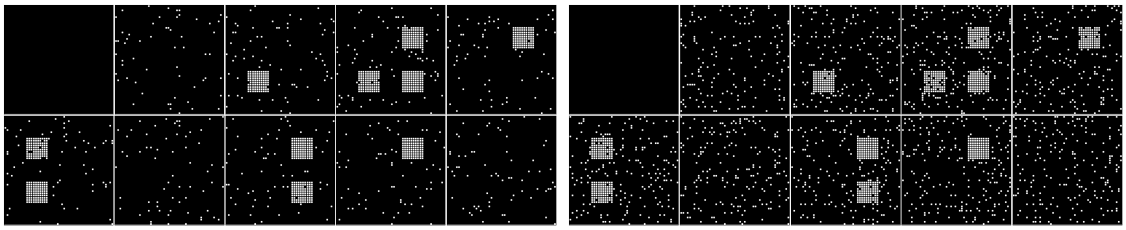
- Bottom left; a combination of a Dirac pulse and a Heaviside, both of equal magnitude.
- Bottom right; two Heaviside responses of equal and opposite magnitude, describing a Rectangular function.

The stacks are analysed for changes with the Differencing, CUSUM, ANOVA and PELT CPD methods. The results of which can be seen in Figures 5.2-5.5. These figures show the binary change masks when the four algorithms are applied on the Gaussian and Rayleigh stack. White points indicate that a change point has been detected. The corresponding confusion matrices of the simulation runs are provided in Table 5.1. The change regions have a known offset, making Differencing and CUSUM in particular perform much better than with real data. Moreover, these offsets are arbitrary unitless values, which should be interpreted as the magnitude of change of consecutive data points. For example, a time series described by $y_{1:3,8:10} = 2$ and $y_{4:7} = 1$ resembles a temporal change signature between data points 4 and 7 that is twice as small as the 'normal' state of the system.



(a) Gaussian. (b) Rayleigh.

Figure 5.2: The Differencing algorithm applied on the Gaussian stack and Rayleigh stack, with a threshold of 0.5. The results show a high TP rate, because the offsets are known. This allows for perfect tuning of the algorithm. However, many FPs are also observed, especially for the Rayleigh stack.

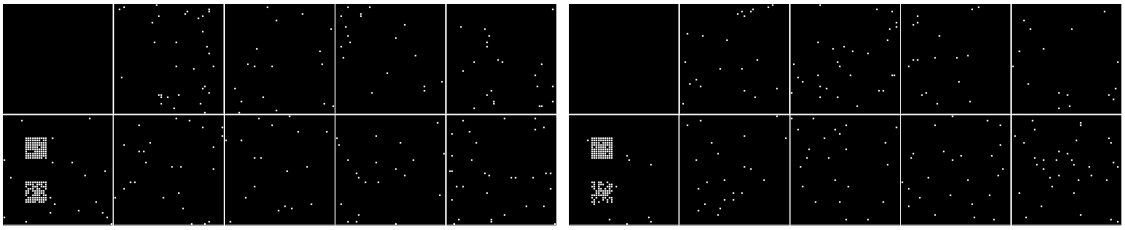


(a) Gaussian. (b) Rayleigh.

Figure 5.3: The CUSUM algorithm applied on the Gaussian stack and Rayleigh stack, with a threshold of 0.5 and a drift of 0.1. The results show a similar result when compared to Differencing. Most of the changes in the change regions are detected, because of the known offsets. More FPs are detected in non-Normally distributed data, than in a data set following a Normal distribution.

First, the CPD results of the Differencing and CUSUM algorithms are examined, as these two algorithms show some similarity in approach. Both show a high rate of TPs for both the Gaussian and Rayleigh stack. This high success rate is largely due to knowing the imposed offset of the change regions, resulting in a near-optimally tuned threshold for both. Especially, the Normally distributed stack benefits from this. The Rayleigh stack shows a significantly higher FP rate for both algorithms due to the asymmetrical distribution of the underlying texture. The Differencing and CUSUM algorithm don't show an apparent bias towards better or worse detection of a particular change pattern, i.e. Dirac, Heaviside or Rectangular.

Even with a near-optimal threshold a lot of FPs are detected, caused by the simulated speckle. This indicates a high sensitivity of both algorithms to speckle noise. With visual analysis alone, it is very difficult to see the differences between the Differencing and CUSUM algorithm's solutions. However, the confusion matrices for Differencing and CUSUM in Table 5.1 do show some differences. CUSUM shows a smaller FP rate for the Gaussian stack and a significantly smaller FP rate for the Rayleigh stack. Meanwhile, the TP rate of CUSUM and Differencing are comparable in both instances. This indicates that the drift correction and the partial consideration of the temporal evolution of a pixel in the CUSUM algorithm already provide significant robustness against speckle.

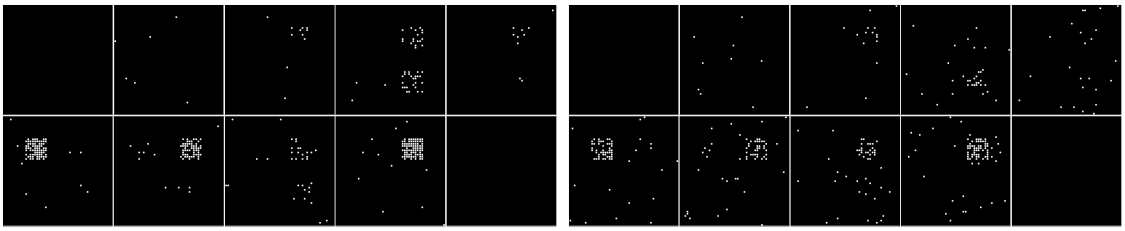


(a) Gaussian. (b) Rayleigh.

Figure 5.4: The ANOVA algorithm applied on the Gaussian stack and Rayleigh stack, with a significance level of 0.01 and a guard of 1. ANOVA only detects two change regions, in the middle of the time series. Both coincide with permanent change. Moreover, the FP rate for the Rayleigh stack is only slightly larger than for the Gaussian stack.

The change map of ANOVA-based change detection shows very few detected changes. Only two change regions are properly detected, both of which related to a single Heaviside response. If the bottom right change region is examined, one might expect an equally well detection rate there, due to the Rectangular temporal signature. Therefore, the near-zero TPs for this square is surprising at first. However, Equation (4.10), with the $K = 2$ groups, tells us that effectively the means of two groups are compared. As the examined temporal behaviour of the bottom right square is described as $y_{4:7} = 1$ and $y_{1:3,8:10} = 2$ in a noise-free state, the reason for the poor detection rate is explained. The most likely change point τ_i with the highest F-value will be detected at y_3. This means the first 3 values are compared to the rest of the time series. As the latter is a combination of both $y_{4:7} = 1$ and $y_{8:10} = 2$ a relatively similar mean to the mean of $y_{1:3}$ is observed. For this particular case a $K = 3$ run would yield a far better result as it would compare the means of the three regions at once, returning a much higher F-value.

A similar reasoning holds for detecting Dirac pulses with ANOVA, e.g. the top right change region. When a single data point in a time series is analysed with ANOVA it has a within-group variability of 0, as the actual value of the measurement is compared to the mean of the measurement. Assuming that the rest of the series has a reasonable within-group variability, and the fact that the total within-group variability (MS_{wg}) is the denominator in Equation (4.10), a relatively large F-value will be returned. However, with the $K = 2$ groups approach, it is not unlikely that a Dirac pulse is part of a larger cluster of values. Lastly, what these results also show is how robust ANOVA actually is to non-normality. Both the number of TPs and the FPs for the Gaussian stack and Rayleigh stack with ANOVA are comparable.



(a) Gaussian. (b) Rayleigh.

Figure 5.5: The PELT algorithm applied on the Gaussian stack and Rayleigh stack, with a constant variance of 0.1. The results of the PELT algorithm with the Gaussian stack show a high success rate and minimal FPs. However, when applied on the Rayleigh stack, the results are noticeably worse.

The CPD with PELT shows vastly different results. Firstly, it should be noted that for both simulation runs a cost function for Normal distributed data was used, as explained in Section 4.2.5. The (noise-free) Dirac pulses described as $y_4 = 1$ and $y_8 = 2$ for the top right square show quite a significant dependency on the magnitude of the signal response for PELT. Another interesting observation is the small amount of TPs in the bottom right change region. For the first 7 data points the signals from the top left and bottom right square behave the same. This lack of segmentation of the bottom right change region is due to the recursive nature of the PELT algorithm. Up until y_8 a segmentation at y_3 minimizes the penalised case. However, when the remaining data points are taken into account, the no-change classification gains the upper hand.

Furthermore, it shows that PELT doesn't seem to favour either Dirac or Heaviside signals. Moreover, PELT shows an apparent bias to segmentation at the first change point in the case of equal and opposite temporal

change behaviour, if the cost is minimized for the single transition instant. Some of these observations will be further examined in a *sensitivity analysis* in Section 5.3. Lastly, please note that the confusion matrices in Table 5.1 and the binary change maps in Figures 5.2-5.5 shouldn't be used to form a conclusion on what algorithm performs better or worse than the others! The simulation serve the purpose of observing particular strengths and weaknesses of the four algorithms, when applied on different data sets and change signals.

Table 5.1: The confusion matrices of the Differencing, CUSUM, ANOVA and PELT algorithms applied on the simulated stacks. A total of 1000 change points are present in the simulated stacks, which is 4.4% of the 22500 data points.

(a) Differencing Gaussian stack.

Truth \ Classified	**Positive**	**Negative**
Positive	983 (4.37%)	17 (0.07%)
Negative	832 (3.70%)	20668 (91.86%)

(b) Differencing Rayleigh stack.

Truth \ Classified	**Positive**	**Negative**
Positive	966 (4.30%)	34 (0.15%)
Negative	2120 (9.42%)	19380 (86.13%)

(c) CUSUM Gaussian stack.

Truth \ Classified	**Positive**	**Negative**
Positive	985 (4.38%)	15 (0.07%)
Negative	576 (2.52%)	20924 (93.03%)

(d) CUSUM Rayleigh stack.

Truth \ Classified	**Positive**	**Negative**
Positive	960 (4.27%)	40 (0.18%)
Negative	1575 (7.00%)	19925 (88.55%)

(e) ANOVA Gaussian stack.

Truth \ Classified	**Positive**	**Negative**
Positive	202 (0.90%)	798 (3.42%)
Negative	861 (3.83%)	20639 (91.85%)

(f) ANOVA Rayleigh stack.

Truth \ Classified	**Positive**	**Negative**
Positive	229 (1.02%)	771 (3.43%)
Negative	905 (4.02%)	20595 (91.53%)

(g) PELT Gaussian stack.

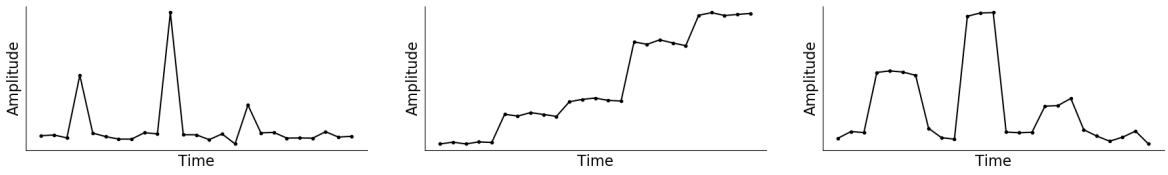
Truth \ Classified	**Positive**	**Negative**
Positive	230 (1.02%)	770 (3.42%)
Negative	118 (0.53%)	21382 (95.03%)

(h) PELT Rayleigh stack.

Truth \ Classified	**Positive**	**Negative**
Positive	154 (0.68%)	846 (3.76%)
Negative	180 (0.80%)	21320 (94.76%)

5.1.2. Individual signal types

A general overview of the algorithms performance in a simulated setting, considering all defined change signals, has been established. Now, each individual signal type is approached individually. The time series of the simulated change behaviour can been seen in Figure 5.6. It can been seen that only one change type is present in each simulation run this time; Dirac change behaviour in Figure 5.6a, Heaviside in 5.6b and temporally Rectangular change behaviour in Figure 5.6c. The Rectangular temporal change behaviour only spans on average three data points each time, which is fairly short. Section 5.3 discusses the 'grey' area between Dirac-like events and temporal change behaviour observed over only a few acquisitions.



(a) Dirac simulation. (b) Heaviside simulation. (c) Rectangular simulation.

Figure 5.6: Examples of the simulated responses linked to change detection in SAR amplitude time series, as used to test the Differencing, CUSUM, ANOVA and PELT algorithms' performance against temporary, permanent and event-like change.

The results of the simulation runs are captured in three quality metrics; the accuracy and precision (Table 5.2), and ROC curves (Figure 5.7), introduced in Section 3.3.1. The thresholds for the ROC curves are: $\tau = [1 \times 10^{-1}, 9.0]$ and $D = 1$ for Differencing; $\tau = [1 \times 10^{-1}), 9.0]$ and $v = [5 \times 10^{-2}, 4.5]$ for CUSUM; $\alpha = [1 \times 10^{-5}, 5 \times 10^{-1}]$ and $D = 1$ for ANOVA; and $\sigma = [1 \times 10^{-1}, 9.0]$ for PELT. The accuracy gives a ratio of correctly identified change and no-change points to all the points, whereas the precision gives a ratio of correctly identified change points to the total change points. With a high number of no-change points, it is fairly

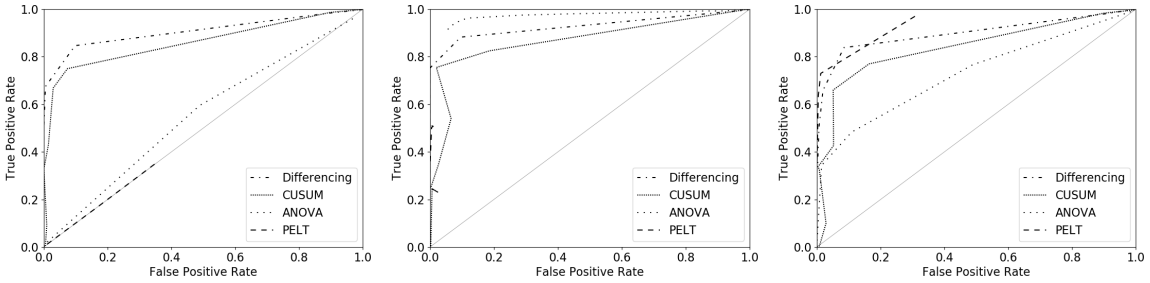
easy to obtain a high accuracy, but much harder to have good precision. When the results are visualized via ROC curves, by plotting the true positive rate (or sensitivity) against false positive rate (or fall-out) at different thresholds, the conclusions following the accuracy and precision are substantiated. The perfect detection rate is at the (0,1) coordinate, where the true positive rate is 100%, hence a good segmentation will reside in the top left corner of the ROC curve. A (near) horizontal curve means that an increase in threshold will increase the false positive rate rapidly for little to no additional correctly detected change points.

The accuracy and precision of Differencing for the different signal types is fairly consistent across the board, implying no preference of signal type. The same holds for CUSUM, where no significant difference between accuracy and precision for different signal types is observed. ANOVA, on the other hand, does not distinguish between no-change and Dirac change behaviour for the simulated signals, as seen by the 0 precision. It is much more reliable in detecting a Rectangular change signal. The relatively low accuracy is likely due to the limited amount of offset points for both temporary changes, namely 3. ANOVA allows for near-perfect detection of Heaviside change behaviour in this particular simulation setting. PELT shows a better detection of Dirac behaviour when compared to ANOVA, at the expense of worse detection of permanent change modelled by the Heaviside signal. The detection rate for Rectangular signal types are comparable with a perfect detection rate at the expense of some FPs for both.

Table 5.2: Accuracy and precision of the algorithms applied on simulated Dirac, Heaviside and Rectangular change behaviour, as shown in Figure 5.6. The results have to be interpreted per algorithm, rather than compared between algorithms, as threshold selection and simulation setup are determining factors in the success rate of an algorithm. Differencing and CUSUM don't show a preference of temporal change behaviour, whereas ANOVA and PELT do.

	Differencing		CUSUM		ANOVA		PELT	
	Accuracy	Precision	Accuracy	Precision	Accuracy	Precision	Accuracy	Precision
Dirac	0.96	1	0.98	1	0.75	0	0.75	0.50
Heaviside	0.98	1	0.98	1	0.99	0.99	0.91	1
Rectangular	0.96	1	0.96	0.99	0.89	1	0.93	1

The ROC curves show a huge dependency of Differencing and CUSUM on threshold selection for all signal types, as a too low of a threshold will lower the true positive rate really quickly and a larger threshold will soon introduce a much higher false positive rate at only marginally better true positive rates. For the Heaviside response ANOVA, even at a very low level of significance of 10^{-5}, the true positive rate is still around 90%. For Dirac-like change behaviour, the ANOVA and PELT ROC curves nearly follow the line of equality, which coincides with a very poor successful detection rate. PELT does show the best ratio of true positives to false positives for Rectangular responses, with a relatively low false positive rate at a nearly perfect detection of the superimposed change points.



(a) Dirac simulation. (b) Heaviside simulation. (c) Rectangular simulation.

Figure 5.7: ROC curves of the Differencing ($\tau = [1 \times 10^{-1}, 9.0]$ and $D = 1$), CUSUM ($\tau = [1 \times 10^{-1}), 9.0]$ and $\nu = [5 \times 10^{-2}, 4.5]$, ANOVA ($\alpha = [1 \times 10^{-5}, 5 \times 10^{-1}]$ and $D = 1$) and PELT ($\sigma = [1 \times 10^{-1}, 9.0]$) algorithms applied on the simulated Dirac (Figure 5.6a), Heaviside (Figure 5.6b) and Rectangular (Figure 5.6c) data. The entire range of used thresholds can be found in Appendix B. A good detection rate will reside in the top left corner, near the (0,1) coordinate [Dotted/Dashed lines: ROC curves, Solid line: line of equality].

5.2. Application on Sentinel-1 and TerraSAR-X SLC data

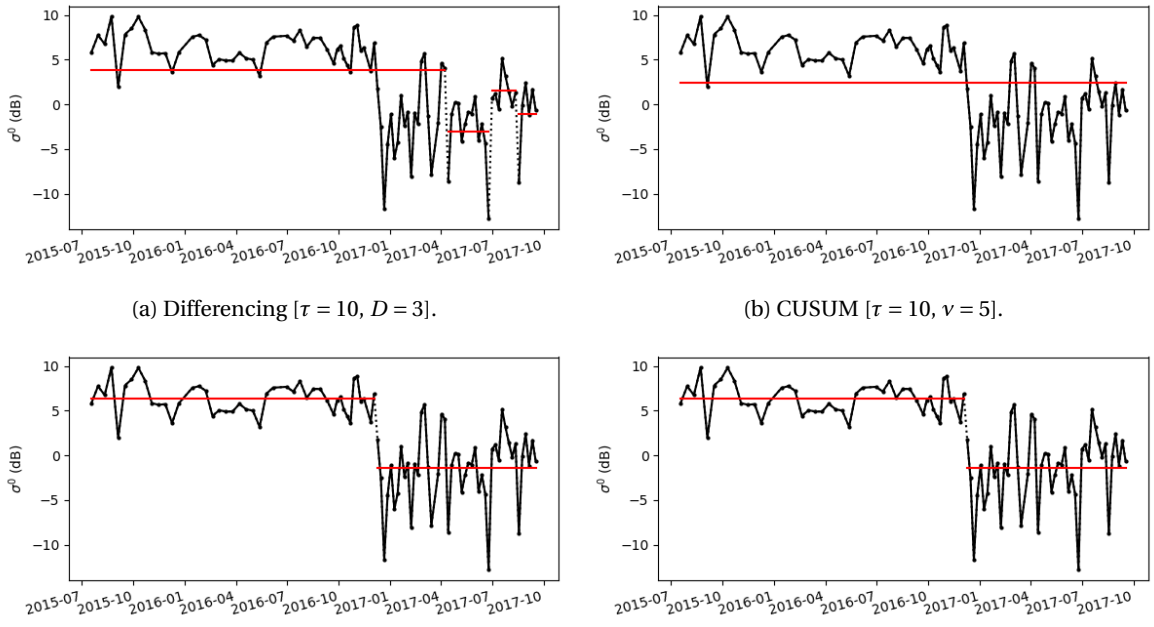
5.2.1. Houston

The Houston test site, sensed with Sentinel-1a and Sentinel-1b, has been introduced in Section 4.1.1 (Figure 4.2a). First, some individual time series are segmented using the four described algorithms, followed by a spatial approach in the form of change density maps. Four different types of time series are analysed for change points using all four algorithms, as shown in Figures 5.8, 5.10, 5.11 and 5.12. For convenience, the time series are referred to as *TS1*, *TS2*, *TS3* and *TS4*, respectively. These time series were selected because of their change behaviour:

- TS1 shows 2 distinct different states of the time series. From a high sigma nought with moderate variance to a much weaker response and a higher variance. This is a good example of permanent change.
- TS2 is a good example for temporary change. The time series shows at least 4 different states and arguably even more.
- TS3 is sensed over water. The positive outlier corresponds with a ship, i.e. an event type signal.
- TS4 shows 3 different states. A strong response with moderate variance, a much weaker Dirac-like response, followed by very stable and strong sigma nought all throughout the rest of the time series. In other words, a combination of different types of events.

Optic satellite images from Google Earth of the observed ground areas corresponding to each time series can be found in Appendix A. All sigma nought time series are on a decibel scale and have not been filtered or smoothened in any way. Furthermore, it is important to note that the same (threshold) parameters have been used for each algorithm applied on all four time series. These are specified in the caption of each figure.

TS1 shows the most likely breakpoint at December, 2016. Differencing shows the worst segmentation of the time series. The higher variance after December, 2016 causes three FPs. The CUSUM algorithm finds no change points in the time series, which might be surprising at first. Visual interpretation finds that ANOVA and PELT return the best segmentation for this time series, as both segment the time series at the expected breakpoint.



(a) Differencing [$\tau = 10$, $D = 3$].

(b) CUSUM [$\tau = 10$, $\nu = 5$].

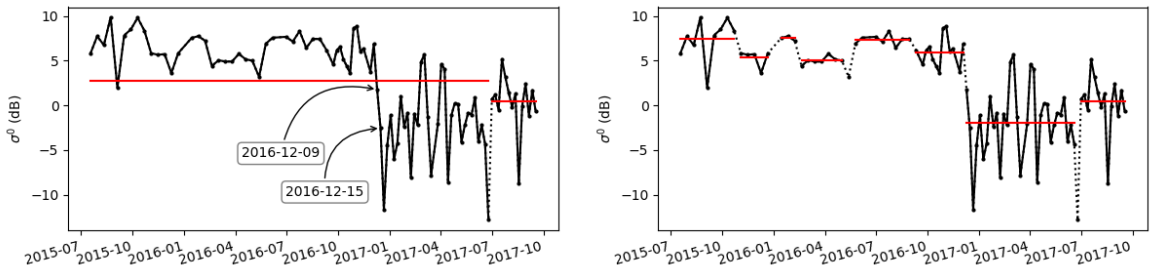
(c) ANOVA [$\alpha = 0.005$, $D = 3$].

(d) PELT [$\sigma = 10$].

Figure 5.8: TS1: a Sentinel-1 sigma nought signal following a Heaviside response with two distinct different mean values and variances [Dotted line: the SAR original time series, Black solid line: segments following CPD, Red solid line: mean of associated segment].

