

RAM

● ROBOTICS
AND
MECHATRONICS

Towards feature-based underground void detection with Ground Penetrating Radar from within sewers using Image Processing

M. (Matthijs) van Delft

MSC ASSIGNMENT

Committee:

H. Noshahri, MSc
dr. ir. E. Dertien
J.A. Hempenius, MSc
dr. ir. L.L. Scholtenhuis
dr. ir. J.F. Broenink

October, 2019

046RaM2019
Robotics and Mechatronics
EEMathCS
University of Twente
P.O. Box 217
7500 AE Enschede
The Netherlands

UNIVERSITY
OF TWENTE.

TECHMED
CENTRE

UNIVERSITY
OF TWENTE.

DIGITAL SOCIETY
INSTITUTE

Contents

| | |
|--|-----------|
| Acronyms | iv |
| 1 Introduction | 7 |
| 1.1 Problem context | 7 |
| 1.2 Problem statement | 10 |
| 1.3 Research questions | 11 |
| 1.4 Approach | 11 |
| 1.5 Outline | 12 |
| 2 Background | 13 |
| 2.1 Ground Penetrating Radar | 13 |
| 2.2 Ground Penetrating Radar simulation software | 15 |
| 3 Analysis | 17 |
| 3.1 Feasibility study on gprMax | 17 |
| 3.2 Void characterisation | 18 |
| 3.3 Wave propagation inside a sewer | 19 |
| 4 Design and Implementation | 23 |
| 4.1 Radargram acquisition | 23 |
| 4.2 Pre-processing | 26 |
| 4.3 Label radargrams | 28 |
| 4.4 Feature extraction | 31 |
| 4.5 Label observations | 38 |
| 4.6 Feature set | 39 |
| 4.7 Classification model | 40 |
| 5 Results and discussion | 43 |
| 5.1 Test 1: experimental versus simulation radargram results | 43 |
| 5.2 Test 2: feature selection on planar and in sewer model | 46 |
| 5.3 Test 3: detection accuracy by feature selection | 51 |
| 6 Conclusion and Recommendations | 54 |
| 6.1 Conclusion | 54 |
| 6.2 Recommendations | 56 |
| Bibliography | 59 |
| A Appendix 1 - Code repository read me | 62 |
| B Appendix 2 - Experimental setup test log | 67 |
| C Appendix 3 - Feasibility study gprMax | 69 |
| D Appendix 4 - Test 1 additional results | 81 |
| E Appendix 5 - Test 2 additional results | 85 |
| F Appendix 6 - Test 3 additional results | 87 |

Acronyms

| | |
|----------------|--|
| BRISK | Binary Robust Invariant Scalable Keypoints. 32 |
| CCTV | Closed-Circuit Television. 9 |
| CLAHE | Contrast-Limited Adaptive-Histogram Equalisation. 27, 42 |
| CMP | Common Midpoint. 19, 21, 28 |
| CPU | Central Processing Unit. 15 |
| ECOC | Error-Correcting Output Codes. 42 |
| ERT | Expected Reverberation Time. 21 |
| FAST | Features from Accelerated Segment Test. 32 |
| FDTD | Finite-Difference Time-Domain. 15, 57 |
| GPR | Ground Penetrating Radar. 1, 3–6, 9–17, 19, 20, 22–24, 28, 42, 44–46, 50, 54–57, 67, 72 |
| GPU | Graphics Processing Unit. 15, 24 |
| HPC | High-Performance Computing. 2, 15, 24, 25, 71 |
| IMMSE | Image Mean Squared Error. 69–76, 80 |
| KCV | k -fold Cross Validations. 40, 41 |
| LBFGS | Limited memory Broyden-Fletcher-Goldfarb-Shanno. 39 |
| ML | Machine Learning. 11, 12, 23, 38, 42, 56, 57 |
| MSER | Maximum Stable Extremal Regions. 31–35, 42, 46, 57 |
| NCFS | Neighbourhood Component Feature Selection. 39, 48 |
| ORB | Oriented FAST and Rotated BRIEF. 32 |
| PNG | Portable Network Graphics. 23, 25, 28 |
| PSNR | Peak Signal to Noise Ratio. 69–76, 80 |
| RaM | Robotics and Mechatronics. 2, 7, 62 |
| SGD | Stochastic Gradient Descent. 39 |
| SSET | Sewer Scanning and Evaluation Technology. 9 |
| SSIM | Structural Similarity Index. 69, 70 |
| SURF | Speed Up Robust Features. 32 |
| SVM | Support Vector Machine. 41, 42, 51, 54, 57 |
| TISCALI | Technology Innovation for Sewer Condition Assessment – Long-distance-Information-system. 3, 5, 7 |
| TWTT | Two-Way Travel Time. 13, 14, 18, 21, 28, 29, 35, 46 |
| TZC | Time Zero Correction. 21, 26, 28, 29 |

Abstract

Non-invasive inspection techniques become more important for the rehabilitation of underground utilities. This research focusses on in-sewer Ground Penetrating Radar (GPR) to detect voids behind the sewer walls. Detection of small voids facilitates for maintenance to avert collapsing sewers and roads as the result of the growing voids. The comparison between ground surface GPR (planar) and in-sewer GPR (cylindrical) , helps to give insight into the effects of both topologies on simulated GPR radargrams. Image processing techniques are used to extract void characteristics as features, which are used for feature selection and the calculation of the classification accuracy of voids. The resulting feature set proves to contain similar features and classification accuracies for both the planar and cylindrical topology. Hence, the closed-off cylindrical shape of the sewer does not pose an issue in the classification of voids behind the sewer wall using the extracted features. Therefore, based on the simulations, the in-sewer GPR proves to have potential in void detection. More research regarding the application and environment is necessary to determine the potential in real-life circumstances.

Acknowledgement

Throughout the research, I have been lucky to receive support on a professional and personal level.

I would first like to thank my supervisors, H. Noshahri, J.A. Franco Hempenius, dr. ir. E.C. Dertien, dr. ir. L.L. Olde Scholtenhuis and dr. ir. J.F. Broenink for sharing invaluable knowledge and time with me. The weekly meetings were very convenient, and they made me part of the greater sense of the project, which has been a great motivation.

Also, I would like to thank the staff of the RaM department at the University of Twente for sharing their knowledge and resources during the research. Additionally, I would like to thank members of staff working on the HPC cluster of the University of Twente. Without the availability of this resource, it would not be possible to execute the test as extensively as included in the research.

I would also like to thank my friends and fellow students for bearing with me on the grumpy days. The peer reviews on the concepts meant a great deal, and the hours of distraction even more. The numerous reminders to relax kept me mentally sharp in times of need.

Besides, I would like to thank my parents for always being there for me when I need it. You made me who I am today, and are a great example of humble, hard-working people. Although you might not always understand what I do, your listening ear takes away a huge load of my shoulders.

Matthijs van Delft, 3-10-2019 Enschede, The Netherlands

Summary

An enormous challenge in the 21st century is the maintenance and rehabilitation of underground utilities. Nowadays, subsurface environments are generally evaluated by excavation, followed by manual inspection. The process of excavation leads to a higher probability of damage and nuisance to residents. This thesis will examine an alternative to excavation to reduce additional damage and nuisance.

This research is part of the Technology Innovation for Sewer Condition Assessment – Long-distance-Information-system (TISCALI) project. A part of TISCALI researches the use of non-invasive techniques from within the underground utilities. Ground Penetrating Radar (GPR) has outstanding penetrative abilities and a lot of work and research exists with GPR from the ground surface. Changing the ground surface GPR approach to GPR from within a sewer, reduces the necessity for excavation, and gets the GPR closer to the object behind the sewer wall. Decreasing the distance to object under inspection leads to higher resolution measurements and results in a more accurate inspection.

Collapsing sewers are one of the issues in the Dutch subsurface infrastructure. A cause for sewer collapse is the natural deterioration and initial damage to the sewer, leading to the intrusion of the surrounding soil resulting in growing voids behind the sewer wall. Ultimately, these voids might lead to a collapse of the sewer or the road above.

Voids and objects behind the sewer wall create reflections in GPR radargrams. An algorithm detects the reflection from the radargram and calculates the features from the reflections. A classification model classifies reflections, e.g., a void, another subsurface object or nothing. This research aims to find a set of features able to classify the reflections as accurately as possible from a radargram acquired by an in-sewer GPR.

GPR radargram acquisition in an experimental setup requires a significant amount of time, and is considered unfeasible for this research. Instead, GPR simulation software is used to design and simulate GPR in a sewer and calculate the corresponding radargram. A comparison between experimental radargrams and approximated simulated radargrams gives insight into the potential of simulating experimental setups. The sewer is simulated as a flat object and as a cylindrical object. Consider this a cylindrical to planar transformation. Studying the difference in wave propagation of both topologies helps to get insight into possible artefacts in radargrams when using GPR in a sewer. Two encountered artefacts in the cylindrical topology are a reverberation pattern and a longer reflection path through the sewer wall. The reverberation pattern leads to additional reflections band in the radargram.

The GPR radargrams contains reflections of different shapes and intensities. To be able to classify the voids, it is vital to get a better understanding of how voids differ from other objects behind the sewer wall. The feature extraction algorithms extract the region features and the hyperbola features from regions in radargrams. The resulting features are useful to determine the most optimal subset of features to classify voids behind the sewer wall for planar and cylindrical topologies. The results show that roughly half of the total number of features are in the selected feature set. The selected features lead to an almost identical classification accuracy compared to all features combined. Therefore, the selected feature set is capable of capturing the characteristics of voids while reducing the number of features.

A dummy classification model calculates the classification accuracy of the features for both topologies. The correct classification results of the planar topology (72%) and cylindrical topology (78%) differ slightly. Interestingly, the accuracy is higher for the cylindrical topology, proving that the reverberation pattern does not have a negative influence on the classification results. It seems that the cylindrical topology has a slightly better classification accuracy because of more distinctiveness in reflection. The reflections in the cylindrical topology are perceived differently because of the non-linear GPR movement, which results in steeper reflections and thus less overlap and interference.

Although the classification accuracies for the topologies are near-identical, the feature sets for the topologies differ quite significantly. The planar topology depends evenly on region and hyperbola features, but the cylindrical topology depends for 70% on regions features. Therefore, it seems that the reverberation pattern in the cylindrical topology introduces hyperbolic noise which is countered by depending more on the region features than on the hyperbola features.

Samenvatting

Een enorme uitdaging in de 21ste eeuw is het onderhoud en de renovatie van ondergrondse utiliteiten. Tegenwoordig worden ondergrondse omgevingen meestal geëvalueerd doormiddel van afgravingen gevolgd door manuele inspectie. Afgraven leidt tot een hogere kans van schade of verstoring van omwonende. Deze scriptie onderzoekt een alternatief van afgraving om de kans op schade en verstoring te verminderen.

Dit onderzoek is deel van het Technology Innovation for Sewer Condition Assessment – Long-distance-Information-system (TISCALI) project. Een deel van TISCALI onderzoekt het gebruik van niet-invasieve technieken vanuit ondergrondse utiliteiten. Ground Penetrating Radar (GPR) heeft uitstekende penetratieve eigenschappen en er is veel werk en onderzoek beschikbaar omtrent het gebruik van GPR op grondoppervlak. Overstappen van het gebruik op grondoppervlak naar het gebruik van GPR in een riool reduceert de behoefte aan afgraving en positioneert de GPR dicht bij de objecten achter de rioolwand. Het verminderen van de afstand tot de objecten onder inspectie leidt tot metingen met hogere resolutie en resulteert in een secuurdere inspectie.

Instortende rioolpijpen behoren tot één van de hoofdproblemen in de ondergrondse infrastructuur van Nederland. Één van de oorzaken is natuurlijke slijtage en initiële schade aan de rioolpijp, wat leidt tot indringing van omliggende grond en tot groeiende lege ruimtes achter de rioolwand. Uiteindelijk kunnen deze lege ruimtes resulteren tot het instorten van de rioolpijp of de weg erboven.

Lege ruimtes en objecten achter de rioolwand creëren reflecties in de GPR radargrammen. Een algoritme detecteert de reflecties vanuit de radargram en berekend de eigenschappen van de reflectie. Een classificatiemodel voorspelt de klasse van de reflectie, bijvoorbeeld, een lege ruimte, een ander object of niks. Dit onderzoek heeft als doel om een set aan eigenschappen te vinden met een zo hoog mogelijke precisie in classificeren van radargram reflectie gemeten vanuit een riool.

Radargram acquisitie in een experimentele opstelling kost veel tijd en wordt beschouwd als onpraktisch voor dit onderzoek. In plaats daarvan wordt GPR simulatie software gebruikt voor het simuleren van een vooraf ontworpen omgeving en het berekenen van de bijbehorende radargram. Een vergelijking tussen experimentele radargrammen en gesimuleerde radargrammen geeft inzicht in de potentie van het simuleren van experimentele opstellingen. De rioolbuis is gesimuleerd als een vlak object en als een cilindrisch object. Zie dit als een planair naar cilindrisch transformatie. Het bestuderen van het verschil in golf propagatie in beide topologieën helpt inzicht te verkrijgen in het ontstaan van mogelijke artefacten in radargrammen gemeten vanuit een riool. Twee resulteerde artefacten in de cilindrische topologie zijn een weerkaatsingspatroon en een langer signaal pad door de rioolwand. Het weerkaatsingspatroon leidt tot bijkomende reflectie in de onderste helft van de radargram.

De radargrammen van de GPR bevatten reflectie in verschillende vormen en intensiteiten. Voor het voorspellen van lege ruimtes is het van belang om inzicht te krijgen hoe lege ruimtes verschillen van andere ondergrondse objecten achter een rioolwand. Het eigenschap extractie algoritme berekend de regio eigenschappen en hyperbool eigenschappen uit de regio's in een radargram. De resulterende eigenschappen zijn nuttig voor het bepalen van een subset van

eigenschappen geschikt voor het voorspellen van lege ruimtes voor de planair en cilindrische topologie. Ongeveer de helft van het totaal aantal eigenschappen zijn deel van de geselecteerde eigenschappen. De set met geselecteerde eigenschappen voorspelt met een bijna identieke voorspellingsnauwkeurigheid in vergelijking tot een voorspelling met alle eigenschappen. Daarom zijn de geselecteerde eigenschappen geschikt voor het beschrijven van de kenmerken van lege ruimtes terwijl het de hoeveelheid benodigde eigenschappen verminderd.

Een proef classificatie model berekend de classificatie nauwkeurigheid van de eigenschappen voor beide topologieën. De correcte classificatie resultaten voor de planair topologie (72%) en cilindrische topologie (78%) verschillen nauwelijks. Interessant is dat de nauwkeurigheid hoger is voor de cilindrische topologie, wat bewijst dat het weerkaatsingspatroon geen negatieve invloed heeft op de classificatie nauwkeurigheid. Het lijkt dat de cilindrische topologie een betere nauwkeurigheid heeft omdat de reflecties beter te onderscheiden zijn. De reflecties in de cilindrische topologie worden anders waargenomen door het non-lineaire GPR pad, wat resulteert in steilere reflecties en dus minder overlap en interferentie.

Desondanks de classificatie resultaten van beide topologieën bijna identiek zijn, zit er verschil in de eigenschappen van de geselecteerde sets. De planair topologie is voor 70% afhankelijk van regio eigenschappen. Daarom lijkt het erop dat het weerkaatsingspatroon hyperbolische ruis introduceert in de cilindrische topologie wat afgevangen wordt door meer afhankelijk te worden op regio features.

1 Introduction

This research comes forth out of the Technology Innovation for Sewer Condition Assessment – Long-distance-Information-system (TISCALI) project¹ from the Robotics and Mechatronics (RaM) department of the University of Twente, Enschede, The Netherlands. This chapter gives insight into the importance of the research topic. It explains the context of the problem. Subsequently, it describes the research problem and presents the research questions to tackle individual parts of the problem. Finally, the approach helps to understand the choices made in this research.

1.1 Problem context

An enormous challenge in the urbanising world of the 21st century is the proper maintenance and rehabilitation of underground utilities. A growing infrastructure results in an increased probability of defects relating to underground utilities in areas where it is not feasible to inspect with invasive techniques. High environmental complexity in, e.g., old city centres and crowded intersections, make for difficult inspection using excavation. Excavation in such places is a nuisance for residents, but might also lead to dangerous working conditions and prolonged deadlines. Excavation is generally unwanted if not necessary. Therefore, TISCALI proposes inspection from within underground utilities.

Lack of accurate and sufficient information about large multi-layer underground network has made localising and assessing the condition of any subsurface pipe or cable system difficult. RIONED foundation² is the interest group for urban drainage concerns in the Netherlands. The RIONED database is consulted to create a better understanding of possible sewer defects. The reports and statistics gathered by RIONED describe two main reasons for sewer damage: excavation damage and natural deterioration.

Excavation damage caused by human interference, mainly with excavators, come forth out of the rapid changing subsurface infrastructure (van der Werf et al., 2017). Excavation damage generally has three main causes. First, incorrect or incomplete information of the environment and the location of cables, pipes, and sewers results in an increased probability of excavation damage. Secondly, the primary drainage often lays parallel to the grid of cables and pipes, and the sewer system is in the middle of the road, which makes them extra vulnerable. Thirdly, drainage replacement projects are often happening within city centres and old neighbourhoods, where the composition of subsurface objects is generally more complex.

Besides the damage caused by human interference, different natural factors influence the conditions of the sewer as well. Natural deterioration can weaken the structural integrity of the sewers to a point where it breaks down or needs maintenance. Distinctive features in and around the sewer, such as cracks, voids, scours, chemical reactions, tree roots, and groundwater levels, influence the deterioration process. Consequently, it is hard to predict when a sewer will collapse. However, different initial defects lead to a loss of support from the surrounding soil and thus the creation of voids (Davies et al., 2001). Therefore, the voids can be considered a potential sign of the probability of collapse, no matter the initial source of the damage.

¹<https://www.utwente.nl/en/tiscali/proposal/> (URL accessed on 28-10-2019)

²<https://www.riool.net/english> (URL accessed on 28-10-2019)

A sewer collapse is the collapse of the sewer, the road above the sewer or both. A sewer collapse often consists of three stages depicted in figure 1.1. Firstly, initial damage occurs by human interference or natural deterioration, which eventually evolves to a point where material intrudes the sewer. Secondly, support material around the sewer washes away, resulting in a void filled with water or air. Thirdly, the collapse is often triggered by a random event that is not related to the cause of the initial damage. Ideally, sewer inspection should be able to detect small voids directly behind a sewer wall. Early maintenance at the occurrence of voids help prevents collapsing sewers.

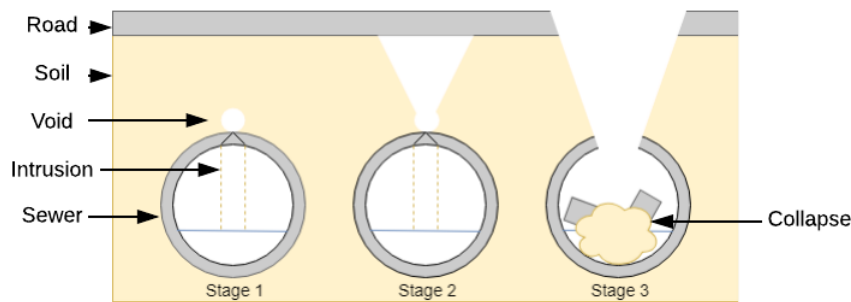


Figure 1.1: The three stages of a sewer collapse visualised by an underground sewer cross-section.

RIONED performed a national questionnaire in which municipalities reported their sewer collapses between 2008 and 2018 to gain insight into sewer collapses. In the questionnaire, the definition of a sewer collapse restricts to sewer pipes with a minimum diameter of 200mm, house connection culverts excluded. Of all cooperating municipalities (47%), a significant amount (81%) reported a collapse within the last ten years. From all reported sewer collapse, the main cause (19.4%) is inferior conditions of the sewer, followed by issues with pressure sewers (18.2%). The sewer diameter and material correlate with the number of collapses. More than 80% of the reported collapses are concrete sewers. Since the Dutch sewage infrastructure consists of for 70% of concrete sewers, it seems assumable that these percentage of collapsed concrete sewer reflect the percentage of existing concrete sewers. The most common diameters of collapsed sewers are between 300mm and 500mm, which correlates with the diameters of existing sewer in The Netherlands in 2013, see figures 1.2 and 1.3. The finding suggests that it is most likely for a sewer with a diameter between 300-500mm to collapse.

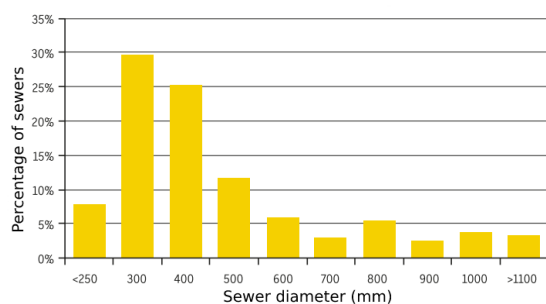


Figure 1.2: Histogram of collapsed sewer in the Netherlands in 2018 (Wonink and van der Werf, 2019).

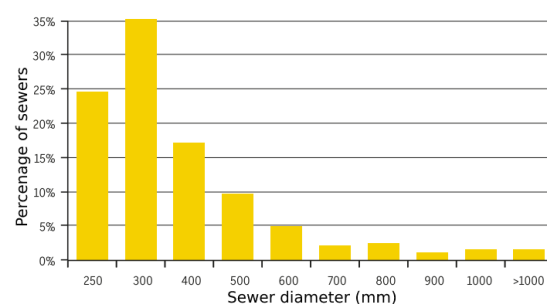


Figure 1.3: Histogram of existing sewer in the Netherlands in 2013 (Wonink and van der Werf, 2019).

The thickness of the sewer wall grows with the diameter. De Hamer Beton B.V. is a Dutch manufacturer of sewers. Their specifications³ helps to create an expectation of likely sewer wall

³<https://www.dehamer.nl/producten/buizen/ronde-buizen/300/hapro-300-x-2400/> (URL accessed on 28-10-2019)

| Sewer inner diameter c | Sewer wall thickness w |
|--------------------------|--------------------------|
| 300mm | 55mm |
| 400mm | 55mm |
| 500mm | 65mm |

Table 1.1: Inner sewer diameters c with the corresponding thickness w .

thickness for specific inner diameters. Table 1.1 includes inner sewer diameters and sewer wall thickness data. Note that the table includes only sewer diameters of 300-500mm since these are statistically most vulnerable to collapse. The research continues with a sewer wall thickness around 55mm.

An additional property of sewers is that some include reinforced steel rebar in the concrete. Non-reinforced sewer are generally produced with 300-1000mm diameter where reinforced sewer allows for 300-3500mm diameter⁴. The research focusses on non-reinforced sewers although the diameter of interest 300-500mm could be either non-reinforced or reinforced. Steel rebar in the concrete lead to strong reflection in the radargram, which interferes with reflections from objects behind the sewer wall (Cassidy et al., 2011). This study aims at sewers without rebar to create a stable base for future research in which more complexity, such as rebar, may be introduced.

A non-invasive penetrative technique is essential to detect small voids behind the sewer wall and reduce the need for excavation. Koo and Ariaratnam (2006) summarises multiple non-invasive inspection techniques: Closed-Circuit Television (CCTV), Sewer Scanning and Evaluation Technology (SSET), sonar systems, laser scanning systems, and GPR. CCTV, SSET and laser scanning systems lack penetrative power, which makes them only feasible for inspection of the inner surface of the sewer wall. GPR its outstanding features is its penetrative ability which outperforms sonar techniques. Depending on the frequency of the antennas and the properties of inspected material, GPR can measure the reflections up to 50m deep⁵. Penetrative power can obtain information about the properties of the object behind other objects. The information is useful to assume the material, dimensions, and location of objects behind the sewer wall. Therefore, this research uses GPR to acquire data of the subsurface environments.

Ideally, an automatic algorithm performs the classification of the reflections in the GPR radargrams. The advantage of an algorithm is that it is objective, which makes it more consistent in its results. In contrast, an inspection specialist might unknowingly be subjective. Subjectivity in an assessment might result in inconsistent evaluation results. The introduced inconsistency might lead to wrong decisions relating maintenance of the sewage systems and thus increase the risk of defects. Therefore, it is vital to reduce the subjectivity of the assessment procedure as much as possible.

⁴<https://febe.be/frontend/files/userfiles/files/Andere%20Publicaties/publication-techniques/15%20goede%20redenen%20om%20te%20kiezen%20voor%20rioleringsystemen%20in%20prefab%20beton.pdf> (URL accessed on 28-10-2019)

⁵<https://www.sensoft.ca/wp-content/uploads/2015/11/pulseEKKO-Users-Guide-2005-00040-08.pdf> (URL accessed on 28-10-2019)

1.2 Problem statement

The potential of algorithms detecting voids located directly behind the sewer wall using GPR from within the sewer is yet unknown. This research aims to extract features from GPR radargrams that can describe a void located directly behind the sewer wall. Knowing which features capture the characterisation of voids is useful in selecting a classification model to capture similar objects or properties. Additionally, the features are useful as input for certain applicable classification models. Extracting a robust set of features contributes to future research in the field of detecting sewer defects. Therefore, the main research result will be a feature set with the highest theoretical potential in detecting voids directly behind the sewer wall using in-sewer GPR and the proposed algorithm.

The research problem is split up in multiple subproblems which lead the research questions in section 1.3.

1. Detecting voids behind a sewer wall using in-sewer GPR is a particular problem. Research in the field of underground GPR is scarce. As a result, a relevant dataset is hard to find but still essential in this research. Simulation of GPR is vital to create artificial radargrams to replace the lack of experimental data.

Acquisition of experimental radargrams is important in determining the feasibility of simulating radargrams. A comparison between experimental and simulated radargrams helps to get insight into the potential of GPR simulation. Therefore, a feasibility study compares the experimental and simulated GPR results.

2. Currently, literature about GPR is mostly related to detecting subsurface objects or artefacts from a ground surface environment. Placing a GPR in a closed-off underground environment, such as a sewer, might introduce artefacts as results of unknown behaviour of the emitted waves. Therefore, it is crucial to research the difference in wave propagation of using a GPR in a ground surface environment compared to an in-sewer environment. The resulting difference help to determine the potential of GPR in detecting voids from within a sewer.
3. For studying future detection algorithms, it is convenient to know what features capture the characterisation of voids. A feature extraction algorithm calculates a pool of feature from the radargram. The pool of features needs to be narrowed down to a feature set able to capture the difference between voids and non-voids. The process of feature selection minimises a classification loss function. Hence, when a classification model trains on the features, it minimises the misclassification probability resulting in accurate classification during the evaluation of the model.
4. Evaluation of the selected feature set is necessary to conclude if it shows true potential in detecting voids. A dummy classification model calculates the classification accuracy using dataset consisting out of a training and test set. The detection accuracies are calculated using all features and the selected features. The detection accuracy of the selected features should be around equal or higher compared to using all features, proving that a limited set of features can capture the characteristics of voids.
5. The behaviour of GPR inside a sewer might prove fundamentally different compared to ground surface GPR. In-sewer GPR radargram could contain additional artefacts which could increase the difficulty of detecting the voids. Ideally, the detection accuracies are similar for an in-sewer environment and a ground surface environment. A comparison in detection accuracy between both environments help to get insight into the additional

complexity when putting GPR in a sewer. Both environments relate by a Cartesian to Polar transformation and are referred to as the planar topology and cylindrical topology. Consequently, when having one topology, the other can be approximated based on the transformation. Feature extraction and selection on both topologies result in a dataset for each topology. The classification accuracy is calculated based on the selected features for each topologies using the dummy classification model. Ideally, both detection accuracies have similar detection results, meaning that the model generalises well between the cylindrical a planar topology based on the selected features. If this is the case, the designed algorithm shows potential in detecting voids from within sewers using GPR.

1.3 Research questions

The following research questions each tackle a sub problem given is section 1.2. This research aims to presents answers to all research questions.

1. How do experimental radargrams correlate with simulated radargrams when measuring voids behind a concrete sewer wall?
2. What is the difference between radargrams of a planar and cylindrical topology when measuring voids behind a concrete sewer wall?
3. What features in GPR radargrams are most feasible for detecting voids behind a concrete sewer wall?
4. What is the difference in detection accuracy between all features and the selected feature set?
5. What is the difference in detection accuracy on a planar and cylindrical topology?

1.4 Approach

Nowadays, specialists perform inspections to find sewer defects. In the future, the inspection of sewer defects might be a fully automatic process which does not require any use for a specialist. Changing from manual inspection to automatic inspection is a big step which could receive resistance by experts in the field. Therefore, the next step is to work towards a semi-automatic detection of voids using in-sewer GPR. The symbiotic relationship between the inspection specialist and the algorithm is desirable to take the step to full automatic detection of sewer defects in the future. Ideally, the algorithm creates visualisations of the detection process. The inspection specialist can give a second opinion on the detection and track the performance of the detection algorithm.

To realise a symbiotic relationship between specialist and algorithm, feature extraction is realised using image processing, enabling easy visualisation of features. Image processing calculates features from the radargram after which a feature selection algorithm calculates the corresponding weights to the features. The resulting feature weights might prove useful in future research aiming to capture desired void characteristic with a specified set of features.

Although this research focuses on feature extraction based on image processing, it also creates a strong connection point with Machine Learning (ML). Firstly, the resulting features can be used as input for an appropriate ML classification model to increase the classification accuracy. Hence, a classification model is tied to the feature extraction algorithm to optimise the classification accuracy. Secondly, the set of features can be used to select a ML model able to

do feature extraction and capture similar features. Consequently, image processing is replaceable by a ML in further research. A study of the advantage and disadvantage of the approach is vital in choosing a viable ML replacement.

1.5 Outline

This chapter described the importance of the research. The subsequent chapter 2 informs about the necessary background information relating to GPR and the process of simulating radargrams. Chapter 3 includes the feasibility study of GPR simulation software and analyses the wave propagation of a GPR from within a sewer. The design and implementation decisions are part of chapter 4. Chapter 5 describes the tests and corresponding results. Answers to the research questions are given in chapter 6, followed by recommendations regarding this research and future research.

2 Background

This chapter contains essential background information to understand the following chapters. Section 2.1 focuses on the theory behind GPR. Section 2.2 explains the how software simulates GPR.

2.1 Ground Penetrating Radar

This section explains the fundamental principles of GPR. The principles covered are the reflection amplitude and polarity, the origin of hyperbolic reflections, the horizontal and vertical resolution of GPR and wave velocity through the material. The theory in this section is essential to understand concepts introduced later on.

GPR assesses subsurface environments in a non-invasive manner. The GPR system emits electromagnetic waves that propagate through the material and partially reflect off a subsurface object with contrast in dielectric properties. The antenna receives the reflection after which the system creates a radargram based on the amplitude, polarity, and travel time of the signal.

A trace is a measurement within a specified time window from a single GPR location. A GPR system emits a waveform after a certain travelled distance, called the trace spacing. The wave reflects off subsurface objects, after which the antenna receives the reflection. A time window for the antenna specifies how long the receiver has to measure at a location. The traces from each location merge into a radargram, describing the reflections over a travelled distance. Hence, a directional GPR measurement consists out of multiple traces.

