


UNIVERSITY OF BIRMINGHAM

# Novel AI-assisted computational solutions for GPR data interpretation and electromagnetic data fusion to detect buried utilities

By

**Arasti Afrasiabi**

A thesis submitted to The University of Birmingham
For the degree of
DOCTOR OF PHILOSOPHY

School of Engineering
College of Engineering and Physical Sciences
University of Birmingham
July 2023

UNIVERSITY OF
BIRMINGHAM

## University of Birmingham Research Archive

### e-theses repository

This unpublished thesis/dissertation is copyright of the author and/or third parties. The intellectual property rights of the author or third parties in respect of this work are as defined by The Copyright Designs and Patents Act 1988 or as modified by any successor legislation.

Any use made of information contained in this thesis/dissertation must be in accordance with that legislation and must be properly acknowledged. Further distribution or reproduction in any format is prohibited without the permission of the copyright holder.

# ABSTRACT

This research presents a number of novel computational solutions using artificial intelligence (AI) to interpret ground penetrating radar (GPR) data as well as fusing GPR data with data from other sensing modalities, including electromagnetic conductivity (EMC) and electromagnetic locating (EML). The application of the proposed computational solution is predominantly for detecting and locating buried utilities (e.g. pipes and cables) and ground anomalies (e.g. ground disturbances) in the shallow subsurface environment although the work can be extended to detect other buried anomalies. Processing GPR data is usually a subjective and time-consuming practise which involves expert intervention. Thus, the quality of the interpretation of such data depends on user experience and knowledge. Whilst several numerical approaches are available in the literature for post-processing GPR data, they all suffer from various shortcomings including lack of accuracy and/or excessive computational time. The issue is similar (or often worse) for data fusion between GPR and other sensors e.g. EMC and EML. To tackle some of these issues, in this research, four new computational procedures were developed. Three of these computational procedures are based on Kalman Filtering (KF), a less-studied approach to process GPR radargrams despite its great potential in efficient data analysis, and genetic algorithm (GA) as a machine learning based global optimisation tool. The final computational procedure combines finite element modelling and genetic algorithm to infer fused EML-GPR data. For the first two numerical methods, new algorithms were developed to optimise KF parameters using GA to remove noises from GPR radargrams and detect targets. The proposed procedures were validated against data from field and their performance was assessed against additional unseen dataset different to that of the validation to identify their potential limitations. Furthermore, their performances were compared against existing GPR data processing methods and differences were highlighted.

The other two computational packages focused on data fusion from GPR and EMC/EML. The first of these two, extended the above KF algorithm to fuse data from GPR and EML as well as GPR and EMC. The results showed that the proposed data fusion algorithm significantly enhanced the quality of locating conductors and conductive regions in the subsurface compared to the individual techniques which were either incapable of defining the material of the buried target or the geometry of conductive anomalies. Finally, a novel inversion algorithm was developed by integrating finite element modelling of a coupled magnetic field and GA for detecting and locating buried live cables using GPR and EML. It was demonstrated that the proposed inversion can successfully detect the location of the buried cables as well as their intensity.

To *Nastaran*

The love who kept me sane

With whom I know no pain

My wonderful wife

The fountain to avoid me drain

~AND~

To my beloved mother

*Asefeh*

who took paths in life which no angle did dare to take

# ACKNOWLEDGEMENTS

The author would like to thank his academic supervisors, Dr Asaad Faramarzi and Professor David Chapman for their continuous support and guidance through this research project.

The author would also like to express his gratitude to all his friends and people who directly or indirectly helped him through this journey. In particular, Mr Alireza Keshavarzi, Dr Farough Rahimzadeh, Dr Koohyar Faizi, Mr Ali Monzer, and Mr Mathew Stringfellow.

The author acknowledges the Engineering and Physical Sciences Research Council (EPSRC) in the UK, with sincere gratitude, for the provided scholarship and financial support.

# Table of Contents

List of Figures ..... X
List of Tables ..... XVIII
Nomenclature ..... XIX
Notations ..... XX
**Chapter 1 INTRODUCTION ..... 1**
1.1 Background ..... 1
1.2 Problem statement ..... 7
1.3 Aim and Objectives ..... 8
1.4 The original contribution of this research ..... 9
1.5 Structure of the thesis ..... 11
**Chapter 2 LITERATURE REVIEW ..... 13**
2.1 Introduction ..... 13
2.2 Overview and Classification of Nonintrusive Methods ..... 14
2.3 Ground Penetrating Radar (GPR) ..... 18
2.3.1 Introduction ..... 18
2.3.2 GPR data ..... 21
2.3.3 Processing GPR data ..... 23
2.3.4 GPR data interpretation ..... 38
2.4 Data Fusion and multi remote sensing ..... 54
2.4.1 Introduction ..... 54
2.4.2 The low-level data fusion (data-oriented or signal level) ..... 55
2.4.3 Intermediate level or feature level data fusion ..... 56
2.4.4 High level or decision level data fusion ..... 56
2.4.5 Fusion of nonintrusive NDT geophysical methods ..... 58
2.5 Mathematical estimators for data fusion ..... 64
2.5.1 Kalman Filter (KF) ..... 75
2.5.2 Kalman Filter in GPR application ..... 82
2.5.3 Optimisation of KF parameters in GPR application ..... 88
2.6 EM-based techniques for data fusion with GPR ..... 90
2.6.1 Electromagnetic locator (EML) – Cable avoidance tool (CAT) ..... 90
2.6.2 Ground Conductivity Measurement (Train Conductivity) ..... 100

2.7 Gaps identified from the literature ..... 106
**Chapter 3 METHODOLOGY ..... 111**
3.1 Introduction ..... 111
3.2 The journey of this research project ..... 113
3.2.1 Why Kalman Filter? ..... 114
3.3 Semi-autonomous algorithm for GPR data processing using Kalman filter ..... 115
3.3.1 Overview of the semi-autonomous algorithm ..... 115
3.3.2 Kalman Filter in GPR applications ..... 117
3.3.3 Overview of the linear discrete Kalman filter algorithm (LDKF) ..... 123
3.3.4 Incorporation of basic signal pre-processing filters ..... 127
3.3.5 A novel target detection method using Normalised Innovation Squared values (NIS) ..... 144
3.4 Overview of the proposed optimisation algorithm ..... 155
3.4.1 Multi-objective genetic algorithm (MOGA) and Pareto solutions ..... 156
3.4.2 Optimisation of noise covariances Q and R ..... 157
3.4.3 Optimisation of Fourier frequencies (FFreq) ..... 158
3.4.4 The solution to 'identify the target width' ..... 162
3.5 Autonomous algorithm using Kalman filter and wavelet transform ..... 164
3.5.1 Overview of the autonomous algorithm ..... 164
3.5.2 Wavelet transform as a multi-resolution signal analysis ..... 166
3.5.3 Analysing NIS signal using wavelet transform ..... 172
3.5.4 Selection of mother wavelets ..... 174
3.5.5 Optimisation algorithm using single-objective optimisation technique ..... 178
3.6 Experiments and data collection ..... 180
3.6.1 Experiment setup: EML and GPR ..... 184
3.6.2 CAT and Genny survey ..... 184
3.6.3 GPR surveys ..... 190
3.6.4 EMC and GPR data provided by a third party ..... 191
3.7 Fusion of GPR, EML, and EMC data ..... 193
3.7.1 Multiple sensors data alignment and data correlation ..... 193
3.7.2 Kalman-based fusion of data (sensor level) ..... 196
3.8 Fused GPR-EML data inversion using FEM and GA ..... 201
3.8.1 Forward model of EM field due to multiple buried utilities ..... 202

3.8.2 The inversion algorithm using the FE forward model and a GA ..... 214
3.9 Summary ..... 219
**Chapter 4 RESULTS AND DISCUSSION ..... 222**
4.1 Introduction ..... 222
4.2 GPR survey data ..... 223
4.3 The semi-autonomous algorithm ..... 225
4.3.1 Optimisation of the GPR outcome using the MOGA analysis ..... 225
4.3.2 Evaluating the performance of the MOGA optimisation of KF ..... 229
4.3.3 Generated ROC curves from Pareto solutions (PSs) ..... 231
4.4 The autonomous algorithm ..... 234
4.5 Verification of the proposed algorithms ..... 243
4.5.1 Case study one ..... 243
4.5.2 Case study two ..... 248
4.5.3 Case study three ..... 252
4.6 Evaluation of the proposed semi-and-autonomous algorithms ..... 259
4.6.1 Evaluation using the chi-squared hypothesis testing method ..... 259
4.6.2 Evaluation using REFLEX software ..... 262
4.7 Fusion of data ..... 264
4.7.1 Fusion of experimentally obtained GPR and EML data ..... 264
4.7.2 Fusion of experimentally obtained GPR and EMC data ..... 269
4.7.3 Inversion of fused GPR and EML using FE forward model and GA ..... 272
4.8 Summary ..... 277
**Chapter 5 CONCLUSIONS AND RECOMMENDATIONS FOR FUTURE WORK ..... 281**
5.1 The semi-autonomous algorithm ..... 282
5.2 The autonomous algorithms ..... 284
5.3 The fusion of multiple geophysical data ..... 285
5.4 The inversion of multi-sensing data in complex buried utility scenarios ..... 286
5.5 Recommendations for future work ..... 286
**References 289**
Appendix I LITERATURE REVIEW ..... XXVI
Appendix II THE JOURNEY OF THIS RESEARCH PROJECT ..... XLVI
Appendix III AUC TABLES OF ROC CURVES ..... XXIX

# List of Figures


Figure 2-1. Overview of classified non-destructive methods for detection and condition assessment ..... 16

Figure 2-2. a) A-scan, b) Traveling time and depth, c) Radargram (B-scan) ..... 22

Figure 2-3. a) Experimentally collected A-scan, b) Associated radargram (B-scan) showing sand layers and a hyperbola signature for a buried concrete pipe ..... 23

Figure 2-4. a) Original GPR radargram prior to processing. The time zero (surface reflection) may not perfectly align with the top of the radargram. b) GPR radargram after applying time zero correction. The surface reflection is adjusted to the 'zero-time' reference, eliminating the offset seen in the unprocessed radargram and providing a more accurate depth interpretation of subsurface features (After Ciampoli et al., 2019) ..... 28

Figure 2-5. a) GPR radargram post time-zero correction, prior to signal gain adjustment. b) GPR radargram after both time-zero correction and signal gain adjustment, demonstrating enhanced feature visibility (After Ciampoli et al., 2019) ..... 30

Figure 2-6. a) GPR radargram prior to the application of a horizontal band removal filter. b) GPR radargram after application of a horizontal band removal filter, showcasing reduced horizontal banding noise (After Olhoeft, 2000) ..... 32

Figure 2-7. a) GPR radargram prior to the application of a band-pass filter. b) GPR radargram post application of a band-pass filter, demonstrating filtered frequency content (After Xie et al., 2013b) ..... 34

Figure 2-8. The principle of the wavelet transform as a multiresolution analysis of a signal. The chart illustrates a level 4 signal decomposition into approximations and detail components ..... 36

Figure 2-9. Example of feature integration data fusion ..... 57

Figure 2-10. Example of feature classification data fusion ..... 57

Figure 2-11. Response surfaces of the objective function for EM, GPR, and EMI-GPR data fusion. The surface depends on the input data and fusion formulation, i.e. using GPR and EMI data separately or together. A white star marks the true parameter value (After Moghadas, 2010) ..... 61

Figure 2-12. Example of automatic detection using curvelet and adaptive detection methods on a GPR radargram obtained from Savigny-Sur-Orge site, Paris. D) The interpreted GPR radargram by both methods (After Hafsi et al. 2018) ..... 62

Figure 2-13. Hidden Markov Model ..... 77

Figure 2-14. Optimal estimate through prediction and correction steps in a Gaussian world ..... 78

Figure 2-15. Kalman Gain distribution ..... 80

Figure 2-16. The general framework of Kalman Filtering ..... 81

Figure 2-17. Performance of the KF for detection of land mine in comparison to other methods using ROC curves (After Zoubir et al., 2002) ..... 83

Figure 2-18. GPR radargram before (left) and after applying the KF for reducing the clutter (right) (After Smitha and Singh, 2016) ..... 84

Figure 2-19. Signal response from a conductor to EMF induction using cable avoidance tool (CAT) based on (After Radiodetection, 2008) ..... 94

Figure 2-20. a) Direct connection to induce AC, b) indirect connection by clamping the AC into the conductor (After Radiodetection ,2008) ..... 95

Figure 2-21. Example of distorted electromagnetic field (coupling effect) caused by an adjacent utility (After Siu and Lai, 2019) ..... 95

Figure 2-22. The principle of utility detection using cable avoidance tool (CAT) (After Radiodetection, 2008) ..... 97

Figure 3-1. Flowchart of the semi-autonomous algorithm for processing GPR data ..... 116