Now, TS1 is further examined with CUSUM and ANOVA to highlight the subtilties in threshold selection and the limitations of simple algebraic CPD methods. The same algorithms are used with different threshold

values. For CUSUM a more generous threshold and smaller drift was selected and ANOVA was performed with a higher significance level. Figure 5.9 shows the results.



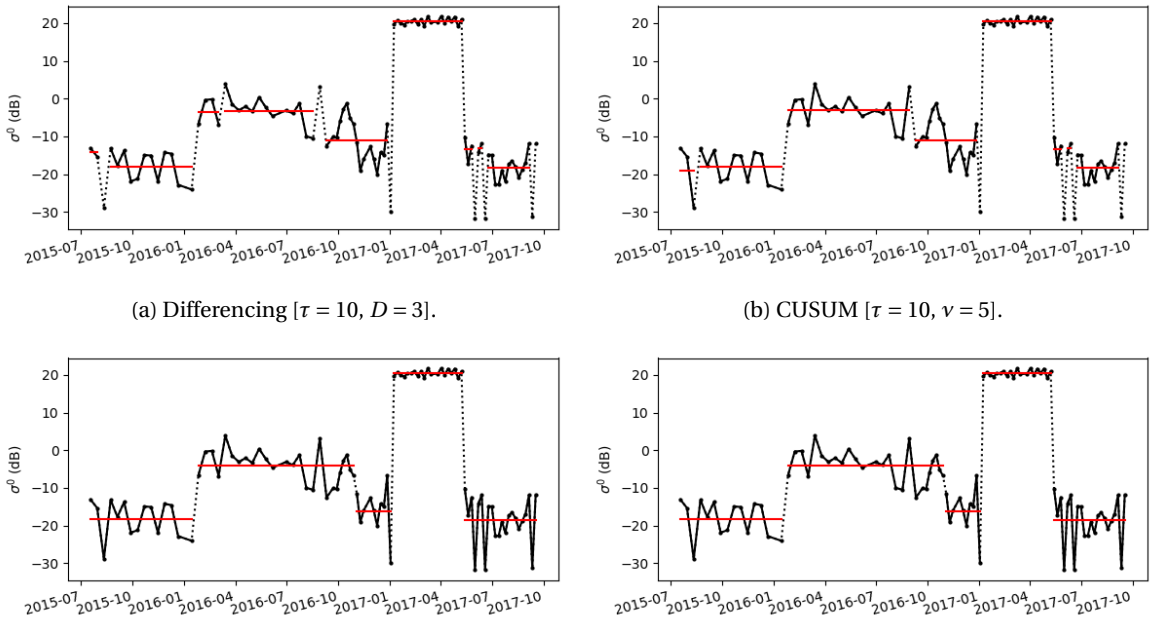
(a) CUSUM [$\tau = 8$, $\nu = 4$].

(b) ANOVA [$\alpha = 0.05$, $D = 3$].

Figure 5.9: Different thresholds for the CUSUM and ANOVA algorithms to segment the time series presented in Figure 5.8 [Dotted line: the SAR original time series, Black solid line: segments following CPD, Red solid line: mean of associated segment].

CUSUM still doesn't show a break point at the (visually) expected time. This can be attributed to the two highlighted intermediate point on 2016-12-09 and 2016-12-15. The algorithm makes two additional calculations that involve the subtraction of the drift twice, resulting in a significantly lower test statistic value than without the two intermediate points. ANOVA, on the other hand, shows a far worse segmentation even with a cut-off F value at the 99.95 quantile. This underlines the importance of proper threshold selection, or tuning, of a CPD algorithm. However, a better solution for one time series, might result in a worse segmentation for another.

TS2 shows more similar results for all algorithms. Between January and June, 2017 the signal shows a consistency in backscatter linked to point scatterers; a relatively high and stable backscatter response. All four algorithms correctly identify this time interval as a separate segment. Again, both ANOVA and PELT arguably perfectly segment the time series. The largest outliers, likely to be speckle related, are identified as change point with Differencing, and with the CUSUM algorithm to a lesser extent. These outliers are not detected with the ANOVA and PELT algorithm.



(a) Differencing [$\tau = 10$, $D = 3$].

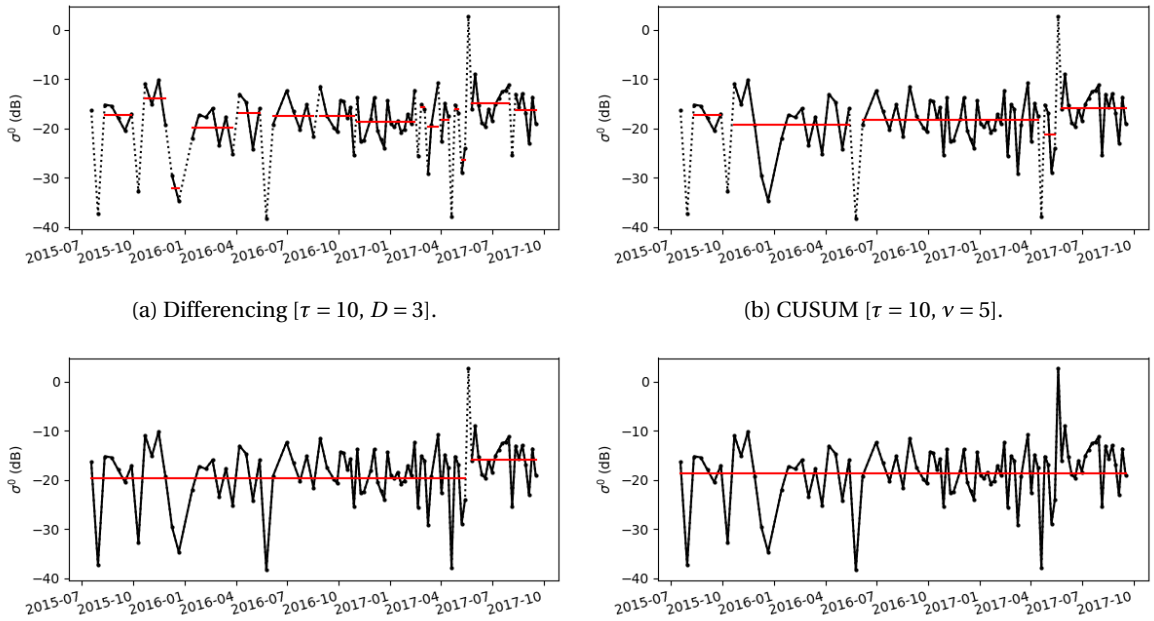
(b) CUSUM [$\tau = 10$, $\nu = 5$].

(c) ANOVA [$\alpha = 0.005$, $D = 3$].

(d) PELT [$\sigma = 10$].

Figure 5.10: TS2: a Sentinel-1 sigma nought signal with fairly constant variance except between January-June 2017, best described as a combination of a Heaviside response and a Rectangular response [Dotted line: the SAR original time series, Black solid line: segments following CPD, Red solid line: mean of associated segment].

The time series TS3 has been recorded over water, in a canal that sees much ship traffic. The Dirac pulse in June, 2017 can be traced back to the reflection of a ship, creating an interesting signal with a very weak response averaging around -20 dB, which is consistent with average VV backscatter measurements for 'river water' as documented in Manjusree et al. (2012). Only ANOVA correctly segments TS3, whereas PELT doesn't show any change points and both Differencing and CUSUM show a severe over-segmentation. TS3 imposes an interesting consideration when using a decibel scale in combination with Differencing, and CUSUM to some lesser extent. On a linear scale the time series sigma nought values are $[1 \times 10^{-1}, 1 \times 10^{-3}]$, with an outlier of around 3. In that case a threshold of 3 would surely segment TS3 properly. This comes at the expense of severe over-segmentation of higher valued segments in a time series (see e.g. Figure 2.13). This is a major drawback in using a static threshold method in combination with the large dynamic range of backscatter values inherent to heterogenous areas sensed with SAR.



(a) Differencing [$\tau = 10$, $D = 3$].

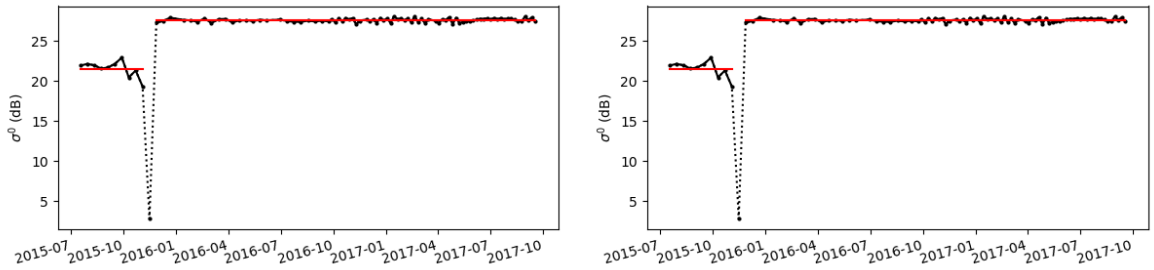
(b) CUSUM [$\tau = 10$, $\nu = 5$].

(c) ANOVA [$\alpha = 0.005$, $D = 3$].

(d) PELT [$\sigma = 10$].

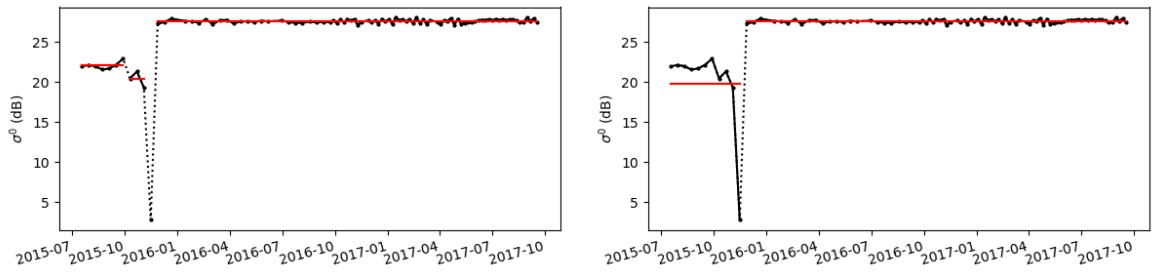
Figure 5.11: TS3: A Sentinel-1 sigma nought signal with a fairly stable mean and a large variance, including at least one significant outlier in mid-July 2017 [Dotted line: the SAR original time series, Black solid line: segments following CPD, Red solid line: mean of associated segment].

Lastly, TS4 is the response from an industrial site and most likely shows the construction, repair or renovation of a factory or industrial building. After completion the resulting object returns a very stable backscatter over time. The Differencing and CUSUM algorithms actually show the most convincing result. This highlights their use for detecting changes in consistent (point) scatterers. PELT shows the worst segmentation, after showing consistently reliable results in the previously discussed time series. It only detects a breakpoint at the right side of the outlier, which has a larger offset magnitude and a more stable response subsequently.



(a) Differencing [$\tau = 10$, $D = 3$].

(b) CUSUM [$\tau = 10$, $\nu = 5$].



(c) ANOVA [$\alpha = 0.005$, $D = 3$].

(d) PELT [$\sigma = 10$].

Figure 5.12: TS4: a Sentinel-1 sigma nought signal of a very stable scatterer, with on significant outlier in late 2015 [Dotted line: the SAR original time series, Black solid line: segments following CPD, Red solid line: mean of associated segment].

In Section 4.4.1, the macro evaluation with temporal and spatially weighted change density maps was already mentioned. Figures 5.13 and 5.14 capture the change behaviour of the non-filtered Houston stack between July 2015 and October 2017. For each pixel i the detected number of changes over time are summed up to form the *temporal change density maps* in Figure 5.13. For a pixel i the value for the *spatially weighted change density map* is created by adding the number of changes detected in the second order neighbourhood of a pixel i with that of the binary change map value of pixel i, i.e. the number of detected changes in a 3x3 window are summed. Here, every pixel in the second order neighbourhood, including pixel i, is assigned the same weight of $\frac{1}{9}$. This procedure is repeated for every binary change map and summed over time. The results can be seen in Figure 5.14. If both change density maps show similar results for some pixel i, it is more likely that some larger object is observed.

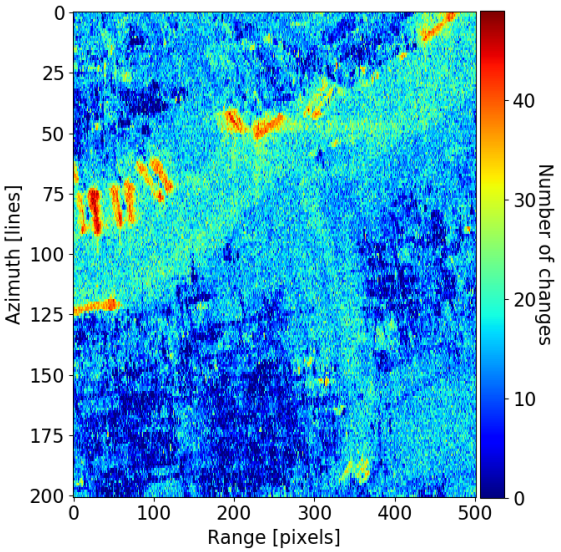
First, the results of Differencing are analysed. The results from the temporal density change map shows a good detection rate of single events, caused by ships passing by. Moreover, many change points are identified in the canal, to the point where the actual shape can be made out. Looking at the time series of the water body on a decibel scale, a constant mean and high variance is noted. This causes a lot of false alarms. At this point we can safely conclude that a simple arithmetic operation with a static threshold isn't suited for reliable change detection using SAR amplitude time series, due to the large dynamic range of the backscatter.

On first sight, CUSUM slightly outperforms Differencing. Looking at the actual number of changes detected, CUSUM actually has a higher maximum. This can be explained by the additional guard used with Differencing, that doesn't allow for segmentation of time series with 3 or less data points in this case. For CUSUM, the drift correction, that acts as a guard against over-segmentation, is not sufficient for the highly variable time series observed in water bodies. Similarly to Differencing, CUSUM performs the best with event type changes. This isn't surprising, given the simplicity of the temporal trajectory usage through the cumulative sum of the sigma nought differences.

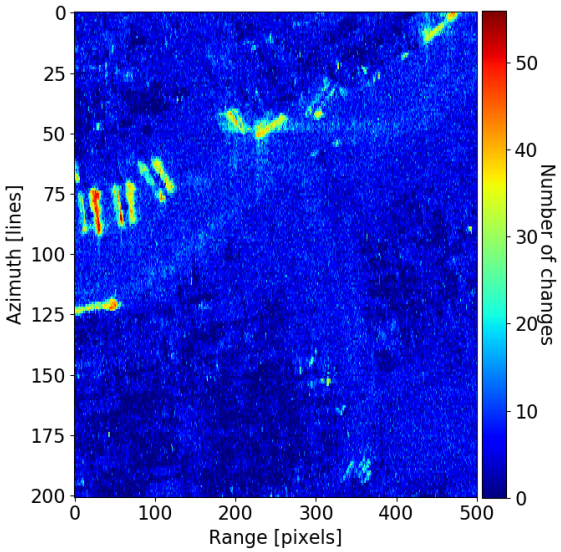
Unlike Differencing and CUSUM, the change maps following ANOVA don't show the outlines of the ships. This is due to the earlier noted poor detection rate of ANOVA of Dirac-like outliers. With a maximum of 10 changes detected over the span of over 2 years, all change hotspots over land and the presence of clusters of the same number of changes observed, visual interpretation the ANOVA-based change maps shows promising segmentation results. Especially when the spatially weighted change map is analysed, some actual buildings can be made out.

PELT shows a somewhat intermediate result between CUSUM and ANOVA, with detection hotspots mapping the ships and a couple of buildings. It was shown in Section 5.1.1 that PELT is less sensitive to Heaviside responses and hence has fewer bright hotspots on land when compared to ANOVA. Moreover, PELT and ANOVA showed similar poor results for Dirac response detection. It might be therefore surprising to see this discrepancy between the detection of the ships between ANOVA and PELT. However, closer inspection of the ship events with PELT finds that the majority of the detected ships were observed over multiple SAR acquisitions (not shown). That is to say the observed ship events with PELT usually follow the Rectangular change behaviour.

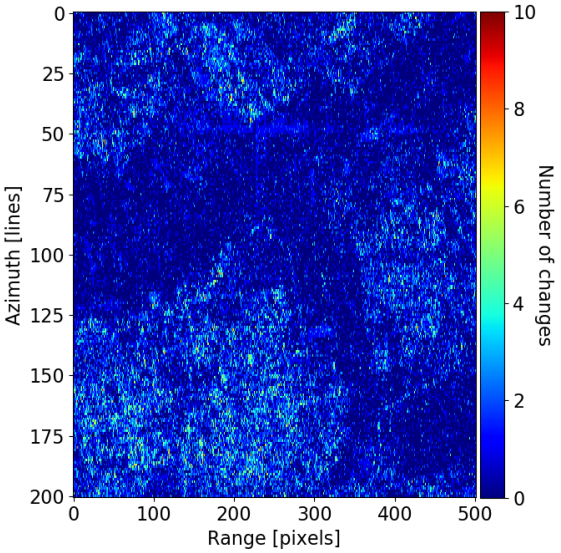
Lastly, note the cross originating from line 50, pixel 240 present in all change maps. This can be traced back to a strong reflector causing severe side lobes. This can be seen best in Figure 5.16b. These type of events, inherent to radar, cause FPs if not properly taken care of because of the large backscatter response compared to the default state.



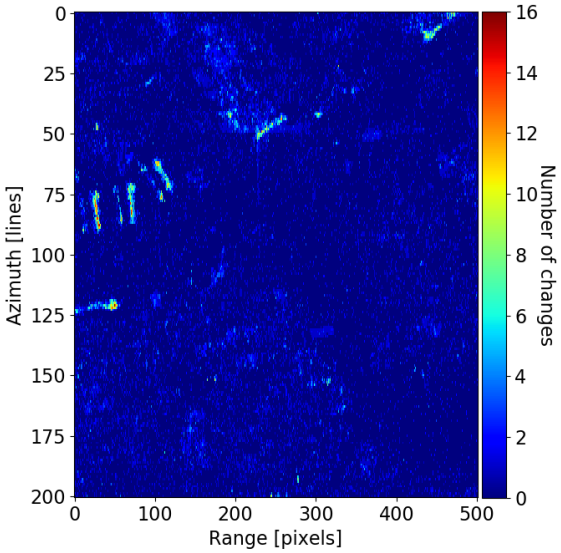
(a) Differencing [$\tau = 10$, $D = 3$].



(b) CUSUM [$\tau = 10$, $\nu = 5$].

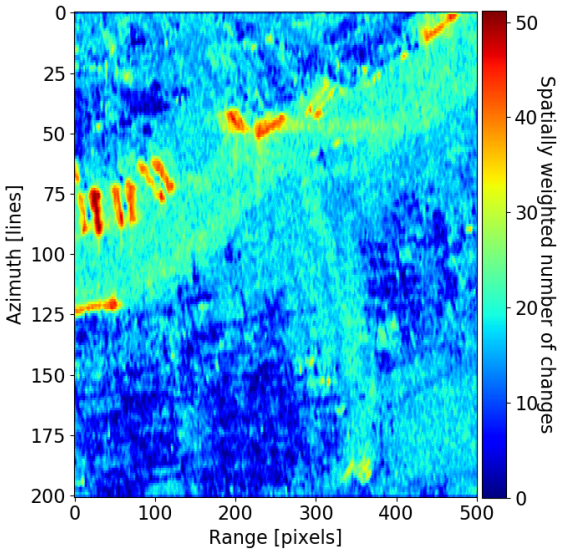


(c) ANOVA [$\alpha = 0.005$, $D = 3$].

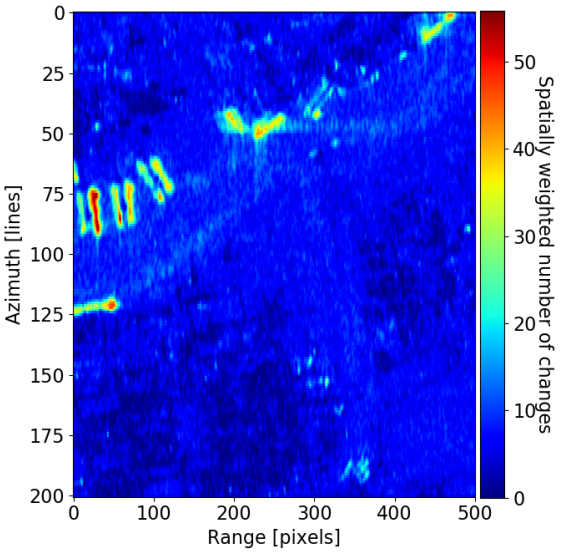


(d) PELT [$\sigma = 10$].

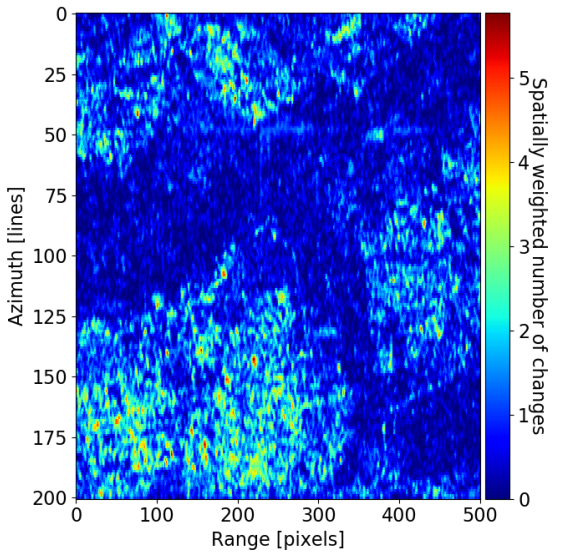
Figure 5.13: The temporal change point density heatmaps of the four algorithms applied on the Houston stack between July 2015 and October 2017. Beware that the colour bars have different scales. Differencing best identifies Dirac-like change events and shows severe over-segmentation especially over water. CUSUM detects similar events, however is much more robust against variability in a time series. ANOVA mostly detects changes over land, coinciding with temporal and permanent change. The solution following PELT detects changes over land and water of Dirac-, Heaviside-, and Rectangular-like change behaviour.



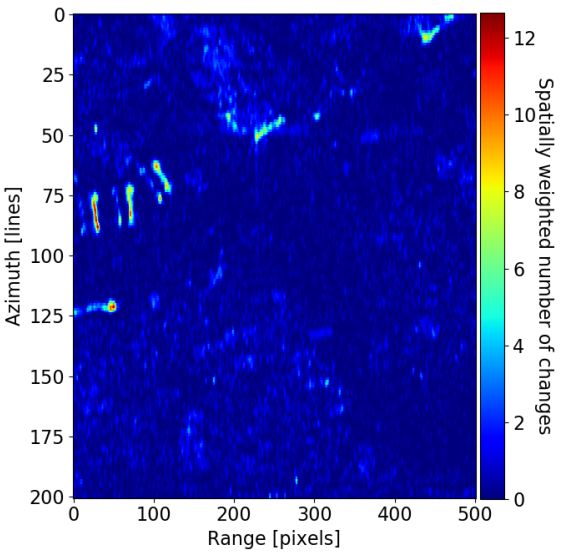
(a) Differencing [$\tau = 10$, $D = 3$].



(b) CUSUM [$\tau = 10$, $\nu = 5$].