The traces and radargrams contain the measured polarity and amplitude of reflections in the time window. The amplitude of the reflection, called the reflection amplitude R , is a product of the difference between the dielectric properties of two materials. Equation 2.1 describes reflection amplitude R , where ϵ_1 is the relative dielectric constant of medium (material) one, and ϵ_2 is the relative dielectric constant of medium (material) two. Hence, equation 2.1 describes the amplitude of the reflection coming off a wave traversing from material one into material two.

$$R = (\sqrt{\epsilon_2} - \sqrt{\epsilon_1}) / (\sqrt{\epsilon_2} + \sqrt{\epsilon_1}) \quad (2.1)$$

An example of a radargram is depicted in figure 2.2. This radargram is the result of a simulation based on the environment in figure 2.1. The GPR antenna moves from left to right through the air (red), emitting a signal through a layer of concrete (orange) located on a layer of sand (grey). The signal reflects from the subsurface object (blue) back to the antenna where the reflection is received. The antenna traverses over the object, resulting in a changing distance (green lines) from the antenna to the subsurface object for the different measurement locations.

The green lines in figure 2.1 depict the shortest distance path for the emitted signal to travel back and forth. The distance between the GPR and subsurface objects influences the Two-Way Travel Time (TWTT) of the signal. Consequently, the resulting amplitude occurs later in a trace when the GPR is located further from the subsurface object. Therefore, when the GPR traverses over an object, the amplitudes form a hyperbolic shape in the radargram. Figure 2.2 depicts the hyperbola created by the changing distance from the GPR to the subsurface object.

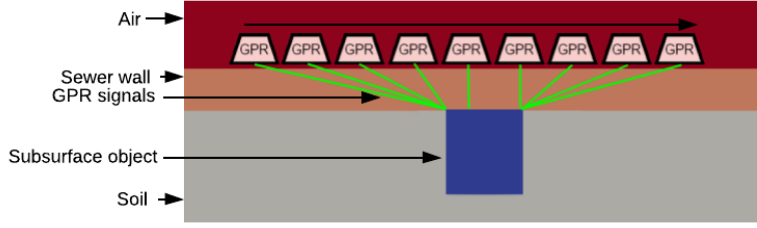


Figure 2.1: The GPR traversing over a subsurface object denoting the distance from the GPR to the object as green lines.

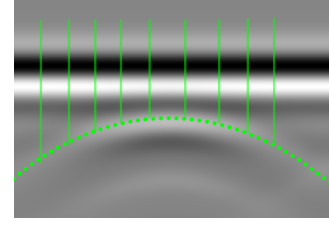


Figure 2.2: Hyperbolic reflection in a radargram.

A challenging aspect of using GPR is the trade-off between resolution and penetrative power. Measuring deep subsurface objects or objects in soil with a high permittivity is only feasible with low-frequency antennas. With a high-frequency antenna, the signal attenuates too quick, and reflections are not received. Although low-frequency antenna's result in more penetrative power, their resolution is lower since the increasing wavelength results in a less sensitive system. Therefore, a lower resolution means that the minimum object dimensions need to be larger to be detected by GPR.

The vertical and horizontal resolution defines the sensitivity of GPR. The vertical and horizontal resolution determine the minimum size of the objects. The vertical resolution is one-quarter of the wavelength. The wavelength λ is equal to the wave velocity V , divided to the GPR operating frequency f (Reynolds, 2011).

$$\lambda = V / f \quad (2.2)$$

The horizontal resolution is usually considered equal to the width of the first Fresnel zone. The first Fresnel zone is defined as the radius r (Reynolds, 2011). Hence, the resolution equals two times the radius of the first Fresnel zone.

$$r = ((\lambda_c/4)^2 + \lambda_c z/2)^{1/2} \approx (z\lambda_c/2)^{1/2} \quad (2.3)$$

Where λ_c is the centre-frequency wavelength, and z is the distance from the source to the point of reflection.

The use of GPR generally comes with a trade-off between penetrative power and resolution. This research focuses on in-sewers GPR to measure voids directly behind the sewer wall. The need for penetrative power is less important to the ground surface approach since the GPR is relatively close to the voids. The reduction in penetrative power facilitates for higher frequency GPR and therefore higher resolutions.

A reflection is not only described by the reflection amplitude R , but also by the TWTT. The TWTT depends on the material properties of the subsurface that make for a certain wave velocity. Each type of material has a specific loss factor P , describing the level of attenuation of the wave travelling through the material. Consequently, the velocity of a wave through the material

V_m is derived from the loss factor and the dielectric properties as stated the following equation (Reynolds, 2011).

$$V_m = c/(\epsilon_r \mu_r / 2)((1 + P^2) + 1)^{1/2} \quad (2.4)$$

Where c is the speed of light in free space, μ_r relative permeability (equal one for non-magnetic materials), P the loss factor such that $P = \sigma/\omega\epsilon$, σ is the conductivity, $\omega = 2\pi f$, ϵ the permittivity equals $\epsilon_r \epsilon_0$ and ϵ_0 is the permittivity of free space $8.854 \cdot 10^{-12} F/m$.

2.2 Ground Penetrating Radar simulation software

Warren et al. (2016) proposes gprMax, an open-source software pack that simulates electromagnetic wave propagation which uses Yee's algorithm to solve Maxwell's equation using the Finite-Difference Time-Domain (FDTD) method. The software is specifically designed to simulate GPR, and model electromagnetic wave propagation in fields such as engineering, geophysics, archaeology, and medicine. gprMax is command-line-driven and built around python. It runs on Central Processing Unit (CPU), Graphics Processing Unit (GPU) and High-Performance Computing (HPC) environments using OpenMP¹ and OpenMPI².

The difficult nature of simulating a GPR problem is the result of an infinite space in which the emitted wave typically attenuate. Infinite space is infeasible in the simulation since it would require infinite computation space. gprMax discretizes both the space and time continua using a FDTD grid consisting out of Yee cells (Yee, 1966). Yee cells are the building blocks for any arbitrary shape in the simulation space, including non-rectangular shapes. The non-rectangular shapes are an approximation since they consist of cubic Yee cells. The smaller the Yee cell, the more smooth the curves of an object. For example, pixels in a display are used to show curves. A higher resolution display results in a smoother curve.

gprMax restricts the modelling of objects to elementary shapes, e.g., circles, cylinders and rectangles. Although Yee cells can approximate arbitrary shapes, gprMax provides no feasible option to visually design complex shapes and convert them into a simulation. Hence, complex object shapes are practically impossible to implement. Therefore, this research restricts to modelling objects of elementary shapes.

Besides the shape, the object requires a location and material properties. Coordinates in a Cartesian space describe the location of the object. Also, the antenna position is described using Cartesian coordinates. The performance of the antenna relies on the frequency, waveform, antenna separation and movement of the GPR. The objects are characterised by their material properties: relative permittivity ϵ_r , conductivity (Siemens/metre) σ , relative permeability μ_r and the magnetic loss (Ohms/metre) σ_* . Reynolds (2011) describes many dielectric properties of earth materials. Therefore, a small set of materials often encountered in sewer environments are summarised in table 2.1. Note that, throughout this research, these are the parameters used when referring to a certain material.

A disadvantage of simulation is that it might be optimistic in terms of noise and contrast. For example, it might be difficult to distinguish a reflection from an experimental radargram, whereas in the simulated radargram the reflection distinctive. As a result, a detection algorithm

¹<https://www.openmp.org/> (URL accessed on 6-11-2019)

²<https://www.open-mpi.org/> (URL accessed on 6-11-2019)

| Material | Relative permittivity ϵ_r | Velocity $V(mm/ns)$ | Conductivity $\sigma(mS/m)$ |
|----------------|------------------------------------|---------------------|-----------------------------|
| Air | 1 | 300 | 0 |
| Water (fresh) | 81 | 33 | 0.5 |
| Concrete | 4-30 | 55-250 | 1-100 |
| Sand (dry) | 3-6 | 122-173 | $10^{-4} - 1$ |
| Sand (wet) | 10-32 | 53-95 | 0.1-10 |
| Average (soil) | 16 | 75 | 5 |

Table 2.1: Properties of materials that recur in this research.

might perform well on simulated radargrams and wrong on experimental radargram. Additionally, the simulation uses elementary shapes and homogeneous materials, contradictory to real materials, which are heterogeneous and have all kinds of shapes. As a result, it is challenging to simulate any arbitrary subsurface environment in great detail.

The advantage of simulating the GPR radargram is the high modularity and scalability in the creation of a dataset. Simulation input files can be generated as desired and automatically simulated by as many resources necessary. In an experimental setup, for every measurement, the voids have to be changed and manually scanned with the GPR. Another advantage of simulation is that the radargrams are inherently labelled. The input files for the simulation already contain the location, dimension, and material of the object. Using the data from the input files, it is possible to calculate the expected location of reflection. Hence, the simulation input files are labelled. As a result, the detection problem equals a supervised (ground truth is known) classification problem. Therefore, further research mainly depends on simulated radargram from gprMax.

3 Analysis

This chapter describes the analysis of different important topics of the research. The goal of this chapter is to analyse the theory that leads to answering the research questions. Section 3.1 explores the differences between experimental and simulation results by using the gprMax simulation software. The feasibility study researches the trade-off between radargram quality and simulation time for different values of the Yee cell size and trace spacing. The goal of section 3.2 is to get more insight into the characterisation of voids in radargrams. The characterisation helps to decide on potential types for the features extraction process. Section 3.3 simulates planar and cylindrical topologies to determine the differences in wave propagation and additional effects on the radargrams.

3.1 Feasibility study on gprMax

The feasibility study on gprMax analyses the potential of gprMax for radargram acquisition from within a sewer by simulation. A comparison between gprMax radargrams and experimental radargrams provides more insight into the pros and cons of both acquisition techniques. The results help to support the decision on the use of gprMax.

The main question answered by the feasibility study is: *what is the most feasible configuration in the simulation radargrams?* Ideally, the radargrams resulting from the most feasible configuration matches with the experimental results. Three different tests help determine the feasibility of gprMax. For the full feasibility study, see appendix C.

1. Test 1: create a comparison between duplicate experimental results to determine the consistency of acquisition in the experimental setup. Additionally, test if an introduced inconsistency influences the numerical quality metric, and thus determine if the inconsistency, resulting in different radargrams, is measurable.
2. Test 2: compare the simulation times and quality results of different simulation configuration for each 2D and 3D radargrams separately. Different values for Yee cell size and trace spacing result in a trade-off between radargram quality and simulation time.
3. Test 3: compare the simulation times and quality results of different simulation configuration between 2D and 3D radargrams. The comparison explores how 2D simulations can capture the results of a 3D model.

To summarise the conclusion of the feasibility study in appendix C. The findings contribute to the understanding of the feasibility of gprMax and provide a set of parameter values for further simulations. Test one concludes that experimental measurements are stable, and the quality metrics can measure a slight difference in the difference between radargrams. During measurements in the experimental setup, the GPR was lifted slightly at the end of the measurements to cover the gap without concrete tiles, see figure C.2. The inconsistency in the radargram is due to the erratic lifting movement.

Test two combines different values for the Yee cell size and trace spacing resulting in the most feasible simulation configuration. Simulation in 2D is most feasible with a Yee cell size of 2-

4mm and a trace spacing of 4-8mm. 3D simulation is most feasible with Yee cell size of 5mm and trace spacing 5-10mm.

Test 3 concludes that 3D simulation is 40-50 times slower than 2D simulation. If a 2D simulation captures similar reflections to 3D, it might be more convenient to simulate in 2D. The reduction in simulation time would significantly increase the feasibility of gprMax. A comparison between simulation time and the quality metrics results in two different clusters of tests. The difference in tests proves that some tests perform significantly worse than others. The difference in quality metric values is the result of a considerable difference in pixel intensity in the homogeneous parts of the radargram. Although the numerical quality decreases, on visual inspection, the simulations prove to simulate the reflections adequately. Consequently, 2D simulations are considered to capture the reflections from the more realistic 3D simulations while reducing the simulation time. Therefore, this research prefers simulations in 2D. More research on the difference in shape, location, polarity, and reflection coefficient would help to establish a higher degree of accuracy on determining the feasibility.

3.2 Void characterisation

Unique characterisation of voids is essential to separate voids from other subsurface objects. A reflection is described by the reflection amplitude R and the TWTT. Therefore, the characterisation depends on reflection amplitude and the TWTT and any new data calculated from both, such as the hyperbolic shape of the reflection, explained in section 2.1.

One example of the use of the reflection amplitude in the characterisation of objects is creating an expectation of the possible reflections. The contrast in material influences the amplitude and polarity of the reflections, see equation 2.1. Vice versa, the reflection amplitude is also useful in making assumptions about the materials leading to the reflections. Therefore, extracting the reflection coefficient from radargrams could help make assumptions about the subsurface materials. For example, consider the hypothetical environment where air and a water void are located directly behind a concrete sewer wall surrounded by average soil. Substituting the material properties from table 2.1 into the equation 2.1 results in a reflection coefficient for the water void $\sim 0.33 < R < \sim 0.69$ and for the air void $\sim -0.64 < R < \sim -0.24$. Calculating the reflection coefficient in a theoretical environment gives an indication of what intensity reflection to expect for certain subsurface objects.

One of the problems with detecting voids is the moisture content in the surrounding soil. Consider water void, it is assumable that the surrounding soil is saturated by water as well. Hence, the dielectric properties of the void and the surrounding soil are similar, resulting in a weak reflection coefficient that might prove not distinctive.

A second difficulty about water voids surrounded by humid soil is the difference in loss factor P during wave propagation. The loss factor, defined by equation 2.4, is different for water and air-filled voids. To calculate the loss factor, substitute the dielectric properties from table 2.1 in the function for the loss factor. The resulting in $P_{air} \approx 0$ for air and $P_{water} \approx 0.11$ depict the dissipation rate of the wave through the material. An emitted wave keeps reflecting within the void until it attenuates, which is relatively quick for water and slow for air. The phenomenon of multiple reflections, called reverberations, is unique for air voids since the signal attenuates quicker as a result of the higher loss function. Kofman et al. (2006) builds on the assumption that voids can be detected as reverberation patterns if the antenna wavelength is of the same order of magnitude as the void diameter. Additionally, Kofman et al. (2006) states that the reflections from the bottom of the void are ambiguous and mostly of queasy-hyperbolic form;

contradictory to flat or quasi-flat reflections observed from the reflection tops. Therefore, reverberation patterns of voids are challenging to detect since their shape is arbitrary.

To conclude, air voids theoretically produce more characteristics reflections than water voids. Air voids produce relatively stronger reflections and possibly contains reverberation patterns. As a result, it is assumable that air voids are easier to detect compared to water voids. Therefore, this research continues with air voids and any further mention of void denotes an air void.

3.3 Wave propagation inside a sewer

The experimental radargrams are acquired in the planar topology during the feasibility study, see section C.0.1. GPR measurements inside an underground sewer might result in radargrams with additional artefacts. To be able to anticipate on the artefacts, it is important to simulate a sewer in the cylindrical topology and compare the resulting radargram with the corresponding planar topology.

Figure 3.2 shows a cross-section of the cylindrical topology where the grey object is the sewer wall with thickness w . The location of GPR transmitter (red dot), receiver (green dot) and coordinates of the cylindrical topology are in a polar coordinate space (r, θ, z) . Where r is the radius of the GPR from the middle point (blue dot) of the sewer, θ denotes the rotation around the middle point, and z is the translation along the sewer. The antenna separation a denotes the difference between the transmitter and the receiver. The distance between the middle point and the sewer wall equals c . A subsurface object is located at the Common Midpoint (CMP) between the transmitter and receiver with a distance d to the outer sewer wall. The transmitter rotates in 360° , and the receiver follows based on the antenna separation. When antenna finishes the full circle, it is translated to the next location z in the sewer.

Figure 3.1 resembles the sewer cross-section but in the planar topology. The planar topology is the result of the Cylindrical to Planar transformation, which means it is the "rolled out" version of the cylindrical topology. The variables are kept equal to the cylindrical topology for easier comparison. Note that each rotation inside the sewer corresponds to a line measurement in the planar topology.

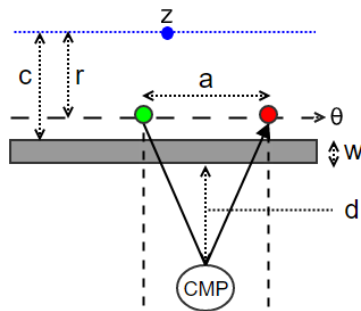


Figure 3.1: Planar topology.

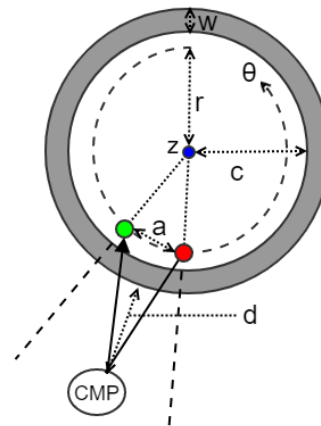


Figure 3.2: Cylindrical topology.

Figures 3.3 and 3.4 overlays the topologies parameters on two visualisations of gprMax simulations. The colours describe the different materials where the air is red, sewer wall brown, subsurface objects blue and the surrounding soil grey. Figure 3.4 denotes the non-adjusted

simulations results from gprMax. Figure 3.3 underwent a Cylindrical to Planar transformation. Therefore, the planar topology might contain additional artefacts.

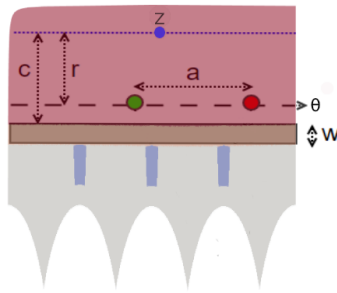


Figure 3.3: Sewer cross section of the planar topology.

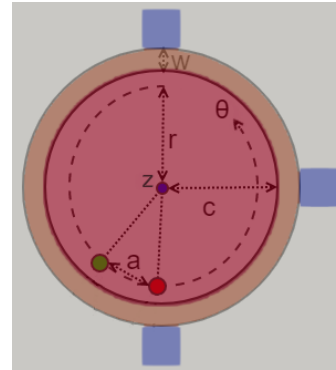


Figure 3.4: Sewer cross section of the cylindrical topology.

Firstly, figure 3.3 includes parabolas in the bottom of the visualised environment. These parabolas come forth out of the transformation, and describe the different magnitudes from the centre of the cylindrical figure 3.4 to the image border. The distance from the middle points to the image borders gets greater when moving to the outer corner of the image. Hence, it results in a parabola when applying a Polar to Cartesian transformation.

Secondly, the shape of the subsurface object changes, see the blue objects. The objects are thinner since the "unrolling" of the cylindrical topology is compressed in a rectangular image. Theoretically, figure 3.3 should be wider, but by transforming the ratio is kept similar to figure 3.4. More important is that the subsurface objects are no longer rectangular in the planar topology. The objects are slightly thinner in the lower section as a result of the different projection on an object when the GPR rotates in the cylindrical topology.

Thirdly the objects in the planar contain more noise as a result of rounding during the transformation. Calculating the new location of each pixel from the cylindrical topology to the planar topology results in rounding mistakes since an image is a discrete domain. Unfortunately, this noise is not visible in the planar topology, see figure 3.3.

An additional problem of transforming the environment from cylindrical to planar topology is the introduction of non-elementary shapes. The special shapes can not be designed using basic gprMax shapes. Subsequently, gprMax does not include a way to import an image to a simulation input file. Therefore, a Polar to Cartesian transformation is not feasible. Instead, consider an approximation of the expected planar topology, see figure 3.5. The approximation only transforms the location of subsurface objects from the Polar space into a Cartesian space.

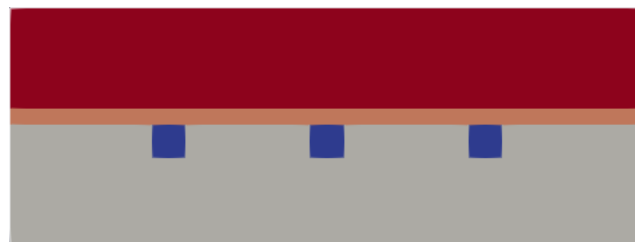


Figure 3.5: Sewer cross section of the approximate planar topology.

Figures 3.6 and 3.7 depict the radargrams from the topologies in figure 3.5 and 3.4 respectively. The radargram resulting from the cylindrical reveals a significant amount of additional horizontal reflection in comparison with the radargram from the planar topology. The reflections start occurring after $5ns$, which is the assumed reverberation time. If the reflections are the result of reverberating waveforms, the diameter of the sewer would determine the reverberation time. The reverberation time is the product of the sewer diameter and the velocity of the wave. In this simulation, the sewer radius c equals $300mm$, and the antenna rotates in the sewer with a radius of r of $250mm$. The clearance between the antenna and the sewer wall is equal to $50mm$.

The Expected Reverberation Time (ERT) is equal to two times the distance from the antenna to the opposite sewer wall divided by the wave velocity through air V_{air} , see table 2.1. Additionally, sum time zero to Expected Reverberation Time (ERT) offset to compensate for the Time Zero Correction (TZC) in the emitted wave. The following equation calculates the Expected Reverberation Time (ERT) for the described setup in figure 3.4.

$$c = 300mm; \quad r = 250mm; \quad V_{air} = 300mm/ns; \quad timezero = 1.41ns; \quad (3.1)$$

$$ERT = 4 \cdot r / V_{air} + timezero$$

$$ERT = 1000/300 + 1.41 \approx 4.74ns \quad (3.2)$$

The location of the ERT in the radargram is proportional to the time window of the radargram. The product of the ERT and the vertical resolution divided by the time windows results in the start of the expected reverberation location in a radargram. The green line in figure 3.7 depicts the Expected Reverberation Time (ERT) of $4.74ns$ in the radargram with a time window of $10ns$. The line accurately describes the time where the first of multiple horizontal reflections are measured. Therefore, the additional reflections of the cylindrical topology are assumed to be the result of reverberation in the sewer.

The reverberation patterns in the radargram are undesired since they could mask reflections of subsurface objects. Fortunately, only objects with a TWTT within the reverberation band are affected. In other words, the region of interest does not have to overlap with the reverberation pattern. Since this research focuses on voids close behind the sewer wall, it is assumable that the reverberation pattern is not a potential issue. Therefore, this research applies no filtering to reduce the reverberation pattern.



Figure 3.6: Radargram of the planar topology.

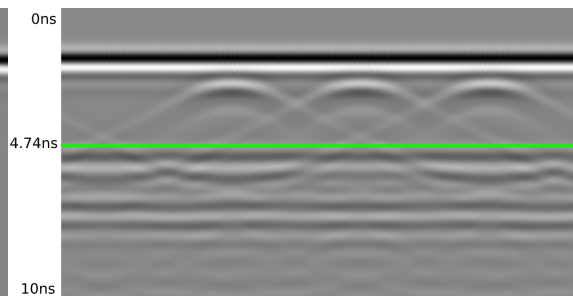


Figure 3.7: Radargram of the cylindrical topology.

A second difference between the planar and cylindrical topology is the length of the reflection path of a subsurface object. The antenna emits a signal which propagates through the sewer wall and reflects off the object located at the CMP, again it propagates through the sewer wall to

be received by the antenna. The length of the reflection path is equal to the difference between the signal intersection with the inner and outer sewer wall. The signal path is depicted as a solid black line on both topologies, see figure 3.1 and 3.2. The curvature of the cylindrical topology results in a longer reflection path through the sewer wall compared to the planar topology. The difference is that the result of the signal not being perpendicular on the sewer; hence, the path through the sewer wall is longer.

For example, consider a sewer with inner radius c of 600mm, a sewer wall w of 50mm thick and a subsurface object located precisely between the transmitter and the receiver. The environmental specifications result in the following length of the reflection path for a cylindrical and planar topology. It is beneficial to propagate through as less material of the sewer wall as possible since it would only add to the attenuation of the signal and thus reduce the penetrative power. Figure 3.8 shows that the length of the reflection path for a cylindrical topology grows faster than for the planar topology. Therefore, when using GPR in a sewer, it is essential to keep the antenna separation as small as possible to reduce the loss in penetrative power.

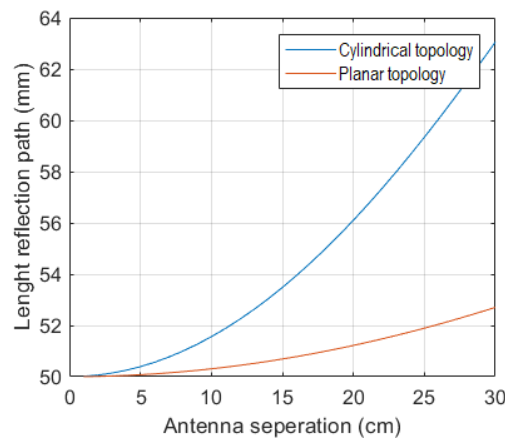


Figure 3.8: Length of the reflection path versus the antenna separation.

In summary, in sewer GPR measuring introduces reverberation. The reverberation reflections do not necessarily overlap with the reflection of subsurface objects directly behind the sewer wall. Therefore, it is considered unnecessary to filter the reverberation reflections to detect objects directly behind the sewer wall. The non-linear movement of the GPR in the cylindrical topology results in different projections of subsurface objects. Subsequently, the cylindrical shape of a sewer results in a longer reflection path through the concrete sewer wall. Hence, it is essential to keep the antenna separation to a minimum to minimise the loss in penetrative power.

4 Design and Implementation

The goal of the algorithm proposed in this chapter is to acquire features with the potential to describe voids directly behind the sewer wall. Many different techniques apply to the process of feature detection based on images, but the vast majority categorises in two groups: Image Processing techniques and Machine Learning (ML) techniques.

Both techniques have pros and cons relating to this research. Image processing algorithms generally contains many parameters, which makes it difficult to tune for an optimal result. Even when tuned, the algorithm could not generalise well to different input images. The advantage of image processing techniques is that it makes for inherent visualisation of intermediate results. The visual results are advantageous for evaluation and backtracking of any undesired artefacts.

The advantage of machine learning techniques is that they calculate the feature weights by minimising a loss function. The process of training the model reduces the need for manual tuning. Also, the level of generalisation of machine learning can be improved by training on a properly designed training set. A disadvantage is that the model behaviour depends on the number of training iterations and quality of the dataset. Therefore it is difficult to argue for the classification outcome (void, other objects or non-void) of observations. More advanced machine learning techniques tend to become more abstract, and therefore often referred to as a black-box model.

This chapter describes the choices made in the design and implementation of an algorithm for feature extraction, selection and classification. The outline of this chapter follows a chronological order, starting at radargram acquisition in section 4.1, followed by describing the pre-processing steps of the radargrams in section 4.2. After pre-processing, the radargrams are labelled to create a ground truth to determine the accuracy of the classification, see section 4.3. The process of feature extraction determines interesting regions and hyperbolas and combines both types of features into observations, see section 4.4. Section 4.5 explains the process of labelling the observations based on the labels from the radargram. Section 4.6 explains the process of feature selection to get an optimal set of features. Finally, section 4.7 explains the details for a dummy classification model to calculate the classification accuracy and prove that the selected features are capable of describing the void.

Figure 4.1 depicts the flowchart of the design of the full algorithm. Each section describes its process in more detail. The processes, denoted by rectangles, are explained in the sections of this chapter which include more detailed flowcharts. The detailed flowcharts of the processes can be substituted in this general flowchart to get a total view of the design.

4.1 Radargram acquisition

During the feasibility study, see section 3.1, a small set of experimental radargrams is acquired by doing manual measurements using a GPR system in the experimental setup. The resulting experimental radargrams are SGY files, which is a file extension for various seismic tools. The SGY files are converted into a 1920×1080 Portable Network Graphics (PNG) image of the corresponding radargram. The PNG images are imported into MatLab for further processing of the radargrams.

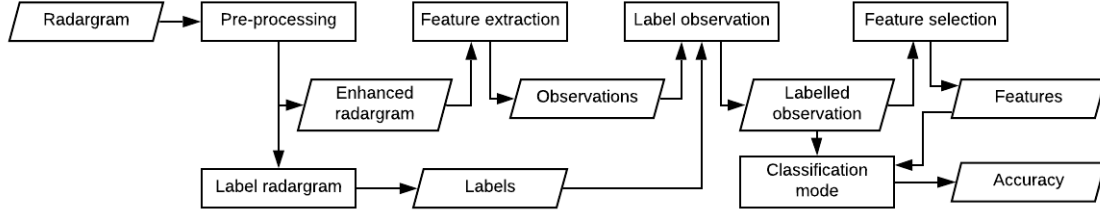


Figure 4.1: Flowchart of the general design.

The gprMax simulations are offloaded to the High-Performance Computing (HPC) cluster of the University of Twente. Utilising the Graphics Processing Unit (GPU) on the HPC cluster drastically decrease the simulation times (Warren et al., 2019). The simulations are arbitrarily scheduled over either multiple Nvidia’s Geforce GTX 1080 Ti GPUs, GeForce GTX TITAN X GPUs or Tesla P100 GPUs.

The gprMax simulation input files contain all information about the simulation, including the domain size, Yee cell size, operating waveform type, frequency, transmitter and receiver path and the objects describing the environment, see section 2.2. A dedicated python script generates a specified amount of input files, including a static environment and a specified number of subsurface objects. The python script generates the input files for both the cylindrical topology and the approximate planar topology, see section 3.3. During generation, the environments of both topologies are split up in a number of equally sized section. The number of sections is equal to the specified number of subsurface objects. Each section includes an arbitrarily located circle between a range of 50mm-100mm behind the sewer wall. The material of the subsurface objects is randomly picked from a few hard-coded options, see table 2.1. Hence, during generation, the subsurface objects are arbitrarily located in the environment of the topologies. Note that the environments for a generated simulation match for both topologies. For example, the n -th planar simulations match the n -th cylindrical simulation by the Cylindrical to Planar transform approximation, see section 3.3

Voids in an early stage are small of size. Consequently, the size of the circles representing the voids in simulation should be small, while ensuring that the reflections are measurable for the GPR. To make sure the object is measurable, it should be bigger than the vertical and horizontal resolutions explained in section 2.1. For calculating the vertical resolution, see equation 4.1, consider the velocity of radio-waves through dry sand, which matches the experimental setup from the feasibility study. The velocity of dry sand equals $122 \leq V_{dry_sand}(mm/ns) \leq 173$, see table 2.1. For conveniences, take the average of the velocity values $V_{dry_sand} = 148mm/ns$. Substituting the values in equation 4.1 results in a vertical resolution of 37mm.