Figure 3-2. GPR survey model ..... 119

Figure 3-3. Sample radargram from an experiment at NBIF at university of Birmingham consisting of a concrete pipe (d=30 cm) buried 30 cm deep in compacted sand ..... 120

Figure 3-4. Integration of basic, but fundamental, GPR signal processing filters ..... 126

Figure 3-5. Example of a noisy GPR radargram consisting of a tunnel and adjacent utilities ..... 127

Figure 3-6. Example of a noisy GPR radargram consisting of five buried pipes in shallow depth ..... 127

Figure 3-7. Overview of prediction of time zero position ..... 130

Figure 3-8. Radargram with a new time zero position after utilising the time zero correction ..... 130

Figure 3-9. Unprocessed radargram without applying bandpass filter. ..... 131

Figure 3-10. Radargram with applied bandpass filter ..... 132

Figure 3-11. Example of a radargram containing horizontal bandings associated with both informative (soil layers) and redundant HB (antenna itself and signal saturation between the antenna and the ground surface) ..... 133

Figure 3-12. Radargram with applied moving average filter ..... 133

Figure 3-13. Deviation of signals from zero by increasing the depth after utilising a gain function, a) original signal (randomly selected trace from radargram in Figure 3-6, b) graph of the applied gain function (exponential), c) Signal deviation (drift) after utilising the gain function ..... 135

Figure 3-14. Unprocessed radargram selected from the dataset ..... 136

Figure 3-15. Gain adjustment of a $40^{th}$ trace of the radargram. a) unprocessed trace, b) applied gain functions, c) gain adjusted trace ..... 138

Figure 3-16. The selected radargram after utilising the linear-exponential signal gain adjustment. ..... 138

Figure 3-17. Signal drift after utilising gain function ..... 139

Figure 3-18. Solved signal drift issue using windowed moving average ..... 139

*Figure 3-19.* Applying linear gain function, a) unprocessed trace, b) applied gain functions, c) gain adjusted trace, d) linear gain function applied on a selected radargram ...140

Figure 3-20. Applying exponential gain function, a) unprocessed trace, b) applied gain functions, c) gain adjusted trace, d) exponential gain function applied on a selected radargram ...141

Figure 3-21. Applying polynomial gain function, a) unprocessed trace, b) applied gain functions, c) gain adjusted trace, d) polynomial gain function applied on a selected radargram. ...142

Figure 3-22. Linear discrete Kalman filter flowchart. In the above equations, $\boldsymbol{xk}-$ is the a priori estimate, $\boldsymbol{Pk}-$ is the a priori state estimation error covariance, $\boldsymbol{xk}+$ is the a posteriori estimate, and $\boldsymbol{Pk}+$ is the a posteriori estimation error covariance. $\boldsymbol{K}k$ is the Kalman gains, $\boldsymbol{A}$ is the 'dynamic model's transition matrix', $\boldsymbol{B}$, $\boldsymbol{Q}$, are the 'input', 'system noise covariance', $\boldsymbol{P}$ is the 'estimation-error covariance', $\boldsymbol{H}$ is the 'measurement model matrix', and $\boldsymbol{R}$ is the 'sensor measurement noise covariance matrix' ...143

Figure 3-23. Example of innovation values in their vector form ...145

Figure 3-24. Example of created NIS signal using scalar NIS values by collecting them and putting them together in a new matrix. ...146

Figure 3-25. Normalised innovation (NIS) evaluation chart ...147

Figure 3-26. Detailed configuration of the buried pipes on a trial site ...148

Figure 3-27. a) Numerical GPR model of the trial site, b) Experimentally obtained GPR radargram from the trial site, c) Noise-free NIS-signal from numerical radargram, d) Noisy NIS-signal from the trial site ...149

Figure 3-28. Result radargram with allocated dashed lines produced by the KF after analysing the numerical model generated by the matGPR software. ...150

Figure 3-29. Experimental determination of the global NIS ...153

Figure 3-30. Overview of the multi-objective genetic algorithm (MOGA) optimisation procedure of the Kalman filter ...155

Figure 3-31. An example of a Preto solution curve for two cost functions $f1$ and $f2$ ...156

Figure 3-32. Pareto solutions of 6th radargram based on AuC of ROC curves and Fourier frequency (FFreq) ...162

Figure 3-33. Pseudocode for detection of target width ...163

Figure 3-34. Flowchart of the autonomous algorithm for processing GPR data ...165

Figure 3-35. An example of a) Time-domain non-stationary signal and b) the associated frequency-domain of the same signal ...167

Figure 3-36. Stationary sine wave with constant amplitude over time associated with a Fourier transform ...168

Figure 3-37. Example of a non-stationary NIS signal (signal of interest) to be processed by Fourier and wavelet transforms ...168

Figure 3-38. The expected Fourier transform's weighing function (weighted area) for the NIS signal after computing the dot product between the signal and the sine wave 168

Figure 3-39. Wavelet example (Morlet wavelet) with zero tapering at both ends ..... 169

Figure 3-40. The expected weighting function after computing the dot product between the NIS signal and a real-valued wavelet (e.g. Morlet wavelet here) yields spectrally specific information at the position of the wavelet. ..... 169

Figure 3-41. An example of a complex wavelet consisting of zero tapered sine and cosine waves corresponded to real and imaginary parts ..... 170

Figure 3-42. The expected weighting function after computing the dot product between the NIS signal and a complex wavelet (e.g. Morlet wavelet here) yields spectrally specific information at the position of the wavelet. ..... 171

Figure 3-43. An example of a wavelet coefficient consisting of real and imaginary parts ..... 171

Figure 3-44. Example of multiresolution analysis of NIS signal (level=4) ..... 173

Figure 3-45. Examples of evaluated wavelets and scaling functions: a) Daubechies, b) Symlet, c) Discrete Meyer ..... 176

Figure 3-46. The test pit at NBIF and buried HDPE pipe ..... 182

Figure 3-47. Buried pipes in the pit (NBIF), a) PVC water pipe buried at 1 m depth, b) Galvanised steel pipe buried at 0.3 m depth ..... 183

Figure 3-48. Cross- section showing the configuration of buried pipes and void in the test pit in NBIF ..... 183

Figure 3-49 Experiment setup in NBIF – surveyed points ..... 184

Figure 3-50. Schematic showing the gridlines for the EML and GPR surveys ..... 186

Figure 3-51. Direct connection of the Genny with the buried metallic pipe ..... 187

Figure 3-52. Example of an EM locator measurement using the RD8200G device with direct Genny connection to the metallic pipe ..... 187

Figure 3-53. Visualised signal of the EM locating survey as: a) strips to replicate the measurement direction, b) top view in which the buried pipe can be identified. ..... 188

Figure 3-54. Overlayed interpolated signals reconstructed from EML measurements ..... 189

Figure 3-55. Utilised GPR antennas for collecting data; a) IDS duo, b) Multi-channel Stream C. ..... 191

Figure 3-56. a) visualisation of the EMC survey indicating the dominance of inspection chambers with relatively higher apparent conductivities compared to buried utilities. b) the plan of the site showing all the buried utilities some of which run through inspection chambers (coloured blocks) ..... 192

Figure 3-57. Overview of data alignment and correlation in this project ..... 194

Figure 3-58. Schematic of data alignment of an (a) example GPR radargrams (510 sampling point and 1300 traces) with a signal penetration depth of e.g. 5 m and (b) EM38-MK2 conductivity meter data with significantly lower cover depth (0.75 m or 1.5 m), (c) the aligned data ..... 195

Figure 3-59. Example conductivity map from EMagPy software for a selected subset of the EMC data used in this thesis, showing the conductive utilities including the dominating inspection chamber on the right-hand side ..... 199

Figure 3-60. Schematic of the forward model of directly buried utilities utilised in the proposed inversion framework. $D1$ and $D2$ represent the cover depths of buried cables, $D3$ represents the separation distance between them. The diameters of the buried cables are represented by $Dp1$ and $Dp2$, respectively. $rp1$ and $rp2$ represent the radial distances of each cable from the origin of the coordinate system, and $\theta p2$ and $\theta p1$ are the angular coordinates of each cable with respect to the main coordinate system. ..... 206

Figure 3-61. Schematic of discretisation of the model using user-controlled meshing with a fine element size ..... 208

Figure 3-62. Schematic of discretisation of the model using physics-controlled meshing with a fine element size ..... 208

Figure 3-63. Statistic results produced by COMSOL associated with the user-controlled discretisation mode ..... 209

Figure 3-64. Statistic results produced by COMSOL associated with the Physics-controlled discretisation mode ..... 210

Figure 3-65. The experiment scenario by Siu and Lai (2019) for investigation of the electromagnetic field pattern caused by the coupling effect of adjacent buried conductive utilities ..... 212

Figure 3-66. Example results obtained from the FE model of the experiment reported by Siu and Lai, a) Electromagnetic field pattern, b) The magnetic flux density measured on the ground surface ..... 213

Figure 3-67. Flowchart of the proposed inversion algorithm consisting of surveys, numerical simulation using a FE forward model, constraining the variables using GPR information (data fusion), and the genetic algorithm optimisation ..... 218

Figure 4-1. Brocken pipe section (image: RSK) ..... 223

Figure 4-2. Selected radargrams: a) Radargram 9 with no defective pipes (five fully intact pipes), b) Radargram 13 with a broken pipe section at approximately trace 650 (four fully intact pipes and a broken soil filled pipe). ..... 224

Figure 4-3. a) Original radargram 009 with allocated dashed lines using PS4, b) By the KF processed radargram 009 using PS4 ..... 226

Figure 4-4. a) Original radargram 009 with allocated dashed lines using PS1, b) By the KF processed radargram 009 using PS1 ..... 227

Figure 4-5. (a) Original radargram 013 with allocated dashed lines using PS4, (b) By the KF processed radargram 013 using PS4 ..... 228

Figure 4-6. (a) Original radargram 013 with allocated dashed lines and detected ground anomaly by KF algorithm using PS1, (b) By the KF processed radargram 013 and detected ground anomaly using PS1 ..... 229

Figure 4-7. (a) NIS signal for experimentally obtained Q-R from radargram 009, (b) NIS signal for experimentally obtained Q-R from radargram 013 ..... 230

Figure 4-8. a) NIS signal for optimised Q-R from radargram 009 using PS4, b) NIS signal for optimised Q-R from radargram 013 using PS4 ..... 231

Figure 4-9. (a) Generate ROC curves for all radargrams in dataset using PS4, (b) Generate ROC curves for all radargrams in dataset using PS1 ..... 233

Figure 4-10. a) Radargram 9 with no defective pipes (five fully intact pipes), b) Radargram 13 with a broken pipe section at approximately trace 650 (four fully intact pipes and a broken soil-filled pipe), c) Radargram 9 with the removed background. d) Radargram 13 with removed background ..... 235

Figure 4-11. Processed NIS signals (left) and associated identified anomalies (right) for radargram 9 with five fully intact pipes, using db10 Daubechies wavelet under decomposition levels of a) l = 3, b) l = 6, and c) l = 7, showing the effect of wavelet decomposition level on the sensitivity of the algorithm ..... 238

Figure 4-12. Processed NIS signals (left) and associated identified anomalies (right) for denoised radargram 9 (FINDIT data) using the proposed autonomous algorithm with three different mother wavelets under the decomposition level (l) = 6, a) db20 wavelet, b) Sym20 wavelet, c) dmey wavelet ..... 239

Figure 4-13. Processed NIS signals (left) and associated identified anomalies (right) for denoised radargram 13 with 4 intact pipes and 1 broken pipe at approximately trace 650 (FINDIT data) using the proposed autonomous algorithm with three different mother wavelets under the decomposition level (l) = 6, a) db20 wavelet, b) Sym20 wavelet, c) dmey wavelet ..... 240

Figure 4-14. Heatmaps are associated with the scoring function for identifying the best mother wavelet for analysing the NIS signal. a) Heatmap for evaluation of the Daubechies wavelet, b) Heatmap for evaluation of the Symlet wavelet, c) Heatmap for evaluation of the Discrete Meyer wavelet ..... 242

Figure 4-15. Pareto solutions for the north section of the first survey site with different buried utilities ..... 244

Figure 4-16. Radargrams from the north section of the first survey site for two different cases of buried utilities. a) Original radargram 7, b) processed radargram 7, c) original radargram 10, d) processed radargram 10 ..... 245

Figure 4-17. Processed NIS signals (left) and associated identified anomalies (right) for denoised radargram 7 (case study one) using the proposed autonomous algorithm with three different mother wavelets under the decomposition level (l) = 6, a) db20 wavelet, b) Sym20 wavelet, c) dmey wavelet ..... 246

Figure 4-18. Processed NIS signals (left) and associated identified anomalies (right) for denoised radargram 10 (case study one) using the proposed autonomous algorithm with three different mother wavelets under the decomposition level (l) = 6, a) db20 wavelet, b) Sym20 wavelet, c) dmey wavelet ..... 247