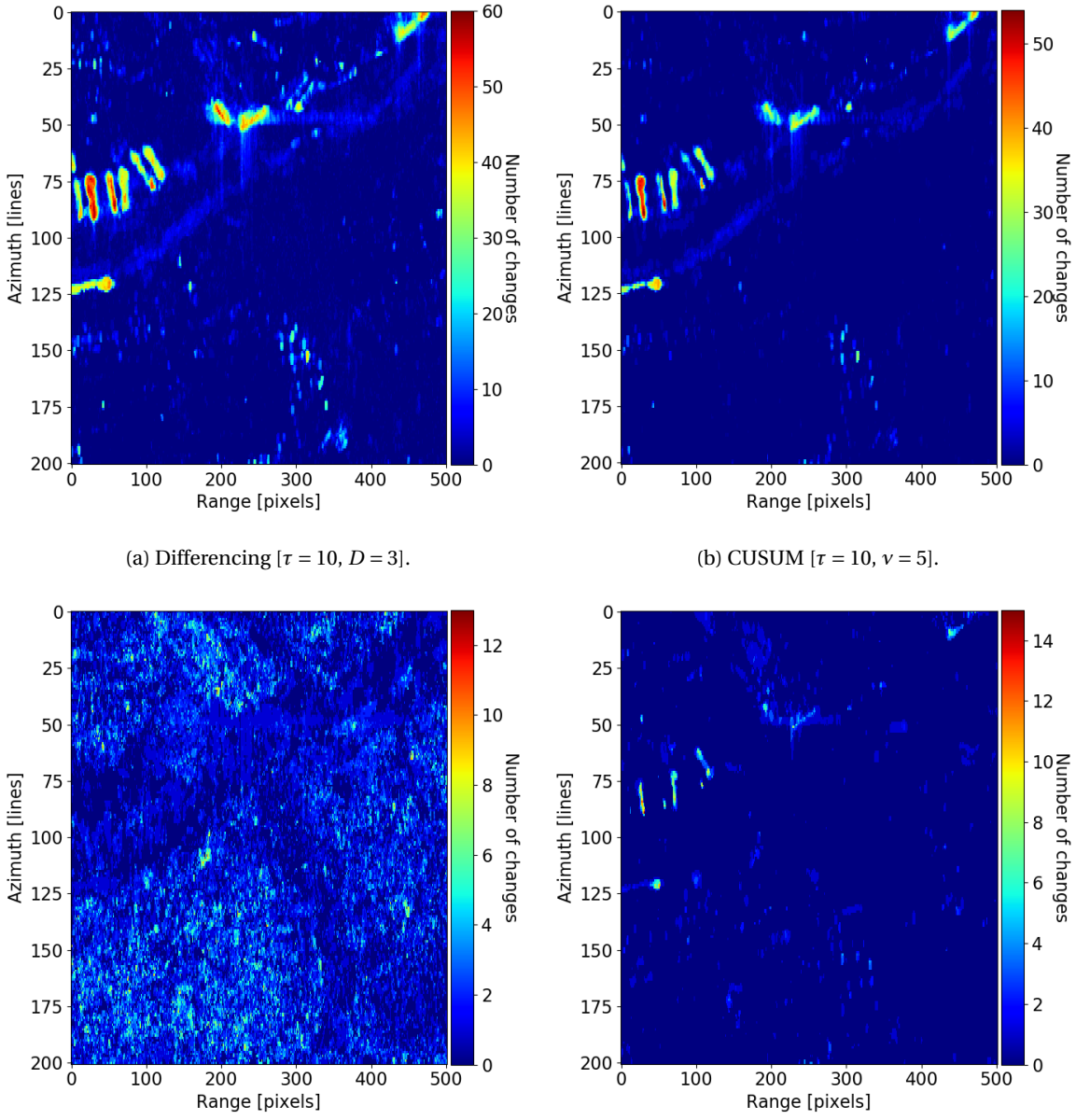
(c) ANOVA [$\alpha = 0.005$, $D = 3$].



(d) PELT [$\sigma = 10$].

Figure 5.14: The spatially weighted temporal change point density heatmaps of the four algorithms applied on the Houston stack between July 2015 and October 2017, considering the second order neighbourhood with equal weights. Beware that the colour bars have different scales. Areas of similar change behaviour in the spatial sense will show as bright spots, or 'hotspots', in the images.

Lastly, the same CPD methodology was applied on a moving average spatially filtered Houston stack, with a window size of 3x3. The results can be found in Figures 5.15 and 5.16. For Differencing and CUSUM a very significant increase in segmentation quality can be observed. Moreover, the PELT-based change map also shows a less noisy result, with more clearly defined groups of change areas. ANOVA seems to benefit the least from the spatial filtering, as it still identifies 1 change point for many of the pixels located over water. However, some very clear on-land clusters of change points can be observed, coinciding with larger buildings or similar objects.



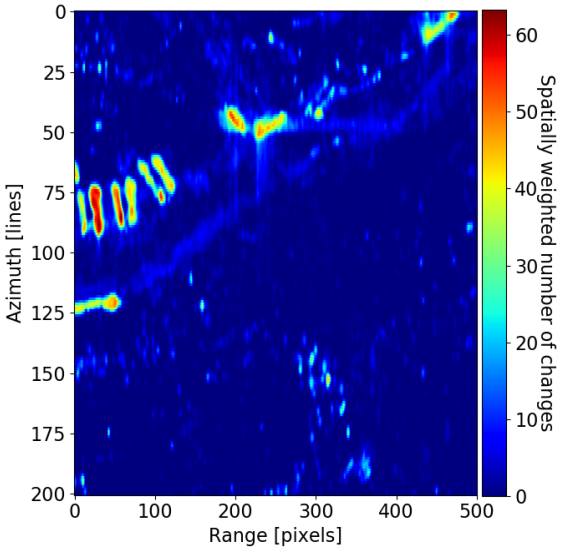
(a) Differencing [$\tau = 10$, $D = 3$].

(b) CUSUM [$\tau = 10$, $\nu = 5$].

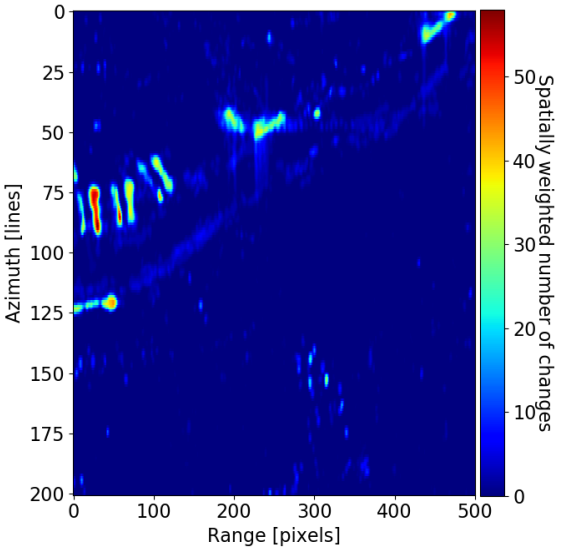
(c) ANOVA [$\alpha = 0.005$, $D = 3$].

(d) PELT [$\sigma = 10$].

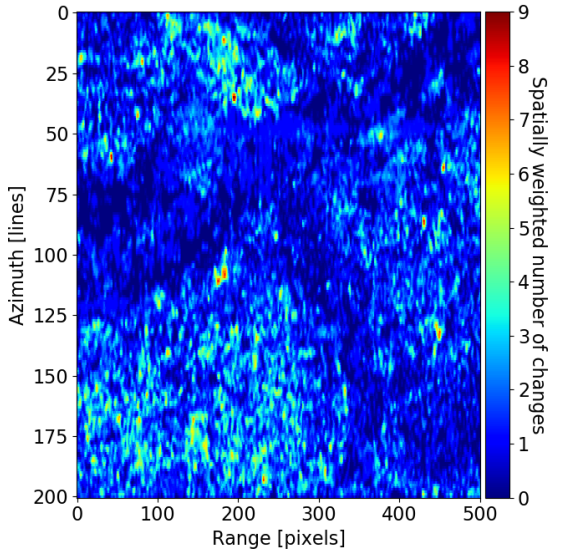
Figure 5.15: The temporal change point density heatmaps of the four algorithms applied on the Houston stack between July 2015 and October 2017, that was filtered with a spatial moving average filter with a window size of 3x3. Beware that the colour bars have different scales. Differencing, CUSUM and PELT benefit the most from this simple filtering procedure.



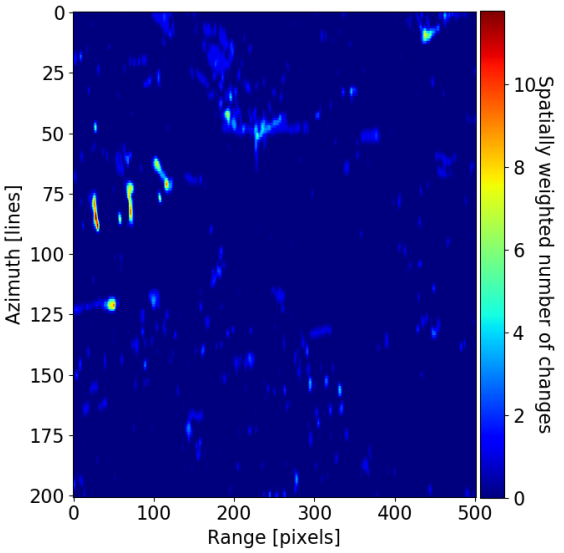
(a) Differencing [$\tau = 10$, $D = 3$].



(b) CUSUM [$\tau = 10$, $\nu = 5$].



(c) ANOVA [$\alpha = 0.005$, $D = 3$].



(d) PELT [$\sigma = 10$].

Figure 5.16: The spatially weighted temporal change point density heatmaps of the four algorithms applied on the Houston stack between July 2015 and October 2017, that was filtered with a spatial moving average filter with a window size of 3x3. Beware that the colour bars have different scales. Areas of similar change behaviour in the spatial sense will show as bright spots, or 'hotspots', in the images.

5.2.2. Rotterdam

For the Rotterdam stack, observed between April 2009 and September 2017, only the spatial solutions in the form of the temporal (Figure 5.18) and spatially weighted temporal (Figure 5.19) change maps are shown. The Rotterdam stack is comparable to the Houston stack in the sense that both areas of interest are located in a very active harbour with a canal intersecting the image. Moreover, both are co-polarized data sets and captured in a descending orbit geometry. Therefore, the same (threshold) parameters are used as in the Houston-stack. The Rotterdam stack does have a superior resolution and consists of a much longer time series, 238 images to Houston's 94.

A horizontal line around the 450 line is clearly visible in the Differencing and CUSUM solutions. Moreover, in the same figures a very pronounced horizontal line around the 470 line, between the 450-500 pixels. These are likely due to pre-processing errors prior to calibration, as these lines are found in roughly 1 out of 5 images in both the sigma nought notation and the uncalibrated SLC amplitude images.

A very large number of changes are detected with Differencing and CUSUM; up to 147 and 115 for a single time series for both CPD methods respectively. Again, this can be attributed to a combination of a static threshold and the large dynamic range of values. This becomes especially troublesome over water areas, where up to some 120 and 50 change points are detected for Differencing and CUSUM over time. Of course, the vast majority of these are falls alarms, caused by e.g. changing wind conditions causing ripples and waves with a higher reflectivity, and speckle.

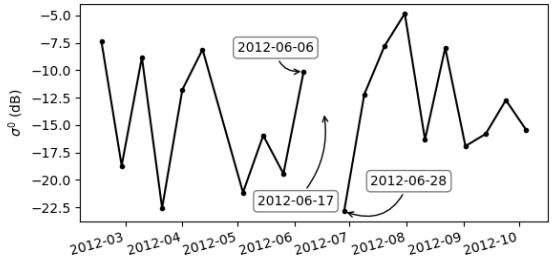
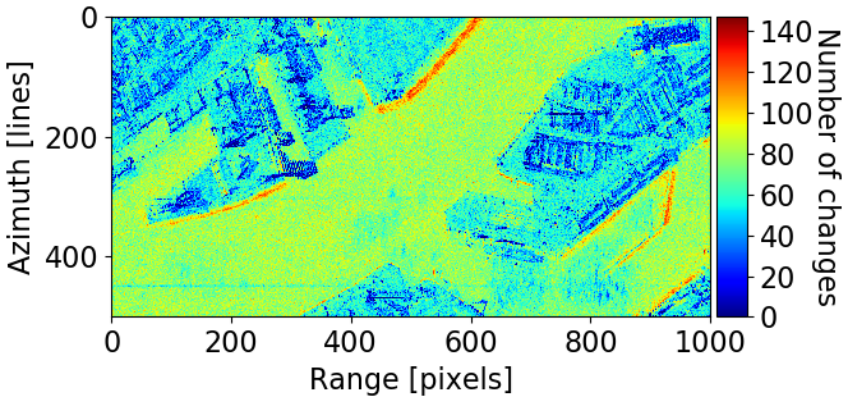


Figure 5.17: Subset of the time series from Rotterdam showing severe oversegmentation with PELT, following a pre-processing error.

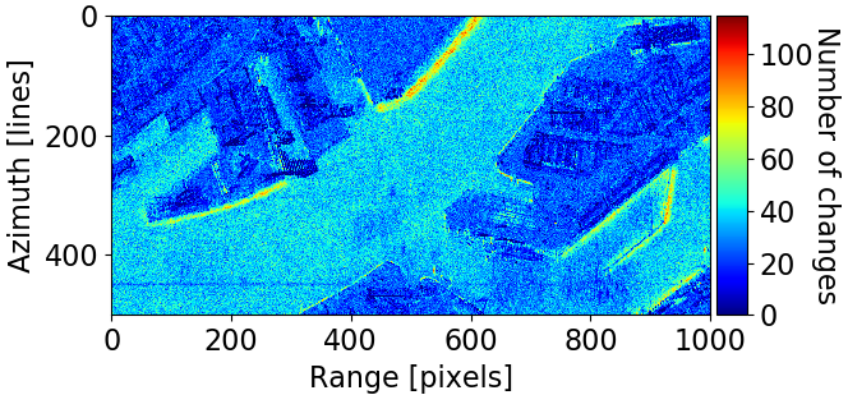
Unlike the reasonable solution for the Houston stack using PELT, a surprisingly large number maximum of changes is detected with PELT in the Rotterdam stack compared to the PELT-based Houston temporal change maps; 16 to 71 actually. Even while taking into account the longer time series from the Rotterdam stack, this is an excessive number of changes. At closer inspection, this unusually high number of changes is a severe outlier, as the second largest number of changes is 17. Therefore, Figure 5.18d would hardly show any distinctive power due to a lack of contrast caused by this outlier. Therefore, the outlier is removed. Figure A.2 in Appendix A shows the original temporal change map resulting from PELT, with the outlier. The spatiotemporal counterpart in Figure 5.19d is much better for identifying regions of change without the need to remove the outlier. The cause of this sever over-segmentation is missing data point (NAN value) in the sigma nought time series, as seen in Figure 5.17. This is the result of an error in the calibration, i.e. a pre-processing error.

Both the ANOVA and PELT change map results show a clear outline of the area, where the general structure of the image can be made out. This is due to many pixels showing a couple of changes along its temporal axis. This might be indicative of an image wide event, or multiple events, taking place. When comparing the 10 acquisition dates that prompted the most break point with ANOVA and PELT, it is found that 5 dates overlap, namely 2009-04-08, 2009-04-19, 2009-11-25, 2009-12-06 and 2010-04-06. The most probable explanation is it being linked to rain events, which are not accounted for in the absolute calibration. Data from Weergegevens.nl (2018) supports this hypothesis. On 2009-04-08 some 5.5 hours of rain over Rotterdam was observed, whereas during the next acquisition, 2009-04-19, no rain event was recorded. The same holds for the consecutive observations on 2009-11-25 and 2009-12-06. Lastly, 2010-04-06 also follows this pattern, as on 2010-03-26, the date of the previous acquisition, there was nearly 1.5 hours of rainfall.

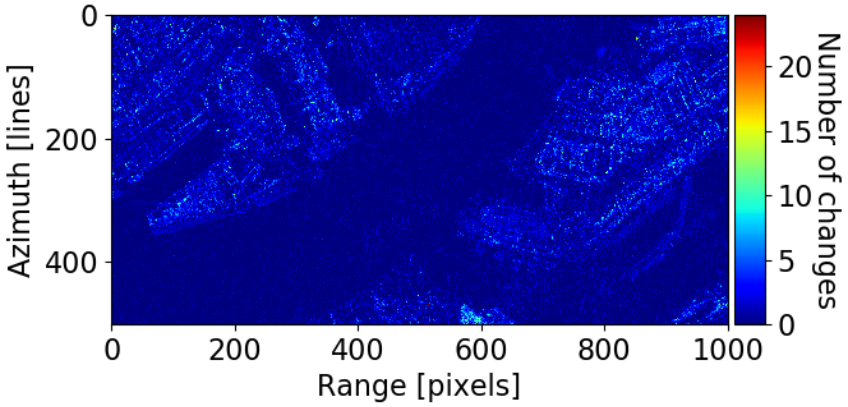
Besides the Rotterdam stack showing more details than the Houston stack, based on visual interpretation there is no clear indication that a significant correlation between the resolution and the quality of the break points solution exists. The Rotterdam stack change maps suffer from the same problems as the Houston change maps, e.g. over-segmentation over water bodies and sensitivity to pre-processing errors. It is very likely that the resolution becomes more important with a spatiotemporal approach, rather than segmenting based only on the temporal trajectory of a pixel, as done by the four different algorithms.



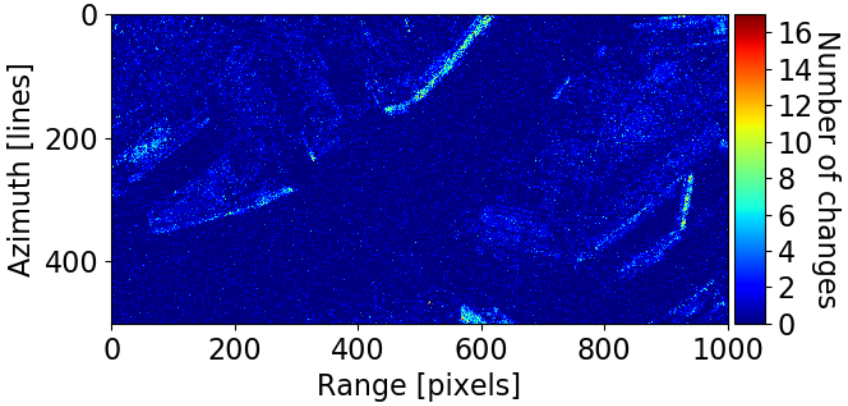
(a) Differencing [$\tau = 10$, $D = 3$].



(b) CUSUM [$\tau = 10$, $\nu = 5$].

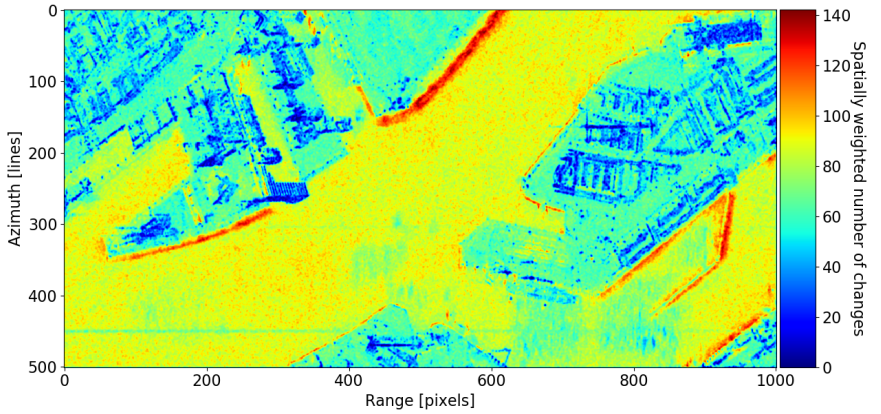


(c) ANOVA [$\alpha = 0.005$, $D = 3$].

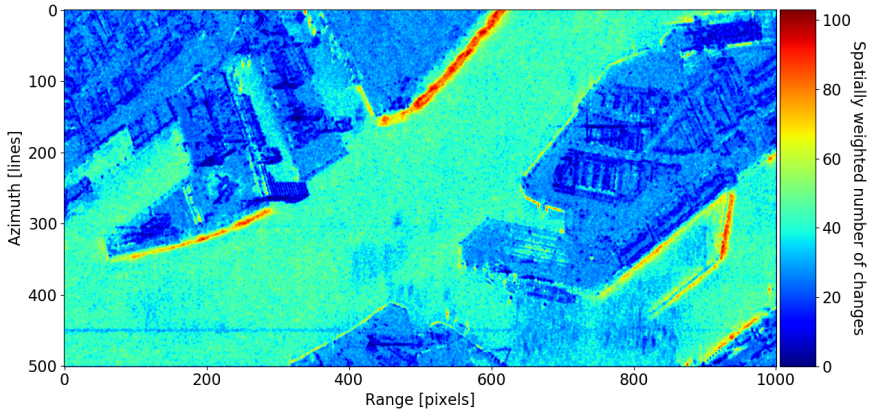


(d) PELT [$\sigma = 10$].

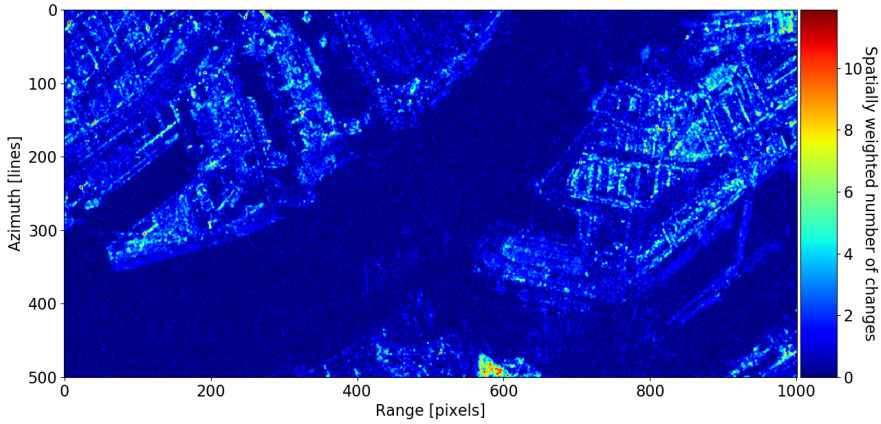
Figure 5.18: The temporal change point density heatmaps of the four algorithms applied on the Rotterdam stack between April 2009 and September 2017. Beware that the colour bars have different scales. Similar results to Figure 5.13 are observed.



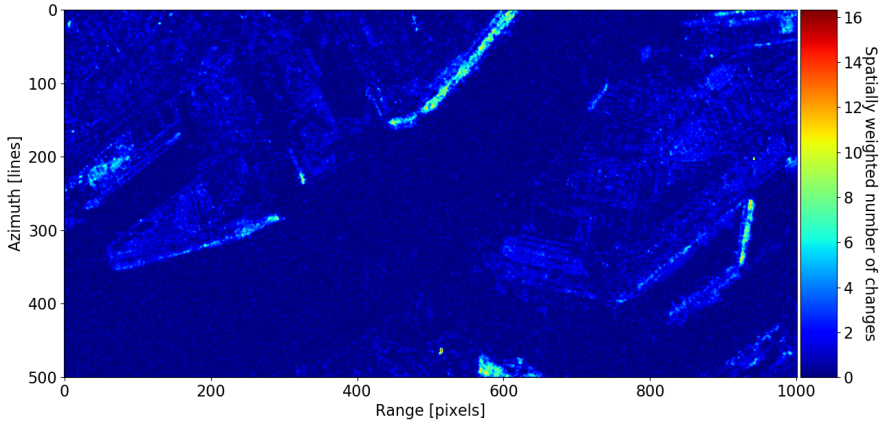
(a) Differencing [$\tau = 10$, $D = 3$].