The horizontal resolution depends on the distance from the source of the emitted wave to the point of reflection, see equation 4.2. For the simulations, the clearance between the GPR and the sewer wall c equal 50mm. The sewer wall thickness w also equal 50mm and the distance between the sewer wall and the object d is an arbitrary value between 50mm and 100mm. Hence, the total distance varies between 150mm and 200mm. The largest distance, $max(c + w + d) = 200mm$, is used to define the smallest possible object which results in a horizontal resolution of approximately 10mm. The object its size is set to roughly three times the biggest resolution, in this case, the vertical resolution of 37m. Hence, all circles have a 100mm diameters. An object with a size a few times bigger than the resolution guarantees that it forms reflections measurable by the GPR.

$$\text{Vertical resolution} = 0.25\lambda = 0.25 \cdot (V/f) = 0.25 \cdot (148/1) = 37\text{ mm} \quad (4.1)$$

$$\text{Horizontal resolution} = ((\lambda_c/4)^2 + \lambda_c(c+w+d)/2)^{1/2} = ((1/4)^2 + 1 \cdot 200/2)^{1/2} \approx 10.0031\text{ mm} \quad (4.2)$$

Figure 4.2 depicts the visualisation of a gprMax input file. The Slurm workload manager¹ schedules the input files on a specified HPC node. The resulting radargram is saved as a 1920×1080 PNG file. Based on the results of the feasibility study in section 3.1, all simulations are in 2D. Simulations are 2D require significantly less simulation time which makes it more feasible to create a big dataset useful in the later stage of this research.

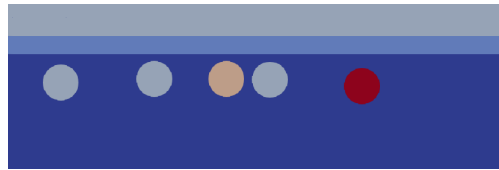


Figure 4.2: Visualisation of the input file.

Figure 4.3 depicts the radargram resulting from the environment in figure 4.2. Any visualisation of the image processing step in this chapter will be visualised using the same radargram. Note that this is a 2D simulated radargram, hence the result might significantly differ on 3D simulated radargrams or experimental radargram.



Figure 4.3: gprMax simulated radargram.

¹<https://slurm.schedmd.com/documentation.html> (URL accessed on 6-11-2019)

4.2 Pre-processing

This section explains the steps of pre-processing the radargrams. Pre-processing consists of two steps, Time Zero Correction (TZC) and contrast enhancement. Pre-processing is necessary to optimally extract features from the radargrams in a later stage. Figure 4.4 depicts flowchart of pre-processing. TZC and contrast enhancement are discussed in the following sections, respectively.

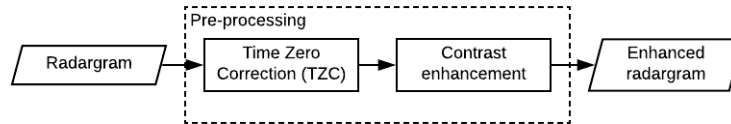


Figure 4.4: Flowchart pre-processing.

4.2.1 Time zero correction

TZC sets time zero position at ground surface level instead of at the point of the emitted wave. In other words, TZC compensates for the height of the antenna above the ground surface. As a result, determining time zero improves the accuracy in shallow depth measurements since the height of the antenna above the ground surface might be significant in comparison with the depth of the measurable objects (Yelf and Yelf, 2006).

Time zero is set equal to the first positive peak of the direct wave. The first positive peak is generally the result of the emitted wave travelling directly from the transmitter to the receiver. The direct wave is the largest amplitude of the first positive peak. Since there are no obscurities in the signal path, it is generally easy to identify and favoured over other techniques (Yelf and Yelf, 2006).

The location of the first positive peak is different for each trace. That is why TZC calculates the location of the first positive peak for each trace. Figure 4.5 denotes a trace with a red intersecting line describing the first positive peak. Note that the x -axis depicts the vertical pixels in a radargram and the y -axis the intensity of the pixels. Therefore, the x value of the red line denotes the time zero pixel on the y -axis of radargram the trace belongs to.

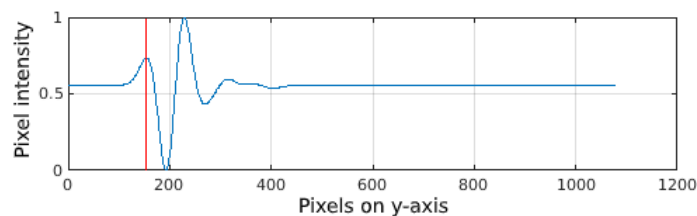


Figure 4.5: Section of an enhanced radargram.

The process of calculating the location of the first positive peak repeats for all traces. The TZC of a radargram is equal to the mean of all first positive peak locations. Figure 4.8 depicts a radargram including the TZC as a green line.

4.2.2 Contrast enhancement

Contrast enhancement is beneficial in detecting reflections. Whenever the absolute reflection coefficient is near zero, the resulting reflection blends in with the homogeneous pixel intensity. Local contrast enhancement is applied to enhance relative weak reflections.

Many applications in the field of seismological surveillance have options for manually setting the gain of different horizontal bands in the radargram. Increasing the gain in a particular band increases the contrast in the radargram, which might reveal weak reflections that were not visible before. Since the research goes towards an automatic algorithm, it is unwanted to determine the gains manually. Therefore, this section introduces an automatic contrast enhancement technique.

Contrast-Limited Adaptive-Histogram Equalisation (CLAHE) (Zuiderveld, 1994) calculates the contrast transform function on small regions called tiles. The contrast of each tile is enhanced individually, resulting in a tile histogram matching a specific distribution function. Together, all tiles form a full image. The operation on separate tiles introduces edges on the border of the tiles. The resulting images, including edges, is blurred using bi-linear interpolation to reduce the edge contrast and get a smooth image.

The advantage of CLAHE is the possibility to limit the contrast stretch strength. By setting the parameters correctly, it is possible to enhance contrast in the region of interest with minimum noise amplification. Hence, it is of significant importance to reduce noise amplification since noise could mask reflections or lead to false classifications.

The parameters of the CLAHE are the number of tiles, the contrast enhancement scalar, contrast enhancement range, and histogram distribution. The algorithm is tuned visually to enhance the contrast as much as possible while reducing the amplification of noise. The distribution of the histogram is uniform with a maximum stretch range. Therefore, the expected contrast is maximised over the possible range of pixel values.

Figure 4.6 depicts the enhanced version of the original radargram in figure 4.3. The enhanced radargram shows a significant amount of contrast compared to the original radargram. Because of the higher contrast, it is relatively easy to recognise the reflections. Therefore, the reflections become more distinctive for the feature extraction algorithm.

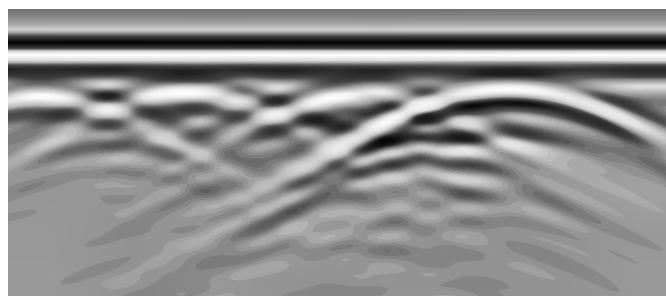


Figure 4.6: Section of an enhanced radargram.

4.3 Label radargrams

A sense of ground truth is necessary to create a labelled dataset which describes to which class each observation belongs. Simulation using gprMax requires input files which describe the environment under simulation. The input file contains the description of the discrete environment using different parameters, see section 4.1. The parameters of the gprMax input file lead to an expectation of the reflection in the radargram, making the simulations inherently supervised. The parameters are read from the input file to software variables to calculate the labels denoting the expected location of the reflection in the radargram. Hence, using the object coordinates and material properties variables, a label is calculated for each simulated subsurface object.

The location of the label is described in two dimensions since the radargram is a two-dimensional PNG image. Assume a GPR in a planar topology. The horizontal axis of the radargrams describes the movement in a certain direction, and the vertical axis describes the time for a signal to reflect off a subsurface object and be received by the GPR. The horizontal location of a label is proportional to the travel distance of a GPR and the dimension and location of the subsurface object. The subsurface objects, limited to a box and cylindrical shapes in gprMax, consists out of vertices. The *box*(v_1, v_2) is described between two vertices in a 3D space: vertex one $v_1(x, y, z)$ and vertex two $v_2(x, y, z)$. Note that even 2D simulations require 3D parameters, which is the result of 2D simulations consisting of a single layer of 3D Yee cells. The *cylinder*(v_1, v_2, r) consists out radius r and two vertices v_1, v_2 describing the midpoint of the two faces of the cylinder. Additional variables are the travel distance of the receiver Δd_{rx} , and the radargram image resolution in two dimensions res_x and res_y .

The horizontal location of the label in the radargram is proportional to the object dimension and location in the measurement direction. The first step is to scale the object locations in the environment to the measured distance and the horizontal radargram size. Consider the point of reflection as the CMP between the transmitter and receiver. Assuming the radargram is a 1920×1080 image, the label is defined between horizontal coordinates ℓ_{x1} and ℓ_{x2} on the x -axis of the radargram, see equation 4.3. Where x_{v1} is the x coordinate of vertex $v1$ and where x_{v2} is the x coordinate of vertex $v2$.

$$\begin{aligned}\ell_{x1} &= res_x \cdot (x_{v1} - CMP) / \Delta d_{rx} \\ \ell_{x2} &= res_x \cdot (x_{v2} - CMP) / \Delta d_{rx}\end{aligned}\tag{4.3}$$

The vertical location of the label in the radargram results from the TWTT, TZC and the radargram image resolution. The parameters in the input file reveal the dielectric properties of the layers between the GPR and subsurface objects. The dielectric constant ϵ_r is substituted into a simplified equation for the wave velocity through the material, see equation 4.4. This equation assumes low loss materials, thus $P \approx 0$ (Reynolds, 2011). The simplified wave equation is preferred since the surrounding soil is dry sand, which has a relatively low loss factor.

$$V_m = \frac{c}{\sqrt{\epsilon_r}}\tag{4.4}$$

Next, the resulting wave velocity V_m in mm/ns is converted to ns/mm . The conversions allow multiplying the velocity through a material V_m by the thickness τ of the respective layer of material, see figure 4.7. The result is the time for a wave to propagate through a layer of material

(object), called the propagation time Pt_{obj} , see equation 4.5. Multiply the sum of the propagation time of all layers between the receiver and reflection by two to calculate the TWTT of the emitted wave, see equation 4.6. The calculation for label height on the y -axis defined between points y_1 to y_2 , is a combination of the TZC plus the TWTT scaled to the vertical resolution res_y of the radargram based on the time window Δt of the simulation, see equation 4.7.

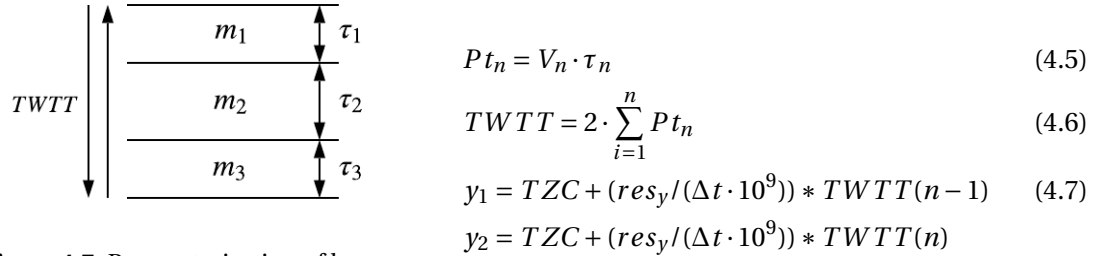


Figure 4.7: Parameterisation of layers.

The above theory explains the calculation of labels for the planar topology. The cylindrical topology is very similar, but instead of having rectangular layers, it contains out of circular layers. A polar coordinate system describes the position of subsurface objects from the centre point of the sewer. The space consists of angle θ is equal to $0 \leq \theta \leq 2\pi$ and magnitude r . The location of the subsurface objects is transformed from the cylindrical to the planar topology using the trigonometric equations 4.8. When calculating the labels for a cylindrical topology, first transform the location to the planar space using trigonometry. Secondly, use the planar locations to calculate the label position in the radargram.

$$x = r \cdot \cos(\theta); \quad y = r \cdot \sin(\theta) \quad (4.8)$$

The position of the label gives a good indication of where to expect a reflection in the radargram. Still, the reflections from a subsurface object may befall outside of the corresponding labels. Depending on the material properties, the reflection lasts a while before the signal fully attenuates, resulting in multiple similar reflections on the vertical axis of the radargram. Although the location of the reflection is not within the label, the reflection can still be the result of the subsurface object described by the label. To counter this effect, the bounding box should contain vertical padding which reduces the probability of reflections of an object falling outside the label. The reflection depicting a subsurface object is generally wider than the object itself. Therefore, the label width is equal to two times the width of the original calculated label size for a reflection.

The label describes not only the location but also the class of the expected reflection. The material is of importance when determining if a classification model can distinguish reflections as a result of voids or other materials. The bounding box of the label depicts the different material as by colours. White is an air void and red other subsurface materials. Figure 4.8 depicts the enhanced radargram, including the calculated labels. Note that that label three in the radargram describe a location where there is not a visible reflection as a result of overlapping reflections.

Although the labelling technique described in this section can give a reasonable expectation of the occurring reflection, there are a few nuisances which can not be taken account for. Firstly, the label can partly overlap, resulting in a reflection that is considered a result of two different subsurface objects. E.g., figure 4.8 shows that the reflections of labels two and four overlaps the expected reflection at label three. Secondly, the rectangular label describes the location of a hyperbolic reflection by a rectangular shape. Reflections artefacts may be located within the

label but are not related to the reflection. Consequently, the algorithm labels the reflections as subsurface objects while it is not related to the subsurface object. Hence, it is called a false positive. Thirdly, if a virtual environment contains subsurface objects with almost no dielectric contrast, it results in labels in the radargram although the reflections might not be visible, resulting in false negatives.

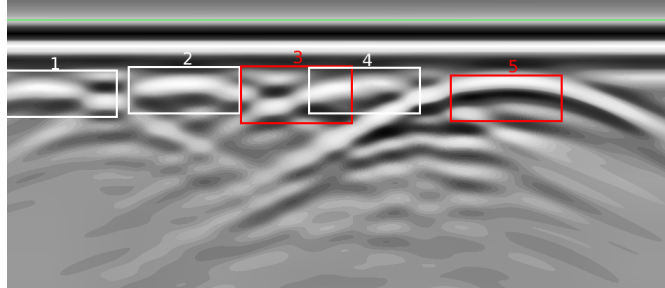


Figure 4.8: Labelled enhanced radargram including a green line depicting time zero.

Although the labelling is supposed to be ground truth, it can not be guaranteed. Still, the labelling technique makes for fast data generation and solve the ambiguity introduced when labelling manually. Therefore, the research continues with all simulations being automatically labelled using the described algorithm.

4.4 Feature extraction

Feature extraction consists out of two different techniques: region feature extraction and hyperbola feature extraction, see section 4.4.1 and 4.4.3 respectively. The two feature extraction techniques are connected sequentially. Firstly, region feature extraction calculates the regions based on the Maximum Stable Extremal Regions (MSER) algorithm. The region features are calculated from the resulting regions, for example, area, centroid and circularity. Secondly, A region to line transformation calculates lines from the regions. Thirdly, the hyperbola feature extraction required the lines as input to calculate the hyperbolic observations plus features. The algorithm determines if a line is a hyperbola by calculating the hyperbola features, and determining if the values are within a specific range.

Figure 4.9 depicts the flowchart of the feature extraction algorithm. The final results of the feature extraction algorithm are features for each hyperbolic observation. An observation is a MSER region creating a line that is also a hyperbola. Hence, the observations consist of both types of features.

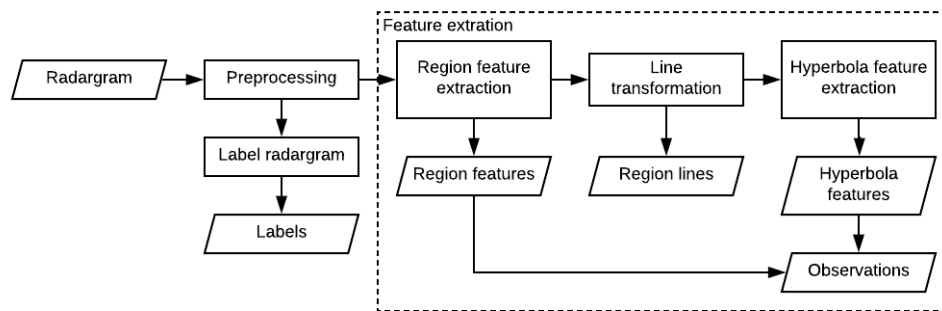


Figure 4.9: Flowchart feature extraction.

4.4.1 Region feature extraction

Region feature extraction consists of two steps. First, detecting the regions of interests. Secondly, the calculation of the features of the corresponding regions.

A region of interest in an image is often a corner, blob or a region of uniform pixel intensity. The algorithm detecting the region of interest typically detects matching points or objects and can be scale and or rotational invariant. A corner feature type is a point in an image where structures with different pixel intensity and orientation meet. Sets of multiple points are useful in tracking and transformation purposes. A corner itself does not contain much information except its location and pixel intensity, which makes it challenging to calculate more numerical metrics from it. Hence, blobs and regions with uniform intensity are more interesting to detect since they contain more data in the form of pixels. Table 4.1 describes the most common detectors.

The feature detector needs to be scale-invariant. In a radargram, the scale of the reflection is a result of the dimension of subsurface objects and the dielectric properties of soil and subsurface objects. A scale-invariant detector is required to make sure the scale does not influence the accuracy of the detection.

| Feature detector | Feature type | Scale independent |
|---|-------------------------------|-------------------|
| Features from Accelerated Segment Test (FAST) (Rosten and Drummond, 2006) | Corner | No |
| Minimum eigenvalue algorithm (Shi and Tomasi, 1994) | Corner | No |
| Corner detector (Harris and Stephens, 1988) | Corner | No |
| Speed Up Robust Features (SURF) (Bay et al., 2006) | Blob | Yes |
| KAZE (Alcantarilla et al., 2012) | Blob | Yes |
| Binary Robust Invariant Scalable Keypoints (BRISK) (Leutenegger et al., 2011) | Corner | Yes |
| Maximum Stable Extremal Regions (MSER) (Matas et al., 2004) | Region with uniform intensity | Yes |
| Oriented FAST and Rotated BRIEF (ORB) (Rublee et al., 2011) | Corner | No |

Table 4.1: Most common feature detectors.

The requirements for the detector to detect a blob or regions of uniform interest while being scale independent leads to the following detectors: SURF, KAZE, and MSER. The advantage of MSER over SURF and KAZE is that inherently leads to a list of pixels. SURF and KAZE describe the blob by using the location of one pixel. A region contains more data in the form of a pixel list compared to a point which is only one pixel. Using the pixel list, additional features, e.g., area, eccentricity and circularity, can be calculated. Therefore, the MSER detector has an advantage over the other descriptors and is used to extract the regions.

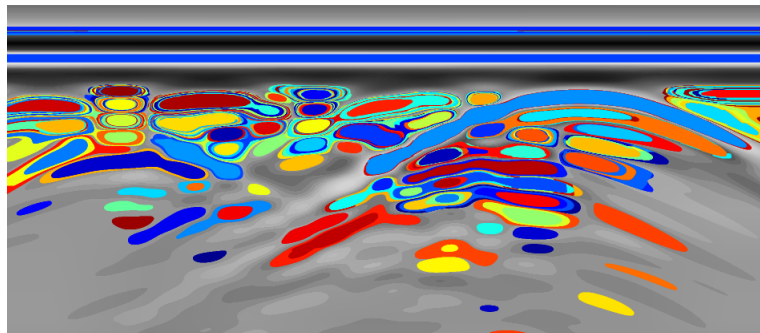


Figure 4.10

MSER (Matas et al., 2004) detect stable regions by stepping incrementally through the intensity range of the radargram using a variable threshold delta. The areas from the resulting thresholds vary in size. The variation in area size must be lower than the area variation function parameter, to be considered a stable region.

Figure 4.11 visualises MSER regions as layers of water in a bucket². Think of the threshold delta as a cup filling the bigger bucket. The smaller the cup, the more time it needs to fill the bucket

²<https://nl.mathworks.com/help/vision/ref/detectmserfeatures.html> (URL accessed on 28-10-2019)

resulting in more layers of water and thus more regions. Changing the threshold iteratively results in regions that overlap each other.

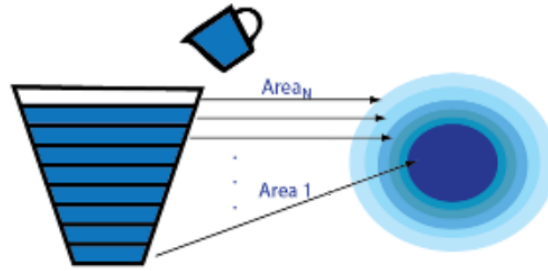


Figure 4.11: Bucket example of visualising MSER regions. Left is a cross section from the side of the bucket depicting the layers of water. Right is the top view of the bucket, which is comparable a contour plot depicting the layers of water.

The resulting regions denote as matrices of two column vectors containing x and y coordinates. The matrices containing the pixel coordinates are called a pixel lists $PL_{n,2}$, where n is the number of pixels in the region. Basic region properties are calculated from the pixel lists and used as features. Table 4.2 describes all resulting region features from pixel list $PL_{n,2}$. Since it is still undetermined which features are adequate for capturing the void characteristics, all single numerical region properties are considered features.

| Feature name | Description |
|-------------------------|---|
| Area | The number of pixels in the region. |
| Centroid | Center of the mass of the region consisting out of $[x, y]$ coordinates. $Centroid_1$ denotes the value on the x -axis and $Centroid_2$ the value on the y -axis. |
| Convex Area | Number of pixels in the convex image. The convex image is a binary image of the convex hull. |
| Convex hull | Smallest polygon containing the region. |
| Circularity | The roundness of the object. |
| Eccentricity | Ratio of the foci of the ellipse and its major axis length as a scalar. The ellipse has the same second-moments as the region. |
| Equivalent Diameter | Scalar equal to the diameter of a circle with the same area as the region. |
| Extent | Ratio of pixels in the region to pixels in the regions bounding box. |
| Max/min Feret Diameter | Maximum/minimum distance between two boundary points on the antipodal vertices of the convex hull containing the region. |
| Max/min Feret Angle | Angle of the maximum/minimum Feret diameter with respect to the horizontal axis of the image. |
| Major/minor Axis Length | Pixel length of the major/minor axis of the ellipse that has the equal second-moments to the region. |
| Orientation | angle between the major axis of the ellipse and the x -axis. |
| Perimeter | distance around the boundary of the region. |
| Solidity | Ratio of pixels in the convex hull to the pixels that are also in the region. |

Table 4.2: Region features.

4.4.2 Line transformation

A line is a set of successive pixels that are eight-connected. Eight connect meaning that the pixels are successive either vertically, horizontally or diagonally. The input of the hyperbola feature extraction algorithm is a map with lines. Therefore, this step transforms the MSER regions from the radargram into lines.

A common technique to get lines is edge detection. Edge detection is the process of finding the boundary of regions with different contrast within an image. The boundaries are detectable by finding discontinuities in brightness. Different well-known edge detectors are Canny, LoG, Sobel, Prewitt and Roberts (Maini and Aggarwal, 2009).

The MSER regions are regions of uniform intensity. Hence, the edge of a MSER region describes the point where energy transverses to either more or less energy. Ideally, the line describes the point of the local extrema in the region, which is far from the edges. Consider the following example of regions in the bucket of figure 4.11. Different colours denote all MSER regions. Note that regions overlap, meaning that one region might continue under another one. Interpret the regions as a contour map, where the inner contour with the darkest colour describes local extrema, in this case negative. When interpolating contour plot into a 3D Gaussian the minima is the exact middle point of the contour plot which is farthest away of all edges of the contour. Therefore, the line transformation should detect the central line of the region and not the edges.

Lee et al. (1994) propose a thinning algorithm called skeletonise, which preserves the topological and the geometrical conditions. Such a line is advantageous for the hyperbola algorithm since the line still resembles the shape of the region. The algorithm transforms all regions, including overlapping regions into lines.

Inherently, high-intensity reflections contain more lines than homogeneous areas in the radargram. The local extrema contain more overlapping MSER regions. Since every MSER region results in one line, these areas contain relatively more lines. Therefore, the probability of finding a hyperbola at local minima or maxima is higher in comparison to a region in the radargram with a more homogeneous pixel intensity. Figure 4.12 depicts a line plot with a high concentration of similar lines in regions overlying the high intensities of the corresponding radargram, see figure 4.10.



Figure 4.12: Line plot of the MSER regions.

4.4.3 Hyperbola feature extraction

A hyperbola detection is performed on the lines of the regions acquired by the MSER algorithm. Hyperbola detection is based on an automated detection algorithm for GPR reflection hyperbolas (Mertens et al., 2016). The hyperbola in a radargram, see figure 4.13, is described by equation 4.9.

$$\frac{t^2}{a^2} - \frac{|y - y_0|^2}{b^2} = 1 \quad (4.9)$$

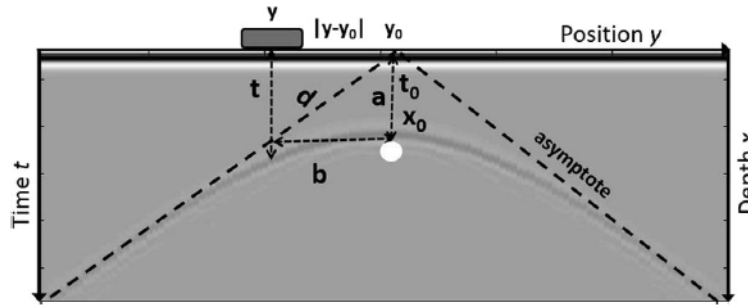


Figure 4.13: Physical representation of hyperbola parameters (Mertens et al., 2016).

Figure 4.13 depicts hyperbola its apex location coordinates are y_0 on the y -axis and x_0 on the x -axis. The distance between point y to the apex of the hyperbola on the y -axis equals $|y - y_0|$. The TWTT of the emitted signal is t . a is the antenna position of the apex on time t -axis and b is the location of the asymptotic on the y -axis where $t = a$. The equation is rewritten using the velocity in a medium based on the speed of light c and the relative permittivity ϵ_r resulting in equation 4.10 (Mertens et al., 2016).

$$1 = \frac{t^2}{t_0^2} - \frac{|y - y_0|^2}{\frac{t_0^2}{4} \left(\frac{c}{\sqrt{\epsilon_r}}\right)^2}; \quad a = t_0; \quad b = (t_0/2)(c/\sqrt{\epsilon_r}) \quad (4.10)$$

The relative permittivity ϵ_r is ranging between a pre-defined minimum ϵ_r^{min} and maximum ϵ_r^{max} , resulting in hyperbolas with different slopes for the same apex location. Comparing the theoretical hyperbola with the line plot results in numerous hyperbola features. The resulting features describe the similarity of the lines to the theoretical hyperbolas. Whenever a feature is not in between the accepted values, no hyperbola is detected, and the algorithm continues to the next pixel of the line. Figure 4.14 depicts a flow chart describing steps of hyperbola feature extraction.

The first feature is the relative density ($drel_i$) between the theoretical hyperbola and the line on each side of the apex $i = 1, 2$. Divide the number of pixels within the buffer zone of the theoretical hyperbola (N_{data}) by the number of pixels used to describe the theoretical hyperbola (N_{theo}), see equation 4.11. Accepted value are $50 \geq drel_i \geq 270$.

$$drel_i = 100 \cdot \frac{N_{data}}{N_{theo}} \quad (4.11)$$

The second hyperbola feature is the normalised symmetry (*sym*) of the region line. The symmetry is the absolute difference between the relative densities of each side of the hyperbola normalised by the highest relative density, see equation 4.12. Accepted values are $sym \geq 0.6$.

$$sym = \left(1 - \frac{|drel_1 - drel_2|}{\max(drel_1, drel_2)} \right) \quad (4.12)$$

The third feature is the relative data range (drt_i and dry_i) on the vertical (t) and horizontal axis (y). Both ranges are equal to the difference between pixel ranges in the region line and the range of the theoretical hyperbola. Calculate the difference between the data range of theoretical hyperbola try_i and the absolute difference in the theoretical buffer zone $adry_i$. Divide the resulting difference by trt_i leading to equation 4.13 for the respective axis. Accepted values are $drt_i \geq 0.7$ and $dry_i \geq 0.7$.

$$drt_i = 1 - \frac{trt_i - adrt_i}{trt_i}; \quad dry_i = 1 - \frac{try_i - ady_i}{try_i} \quad (4.13)$$

The difference in distribution (*dis*), between the region line and the theoretical hyperbola (*th*) comes forth out of histogram operations *hist()*. Sum the absolute difference between each bin of the histogram. Divide the results of the sum by the number of data points (n_{hist}) in the histogram. Accepted values are $dis \geq 0.7$.

$$dis = \frac{\sum_1^{n_{hist}} |hist_n(data) - hist_n(th)|}{n_{hist}} \quad (4.14)$$

The final hyperbola feature is the relative number of pixels surrounding the theoretical buffer zone (*sur*). The surrounding number of pixel (N_{sur}) divided by the number within the buffer zone (N_{buf}) leads to equation 4.15. Accepted value are $sur \leq 1.20$.

$$sur = 1 - \frac{N_{sur}}{N_{buf}} \quad (4.15)$$

A hyperbola is accepted when it fits the criterion for each feature. Since the area around the region line its apex might contain multiple accepted hyperbolas, post-processing is necessary to reduce the clutter of hyperbolas to ideally one. A mask is located with its centre point at apex a and has a height of $[t_a - T, t_a + T]$ and width $[y_a - 5cm, y_a + 5cm]$ where $T = 1/f_{GPR}$. If multiple

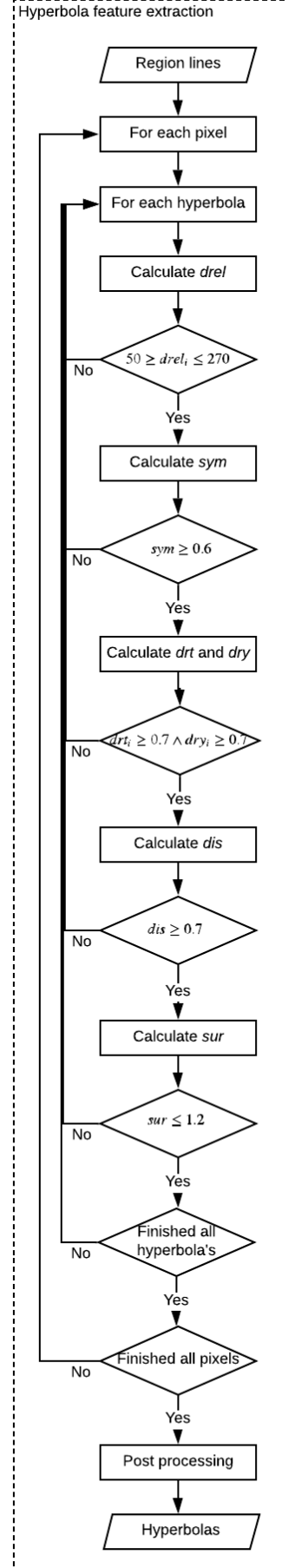


Figure 4.14: Flowchart of hyperbola feature extraction.

hyperbolas are within the mask boundaries, the hyperbola with the highest pixel intensity at its apex is considered the real hyperbola.