Figure 4-19. Original radargram 4 from the experiment in NBIF consists of three pipes and a void buried in sand ..... 248

Figure 4-20. Pareto solutions for the NBIF GPR experiment ..... 249

Figure 4-21. Processed radargram using the semi-autonomous algorithm in case study 2 ... 250

Figure 4-22. Processed NIS signals (left) and associated identified anomalies (right) for denoised radargram 4 from the NBIF experiment using the proposed autonomous

algorithm with three different mother wavelets under the decomposition level (l) = 3, a) db20 wavelet, b) Sym20 wavelet, c) dmey wavelet ..... 251

Figure 4-23. Pareto solutions for the additional validation, case study 3 ..... 253

Figure 4-24. Additional validation of the performance of the proposed semi-autonomous algorithm using three different radargrams from the third site survey. All three processed radargrams are showing a successfully removed background, identifiable hyperbola signatures, and allocated dashed lines at the position of potential targets or ground anomalies. (a) original radargram 60, (b) processed radargram 60 (c) original radargram 19, (d) processed radargram 19, (e) original radargram 39, (f) processed radargram 39. ..... 254

Figure 4-25. Processed NIS signals (left) and associated identified anomalies (right) for denoised radargram 19 (case study three) using the proposed autonomous algorithm with three different mother wavelets under the decomposition level (l) = 5, a) db20 wavelet, b) Sym20 wavelet, c) dmey wavelet ..... 256

Figure 4-26. Processed NIS signals (left) and associated identified anomalies (right) for denoised radargram 39 (case study three) using the proposed autonomous algorithm with three different mother wavelets under the decomposition level (l) = 6, a) db20 wavelet, b) Sym20 wavelet, c) dmey wavelet ..... 257

Figure 4-27. Processed NIS signals (left) and associated identified anomalies (right) for denoised radargram 60 (case study three) using the proposed autonomous algorithm with three different mother wavelets under the decomposition level (l) = 5, a) db20 wavelet, b) Sym20 wavelet, c) dmey wavelet ..... 258

Figure 4-28. Comparison of the Carevic (1999) algorithm based on chi-squared hypothesis testing with the proposed algorithm in this research; a) user-defined window =300, b) user-defined window =150, c) user-defined window =75, d) result from semi-autonomous algorithm, e) result from autonomous algorithm. ..... 261

Figure 4-29. Comparison of the results between (a) the proposed autonomous algorithm in this research and (b) the REFLEX software for a sample case study, and (c) the associated NIS signal. ..... 263

Figure 4-30. a) Cable avoidance tool measurements (6 measurement lines), b) the measured signal strength associated with the measurement line 4 ..... 265

Figure 4-31. a) unprocessed radargram 4 from NBIF experiment consisting of three buried pipes and a void, b) the radargram 4 with applied zero correction, moving average filters (gain adjustment was not required due to the quality of data), c) result of the fusion of EML measurements, d) 3D view of the fusion of GPR and EML data ..... 267

Figure 4-32. a) unprocessed radargram 5 from NBIF experiment consisting of three buried pipes, b) the radargram 5 with applied zero correction, moving average filters (gain adjustment was not required due to the quality of data), c) result of the fusion of EML measurements in Figure 4-30 (measurement line 5) and the radargram 5. ..... 268

Figure 4-33. Fusion result. a) unprocessed radargram 19 from the trial site, b) EMC map inverted by the EMagPy software showing two conductive areas around 1100th and 1600th points of the X-axis, c) fused GPR and EM data ..... 270

Figure 4-34. Fusion result. a) unprocessed radargram 10 from the trial site, b) EMC map inverted by the EMagPy software showing two conductive areas around 400$^{th}$ and 2100$^{th}$ points of the X-axis, c) non-inverted EMC measurement associated with the radargram 10, d) fused GPR and EM ..... 271

Figure 4-35. Four different scenarios of two buried cables modelled using the developed FE forward model in section 3.8.1.3 (see Figure 3-57) to demonstrate different types of MF coupling effects, a) 30 cm horizontal separation between the two cables with the same burial depth, b) different burial depts with 30 cm separation, c) large separation (1 m) to minimise coupling effects, d) very short separation (1 cm) to demonstrate the maximum coupling effect. ..... 273

Figure 4-36. Example of numerically generated GPR radargrams using matGPR software associated with the corresponding forward model in Figure 4-35a ..... 274

Figure 4-37. Expected values versus estimated values (i.e. estimated positions and magnetic fields) by the numerical inversion-fusion algorithm associated for four different numerical scenarios ..... 275

# List of Tables

Table 2-1. References associated with the methods in Figure 2-1 ..... 17

Table 2-2. Overview of the most common and basic processing steps on GPR data ..... 25

Table 2-3. Overview of publications with focus on automated hyperbola recognition in GPR radargrams ..... 44

Table 2-4. Comparison of methods for hyperbola detection in GPR radargrams ..... 47

Table 2-5. List of works focused on remote sensing data fusion ..... 58

Table 2-6. Comparison of mathematical estimators in GPR application ..... 69

Table 2-7. Summary of works on GPR in correlation with Kalman filtering ..... 85

Table 2-8. Comparison of the relative advantages of EML and GPR ..... 91

Table 2-9. List of published work that has utilised inverted EMC data ..... 103

Table 3-1. Relationships between Kalman Filter parameters i.e., estimates and uncertainties. (After Simon, 2006) ..... 123

Table 3-2. Input variables of KF in GPR application ..... 126

Table 3-3. Overview of ROC curves and the area under these curves (AuC) for GPR application in terms of potential false alarms and missing targets associated with high sensitivity and high specificity setup of the system, respectively ..... 160

Table 3-4. Definition of terms used for generating the ROC curves (Carter et al., 2016) ..... 161

Table 3-5. Parameters of the soil and buried pipes in the NBIF experiments ..... 181

Table 3-6. Gridline set up and measurement point distances ..... 186

Table 3-7. Mathematical model of the system consisting of three sensors (GPR, EM, and EML) ..... 197

Table 3-8. Results associated with a sensitivity analysis for determining the computational cost of different element sizes ..... 211

Table 3-9. Example of suitable GA control options determined for the optimisation problem in this research ..... 217

Table 4-1. Parametric results associated with FE forward models in Figure 4-35 and Figure 4-37 ..... 276

# Nomenclature

| | |
|---|---|
| Area under the Curve | AuC |
| Cable avoidance tool | CAT |
| Continuous wavelet transform | CWT |
| Discrete wavelet transform | DWT |
| Electromagnetic | EM |
| Electromagnetic conductivity | EMC |
| Electromagnetic field | EMF |
| Electromagnetic locator | EML |
| Evolutionary algorithm | EA |
| Finite element method | FEM |
| Finite difference method | FDM |
| Fourier transform | FT |
| Fast Fourier transform | FFT |
| Fourier frequencies | FFreq |
| Genetic Algorithm | GA |
| Global NIS | GNIS |
| High voltage | HV |
| Infrared | IR |
| Infrared Thermography | IRT |
| Kalman Filter | KF |
| Low induction number | LIN |
| Low voltage | LV |
| Magnetic field | MF |
| Multi objective optimisation algorithm | MOGA |
| Neural Network | NN |
| Normalised Innovation Squared | NIS |
| Non-destructive testing | NDT |
| Pareto solution | PS |
| Receiver operating characteristic | ROC |
| Scan-cross-section | SCS |
| signal to noise ratio | SNR |
| Two Way Travel Time | TWTT |

# Notations

| | |
|---|---|
| $u$ | Measurement equation |
| $\xi$ | New measurement equation |
| $s^b$ | Background signal |
| $s^t$ | Target signal |
| $b$ | Bias |
| $v$ | Model noise |
| $\omega$ | Measurement noise |
| $\nu$ | Innovation parameter |
| $\sigma_v^2$ | Model noise covariance |
| $\sigma_\omega^2$ | Measurement noise covariance |
| $\hat{x}_k^-$ | A priori estimate |
| $\hat{P}_k$ | Estimation error covariance |
| $\hat{x}_k^+$ | A posteriori estimate |
| **A** | Dynamic model's transition matrix |
| **B** | Input matrix |
| **H** | Measurement model matrix |
| **Q** | System noise covariance |
| **R** | Sensor noise covariance |
| $K$ | KF Gain |
| $\theta$ | Phase angle |
| $\psi$ | Mother wavelet |
| $\hat{f}$ | Fourier transform |
| $G$ | Gain filter |
| α | Gain factor |
| **S** | Innovation covariance |

# Chapter 1 INTRODUCTION

## 1.1 Background

Buried infrastructure are important part of modern communities using which vital contents for the daily life of all community members are supplied (e.g. water and electricity). Millions of kilometres of utilities such as sewage, water mains, oil/gas transportation pipes, and telecommunication cables are distributed worldwide below the ground to provide services and resources to members of communities. Currently, in the UK alone there are more than 4 million kilometres of buried infrastructure utilities (Bricker and Frith, 2017).

Similar to above-ground infrastructure utilities, buried utilities are also subject to ageing and deterioration problems. Often lack of proper investigation, health monitoring and intervention can lead to deterioration and failures of buried infrastructure due to the ageing process (Clair and Sinha, 2012). Hence, continuous maintenance, repair, and replacement are essential in order to preserve an appropriate level of serviceability of these utilities proportionate to the needs of modern society (Davis et al., 2008). The complexity of today's modern infrastructure in populated urban areas is associated with a common issue of accessibility to these buried utilities. The challenge of accessing buried utilities starts with detecting and locating them at the first stage, as many are installed decades or, in some cases, centuries ago (Rogers and Overton, 2018). On the other hand, the lack of well-documented maps that provide information about accurate locations of buried utilities is a common issue (Metje et al., 2019).

Traditional ground investigation techniques are based on breaking the ground using for example trenches, and boreholes which can be costly, inaccurate, and destructive for both

buried utilities and their supporting ground (Makana et al., 2018). The cost associated with traditional ground investigations is not only limited to the project, but also the potentially enormous costs associated with disruption and delays in services such as traffic management systems (Metje et al., 2019). In addition to the financial disadvantages, as reported by the utility strike avoidance group, (2020) serious human injury and hundreds of fatal incidents worldwide can result from utility strikes, i.e. services such as high voltage cables, or gas pipes are hit by excavators (USAG, 2020). In the UK alone, 3000 utility strikes cases were reported in 2017/2018 followed by 2500 cases in 2019 in association with 10,000 and 12,000 construction works taking place, respectively (USAG, 2020, 2019). However, these numbers are limited to the officially reported cases, which could be higher if non-reported cases are considered.

In contrast to intrusive investigations, geophysical non-destructive testing/evaluation (NDT/NDE) methods, investigate the subsurface remotely by either focusing on mechanical properties, or searching for contrasts in the physical properties of the materials making up the subsurface (Hao et al., 2012). Utilising geophysical based methods for identifying the location and assessing the condition of buried utilities can be beneficial. The benefits include, more rapid and accurate data gathering, cost efficiency, non-destructive (e.g. trenchless or borehole free), decreasing the unnecessary disruptions in services caused by traditional methods, and finally reducing the risk of utility strikes as a result of unseen or unknown potential risk factors on each site (Milsom and Eriksen, 2011). In general, NDT techniques can characterise the soil properties in terms of spatial variability as well as detecting and locating shallow targets and anomalies such as pipes and voids (Doolittle and Brevik, 2014; Goodman and Piro, 2013a). Amongst all the available geophysical NDT methods, interest in utilising electromagnetics based techniques for investigating the shallow underground space (i.e. within the top 5 m) is

increasing (Reynolds, 2011). In particular, they are becoming popular options with geophysical surveyors due to a number of factors including *i)* being a relatively fast approach compared with other available geo-sensing technologies such as electrical resistivity tomography (ERT) that involves the measurement of the resistance of subsurface materials to the flow of an electrical current. ERT can be slow because it requires the deployment of specialised equipment (Pringle et al., 2021), *ii)* their ability to survey ground without a need for a direct contact *iii)* the relatively easy process to set up and use and *iv)* their relatively light weight which make them suitable for transporting to remote sites. Some of the commonly used EM based methods in this field are ground penetrating radar (GPR), electromagnetic locating (EML) also known as cable avoidance tools (CAT), and electromagnetic conductivity (EMC) which are discussed below. It is worth noting that despite their large potentials, the electromagnetic based NDT methods have their limitations such as limited depth penetration in different ground and operational conditions.