(b) CUSUM [$\tau = 10$, $\nu = 5$].



(c) ANOVA [$\alpha = 0.005$, $D = 3$].



(d) PELT [$\sigma = 10$].

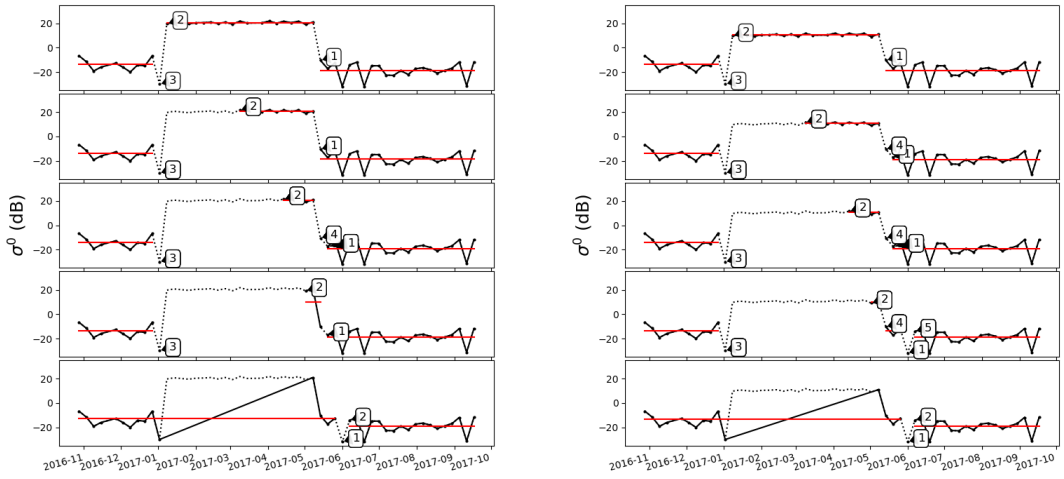
Figure 5.19: The spatially weighted temporal change point density heatmaps of the four algorithms applied on the Rotterdam stack between April 2009 and September 2017, considering the second order neighbourhood with equal weights. Beware that the colour bars have different scales. Areas of similar change behaviour in the spatial sense will show as bright spots, or 'hotspots', in the images.

5.3. Sensitivity analysis

In the sensitivity analysis the observed and expected limitations of the algorithms are further examined. Some algorithms have shown difficulties detecting Dirac pulses, or outliers if you will, for example. With that in mind, the 'grey' area between the Dirac pulse and a Rectangular response is explored. This is done to answer questions such as: "how many measurements have to show a certain offset to be classified as a change point/ region?" and "is that detection dependent on the magnitude of the change?". Moreover, the sensitivity to non-Normally distributed data is assessed. Differencing and CUSUM are excluded for these tests due to the algorithm's limitations. The Differencing algorithm only looks at changes between consecutive measurements and CUSUM includes the temporal component only via the cumulative sum derived from the aforementioned amplitude changes. Non-normality in the data will simply yield more changes detected at the skewed side of the data distribution, i.e. the lower valued backscatter for the Houston and Rotterdam case.

Recall the time series presented in Figure 5.10. A subset from late 2016 to late 2017 is selected, hence creating a time series showing two different states over three time intervals with mean values of $y_{1:11} \approx -16$, $y_{12:31} \approx 20$ and $y_{32:53} \approx -16$ dB, i.e. a Rectangular response or temporary change. The PELT and ANOVA algorithm are now used to search the time series for change points in this subset. Now observation points from the time series are incrementally removed in the second segment, effectively creating a shorter intermediate state with each removed sample. This is done several times, each time searching for change points with the same algorithm parameters. For example, the top subfigure in Figure 5.20 is the original subset, the second subfigure has had 10 observations removed and so forth. Moreover, this experiment is repeated with a superimposed offset in the second segment, by subtracting 10 dB from all observations in this section. The results of this experiment are shown in Figure 5.20 for ANOVA and Figure 5.22 for PELT.

For the ANOVA-based segmentation three change points are identified for the complete subset of the time series (top subfigures in Figure 5.20). ANOVA yields change points at τ_{12} ('2') and τ_{32} ('1'), and an additional change point at τ_{11} ('3'). The second subfigures, having removed 10 out of the 20 observations of the second segment, shows an additional change point at τ_{33} (now '1') for the time series with the superimposed offset. The middle subfigures show the same segmentation. The second to last and last subfigures in Figure 5.20 show poor segmentation of the second segment and clear differences in segmentation between the no-offset experiment (Figure 5.20a) and the superimposed offset experiment (Figure 5.20b) for the second to last subfigure.



(a) ANOVA [$\alpha = 0.05$, $D = 3$].

(b) ANOVA -10 dB [$\alpha = 0.05$, $D = 3$].

Figure 5.20: Sensitivity analysis of ANOVA with a superimposed offset of -10 dB for January-May, 2017 in Subfigure 5.20b. The top figures are the original time series, the second figures had 10 data points removed from the left side of the January-May offset region, the third figures had an additional 5 points removed and so forth. The order of change point detection is shown to be a determining factor for the ANOVA algorithm. Much more than the actual magnitude of the change offset [Dotted line: the SAR original time series, Black solid line: segments following CPD, Red solid line: mean of associated segment, Numbers: order of detection].

At this point it is important to recall the methodology of ANOVA (Section 4.2.4), as one might jump to the conclusion that the difference in results is purely due the difference in magnitude of the offset. ANOVA

uses the highest F-value of the time series to compare to the critical F-value, assuring the best segmentation for the first iteration. However, this doesn't guarantee an optimally segmented time series after multiple iterations. Moreover, ANOVA loses context after segmentation, that is to say each newly created segment is analysed without context of the time series as a whole. Therefore, the order of change point(s) detection plays a big role. For each subsequent CPD run the CPD order is annotated. If the second run of Subfigure 5.20a is compared to that of Subfigure 5.20b, the impact of this segmentation order can be seen. Note point '4' in Subfigure 5.20b; this point is seen as a separate segment in the test run with the superimposed -10 dB. This is due to prior segmentation at point '1', attributing point '4' to a segment with the observation points preceding it that have a much higher backscatter value.

The aforementioned example showing the preferred first segmentation at τ_{33} rather than τ_{32} can easily be explained as a combination of the $K = 2$ approach and the fewer samples that are present in the intermediate interval ($y_{12:31}$). This results in a smaller within-group-variability for the first segment if split at τ_{33}. This also implies that there is a dependency on the location of a Dirac pulse and the ability of ANOVA to detect it. Figure 5.21 is the result of a simulation that assumes a constant value for $y_{1:50}$ and superimposes a Dirac pulse at some starting index i, with $1 \leq i \leq N$. If the algorithm isn't able to detect the offset, the same offset is added to the next point until a change point is found. This is done for different significance levels.

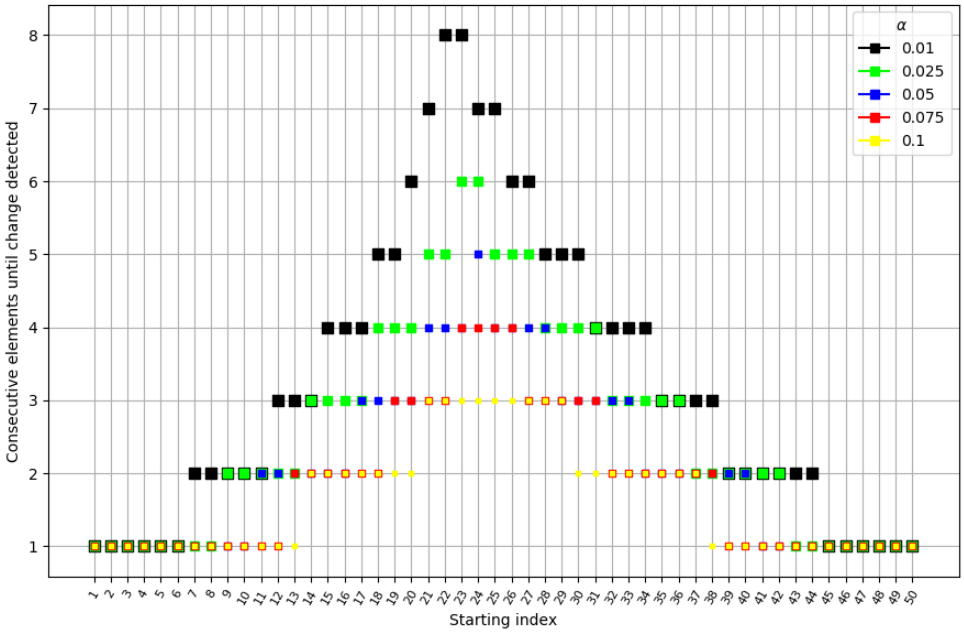


Figure 5.21: Dirac sensitivity test for ANOVA with an offset of 10% of the signal strength, for different levels of significance α. The y-axis shows the consecutive elements, with 1 point being a Dirac pulse and more than 1 point a Rectangular-like response. The plot shows the relation between the starting index of the change and the number of needed offsets for ANOVA to identify the change behaviour. The sensitivity of the implementation of the ANOVA algorithm is highest when the outlier is located at the outsides of the time series.

Figure 5.21 shows that ANOVA's capability to detect a Dirac pulse is much better when the Dirac pulse is located at the edge of a time series. This makes sense as comparing a Dirac pulse against the rest of a time series yields a much higher between-group-variability and smaller within-group-variability as can be seen in the notation for the sum-of-squares (SS) for these two quantities

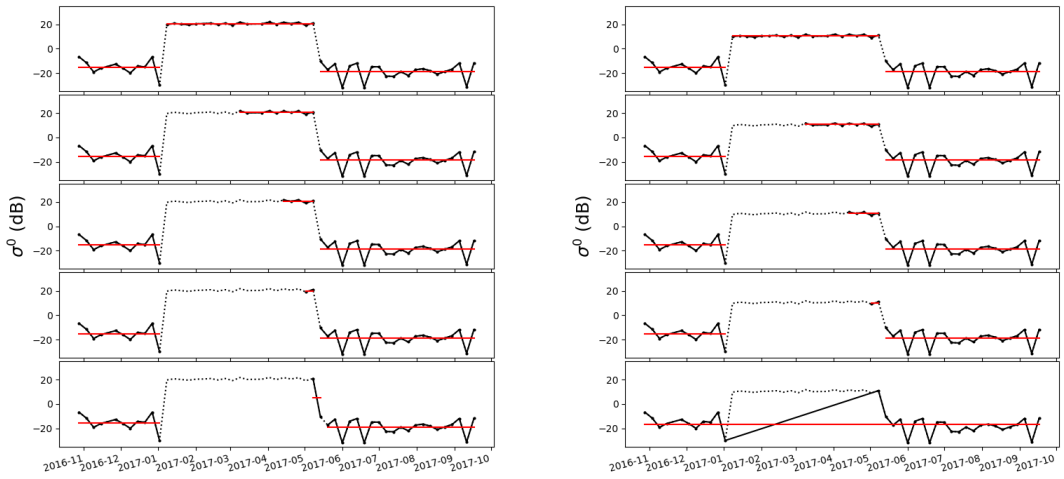
$$SS_{bg} = \sum_{i=1}^{K} \left(N\left(\mu_i - \mu\right)\right)^2, \tag{5.1a}$$

$$SS_{wg} = \sum_{i=1}^{K} \sum_{j=1}^{N} \left(A_{ij} - \mu_i\right)^2. \tag{5.1b}$$

Secondly, the need for a relatively larger increase in significance level to detect Rectangular-like temporal changes in the centre of a time series, compared to the detection of Dirac pulses at the edge of a time series.

Contrary to ANOVA, PELT isn't influenced by the order of the CPD due to its recursive nature and optimality. However, if the last test runs in Figures 5.22a and 5.22b are compared the dependency of PELT on

the magnitude of change can be seen. In other words, the detection power of PELT increases with increasing offset values. This observation is confirmed by repeating the experiment with different offset values. The expected standard deviation and the length of the time series play a big part in this change magnitude dependency. A relatively smaller standard deviation will increase the cost for segmentation quite rapidly, while the penalty term becomes larger with sequence length.



(a) PELT [$\sigma = 10$].

(b) PELT -10 dB [$\sigma = 10$].

Figure 5.22: Sensitivity analysis of PELT with a superimposed offset of -10 dB for January-May 2017 in Subfigure 5.22b. The top figures are the original time series, the second figures had 10 data points removed from the left side of the January-May offset region, the third figures had an additional 5 points removed and so forth. The PELT algorithm has a dependency on the magnitude of the change offset [Dotted line: the SAR original time series, Black solid line: segments following CPD, Red solid line: mean of associated segment].

A similar experiment to Figure 5.21 can be done for PELT, as shown in Figure 5.23.

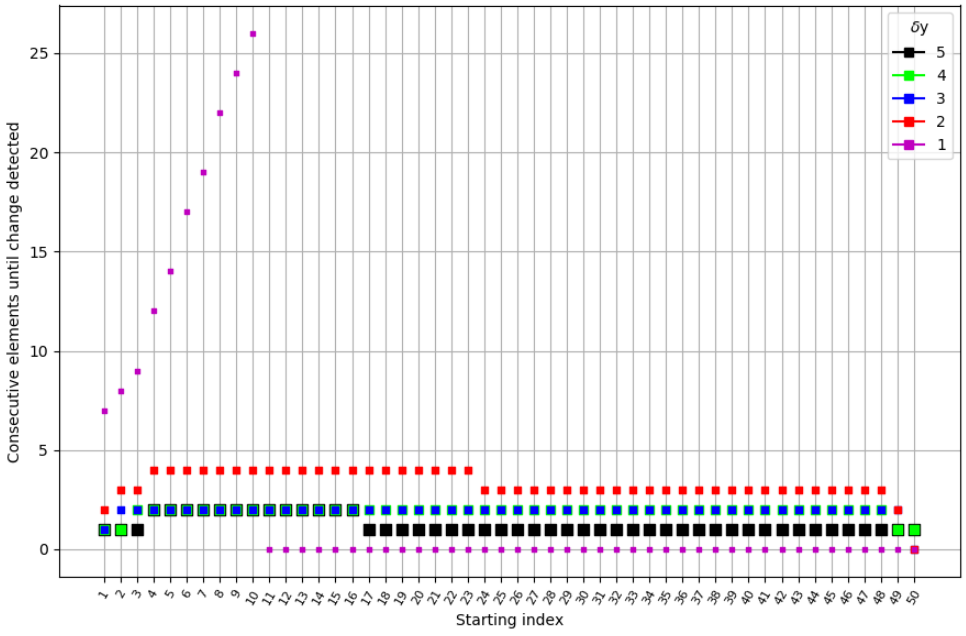
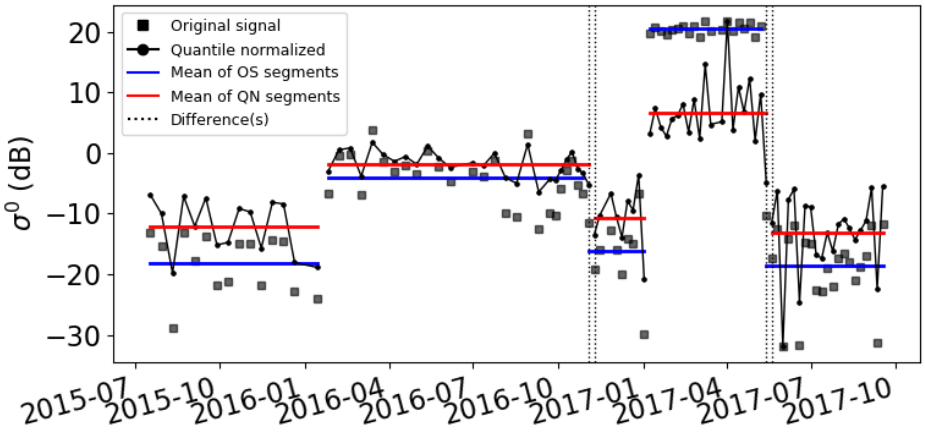


Figure 5.23: Offset magnitude sensitivity test for PELT with varying offset levels and constant algorithm parameters. The y-axis shows the consecutive elements, with 1 point being a Dirac pulse and more than 1 point a Rectangular-like response. The plot shows the relation between the starting index of the change and the number of needed offsets for ANOVA to identify the change behaviour. The smaller the offset, the more consecutive elements needed for PELT to detect a change point. When the expected standard deviation is (nearly) the same as the magnitude of change, a very large dependency on the direction of time series analysis is observed.

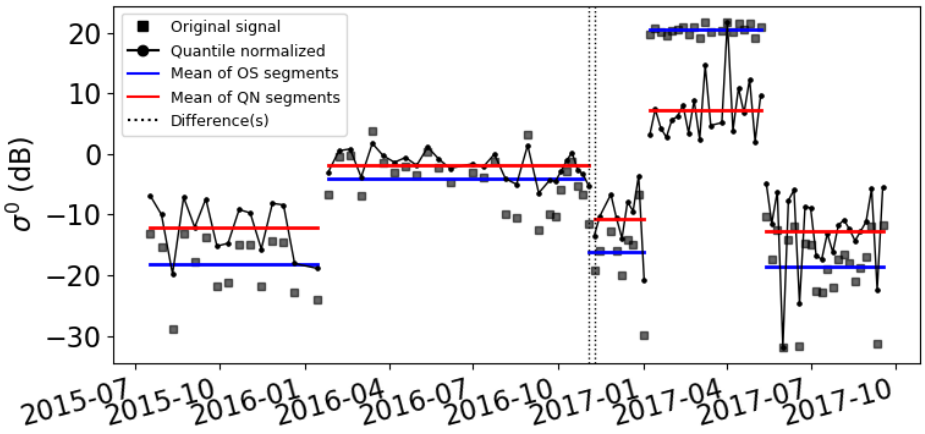
Here, a constant standard deviation σ of 1 is used, with a changing offset δy. The difference between $\delta y = 1$ and $\delta y = 2$ shows exactly how much better PELT performs with a larger offset. When $\sigma = \delta y$, the number of offset samples needed for PELT to detect a change transition increases quite rapidly with starting index up until it plateaus at a maximum number of 26 consecutive samples. At this point no more changes are found. This indicates that PELT is highly sensitive to the direction in which a sequence is analysed for offsets similar to the expected standard deviation, as a direction unbiased solution would yield a symmetrical shape over the starting index.

Change points with an offset close to the expected standard deviation should therefore be interpreted with caution. Moreover, the increasing smaller left side bias of $\delta y \in (2,3,4,5)$ indicates an decreasing dependency on the direction of time series analysis with increasing offset. From Figure 5.23 it is also found that CPD is nearly independent of change point location, given a large enough magnitude of change. Only a slight bias to the edges of the time series is observed.

Lastly, in Section 4.4.2 the methodology of quantile normalization to assess the sensitivity to non-normality of ANOVA and PELT was discussed. Figure 5.24 shows the results of this approach for TS2 from the Houston stack. The results for TS1, TS3 and TS4 can be found in Appendix A. The squares denote the original data points in sigma nought notation and the continuous time series is the quantile normalized time series of these measurements.



(a) ANOVA [$\alpha = 0.005$, $D = 3$].



(b) PELT [$\sigma = 10$].

Figure 5.24: Sensitivity to non-normality of ANOVA and PELT following the quantile normalized time series TS2 from the Houston stack, as shown in Figure 5.10 [Black squares: data points of original SAR time series, Black solid line: the quantile normalized time series, Blue solid line: mean of associated segments original time series, Red solid line: mean of associated segment quantile normalized series, Dotted line: differences between segmentations original and quantile normalized series].

The results of the quantile normalization experiment, combined with the confusion matrices following the simulated data in Section 5.1.1, Table 5.1, indicate that both ANOVA and PELT are relatively robust against non-normality. The segmentation of TS2 and the quantile normalized TS2 is very similar. Moreover, the confusion matrices of the simulated Gaussian stack and Rayleigh stack show very similar percentages of FPs, FNs, TPs and TNs. However, recall the methodology of the quantile normalization and the ANOVA-based change detection. The quantile normalization, as used, assures a Normal distribution in the entire time series; whereas ANOVA compares combinations of $K = 2$ groups within that time series, which ideally should both be Gaussian distributed.

Instead of quantile normalizing the entire time series in one go, one could quantile normalize both of the $K = 2$ compared groups in ANOVA separately for all combinations. In other words, each possible segment has been normalized independent of the other segments. The results of this are shown in Figure 5.25. As can be seen, more segments are identified with ANOVA when both groups follow a Normal distribution. This makes sense, as the within-group variability is smaller, resulting in larger F values.

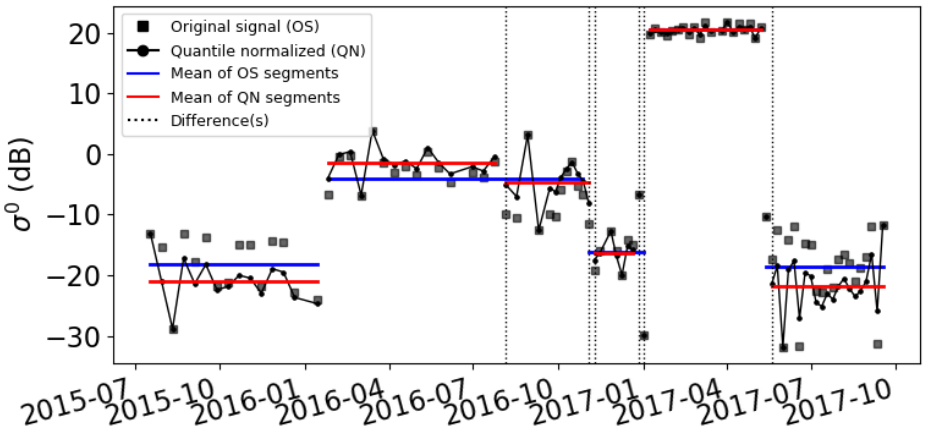


Figure 5.25: Results of ANOVA-based change detection using separate quantile normalization for both groups in each possible segmentation [Black squares: data points of original SAR time series, Black solid line: the quantile normalized segments, Blue solid line: mean of associated segments original time series, Red solid line: mean of associated segment quantile normalized segments, Dotted line: differences between segmentations original and quantile normalized segments].

5.4. Assessment of change detection performance

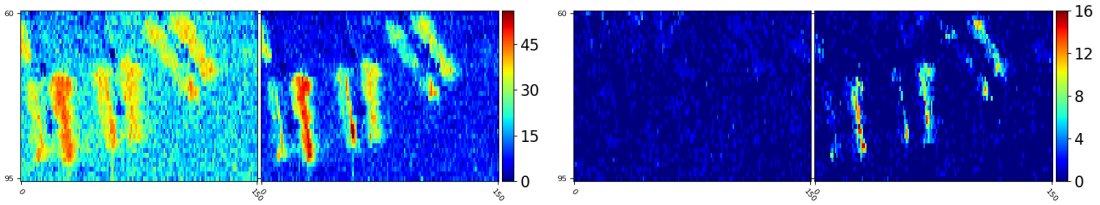
Sections 5.2.1 and 5.2.2 showed temporal and spatially weighted temporal change maps for all methods applied on the Houston and Rotterdam stack. This section is aimed at assessing the reliability of the detected changes, without having actual truth maps to compare them to. Google Earth optical satellite and aerial images are used and compared to the change maps, to this end. Regions that exhibit a high change density ('hotspots' in the images) in both the temporal and spatial domain should be easily explainable. Subsets of the temporal change maps are provided for each discussed change region or specific pixel. In total 5 cases will be evaluated based on hotspots found in the temporal and spatiotemporal change maps, 3 for Houston and 2 for Rotterdam. These turn out to correspond to:

1. the docking and undocking of ships;
2. the destruction of a building;
3. a temporary hangar;
4. a biannual fair; and
5. a highly variable container storage area.