Besides the features proposed by Mertens et al. (2016), the algorithm extracts some extra features with the potential of characterising voids. A second-order polynomial fit extracts the polynomial parameters ($pol2$, $pol1$, pol), which are useful in describing the hyperbolic curve without the pixels list. The location of the polynomial on the x -axis is from $minx$ to $maxx$. Also, the pixel intensity (int) of the radargram at the location of the apex is a feature.

Subsequently, the algorithm extracts the average velocity of the reflection (vel). The average velocity is a product of the dielectric constant ϵ_r and the speed of light c . Assuming that the subsurface consists out of low-loss material, the velocity equals equation 4.4. When not consider low loss material, the velocity is the result of equation 2.4. Therefore, the average velocity is equal to the mean of the velocity through multiple layers.

Figures 4.15 and 4.16 depict the hyperbolic observation overlaying the line plot and the enhanced radargram respectively. The red line depicts the accepted hyperbola located at the apex denoted by the green dot. Some of the reflections are detected and some not. The reflection resulting from the fifth object, see figure 4.8, is detect perfectly. The other reflections overlap each other so much that the line the hyperbolas are no longer distinctive in the line plot. Hence, the hyperbola extraction algorithm does not consider these lines hyperbolas.

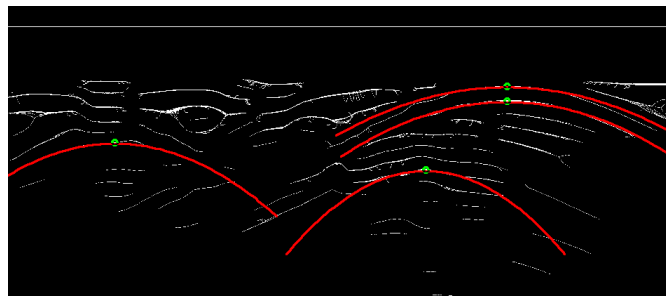


Figure 4.15: Hyperbolic observations overlaying the line plot.

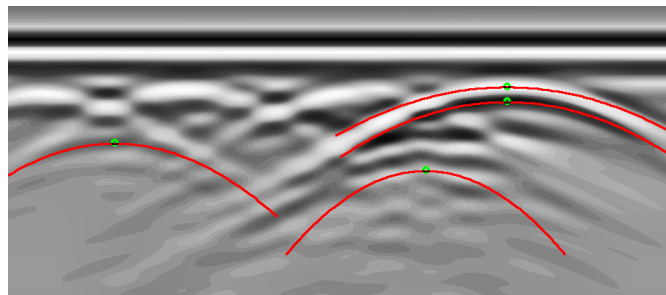


Figure 4.16: Hyperbolic observations overlaying the enhanced radargram.

4.5 Label observations

Each simulated radargram receives labels describing the location of expected reflection as described previously in section 4.3. For the ML problem to be supervised, each observation requires a label. Which label depends on the overlap of the observation with the pre-defined labels in the radargram.

The feature extraction algorithm calculates the features from an observation(O), see section 4.4. The location of the object is generally considered to be located directly under the apex of the hyperbolic reflection. Hence, if the apex and the label (ℓ) overlap, the corresponding observation is considered a reflection of that subsurface object denoted by the label. Therefore, the radargram label is assigned to the observation when the pixel list of the label and the apex of the hyperbola observation intersect, see equation 4.16.

$$\sum (O_{apex} \cap \ell_{pixellist}) \neq 0 \rightarrow O_{class} = \ell_{class} \quad (4.16)$$

$$\ell_{class} \in \{Void, Other, None\}$$

When the pixel lists do not intersect, the comparison continues with a newly selected label. The process repeats until there are no labels left. The observation its class is set to *None* when there is no intersection with any of the labels.

Figure 4.17 depicts the enhanced radargram including labels and hyperbolic observations. Some of the observation its apexes are located inside the label. Therefore, these observations are set equal to the class of the corresponding label. The observations outside the labels are equal to class *None*.

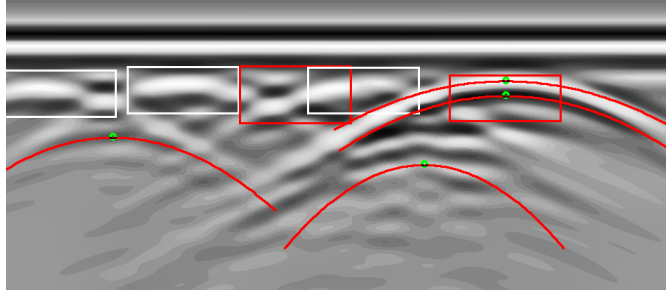


Figure 4.17: Enhanced radargram overlaying by the labels and the hyperbolic observations.

When every observation contains a label, the problem is a supervised classification problem. A supervised classification model uses the observations to train the weights to each feature based on the expected results. Subsequently, a test set is used as input on the trained model to determine the classification performance. The classification outcome is compared to the expected outcome to determine the accuracy of the classification.

4.6 Feature set

A selective set of features should be able to describe the characteristics of voids. It might be that some features are not able to capture the characteristics of the voids, which makes them irrelevant. Ideally, the feature set is as small as possible while still able to distinguish voids from other objects. The process of feature selection calculates the features needed to describe such a set, see section 4.6.1

Not all features contribute to the classification of voids and other objects. Some features might correlate to much and thus be redundant. Other features might not capture a void characteristic and do not contribute to adequate classification. The goal is to determine the features with a high contribution to the future classification of voids. Knowing the significance of features helps to argue for a feature set. Ideally, the accuracy is kept as high as possible while reducing the number of features.

4.6.1 Feature selection

The contribution of features to the classification problem is measurable using different techniques. Yang et al. (2012) propose Neighbourhood Component Feature Selection (NCFS). The NCFS algorithm maximises the expected leave-one-out classification accuracy using gradient ascent with a regularisation term. The regularisation term intends to derive the irrelevant features to zero. The weight corresponding to a feature describes its contribution to the classification. Therefore, the features with a weight close to zero are considered irrelevant features. The higher the weight, the more relevant the feature compared to the others.

The NCFS algorithm is pre-implemented as FSCNCA in the MatLab statistics and machine learning toolbox³. This method includes the regularisation term, different solvers and loss functions as method parameters. Since the number of observation is relatively small $n < 1000$, the method proposes Limited memory Broyden-Fletcher-Goldfarb-Shanno (LBFGS) instead of Stochastic Gradient Descent (SGD) to minimise the loss function and estimate the features. Since the detection of voids is a classification problem, the loss function (l) equals the misclassification error. The objective function describes the minimising the loss function as follows.

$$f(w) = \frac{1}{n} \sum_{i=1}^n \sum_{j=1, j \neq i}^n p_{ij} l(y_i, y_j) + \lambda \sum_{r=1}^p w_r^2; \quad l(y_i, y_j) = \begin{cases} 1 & \text{if } y_i \neq y_j \\ 0 & \text{otherwise} \end{cases} \quad (4.17)$$

Where w are the features, n is the number of observations and p the features. p_{ij} is the probability that x_j is the reference point for x_i (Yang et al., 2012).

The advantage of NCFS is the addition of the regularisation term λ . Tuning the regularisation term results creates the possibility to minimise the loss function by pushing feature weights to zero. As a result, irrelevant features values are set towards zero. Many other techniques, such as Relieff proposed by (Kononenko et al., 1997), do not contain a regularisation term. Hence, Relieff result in a set containing more features and a more homogeneous weight distribution compared to NCFS. Ideally, the feature set has the lowest possible loss function while reducing the number of selected features. Therefore, the features are extracted using NCFS.

³<https://nl.mathworks.com/help/stats/neighborhood-component-analysis.html> (URL accessed on 28-10-2019)

4.7 Classification model

The selected features should have a similar or increased classification accuracy compared to all features. A dummy classification model determines the difference in classification accuracy between all features and the selected features. The observations obtained by the feature extraction algorithm are divided into a training and test set. The dummy model uses the training set to determine the feature weights. The test set evaluates the model performance. The process repeats twice, once using all features of the observations and once with only the selected features. Figure 4.18 depicts the flowchart of the classification model.

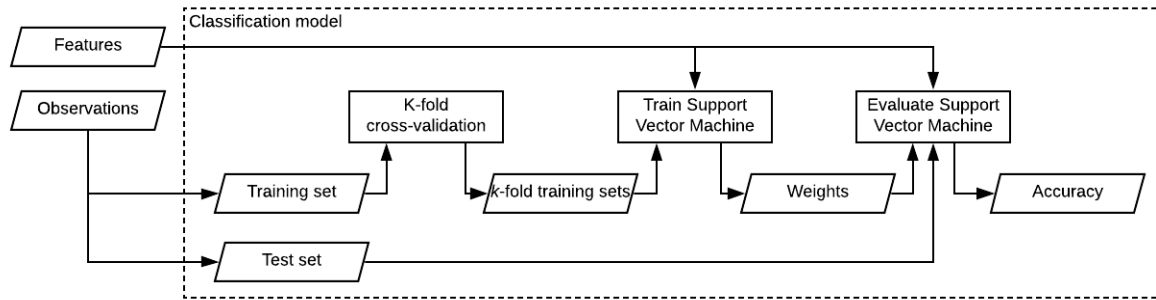


Figure 4.18: Flowchart of the classification model.

Every observation has a label (ℓ), given by the labelling algorithm, see section 4.3. The labels are a void, a different object or not an object, $\ell_{class} \in \{Void, Other, None\}$. Hence, the input (features) and output (labels) are prior knowledge making it a supervised classification problem.

4.7.1 K-fold cross-validation

The k -fold Cross Validations (KCV) technique splits all extracted observations in training and test sets (Arlot and Celisse, 2010). Generally, a classification model trains on 70-80% of the observations. The rest of the observations is for evaluating the model, called the test set. The goal of the training set is to minimise the loss functions.

The dataset is "folded" k times resulting in k training iterations called splits and dividing the training set in k folds, see figure 4.19. For each split, a new test fold is selected, resulting in $k-1$ times the folds for training (green) and one fold for testing (blue) the split. The weights are trained and evaluated for each split. When the splits are processed, the initial test set is used as input to determine the model performance.

Figure 4.19 depicts an example of 5-folds cross-validation. Note that the process of KCV only uses the training set to process the splits and excludes the test set for the final evaluation. Hence, the model does not get biased by training on evaluation data. The goal of excluding the test set from the training set is to reduce the probability of over-fitting the model. Training on the test observations results in a biased model that recognises the test data.

KCV is a re-sampling procedure since it uses observations more than once. The procedure extends the original dataset, and it is therefore efficient for problems containing relatively small

⁴https://scikit-learn.org/stable/modules/cross_validation.html (URL accessed on 28-10-2019)

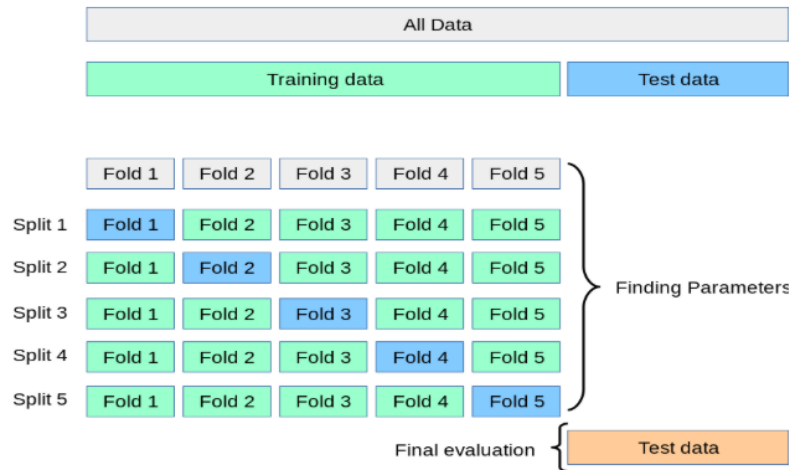


Figure 4.19: Visualisation of a KCV.⁴

datasets. The explained 5-fold cross-validation is used to train the classification model during the future tests, see chapter 5.

4.7.2 Support Vector Machine

The dummy classifier is a multi-class Support Vector Machine (SVM). As the name implies, a multiclass SVM is able to classify multiple classes, whereas a normal SVM can only predict two classes, making it a binary problem.

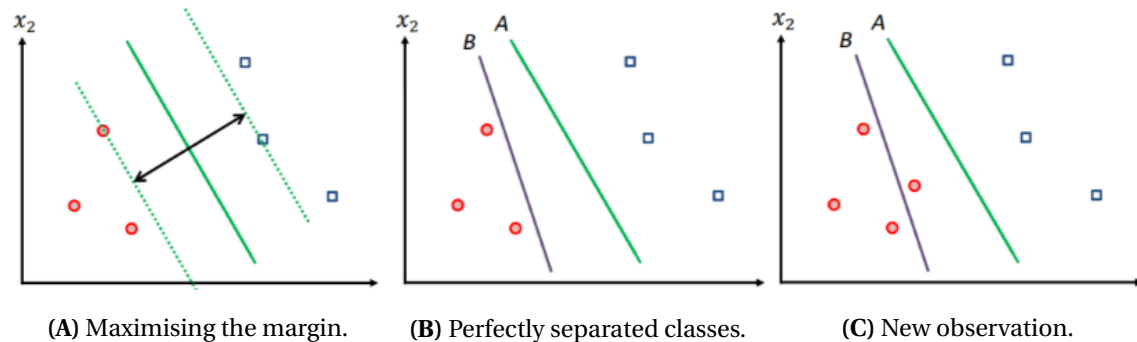


Figure 4.20: Visualisation of a SVM.

A SVM is a linear binary classifier. A SVM separates classes by drawing a linear boundary between observations. The goal of this boundary is to minimise the classification loss of new observations by placing the linear boundary between the two clusters of classes. Figure 4.20B depicts boundary A and B between the red and blue classes. Both boundaries have a loss function of zero since they can separate the two classes linearly.

Intuitively, the most reliable boundary would be between the classes and as far away from both classes as possible. The process of determining such a boundary is called maximising the margin between two classes, see figure 4.20A. Maximising the margin reduces the probability of misclassification of new observations. Figure 4.20C depicts a newly classified red observation which is wrongly classified by B but correctly classified by the line A, which maximises the margin.

Although the graphs in figure 4.20 describe a binary classification problem, the principle is similar for multiple classes. Generally speaking, SVM are effective on high dimensional data. The multi-class SVM is chosen as dummy classifier because the amount of training data is still limited, and there is a relatively high number of features.

The multi-class SVM is implemented by using MatLab *fitcecoc* functionality. The Error-Correcting Output Codes (ECOC) model reduces the multi-class problem to a set of binary classification problems. The coding design describes how each binary classification problem aggregates to describe the total multi-class problem. The one-versus-one coding design for three classes looks as follows.

| | Learner 1 | Learner 2 | Learner 3 |
|---------|-----------|-----------|-----------|
| Class 1 | 1 | 1 | 0 |
| Class 2 | -1 | 0 | 1 |
| Class 3 | 0 | -1 | -1 |

Table 4.3: Coding design scheme.

Note that the learners in table 4.3 are individual SVMs. The first step is for learner one to train on positive class one and negative class two. The second step is to keep the class one positive and select the following negative class, in this case, the third class. When all negative classes are evaluated, the next positive class is selected. When the SVM is trained, it classifies based on the calculated boundaries. The test set is fed into the model to determine the evaluation. Finally, confusion matrices and other accuracy metrics can represent the performance of the model numerically.

4.7.3 Final deliverable

In summary, this chapter presents an algorithm to extract features from a radargram. Also, it describes the steps to calculate a useful feature set and classification accuracy. The process consists of two parts: calculating features from the radargrams and using the features to calculating the feasible feature set plus the classification accuracy.

Firstly, CLAHE enhances the contrast of the radargram. The labels in the radargrams are calculated using the environmental variables of the *gprMax* input files for both topologies. The feature extraction process calculates the MSER regions and region features. Subsequently, the process skeletonises the regions into lines for calculating the hyperbolas. The hyperbolas lead to calculating the hyperbola features and combine them with the features of the corresponding region into observations. A comparison between the observation location and label locations results in observations receiving the appropriate class *Air*, *Other*, or *None*. Hence, the dataset is labelled and applicable to supervised ML models.

Secondly, the process of feature selection calculates the feature set of both topologies minimising the loss function. A multi-class SVM uses the selected feature to determine the classification accuracy. The resulting feature weight and classification performance help to argue for the most feasible features for the classification of voids directly behind the sewer wall using GPR from within sewers.

⁴<https://nl.mathworks.com/help/stats/fitcecoc.html> (URL accessed on 28-10-2019)

5 Results and discussion

This chapter consists of three separate. All the results are discussed to determine if the outcome is as expected and why. The results of each test help to argue for a different research question. Each test starts with an introduction of the research questions and works towards the results answering the question.

5.1 Test 1: experimental versus simulation radargram results

This first test calculates the hyperbolas from the experimental dataset and the respective simulated dataset. The goal of this test is to get insight in research question one. *How do experimental radargrams correlate with simulated radargrams when measuring voids behind a concrete sewer wall?* This test measures the correlation between the experimental and simulated radargrams in terms of the difference in the number of observations matching radargram labels. The environments of the simulated radargrams approximate the environments of the experimental setup from the feasibility study, see section 3.1. Appendix B includes the layout of the environments.

One of the issues emerging from the feasibility study on gprMax is the optimal resolution of the model, section 3.1. The Yee cell size can not be smaller than 2mm since it will exceed the maximum GPU memory. The 2mm Yee cell size does not match the experimental radargram, which is acquired using a 1mm, 3mm and 10mm trace spacing. Therefore, it is most efficient to acquire the simulated and the experimental results with a 3mm Yee cell size and 3mm trace spacing. Hence, the experimental and simulated radargrams both have a 3mm trace spacing which satisfies the trace spacing constraint, meaning that the minimum trace spacing should be equal to the Yee cell size or a multitude. Thus, gprMax simulates 2D and 3D simulations of the experimental environments with a 3mm Yee cell size and 3mm trace spacing, making them comparable.

The parameters of the feature extraction algorithm are identical regardless of the type in radargram. As a result, the performance of the feature extraction purely depends on the input radargram, thus the 2D/3D simulated or experimental radargrams. Therefore, the number of observations matching a label is comparable for the three respective types of radargrams.

Not all experimental radargrams contain visible reflection. Visual inspection leads to a subset only including the experimental radargrams containing a visible reflection. The subset consists out of radargrams from environments: 3, 5, 6, 10, 11, 14, 18, 19, 20, 22 and 26. Hence, the results of this test depend on this subset of radargrams.

Ideally, hyperbolic observations describe the same reflections over all types of radargrams. The result would be a similar amount of observations inside and outside the labels for all three types of radargrams. In this case, the feature extraction algorithm does not make a distinction between the 2D/3D simulated or experimental radargrams as input. Therefore, the feature extraction algorithm can generalise the simulations to the experimental results. The generalisation is in terms of the number of observations inside and outside the labels.

The feature extraction algorithm calculates hyperbolas from all radargrams. The hyperbolas occur inside or outside the labels, see figure 5.1 and 5.2 respectively. The box denotes the inner

25% to 75% of the sample values, also called the interquartile range. The lines extending above and below the box describe the maximum and minimum values. The red line is the median, and the red plus depicts outliers defined by more than 1.5 times the interquartile range away from the top or bottom of the box.

Important to note when interpreting the results is that labels might overlap. One observation might occur inside two labels and theoretically, describe two objects. If an observation identifies with two labels, it counts as one detection. Hence, an observation can detect not more than one label. Therefore, the number of observations inside and outside the label sum to the total number of found observations.

Figure 5.1 shows that most observations within the label are found in 2D and 3D simulated radargrams, followed by the experimental setup. The difference between 2D and 3D simulations are the interquartile ranges, which is overall higher for the 2D simulations. The detailed feasibility study in appendix C mentions a reduction in reflection intensity when changing from 2D to 3D simulations. The feasibility study assumes that the difference in intensity is the result of energy loss in the third dimension. The additional dimension results in more energy attenuation and less energy reflected to the GPR. The reduction in energy leads to less contrast in 3D radargram, which makes them reflections more challenging to distinguish. Therefore, 3D simulations might result in fewer observations inside the label compared to 2D simulations.

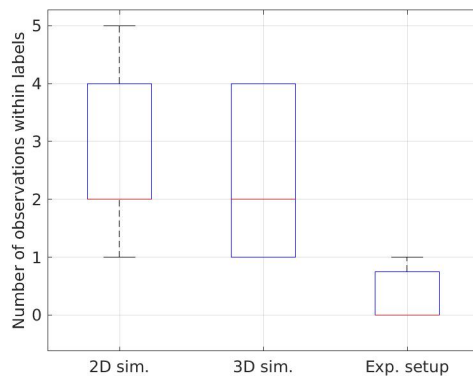


Figure 5.1: Box plot describing the number of observations found inside the labels.

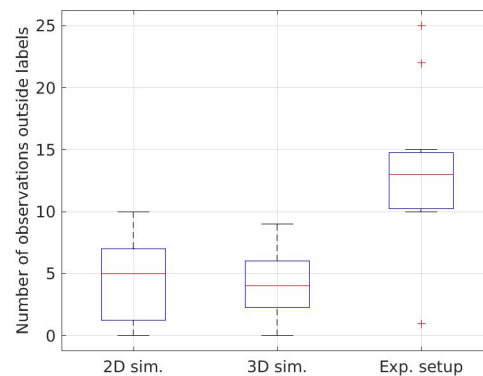


Figure 5.2: Box plot describing the number of observations found outside the labels.

The box plot in figure 5.2 shows the highest median of the number of observation outside the labels for the experimental setup. The value goes slightly down for the 2D and 3D simulations, respectively. Interestingly, the 2D and 3D simulations depict a higher number of observations within the labels and a low number outside the label. Contradictory the experimental setup describes a low number of observations within the label and a high number outside the label. Therefore, the numbers of labels inside and outside the label for simulated and experimental radargrams seem to be in an inverse relationship. For more details about the results per test environment, see the bar graphs in appendix D.

The number of observations outside the labels drastically increase for the experimental setup. This behaviour might be the result of additional noise in experimental radargrams. The noise appears in the lower parts of the experimental radargram. The research focuses on detecting void directly behind the sewer wall. Hence, the labels occur in the higher parts of the radargram. As a result, the observations from the hyperbolic noise fall outside of the labels. Figure 5.3 depicts the experimental radargram of test 19 including the hyperbolic observations. As

described, most observations are in the lower parts of the radargram resulting from hyperbolic noise.

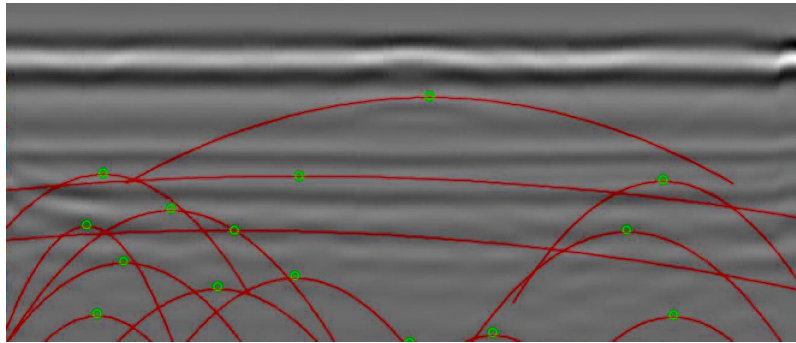


Figure 5.3: Experimental radargram of test 19 including the hyperbolic observations denoted as red lines.

The radargrams from the experimental setup are significantly different from simulated radargrams. Experimental radargrams show less distinctive reflections compared to simulated radargrams. Also, the experimental radargrams show more hyperbolic noise in the lower parts of the radargrams which might be the result of additional noise and artefacts due to the GPR system, non-homogeneous soil and subsurface objects and objects surrounding the experimental setup. The reduction in the distinctiveness of the reflections and addition of noise makes it difficult for the feature extraction algorithm to find observations located inside the labels.

The feature extraction algorithm does not generalise properly between simulated and experimental radargrams in test one. Multiple observations inside the labels are found for the simulated radargrams and near to none for the experimental radargrams. For now, the feature extraction algorithm seems not sensitive enough to find well-hidden hyperbolic reflections of the experimental radargrams.

Extensive tuning of the accepted ranges of the region and hyperbola features might increase the level of generalisation between simulated and experimental results. The accepted ranges should include reflections and exclude as much noise as possible. A parameter optimisation algorithm like a random search or grid search might improve the performance of the feature extraction algorithm. The goal of the feature extraction algorithm is to find as many observations in all labels of the radargram. A future consecutive classification model can determine which observation is correct and which not. It is essential to understand that whenever the feature detection algorithm does not find an observation, it does not participate in the training of a classification model, and it does not contribute to improving the classification accuracy.

Unfortunately, the experimental dataset is too small and results in too little observations to determine feature weights (195 observations of class *None*, but only three of class *Air* and one of class *Other*). The total number of observations is scarce, and the number of observations within the labels is near zero. Therefore, the algorithm is not able to properly reduce any loss function. As a result, it is impossible to determine the feature weights of the simulated radargrams and the experimental radargrams. Therefore, research into the optimal size of the experimental dataset is vital to choose a classification model to optimally classify experimental data in the future.

5.2 Test 2: feature selection on planar and in sewer model

This test determines which void characteristics and features prove to have a significant contribution to the detection of voids. The goal of this test is to answer the following two research questions. *What is the difference between radargrams of a planar and cylindrical topology when measuring voids behind a concrete sewer wall? What features in GPR radargrams are most feasible for void detection?.* This test describes the research question in the respective order.

Firstly, one of the difference between the radargrams of a planar and cylindrical topology is the reverberation pattern introduced in the cylindrical topology. The effect of the reverberation pattern is visible when extracting the MSER regions for the planar and cylindrical topology, see figure 5.4 and 5.5 respectively. The cylindrical topology results in significantly more regions. Each region results in a line for hyperbola feature extraction. Hence, there is a higher probability of hyperbolic noise in the cylindrical topology. As a result, the cylindrical topology relies more on different features to be able to filter out the noise created by the reverberation pattern. Therefore, the cylindrical topology is considered to be more complex than the planar topology.

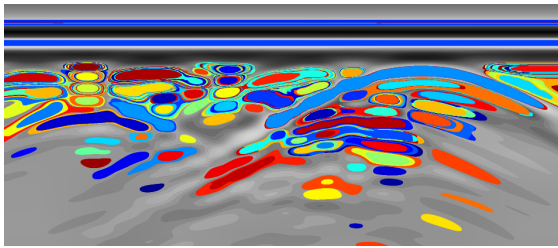


Figure 5.4: Planar topology contrast enhanced radargram plus MSER regions.

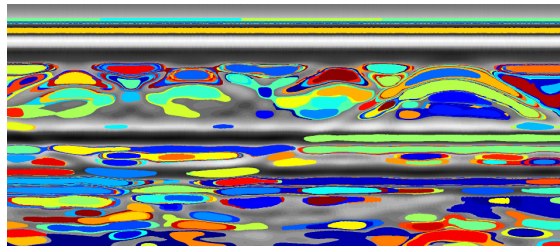


Figure 5.5: Cylindrical topology contrast enhanced radargram plus MSER regions.

Additionally, movement of the GPR and the length of the reflection path of a reflection are different for the cylindrical topology than for the planar topology. As a result both topologies percept subsurface objects differently which makes for different reflections, see the reflections in the planar and cylindrical topology, figure 5.6 and 5.7 respectively. The planar topology shows wider shaped hyperbolic reflections. The cylindrical topology depicts reflections that have slightly less intensity and are a bit steeper compared to the reflections in the planar topology. The lower intensity of the cylindrical topology might be the result of a longer length of the reflection path, which results in more attenuation and thus weaker reflection. The steeper hyperbolas are the result of the non-linear motion of the GPR in the cylindrical topology. Because of the motion, the distance between the GPR and the subsurface object increases quicker, resulting in an increasing TWTT and a steeper hyperbola. Steeper hyperbolas result in less overlap in the corresponding reflections. Therefore, the visual distinctiveness of the reflections is considered higher for the cylindrical topology than the planar topology.

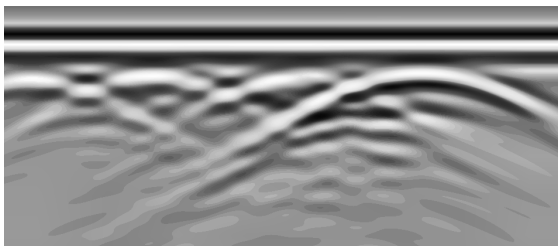


Figure 5.6: Planar topology contrast enhanced radargram.

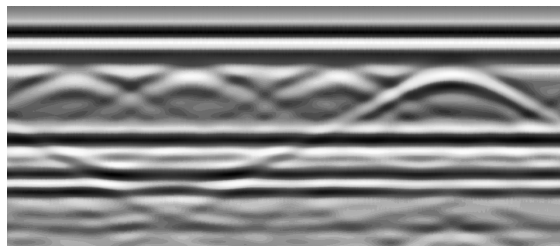


Figure 5.7: Cylindrical topology contrast enhanced radargram.

Secondly, in order to determine the feasible features, a dataset of 250 2D radargrams is simulated in gprMax for each the planar and cylindrical topology. The planar radargrams are approximations of the cylindrical radargrams, see section 3.3. The previous paragraph shows that the reflections of both topologies are not identical but similar to some extent. Hence, it seems assumable that the selected feature set is also, to some extent, similar.

In order to examine this hypothesis, firstly detect the number of observations for both topologies overall 250 simulations. The numbers should give a general idea of how the algorithm generalises over both topologies. Subsequently, the numbers might help to explain any future classification results. Therefore, it is vital to know the number of observations inside and outside the labels for both topologies.

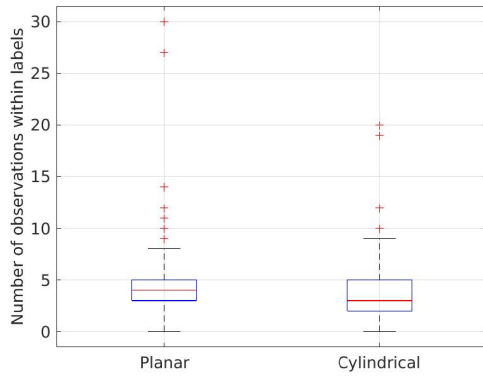


Figure 5.8: Box plot describing the number of observations inside the labels.

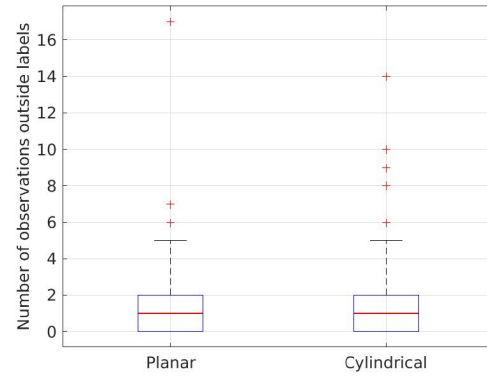


Figure 5.9: Box plot describing the number of observations outside the labels.