Amongst the EM based NDT techniques, GPR has been one of the most common techniques used for exploring and mapping the underground space for decades. In addition to the advantages of EM based techniques (some listed above), they have simple data collection, potentially capable of delivering high-resolution data (e.g. GPR), and delivering simultaneous live results during a survey which makes GPR a suitable method for a wide range of circumstances (Daniels, 2004). In a GPR survey, an electromagnetic pulse is transmitted into the ground and the reflected signal, called the backscattered signal, is sampled. The strength of the backscattered signals is proportional to the level of contrast in the physical properties (e.g., electrical conductivity, and dielectric permittivity) of the subsurface materials (Reynolds, 2011). The variation in the ground composition and physical properties (e.g., different soil layers), or the existence of subsurface targets and anomalies (e.g., buried utilities and voids,

respectively) has different impacts on the backscattered signals. The efficient signal penetration depth associated with each individual GPR antenna is proportional to the physical properties of the host medium and the frequency of the transmitted signal, called the central frequency (Jol, 2009). The outcome of GPR scans is usually presented in form of so called radargrams or B-scans. The processing of the captured data and the interpretation of the noisy GPR radargram images of the subsurface, are the biggest challenges of using GPR and is usually done by experts and therefore can be subjective and time consuming (Ciampoli et al., 2019). To assist practitioners, the data processing is carried out using a series of algorithms called filters that reduces or eliminates the unwanted signals (Daniels, 2004; Utsi, 2017a). Several filtering techniques have been developed some routinely used for commercial purposes with various degrees of success. Some examples are, particle filtering, (Ng et al., 2008), spectral domain features (Ho et al., 2004), maximum likelihood (Ho and Gader, 2002; Smitha and Singh, 2016), eigenvalue-based background removal (Khan and Al-Nuaimy, 2010), median filtering (Kim et al., 2007), parametric modelling (Van Merwe and Gupta, 2000), adaptive ground bounce (Wu et al., 2001), and singular value decomposition (Liu et al., 2017).

In parallel to the above, it is known that despite there being a wide range of NDT methods available for the detection, location and condition assessment of buried utilities, in isolation none of these methods is capable of capturing a comprehensive image of the subsurface (Metje et al., 2007). The ambiguity in the characteristics of the subsurface conditions (e.g., structural, and physical properties), is the main reason for these limitations. This issue can potentially be addressed by the fusion of data to enhance the quality of the analysis by increasing the level of confidence (Gros, 1997a). This has been recognised in standards both in the UK (Institution of Civil Engineers, 2014) and worldwide e.g. ASCE 38-02 (American Society of Civil Engineers, 2022), French DICT (Ministère de la Transition Écologique et

Solidaire, 2019), and Canadian CSA S-250 (Professional Surveyors Canada, 2019). PAS 128 highly recommend the use of at least two sensing techniques for higher quality subsurface surveys. Therefore, often site surveys with GPR are complemented with at least another sensing modalities e.g. EML or EMC.

EML is an EM based technique for detection and locating buried conductive masses i.e. masses capable of carrying electric currents such as metallic pipes and cables (Bryakin and Bochkarev, 2019). EML devices typically consist of two wire coils (transmitter and receiver). A time-varying electromagnetic field (EMF) is transmitted into the buried utility-ground system, causing electric current to follow along the conductive buried target. Consequently, a secondary EMF is then created around the conductive buried target which will be detected (measured) as an anomaly in Earth's magnetic field. Similar to GPR, the characteristics of the measured secondary EMF (signals), for instance, signal strength and frequency, depend on the characteristics of the transmitted signals and the physical property of both ground composition and the buried target. The location and potential depth of the buried metallic target can be estimated by the measured signals (Huang et al., 2019; Radiodetection, 2018; Waite and Wellaratna, 2006). EML can be operated in both active and passive modes i.e. actively generated EMF using transmitters or those generated by external sources such as the EMF around a live cable (power network) carrying electric current. A commonly used sensing device that works based on EML is called cable avoidance tool (CAT) which is used to detect buried conductors, e.g. live cables, by sensing the magnetic field due to the flowing current in the conductor. However, for EML techniques, such as CAT, to detect a buried conductor the current in the conductor is required to produce a magnetic field that reaches the surface. In addition, EML techniques are susceptible to coupling effect of magnetic fields which limits the

detection of multiple buried conductors in EML survey (Radiodetection, 2018; Siu and Lai, 2019).

Electromagnetic conductivity (EMC) is another commonly used EM based method used in industry for the measurement of the conductivity responses of the subsurface. The EMC method is used for mapping the underground space, detecting and locating of conductive materials such as metallic objects and objects containing sulphide ores (Reynolds, 2011). The principle behind this technique is the induction of controlled electromagnetic fields into the ground and measuring the secondary fields produced by electric currents flowing through the conducive materials in the ground (McNeil, 1980). EMC data typically consist of thousands of sampled data points distributed over the survey area down to the maximum penetration depth of the utilised EMC device (Guillemoteau et al., 2017). It should be noted that the EMC method is limited to detection and location of objects containing ferrous material such as metallic pipes or cables (Reynolds, 2011). There are various methods of inverting EMC data, which is the process of using EMC data to create a model of the subsurface, in order to reconstruct the relative conductive nature of the ground. These methods are discussed in more detail in the next chapter. However, the reliability of each inversion technique depends on the amplitude range of the acquired EMC data because the amplitudes can affect the accuracy of the model being created (Milsom and Eriksen, 2011). Furthermore, this technique requires prior assumptions for data analysis specific to each test site. These cause the data analysis in EMC technique to be a subjective practice and highly dependent on the knowledge of the expert user that can be subject to human error and bias (Moghadas, 2019; Nabavi, 2019).

## 1.2 Problem statement

As explained above, a major concern associated with the GPR, EML, and EMC is the need for time consuming and complex pre- and post-processing to analyse the data, which needs to be carried out by experienced surveyors. In addition, the need for human assessment and intervention can introduce errors that may lead to additional costs. Furthermore, data fusion from these sensing techniques suffers from additional limitations such as incompatibility, e.g. data sets with different units, and misalignment due to different characteristics of data in each technique, e.g. different penetration depth, recording frequency, and signal frequency of GPR, EML, and EMC data (Kvamme et al., 2019; Lenzerini, 2002).

A classic method to process GPR data for detecting buried targets is ground-signal removal (or background removal), in which the background signals in radargrams are estimated and then subtracted from noisy sampled signals (Tesfamariam, 2013a). Various algorithms have been proposed to remove the background and detect targets of interest. The main problem with these algorithms is the need to calibrate and adjust various parameters that require expert interventions at several stages of the process. Selecting best approach and choosing correct values of each algorithm require several trial and errors therefore, it is often a time-consuming job and may even lead to human errors. At the same time, the current practice for data fusion of GPR with EML/EMC involves a manual and laborious process where a geophysics expert compares and interrelates data from different surveys in order to infer the information of interest (Kvamme et al., 2019; Lenzerini, 2002). Several numerical procedures have been developed to assist with the data fusion processing. These numerical algorithms, whilst can assist with reducing the time required for the analysis, lack the necessary accuracy and/or are restricted to the range and characteristics of the data to be able to perform multi-sensing data

fusion (Saey et al., 2015b). The above collectively, makes the post-processing of site surveys using electromagnetic sensing techniques a challenging process and limit their applications.

## 1.3 Aim and Objectives

The aim of this research is to develop a number of novel, robust autonomous computational solutions using a combination of classic and artificial intelligence (AI) based processing algorithms to process ground penetrating radar (GPR) data as well as effectively fusing complementary data from non-destructive geophysical techniques, namely GPR, electromagnetic locator (EML) and electromagnetic conductivity (EMC), to improve detecting and locating of shallow buried utilities (e.g. pipes and cables) and ground anomalies (e.g. ground disturbance).

The identified objectives to achieve the aim of this research can be summarised as:

1. Critically review the literature on the current electromagnetic (EM) based geophysical non-destructive techniques and their hybrid use to identify the gaps in knowledge and areas which require improvement.

2. Through critically reviewing the literature, identify suitable geophysical data processing methods for noise reduction and decluttering of EM based data, detection of buried pipes and ground anomalies, and data fusion of EM based techniques.

3. Identify and collect suitable dataset through literature and/or acquire experimental data for evaluating the data processing methods proposed in this thesis.

4. Develop novel optimised pre- and post-processing computational solutions of GPR data making use of machine learning global optimisation.

5. Develop novel autonomous algorithms for pre- and post-processing of GPR data making use of the proposed computational solutions in Objective 4 in combination with classic signal processing techniques such as Fast Fourier Transform and wavelet decomposition.

6. Develop a novel algorithm for automating the fusion of GPR and EMC/EML data making use of the proposed algorithms in Objective 5.

7. Develop a novel algorithm for inversion of EM multi-sensing data in complex buried utility scenarios (e.g. multiple buried utilities with coupling effect) using numerical models and machine learning.

8. Validate and evaluate the performance of the developed algorithms against field and synthetic data and by comparing the results to conventional processing methods currently used in industry.

## 1.4 The original contribution of this research

A brief description of the original contributions of this research project is presented in this section. The research described in this thesis proposes methods to address and overcome the shortcomings of existing methods to both improve the quality of GPR data processing and to fuse data from GPR, EML and EMC techniques effectively and efficiently. To address the problems associated with data analysis and post-processing of survey data, machine learning (ML) techniques (in form of global optimisation search) were integrated with physics informed and classic data processing techniques. ML reduces/eliminates the need for human intervention whilst the use of classic data processing algorithms ensures the output of the algorithms are not in contradiction with the practicality of the problem.

To remove background noise from GPR radargrams, whilst improving buried pipes and ground anomaly detection, a new signal function was proposed that was implemented in a Kalman Filter (KF) algorithm which is a well-established numerical tool for estimation and prediction of state of a system from measured sensor data. Through this, a new KF process was developed. A multi-objective optimisation technique, i.e. multi-objective optimisation genetic algorithm (MOGA) was used to determine the parameters of the proposed KF by minimising the mean of innovation parameter of the KF, which is the residual of KF estimates and sensor measurements, on one hand and maximising the area under the curve of operator characteristic curves (ROC) on the other. The integration of MOGA and KF uses the best features of these techniques in order to interpret the location and condition assessment of shallow pipes and their supporting ground. Additionally, a further novel algorithm was proposed that combines a single objective genetic algorithm (SOGA), and wavelet transform into the Kalman filter. As a result, the computational time for processing GPR scans could be significantly reduced compared with available approaches without compromising the performance of the algorithm in terms of detection rate and accuracy.

Data from GPR, EML, and EMC, are fused using Kalman filtering (as a multi-sensing estimator) to overcome the shortcomings of each of these methods. An innovative numerical procedure is introduced to fuse data from GPR, EML, and EMC using Kalman filtering in an effective and efficient approach. A potential application could be where GPR is not capable of detecting and locating very small diameter cables, EML and EMC can be useful complementary methods such that thin metallic cables can be identified.

Fourth and final contribution of this thesis is an inversion algorithm that combines a finite element model of a coupled magnetic field as a forward model and a genetic algorithm to

detect buried live cables in close adjacent and cannot be distinguished by EM locators otherwise.

## 1.5 Structure of the thesis

This thesis is organised in five chapters, which are described as follows:

**Chapter 1** introduces the thesis containing: a brief description of non-destructive testing methods for detecting and locating buried utilities, the problem statement, the aim and objectives, the contribution of this research, and finally the structure of the thesis.

**Chapter 2** presents a detailed critical literature review of most common geophysical nonintrusive testing method namely, ground penetrating radar (GPR), GPR data processing and interpretation, data fusion methods, mathematical estimators for data fusion, Kalman filter (KF), further electromagnetic based method i.e. electromagnetic locating (EML), electromagnetic conductivity (EMC), and EMC inversion methods. Finally, a summary of the gaps in knowledge identified from the review is presented.

**Chapter 3** is the methodology chapter that briefly explains the use of Kalman filtering as a stochastic estimation method for processing GPR data and for the fusion of GPR, EML, and EMC data. Furthermore, a brief explanation of KF parameters such as innovation vector, normalised innovation squared (NIS), receiver operator characteristic (ROC), Fast Fourier Transforms (FFT), optimisation methods making use of multi- and SOGA, wavelet transforms, inversion of EMC data, and numerical modelling and inversion of EML data are presented. Chapter 3 proposes four algorithms for this research. Two of which are for processing the GPR data in a semi- and autonomous fashion followed by the third algorithm for the fusion of GPR, EML and EMC data, aimed at the detection and location of shallow pipes and potentially

ground anomalies. Finally, the fourth algorithm which is a numerical modelling-based inversion-fusion algorithm making use of genetic algorithm (GA) and COMSOL Multiphysics for the inversion of the EM locator data by fusing the information obtained from the GPR. Chapter 3 concludes by summarising the proposed computational solutions for detecting and locating shallow pipes.

**Chapter 4** is the results and discussion chapter that presents the results and discussions from the above listed algorithms. These include: *i)* an analysis of the effect of employing various input parameters of Kalman filter and Fast Fourier frequencies obtained from pareto solutions of multi objective genetic algorithms (MOGA) on GPR data; *ii)* an analysis of the effect of various mother wavelets in association with the optimised Kalman filter, making use of a single objective genetic algorithm (SOGA), for detecting and locating buried pipes; *iii)* validation and evaluation of proposed algorithms, and *iv)* the results and discussion for the fusion of GPR data with the EML and EMC data, as complementary sensors, for improving the detection and location of conductive pipes. All sections have summaries of the main findings and results.