The docking and undocking of ships

The easiest change behaviour to analyse is the docking and undocking of large (tanker) ships in the Houston stack. As can be seen in Figure 5.26, these are observed with the Differencing, CUSUM and PELT methods.

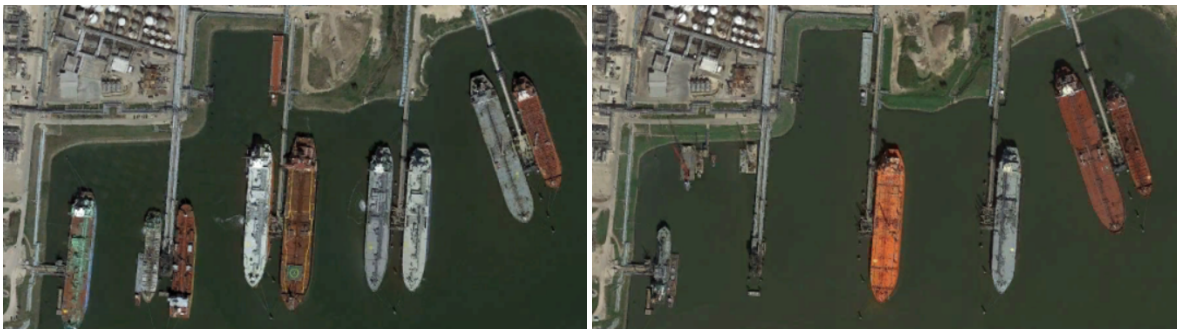
PELT seems to provide the most reliable result, with some 8-16 docking and undocking events in the moments captured in a two-year timespan. This can be visually confirmed by looking through the radiometric images in the Houston stack. The Differencing- and CUSUM-based change maps observe up to 45-50 changes. This is a severe over-segmentation, going by the individual radiometric images, the sparsely available ground truth and the fact that the surrounding water pixels show up to 15 and 30 changes for the CUSUM and Differencing solution, respectively. Figure 5.27 shows two snapshots in time of the ships in the Houston harbour.



(a) Differencing (left) and CUSUM (right) temporal change maps.

(b) ANOVA (left) and PELT (right) temporal change maps.

Figure 5.26: Case 1 - the docking and undocking of ships in Houston shown by temporal change maps. Beware that the colour bars have different scales (subset from Figure 5.13, lines 60-95 and pixels 0-150).



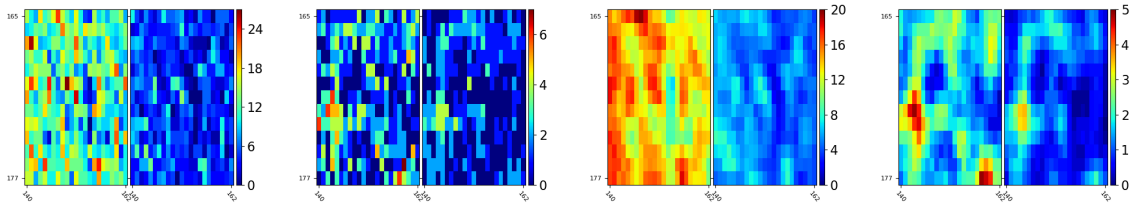
(a) 2016-03-04.

(b) 2017-10-29.

Figure 5.27: Satellite and aerial optical images for case 1.

The destruction of a building and a temporary hangar

Secondly, two cases on the main land of Houston are analysed; indicated in Figure 5.28e. The first case shows the multi-stage destruction of a fairly large building being demolished in the span of a couple of months (Figure 5.32). The second one follows a temporary hangar-like object, subject to much truck traffic in its close vicinity (Figure 5.33). For both test cases an example time series is provided.

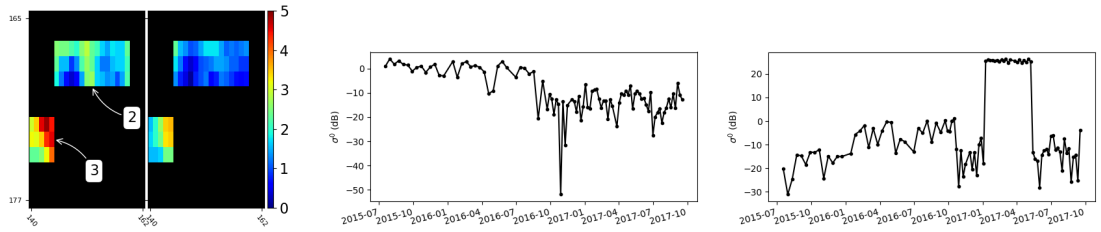


(a) Differencing (left) and CUSUM (right) temporal change maps.

(b) ANOVA (left) and PELT (right) temporal change maps.

(c) Differencing (left) and CUSUM (right) spatially weighted change maps.

(d) ANOVA (left) and PELT (right) spatially weighted change maps.



(e) Masked and annotated Figure 5.28d.

(f) A time series indicative of case 2.

(g) A time series indicative of case 3.

Figure 5.28: Case 2 and 3 - destruction of a building (2) and a temporary hangar (3) in Houston shown by temporal change maps and spatially weighted temporal change maps, with example time series for both cases. Beware that the colour bars have different scales (subsets from Figures 5.13 and 5.14, lines 165-177 and pixels 140-162).

These events are detected in the spatially weighted change maps for ANOVA and PELT, where especially ANOVA detects neighbouring pixels showing the same change behaviour over time. For these cases Differencing and CUSUM lack this spatial correlation in the spatially weighted change maps. Therefore, the Differencing- and CUSUM-based segmentations are not considered for case 2.

Case 2 is examined first. Both ANOVA and PELT show convincing segmentation results on pixel level, with ANOVA identifying 1-2 more change points on average (Figure 5.28e). The PELT algorithm identifies all the major change points in the time series. ANOVA-based CPD seems to also identify some smaller destruction related events, such as the parking of trucks and a favourable orientation of piles of sand, as can be seen in Figure 5.32b. Both the ANOVA- and PELT-based solutions show fewer identified change through time for some of the lower left points. Figure 5.32a shows why this is. These points are situated in a zone subject to radar shadow caused the elevated middle section of the building when sensed in a descending orbit geometry.

Further analysis of the time series segmentation with ANOVA and PELT seems to indicate a multi-stage destruction of the building; most likely in 3 stages, see Figure 5.29. This makes sense, going by the size of the building at roughly 80 by 35 m. The building can be divided in three parts; the right wing, the left wing and a middle section separating the two.

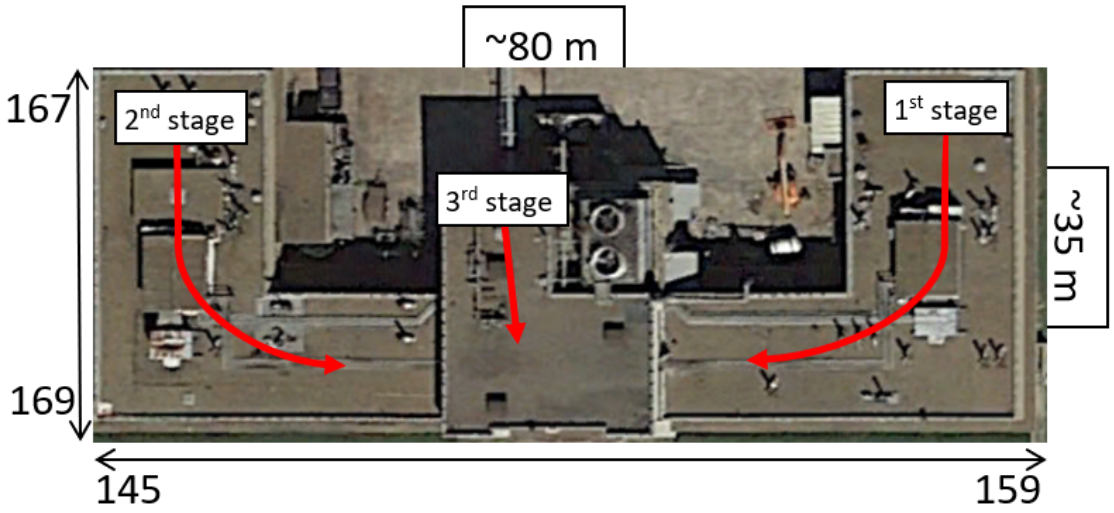


Figure 5.29: The properties of the building of the second test case, where the expected order of the 3 stages of the destruction are indicated. The x-axis and y-axis give the expected range of pixels and lines, that correspond with Figure 5.28.

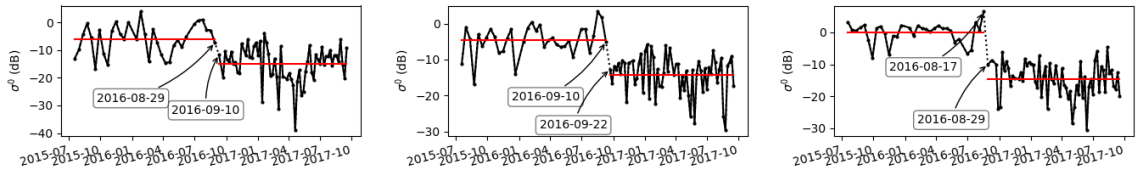
If the individual time series segmentations for each section of the building are examined, these 3 stages can be clearly identified. The destruction of the building started with the right wing, closely followed by the left wing and was finalized by tearing down the middle section, yielding three stages of destruction:

1. the demolishment of the right wing started between August, 17 and August, 29 (Figure 5.30c), and was finalized before September, 10;

2. the left wing was destroyed shortly after, starting between August, 29 and September, 10 (Figure 5.30a), and was likely finalized before September, 29; and

3. lastly, the middle section started being demolished between September, 10 and September, 22 (Figure 5.30b), with no clear ending date.

Recall the lower left pixels showing little to no change points, partly due to radar shadowing. The order in which the building was destroyed also aggravated this inability to sense the destruction of this part of the building, as it is not unlikely that the right part of the left wing was already demolished before the destruction of the middle part. This effectively means that the radar was blind to the change in that area in a descending orbit geometry. This can be verified by looking at the segmented time series of the different building sections in Figure 5.30.



(a) Segmentation of left wing. (b) Segmentation of middle section. (c) Segmentation of right wing.

Figure 5.30: The segmentation with ANOVA ($\alpha = 0.005$ and $D = 3$) of three time series indicative of the left, right and middle section of the building shown in Figure 5.29. Note that segmentation following PELT yields the same results and the x- and y-axis labels have been removed for increased visibility.

Case 3, also highlighted in Figure 5.28e, shows the temporary instalment of a hanger-like object, as seen in Figure 5.33. Looking at the time series of Figure 5.28g, it becomes very clear that the temporal evolution shows at least 3 different states; a low backscatter response, a point scatterer response and back to a low backscatter response. Moreover, 2 additional state changes could be argued to take place at February and November, 2017. This becomes more obvious when TS2 from Section 5.2.1 is also considered, which is the neighbouring pixel of the time series shown in Figure 5.28g. The segmentation of TS2, shown in Figure 5.10, indeed shows that all algorithms identify all aforementioned segments. An over-segmentation for Differencing and CUSUM is observed, and a subjectively perfect segmentation for ANOVA and PELT, based on visual analysis and comparison against Google Earth imagery.

In Figure 5.31, the PELT-based segmentation is shown. A total of 5 segments are indeed identified. Going by fairly sparse available ground truth of case 3, a working theory can be developed. Before construction began, the area was not really used all that often. With the installation of the neighbouring hangar, between 2015-08-01 (Figure 5.33a) and 2016-03-04 (Figure 5.33b), the area of interest had much more truck traffic, which explains the increase in mean backscatter observed in segment 2. Segments 3 and 5 show a dip in mean backscatter, back to the mean value observed in segment 1. This means that at those time intervals the backscatter response was mainly dominated by the flat ground. The ANOVA algorithm yielded an almost identical segmentation, with the exception of an additional segment identified for the first 3 data points. Meanwhile, Differencing and CUSUM did not identify the change in state between segment 1 and 2, as the change is to gradual for the algorithms to detect. Moreover, segments 3 and 5 were subject to over-segmentation.

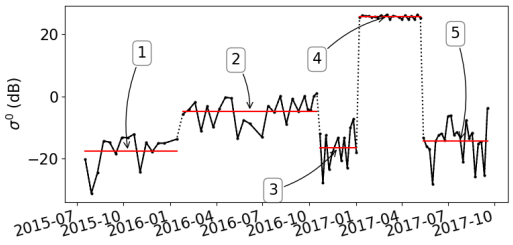
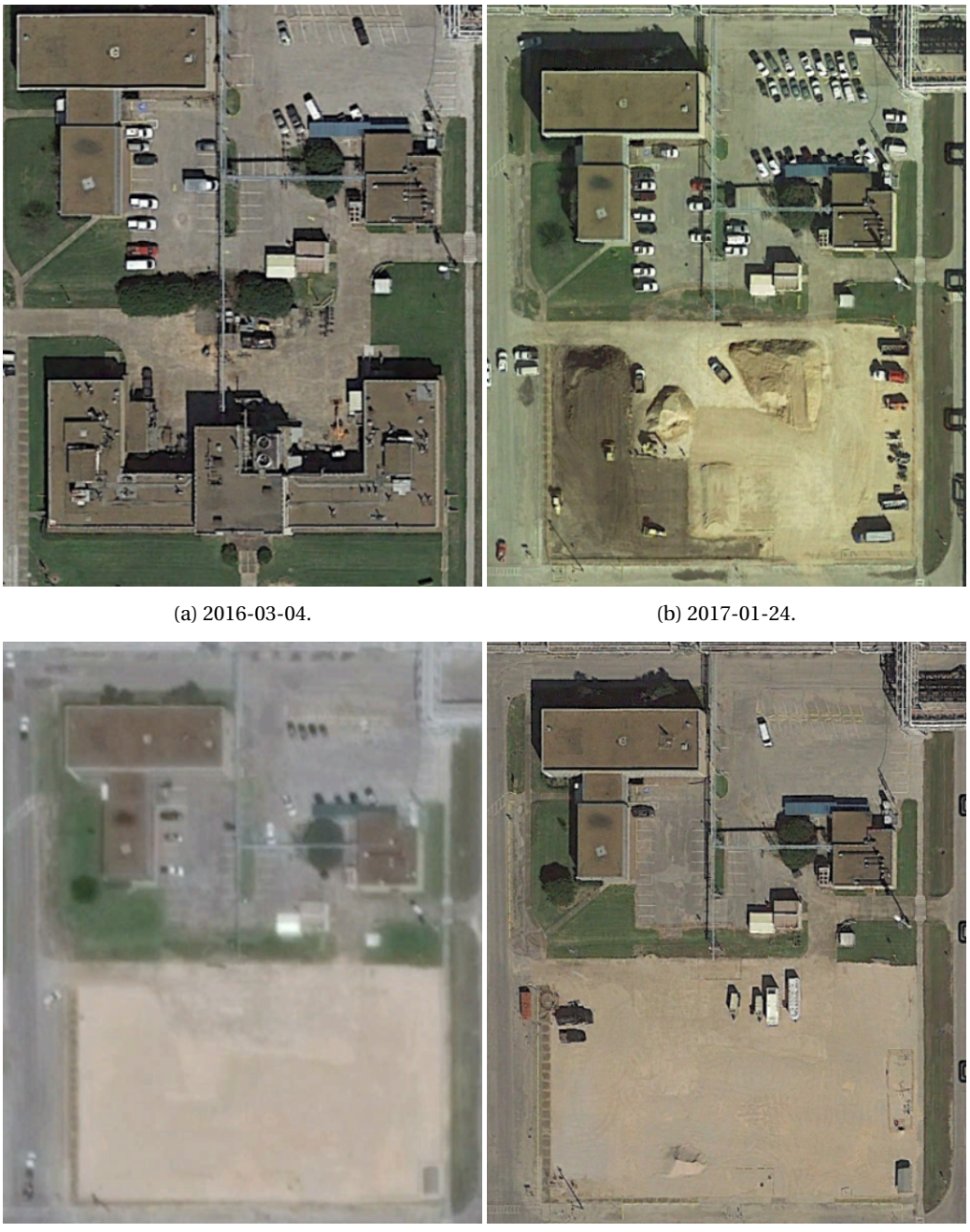


Figure 5.31: The segmentation with PELT ($\sigma = 10$) of the time series shown in Figure 5.28g. A total of 5 segments are defined by PELT, which are highlighted in the figure.



(a) 2016-03-04. (b) 2017-01-24.

(c) 2017-04-06. (d) 2017-10-29.

Figure 5.32: Satellite and aerial optical images for case 2.

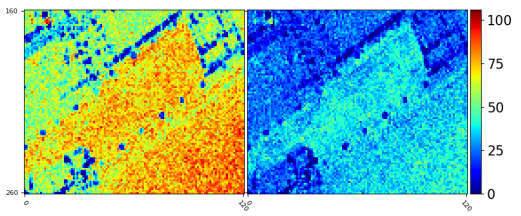
(a) 2015-08-01. (b) 2016-03-04.

(c) 2017-04-06. (d) 2017-10-29.

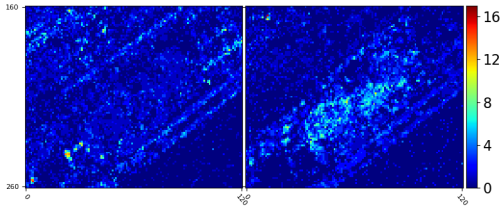
Figure 5.33: Satellite and aerial optical images for case 3.

A biannual fair

The fair in Schiehaven (case 4), Rotterdam is a biannually recurring event in late April to early May (the 'Koningskermis') and in late October to early November (the 'Najaarskermis'), and is being held in the same area every time (Festipedia, 2018). Figure 5.35c shows the default ground condition, without a fair being held, Figure 5.35a when the fair is being held and Figure 5.35b the breaking down of the fair. Over the time span sensed by TerraSAR-X up to 17 fairs have been held, assuming this biannual schedule. From the temporal change maps in Figure 5.34 it can be seen that Differencing detects some 75 change points on average, CUSUM around 50, ANOVA 2-3 on average and PELT yields up to some 12 change points.

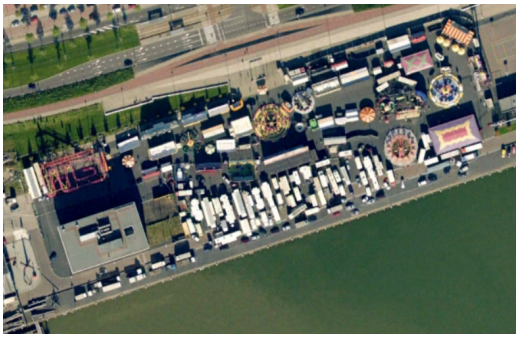


(a) Differencing (left) and CUSUM (right) temporal change maps.

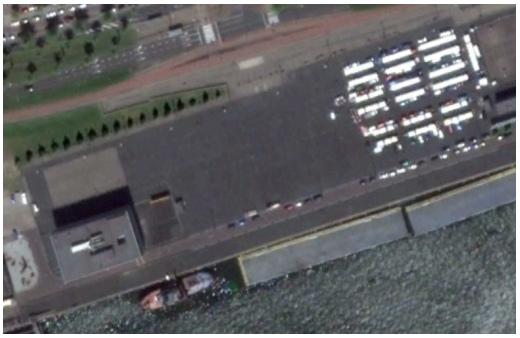


(b) ANOVA (left) and PELT (right) temporal change maps.

Figure 5.34: Case 4 - the biannual fair in Rotterdam shown by temporal change maps. Beware that the colour bars have different scales (subset from Figure 5.18, lines 160-260 and pixels 0-120).



(a) 2007-10-31.



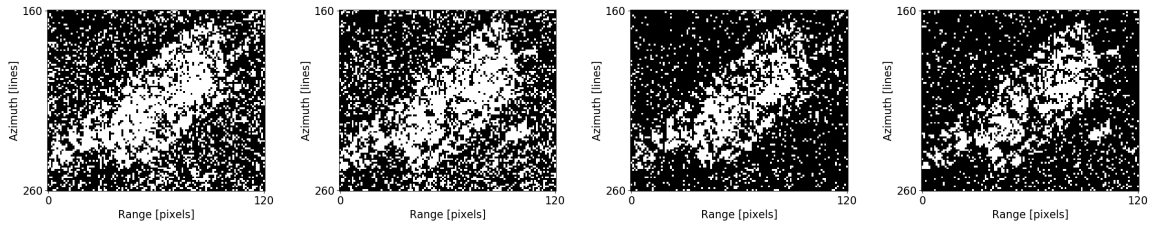
(b) 2012-05-27.

(c) 2017-05-27.

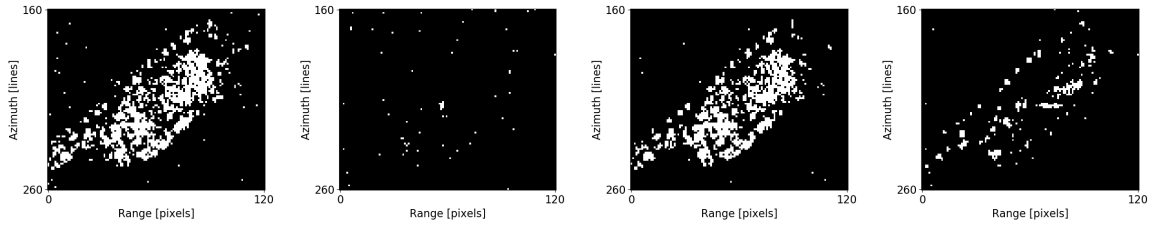
Figure 5.35: Satellite and aerial optical images for case 4.

The first (2009-05-11) and fifth (2011-05-08) fair in the time series are now analysed to determine the algorithms performance in detecting these events (Figure 5.36). It has to be noted that the first fair was observed for 3 consecutive measurements on average, and the fifth fair for 2. A very noisy Differencing result is observed, for both dates. However, the fair events do seem to be segmented properly. A slightly less noisy result is obtained with CUSUM. ANOVA-based CPD yields convincing results for detecting the first fair, but fails to properly detect the subsequent events. This goes back to the sensitivity of ANOVA to the placement

of these types of events, as discussed in Section 5.3. PELT shows a good result for the detection of the first event as well, but shows a much better detection of the fifth fair compared to ANOVA. Moreover, the result of PELT-based CPD seem slightly less noisy.



(a) Differencing 2009-05-11. (b) Differencing 2011-05-08. (c) CUSUM 2009-05-11. (d) CUSUM 2011-05-08.