Figure 5.8 is a box plot depicting the number of observations within the labels. The median over 250 simulations equals three for both typologies. Hence, on average only three observations fall into labels. Note that every simulation contains exactly five subsurface objects. Do not draw the wrong conclusion by stating that observations depict three out of five labels. Multiple observations might fall into the same label.

The cylindrical topology contains a lower median compared to the planar topology. Over 250 simulations, fewer observations overlap with labels for the cylindrical topology compared to the planar topology. It might be that the feature extraction algorithm performs better on the planar topology. The cylindrical topology might introduce artefacts being extracted as hyperbolic observations which could coincidentally overlap with labels. Still, the difference in the numbers inside and outside the labels is near zero.

Figure 5.9 depicts the number of observations outside the labels. There is no difference in the median of the number of observations between the planar and cylindrical topology since both topologies have a median equal to one. Also, the interquartile ranges for both topologies are identical. Consider the assumption made in test one, section 5.1: "the difference in trends between the number of observations inside and outside the labels seems to be an inverse relationship". If the assumption would hold, then the number of observations outside the label for the cylindrical should go down compared to the planar topology. Since this is not the case, it strengthens the idea that the cylindrical topology might introduce artefacts being extracted as hyperbolic observations coincidental overlapping labels.

The second part of this test aims to determine the feature weights for both topologies. To make the features comparable, it is essential to have a heterogeneous dataset. A homogeneous data-

set will result in a biased classification model that only classifies one class. The dataset is not inherently heterogeneous since the feature extraction algorithm does not necessarily find the equal amount of observation for different classes. To ensure a heterogeneous dataset, it is crucial to know the size of the less frequent class in both topologies. The dataset is enforced to contain the same number of randomly selected observation for each class. For this test, the smallest class is *Air*, containing 319 observation over 250 radargrams. Therefore, the dataset contains the same number of observations (319) for the three classes, setting the total number of observations equal to 957 observations per topology.

The NCFS model calculates the feature weights by maximising the leave-one-out classification accuracy using gradient ascent with a regularisation term, see section 4.6.1. Tuning the regularisation term λ results in a set of features with the lowest possible loss function. Figures 5.10 and 5.11 depict the normalized maximum square error loss to the lambda for the planar and cylindrical topology respectively. Tuning λ for the planar topology results in a minimum loss of 0.19 and 0.15 for the cylindrical topology.

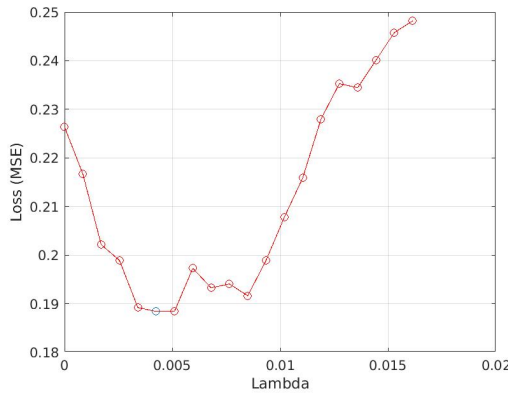


Figure 5.10: Minimising the loss function by tuning regularisation term for the planar topology.

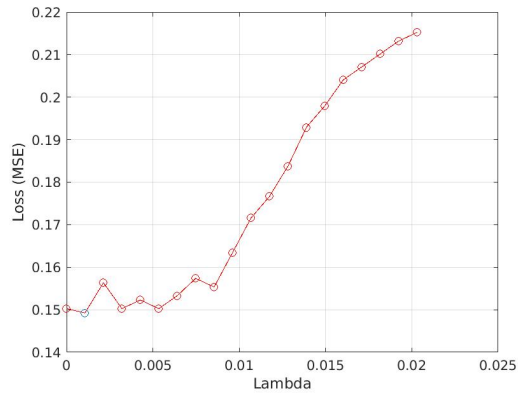


Figure 5.11: Minimising the loss function by tuning regularisation term for the cylindrical topology.

Both topologies result in different selected feature sets. Table E.1 in appendix E includes the feature weights. The features depicted by green are included in the selected feature set of the specified topology. Many of the features are considered irrelevant for both of the topologies. The irrelevant region features are: *Area*, *Centroid₁*, *ConvexArea*, *Perimeter*, *MaxFeretDiameter*, *Sym*, *Drt₂*, *Dis₁*, *Eps*, *Pol2*, *Pol1*, *Pol*, *MinX*, *MaxX*. Since these features are irrelevant for both topologies, they are considered insufficient in the classification of voids for further research.

Similarly, table E.1 includes the features relevant for both topologies. The features relevant for both topologies are: *Centroid₂*, *MajorAxisLength*, *MinorAxisLength*, *Orientation*, *EquivDiameter*, *Solidity*, *Extent*, *ApexY*, *Drt₁*, *Dis₂*, *Int*, *Velocity*. Since these feature are relevant for both topologies they are considered the features able to generalise between the planar and cylindrical topology. Therefore, these features are called topology independent features set.

Subsequently, some features are relevant to only one of the two topologies. The features are for the planar topology are: *Circularity*, *MinFeretDiameter*, *ApexX*, *Dry₁*, *Dry₂*, *Sur*. For the cylindrical topology the features are: *Eccentricity*, *MaxFeretAngle*, *MinFeretAngle*, *Drel₂*. These features are relevant in minimising the loss function for the corresponding topology and are called the topology dependent features set. Since the features are only relevant for one topology, they are not able to generalise between both topologies.

Figure 5.12 depicts the normalised weights for the planar and cylindrical topology respectively. Table E.1 includes the values of the feature weights for both topologies. The specific weights are bound to the regularisation parameters. Therefore, when selecting another λ , the weights of the features will change.

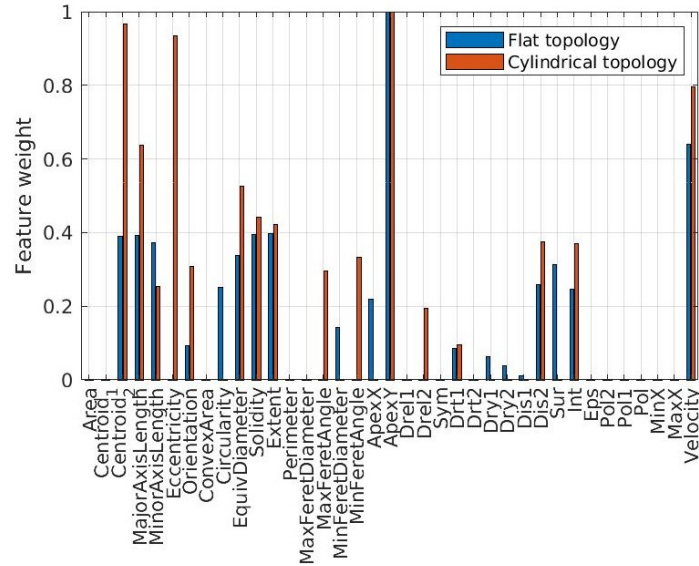


Figure 5.12: Feature weights of the planar and cylindrical topology scaled between zero and one.

Before directly comparing the weights, note that the weights in each graph are relative to each other. The bar graph depicts clear differences in the features sets. For both topologies, the location of the hyperbola its apex on the y -axis of the radargram (*ApexY*) has the highest weights. Since all subsurface objects are simulated directly behind the sewer wall, it seems assumable that *ApexY* is essential for the classification of classes. If the observation is not located in the range directly behind the sewer wall, it can not be of class *Air* or *Other*, and thus must be of class *None*. Therefore, feature *ApexY* is of utter importance in the classification of objects directly behind the sewer wall.

Interestingly, for the planar topology, all other features have a relatively low feature weight compared to *ApexY*. The relative high *ApexY* in the planar topology could be the result of observations having similar hyperbolic shapes. In the simulated training data, the subsurface objects are all circles with roughly a similar size on a similar depth, see figure 4.3. Since all reflections have similar shape sources, the hyperbola shape is the same, but the location and the reflection coefficient of the hyperbolas differs. As a result, many features can not distinguish reflection since they describe similar shaped hyperbolas. Thus, features that are invariant of the shape, such as *ApexY*, get a relatively higher weight in the planar topology.

The hypothesis of the relatively high *ApexY* in the planar topology raises the question of why this is not the case in the cylindrical topology. Similarly to the planar topology, the cylindrical topology also includes identical subsurface objects and thus also result in similar hyperbolic shapes. Interestingly, the cylindrical topology the *ApexY* is less distinctive since *Centroid2*, *Eccentricity* and *Velocity* got similar feature weights. One of the possible arguments for the difference in feature weights distribution compared to the planar topology might be occurring reverberation pattern in the cylindrical topology, see section 3.3. The reverberation pattern in the cylindrical topology creates additional reflections which are no longer only the result of the subsurface objects. The reflections result in a variety of differently shaped hyperbolas while the planar topology results in similarly shaped hyperbolas. The different shapes result in a broader

range of hyperbola feature values for observations of the cylindrical topology. As a result, more different features are necessary to minimise the classification loss function. Therefore, more features receive higher weights in the cylindrical topology.

The most feasible type of features equals the set with the highest total weight. Table 5.1 includes the weight of the region and hyperbola features for both topologies. The selected features stand for all features denoted by green in table E.1 for the topology. The topology independent type only considers features relevant for both topologies. Looking at the values reveals that the sum of the weights of the hyperbola features stays similar for both topologies, contradictory the region features. The sum of the weights of the region features is significantly higher in the cylindrical topology compared to the planar topology. In the cylindrical topology, the region features have a higher contribution to the classification in terms of weights. Therefore, the region features are considered most feasible for the cylindrical topology. The regions and hyperbola features are equally feasible for void detection in the planar setup.

| | Planar topology | Cylindrical topology |
|---|-----------------|----------------------|
| Selected region features | 7.7574 | 14.5747 |
| Selected hyperbola features | 8.0259 | 8.0633 |
| Topology independent region features | 6.6565 | 10.1257 |
| Topology independent hyperbola features | 6.2483 | 7.5123 |

Table 5.1: Sum of the region and hyperbola features per feature set and topology.

In summary of this test, firstly, the main differences between radargrams of a planar and cylindrical topology are the reverberation pattern introduced in the cylindrical topology. As assumed in section 3.3, the reverberation pattern does not overlap with reflections from objects directly behind the sewer wall. These findings suggest that the reverberation pattern does not reduce the classification performance for the cylindrical topology. Additionally, the non-linear motion of the GPR results in different reflection shapes and intensities for both topologies. The steeper shape hyperbola in the cylindrical topology result in less overlap between the reflection and increase the distinctiveness of the individual reflections. It is, therefore, possible that the hyperbolic observation from the cylindrical topology matches the reflection in the radargram with greater detail than the observations from the planar topology. As a result, it might be that the classification accuracy of the cylindrical topology is higher compared to the planar topology.

Secondly, the feature in the selected feature set varies per topology. The topology independent features are: *Centroid₂*, *MajorAxisLength*, *MinorAxisLength*, *Orientation*, *EquivDiameter*, *Solidity*, *Extent*, *ApexY*, *Dr_{t1}*, *Dis₂*, *Int*, *Velocity*. Additionally, the topology dependent features reduce the loss function of the corresponding topology. For the planar topology the additional features are relevant: *Circularity*, *MinFeretDiameter*, *ApexX*, *Dry₁*, *Dry₂*, *Sur*. For the cylindrical topology the additional features are relevant: *Eccentricity*, *MaxFeretAngle*, *MinFeretAngle*, *Drel₂*.

Determining the most feasible type of features for the classification problem is based on the sum of the weights of the selected features. The planar topology results in an even distribution of the weights of regions and hyperbola features. The total weight feature of the cylindrical topology consists for %70 regions features and %30 of hyperbola features. Therefore, the region features seem slightly more feasible in the detection of voids. However, both regions and hyperbola features are required to minimise the classification loss function to get an optimal classification accuracy.

5.3 Test 3: detection accuracy by feature selection

This test calculates the detection accuracy of both topologies for all features and the selected features resulting from test two, see section 5.2. The goal of this test is to help answer the following research questions. *What is the difference in detection accuracy between all features and the selected feature set? What is the difference in detection accuracy on a planar and cylindrical topology?*

The model calculating the classification accuracy is a multi-class SVM, for more information on the model see section 4.7. The SVM is trained on the heterogeneous dataset described in the previous section by using 5-fold cross-validation, see section 4.7.1. Figure 5.13 and 5.14 depict the confusion matrices of the planar and cylindrical topologies respectively. Table 5.2 contains the numerical performance metrics accompanying the confusion matrices.

The confusion matrices consist out of two parts, the left 3×3 matrix being the actual confusion matrix and the right 2×3 matrix denoting the percentage of classification accuracy per class. The confusion matrix describes the number of classifications outcome of all classes. For example, consider the confusion matrix of figure 5.13A. The second element of the first row equals 31, meaning that 31 classification of *None* were actually of class *Air*. Note that the diagonal of the confusion matrix describes correct classification and the other elements of a row depict the wrong classifications. The colour blue stands for correct classifications and the colour red stands for wrong classifications. The darker colour stands for more numerous classifications. For example, when an element is dark blue, there are relatively more correct classifications than wrong classifications.

Figure 5.13A and 5.13B depicts the confusion matrices for the planar topology using all features and the selected features, respectively. The confusion matrix for the planar topology using all features shows the classifications of the class *None* are most accurate, followed by class *Air* and *Other*. There is only a small difference in performance using all features or the selected features, compare both confusion matrices in figure 5.13. The performance of classification using the selected features goes slightly down for *Air* and slightly up for *Other* and *None*. The difference in performance is considered insignificant. Therefore, the performance of the planar topology using all features and the selected features is near to identical.

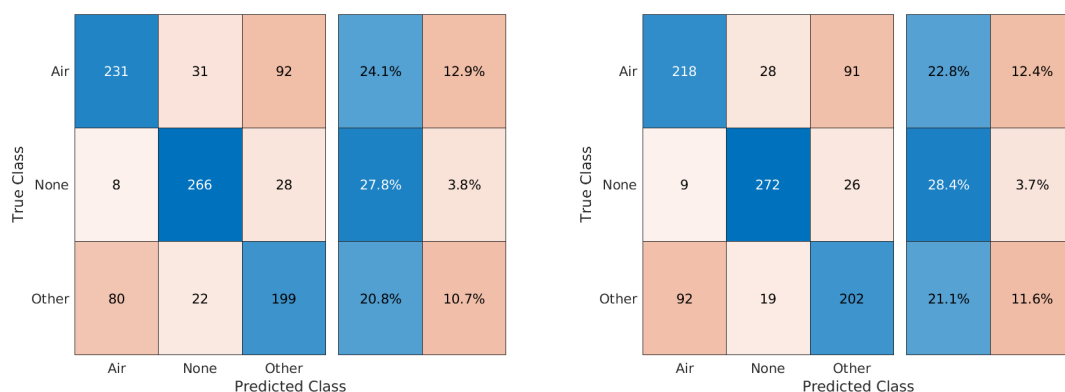
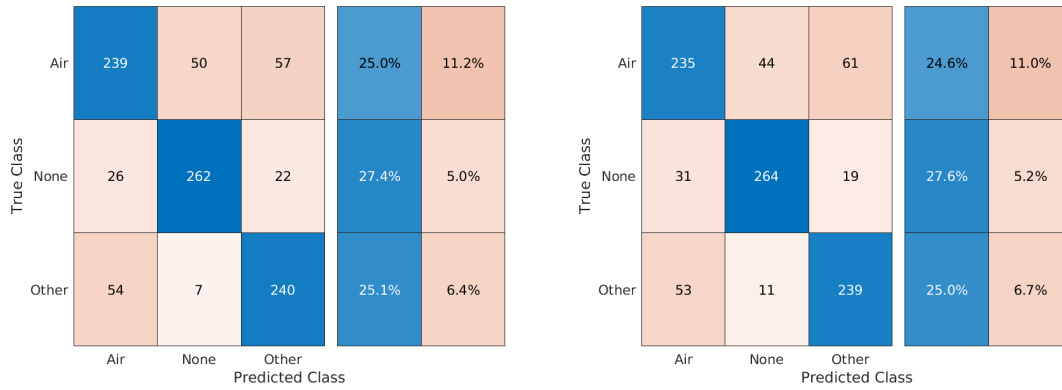


Figure 5.13

Figure 5.14A and 5.14B depicts the confusion matrices for the cylindrical topology using all features and the selected features, respectively. Note that the performance is slightly better than the planar topology described previously. Similar to the result of the planar topology, the best-classified class is *None*, followed by similar results for *Other* and *Air*. As a result, the similarity in confusion matrix values makes it necessary to visualise the performance more compactly.



(A) Confusion matrix of the cylindrical topology using all features. (B) Confusion matrix of the cylindrical topology using the selected features.

Figure 5.14

Table 5.2 describes normalised performance metrics to give a clear visualisation of the performance between both topologies and the different feature sets. The prevalence denotes the ratio of class occurrences in the dataset. The multi-class model describes three classes in a heterogeneous dataset. Hence, the number of occurrences of each class in a dataset are identical. Therefore, the prevalence is equal to the ratio of class occurrences in the dataset, which is always 1/3 of the total dataset.

| | Planar all features | Planar selected features | Cylindrical all features | Cylindrical selected features |
|--------------|---------------------|--------------------------|--------------------------|-------------------------------|
| Correct rate | 0.727 | 0.723 | 0.774 | 0.771 |
| Error rate | 0.273 | 0.277 | 0.226 | 0.229 |
| Prevalence | 0.333 | 0.333 | 0.333 | 0.333 |
| Sensitivity | 0.834 | 0.853 | 0.821 | 0.828 |
| Specificity | 0.944 | 0.945 | 0.925 | 0.922 |

Table 5.2: Model accuracy metrics for the different feature sets of both topologies.

The sensitivity, also known as the true positive rate or the recall, describes the ratio of a true classification on a positive label. The specificity, also known as the true negative rate, describes the ratio of true classification on a negative label. Equation 5.1 depicts the sensitivity and specificity. Figure 5.15 depicts the model accuracy values of both topologies using all features and the selected features. The radar plot makes for easy comparison in the performance difference of the feature sets for both topologies. The radar plot shows there is almost no difference between the planar topology using all features and the planar topology using the selected features set.

$$sensitivity = \frac{TruePositive}{ActualPositive}; \quad specificity = \frac{TrueNegative}{ActualNegative} \quad (5.1)$$

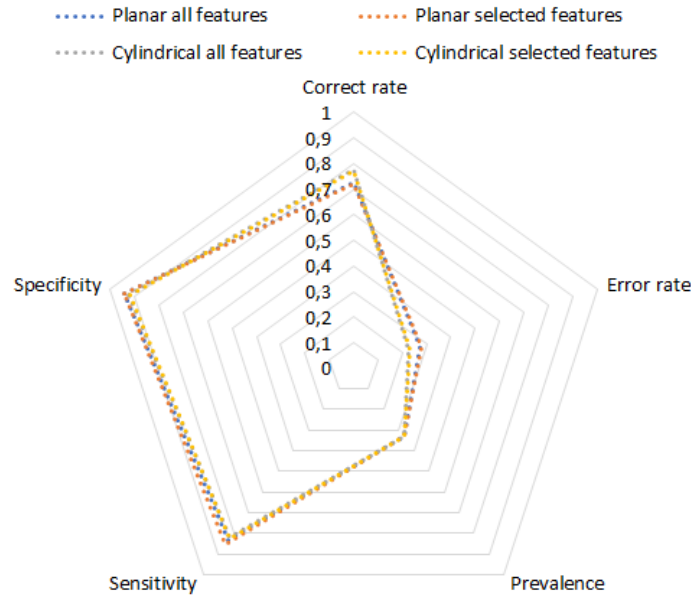


Figure 5.15: Radar plot depicting the accuracy metrics of the different topologies.

To conclude, firstly, the difference in detection accuracy between all features and the selected features is near to none. The radar plot shows that the values of the correct rate for the different feature sets of a topology are almost identical. The difference in detection accuracy between using all features and the selected feature set for both topology are close to none. The performance is as expected since the irrelevant feature weights are put near to zero, meaning that they are inherently not taken into account during classification. Therefore, the selected feature sets are capable of reducing the number of features of the classification problem without losing significant accuracy.

Secondly, the radar plot shows that the performance between the planar topology and the cylindrical topology are almost identical. The positive identical performance results come forth of the feature sets being able to capture the characteristics of the subsurface objects, including voids. Since the results are identical for both topology, it seems that the feature extraction algorithm can generalise between the planar topology and cylindrical topology. Therefore, the algorithm shows potential in extracting the useful features om experimental setups similar to the simulations in this research.

The cylindrical topology has a 4–5% higher correct rate than the planar topology. Test 2, section 5.2, mentions the increasing distinctiveness of reflection in the cylindrical topology as a result of the steeper shape of the reflection. Furthermore, it assumes that the classification accuracy of the cylindrical topology might be higher. This test confirms indeed that the classification accuracy of the cylindrical topology is slightly higher than the planar topology. However, it is not proven that increasing distinctiveness of the reflections in the cylindrical topology directly influences the classification accuracy.

6 Conclusion and Recommendations

This chapter draws the conclusions of this research directly related to the research questions, see section 6.1. Additionally, recommendation to this research and future research are given in section 6.2.

6.1 Conclusion

Firstly, this section answers each research questions using the results of the analysis and the tests. Secondly, a more general conclusion overarches the individual research questions.

1. *How do experimental radargrams correlate with simulated radargrams when measuring voids behind a concrete sewer wall?*

Test one aims to determine how the experimental and simulated radargram correlate. The correlation is expressed in the number of observations inside and outside the label. Firstly, test one proves that the number of observations inside and outside the labels significantly differ for the 2D and 3D radargrams, and experimental radargrams. The different distinctiveness in reflections for the different types of radargrams might be the cause. When a reflection is below a certain level of distinctiveness, the feature extraction algorithm is not able to detect the reflection as an observation. The only correlation in terms of numbers of observations inside and outside the labels is an inverse relationship. Hence, when there are more observations inside the label, there seem to be less outside the label.

The experimental radargram dataset is too small for feature selection. Feature extraction on the experimental radargrams results in only a few observations of the three classes. Therefore, there are not enough observations for feature selection or calculating the classification accuracy using the SVM.

2. *What is the difference between radargrams of a planar and cylindrical topology when measuring voids behind a concrete sewer wall?*

Test two, section 5.2, and the analysis of the wave propagation in section 3.3, give insight into this research question. The analysis proves that a cylindrical topology and the approximate planar topology create different radargrams. The cylindrical topology contains a reverberation pattern and has a non-linear GPR motion which results in a different length of the reflection path through the sewer wall. These findings influence the reflections in multiple different ways.

The reverberation pattern creates a band of noise in the radargram. As assumed in the analysis, the reverberation pattern does not overlap with reflections of voids and objects directly behind the sewer wall. Therefore, the finding suggests that the reverberation pattern does not influence the classification performance for the cylindrical topology.

The non-linear GPR motion in the cylindrical topology results in steeper shaped hyperbolas. The steeper shape makes for less overlap of reflections in the radargram and thus better distinctiveness of the reflections. Additionally, the longer length of the reflection path through the sewer is a result of the wave not entering perpendicular on the sewer wall. As a result, the signal attenuates more, and the intensity of the reflections get slightly weaker than in the planar topology. The difference in intensity is small and therefore neglected in terms of distinctiveness of the reflections. Hence, it may be the

case that the classification performance of the cylindrical topology increases as a result of the increasing reflection distinctiveness by having less overlap in reflections.

3. *What features in GPR radargrams are most feasible for void detection?*

The most feasible characteristics of voids depend on the topology of the problem. The topology independent features are: *Centroid₂*, *MajorAxisLength*, *MinorAxisLength*, *Orientation*, *EquivDiameter*, *Solidity*, *Extent*, *ApexY*, *Drt₁*, *Dis₂*, *Int*, *Velocity*. Additionally, the topology dependent features reduce the loss function of the corresponding topology. For the planar topology the additional features are: *Circularity*, *MinFeretDiameter*, *ApexX*, *Dry₁*, *Dry₂*, *Sur*. For the cylindrical topology the additional features are: *Eccentricity*, *MaxFeretAngle*, *MinFeretAngle*, *Drel₂*. Determining the most feasible features is based on the sum of the selected features.

The most contributing type of feature is the type with the highest total weight of features. The sum of the hyperbola feature weights stays similar for both topologies, contradictory to the region feature weights. The sum of the region feature weights equals $\sim 70\%$ on the total feature weights in the cylindrical topology, compared to $\sim 30\%$ in the planar topology. In the cylindrical topology, region features have a higher contribution in terms of weights to the classification problem. The results suggest that the region features are most feasible for the cylindrical topology. The regions and hyperbola features are equally feasible for void detection in the planar topology. However, both regions and hyperbola features are required to extract observations and minimise the classification loss function to get an optimal classification accuracy.

4. *What is the difference in detection accuracy between all features and the selected feature set?*

The radar plot in figure 5.15 reveals that the values of the correct rate for the different feature sets of a topology are almost identical. The difference in detection accuracy between all features and the selected feature set for both topology is close to none. Hence the selected feature sets are capable of reducing the number of features of the classification problem without losing significant accuracy. The performance is as expected since the irrelevant feature weights are put near to zero, meaning that they are inherently not taken into account during classification using all features.

5. *What is the difference in detection accuracy on a planar and cylindrical topology?*

The cylindrical topology has a correct classification rate of $\sim 0.73\%$ and the cylindrical topology of $\sim 0.78\%$. The feature extraction is executed with the same parameters to be able to compare the results. Therefore, the feature extraction algorithm generalises well between the planar topology and the cylindrical topology.

The similar performance strengthens the assumption that reflections of the object directly behind the sewer wall and the reverberation pattern do not overlap. If they would, it seems assumable that the reverberation pattern would hide the reflections of the sub-surface objects, which would result in lower classification accuracy for the cylindrical topology. Hence, when looking for objects directly behind the sewer wall, the reverberation pattern does not necessarily have to be an issue.

Surprisingly, the cylindrical topology even has a slightly higher 4 – 5% classification accuracy than the planar topology. The difference might be the result of the better distinctiveness of reflection in the cylindrical topology mentioned in section 5.2. As a result, the dataset of the cylindrical topology might contain higher quality observations. The observations are easier to classify and, therefore, the classification accuracy of the cylindrical topology is better than the planar topology. However, it is not proven that this is the result of the steeper hyperbolas. Despite the assumption, more research is necessary to prove this.

To summarise, this research lays the brick for future research in the field of in-sewer GPR to obtain info about surrounding objects. It is difficult to conclude about the feasibility of GPR to inspect defects from within an underground sewer. The research concludes that classification results for a planar and cylindrical topology are almost identical. Since GPR is already useful for inspection on the ground surface, it seems that there is potential in inspection from within underground pipes as well. However, there are still many recommendations to take into account and topics to research before it is possible to use GPR for in-sewer inspection. Section 6.2 describes the recommendations and topics on this work and future research.

The findings prove that GPR contains the potential for the use of in-sewer inspection. Not only will in-sewer GPR reduce the necessity of excavation, the GPR is also located closer to the sewer wall compared to a ground surface GPR measurement. Hence, the approach has the potential to inspect with a higher resolution, which results in a more detailed inspection.

Additionally, multiple different techniques, such as image processing and ML apply to this inspection problem. Applying these techniques introduces numerous possibilities to advance in the inspection problem. Currently, most inspections are executed manually by an inspection specialist. It does not seem very reasonable to replace this inspection specialist with a black box inspection algorithm. Ideally, the inspection algorithm and inspection specialist work together in a symbiotic relationship. For example, the algorithm visualises (intermediate) results to the inspection specialist, who can evaluate the performance and tune the algorithm to increase performance. Subsequently, the ML algorithm trains and inspects the incoming radar-grams, which can be a second opinion to the inspection specialist or the inspection specialist is a second opinion of the algorithm. Such a symbiotic relationship between the algorithm and inspection specialist would bridge the step toward fully automatic sewer inspection.

6.2 Recommendations

The recommendations is divided into two sections. Firstly, the recommendations directly applicable to the research in this project, see section 6.2.1. Secondly, the recommendations on future research relating this topic, see section 6.2.2.

6.2.1 Current research

Firstly, after finishing the feasibility study, it appeared that the emitted waveform of the 1Ghz GPR system, used in the experimental setup, was not of the expected type. The waveform should resemble the exponential growth and decay of a capacitor with a pulse duration of 4ns (Daniels, 2004), instead of a Ricker wave of this research. A different waveform results in different reflective behaviour during the simulations with gprMax. Therefore, it would be interesting to redo the simulations and feasibility study with a virtual GPR matching the waveform to the waveform of the GPR in the experimental setup.

Secondly, gprMax contains different options to create more realistic simulations. The lack of knowledge concerning the state of sewers and material science have been the main arguments in disregarding research towards realistic simulations. A natural progression of this work is to team up with a specialist to analyse the relevant sewers and design a more realistic and matching simulation. Two interesting options to make the simulations more realistic are adding the dispersion of the signal and the introduction of soil mixture models. gprMax includes the dispersion of the signal by using multi-pole Debye, Lorentz of Drude formulations. Subsequently, the mixing model for soils might be useful to create soils with more realistic dielectric and geo-

metric properties. Therefore, gprMax contains options yet to be explored to create realistic simulations in further research.

Thirdly, Giannakis et al. (2018) is taking steps to simulate radargrams using ML in gprMax. The approach overcomes the time-consuming FDTD approach for simulations. Using this technique might lead to significantly bigger radargram dataset in a small amount of time. Therefore, when planning to acquire radargrams in the near future, it might be feasible first to research the steps made in the field of simulating radargrams using ML.

Fourthly, the proposed algorithm needs adjusting and tuning to be able to create an optimal dataset for classification. Test 1, section 5.1, concludes that the algorithm has difficulties finding regions in experimental radargrams. Tuning the pre-processing step to increase contrast in locations of interest is a logical first step in tuning the algorithm. This research focuses on voids directly behind the sewer wall, which results in reflections high up in the radargram. It is possible to incorporate the prior knowledge of the objects under detection in the pre-processing step. For example, the reflections of voids directly behind the sewer wall would occur high up in the radargram. Therefore, the pre-processing should enhance the contrast up in the radargram and gradually decrease when going down in the radargram. An advantage of this technique is that noise, which often occurs more down in the radargram, is not enhanced. Also, reducing the number of found MSER regions reduces the calculation of observations and thus drastically decreases the processing time.

Fifthly, despite the promising results, it would be good to tune the parameters to optimise the performance of the feature extraction algorithm. Sensitivity analysis helps to get insight into the influences of the parameters to the output of the algorithm. During this research, the parameter sets have been identical for both topologies to keep the results comparable. Specific tuning for the cylindrical topology might increase the quality of the observations. The classification accuracy may get higher with the right parameter settings for the feature extraction algorithm. Future studies on the sensitivity analysis will be needed to determine the potential increment of the classification accuracy.