**Chapter 5** presents the conclusions of this research project and provides recommendations for further research.

# Chapter 2 LITERATURE REVIEW

## 2.1 Introduction

For decades, employing geophysics for the development of remote nonintrusive non-destructive sensing methods, and equipment has been an active area of research for exploring underground space. This chapter focuses on critically reviewing the existing works on nonintrusive non-destructive methods for the exploration of the shallow subsurface. It should be noted that the term 'shallow' in this research refers to the near-surface underground space which is defined by Boddice et al. (2019) as the top 5 m from the surface. Some of the important applications of these methods are: producing models showing the structure of the underground space, detecting and locating buried infrastructure assets (e.g., tunnels, pipes, sewers, cables), identifying subsurface anomalies (e.g., voids, cavities, and soil discontinuities), and finally health monitoring of buried utility services as a proactive approach for their maintenance (Costello et al., 2007). The benefits and drawbacks associated with conventional subsurface exploration methodologies are also presented. In addition, the fusion of data obtained from these methods is reviewed.

Section 2.2 of this chapter presents a general overview and classification of the conventional methodologies for detecting and locating buried infrastructure utility services, as well as the health monitoring methods for the ground and buried assets. The background, theory, and definitions behind these methods with a specific focus on electromagnetism are presented in Appendix I. Section 2.3 reviews the Ground penetrating radar as a nonintrusive electromagnetic based methods in subsurface exploration with a focus on challenges in GPR data processing and interpretations. Section 2.4 reviews the data fusion of nonintrusive

techniques and the existing work that exploits data fusion in subsurface exploration. Section 2.5 focuses on the mathematical estimator as a powerful contributor for reducing noise and clutter from GPR data and a fusion tool for the data obtained from different NDT techniques Section 2.6 reviews two commonly used electromagnetic based techniques, namely electromagnetic locating (EML) and electromagnetic conductivity (EMC). Finally, section 2.7 presents a summary of this chapter.

## 2.2 Overview and Classification of Nonintrusive Methods

Subsurface explorations are typically performed to achieve specific goals, such as detecting and locating buried infrastructure and ground anomalies, mapping and reconstructing the subsurface structure, detecting and locating voids and cavities, and assessing the condition of buried infrastructure and their surrounding ground (Burger et al., 2006; Reynolds, 2011). All materials on Earth are characterised by their physical properties, which are any measurable property of the material that describes the state of a physical system (Milsom and Eriksen, 2011). The diversity of physical properties in ground materials (whether this is the ground itself or buried anomalies within it) provides valuable information, and researchers have developed a wide range of non-intrusive techniques utilising these properties to remotely investigate the subsurface (Reynolds, 2011). The success of a subsurface exploration project is highly dependent on the selection of a suitable technique that is in good agreement with the physical properties of the ground. Therefore, it is crucial to understand such properties and neglecting this importance can lead to the selection of ineffective techniques for certain scenarios (Jeong et al., 2003; Rogers et al., 2012).

The results obtained from each technique vary in different ways, but all are associated with the physical properties of the materials they are investigating. Some of them deliver an estimate of different parameters of the subsurface, whereas some produce spatial images (one- two- or three-dimensional images). For instance, ground penetrating radar (GPR) as an electromagnetic technique delivers subsurface images by utilising the electrical properties of materials in the ground composition by sampling the EM waves reflected from the subsurface (Annan, 2009; Koppenjan, 2009; Stern, 1929), while other method such as impact echo method focuses on mechanical properties of the ground, particularly the elastic properties, by sampling the refracted seismic waves on the ground (Lenham et al., 2005).

From this starting point, a summary of conventional geophysical non-destructive techniques in the literature for both detection and condition assessment of buried assets, classified based on the governing physics and physical properties is presented in Figure 2-1. The associated references are presented in Table 2-1.

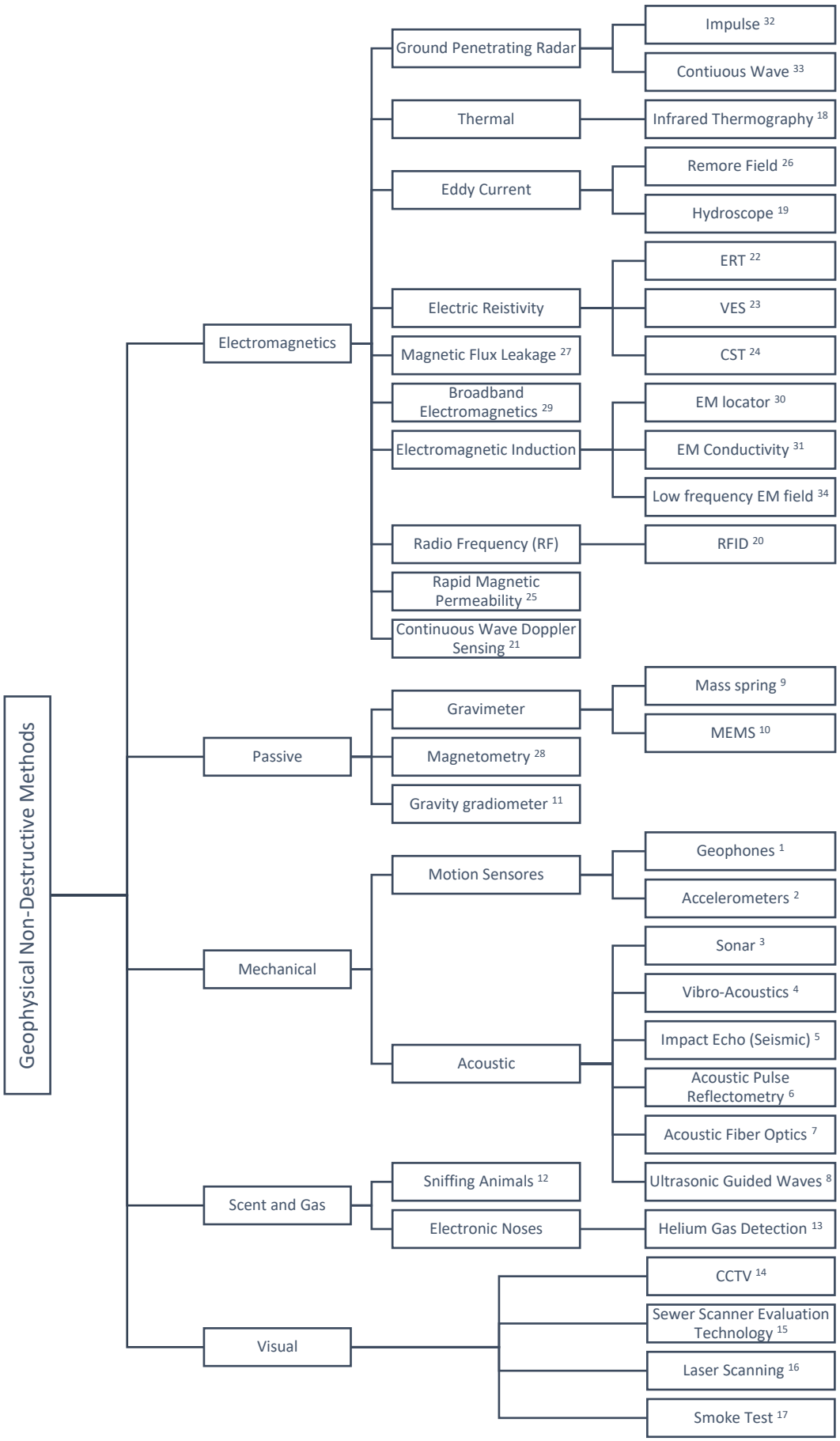



*Figure 2-1. Overview of classified non-destructive methods for detection and condition assessment*

*Table 2-1. References associated with the methods in Figure 2-1*

| | | | | | |
|---|---|---|---|---|---|
| **1** | (Nayana et al., 2021) | 13 | (SUEZ, 2021) | 25 | (Roubal, 2018) |
| **2** | (Pal et al., 2010) | 14 | (Haurum and Moeslund, 2020) | 26 | (Li, 2015) |
| **3** | (Fox, 2011) | 15 | (Wirahadikusumah et al., 1998) | 27 | (Jiles, 1990) |
| **4** | (Muggleton and Rustighi, 2013) | 16 | (Gunatilake et al., 2021) | 28 | (Vo et al., 2020) |
| **5** | (Kang et al., 2017) | 17 | (Ratliff, 2003) | 29 | (Zheng and Liu, 2011) |
| **6** | (Mohd Shahrul Azmi, 2019) | 18 | (Carreño-Alvarado et al., 2014) | 30 | (Siu and Lai, 2019) |
| **7** | (Muggleton et al., 2020) | 19 | (Makar and Chagnon, 1999) | 31 | (Moghadas, 2019) |
| **8** | (Sun and Zhu, 2020) | 20 | (Abdelnour et al., 2018) | 32 | (Pringle et al., 2020) |
| **9** | (Middlemiss et al., 2016) | 21 | (Bimpas et al., 2010) | 33 | (Pasternak and Kaczmarek, 2019) |
| **10** | (Tlemçani et al., 2018) | 22 | (Perrone et al., 2014) | 34 | (Tabbagh and Panissod, 2000) |
| **11** | (Difrancesco et al., 2009) | 23 | (Perrone et al., 2014) | | |
| **12** | (Hadagali and Suan, 2017) | 24 | (Perrone et al., 2014) | | |

A thorough literature review was conducted on all techniques presented in Figure 2-1. However, the hypothesis of this thesis is to identify gaps in electromagnetic-based methods, specifically in regard to the use of ground-penetrating radar as a standalone technique or in conjunction with other non-invasive methods. Consequently, the literature review pertaining to techniques not related to this research, as well as the underlying principles and theory of electromagnetics (including aspects of GPR), can be found in Appendix I of this thesis.

# 2.3 Ground Penetrating Radar (GPR)

## 2.3.1 Introduction

Ground penetrating radar (GPR) is a widely used geophysical non-destructive remote sensing technique for imaging and investigating the subsurface (Gebremichael, 2013). As an electromagnetic method, GPR works by transmitting electromagnetic waves into the ground and measuring the reflected waves to obtain information about the subsurface (Daniels, 2004; Goodman and Piro, 2013b). Different antennas and a range of central frequencies (up to several GHz) can be used to achieve different penetration depths and resolutions, which are proportional to the antenna's central frequency (Benedetto et al., 2017).

The first patent for a radar device related to GPR was submitted in 1904 in Germany by Christian Hülsmeyer, called the 'Telemobiloskop' (patent DE 165546). One of the earliest recorded GPR surveys was conducted in 1929 to measure glacier depth in the Austrian Alps (Stern, 1929). However, the routine use of GPR for surveying began in the 1960s as scientists started using radio waves of various frequencies to probe the dielectric properties of soil materials (Annan, 2009).

GPR has been successfully applied in a variety of subsurface investigations. Some common examples include identifying water tables in the ground (Bentley and Trenholm, 2002), mapping soil water content and moisture (Mahmoudzadeh Ardekani, 2013), glaciology (Binder et al., 2020), differentiating layers and interfaces in ground composition, particularly with regard to layer thicknesses (Binder et al., 2020; De Coster et al., 2018; Khamzin et al., 2017), detecting and locating buried infrastructure and service utilities (De Coster et al., 2019; Prego et al., 2017), water leakage detection (Bimpas et al., 2010; Lai et al., 2016), forensic investigations (e.g., graves and homicide) (Pringle et al., 2021, 2020, 2012), subsurface

mapping (Metje et al., 2019), detecting and locating voids and cavities (Baek et al., 2018; Luo and Lai, 2020), heritage and archaeological explorations (Karamitrou et al., 2020), unexploded ordnance (UXO) detection (Abdelgwad and Nashat, 2017; Núñez-Nieto et al., 2015), concrete structure scanning and assessments (Sbartaï et al., 2012), roads pavement inspection (Zhao and Al-Qadi, 2017), asphalt crack detection (Torbaghan et al., 2020), and recently used by NASA and incorporated into their rover robot to investigate the Mars subsurface (Hamran et al., 2020). Moreover, GPR is also a suitable technique for scanning and investigating the structural elements such as walls, decks, slabs, tunnels etc. made of different materials (e.g., concrete, masonry) due to its minimal access requirement (i.e. single-sided access).