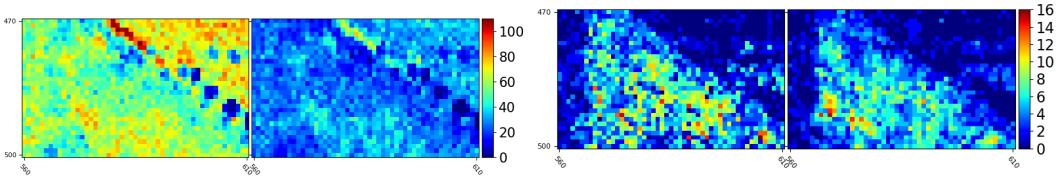


(e) ANOVA 2009-05-11. (f) ANOVA 2011-05-08. (g) PELT 2009-05-11. (h) PELT 2011-05-08.

Figure 5.36: The binary change masks for two fair events in Rotterdam. At 2009-05-11 the first sensed fair was broken down and 2011-05-08 is the date of the end of the fifth fair in the time series. White indicates an observed change point and black the no-change case.

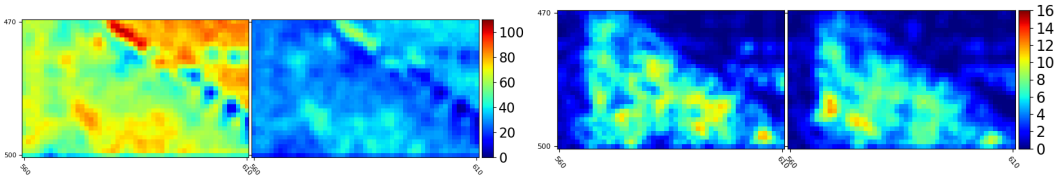
A highly variable container storage area

Case 5 describes a highly variable and unpredictable area which stores large sea containers. This makes for some very erratic behaviour in the time series. The temporal change maps in Figure 5.37b show a significant amount of changes detected for ANOVA and PELT, which have shown to be much more conservative than Differencing and CUSUM. Looking back at Figures 5.18c and 5.18d, this area shows up as a hotspot on both the ANOVA- and PELT-based images, for the entire Rotterdam subset. Therefore, it is good to take a look at the ground truth in this area, as shown in Figure 5.38.



(a) Differencing (left) and CUSUM (right) temporal change maps.

(b) ANOVA (left) and PELT (right) temporal change maps.



(c) Differencing (left) and CUSUM (right) spatially weighted change maps.

(d) ANOVA (left) and PELT (right) spatially weighted change maps.

Figure 5.37: Case 5 - highly variable container storage area in Rotterdam shown by temporal change maps and spatially weighted temporal change maps. Beware that the colour bars have different scales (subset from Figures 5.18 and 5.19, lines 470-500 and pixels 560-610).

The many detected change points, the lack of ground truth maps and the absence of some sort of pattern in the time series, makes it really difficult to assess the validity of these segmentations. Hence, the spatially weighted change maps are used to at least get a feeling for the quality of the results. A pixel showing similar values for the total changes over time (Figures 5.37a and 5.37b) and the second order neighbourhood weighted changes over time (Figures 5.37c and 5.37d) has a much higher chance of being correctly identified as a change point, as neighbouring pixels exhibit the same results. It is surprising to see the high level of spatial correlation between changes detected with Differencing and CUSUM. Another interesting observation is the fact that CUSUM actually has a much smaller spatial correlation than Differencing over water, seen in the top right of the subfigures in Figure 5.37c. Still, both algorithms detect way to many changes; up to 110 for Differencing. ANOVA and PELT show fairly similar result, both in shape of the change area and the number of changes detected in the area. Moreover, the spatially weighted change maps indicate a high spatial correlation between detected changes.

(a) 2005-12-31.

(b) 2016-09-08.

(c) 2017-05-27.

Figure 5.38: Satellite and aerial optical images for case 5.

5.5. Discussion

So far, a bifurcation in Section 5 exists in the sense that the simulated data stacks and the real SAR date cubes segmentations are discussed separately. This section aims to bring together all the results and to reflect upon them; essentially converging all the information in Section 5 to one analysis.

The computed accuracy, precision and ROC curves for the defined discontinuity functions, i.e. Dirac, Heaviside and Rectangular functions, show that Differencing and CUSUM perform roughly the same for each defined change type. PELT, and especially ANOVA, are shown to be much less sensitive to Dirac-type change behaviour. This also means that these algorithms are much less sensitive to speckle, evident from the fewer segmentations of pixels observed over water for example. It is important to realise that when one is analysing time series with significant outliers, these data points can both be attributed to an actual change event (Dirac-like) and speckle. This relative insensitivity to speckle can be traced back to the equations of the PELT and ANOVA algorithms. Both algorithms analyse the entire SAR time series before segmenting the series; ANOVA by comparing all combinations of the 2-group configuration and PELT through recursion. CUSUM also takes the temporal evolution into account, but only through the cumulative sum of the amplitude differences be-

tween consecutive measurements. This makes it more robust against speckle and the better choice for change detection of Dirac-like change behaviour than Differencing, which is also evident from the accuracies given in Table 5.2. In general, taking the temporal evolution of a pixel into account is shown to be very beneficial, especially if the entire time series is examined prior to segmentation.

For online ('real time') applications, where new data is added once available, the four offline CPD algorithms can be applied. The repeat cycle of Sentinel-1 and TerraSAR-X is 6 and 11 days, respectively. This means that new observations in a the same orbit geometry are available every 6 or 11 days, which is more than sufficient time to process large data sets. These new data point can be appended to the data points of the last identified segment, which will then be used as input data for the algorithms. This can even be done in an automated way, for continuous monitoring applications for example.

The implemented algorithms are all pixel-based methods, not considering the spatial component in the decision rule. However, the spatially weighted change maps show that a lot of information resides in the spatial neighbourhood of a pixel. The superior spatial resolution of TerraSAR-X data, compared to Sentinel-1 data, makes it so that change events with a smaller spatial footprint are easier to identify, as seen in Figure 5.37 for example. The longer time series that is available in the Rotterdam stack, 240 compared to Houston's 94, also makes the PELT algorithm and especially ANOVA-based CPD less sensitive to speckle. This is best seen by comparing the ANOVA-based segmentation over water in Figures 5.13c and 5.18c. Subsequently, the ANOVA and PELT algorithm become less likely to pick up a single acquisition change event in a longer time series. Combine this with a strong dependency of ANOVA on the position of the change in the time series, observed in Figure 5.21, and the poor detection of the fair events in Rotterdam (Figure 5.36) with ANOVA is explained.

Erroneous data points, introduced in the pre-processing (e.g. absolute calibration error) or because of inherent radar characteristics (e.g. side lobes), are very apparent in the change maps. The effects of four such error sources are seen in the change maps. Firstly, the horizontal line of changes in the Rotterdam stack around the 450 line is likely caused by a pre-calibration error, best seen in Figure 5.19b. Secondly, the over-segmentation of PELT due to a single faulty data point following calibration, seen in Figure 5.17 and Figure A.2. Thirdly, the side lobes originating from the 50 line 250 pixel point in Figure 5.13c among others. And lastly, the effects of speckle, most noticeable in change maps following the Differencing algorithm.

Applying a moving average spatial filter to smooth the SAR images in the Houston stack shows a big improvement in segmentation over non-filtered images, especially for the Differencing, CUSUM and PELT algorithms; albeit no improvement is seen for the pixels affected by the side lobes, as can be seen in Figure 5.15. However, using spatial filtering comes at a loss of spatial resolution. A temporal filter is a good alternative to spatial filters to preserve spatial resolution. Preliminary tests with using a moving average filter in the temporal domain (not shown) indicate that such a filter is not compatible with SAR-based CPD. More adaptive temporal filters are required, as previously clearly defined break points will over-smooth with a non-adaptive temporal filter. Without filtering, CPD methods in multi-temporal change detection that use the temporal evolution in such a way that the influence of speckle noise is minimal, are preferred. Between the four tested CPD algorithms, ANOVA is the least sensitive to speckle, closely followed by PELT. The CUSUM algorithm and especially Differencing are much more sensitive to speckle than ANOVA and PELT.

Lastly, the transformation of the Rotterdam and Houston data sets to a near-Normal distribution from a negative Exponential distribution, by using the decibel notation of the sigma nought input values for the change detection algorithms, shows the added value of data transformations in general and that taking into account the distribution of the data is very important. The confusion matrices in Table 5.2 show that all four CPD methods perform significantly better for Normally distributed data compared to Rayleigh distributed data. Moreover, the quantile normalization results for the ANOVA and PELT algorithm indicate that both algorithms are fairly robust against non-normality, see Figure 5.24 for example. Differencing and CUSUM are very sensitive to any deviations from Normally distributed data, because of the algebraic approach these algorithms rely on. Preferably, the data would follow a fairly narrow distribution, i.e. a rather homogeneous observed area. The left-side tail in the transformed Houston stack histograms, seen in Figure 4.3b, is an early indicator of more likely over-segmentation of Differencing and CUSUM over small backscatter surfaces, such as water.

6

Conclusions and recommendations

6.1. Conclusions

A comprehensive and comparative research is conducted to determine how to optimally use a stack of co-registered Synthetic Aperture Radar (SAR) images to detect abrupt changes in the statistical properties of the time series in the data cube. Since the SAR phase doesn't hold that much information and there is no emphasis on a particular application, a focus on an unknown number of changes in the mean amplitude on pixel-level is applied. Based on analysis of statistical properties of the SAR data and various change point detection (CPD) classification frameworks four CPD algorithms selected and implemented, and compared by applying them on simulated and real SAR data. This is done in order to answer the main research question:

> *How can a SAR amplitude data cube be optimally used for automatic abrupt change detection in urban areas, in both an efficient and accurate manner?*

In order to answer the research question, this section is divided in three parts. First, the per-method considerations and applicability of the implemented algorithms is assessed. Secondly, the research objective is placed in a wider framework by evaluating the general CPD process. And lastly, an overall conclusion is given.

Implementation and applicability of used CPD algorithms

A total of four CPD algorithms are implemented that analyse a SAR time series for change points, namely (simple) Differencing, CUSUM, ANOVA and PELT. First, each CPD algorithm is discussed separately, regarding the main principle and computational cost, the sensitivity to radar inherent characteristics, various error sources and threshold selection, the compatibility with Dirac-, Heaviside- and Rectangular-like change behaviour and the compatibility with a certain level of data heterogeneity. Based on the aforementioned, the general applicability of each method is discussed with relation to the other methods. When taking the temporal component into account, the computational cost of the algorithms is an important factor with the significant increase in data size due to additional images. The actual running time of all the algorithms is also dependent on optimization and the number of detected change points for ANOVA and PELT. The implementations in this research for Differencing and CUSUM have the lowest run time. Moreover, on average PELT is at least an order of magnitude faster than ANOVA.

Differencing is an easy to understand and implement approach to the change detection problem, as it uses a simple algebraic bi-temporal approach with a complexity of $\mathcal{O}(n)$. It is capable of detecting multiple transitions in the time series, because of the iterative implementation. However, it only compares consecutive measurements and therefore completely neglects information in the temporal evolution of the SAR time series. This makes it extremely sensitive to all error sources and other radar inherent characteristics, such as speckle, side lobes and pre-processing errors. Moreover, the large dynamic range observed in SAR time series isn't particularly suited for methods using a static threshold, especially in heterogenous areas. Differencing is best used for a first quick estimation or detection of significantly large events, and preferably in data sets observing fairly homogenous areas, due to the static threshold. Lastly, Differencing is highly sensitive to the threshold selection, as the method is purely based on comparing changes in amplitude of consecutive measurements against a user determined threshold value.

The CUSUM algorithm, as used in this research, considers the temporal evolution through the cumulative sum of previous amplitude differences at a computational cost of $\mathcal{O}(n)$. Because of the similar approach to CPD as Differencing, it is best to discuss CUSUM with respect to Differencing, and highlight the differences. CUSUM is much less sensitive to speckle and threshold selection, because of the drift correction that acts as a guard against overfitting. With comparable run times, CUSUM is preferred over Differencing for abrupt change detection in SAR time series. CUSUM is limited in its ability at detecting temporal and permanent change behaviour when intermediate measurements are involved. That is to say, one or more data points that have a backscatter value between the different temporal or permanent states. Here, the drift correction actually counteracts the ability to detect actual change behaviour.

The implementation of the ANOVA algorithm, iteratively comparing the variability between and within two groups at a computation cost of $\mathcal{O}(n \log n)$, is best used for detecting Heaviside-like changes over time. For these types of signals, the algorithm is sensitive to very small shifts in mean, even at small significance levels. Furthermore, the detectability of Dirac pulses, and Rectangular responses to a lesser extent, with the ANOVA CPD algorithm has a strong sensitivity to the position of the temporal change behaviour in the time series. ANOVA is most sensitive to outliers at the edges of the series. An outlier which is part of a larger segment causes a slight increase for the within-group variability, whereas a single measurement has a within-group variability of 0. Considering that a small within-group variability (denominator) increases the F value, easier detection for Dirac pulses at the edge of time series can be explained. ANOVA is fairly robust against speckle noise and other sources of non-texture related backscatter offsets. This follows from the similar CPD results for the unfiltered and filtered data stack and the general insensitivity to Dirac-like behaviour that is not necessarily linked to actual changes. The combination of robustness against speckle and being sensitive to small shifts in mean backscatter for permanent change, make ANOVA-based CPD suitable for a wide range of near-Normally distributed data sets. Moreover, previous realisations in literature of ANOVA-based change detection in SAR time series either used a fixed 2- or 3-step model approach or only considered the second segment in an iterative approach. Taking both segments, divided by a detected change point, into consideration makes the algorithm more expensive, but also improves the solution. This is evident from the sensitivity of the ANOVA algorithm to the order of change point detection.

The PELT algorithm uses dynamic programming to find the optimal segmentation of a time series by minimizing a penalised cost function, and includes an inequality based pruning step to speed up the algorithm. In the worst case scenario of no pruning it is $\mathcal{O}(n^2)$ and it has a linear computational cost $\mathcal{O}(n)$ in the best case scenario, i.e. regularity in the change point distribution. When the effects of speckle on Differencing and CUSUM are not taken into consideration, PELT and Differencing/CUSUM detect similar single acquisition change events. On the other hand, when compared to ANOVA, PELT only detects temporary and permanent change with a significantly large difference in mean backscatter. This is because the expected standard deviation should be high enough to not trigger too many false positives and low enough to detect change behaviour characterized by a small in- or decrease in backscatter. In other words, the results of PELT-based change detection are more dependent on parameter selection than ANOVA. The PELT algorithm is very versatile in its use, as it can adopt different cost functions that also work with non-Normally distributed data. This, combined with a fairly good insensitivity to error and noise sources, makes PELT suitable for a variety of data sets, both homogenous and heterogeneous.

Differencing and CUSUM share their applicability in change detection in homogenous areas (e.g. pastures) and are best used for severe outlier detection (e.g. ships over water) or for a first CPD estimation at low computational cost. This should be combined with a filter step, as a simple 3x3 non-weighted running mean filter already yields much better segmentation results. The algebraic approach, combined with a static threshold and a large dynamic range in backscatter in heterogeneous areas makes Differencing and CUSUM better used in homogenous areas. CUSUM is nearly always preferred over Differencing, as it takes the temporal evolution of a pixel into account resulting in less false positives.

ANOVA and PELT generally yield much better results than Differencing and CUSUM, but each find their strengths in different applications, with some overlap in use. Hence, the algorithm consideration should be made on case-to-case basis. The ANOVA-based implementation finds its strengths in permanent change (e.g. destruction of a building), temporary change sensed over 3 or more acquisitions (e.g. activities during tourist season) for time series similar to the ones used in this research and event type behaviour that results in a different mean backscatter after the event compared to that of prior to the event (e.g. building expansion or renovation). Using PELT for SAR time series segmentation is preferred for temporary change that has a similar mean backscatter prior and after the change event (e.g. a mobile office container) and in particular changes measured over 2 or less acquisitions (e.g. a fair). PELT offers some distinct advantages over ANOVA, because

of its versatility in use by changing the expected distribution in the cost function, whereas ANOVA is based on the assumption of Normally distributed time series. However, both implementations of PELT and ANOVA are shown to be fairly robust to non-Normally distributed data. Moreover, PELT has a fairly equal detection rate of each change signal type. ANOVA outperforms PELT for permanent change and temporal change sensed over multiple acquisitions.

The CPD process in SAR time series

For proper change detection in SAR time series, one needs to consider much more than just which CPD algorithm to use. In fact, four components are identified to be carefully considered: the *data selection and storage*, *pre-processing steps*, *CPD algorithm selection and implementation* and *post-processing analysis*. The important considerations per aspect of the CPD procedure as a whole are discussed.

First, data selection and storage. The data selection is extensively covered in literature, however data storage often seems overlooked or regarded as an afterthought going by the lack of coverage in CPD literature. Data selection is intertwined with the change detection application, which in its turn determines the change signal behaviour that has to be captured. The spatial resolution, wavelength, polarization, orbit geometry, imaging mode, repeat cycle and availability should all be weighed carefully. Consider wanting to determine the aftermath of an earthquake with a time series of Sentinel-1 or TerraSAR-X data; the extent of the changes is likely captured even with the coarser spatial resolution of Sentinel-1, which means less computational effort and guaranteed availability of the data. Data storage is equally important, as CPD is an iterative process requiring fine tuning through threshold selection and comparing results for unfiltered and filtered data, for example. Use of a structured file format, such as NetCDF-4/HDF5 and HDF5, even without chunking, already show a great increase in time efficiency in the CPD procedure. Smart and efficient storage does not improve the time series segmentation, but it does streamline the process as a whole.

Secondly, the data needs to be properly pre-processed. Calibrating the stack of images is either done by relative or absolute calibration. Absolute calibration to sigma nought is especially beneficial for change detection methods using direct thresholds. This creates a link between the expected physical change and the threshold selection. A SAR data cube calibrated in beta nought does not correct for the local incidence angle, and therefore has a small range dependency. For the Houston stack beta nought and sigma nought values, a 0.6 dB difference between near and far range is observed. With a used threshold of 10 dB for Differencing, CUSUM and PELT, this range dependency is only 2 orders of magnitude smaller than the threshold parameters. Therefore, using beta nought calibrated SAR images covering large surface areas is to be avoided or should be accounted for. Moreover, the absolute radiometric accuracy of the absolute calibration of Sentinel-1 and TerraSAR-X is 1 dB and 0.6 dB, respectively. Data deficiencies, such as errors introduced with calibration, pre-processing artefacts in general and side lobes are important to identify, as these can introduce profound segmentation errors.

Thirdly, the implementation and selection of a suitable CPD algorithm. One is advised to use existing classification frameworks, combined with the probability distribution of the data set, to make an informed decision on what CPD algorithm to use. It is shown that a logarithmic transformation of a data set following a negative exponential distribution, expected for intensity or absolutely calibrated values, can transform the distribution to an approximate Normal distribution. Generally speaking, for SAR time series segmentation in urban areas offline unsupervised algorithms are suggested. Using training data with supervised methods isn't suitable for the speckle induced unpredictive behaviour in SAR time series. Simple arithmetic-based change detection is shown to be insufficient in this context. The repeat cycle of SAR satellites makes the use of online algorithms and their increased computational load an unnecessity. Furthermore, supervised learning requires ground truth for training data, which is often absent or sparsely available in long time series analysis. Once decided on an algorithm suited for the application and applicable to the observed data distribution, proper tuning of the algorithm will decide the quality of the segmentation. Simulated data can be useful in assisting with this, as quality metrics such as the ROC curve can be easily computed that describe the relation between change behaviour, threshold selection and change signal sensitivity.

Lastly, the post-processing analysis. As mentioned, ground truth is hard to come by for multi-temporal change detection. Therefore, the spatially weighted temporal change map is introduced, in order to assess the reliability of change point by taking the spatial relation into account. These maps allow for quick identification of significant change events that can then be compared to auxiliary information, in order to assess the overall change detection performance. It is important to consider the introduced and data inherent error sources throughout the CPD process. ANOVA is shown to be much less dependent on speckle reduction than CUSUM, for example.

Overall conclusion

The most *efficient* and *accurate* way to identify unknown abrupt change behaviour over time is determined in a data cube of co-registered uncalibrated SAR images, by implementing four different pixel-based CPD methods and reviewing the entire CPD process. The most important conclusions with regards to the main research question are given point wise.

- Efficiency and accuracy are often opposite forces. This observation has to be considered when selecting a CPD method to process a SAR data cube. More sophisticated, and hence often computationally expensive, CPD methods generally yield better segmentation results. The accuracy of the generated change maps is not only determined by the CPD method, but also the suitability of a method for a certain application (i.e. the expected temporal change function) and the error sources in the segmented data. Moreover, efficiency is more than just the computational cost of the CPD algorithm, as change detection is a comprehensive multifaceted process. In general, there is no CPD approach that is optimal for all applications, and under all circumstances. However, for multi-transitional change detection in SAR time series it is suggested to use offline unsupervised CPD algorithms.

- The distribution of a SAR data cube determines the applicability of CPD methods. Methods such as ANOVA-based change detection have an underlying assumption of a Normally distributed data for example. Here, data transformations can be very useful. This is shown by the transformation of the sigma nought calibrated data cubes of Houston and Rotterdam to a decibel scale. This transformation yielded a slightly negatively skewed Normal distribution from a negative Exponential distribution. At this point it is important to assess the robustness of a CPD method to deviations in data distribution from the underlying assumption of the method.

- The suitability of simple algebraic methods in urban areas is limited and taking the temporal evolution into account only slightly improves this. The highly heterogeneous urban test areas in Houston and Rotterdam show a large dynamic range of backscatter values, even in decibel notation. Therefore, it is suggested to only use these kind of methods in urban areas for applications yielding a significantly large amplitude offset in combination with a conservative threshold and a spatial filtering step. Probabilistic methods which take the temporal evolution into account are shown to be much more suitable for CPD in urban areas, as they capture changes at all levels of mean amplitude response in a time series while being relatively robust against speckle; even without any filtering of the data. In general, good segmentation of SAR time series isn't only linked to the level of sophistication of a CPD method; the dynamic range of backscatter values in the data set, the statistical distribution of the data and the magnitude of the backscatter offset caused by the change event also play a vital role.

- Often, change detection follows a certain application. The change behaviour linked with this application is described by a certain expected discontinuity function. Different CPD methods have different strengths and weaknesses in detecting certain change events. It is not always immediately apparent what the particular strengths of an algorithm are from just its equation. This is especially true for more complex methods. Using simulated SAR data, following the speckle model for example, is a great way to easily identify the inner workings of a CPD algorithm by computing quality metrics, such as the ROC curve, for various discontinuity functions.

- Choosing suitable data for some application will increase the amplitude offset between the default state and some change behaviour, e.g. by using co-polarized data for water related events. Moreover, the length of the time series should also be considered. This is most apparent from the relation between the ANOVA algorithm, the Dirac-type change event and the position of change event in the time series. Too many data points can cause information to get lost in the abundance of measurements, dependent on the equation of the CPD algorithm. When working with sufficiently large 3D data, allocating time to facilitate smart data storage becomes beneficial. By storing the SAR data cube in structured file formats, such as HDF5, the entire CPD process will become more efficient. It allows for faster evaluation tests and visualization, and it lowers the run time of the CPD algorithm when applied on the entire stack.

- Pre-processing and error reduction steps are very important in pixel-based change detection in a SAR data cube. Calibration of the data is vital to proper change detection, as to eliminate changes between acquisitions not caused by changes on the ground. Moreover, certain radar inherent characteristics should be carefully taken into account and reduced if possible. Speckle, geometric distortions and side

lobes will trigger false positives or negatives, when compared to ground truth. Methods that consider the temporal evolution of a pixel in a profound way are shown to be much more robust against such phenomena. Moreover, erroneous data points can have a great impact on the CPD solution, as evident from the horizontal lines in the Rotterdam change maps and the severe over-segmentation of PELT following a single faulty data point.