6.2.2 Future research

Firstly, this research focuses on capturing the characteristics of voids in a set of features. Although the research uses a dummy classification model, its only aim is to show the relative accuracies between both topologies using the different feature sets. Therefore, one of the next steps is to select a ML model able to increase the classification accuracy. Existing work on combining GPR and ML introduces the numerous possibilities in combining both. The solutions varying from using environmental parameters in logistic regression model Koo and Ariaratnam (2006), assessing railway-ballast conditions using Support Vector Machine (SVM) on GPR data Shao et al. (2011), and the use of Artificial Neural Networks (ANN) to segment regions of interest classified by additional models Travassos et al. (2018). An important decision when working with ML models for classification in radargrams is where to draw the boundary between image processing and ML. For example, relative simple ML models, such as SVM, require extracted features (image processing) such as the features mentioned during this research. Contradictory, more complex models such ANN do not necessarily need extracted features, but instead thrive on higher dimensional data such as all pixels in the radargram. Hence, it is challenging to create a combination optimal for the problem. Therefore, extensive research needs to be conducted to determine an optimal configuration for the void classification problem.

Secondly, another recommendation relating further research is to develop a labelled experimental dataset consisting of radargrams. The small experimental dataset used during this research is not big enough to use for feature selection or classification. A proper sized experimental dataset allows for feature selection and classification, which was not feasible in this research. Also, using these results, it would be possible to compare the results of the algorithm between simulated and experimental radargrams, instead of only simulations. Such a comparison would give better expectations of what to expect in real sewers, instead of only making approximate simulations. Hence, the difference in feature sets between cylindrical simulated and experimental data can be compared. Ideally, the feature sets match, meaning that the approximate simulations depict the experimental dataset in terms of features. Additionally, a classification model might prove useful to determine the classification accuracy for both the experimental and simulated radargrams.

Thirdly, evaluate the labelled experimental dataset, introduced in the previous recommendation, by visualising the intermediate steps of the feature extraction algorithm and show them to and radargram inspection specialist. This method reveals the difference in perception between the inspection specialist and the algorithm. Subsequently, the algorithm can be tuned to increase the level of perception according to the perception of the specialist. Additionally, the algorithm might assist the specialist during inspection to increase the inspection quality. Therefore, the next possible step is to create an algorithm which assists the inspection specialist and can adapt to the level of perceptions.

Bibliography

- P. F. Alcantarilla, A. Bartoli, and A. J. Davison. KAZE features. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, volume 7577 LNCS, pages 214–227. Springer, Berlin, Heidelberg, 2012. ISBN 9783642337826. doi: 10.1007/978-3-642-33783-3{_}16.
- S. Arlot and A. Celisse. A survey of cross-validation procedures for model selection. *Statistics Surveys*, 4:40–79, 2010. ISSN 19357516. doi: 10.1214/09-SS054.
- H. Bay, T. Tuytelaars, and L. Van Gool. SURF: Speeded up robust features. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, volume 3951 LNCS, pages 404–417. Springer, Berlin, Heidelberg, 2006. ISBN 3540338322. doi: 10.1007/11744023{_}32.
- N. J. Cassidy, R. Eddies, and S. Dods. Void detection beneath reinforced concrete sections: The practical application of ground-penetrating radar and ultrasonic techniques. *Journal of Applied Geophysics*, 74(4):263–276, 8 2011. ISSN 09269851. doi: 10.1016/j.jappgeo.2011.06.003.
- D. J. D. J. Daniels. *Ground penetrating radar fundamentals*. The Institution of Electrical Engineers, London, 2nd edition, 2004. ISBN 0 86341 360 9.
- J. P. Davies, B. A. Clarke, J. T. Whiter, and R. J. Cunningham. Factors influencing the structural deterioration and collapse of rigid sewer pipes. *Urban Water*, 3:73–89, 2001. ISSN 14620758. doi: 10.1016/S1462-0758(01)00017-6.
- I. Giannakis, A. Giannopoulos, and C. Warren. A Machine Learning Approach for Simulating Ground Penetrating Radar. In *2018 17th International Conference on Ground Penetrating Radar, GPR 2018*, 2018. ISBN 9781538657775. doi: 10.1109/ICGPR.2018.8441558.
- C. Harris and M. Stephens. A Combined Corner and Edge Detector. In *Proceedings of the Alvey Vision Conference 1988*, pages 1–23, 1988. doi: 10.5244/C.2.23.
- L. Kofman, A. Ronen, and S. Frydman. Detection of model voids by identifying reverberation phenomena in GPR records. *Journal of Applied Geophysics*, 59:284–299, 2006. ISSN 09269851. doi: 10.1016/j.jappgeo.2005.09.005.
- I. Kononenko, E. Šimec, and M. Šikonja. Overcoming the Myopia of Inductive Learning Algorithms with RELIEFF. *Applied Intelligence*, 10:39–55, 1997. URL <https://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.56.4740>.
- D. H. Koo and S. T. Ariaratnam. Innovative method for assessment of underground sewer pipe condition. *Automation in Construction*, 15(4):479–488, 7 2006. ISSN 09265805. doi: 10.1016/j.autcon.2005.06.007.
- T. Lee, R. Kashyap, and C. Chu. Building Skeleton Models via 3-D Medial Surface Axis Thinning Algorithms. *CVGIP: Graphical Models and Image Processing*, 56(6):462–478, 11 1994. ISSN 1049-9652. doi: 10.1006/CGIP.1994.1042. URL <https://www.sciencedirect.com/science/article/pii/S104996528471042X>.
- S. Leutenegger, M. Chli, and R. Y. Siegwart. BRISK: Binary Robust invariant scalable keypoints. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 2548–2555, 2011. ISBN 9781457711015. doi: 10.1109/ICCV.2011.6126542.
- R. Maini and H. Aggarwal. Study and Comparison of Various Image Edge Detection Techniques. *International Journal of Image Processing*, 3(1):1–11, 2009. ISSN 1985-2304.
- J. Matas, O. Chum, M. Urban, and T. Pajdla. Robust wide-baseline stereo from maximally stable extremal regions. In *Image and Vision Computing*, volume 22, pages 761–767. Elsevier, 2004.

doi: 10.1016/j.imavis.2004.02.006.

- L. Mertens, R. Persico, L. Matera, and S. Lambot. Automated Detection of Reflection Hyperbolas in Complex GPR Images with No A Priori Knowledge on the Medium. *IEEE Transactions on Geoscience and Remote Sensing*, 54(1):580–596, 2016. ISSN 01962892. doi: 10.1109/TGRS.2015.2462727.
- J. M. Reynolds. *An Introduction to Applied and Environmental Geophysics*. John Wiley & Sons, Ltd., 2nd edition, 2011. ISBN 9780471485353. doi: 10.1071/pvv2011n155other.
- E. Rosten and T. Drummond. Machine learning for high-speed corner detection. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, volume 3951 LNCS, pages 430–443. Springer, Berlin, Heidelberg, 2006. ISBN 3540338322. doi: 10.1007/11744023{_}34.
- E. Rublee, V. Rabaud, K. Konolige, and G. Bradski. ORB: An efficient alternative to SIFT or SURF. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 2564–2571, 2011. ISBN 9781457711015. doi: 10.1109/ICCV.2011.6126544.
- W. Shao, A. Bouzerdoum, S. L. Phung, L. Su, B. Indraratna, and C. Rujikiatkamjorn. Automatic classification of ground-penetrating-radar signals for railway-ballast assessment. *IEEE Transactions on Geoscience and Remote Sensing*, 49(10 PART 2):3961–3972, 10 2011. ISSN 01962892. doi: 10.1109/TGRS.2011.2128328.
- J. Shi and C. Tomasi. Good features to track. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition CVPR-94*, pages 593–600. IEEE Comput. Soc. Press, 1994. ISBN 0-8186-5825-8. doi: 10.1109/CVPR.1994.323794.
- X. L. Travassos, S. L. Avila, and N. Ida. Artificial Neural Networks and Machine Learning techniques applied to Ground Penetrating Radar: A review. *Applied Computing and Informatics*, 2018. ISSN 2210-8327. doi: 10.1016/J.ACI.2018.10.001.
- S. van der Walt, J. L. Schönberger, J. Nunez-Iglesias, F. Boulogne, J. D. Warner, N. Yager, E. Gouillart, and T. Yu. scikit-image: image processing in Python. *PeerJ*, 2, 2014. ISSN 2167-8359. doi: 10.7717/peerj.453.
- P. van der Werf, G. Dekker, and Ambient. Graafschade voorkomen bij complex rioleringswerk. Technical report, RIONED, Ede, 2017.
- C. Warren, A. Giannopoulos, and I. Giannakis. gprMax: Open source software to simulate electromagnetic wave propagation for Ground Penetrating Radar. *Computer Physics Communications*, 209:163–170, 2016. ISSN 00104655. doi: 10.1016/j.cpc.2016.08.020.
- C. Warren, A. Giannopoulos, A. Gray, I. Giannakis, A. Patterson, L. Wetter, and A. Hamrah. A CUDA-based GPU engine for gprMax: Open source FDTD electromagnetic simulation software. *Computer Physics Communications*, 237:208–218, 2019. ISSN 00104655. doi: 10.1016/j.cpc.2018.11.007.
- P. Wonink and L. van der Werf. Onderzoek instorting vrijvervalriolen. Technical report, RIONED, Ede, 2019.
- W. Yang, K. Wang, and W. Zuo. Neighborhood component feature selection for high-dimensional data. *Journal of Computers*, 7(1):162–168, 1 2012. ISSN 1796203X. doi: 10.4304/jcp.7.1.161-168. URL <http://ojs.academypublisher.com/index.php/jcp/article/view/5076>.
- K. Yee. Numerical solution of initial boundary value problems involving maxwell’s equations in isotropic media. *IEEE Transactions on Antennas and Propagation*, 14(3):302–307, 1966. ISSN 0018-926X. doi: 10.1109/tap.1966.1138693.
- R. Yelf and D. Yelf. Where is True Time Zero? *Electromagnetic Phenomena*, 7(1):158–163, 2006.
- K. Zuiderveld. Contrast Limited Adaptive Histogram Equalization. In P. S. Heckbert, editor, *Graphics Gems*, pages 474–485. Academic Press, 4th edition, 1 1994. ISBN 9780123361561.

doi: 10.1016/b978-0-12-336156-1.50061-6. URL <https://www.sciencedirect.com/science/article/pii/B9780123361561500616>.

A Appendix 1 - Code repository read me

This appendix includes the read me of the code repository¹ containing all software created during this research. The read me describes the directory structure and files included in the repository. Additionally, it gives a few examples of how to run tests and generate data. The repository is located on the RaM gitlab server of the University of Twente.

¹<https://git.ram.ewi.utwente.nl/delftm/thesis> (URL accessed on 6-11-2019)

Thesis Matthijs van Delft

This repository contains software accompanying the MSc thesis: "Towards feature-based underground void detection with Ground Penetrating Radar from within sewers using Image Processing".

Getting Started

These instructions will get you a copy of the project up and running on your local machine for development and testing purposes. See deployment for notes on how to deploy the project on a live system.

Note that all code is developed to be executed on a High Performance Computing (HPC) clusters. Simulations are offloaded the GPUs on the HPC cluster and the feature extraction process is scalable over multiple cores depending on the available nodes on the cluster. Scheduling on the HPC cluster is executing by the SLURM workload manager and the execution resources are defined in the `.SBATCH` files.

For any details about the algorithm see the thesis.

Prerequisites

The repository contains software with the following dependencies:

- Simulation of GPR by `gprMax`: <http://www.gprmax.com/>
- Software language R: <https://www.r-project.org/about.html>
- Conversion of experimental radargrams (.sgy) to images (.png) using: <https://emanuelhuber.github.io/RGPR/>
- Developed feature extraction algorithm in MatLab: <https://nl.mathworks.com/products/matlab.html>
 - Additional Image Processing Toolbox: <https://nl.mathworks.com/products/image.html>
 - Additional Parallel Computing Toolbox: <https://nl.mathworks.com/products/parallel-computing.html?requestedDomain=>

Simulation

GPR simulation software `gprMax` is used to simulate the radargrams. For test 1 the radargrams are in directory `simulation/test1` and for test 2 and 3 in directory `simulation/test2_3`.

Test 1 consists out of experimental radargrams with approximate 2D and 3D simulated radargrams. The input files are included in `simulation/test1`. To re-simulate the input file, run the following command from the simulation directory.

```
#SBATCH schedule_gprmax.sbatch exp_2D_26T 483
```

`#SBATCH` is the command the schedule a job using Slurm Workload Manager. The first argument is the script that is scheduled. The second argument is the name of the input file that is simulated. The third argument is the number of steps in the simulation. The number is steps is calculated by the desired travel distance divided by the step size. Note, for test 1 the number of steps is always equal to 483.

Test 2 consists out of five-hundred radargrams for both the planar and cylindrical topology. All the radargrams are simulated and can be found in directory *simulation/test2_3*. Run the following command from the simulation directory to simulate all data again.

```
#SBATCH job_generate_data.sbatch 250 5
```

The first argument describes the scheduled script. The second argument denotes the number of arbitrary radargrams per topology, in this case five-hundred. The third argument is the number of randomly positioned subsurface objects behind the sewer wall, which equals five per simulation.

The *.sbatch* file describes the resource used for simulation and needs to be changed depending on the resource on the HPC system. Now the simulation executes on two Nvidia Titan-X GPUs. To change the number and type of resource adjust the following lines of code in file *job_generate_data.sbatch*.

```
#SBATCH --partition="gpu_titan-x" // The resource.
#SBATCH --gres=gpu:2 // The number of resources.
python -m gprMax $filename.in -n 608 -mpi 3 -gpu 0 1
```

The number of scheduled workers *-mpi* should equal the number of resources plus one. The selected GPUs denote by zero to the number of resources minus one.

Image processing

The radargram datasets are processed to calculate the region and hyperbola features. Matlab script *image_processing/test1.m* and *image_processing/test2_3.m* calculate the features for the respective tests. Additionally, the script includes any feature selection and prediction model essential to retrieve the results of tests. Directory *image_processing* contains all Matlab script vital for processing the tests. The test is scheduled on the HPC by the following command from directory *image_processing*

```
#SBATCH matlab.sbatch
```

The test file denotes by the last argument in the following line.

```
matlab -nodesktop -nosplash -nodisplay < test2_3.m
```

To change the number of cores used for calculation change the following line in *matlab.sbatch*.

```
#SBATCH --cpus-per-task=64
```

Subsequently, make sure that the *test*.m* script generates and equally sized parallel pool by selecting an equal amount of cores in the following line.

```
parpool('local',64)
```

Both scripts start with the extraction of the features. The features should only be recalculated when the input

radargrams change through simulation, or when the parameters of the feature extraction algorithm are changed. If not, the section denoted by the comment `%% Feature extraction` can be commented out to prevent the redundant calculations.

Note that the feature extraction process is most time-consuming. The feature extraction process is scheduled over multiple cores to reduce the processing time.

Example running test 1

The input files are designed manually to approximate the experimental environments as well as possible. Therefore, there is no need to generate input files for test 1. Simulation of the input files is executed by the execution of the following command for the desired input file. Repeat this for all the desired input files.

```
sbatch schedule_gprmax.sh exp_3D_5T 483
```

Secondly, run the test script *image_processing/test1.m* from directory */image_processing*. Make sure that the right test is selected in the *matlab.sbatch* file

```
#SBATCH matlab.sbatch
```

The data is written to the respective test directory, in this case *simulation/test1*.

Example running test 2-3

Firstly, run the generate and simulate script from the directory */simulation*.

```
#SBATCH job_generate_data.sbatch 250 5
```

Secondly, run the test script *image_processing/test2_3.m* from directory */image_processing*. Make sure that the right test is selected in the *matlab.sbatch* file

```
#SBATCH matlab.sbatch
```

The data is written to the respective test directory, in this case *simulation/test2_3*.

Directory tree

- thesis (repository)
 - compare (directory)
 - experimental_data (directory)
 - Contains all acquired seismological data (.sgy) from the experimental setup.
 - simulation_data (directory)
 - Contains results from the feasibility study, see thesis.
 - compare_experimental.m
 - compare_radargram.m

- compare_simulated.m
- sgy2png.R - Converts the experimental .sgy files to full HD images .png files.
- image_processing (directory)
 - test1.m - Contains all code for test 1 including feature extraction.
 - test2_3.m - Contains all code for test 2 including feature extraction
 - feature_selection.m
 - dummy_classifier.m
 - gather_accuracy.m
 - gather_features.m
 - gpr_deconvolution.m
 - gpr_edge.m
 - label_radargram.m
 - roi_accuracy.m
 - roi_mser.m
 - roi_parabola.m
 - timezero_correction.m
 - matlab.sbatch - Schedules specified test on HPC cluster.
 - dielectrics.mat - Contains dielectric properties of different materials.
- simulation (directory)
 - out2png.py
 - job_generate_data.sbatch - Schedules generation and simulation of radargrams on HPC cluster.
 - job_generate_both2.py - Generate a specified amount of simulation input files. This files is called by "job_generate_data,sbatch".
 - schedule_gprmax.sbatch - Schedule individual input files on HPC cluster. This files is called by "job_generate_data,sbatch".
 - test1 (directory)

Contains all relevant data and results related to test 1 of the thesis.
 - test2_3 (directory)

Contains all relevant data and results related to test 2 and 3 of the thesis.

Authors

Matthijs van Delft
 University of Twente
 7522NB Enschede
 The Netherlands m.vandelft@student.utwente.nl

B Appendix 2 - Experimental setup test log

This appendix includes the log of the test executed with the experimental setup. Note that the log describes all object locations from their top middle point. The experimental setup contains 400mm of dry sand. The GPR permittivity parameter equals five, which is an approximation to the dielectric constant of the used dry sand. The 4mm thick pavement tiles approximate the planar sewer wall on top of the layer of dry sand.

Each test consists out of six runs, two runs done with 10mm trace spacing, followed by 3mm and 1mm respectively. The environment changes per test which, see table B.1 for the log of each separate test. x , y and z are the coordinates which describe the location of the object in the box. Length l , width w , height h and diameter D are the objects dimensions.

| Test | z(cm) | x (cm) | y(cm) | Shape | Material | l(cm) | w(cm) | h(cm) | D(cm) | Comments |
|------|-------|--------|-------|-------------|----------|-------|-------|-------|-------|------------------------------------|
| 1 | NA. | NA. | NA. | NA. | NA. | NA. | NA. | NA. | NA. | Empty |
| 2 | 140 | 22 | 100 | Rectangle | Steel | 11 | 11 | 5 | NA. | Steel block under tiles |
| 3 | 140 | 17 | 100 | Rectangle | Brick | 21 | 10 | 5 | NA. | Brick under tiles |
| 4 | 135 | 17 | 100 | Rectangle | Brick | 21 | 10 | 5 | NA. | Brick deeper |
| 5 | 135 | 22 | 100 | Rectangle | Steel | 11 | 11 | 5 | NA. | Steel deeper |
| 6 | 140 | 18 | 100 | Cylindrical | Wood | NA. | NA. | 14 | 8 | Dry wood log flat under tiles |
| 7 | 140 | 24 | 103 | Parabola | Air | NA. | NA. | 6 | 5 | Small void under tiles |
| 8 | 140 | 22 | 100 | Parabola | Air | NA. | NA. | 12 | 7 | Medium void under tiles |
| 9 | 140 | 20 | 98 | Parabola | Air | NA. | NA. | 13 | 10 | Wider medium void under tiles |
| 10 | 140 | 20 | 98 | Parabola | Air | NA. | NA. | 13 | 10 | Last: more compression and water |
| 11 | 140 | 23 | 72 | Disk | Air | NA. | NA. | 8 | 8 | Last: deeper less water |
| NA. | 140 | 20 | 98 | Parabola | Air | NA. | NA. | 13 | 10 | |
| 12 | 140 | 23 | 72 | Disk | Air | NA. | NA. | 9 | 8 | |
| NA. | 140 | 20 | 98 | Parabola | Air | NA. | NA. | 13 | 10 | |
| 13 | 140 | 23 | 72 | Disk | Air | NA. | NA. | 13 | 9 | Deeper disk |
| NA. | 140 | 20 | 98 | Parabola | Air | NA. | NA. | 13 | 10 | |
| 14 | 140 | 23 | 72 | Disk | Air | NA. | NA. | 13 | 9 | Disk with more water |
| NA. | 140 | 20 | 98 | Parabola | Air | NA. | NA. | 13 | 10 | |
| 15 | 140 | 23 | 72 | Disk | Air | NA. | NA. | 13 | 9 | Same as test 15 but after 1 hour |
| NA. | 140 | 20 | 98 | Parabola | Air | NA. | NA. | 13 | 10 | |
| 16 | 140 | 23 | 72 | Disk | Air | NA. | NA. | 13 | 9 | Disk more water plus little sand |
| NA. | 140 | 20 | 98 | Parabola | Air | NA. | NA. | 13 | 10 | |
| 17 | 140 | 23 | 72 | Disk | Air | NA. | NA. | 13 | 9 | More water and compressed |
| NA. | 140 | 20 | 98 | Parabola | Air | NA. | NA. | 13 | 10 | |
| 18 | 140 | 23 | 72 | Disk | Air | NA. | NA. | 9 | 9 | Add dry sand and compress |
| NA. | 140 | 20 | 98 | Parabola | Air | NA. | NA. | 13 | 10 | |
| 19 | 140 | 23 | 72 | Disk | Air | NA. | NA. | 5 | 9 | Add sand, minimum compression |
| NA. | 140 | 20 | 98 | Parabola | Air | NA. | NA. | 13 | 10 | |
| 20 | 140 | 23 | 72 | Disk | Air | NA. | NA. | 0 | 9 | Fully filled with dry sand |
| NA. | 140 | 20 | 98 | Parabola | Air | NA. | NA. | 13 | 10 | |
| 21 | 140 | 23 | 72 | Disk | Air | NA. | NA. | 0 | 9 | little disk plastic under tile 5 |
| NA. | 140 | 20 | 98 | Parabola | Air | NA. | NA. | 13 | 10 | |
| NA. | 140 | 22 | 132 | Disk | Air | NA. | NA. | 2 | 7 | |
| 22 | 140 | 23 | 72 | Disk | Air | NA. | NA. | 0 | 9 | Disk filled with plastic and water |
| NA. | 140 | 20 | 98 | Parabola | Air | NA. | NA. | 13 | 10 | |
| NA. | 140 | 22 | 132 | Disk | Water | NA. | NA. | 2 | 7 | |
| 23 | 140 | 23 | 72 | Disk | Air | NA. | NA. | 0 | 9 | Smaller disk |
| NA. | 140 | 20 | 98 | Parabola | Air | NA. | NA. | 13 | 10 | |
| NA. | 140 | 23 | 134 | Disk | Water | NA. | NA. | 2 | 5 | |
| 24 | 140 | 23 | 72 | Disk | Air | NA. | NA. | 0 | 9 | Deeper disk |
| NA. | 140 | 20 | 98 | Parabola | Air | NA. | NA. | 13 | 10 | |
| NA. | 140 | 23 | 134 | Disk | Water | NA. | NA. | 4 | 5 | |
| 25 | 140 | 23 | 72 | Disk | Air | NA. | NA. | 0 | 9 | Measure effect of wet tile |
| NA. | 140 | 20 | 98 | Parabola | Air | NA. | NA. | 13 | 10 | |
| NA. | 140 | 23 | 134 | Disk | Air | NA. | NA. | 4 | 5 | |
| 26 | 140 | 23 | 72 | Disk | Air | NA. | NA. | 0 | 9 | Copy test 25 but with dry tile |
| NA. | 140 | 20 | 98 | Parabola | Air | NA. | NA. | 13 | 10 | |
| NA. | 140 | 23 | 134 | Disk | Water | NA. | NA. | 4 | 5 | |
| 27 | 140 | 23 | 72 | Disk | Air | NA. | NA. | 0 | 9 | |
| NA. | 140 | 20 | 98 | Parabola | Air | NA. | NA. | 13 | 10 | |
| NA. | 140 | 25 | 135 | Disk | Water | NA. | NA. | 3 | 5 | |
| 28 | 140 | 23 | 72 | Disk | Air | NA. | NA. | 0 | 9 | Wet sand mix under tile 4 |
| NA. | 140 | 25 | 135 | Disk | Air | NA. | NA. | 3 | 5 | |
| 29 | 140 | 25 | 103 | Parabola | Air | NA. | NA. | 2 | 3 | |
| NA. | 140 | 23 | 72 | Disk | Air | NA. | NA. | 0 | 9 | |
| NA. | 140 | 25 | 135 | Disk | Air | NA. | NA. | 3 | 5 | |
| 30 | 140 | 25 | 103 | Parabola | Air | NA. | NA. | 4 | 5 | |
| NA. | 140 | 23 | 72 | Disk | Air | NA. | NA. | 0 | 9 | |
| NA. | 140 | 25 | 135 | Disk | Air | NA. | NA. | 3 | 5 | |
| 31 | 140 | 25 | 103 | Parabola | Air | NA. | NA. | 4 | 8 | |
| NA. | 140 | 23 | 72 | Disk | Air | NA. | NA. | 0 | 9 | |
| NA. | 140 | 25 | 135 | Disk | Air | NA. | NA. | 3 | 5 | |
| 32 | 140 | 25 | 103 | Parabola | Air | NA. | NA. | 7 | 8 | |
| NA. | 140 | 23 | 72 | Disk | Air | NA. | NA. | 0 | 9 | |
| NA. | 140 | 25 | 135 | Disk | Air | NA. | NA. | 3 | 5 | |

Table B.1: Experimental setup test log

C Appendix 3 - Feasibility study gprMax

The feasibility study on gprMax exploits the potential of gprMax for data simulation of in sewer GPR systems. A comparison between gprMax radargrams and experimental radargrams provides more insight into the pros and cons of both acquisition techniques. A trade-off between the pros and cons support the decision to use gprMax.

The goal of comparing gprMax radargrams and experimental radargrams is to get the highest possible quality radargram by optimizing two main parameters: trace spacing and Yee cell size. First, simulations with different values for trace spacing result in different quality radargrams. The trace spacing describes the interval between each trace. Therefore, a higher trace spacing reduces the samples of the measurement, which leads to a lower resolution. Additionally, the number of emitted waves decreases, and thus the simulation time.

Secondly, the Yee cell size determines the level of discretization of the model. If the cell size is smaller, more cells necessary to describe a specific size model in comparison with bigger cells. The more cells, the closer the simulation is to reality, which means the resolution increases. Subsequently, the number of calculations increases because the wave propagation through more cells leads to increasing simulation time.

The radargrams are numerically compared using MatLab Image Processing toolbox¹. Trials using full and no-reference image quality metrics describe the feasibility of the particular metrics in describing the radargram quality. The gratifying metrics become part of future testing of gprMax and deciding on the feasibility of gprMax.

Full-reference image quality metrics compare the two radargrams with each other and extract a score value describing the level of similarity. The full-reference image quality metrics used are Image Mean Squared Error (IMMSE), Peak Signal to Noise Ratio (PSNR), and Structural Similarity Index (SSIM).

IMMSE calculates the average squared difference between pixel intensities. The results do not align well with the human perception of quality. In the case of IMMSE, lower values mean higher quality. PSNR calculates the ratio of the maximum pixel intensity to the power of distortion. Like IMMSE, it does not align well with the human perception of quality. In the case of PSNR, higher values mean higher quality. SSIM calculates a score based on luminance, contrast, and structure in an image. Unlike IMMSE and PSNR, this correlates well with the human perception of quality.

No-reference quality metrics usually outperform full-reference metrics in terms of agreement with subjective human quality score. PIQE is an opinion-unaware and unsupervised metric. It measures the quality of images with arbitrary distortion by estimating block-wise distortion and measure the local variance of perceptibly distorted block to compute the quality score. The advantage of PIQE over other no-reference metrics is that it does not require a training set, which is not acquired yet.

Testing the SSIM and PIQE on different radargrams results in a score value that hardly varies. The lack of changing scores might be the result of relative homogeneous pixel intensities in the

¹<https://nl.mathworks.com/products/image.html> (URL accessed on 28-10-2019)

simulated radargrams. If the value of the score does not vary on different radargrams, it is not useful for making a comparison. Therefore, the SSIM and PIQE are considered infeasible image quality metrics, and comparison is made using IMMSE and PSNR.

Although some image quality metrics prove not useful in the feasibility study, they might prove useful during future image processing steps of radargrams.

C.0.1 Experimental setup

The experimental setup is a wooden box full of soil and objects. The goal of the setup is to contain an environment of prior knowledge, which is measured using GPR. The experimental setup is located in a controlled environment to diminish the effects of the water table and rain on the GPR measurement.

The soil in the experimental setup is sifted dry sand, typically used in the production of concrete. The advantage of using sifted dry sand is the reduction in unexpected reflections. Ground truth is that this soil does not contain additional objects that could cause reflections, whereas average ground soil could contain a range of unexpected objects, e.g., rocks, branches, plants, or plastics. The experimental setup helps to create a more consistent and therefore, easy to interpret radargrams. Contradictory to GPR measurement outside where external factors could create unexpected reflections in the radargram.

This feasibility study ignores the cylindrical shape of the sewer and considers a flat sewer wall for a more convenient experimental setup. The shape makes for faster data acquisition since there is no need for any rotational GPR mechanism in a standard cylindrical sewer. The sewer wall is simulated using six pavement tiles of $30 \times 30 \times 4$ cm on top of the soil in the box. The soil is fine dry sand that fills the box up to 40 cm. A thin plastic sheet in the box protects the OSB board from soaking up any humidity left in the sand. Soaked wood could weaken the structural integrity of the box.

The box consists out of 18 mm thick Oriented Strand Board (OSB). The inner dimensions are $200 \times 50 \times 50$ cm l \times w \times h). Figures C.1 and C.2 show the design and the experimental setup. The length of the setup allows multiple subsurface objects alongside each other. The GPR system traverses the full length, receiving the reflections of the subsurface objects.

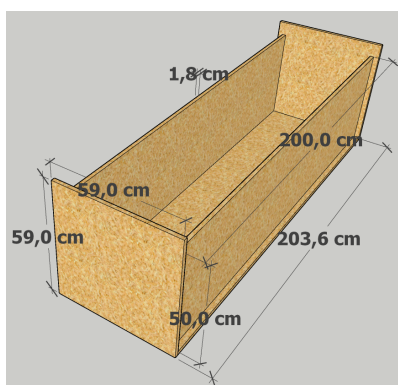


Figure C.1: Design of the experimental setup



Figure C.2: Experimental test setup

The last 20 cm of sand in the box does not include pavement tiles, see figure C.2 since the dimensions of the box do not allow it. Since the GPR systems travel over the tiles using its wheels, it means that the last 20 cm the system needs to be lifted. The lifting motion is a random move-

ment which creates an inconsistency in the measurement. Image crop removes the respective last 20cm from the radargram to remove the inconsistency from the radargram.

The results of the experimental setup are SGY files, which is a file extension for various seismic tools. RGPR² and Scikit-image van der Walt et al. ((2014)) are used to convert the SGY file into a 1920×1080 PNG image of the corresponding radargram. The PNG images are read into MatLab to create a comparison between the experimental and simulated radargram.

C.0.2 Simulation platform

Increasing the resolution of the simulation by reducing the Yee cell size or the trace interval leads to higher model complexity and therefore, a larger simulation time. A consistent comparison between simulations and their respective simulation times is achievable when using exclusive resources for simulation. Hence, the resources are node shared to process other jobs and thus are constantly processing the simulation without any interference.