While GPR is a widely used technique in the industry, processing and interpreting GPR data can be challenging and requires skilled and qualified experts due to the complexity of the data collected, which requires specialised software and techniques for accurate analysis and interpretation (Wahab et al., 2013). As mentioned earlier, the key factor in the success of a GPR survey is the contrast in electrical properties of the subsurface materials. If there is no (or very low) contrast, the method may not be useful (Daniels, 2004). Although GPR can detect anomalous targets, it cannot determine the composition of the material (Jol, 2009). The ability of GPR to detect and locate subsurface anomalies of different sizes at different depths is influenced by the resolution and penetration depth of the electromagnetic waves, which are in turn influenced by the central frequency of the system and the ground conditions (Jol, 2009). These factors can affect the performance of a GPR system in detecting anomalies under different circumstances (Wang et al., 2014). For example, a GPR system with a low central frequency may have difficulty detecting thin buried cables (Dou et al., 2016). The similarity of GPR frequencies with certain signal-generating equipment, such as mobile masts or medical equipment that produce electromagnetic fields, can also create noise, and mask the data. In

addition, electromagnetic waves cannot propagate through highly conductive soils with lossy dielectric characteristics, which can cause rapid absorption of wave energy and result in shallow penetration of the GPR signal (Dou et al., 2016). This is because high conductivity allows the electrical current to flow easily, causing the waves to be absorbed rather than transmitted through the soil (Witten, 2014). Examples of lossy conductive soils include wet clays and salt-saturated soils (Farzamian et al., 2019).

While many research papers have identified water content and saturation of soil as a source of noise and fuzzy areas in GPR radargrams (Bai and Sinfield, 2020; Bi et al., 2018; Ho et al., 2004; Solimene et al., 2014; Wai-Lok Lai et al., 2014), it is worth noting that water and soil saturation are not always a disadvantage for GPR and can be beneficial in certain circumstances. For example, water can improve the detectability of certain targets such as air-filled plastic or concrete pipes, which are normally difficult to detect with GPR due to their similar dielectric properties to the surrounding soil, but the presence of water can make them more distinguishable by increasing the impedance between such targets and the surrounding soil (Bigman, 2012). In a study by Allred et al., (2013), the detectability of air-filled shallow pipes versus water-filled shallow pipes in saturated soil was investigated using a GPR antenna with a 250 MHz central frequency. Despite this, the authors reported that more than 27% of the surveyed agricultural drainage pipes in specific areas, such as in Ohio, USA, could not be detected. The detection rate of water-filled pipes was also reported to be poor, attributed by the authors to the low impedance contrast between water and the pipe material.

As stated by Gonzalez-DIaz et al. (2020), strong detectable signals, i.e. reflected signal after encountering buried targets offering high level of resistance to the flow of an electrical current, resulting from high impedances can also be associated with the presence of metallic objects, which make them ideal candidates for detection by GPR. However, this is not always

an advantage because electromagnetic waves produced by a GPR antenna cannot penetrate metals and will be scattered after interacting with them. As a result, the investigation of the ground beneath a metallic object (e.g., a steel plate) using GPR is often uncertain and targets (depending on their size) or even soil layers beneath may remain invisible (Utsi, 2017a). On the other hand, in some cases, the corrosion of aged metallic objects (e.g., cast-iron water mains) can contaminate the surrounding soil to a large extent, producing highly conductive soil (Pennock et al., 2014). Some studies on the corrosion of cast iron pipes and its impact on GPR detection include Mogharehabed (2014), Schmidt et al. (2006), and Pennock et al. (2014). These studies suggest that the migration of iron ions into the soil can significantly change the dielectric properties of the soil, which can in turn significantly affect the propagation of GPR signals in terms of reflection, energy decay, and attenuation.

### 2.3.2 GPR data

Sampled reflected GPR data can be visualised as an A-scan (Figure 2-2a), a waveform of individual GPR traces that consists of peaks indicating reflectors with permittivity contrast (Skartados et al., 2019). The position of the peaks along the trace is determined by the reflection time (two-way travel time (TWTT)) of the electromagnetic waves (Figure 2-2b). Plotting A-traces against distance creates radargrams (B-scans), which are radar cross-sections (RCS) providing indirect information about the subsurface structure in a scanned area perpendicular to the GPR moving direction (Figure 2-2c). Buried features can be identified through hyperbola-shaped signatures in radargrams created by concatenated A-scans, with the position of the signatures indicating the actual position of the buried object in the subsurface (Benedetto and Benedetto, 2014).

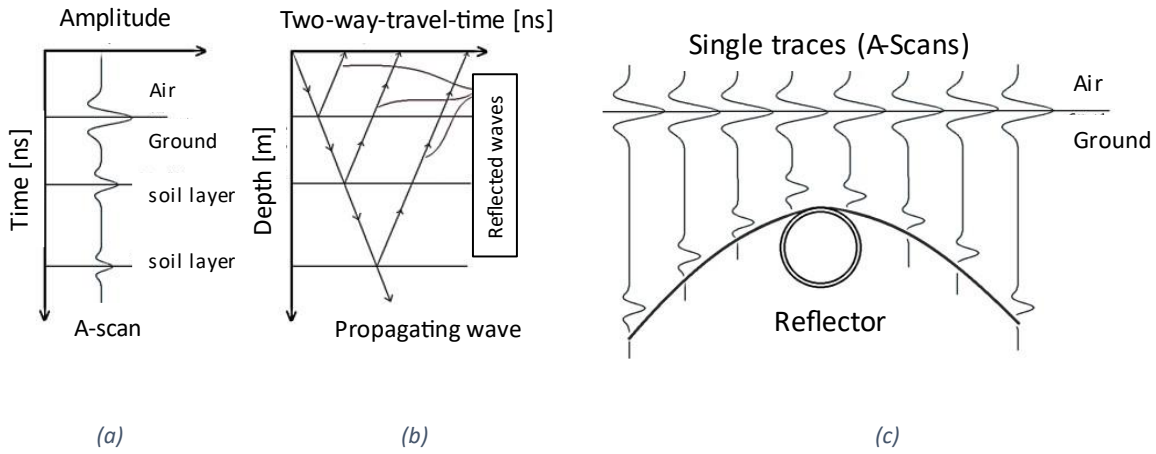
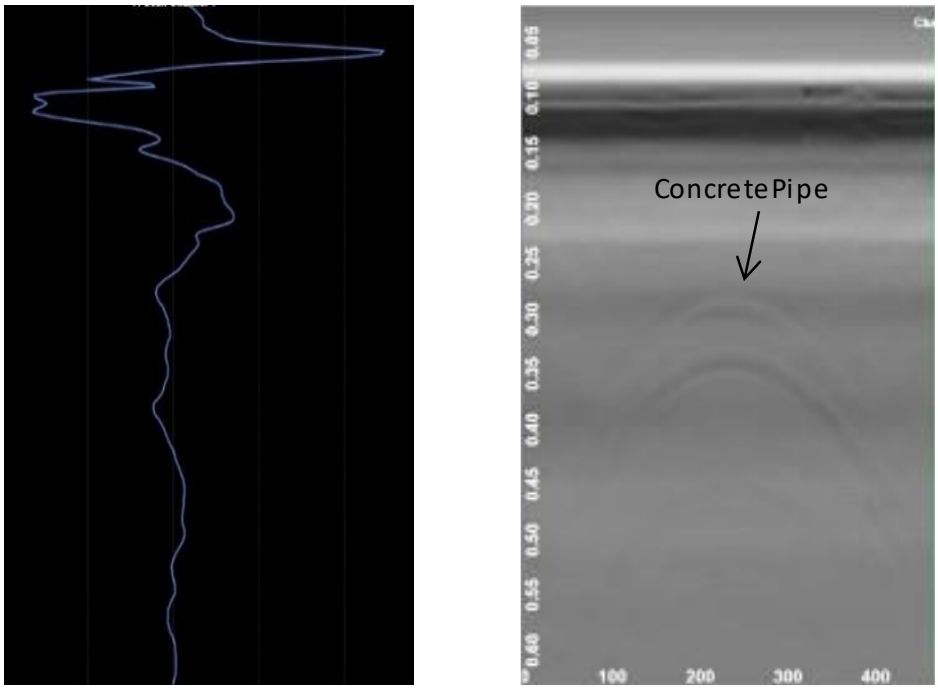



*Figure 2-2. a) A-scan, b) Traveling time and depth, c) Radargram (B-scan)*

However, raw radargrams are often subject to intense noise from the surrounding environment (both on and beneath the surface). Identifying (extracting) hyperbolas in radargrams can be challenging as they do not always have a clear signature shape (Benedetto and Benedetto, 2014). For example, radargrams from dense infrastructure scenarios often have noise and interfere with hyperbola signatures due to close or overlaid utilities such as pipes and cables. Figure 2-3 shows an example of GPR data (consisting of both A-scan and B-scan) with relatively low noise from an experiment conducted under controlled laboratory conditions. The following sections will review different aspects of GPR data filtering and processing.




*(a)* *(b)*

*Figure 2-3. a) Experimentally collected A-scan, b) Associated radargram (B-scan) showing sand layers and a hyperbola signature for a buried concrete pipe*

## 2.3.3 Processing GPR data

Signal processing is a crucial step in geophysical surveys (GPR included) for analysing and improving the clarity of data in the time and frequency domains. It involves using computer-aided algorithms and filters to extract meaningful information and reduce noise from the collected data (Abraham, 2017; Benedetto et al., 2017; Van de Vegte, 2002). Signal processing has applications beyond geophysics and is also used in medical science and aerospace, but these are outside the scope of this research (Bailey and Swarztrauber, 1994; Daubechies, 1990; Mallat, 2009a; Rader and Mailing, 1967). The success of a GPR survey depends on various factors such as selecting an appropriate GPR device, configuring the equipment, and conducting the survey properly (Daniels, 2004). GPR data is often contaminated by noise and clutter, known as background signals, which can obscure important information and make data interpretation difficult (Van Kempen and Sahli, 2001). Dou et al. (2017) divided the

sources of noise in GPR data into three categories: i) system noise, which is generated by the GPR system itself and can include low-frequency components such as signal saturation or "wow noise" that occur during short impulse intervals (Cassidy, 2009a). The causes of signal saturation can include reflections from the ground surface, near-surface reflectors, air waves, and the coupling effect between the transmitter and receiver due to their proximity (Annan, 2009; Tesfamariam, 2013b), ii) non-homogeneity of the host medium, such as multi-layer soils with different physical properties or variations in water content at different depths (Dou et al., 2017), iii) interactions between transmitted and reflected signals, which can mask weaker signals from deeper subsurface anomalies (Pieraccini, 2018).

Noisy data must be pre-processed and post-processed using techniques such as signal and image processing to reduce or eliminate noise and clutter, and improve the quality of data interpretation (Allred et al., 2013). This can significantly improve the readability and ease of interpretation of the data by highlighting healthy signals and hyperbola signatures in radargrams. However, the task of noise cancelling and filtering survey data, such as GPR data, to interpret the subsurface structure and detect and locate buried services such as pipes is a delicate one (Utsi, 2017b). On the one hand, the goal is to remove unnecessary signals (e.g., noise and clutter) from the collected data, but on the other hand, it is important to maintain and not eliminate the crucial information associated with targets during the processing stage. In other words, the processing of survey data must be tailored to the properties of the sampled data. These types of problems are known as supervised problems (Mallat, 2009a).

There are common techniques for processing raw GPR data which enhance its quality and extract useful information. Examples include de-wowing, filtering, and signal gain adjustments (Che Ku Melor et al., 2021; Daniels, 2004; Economou et al., 2017; Rao et al., 2010).

Iswandy et al. (2009) divided the typical process of GPR data processing into three important steps. First, define an appropriate framework that includes all suitable processing steps for each dataset. Second, specify all parameters related to each planned step of the framework. Third, evaluate the results and improve the outcome if necessary.

The most common basic approaches for processing of GPR data are illustrated in Table 2-2.

*Table 2-2. Overview of the most common and basic processing steps on GPR data*

| | | |
|---|---|---|
| **Data Editing** | Applying, modifying, and rearranging information of data | (Park et al., 2019) |
| **Wave Velocity Estimation** | Estimating GPR wave propagation velocity through host medium | (Greaves et al., 1996) |
| **De-Wowing** | Eliminating very low-frequency signals and DC bias components of the data | (Wong et al., 2016) |
| **Time-Zero Correction** | Correction of start time to match with surface position | (Zadhoush, 2021) |
| **Signal gain adjustments** | Applying gains to the signal to improve the visibility and ease of interpretation | (Olhoeft, 2000) |
| **Horizontal Banding Removal** | Correction of data to ensure spatially uniform increments | (Downs and Jazayeri, 2021) |
| **Filtering** | One and two dimensions filtering to improve the signal to noise ratio and visual quality | (Kovalenko and Masalov, 2000) |
| **Deconvolution** | Contraction of signal wavelets to enhance reflection event | (Kadioglu, 2018) |
| **Elevation Correction** | Correcting for the effect of topography | (Busby and Merritt, 1999) |
| **Migration** | Corrections for the effect of survey geometry and spatial distribution | (Özdemir et al., 2014) |
| **Depth Conversion** | Conversion of two-way travel times (TWT) into depths | (Conyers and Lucius, 1996) |
| **Image Analysis** | Using pattern or feature recognition tools and methods | (Szymczyk and Szymczyk, 2013) |
| **Attribute Analysis** | Attributing signal parameters or functions to identifiable features | (Dossi et al., 2015) |
| **Modelling Analysis** | Simulating GPR responses making use of numerical models | (Warren et al., 2016) |

Although there are various techniques for processing GPR data (as shown in Table 2-2), there are four main techniques that are most commonly used by practitioners and are considered the backbone of GPR data processing, addressing essential aspects such as timing and depth precision, signal amplification, noise reduction, and frequency enhancement (Huber and Hans, 2018). These techniques are i) time zero correction, ii) signal gain adjustment, iii) background filter, and iv) Band-pass filtering, which are discussed below in detail.