- Assessing the change maps, following multi-transitional abrupt change detection can be troublesome. Reliable ground truth data is hard to come by and labour intensive to create, even for bi-temporal change detection. Alternatively, with the help of temporal change density maps and spatially weighted change density maps, areas with significant change events can be identified. These 'hotspots' can be used to cross-examine against optical data to assess the reliability of the segmentation.

6.2. Recommendations

The recommendations for further research following the conclusions are:

- Further investigate the usefulness and compatibility of PELT in combination with SAR-based change detection. SAR data following a non-Normal distribution should be evaluated with a compatible cost function.
- Investigate a different ANOVA CPD approach. In the implemented approach the entire time series is divided in two segments and compared based on the segment's variance. Alternatively one could iterate over the time series and compare all previous measurements with the next unseen measurement. If the F value exceeds the threshold a segment is found and the procedure is repeated staring at the point of segmentation. This means a more sensitive method for Dirac and Rectangular detection, but would make segmentations at the begin of the time series less reliant. A guard for overfitting would be needed.
- Improvements of the implemented change detection methods can likely be achieved by analysing the solutions of the algorithms with a more sophisticated adaptive filtering procedure. Results in various other studies have shown simple thresholding methods, similar to the Differencing algorithm, that display much more satisfactory results. These studies all have a large emphasis on the filtering procedure, often in the spatial domain (e.g. Lee filter). Adaptive temporal filters (e.g. Kalman filter) can preserve spatial resolution, while still providing sufficient smoothing for pixel-level CPD methods.
- Apply the algorithms on test data sets, which are often used in literature. Data sets such as the Ottawa (1997), Bern (1999), Pavia (2000), San Francisco (2003/2004) or Yellow River (2008/2009) data set. These change detection problems have available ground truth and allow for quality metric comparison between methods.

6.3. Contributions

The contributions of this research are:

- Created absolute calibration scripts, able to process Sentinel-1 and TerraSAR-X co-registered SLC data. These scripts can be used to calibrate SLC data to beta nought, sigma nought and/or gamma nought, with limited user input and background knowledge.
- Created an interactive tool that can be used to quickly analyse SAR data cubes, by allowing pixel selection, date selection and CPD selection. It also provides useful information such as the mean image and time series statistics.
- Provided an overview of change detection methods, frameworks, operators and quality metrics.
- Implemented four CPD algorithms able to detect multi-transitional changes in SAR imagery, namely Differencing, CUSUM, ANOVA and PELT.
- Improved the ANOVA-based change detection methodology by considering all previously created segments for iterative segmentation.
- Introduced the CUSUM and PELT algorithm in SAR-based change detection.

- Implemented the spatially weighted temporal change map to streamline the assessment of temporal change behaviour results without ground truth.
- Demonstrated the potential of PELT-based change detection with SAR time series.
- Showed the benefits of using a multi-temporal approach over classical bi-temporal change detection approaches in SAR imagery.
- Highlighted the profound implications of pre-processing errors on the segmentation result.

Bibliography

R.P. Adams and D.J.C. MacKay. Bayesian online changepoint detection. 2007.

H. Aghababaee, J. Amini, and Y.C. Tzeng. Improving change detection methods of SAR images using fractals. *Scientia iranica*, 20(1):15–22, 2013.

Airbus. Radiometric calibration of TerraSAR-X data. 2014a.

Airbus. TerraSAR-X image product guide, 2014b.

O.A. Ajadi, F.J. Meyer, and P.W. Webley. Change detection in Syntheric Aperture Radar images using a multiscale-driven approach. *Remote sensing*, 8/2016:1–27, 2016.

H. Akaike. A new look at the statistical model identification. *IEEE transactions on automatic control*, 19(6): 716–723, 1974.

J. Alphonse and V.G. Biju. Recognition of changes in SAR images based on Gauss-log ratio and MRFFCM. *IJASCSE International journal of advanced studies in computer science and engineering*, 4(5):65–71, 2015.

S. Aminikhanghahi and S.J. Cook. A survey of methods for time series change point detection. *Knowledge and Information Systems*, 51:339–367, 2017.

H.H. Arsenault and G. April. Properties of speckle integrated with a finite aperture and logarithmically transformed. *J. Opt. Soc. Am.*, 66(11):1160–1163, 1976.

I.E. Auger and C.E. Lawrence. Algorithms for optimal identification of segment neighborhoods. *Bulletin of Mathematical biology*, 51(1):39–54, 1989.

S. Balanos, D. Stiff, B. Brisco, and A. Pietroniro. Operational surface water detecting and monitoring using Radarsat 2. *Remote sensing*, 8(285), 2016.

R. Bamler. Principles of Synthetic Aperture Radar. *Surveys in Geophysics*, pages 1–6, 2015.

G.A. Barnard. Control charts and stochastic processes. *Journal of the royal statistical society*, 21:239–271, 1959.

Bruzzone Bazi, Y., L., and F. Melgani. Image thresholding based on the EM algorithm and the generalized Gaussian distribution. *Pattern recognition*, 40:619–634, 2006.

Y. Bazi, L. Bruzzone, and F. Melgani. An unsupervised approach based on the generalized Gaussian model to automatic change detection in multitemporal SAR images. *IEEE transactions on geoscience and remote sensing*, 43(4):874–887, 2005.

C. Benedek et al. Multilayer Markov Random Field models for change detection in optical remote sensing images. *ISPRS Journal of Photogrammetry and remote sensing*, 107:22–37, 2015.

A. Bhattacharya. Lecture notes in Speckle filtering/ speckle statistics. URL `http://www.csre.iitb.ac.in/~avikb/GNR647/Lec_11_Speckle.pdf`. [Online: accessed 2018-29-05].

B.M. Bolstad, R.A. Irizarry, M. Astrand, and T.P. Speed. A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics*, 19(2):185–183, 2003.

F. Bovolo et al. A multilevel approach to change detection for port surveillance with very high resolution SAR images. *Proc. Multitemp.*, pages 9–12, 2011.

C. Brantly. Sigma nought and beta nought image from radar image in IMAGINE. [Forum: accessed 2018-01-18], 2016. URL `http://community.hexagongeospatial.com/t5/tkb/articleprintpage/tkb-id/KS_ERDAS_IMAGINE_Q_A/article-id/170`.

D. Brunner, L. Bruzzone, and G. Lemoine. Change detection for earthquake damage assessment in built-up areas sing very high resolution optical and SAR imagery. *Proc. IEEE IGARSS*, pages 2820–2827, 2010.

L. Bruzzone. Change detection in multitemporal remote sensing images, 2016. URL `https://earth.esa.int/documents/973910/2642313/LB1to3.pdf`. Presentation at ESA.

L. Bruzzone and D.F. Prieto. Automatic analysis of the difference image for unsupervised change detection. *IEEE transactions on geoscience and remote sensing*, 38(3):1171–1182, 2000.

L. Bruzzone and S.B. Serpico. An iterative technique for the detection of land-cover transitions in multitemporal remote-sensing images. *IEEE Trans. Geosci. Remote Sensing*, 35:858–867, 1997.

P.S. Bullen. Handbook of means and their inequalities, 2003.

H. Cantalloube and C. Nahum. How to Compute a Multi-Look SAR Image?, 2000.

A. Carter. Lecture notes in Structured file formats: NetCDF and HDF, 2015.

T. Celik. A Bayesian approach to unsupervised multiscale change detection in Syntheric Aperture Radar images. *Signal process.*, 90:1471–1485, 2010.

T. Celik and K.-K. Ma. Multitemporal image change detection using undecimated discrete wavelet transform and active contours. *IEEE transactions on geoscience and remote sensing*, 49(2):706–716, 2011.

H. Chen and N. Zhang. Graph-based change-point detection, 2014.

W.K. Chen. *The electrical engineering handbook*, page 685. Elsevier Academic press, 2005. ISBN 0-12-170960-4.

S.-S. Cheng and F.-J. Yu. A CUSUM control chart to monitor wafer quality. *Journal of Information and Optimization Sciences*, 35(5-6):483–501, 2014.

P. Coppin and M. Bauer. Digital change detection in forest ecosystems with remote sensing imagery. *Rem. Sens. of the Env.*, 13:207–304, 1996.

J. Davis and M. Goadrich. The relationship between precision-recall and ROC curves, 2006.

P. Deer. Digital change detection techniques: civilian and military application, 1999.

R.J. Dekker. Speckle filtering in satellite SAR change detection imagery. *Remote Sens.*, 19(6):1133–1146, 1998.

DEOS. Delft object-oriented Radar Interferometric Software User's manual and technical documentation, 2008. URL `http://doris.tudelft.nl/software/doris_v4.02.pdf`. [Online: accessed 2017-05-23].

O. Dogan and D. Perissin. Detection of multitransition abrupt changes in multitemporal SAR images. *IEEE Journal of selected topics in applied earth observations and remote sensing*, 24(1):73–95, 2010.

D. Eberly. Least squares fitting of data, 1999. URL `https://www.geometrictools.com/Documentation/LeastSquaresFitting.pdf`.

T.F.M. Eineder et al. TerraSAR-X ground segment basic product specification document. 2008.

T. Eltoft. Speckle: modelling and filtering. pages 1–6, 2013.

EO. Copernicus: Sentinel-1 — the SAR imaging constellation for land and ocean services, 2002a. URL `https://directory.eoportal.org/web/eoportal/satellite-missions/c-missions/copernicus-sentinel-1`. [Online: accessed 2018-31-05].

EO. TSX (TerraSAR-X) mission, 2002b. URL `https://directory.eoportal.org/web/eoportal/satellite-missions/t/terrasar-x`. [Online: accessed 2018-31-05].

ESA. Acquisition modes, a. URL `https://sentinel.esa.int/web/sentinel/user-guides/sentinel-1-sar/acquisition-modes`. [Online: accessed 2017-05-25].

ESA. Product types and processing levels, b. URL https://sentinel.esa.int/web/sentinel/user-guides/sentinel-1-sar/product-types-processing-levels. [Online: accessed 2018-26-05].

ESA. Radar course 3, c. URL https://earth.esa.int/web/guest/missions/esa-operational-eo-missions/ers/instruments/sar/applications/radar-courses/content-3/-/asset_publisher/mQ9R7ZVkKg5P/content/radar-course-3-real-aperture-radar. [Online: accessed 2018-18-04].

ESA. Technical guides: Sentinel-1 SAR Performance, d. URL https://sentinel.esa.int/web/sentinel/technical-guides/sentinel-1-sar/sar-instrument/performance. [Online: accessed 2018-01-17].

ESA. Geometry glossary, 2014. URL https://earth.esa.int/handbooks/asar/CNTR5-5.html#eph.asar.gloss.geo:geometry.

V. Fernandez. Detection of breakpoints in volatility. *Estudios de administracion*, 11(1):1–38, 2004.

Festipedia. Kermis Rotterdam. [Online: accessed 2018-07-28], 2018. URL https://www.festipedia.nl/kermis-rotterdam.

T. Fung and E. LeDrew. The determination of optimal threshold levels for change detection using various accuracy indices. *Photogrammetric Engineering and Remote Sensing*, 54(10):1449–1454, 1988.

Schwarz. G. Estimating the dimension of a model. *Ann. Statist.*, 6(2):461–464, 1978.

L. Gagnon and A. Jouan. Speckle filtering of SAR images - a comparative study between complex-wavelet-based and standard filters. 1999. Conference: "Wavelet applications in signal and image processing V".

F. Gao et al. Change detection from Synthetic Aperture Radar images based on neighborhood-based ratio and extreme learning machine. *Journal of applied remote sensing*, 10(4):(040619–)1–14, 2016.

F. Gao et al. Change detection in SAR images based on deem semi-NMF and SVD networks. *remote sensing*, 9/2017:1–20, 2017.

G. Gao. Statisical modeling of SAR images: a survey. *Sensors*, 10:775–789, 2010.

HDF group. FITSIO, HDF4, NetCDF, PDB and HDF5 performance: some benchmarks results, 2002. URL https://www.slideshare.net/HDFEOS/05-2chdf5-performance. Workshop slides in Science data processing [Online: accessed 2018-03-06].

F. Gustafson. *Adaptive filtering and change detection*, pages 19, 65–67. John Wiley & Sons Ltd, 2000. ISBN 0-471-49287-6.

L. Halounova. Lecture notes in Radar basics, 2009. [Online: accessed 2018-05-25].

R.F. Hanssen. *Radar interferometry: data interpretation and error analysis*, pages 9–13, 33–34, 89–91. Kluwer academic publishers, 2002. ISBN 0-7923-6945-9.

R.F. Hanssen. Lecture notes in Geodesy and Natural Hazards, 2017.

E. Hartnett. NetCDF-4/HDF5 file format, 2011.

HDF. HDF5 file format specification version 3.0, 2006. URL https://support.hdfgroup.org/HDF5/doc/H5.format.html. [Online: accessed 2018-02-06].

S. Hido, H. Kashima, T. Id, and H. Matsuzawa. Unsupervised change analysis using supervised learning. *Advances in Knowledge Discovery and Data Mining*, 5012:148–159, 2008.

J. Inglada. Change Detection on SAR Images by using a Parametric Estimation of the Kullback-Leibler Divergence. *Conference: Geoscience and Remote Sensing Symposium*, 6:4104–4106, 2003.

B. Jackson, J.D. Scargle, D. Barnes, S. Arabhi, A. Alt, P. Giomousis, E. Gwin, P. Sangtrakulcharoen, L. Tan, and T. T. Tsai. An algorithm for optimal partitioning of data on an interval. *IEEE signal processing letters*, 12(2):105–108, 2005.

G. Jianya et al. A review of multi-temporal remote sensing data change detection algorithms. *The international archive of photogrammetry, remote sensing and spatial information sciences*, XXXVII-B7:757–762, 2008.

Y.-Q. Jin. A two-threshold EM-MRF algorithm for change detection of multi-temporal ERS-2 SAR images. pages 1–4, 2005.

J.A. Johannessen and Collard. F. Lecture notes in SAR instrument principles and processing, 2013. URL `https://earth.esa.int/documents/10174/950595/01_Wednesday_OCT2013_SAR_principles.pdf`.

JPL. Polarimetry. URL `https://nisar.jpl.nasa.gov/technology/polsar/`.

B.M. Kampes and R.F. Hanssen. Ambiguity resolution for permanent scatterer interferometry. *IEEE Trans. on geo. and r. and s.*, 42(11):2446–2453, 2004.

T. Kasetkasem and P.K. Varsney. An image change detection algorithm based on Markov Random Field models. pages 1–29, 2002.

Y. Kawahara, T. Yairi, and K. Machida. Change-point detection in time-series data based on subspace identification. *IEEE International Conference on Data Mining*, pages 559–564, 2007.

R. Killick. Introduction to optimal changepoint detection algorithms - useR tutorial 2017, 2017. URL `http://members.cbio.mines-paristech.fr/~thocking/change-tutorial/RK-CptWorkshop.html#changepoint-r-package`. [Online: accessed 2018-06-30].

R. Killick, P. Fearnhead, and I.A. Eckley. Optimal detection of changepoints with a linear computational cost. pages 1–24, 2012.

J. Kington. Is there an analysis speed or memory usage advantage to using HDF5 for large array storage (instead of flat binary files)? [Forum post: communicated 2014-12-30], 2014. URL `https://stackoverflow.com/questions/27710245/is-there-an-analysis-speed-or-memory-usage-advantage-to-using-hdf5-for-large-arr`.

Q. Koziol and S. Breitenfeld. Lecture notes in A brief introduction to parallel HDF5, 2015.

T.T. Le, H.T. Pham, X.T. Tran, and V.T. Pham. Change Detection in Multitemporal SAR Images Based on Statistical Similarity Measure. *International Conference on GeoInformatics for Spatial-Infrastructure*, pages 1–6, 2016.

D. Li, H. Sui, and P. Xiao. Automatic change detection of geo-spatial data from imagery. *Int. arch. of pho. and R. S.*, XXXXIV(2):245–251, 2002.

Q. Li et al. High resolution SAR change detection in Hong Kong. *IEEE Xplore*, 2011.

X. Li and J. Bethel. Image coregistration in SAR interferometry, 2008.

S.C. Liew. Effects of atmosphere, 2001. URL `https://crisp.nus.edu.sg/~research/tutorial/atmoseff.htm`. [Online: accessed 2018-05-25].

S. Liu, M. Yamada, N. Collier, and M. Sugiyama. Change-point detection in time-series data by relative density-ratio estimation. *Neural network*, 43:72–83, 2013.

P. Lombardo and C.J. Olivier. Maximum likelihood approach to the detection of changes between multitemporal SAR images. *Proc. Inst. Elect. Eng. - Radar, Sonar Navig.*, 148(4):200–210, 2001.

C. Lopez-Martinez. Speckle noise characterization and filtering in polarimetric SAR data, 2013. URL `https://earth.esa.int/documents/10174/669756/Speckle_Noise_Characterisation.pdf`. [Online: accessed 2018-28-05].

D. Lu et al. Change detection techniques. *International journal of remote sensing*, 25(12):2365–2407, 2004.

S.N. Madsen. *Speckle theory: Modelling, analysis and applications related to Syntheric Aperture Radar Data*. Technical University of Denmark, 1986.

R. Maidstone, T. Hocking, G. Rigaill, and P. Fearnhead. On optimal multiple changepoint algorithms for large data, 2014.

P. Manjusree, P. Kumar, L, C.M. Bhatt, G.S. Rao, and V. Bhanamurthy. Optimization of threshold ranges for rapid flood inundation mapping by evaluating backscatter profiles of high incidence angle SAR images. *Int. J. Disaster risk sci.*, 3(2):113–122, 2012.

C.A. Markowski and E.P. Markowski. Conditions for the effectiveness of a preliminary test of variance. *The American statician*, 44(4):322–326, 1990.

P. Mather and T. Brandt. *Classification methods for remotely sensed data*, page 213. 2001. ISBN 1420090720.

A.T. Mihai, I. Ismail, K. Lowell, O. Karyanto, and M. Santoro. Detecting and Quantifying Forest Change: The Potential of Existing C- and X-Band Radar Datasets. *PLOS ONE*, 10(6):1–14, 2015.

N. Miranda and P.J. Meadows. Radiometric calibration of S-1 level-1 products by the S-1 IPF. 2015.

D.S. Moore and G.P. McCabe. *Introduction to the practice of statistics 6th*, pages 771–774. W. H. Freeman and Company, 2003. ISBN 978-1-4292-1623-4.

G. Moser, Z. Josiane, and S.B. Serpico. SAR amplitude probability density function estimation based on a Generalized Gaussian model. *IEEE transactions on image processing*, 15(6):1429–1441, 2006.

G. Nenes and G. Tagaras. The CUSUM chart for monitoring short production runs, 2005.

NetCDF. An introduction to NetCDF, 2018a. URL `https://www.unidata.ucar.edu/software/netcdf/docs/netcdf_introduction.html`. [Online: accessed 2018-02-06].

NetCDF. Appendix B NetCDF 3 to NetCDF 4 Transition Guide, 2018b. URL `https://www.unidata.ucar.edu/software/netcdf/netcdf-4/newdocs/netcdf-c/NetCDF-3-Transition-Guide.html`. [Online: accessed 2018-02-06].

NRCAN. Radar image distortions, 2015. URL `http://www.nrcan.gc.ca/node/9325`. [Online: accessed 2018-05-22].

NRCAN. Viewing geometry and spatial resolution, 2016. URL `http://www.nrcan.gc.ca/node/9341`. [Online: accessed 2018-05-22].

E.S. Page. Continuous inspection schemes. *Biometrika*, 41:100–115, 1954.

W.D. Philpot. Philpot and philipson: Remote sensing fundamentals. pages 1–6, 2015.

M. Preiss and N. Stacy. Coherent change detection: theoretical description and experimental results, 2006.

S. Quegan et al. Multitemporal ERS SAR analysis applied to forest mapping. *IEEE Trans. Geosci. Remote sensing*, 38(2):741–753, 2000.

G. Quin et al. MIMOSA: an automatic change detection method for SAR time series. *IEEE transactions on geoscience and remote sensing*, 2013.

R.J. Radke et al. Image change detection algorithms: a systematic survey. pages 2–31, 2005.

V.C. Raykar. Scalable machine learning for massive datasets: fast summation algorithms, 2007.

J. Reeves et al. A review and comparison of changepoint detection techniques for climate data. *J. Appl. Meteorol. Climatol.*, 46(6):900–915, 2007.

R. Rew. Chunking data: Why it matters. [Blog post: communicated 2013-01-29], 2013. URL `https://www.unidata.ucar.edu/blogs/developer/entry/chunking_data_why_it_matters`.

C. Rohrbeck. Detection of changes in variance using binary segmentation and optimal partitioning, 2013.

P. Rosin and E. Ioannidis. Evaluation of global image thresholding for change detection. *Pattern recognition letters*, 24(14):2345–2356, 2002.

D. Rudolf, S. Raab, B.J. Doring, M. Jirousek, J. Reimann, and M. Schwerdt. Absolute radiometric calibration of the novel DLR Kalibri transponder, 2015.

A.J. Sccott and M. Knott. A cluster analysis for grouping means in the analysis of variance. *Biometrics*, 30(3): 507–512, 1974.

A. Schmitt, B. Wessel, and A. Roth. Curvelet-based change detection on SAR images for natural disaster mapping. *PFG - Photogrammetrie, Fernerkundung, Geoinformationi*, 6/2010:463–474, 2010.

F. Shao. Research on change information extraction of remote sensing image based on ANN, 2006.

A. Singh. Digital change detection techniques using remotely-sensed data. *Int. jour. of R. S.*, 10(6):989–1003, 1989.

D. Small. Flattening gamma: radiometric terrain correction for SAR imagery. *IEEE transactions on geoscience and remote sensing*, 49(8):3081–3093, 2011.

H. Sofiane and C. Ferdaous. Comparison of change detection indicators in SAR images, 2010.

H. Sofiane and C. Ferdaous. Multi-temporal SAR change detection and monitoring. *International archives of photogrammetry, remote sensing and spatial information sciences*, XXXIX-B7:293–298, 2012.

X. Su et al. NORCAMA: Change analysis in SAR time series by likelihood ratio change matrix clustering. *ISPRS Journal of photogrammetry and remote sensing*, pages 1–14, 2014.

D. Subbarao. Fuzzy clustering with a modified MRF energy function for change detection in Synthetic Aperture Radar images. *International Conference on Industrial Scientific Research Engineering*, 4:164–170, 2014.

M. Tan and M. Hao. Change detection by fusing advantages of threshold and clustering methods. *The international archives of the photogrammetry, remote sensing and spatial information sciences*, XLII-2/W7: 897–901, 2017.

T.L Toan. Lecture notes in Introduction to SAR Remote Sensing, 2007. URL `https://earth.esa.int/landtraining07/D1LA1-LeToan.pdf`.

C.K. Toth, M.G. Lenzano, A. Pereira, and S.A. Miranda. Revista de la Asociación Argentina de Geofísicos y Geodestas. *Geomatics in Earth Sciences*, 39(1), 2014.

UCS. UCS satellite database, 2017. URL `https://www.ucsusa.org/nuclear-weapons/space-weapons/satellite-database#.Wv6uCpdcJPY`. [Online: accessed 2018-18-05].

F.J. Van Leijen. *Persistent scatterer interferometry based on geodetic estimation theory*, pages 8–9. 2014. ISBN 978-94-6186-299-0.