The implementation platform is the High-Performance Computing (HPC) cluster of the University of Twente. Sets of either four Nvidia Geforce Titan-X or four Nvidia Geforce GTX 1080 Ti GPU's execute the simulations. To accurately time the simulations, the jobs are scheduled using the exclusive flag of the Slurm scheduler. Exclusive jobs claim the full node, meaning that other jobs can not interfere with the resources while processing. Therefore, the simulation time is consistent and not influenceable.

Using two sets of different GPU's seems not ideal for comparing simulation times, but the difference in performance between the GPU's insignificant ((Warren et al., 2019)). Therefore, using two sets of different but similar performing GPU's result in concurrent processing of two simulations instead of iterative processing as a result of waiting for free resources.

C.0.3 Comparison tests

Different test cases are developed to get a better understanding of how simulated radargrams compare to radargrams from the experimental setup. This section shortly describes the test and the corresponding results. The implementation and execution of the tests are described in greater detail in appendix B.

Test 1

The first test compares duplicate measurements from the experimental setup. The goal of this test is to get a better insight into the fluctuations of measurement results using the experimental setup. The variable factor during the measurement is the trace spacing. For every subsurface environment, two radargrams are acquired using 1mm, 3mm, and 10mm trace spacing, resulting in six radargrams for each subsurface environment.

The radargrams are denoted as *full* and *cropped radargram*. The full radargram includes the inconsistency of lifting the GPR over the last 20cm of experimental setup where there are no tiles. The cropped radargram only includes the first part of the measurement, excluding the last 20cm of the measurement. Comparing both radargrams determines if the IMMSE and PSNR can describe the inconsistency in the measurement and determine its significance.

²<https://emanuelhuber.github.io/RGPR/> (URL accessed on 28-10-2019)

The first step in determining the fluctuation of experimental radargrams is to calculate the IMMSE and PSNR. The quality metrics compare the duplicate radargrams for a specific trace spacing from a test. The average difference quality for a trace spacing overall tests denotes the variation of quality. Therefore, it proves the influence of the trace spacings on the quality of the radargram. Figure C.3 and C.4 depict the result of the quality metrics for full and cropped radargrams. Note that *TS* in the legend stands for trace spacing.

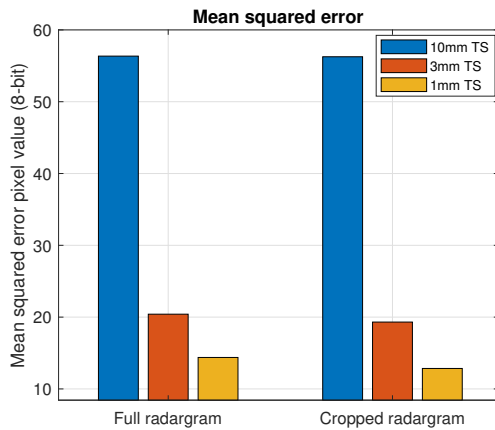


Figure C.3: Result mean squared error full vs. cropped radargram.

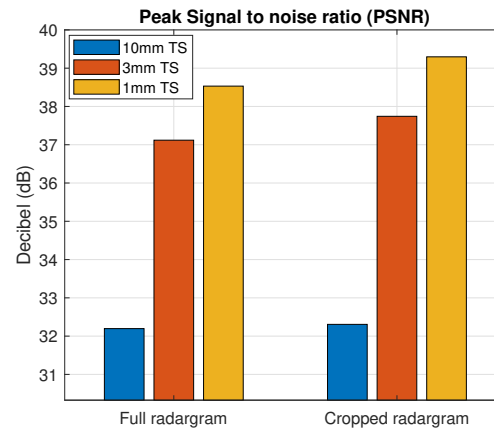


Figure C.4: Result peak signal to noise ratio full vs. cropped radargram.

For IMMSE, lower values denote a higher image quality, vice versa for PSNR. Using a smaller trace spacing results in a higher quality radargram for both IMMSE and PSNR. Hence, trace spacing is an important variable that has to be taken into account, creating a trade-off between quality and simulation time. All numerical values are found in tables C.1 - C.6.

The inconstancy in the full radargram results in a slightly lower quality radargram compared to the cropped radargram. Consequently, lifting the GPR creates inconsistency which that is measurable using both IMMSE and PSNR. From this point on, whenever describing experimental radargrams, they are cropped experimental radargrams if not stated otherwise.

| | IMMSE | PSNR | SSIM |
|--------|--------|--------|-------|
| mean | 56,350 | 32,196 | 0,989 |
| mode | 7,079 | 25,806 | 0,973 |
| median | 44,329 | 31,666 | 0,992 |
| std | 43,533 | 4,090 | 0,007 |

Table C.1: Full radargram 10mm trace spacing.

| | IMMSE | PSNR | SSIM |
|--------|--------|--------|-------|
| mean | 56,260 | 32,307 | 0,989 |
| mode | 5,004 | 25,810 | 0,972 |
| median | 44,547 | 31,645 | 0,991 |
| std | 44,301 | 4,264 | 0,007 |

Table C.2: Cropped radargram 10mm trace spacing.

| | IMMSE | PSNR | SSIM |
|--------|--------|--------|-------|
| mean | 20,410 | 37,119 | 0,994 |
| mode | 3,417 | 27,057 | 0,982 |
| median | 9,390 | 38,423 | 0,995 |
| std | 26,213 | 3,944 | 0,004 |

Table C.3: Full radargram 3mm trace spacing.

| | IMMSE | PSNR | SSIM |
|--------|--------|--------|-------|
| mean | 19,311 | 37,743 | 0,994 |
| mode | 2,314 | 27,049 | 0,983 |
| median | 7,943 | 39,139 | 0,995 |
| std | 26,584 | 4,349 | 0,004 |

Table C.4: Cropped radargram 3mm trace spacing.

| | IMMSE | PSNR | SSIM |
|--------|--------|--------|-------|
| mean | 14,383 | 38,531 | 0,995 |
| mode | 1,975 | 27,696 | 0,987 |
| median | 9,124 | 38,529 | 0,995 |
| std | 21,284 | 3,673 | 0,002 |

Table C.5: Full radargram 1mm trace spacing.

| | IMMSE | PSNR | SSIM |
|--------|--------|--------|-------|
| mean | 12,851 | 39,297 | 0,995 |
| mode | 1,735 | 27,869 | 0,988 |
| median | 7,102 | 39,619 | 0,996 |
| std | 20,859 | 3,819 | 0,002 |

Table C.6: Cropped radargram 1mm trace spacing.

Test 2

This test compares simulated radargram quality for variable Yee cell size and trace spacing in 2D and 3D simulations. The goal of this test is to compose a trade-off between the radargram quality (acrshortimmse/PSNR) and simulation time. Comparisons in this test are within each respective dimensions, e.g., a 2D radargram uses a 2D radargram as reference for the comparison. The following test three compares radargram from different model dimensions.

The input files of the simulations describe the subsurface environment of the experimental setup. The simulation is an approximate discretization of the experimental environment. Therefore, similar reflections are expected to occur in the resulting radargrams.

During this test, the smallest possible yee cell size (2mm) and trace spacing (2mm) is considered to result in a radargram with the highest quality radargram. The previous test proves that a smaller trace spacing results in a higher quality radargram in IMMSE and PSNR. Additionally, a smaller the Yee cell size results in a discretization closer to the real representation of the experimental setup. Therefore, the reference radargram contains as small as possible Yee cell size and trace spacing. Every other radargram is compared to the reference radargram to calculate the IMMSE and PSNR values. Yee cell size and trace spacing smaller than 2mm are not possible, which is clarified later on.

Simulations are run for each feasible combination of Yee cell size and trace spacing resulting in unique radargrams. The Yee cell size ranges from 2mm to 10mm, incrementing 1mm per simulation. The trace space range is also between 2mm and 10mm, but it increments with multiples of the minimum trace spacing. Trial tests show that values exceeding 10mm for cell size or trace spacings result in an exponential growing graph, making it difficult to compare the samples. Consequently, the maximum of the values is 10mm.

The trace spacing can never be smaller than the cell size since it results in a rounded step too small to go into the next discrete cell. Consequently, the virtual GPR can not move and will measure the wave propagation in the same location. Such a radargram is easy to spot since it only contains perfect horizontal reflections.

The trace spacing is a multiple of the Yee cell size. If the value is not a multiple, the step rounds in the discrete domain, which results in inconsistent behaviour. The virtual GPR will move but with different size steps.

The numbers of runs per cell size will differ as to the result of taking multiples of the cell size. For example, 3mm cell size results in three different trace spacings (3,6,9), whereas 4mm cell size results in 2 different trace spacings (4,8). For lower cell sizes, more multiples exist, and thus, it has more influence on the polynomial fit describing the trend. The resulting simulation time and quality metric for both 2D and 3D simulation are depicted in figures C.5 - C.8.

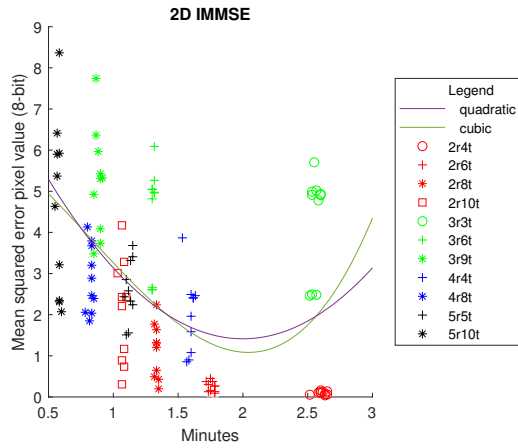


Figure C.5: Samples and trend for 2D IMMSE.

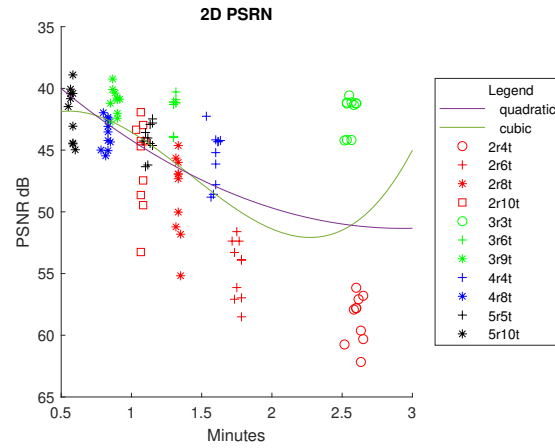


Figure C.6: Samples and trend for 2D PSNR.

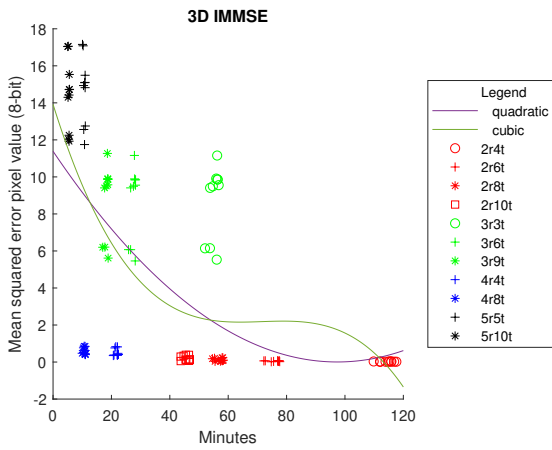


Figure C.7: Samples and trend for 3D IMMSE.

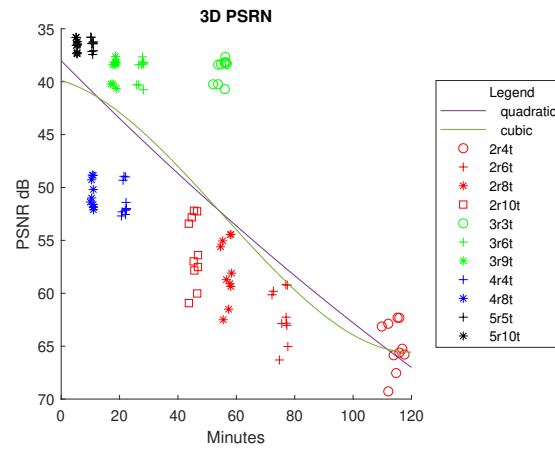


Figure C.8: Samples and trend for 3D PSNR.

Overall, graphs C.5 - C.8 show that a bigger cell size and small trace spacing does not necessarily lead to smallest error. For example, the results of cell size 3mm and trace spacing 3mm in graph C.5 lead to higher IMMSE values in comparison to 4mm cell size and 4mm trace spacing. This surprising observation might be the result of simulated object dimensions being multiples of 2mm instead of 3mm. Consequently, the dimensions of the object are rounded the discrete domain leading to a fundamentally different model, and thus a different quality radargram.

A quadratic and cubic polynomial *polyfit* show the trends of the samples. Table C.7 and C.8 contain the 2D and 3D polynomial functions, the normal of residual values to the data and five configurations closest to the sweet point on the polynomial fit. The sweet point $P2(x_2, y_2)$ on the polynomial fit is the intersection of the shortest vector from the lowest simulation time and metric value to the polynomial $P1(x_1, y_1)$. In case of IMMSE, $P1$ is the point on the lowest simulation time and lowest IMMSE value. For PSNR, $P1$ is the point on the lowest simulation time and highest PSNR value.

The five most convenient combinations of Yee cell size and trace spacing are considered to be the five closest samples to the sweet point of a polynomial. Since the polynomials are different, also different feasible combinations might be expected. The five closest configurations are considered the most efficient for the specific polynomial function, see table C.7 and C.8. Note that in the tables and graphs the resolution is denoted by r and the trace spacing by t .

| Function | Norm of residuals | $P2(x_1, y_1)$ | 5 closest samples |
|---|-------------------|-------------------|-------------------|
| $y_{IMMSE} = 1.72x^2 - 6.9x + 8.31$ | 18.2026 | (1.7630, 1.4914) | 3x2r8t 2x4r4t |
| $y_{IMMSE} = 0.773x^3 - 1.92x^2 - 1.82x + 6.23$ | 18.1556 | (1.8276, 1.2094) | 4x4r4t 1x2r8t |
| $y_{PSNR} = -1.9x^2 + 11x + 35$ | 50.8168 | (50.8073, 2.6500) | 3x2r8t 2x2r6t |
| $y_{PSNR} = -4x^3 + 17x^2 - 15x + 46$ | 50.3552 | (2.2779, 52.7630) | 5x2r6t |

Table C.7: 2D polynomial fit.

| Function | Norm of residuals | $P2(x_1, y_1)$ | 5 closest samples |
|---|-------------------|--------------------|-------------------|
| $y_{IMMSE} = 0.0012x^2 - 0.23x + 11$ | 49.4448 | (7.0319, 9.4420) | 3x5r10t 2x5r5t |
| $y_{IMMSE} = -3.3e - 05x^3 + 0.007x^2 - 0.5x + 14$ | 48.6463 | (8.8783, 10.0895) | 3x5r5t 2x5r10t |
| $y_{PSNR} = -0.00031x^2 + 0.28x + 38$ | 73.5729 | (12.5823, 41.4740) | 4x5r5t 1x3r9t |
| $y_{PSNR} = -2.4e - 0.5x^3 + 0.004x^2 + 0.08x + 40$ | 73.285 | (9.1710, 41.0516) | 5x5r5t |

Table C.8: 3D polynomial fit.

2D and 3D simulations vary significantly in simulation time. The longest simulation time for 2D is 2.7 minutes and for 3D 110 – 120 minutes, making 2D around 40-50 time faster. The performance difference is the result of the additional complexity of 3D models. A 2D model has a depth equal to the Yee cell size, contradictory to the 3D model, which has a depth based on the model domain.

The 2D results resolutions vary from 2-4mm and trace spacing from 4-8mm. The 3D resolution is mainly 5mm with trace spacing between 5-10mm. This test proves that a higher resolution does not necessarily result in a higher performance relating to IMMSE/PSNR and execution time. Furthermore, the 2D and 3D results for the five most efficient configuration do not correlate. Therefore, when chosen for simulating in either 2D or 3D, a different feasible configuration arises.

Test 3

This test compares how radargrams for 2D model compare to radargram from a 3D model. The goal is to determine if 2D radargrams are similar to 3D radargrams of the respective environment. Theoretically, a 3D approximation of a subsurface environment is closer to reality than a 2D approximation since wave propagates in three dimensions. If a 2D approximation describes a subsurface environment similar to the 3D approximation, it might be more convenient in the simulation since it greatly reduces the simulations times, see test C.0.3.

All simulations contain rectangular subsurface objects of different materials. Table C.9 states the test and the location of the subsurface object along the x -axis of the box and the depth under the tiles d . The parameters of the object material are found in table C.10.

Similar to the previous test, the IMMSE and PSNR metrics describe the quality of the radargram. Also, reference radargram has a 2mm Yee cell size and 2mm trace spacing, assuming it results in the highest quality radargram. The result of the comparisons are visualised in figures C.9 and C.10.

Both the IMMSE and PSNR graphs show two different clusters, a horizontal shaped cluster high in the graph and a similarly shaped cluster low in the graph. Some of the tests result in a significantly lower quality 2D radargram compared to other tests. Consequently, not only the trace spacing and Yee cell size influence the quality in this test, but the simulated environment as well.

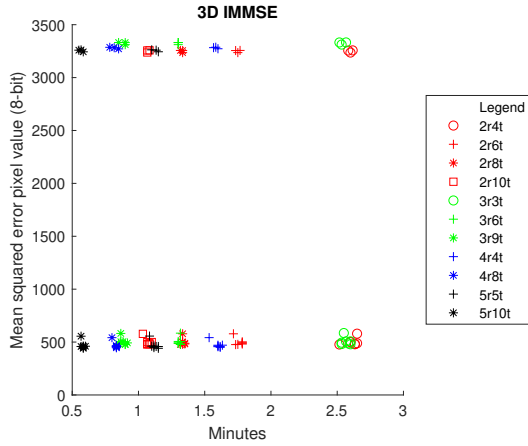


Figure C.9: Result 2D vs. 3D IMMSE.

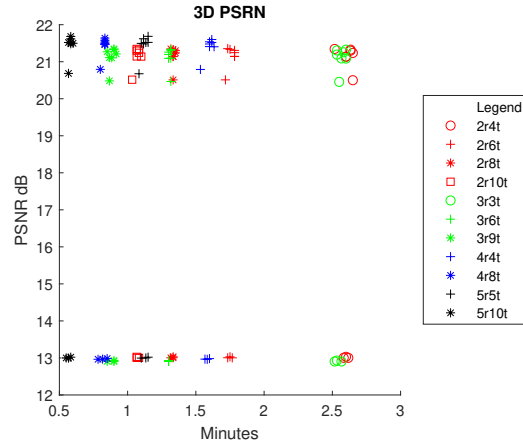


Figure C.10: Result 2D vs. 3D PSNR.

Visualizing the difference between 2D and 3D radargrams is essential to find the source of the different quality results. Firstly, all radargrams from a test in 2D or 3D merge into a composite radargram. Secondly, a difference radargram visualizes the points of similarity of the 2D and 3D composite. The differences denote as high pixel values and similarity as low pixel values. Visually inspection results in three test having a higher homogenous pixel difference, see the radargrams depicted by figure C.15.

The three tests assumed to introduce lower quality radargrams are tests 4, 7, and 8. To prove that these tests are indeed the source of the outliers, they are excluded from the data. After creating a new plot with the new filtered data, see graphs C.11 and C.12, no outliers show up. Therefore, test 4, 7, and 8 indeed result in lower quality radargrams when comparing the 2D radargrams with the 3D reference radargram.

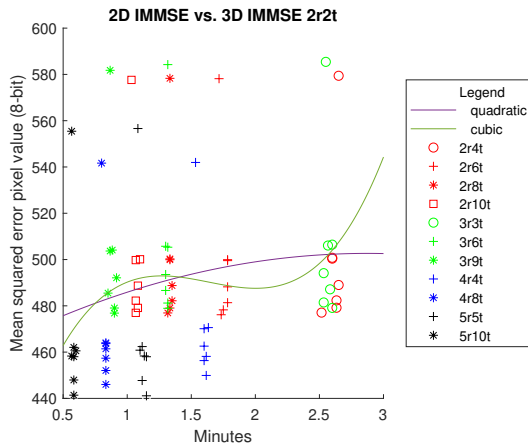


Figure C.11: Result 2D vs. 3D IMMSE filtered.

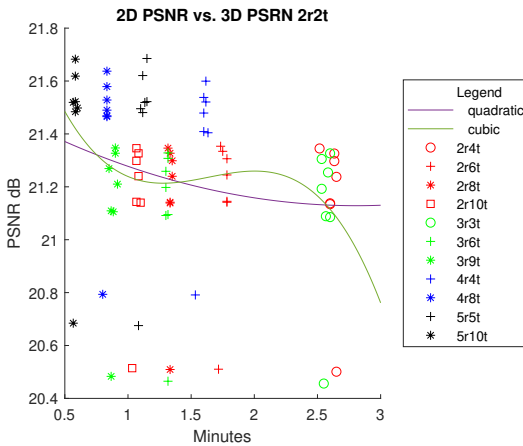


Figure C.12: Result 2D vs. 3D PSNR filtered.

The lower quality radargrams are a result of the higher difference in pixel intensities in the homogeneous regions of the radargram. Although IMMSE and PSNR quality metrics interpret this has a lower quality radargram, it is not per definition. Since the homogeneous regions do not describe any reflections, it does not necessarily influence the difficulty in recognizing reflections from the radargram.

Visual inspection shows that 2D approximation of a 3D environment, e.g. the radargrams figure C.14 and C.13 respectively, creates a similar reflection in the same location. Surprisingly,

the reflection in the 2D radargram is clearly defined, contradictory to the reflection in the 3D radargram that blends more in with the homogeneous pixel intensity. The lack of contrast of reflections in the 3D simulation could be the result of more energy dissipation in the third dimensions. Since the third dimension can not dissipate energy in 2D simulations, the reflections are generally stronger and easier to notice. Although this coincidence might be useful, it does not properly reflect the experimental setup where energy does dissipate in 3 dimensions.



Figure C.13: 3D radargram.

Figure C.14: 2D radargram.

| Test | Object material | Object position (x; d) | Object dimension |
|------|-----------------|------------------------|------------------|
| 1 | pec | 20; 15cm | 10×10×10cm |
| | wood | 40; 15cm | 10×10×10cm |
| | air | 60; 15cm | 10×10×10cm |
| 2 | pec | 30; 15cm | 10×10×10cm |
| | wood | 50; 15cm | 10×10×10cm |
| | air | 70; 15cm | 10×10×10cm |
| 3 | pec | 30; 21.8cm | 10×10×10cm |
| | wood | 50; 21.8cm | 10×10×10cm |
| | air | 70; 21.8cm | 10×10×10cm |
| 4 | pec | 30; 1.8cm | 10×10×10cm |
| | wood | 50; 1.8cm | 10×10×10cm |
| | air | 70; 1.8cm | 10×10×10cm |
| 5 | air | 70; 1.8cm | 10×10×10cm |
| 6 | wood | 50; 1.8cm | 10×10×10cm |
| 7 | pec | 30; 1.8cm | 10×10×10cm |
| 8 | pec | 30; 1.8cm | 10×10×10cm |
| | wood | 50; 1.8xm | 10×10×10cm |
| 9 | wood | 50;1.8cm | 10×10×10cm |
| | air | 70;1.8cm | 10×10×10cm |
| 10 | pec | 30; 1.8cm | 5x10x10cm |
| | wood | 50; 1.8cm | 5×10×10cm |
| | air | 70; 1.8cm | 5×10×10cm |

Table C.9: Test log.

| | f1 | f2 | f3 | f4 |
|----------|-----|------|----|----|
| wood | 2 | 0 | 1 | 0 |
| air | 1 | 0 | 1 | 0 |
| sand | 3 | 0.01 | 1 | 0 |
| concrete | 4.5 | 0.01 | 1 | 0 |

Table C.10: Simulation material properties.

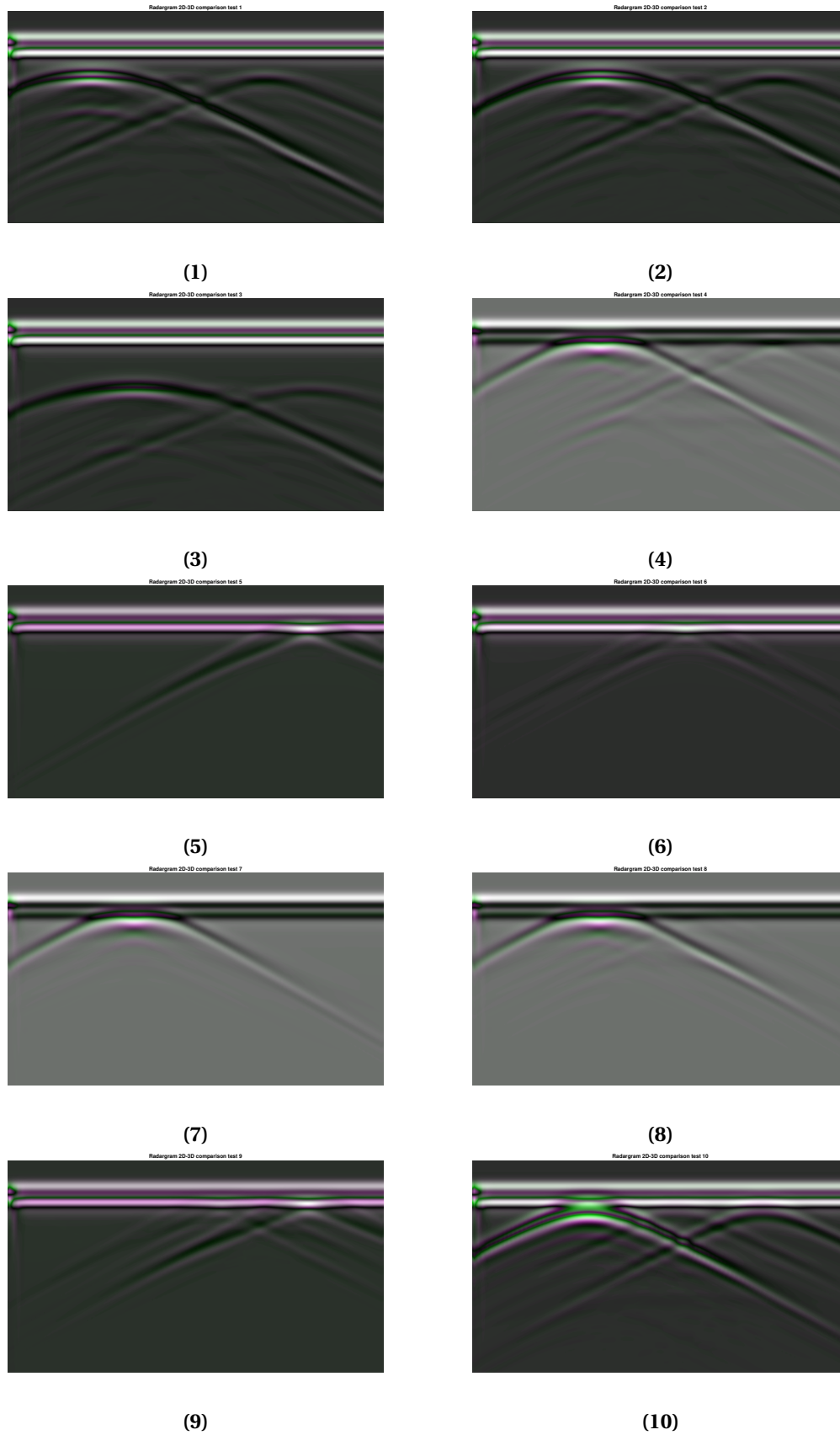


Figure C.15: From top left to bottom right, the radargrams from test 1-10.

C.0.4 Conclusion on feasibility of gprMax

The feasibility of gprMax is a result of three different tests. The first test describes the comparison between experimental radargram duplicates for different trace spacing. The results show that the measurements do not vary much in IMMSE and PSNR value and thus are consistent. Additionally, the lifting of the GPR during the experimental tests result in slightly worse IMMSE and PSNR proving it adds additional inconsistency.

Secondly, simulating 2D and 3D environments with different combinations of Yee cell size and trace spacing result in varying simulations times. The most feasible configuration is a trade-off between image quality in terms of IMMSE/PSNR and simulation time. 2D simulation proves most feasible with a Yee cell size of 2-4mm and trace spacing of 4-8mm, see table C.7. 3D simulation is most feasible with Yee cell size of 5mm and trace spacing 5-10mm, see table C.8. Additional observations during testing show that trace spacing should be equal or a multiple of the Yee cell size to prevent rounding in the discrete domain to result in arbitrary radargrams.

Thirdly, the simulation time for 3D is roughly 40-50 times slower than the approximate 2D simulation. If 2D simulations can capture the reflections adequately, and thus creates similar radargrams to 3D, it would significantly increase the feasibility. When comparing 2D versus 3D results in a significant worse IMMSE and PSNR in comparisons with the quality values of test two. Surprisingly, the 2D radargrams show more contrast in reflections, which makes it easier to detect compared to their 3D counterparts. The high contrast might be a result of energy not dissipating in the third dimension, since only one plane of Yee cells. Hence, more energy is reflected in subsurface objects, producing strong reflections.

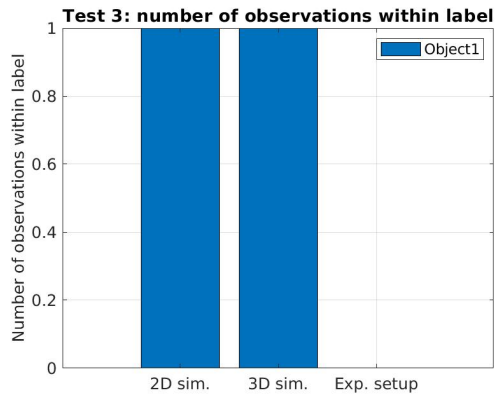
Also, some outliers in metric values prove to be the result of high pixel intensity differences in the homogeneous area of the radargram. Describing the quality using the IMMSE and PSNR values this might result in worse quality radargram. Although the resulting radargram might have a worse quality to the reference radargram, it does not describe the quality of reflections in the radargram.

A limitation of the feasibility study is that it does not contain a numerical comparison between the experimental and simulation radargrams. In a radargram describe reflections, and some describe the noise. In the hypothetical case of a radargram consisting out of mostly noise, IMMSE and PSNR can still describe it as a high-quality radargram when compared to a similar radargram. From the feasibility study can be concluded that determining the quality of the radargram is difficult when there is no prior knowledge of which parts of the radargram contain useful information. Instead of comparing pixel values, the difference in reflection shape, location, polarity, and reflection coefficients of the simulated and experimental reflections might prove as a more reliable quality metric in future research.