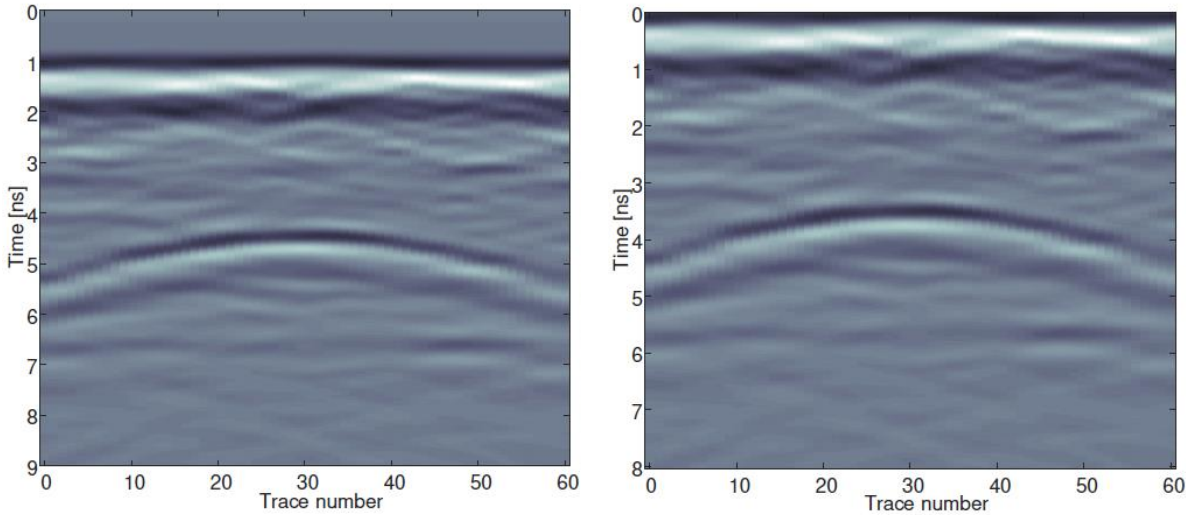
### *Time zero correction (zero offset removal)*

Accurate depth estimation of targets (or anomalies) in the subsurface using GPR requires accurate estimation of the true wave propagation time (travel time) between the transmitter and receiver antennas. However, recorded scattered GPR signals are often subject to the coupling effect between the antenna and the ground surface, which causes an offset of the true time zero in the recorded radar signals (Conyers, 2013; Rasol et al., 2018). The true time zero, at which the first signal reflection event from the ground surface is recorded, can be estimated by eliminating this coupling effect and shifting the first signal reflection event to the zero position (Huber and Hans, 2018).

The time zero correction is an effective technique for processing GPR data where the start time of each trace is adjusted to match the precise moment the radar pulse was transmitted. As highlighted by Zadhoush (2021), even though this technique holds importance in GPR data processing, it comes with certain limitations. For example, the effectiveness of this technique heavily depends on the accurate measurements of the signal velocity, which can pose challenges to its application in highly heterogeneous or layered soils. Additionally, its application can be time-consuming, particularly for large datasets. Furthermore, depending on the goals of the GPR survey and the site characteristics, it may not always be necessary.

Despite these limitations, time zero correction is a crucial step that provides a more accurate time reference, thereby facilitating a more precise interpretation of the depth of subsurface features reflected in the GPR data (Zadhoush, 2021).

As an example, Figure 2-4 shows a comparison of a GPR radargram both before and after the application of time zero correction. Ciampoli et al. (2019) proposed a multi-stage signal processing method for GPR data collected on flexible pavements. This involved time zero correction, signal gain, and band-pass filtering. The researchers demonstrated the effectiveness of time zero correction using a dataset obtained from a rural road in Madrid, Spain, utilising a 1.5 GHz GPR antenna. Figure 2-4a shows the original GPR radargram, prior to any processing. At this stage, the time zero – the exact moment of the radar pulse transmission – may not align perfectly with the top of the radargram due to various factors such as slight timing errors in the system (Ciampoli et al., 2019). This discrepancy can create an offset in the depth interpretation of subsurface features. Figure 2-4b, on the other hand, presents the same radargram after the time zero correction has been applied. The surface reflection is now adjusted to the 'zero-time' reference, eliminating the offset seen in the unprocessed radargram. This alignment facilitates a better interpretation of the depth of subsurface features, and forms a crucial preparatory step for further processing and analysis of GPR data (Ciampoli et al., 2019).




*(a)* *(b)*

*Figure 2-4. a) Original GPR radargram prior to processing. The time zero (surface reflection) may not perfectly align with the top of the radargram. b) GPR radargram after applying time zero correction. The surface reflection is adjusted to the 'zero-time' reference, eliminating the offset seen in the unprocessed radargram and providing a more accurate depth interpretation of subsurface features (After Ciampoli et al., 2019)*

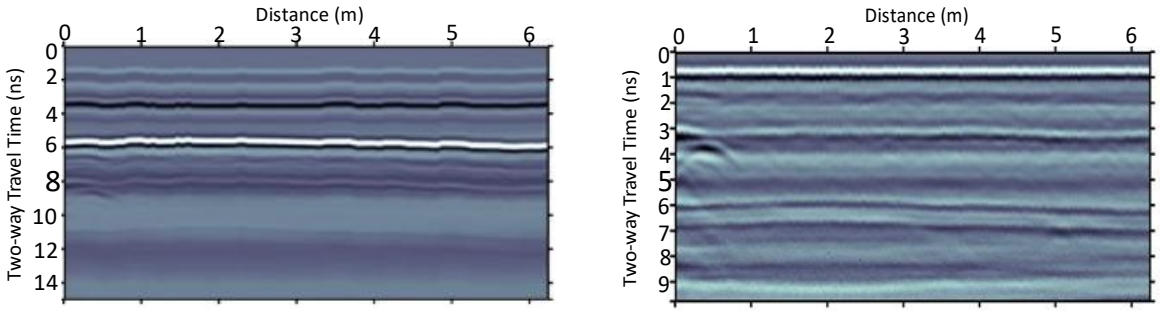
## Signal gain adjustments

The attenuation of GPR signals increases with depth due to the loss of impulse energy as the waves propagate into the ground. In other words, the attenuation of GPR signals increases because the GPR signals encounter various materials as they travel through the ground, which can absorb or scatter the energy of the waves (Cassidy, 2009b). This results in a decrease in the amplitude of GPR signals with increasing depth, making deeper targets and their associated hyperbola signatures difficult to distinguish from background noise and clutter. This results in a decrease in the amplitude of GPR signals with increasing depth, making deeper targets and their associated hyperbola signatures difficult to distinguish from background noise and clutter (Rhee et al., 2021). To address this issue, the scattered signals can be amplified by multiplying the unprocessed initial trace (A-scans) by a gain function or coefficient (Daniels, 2004). Linear and exponential gain functions, or a combination of both,

are effective ways to adjust the scattered signal amplitudes based on the conditions and quality of the data (Goodman and Piro, 2013a; Huber and Hans, 2018).

Signal gain adjustments can be useful to correct variations in signal amplitudes caused by changes in the distance between the antenna and the ground or by variations in the soil properties. However, there are a few limitations to consider when using signal gain adjustments. It can introduce distortions or artifacts into the GPR data if the gain is set too high or too low, and it can be subjective, as the appropriate gain settings may vary depending on the goals of the GPR survey and the characteristics of the site (Everett, 2013). Additionally, it may not be effective for correcting for certain types of variations in signal amplitudes, such as those caused by changes in the frequency of the GPR signals (Goodman and Piro, 2013b).

Figure 2-5, adopted from Ciampoli et al. (2019), illustrates the significant transformation of a GPR radargram via two crucial processing stages, i.e. time-zero correction and signal gain adjustments, for GPR data collected from a survey on flexible pavements along a rural road in Madrid, Spain, using a 2000 MHz GPR antenna. Figure 2-5a, shows the radargram after the application of time-zero correction which was discussed above. Figure 2-5b shows the GPR radargram after both time-zero correction and signal gain adjustment, which demonstrates an enhanced feature visibility. Signal gain adjustment involves the progressive amplification of the signal with increasing depth to counter natural signal attenuation (Ciampoli et al., 2019). The result is an enhanced visual contrast between subsurface features, especially those located deeper, leading to a more discernible presentation of the subsurface.




*(a)* *(b)*

*Figure 2-5. a) GPR radargram post time-zero correction, prior to signal gain adjustment. b) GPR radargram after both time-zero correction and signal gain adjustment, demonstrating enhanced feature visibility (After Ciampoli et al., 2019)*

### Background filter (horizontal banding removal)

GPR radargrams are often subject to constant noise reflections, which appear as horizontal banding (Carrick Utsi, 2017). This can be caused by reflectors at a fixed distance from the antenna, such as a person pushing the system or a structure parallel to the survey traverse (e.g., railway tracks) (Downs and Jazayeri, 2021). Another common source of horizontal banding (HB) is the GPR system itself, which can produce constant noise reflections. HBs on a radargram are proportional to the fixed distance between the antenna and the reflector, resulting in a fixed travel time of the propagated wave. In other words, HBs from above-ground reflectors can intersect and obscure scattered signals from subsurface anomalies if they have the same (or similar) travel time (Downs and Jazayeri, 2021). Background noise in GPR data is a significant source of error and contamination (Daniels, 2004). A common method for eliminating HB artifacts from GPR radargrams is to apply background filters, such as a moving average, to declutter the data by removing the background noise as a pre-processing step (Huber and Hans, 2018). However, this is a delicate process because incorrect background removal can result in losing useful information for detecting buried anomalies (Goodman and Piro, 2013a). Furthermore, it may not be effective for removing all types of HB artifacts, and

it can be time-consuming to apply, especially for large datasets. Additionally, it may introduce errors or distortions into the data if not used carefully. For example, removing a HB associated with a buried pipe in a radargram with a direction similar to the GPR moving direction (Bigman, 2012). Despite these limitations, HB removal can be a useful technique for improving the quality and interpretability of GPR data in certain circumstances (Goodman and Piro, 2013b).

Figure 2-6, adapted from Olhoeft (2000), presents the noticeable effects of a horizontal band removal filter on a GPR radargram. Olhoeft (2000) carried out a research project using a 900-MHz centre frequency GPR antenna, aiming to locate and characterise a buried steel pipe. The antenna collected GPR data over a steel pipe with a diameter of 90 cm, buried at a depth of 37 cm. The relative dielectric permittivity of the surrounding soil was identified as 4.0. As shown in Figure 2-6a, the unprocessed radargram exhibits pronounced horizontal bands or 'stripes', a form of noise common in GPR data. These bands can hamper the clear discernment of subsurface features. The efficacy of the horizontal band removal filter is evidenced in Figure 2-6b. This image showcases the same radargram after the application of the filter (Olhoeft, 2000). The successful reduction of horizontal bands enhances the radargram's clarity and interpretability by mitigating this specific type of noise.