F.J. Van Leijen and J.F Lopez Dekker. [Personal Communication: communicated 2018-02-02], 2018.

N. Varghese and A. Jose. Change detection in satellite images using contourlet transform and RFLICM clustering. *International journal of engineering and computer science*, 3(3):5155–5160, 2014.

G.D. Wambui, G.A. Waititu, and A. Wanjoya. The power of the pruned exact linear time (PELT) test in multiple changepoint detection. *American Journal of Theoretical and Applied Statistics*, 4(6):581–586, 2015.

B. Wang, J. Choi, S. Choi, S. Lee, P. Wu, and Y. Gao. Image fusion-based land cover change detection using multi-temporal high-resolution satellite images. *Remote Sens*, 2017(9):1–19, 2017.

K.D. Ward. Compound representation of high-resolution sea clutter. *Electron. Lett.*, 7:561–565, 1981.

Weergegevens.nl. Weergegevens. 2018. URL `http://weergegevens.nl/weatherdata.aspx?jaar=2010&maand=3&dag=26&station=344`.

M. Wieland and F. Yamazaki. Learning change from Syntheric Aperture Radar images: performance evaluation of a Support Vector Machine to detect earthquake and tsunami-induced changes. *remote sensing*, 8: 1–21, 2016.

P. Yang et al. Adaptive change detection in heart rate trend monitoring in anaesthetized children. *IEEE Trans biomed*, 53(11):2211–2219, 2006.

X. Yang and C.P. Lo. Relative radiometric normalization performance for change detection from multi-date satellite images. *Photo. Engi. and R.S.*, 66(8):967–980, 2000.

H. Zhang et al. Fast SAR image change detection using bayesian approach based on difference image and modified statistical region merging. *The scientific world journal*, 2014(1), 2014.

A

Additional figures

Appendix A holds additional figures, relevant to the research. These were excluded in the body of the thesis, as they are supplementary of nature. Firstly, the range dependency of beta nought is shown in Figure A.1. The temporal change map of the Rotterdam stack, without the removal of the outlier can be found in Figure A.2. Moreover, the position of TS1, TS2, TS3 and TS4 in the Houston stack are shown, accompanied by an impression of the sensed area, in Figure A.3. Lastly, Figures A.4-A.6 are the segmentation results of the quantile normalized time series TS2, TS3 and TS4. These are used to determine the sensitivity of ANOVA and PELT to non-Normally distributed time series.

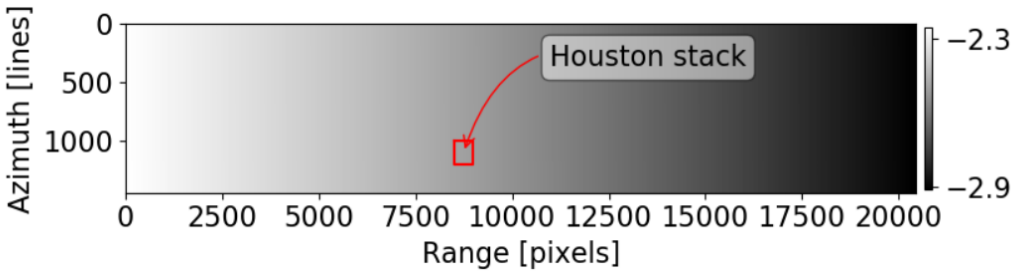


Figure A.1: The difference between sigma nought and beta nought calibrated SAR images, showing the range dependency of beta nought. A ~0.6 dB difference between near and far range for beta nought is observed for the entire Sentinel-1 IW swath.

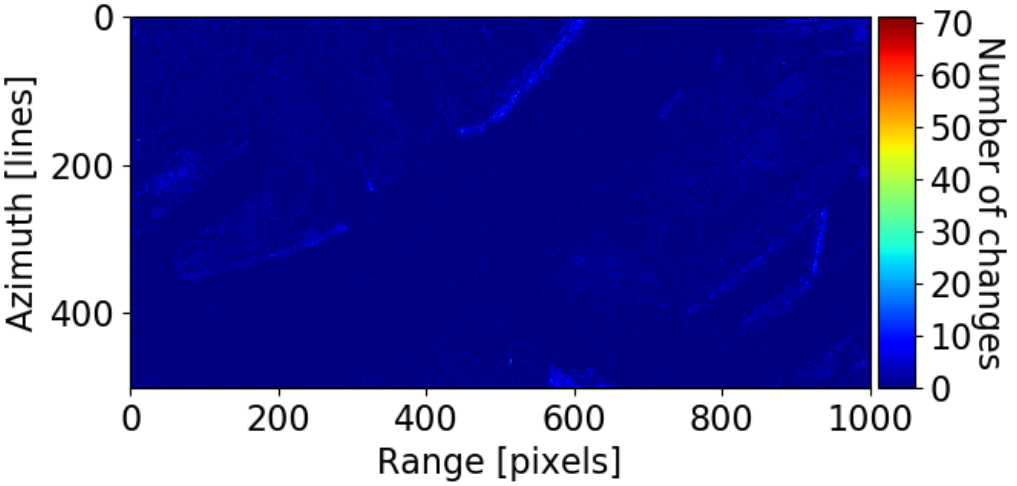


Figure A.2: Temporal change map using PELT [σ = 10] on the Rotterdam stack, without the single outlier removed.

(a) Impression of ground area TS1.

(b) Impression of ground area TS2.

(c) Location of TS1.

(d) Location of TS2.

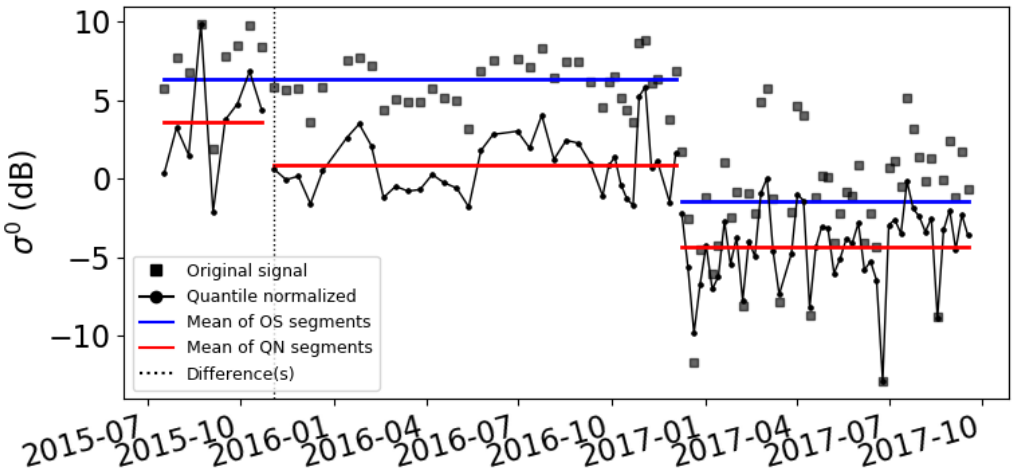
(e) Impression of ground area TS3.

(f) Impression of ground area TS4.

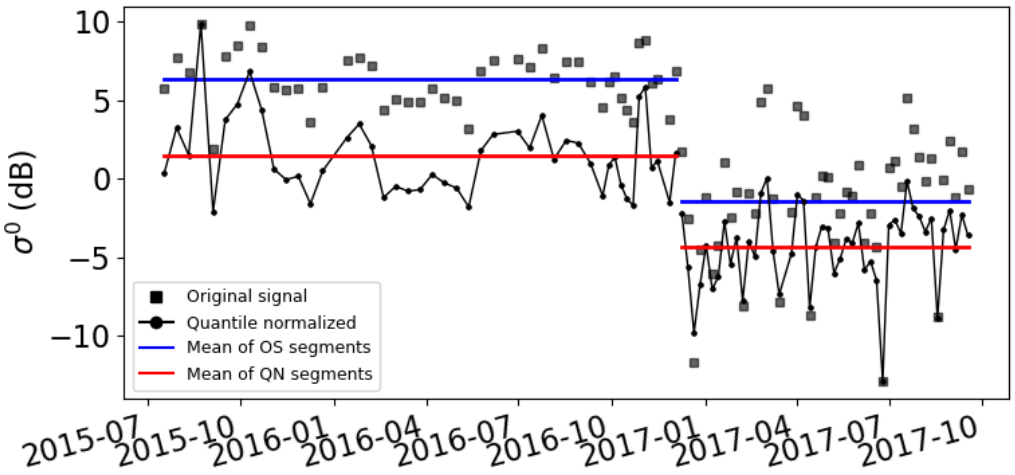
(g) Location of TS3.

(h) Location of TS4.

Figure A.3: Impressions of the sensed areas, as seen in time series TS1 (Figure 5.8), TS2 (Figure 5.10), TS3 (Figure 5.11) and TS4 (Figure 5.12).

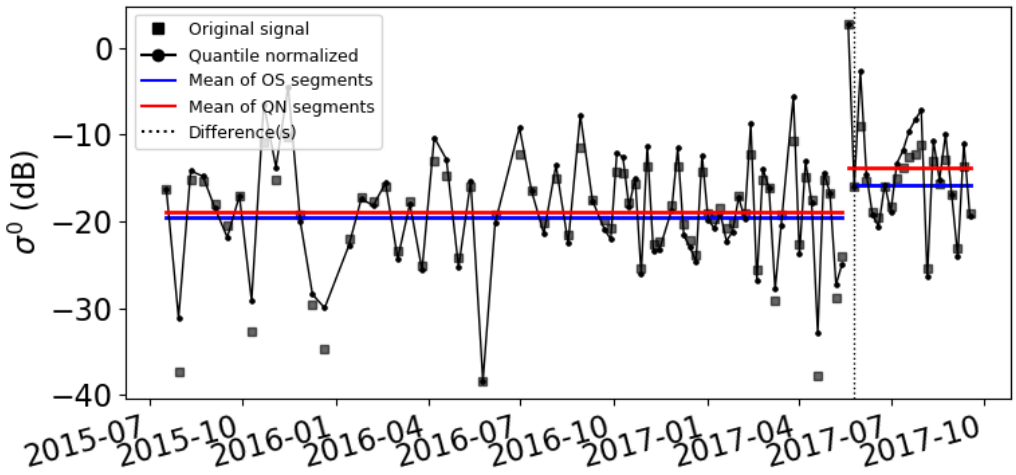


(a) ANOVA [$\alpha = 0.005$, $D = 3$].

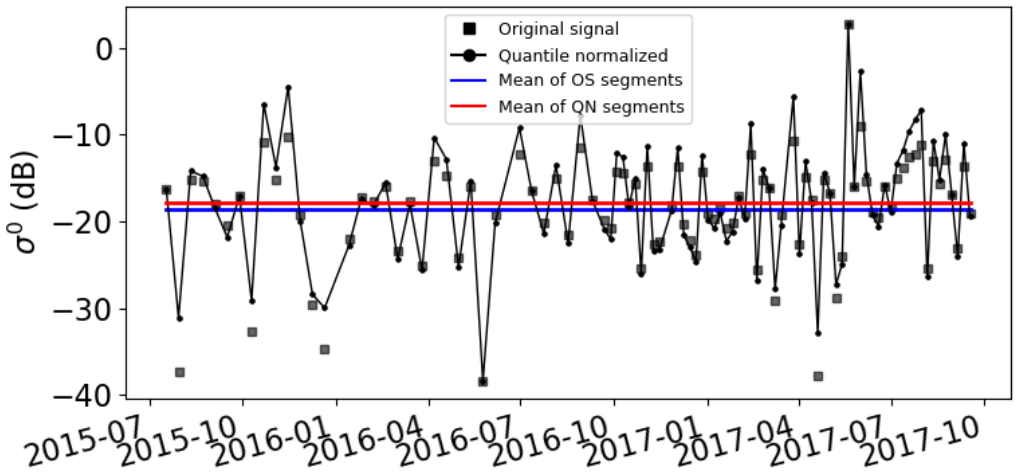


(b) PELT [$\sigma = 10$].

Figure A.4: Sensitivity to non-normality of ANOVA and PELT following the quantile normalized time series TS1 from the Houston stack, as shown in Figure 5.8 [Black squares: data points of original SAR time series, Black solid line: the quantile normalized time series, Blue solid line: mean of associated segments original time series, Red solid line: mean of associated segment quantile normalized series, Dotted line: differences between segmentations original and quantile normalized series].

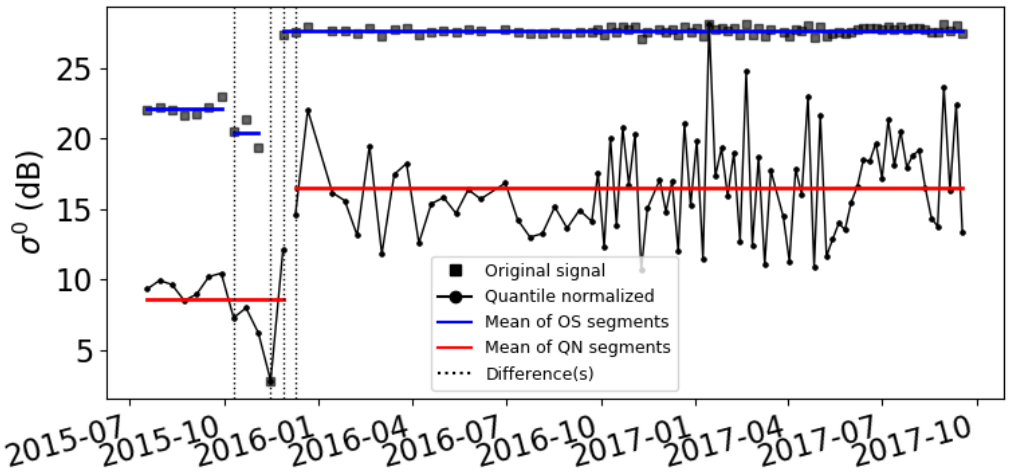


(a) ANOVA [$\alpha = 0.005$, $D = 3$].

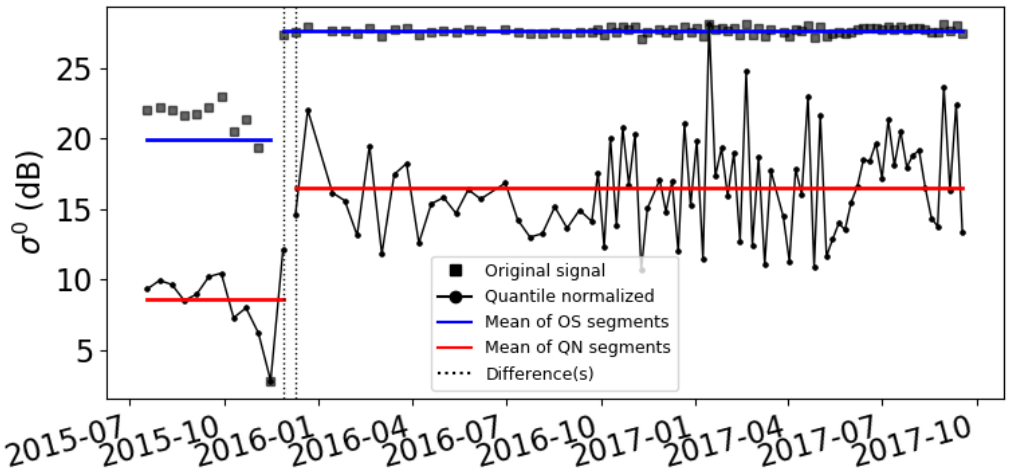


(b) PELT [$\sigma = 10$].

Figure A.5: Sensitivity to non-normality of ANOVA and PELT following the quantile normalized time series TS3 from the Houston stack, as shown in Figure 5.11 [Black squares: data points of original SAR time series, Black solid line: the quantile normalized time series, Blue solid line: mean of associated segments original time series, Red solid line: mean of associated segment quantile normalized series, Dotted line: differences between segmentations original and quantile normalized series].



(a) ANOVA [$\alpha = 0.005$, $D = 3$].



(b) PELT [$\sigma = 10$].

Figure A.6: Sensitivity to non-normality of ANOVA and PELT following the quantile normalized time series TS4 from the Houston stack, as shown in Figure 5.12 [Black squares: data points of original SAR time series, Black solid line: the quantile normalized time series, Blue solid line: mean of associated segments original time series, Red solid line: mean of associated segment quantile normalized series, Dotted line: differences between segmentations original and quantile normalized series].

B

Elaborations equations and threshold settings ROC curves

Appendix B consists of three items, some intermediate steps in the equation to calculate the F-values for ANOVA (Equation (4.11)), the least squares approach to estimate the coefficients of the local incidence plane for TerraSAR-X calibration written in matrix and vector notation (Equation (4.3)), and the used thresholds for creating the ROC curves of the CPD algorithms applied on the simulated data (Figure 5.7).

Elaboration Equation (4.11)**:**

$$DF_{bg} = K - 1 \tag{B.1a}$$

$$SS_{bg} = \sum_{i=1}^{K} \left(N\left(\mu_i - \mu\right)\right)^2 \tag{B.1b}$$

$$MS_{bg} = \frac{SS_{bg}}{DF_{bg}} \tag{B.1c}$$

$$DF_{wg} = N - K \tag{B.1d}$$

$$SS_{wg} = \sum_{i=1}^{K} \sum_{j=1}^{N} \left(A_{ij} - \mu_i\right)^2 \tag{B.1e}$$

$$MS_{wg} = \frac{SS_{wg}}{DF_{wg}} \tag{B.1f}$$

Matrix and vector notation Equation (4.3)**:**

$$\begin{bmatrix} x_1 & x_2 & \dots & x_n \\ y_1 & y_2 & \dots & y_n \\ 1 & 1 & \dots & 1 \end{bmatrix} \begin{bmatrix} x_1 & y_1 & 1 \\ x_2 & y_2 & 1 \\ \dots & \dots & \dots \\ x_n & y_n & 1 \end{bmatrix} \begin{bmatrix} a \\ b \\ d \end{bmatrix} = \begin{bmatrix} x_1 & x_2 & \dots & x_n \\ y_1 & y_2 & \dots & y_n \\ 1 & 1 & \dots & 1 \end{bmatrix} \begin{bmatrix} -z_1 \\ -z_2 \\ \dots \\ -z_n \end{bmatrix}. \tag{B.2}$$

Thresholds Figure 5.7:

- Differencing: $\tau \in [1 \times 10^{-1}, 5 \times 10^{-1}, 1.0, 1.5, 2.0, 2.5, 3.0, 3.5, 4.0, 4.5, 5.0, 6.0, 7.0, 8.0, 9.0]$ and $D = 1$;
- CUSUM: $\tau \in [1 \times 10^{-1}, 5 \times 10^{-1}, 1.0, 1.5, 2.0, 2.5, 3.0, 3.5, 4.0, 4.5, 5.0, 6.0, 7.0, 8.0, 9.0]$ and $\nu \in [5 \times 10^{-2}, 2.5 \times 10^{-1}, 5 \times 10^{-1}, 7.5 \times 10^{-1}, 1.0, 1.25, 1.5, 1.75, 2.0, 2.25, 2.5, 3.0, 3.5, 4.0, 4.5]$;
- ANOVA: $\alpha \in [1 \times 10^{-5}, 2.5 \times 10^{-5}, 5 \times 10^{-5}, 1 \times 10^{-4}, 2.5 \times 10^{-4}, 5 \times 10^{-3}, 1 \times 10^{-3}, 2.5 \times 10^{-3}, 5 \times 10^{-2}, 1 \times 10^{-2}, 2.5 \times 10^{-2}, 5 \times 10^{-2}, 1 \times 10^{-1}, 2.5 \times 10^{-1}, 5 \times 10^{-1}]$ and $D = 1$; and
- PELT: $\sigma \in [1 \times 10^{-1}, 5 \times 10^{-1}, 1.0, 1.5, 2.0, 2.5, 3.0, 3.5, 4.0, 4.5, 5.0, 6.0, 7.0, 8.0, 9.0]$,

C

Overview of implemented Python scripts

Appendix C gives a schematic overview of the scripts used throughout the change detection procedure. These are all written in the high-level programming language Python, mostly in Python 3 and some in Python 2. A complete overview can be found in Table C.1, where a total of 6 sections are defined. These are the *utilities*, *calibration*, *HDF5 storage*, *change detection*, *filtering* and *visualization* scripts. The data was processed in a modular way, carefully examining each processing step with the various visualization tools. This means that no master script is present to execute the entire CPD procedure in one batch. However, a master script for the absolute calibration is created.

Table C.1: Implemented Python scripts for the absolute calibration, filtering, HDF5 storage, change detection and visualization of SAR-based data.

Category	File name	Description
Utilities	read.py	Select and read various file formats used in the absolute calibration of SAR data; .res (Doris log file), .raw (SLC data), .xml (calibration parameters) and .txt (absolute calibration log file).
	write.py	Write absolutely calibrated files to binary files (.raw) as (Re,Im), preserving phase information, alongside a log file (.txt).
	utils.py	Utility functions used across all categories. Notable functions: linear_LS_plane (TSX LIA), moving_average (temporal filter), rolling_mean (spatial filter) and quantile_normalize.
	h5.py	Utility functions for HDF5 storage. Notable functions: getdatasets (generator of absolute paths to data sets), filter (generator of absolute paths based on 1 or more regex input) and makegroups (make tree-structed HDF5 groups based on 1 or more keywords)
Calibration	calibrate.py	Calibrates 1 Sentinel-1 or TerraSAR-X image and outputs in binary format; dependency on reads.py and utils.py
	main_calibrate.py	Interactive script that takes user defined input to calibrate a stack of Sentinel-1 or TerraSAR-X images; dependency on read.py, write.py and utils.py
	main_stack.py	Stack, select an optional subset and store multiple SAR images in binary format (.raw); dependency on read.py.
HDF5 storage	hdf5_p2.py/ hdf5_p3.py	Stack, select an optional subset and store in HDF5 format (.h5); dependency on h5.py and read.py.
Change detection	differencing.py	Functions for Differencing algorithm.
	cusum.py	Functions for CUSUM algorithm.
	anova.py	Functions for ANOVA algorithm.
	pelt.py	Functions for PELT algorithm.
	main_differencing.py/ main_cusum.py/ main_anova.py/ main_pelt.py	Master scripts for the change detection algorithms; outputs a log file with the number of detected changes per image/ epoch (.txt), a pickle file with the indexes of the detected changes per time series (.pkl) and a hash table pickle file with the line-pixel combinations of all changes per epoch (.pkl).
	sensitivity_tests.py	Functions and tests environments to test the change detection algorithms' sensitivity to various conditions (e.g. Dirac pulse); dependency on h5.py, utils.py, differencing,py, cusum,py, anova.py, pelt.py.
	simulated_tests.py	Functions and tests environments to test the change detection algorithms with simulated data (e.g. Gaussian stack); dependency on h5.py, utils.py, differencing,py, cusum,py, anova.py, pelt.py.
	signal_qualitymetrics.py	Compute quality metrics of change detection algorithms on simulated data; dependency on h5.py, utils.py, differencing,py, cusum,py, anova.py, pelt.py.
Filtering	filter.py	Master script for temporal and/or spatial filtering of a stack of images.
Visualization	Interactive.py	Creates an environment for data and change detection result visualization and interaction; dependency on h5.py, utils.py, differencing,py, cusum,py, anova.py, pelt.py.
	Interactive_select.py	Interactive lasso selection and storage of subsets of data.
	Vis_*.py	Various visualization scripts, from change maps to flowcharts.