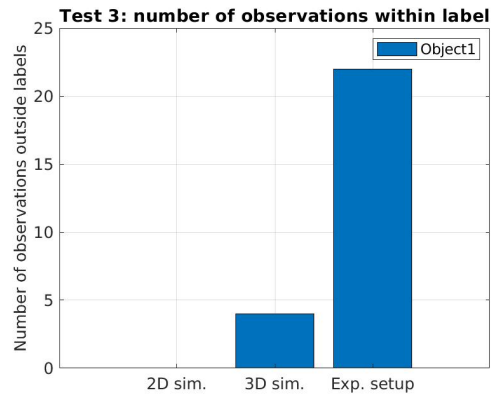
D Appendix 4 - Test 1 additional results

This appendix extends the result of test one, for more details see section 5.1. The results of test one depict the number of observations inside and outside the label over all tests in figure 5.1 and figure 5.2, respectively. Appendix B contains the descriptions of the experimental test environments

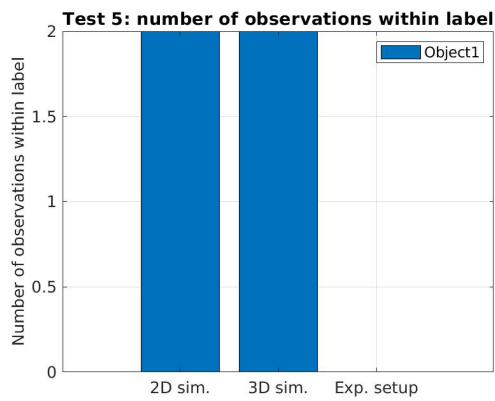
The bar graphs of figure D.1 depicts the number of observations inside and outside the labels per test. The bar graphs in the left columns describe the numbers inside and the right columns the numbers outside the labels. Note that the graph describing the number of observations inside the labels contain multiple bars per type radargram. The multiple bars depict the number of observations inside a specific label.



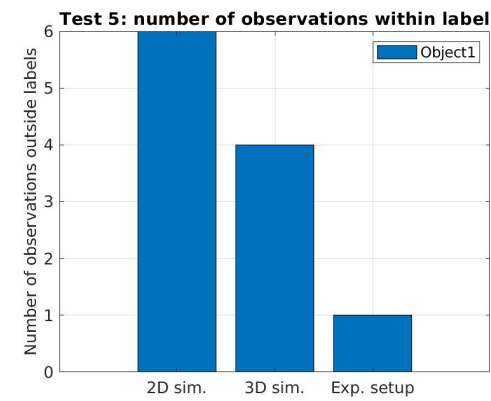
(1)



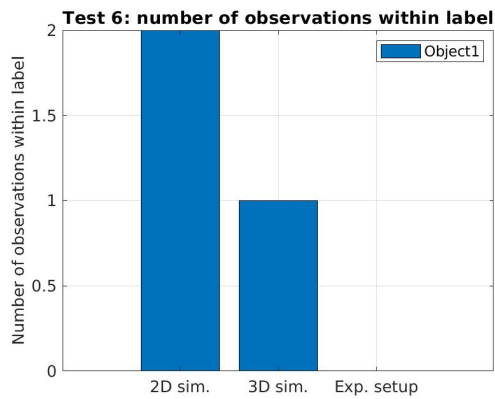
(2)



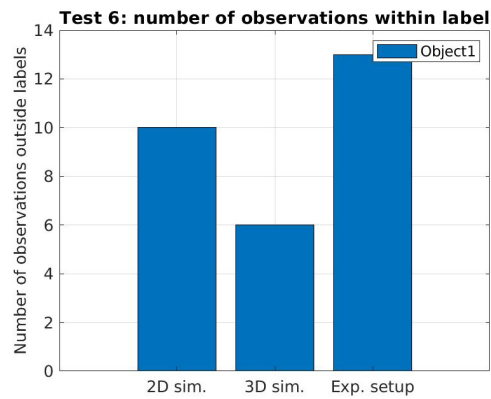
(3)



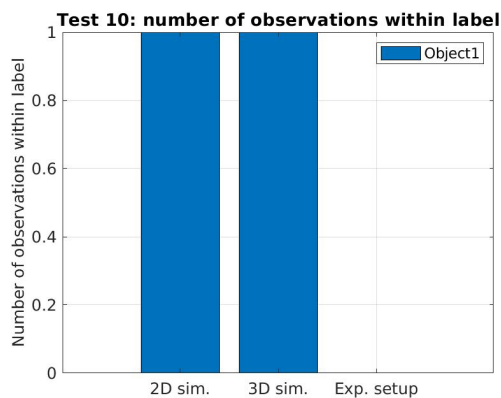
(4)



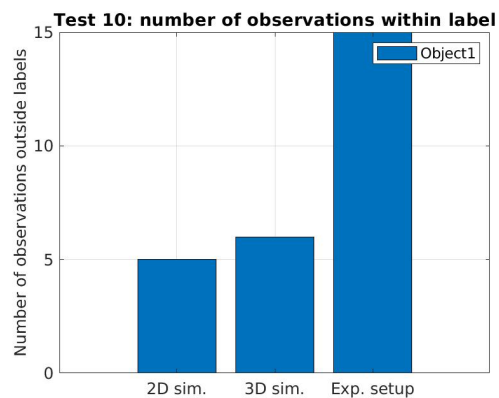
(5)



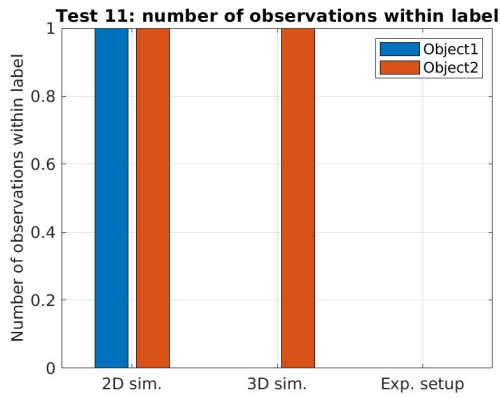
(6)



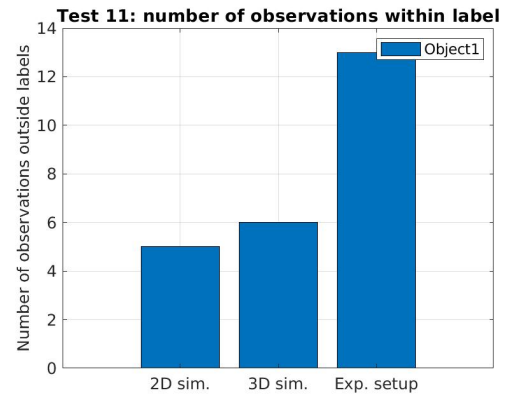
(7)



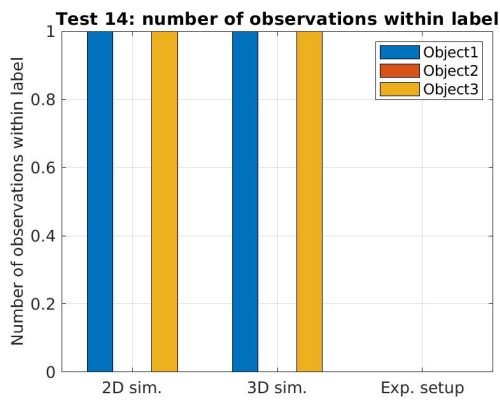
(8)



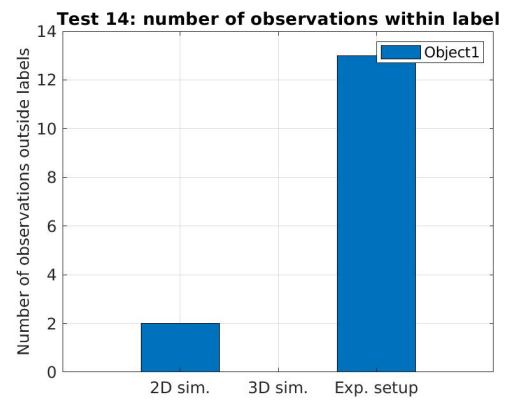
(9)



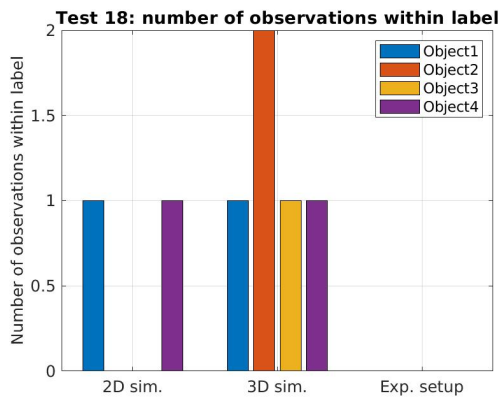
(10)



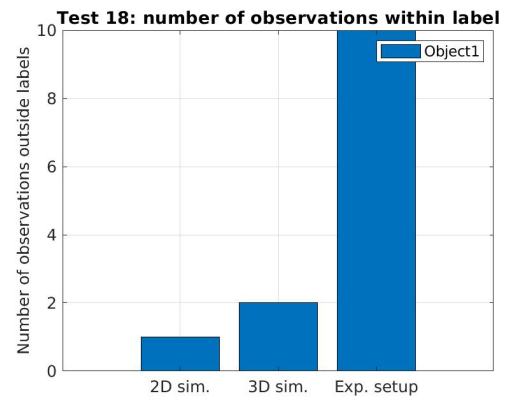
(11)



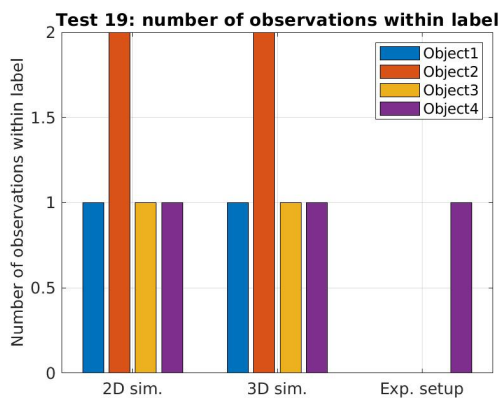
(12)



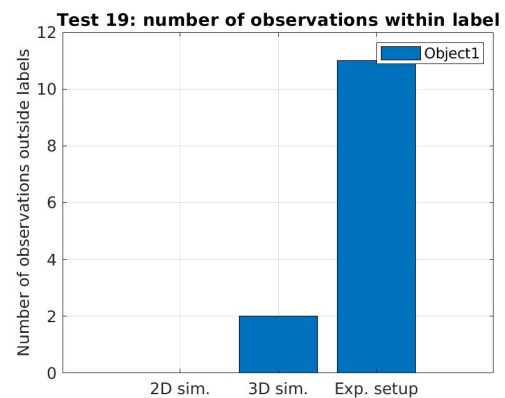
(13)



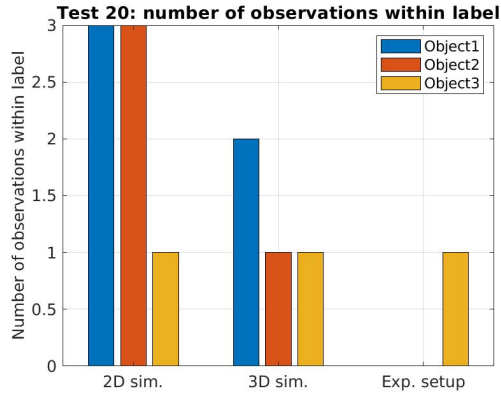
(14)



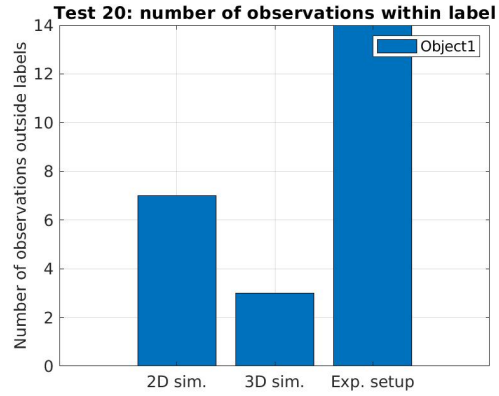
(15)



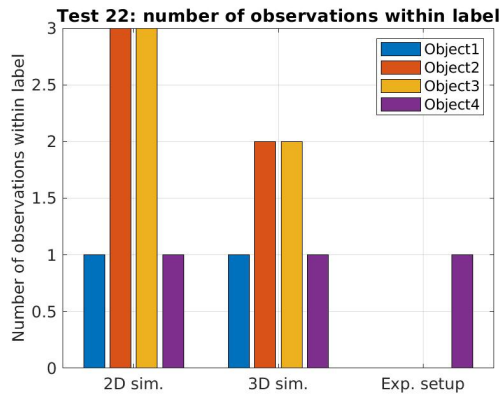
(16)



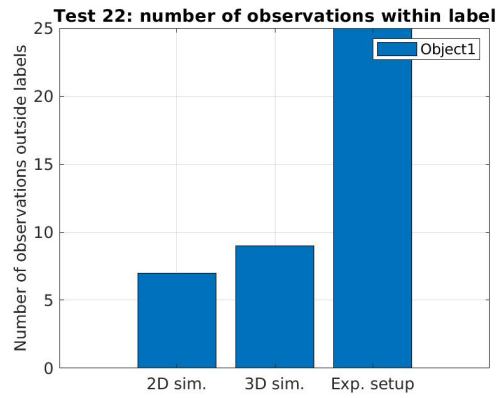
(17)



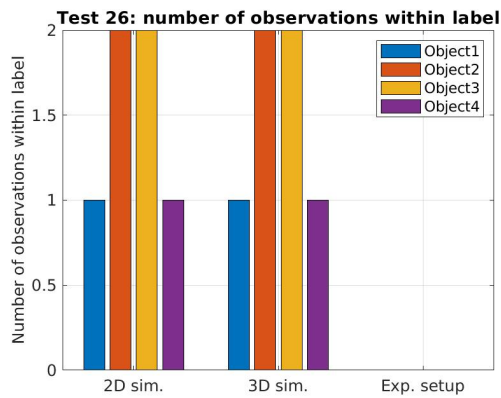
(18)



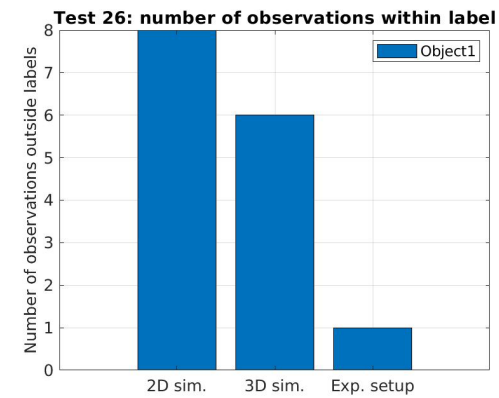
(19)



(20)



(21)



(22)

Figure D.1: Left column depicts the number of hyperbolas within the labels for the different tests. The right column depicts the number of hyperbolas outside of the label for all respective tests.

E Appendix 5 - Test 2 additional results

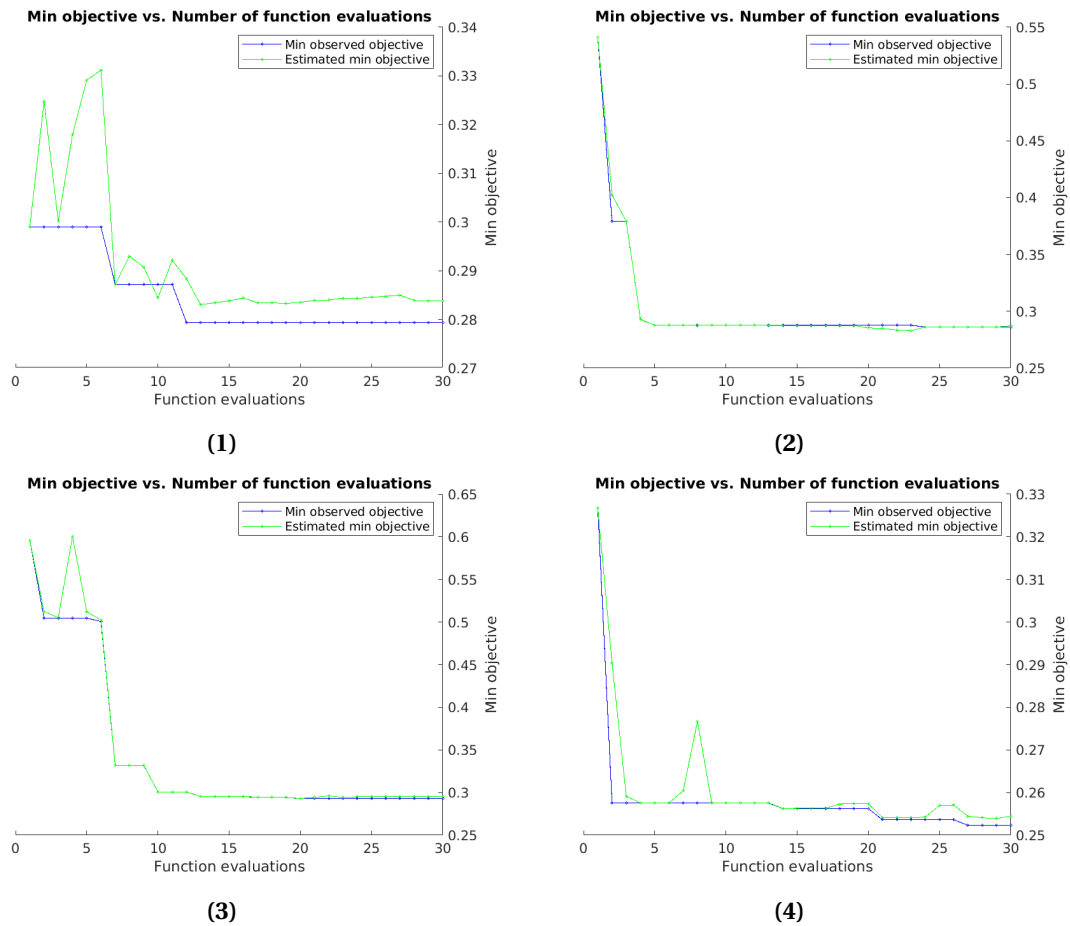
This appendix extends the results of test two, see section 5.2. Table E.1 includes the weights of the features for both the planar and cylindrical topology. The features depicted with green are selected in the feature selection process. If a feature is green for both topologies, it is a topology dependent feature. When a feature is green for only one topology, it is a topology dependent feature. Features which are not green are not selected at all and, therefore, irrelevant in the void detection problem.

| Features | Planar topology | Cylindrical topology |
|------------------|-----------------|----------------------|
| Area | 0,0007 | 0,001 |
| Centroid1 | 0 | 0,0037 |
| Centroid2 | 1,0906 | 2,7491 |
| MajorAxisLength | 1,0992 | 1,8177 |
| MinorAxisLength | 1,042 | 0,7215 |
| Eccentricity | 0,0005 | 2,663 |
| Orientation | 0,262 | 0,8798 |
| ConvexArea | 0,0001 | 0,0005 |
| Circularity | 0,7037 | 0 |
| EquivDiameter | 0,9449 | 1,4958 |
| Solidity | 1,1052 | 1,2577 |
| Extent | 1,1126 | 1,2041 |
| Perimeter | 0,0001 | 0,0002 |
| MaxFeretDiameter | 0,0003 | 0,0005 |
| MaxFeretAngle | 0,0001 | 0,8393 |
| MinFeretDiameter | 0,3972 | 0,0009 |
| MinFeretAngle | 0 | 0,9467 |
| ApexX | 0,6152 | 0,0003 |
| ApexY | 2,8031 | 2,8472 |
| Drel1 | 0,0002 | 0,0006 |
| Drel2 | 0,0002 | 0,551 |
| Sym | 0,0001 | 0 |
| Drt1 | 0,238 | 0,2686 |
| Drt2 | 0 | 0,0001 |
| Dry1 | 0,1761 | 0,0001 |
| Dry2 | 0,1092 | 0,0001 |
| Dis1 | 0,0343 | 0,0001 |
| Dis2 | 0,7232 | 1,0712 |
| Sur | 0,8771 | 0 |
| Int | 0,692 | 1,057 |
| Eps | 0 | 0,0008 |
| Pol2 | 0 | 0,0013 |
| Pol1 | 0,0001 | 0,0002 |
| Pol | 0,0001 | 0,0001 |
| MinX | 0,0001 | 0,0002 |
| MaxX | 0,0013 | 0 |
| Velocity | 1,792 | 2,2683 |

Table E.1: Feature weights of each feature. The feature is green when it is in the selected feature set of the specified topology.

F Appendix 6 - Test 3 additional results

This appendix extends the results of test three. Figure F.1 and F.2 depicts the iterative minimisation of the loss function for the planar and cylindrical topology, respectively. In both figures, the graphs in the left column depict the minimisation for the corresponding topology using all features and the right column using the selected features. The graph depicts an estimated minimum loss function and the observed minimum loss function. The goal is to get the observed loss function as low as possible. After a certain amount of iteration, the minimisation is stable, and the optimum is found.



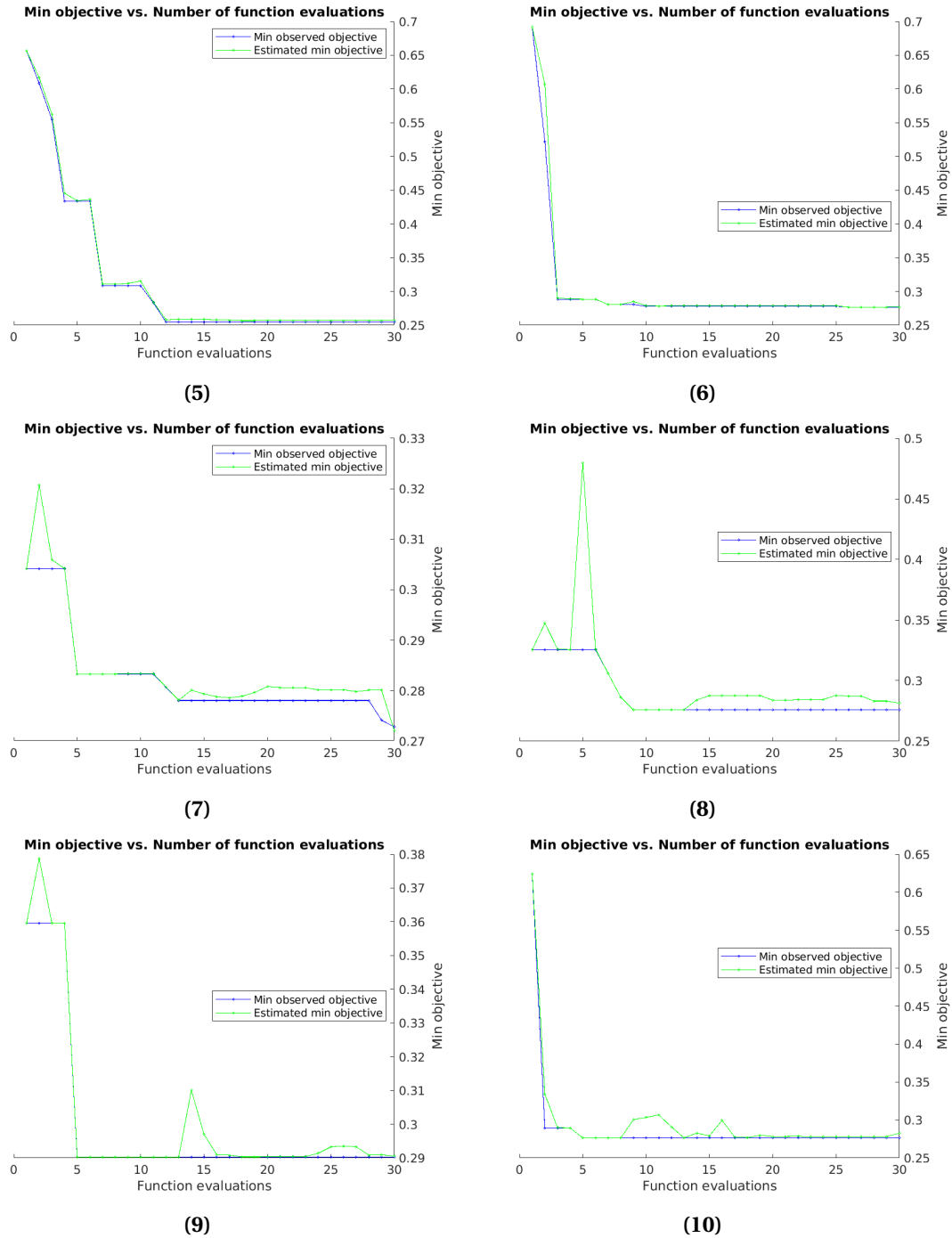
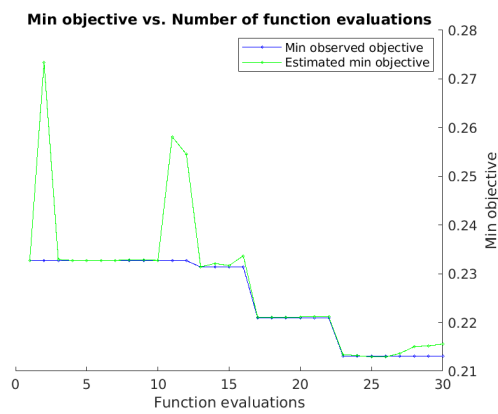
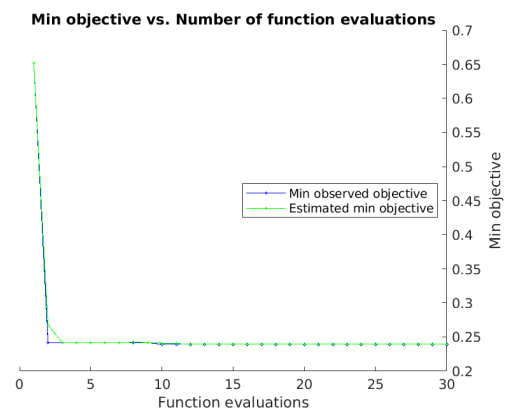


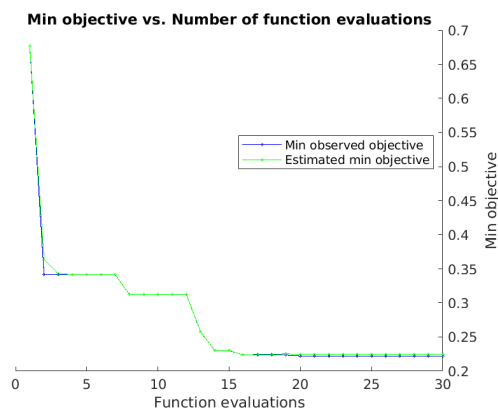
Figure F.1: The left column depicts the minimisation of the objective function on the planar topology using all features. The right column depicts the minimisation of the objective function on the planar topology using the selected features features.



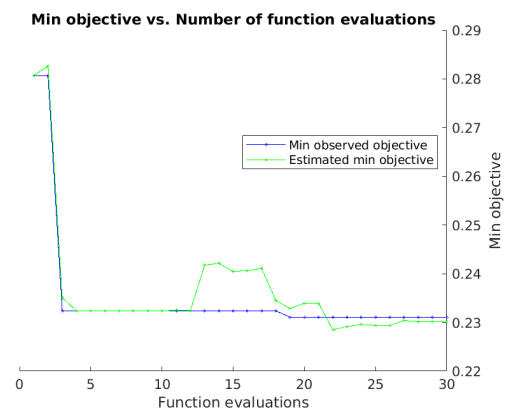
(1)



(2)



(3)



(4)

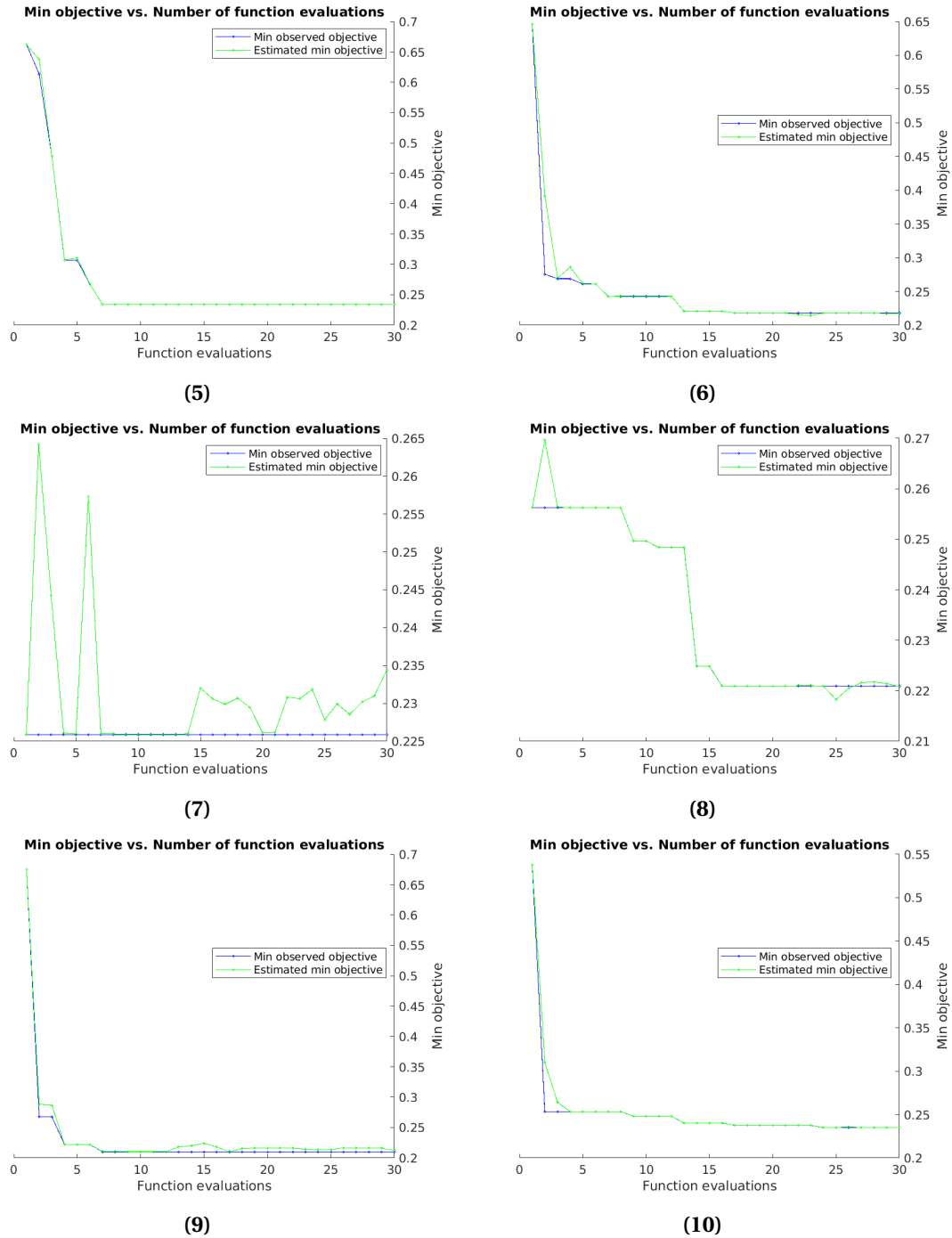


Figure E2: The left column depicts the minimisation of the objective function on the cylindrical topology using all features. The right column depicts the minimisation of the objective function on the cylindrical topology using the selected features features.