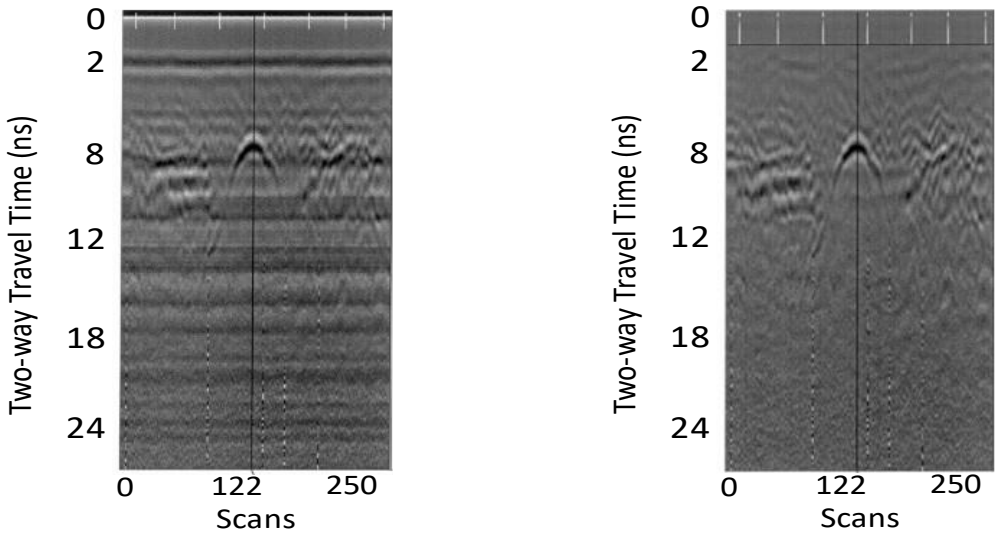



*Figure 2-6. a) GPR radargram prior to the application of a horizontal band removal filter. b) GPR radargram after application of a horizontal band removal filter, showcasing reduced horizontal banding noise (After Olhoeft, 2000)*

### Band-pass filtering

Band-pass filtering is another simple noise reduction method that can be used (if needed) to eliminate undesired signals outside the frequency range of a GPR system (Utsi, 2017b).These undesired frequencies can be generated by an external source and recorded during a survey. Additionally, signal gain adjustments can also cause amplified noise in a radargram (Utsi, 2017b). The frequency range of different GPR systems can vary by manufacturer, but if this range is not provided, a suitable frequency range for the band-pass filter can be determined by considering the central frequency of ±50%, for example, a frequency range of 200 MHz to 600 MHz for data from an antenna with a central frequency of 400 MHz.

Band-pass filtering can help to improve the quality and interpretability of GPR data in certain circumstances. In other words, this technique can be useful for removing noise or other distractions from the data and for highlighting specific features or targets (Liu et al., 2017). However, there are limitations associated with this technique that need to be considered. It

may remove useful information from the data if the filter settings are not carefully chosen (e.g. selecting an inappropriate frequency band), and it can be sensitive to variations in the data, requiring frequent adjustments to the filter settings (Bostanudin, 2013). Additionally, it may introduce distortions or artifacts (e.g. signals from outside sources such as radio masts) into the data if not used carefully (Ciampoli et al., 2019).

Figure 2-7, derived from the research of Xie et al. (2013b), illustrates the impact of applying a band-pass filter in the processing of GPR data. A 500 MHz GPR antenna was utilised to assess the status of grout in shielded tunnels. The researchers used band-pass filtering to amplify the signal-to-noise ratio in the GPR data. It was determined that the relative dielectric constant of the soil is 36, leading to an approximate electromagnetic wave velocity of 0.05 m/ns through the soil. Figure 2-7a displays the original radargram prior to frequency filtering. In this state, the data contains a broad range of frequencies, some of which could interfere with the primary GPR signal. Figure 2-7b shows the same radargram after the application of the band-pass filter. This filter confines the signal to a specific frequency range, thereby gently attenuating undesired high-frequency noise and low-frequency drift. Xie et al. (2013b) demonstrated that the resulting image, while not dramatically different, exhibits a refined signal where the primary reflections from subsurface structures are slightly more discernible.

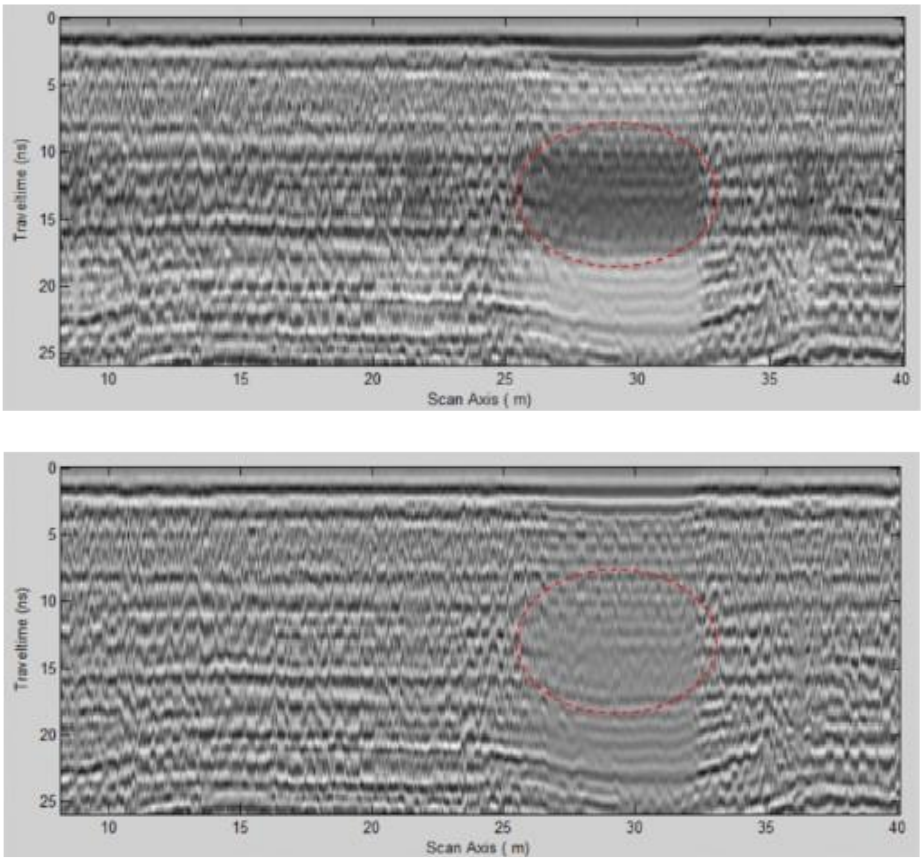



*Figure 2-7. a) GPR radargram prior to the application of a band-pass filter. b) GPR radargram post application of a band-pass filter, demonstrating filtered frequency content (After Xie et al., 2013b)*

## Fourier Transform and Fast Fourier Transform (FFT)

Fourier transform (FT) is a widely used tool for signal processing (frequency domain analysis) that allows for the determination of the frequency components of a signal, such as noise. It decomposes a signal into a sum of trigonometric sine and cosine functions to identify the frequencies and their amplitudes present in the signal (Fourier, 1829; James, 1969; Stein and Shakarchi, 2011). The fast Fourier transform (FFT) is an algorithm that allows for the fast computation of the FT by taking advantage of the algebraic properties of the FT (Bailey and Swarztrauber, 1994; Oshana, 2006; Rader and Mailing, 1967).

The FT and FFT can be applied to raw GPR data or to data that has already undergone other processing steps and be used for identifying specific features or targets within GPR data, as

well as for analysing the frequency content of the data (Zhang et al., 2017). However, it is important to keep in mind that this technique can be computationally intensive, especially when working with large datasets (Xu et al., 2019). Additionally, it may not be effective for all types of GPR data or for all types of features or targets. The results may also be sensitive to certain parameters, such as the window size and overlap (Goodman and Piro, 2013b). Despite these limitations, FT or FFT can be a useful technique for analysing GPR data and identifying specific features or targets within the data in certain circumstances (Goodman and Piro, 2013b; Xu et al., 2019).

### Wavelet Transform (Wavelet Decomposition)

Wavelet transform (also known as wavelet decomposition) is another method for processing and characterising signals, such as GPR signals. It was introduced in 1984 and has seen rapid growth in research across various disciplines in literature. Wavelet transform was developed because the Fourier transform, with its fixed temporal window, is not sufficient to fully analyse signal frequencies (Addison, 2017; Collin and Warnant, 1995; Daubechies, 1990). Wavelet transform is a reversible transformation of a signal that decomposes it into a family of functions (wavelets) - multiple lower resolution scales (approximations and details) of a localised basic wavelike function called the mother wavelet as presented in Figure 2-8 below (Addison, 2017; Foufoula-Georgiou and Kumar, 1994).

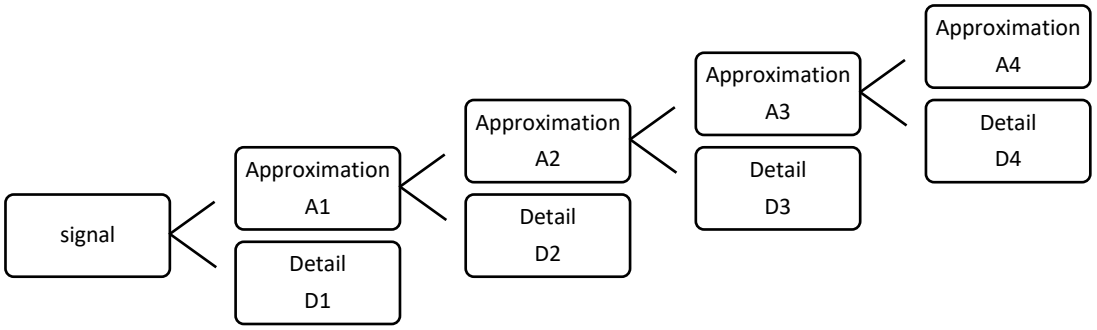



*Figure 2-8. The principle of the wavelet transform as a multiresolution analysis of a signal. The chart illustrates a level 4 signal decomposition into approximations and detail components*

In addition to revealing the frequency components of signals, wavelet transform is localised in both the time and frequency domains (Foufoula-Georgiou and Kumar, 1994). This means that it can identify certain frequency components or patterns hidden in both domains by scaling and shifting the wavelets in a way that allows each frequency element to be analysed and characterised with a resolution corresponding to its level (Bishop et al., 2020; Esfetanaj and Nojavan, 2018). Wavelet transform has been applied in GPR for noise reduction (Baili et al., 2009; Zhu et al., 2017), buried target detection (Kobayashi et al., 2012; Zheng et al., 2014), and identification of changes in energy concentration (Zoubir et al., 2002a). The wavelet transform of a continuous signal, with a mother wavelet $\psi(t)$ chosen based on the problem, can be represented by the equations (Eq. 2-1 and Eq. 2-2) presented by Addison (2017) and Ni et al. (2010). For further details and the theory of wavelet transform, the reader can refer to Addison (2017) and Foufoula-Georgiou and Kumar (1994).

$$T_{a,b} = \frac{1}{\sqrt{a}} \int_{-\infty}^{+\infty} x(t) \psi \left( \frac{t-b}{a} \right) dt \qquad (2\text{-}1)$$

$$x(t) = \frac{1}{C_g} \int_{-\infty}^{+\infty} \int_{0}^{+\infty} T_{a,b} \psi_{a,b}(t) \frac{d_a d_b}{a^2} = \sum_{a=-\infty}^{+\infty} \sum_{b=-\infty}^{+\infty} T_{a,b}\, \psi_{a,b} \qquad (2\text{-}2)$$

Where *a* and *b* are dilation (or scaling) and translation (or shifting) parameters respectively, $x(t)$ the input signal, and $\psi(t)$ the chosen mother wavelet.

It is worth noting that discrete wavelet transform (DWT) is defined as a discretised version of CWT which only differs by the values of dilation and translation parameters $a$ and $b$ (*E*q. 2-1), respectively when implementing the DWT (Baili et al., 2009; Chui and Heil, 1992; Ni et al., 2010) where $l$ is the chosen signal decomposition level and $k \in \mathbb{Z}$.

When it comes to comparing DWT and CWT, both techniques have their own advantages and limitations. DWT is generally more computationally efficient and the results are easier to interpret and visualise compared to CWT (Mallat, 2009a). Additionally, the basic functions used in DWT (also known as wavelet functions or wavelets) are often well-localised in both time and frequency, which can be useful for identifying specific features or targets within the data (Van Drongelen, 2018). However, the accuracy of the results may be lower than that of CWT, especially for signals with rapidly varying frequency content or for signals that are not well-approximated by the chosen wavelet functions (Percival and Walden, 2000a). On the other hand, CWT can provide a more accurate representation of the time-frequency content of the signal, especially for signals with rapidly varying frequency content or for signals that are not well-approximated by the chosen wavelet functions, and the resolution of the time-frequency decomposition can be varied continuously, allowing for more flexibility in the analysis (Addison, 2018). However, the results may be more difficult to interpret and the computational demands may be higher (Abujarad, 2007). Ultimately, the appropriate technique to use will depend on the characteristics of the signal and the goals of the analysis, and it is important to carefully consider the trade-offs between both techniques (Cohen, 2014).

Once the GPR data has been pre-processed, it is ready for interpretation. One common technique for interpreting GPR data is through the detection of hyperbolas. Hyperbolas are a characteristic feature of GPR data and can indicate the presence of buried objects or variations

in subsurface materials. By identifying and analysing these hyperbolas, it is possible to gain insights into the subsurface structure and composition. In the following section, the various methods and techniques used for detecting and interpreting hyperbolas in GPR data will be discussed.

### 2.3.4 GPR data interpretation

As discussed by Dojack (2012), interpreting GPR data can be challenging due to the complexity of the subsurface and various factors that can affect the quality of the data, such as noise, complex subsurface structures, inaccurate or incomplete data, limited depth penetration, and inhomogeneity of the subsurface. Noise reduction techniques, such as signal processing and image processing, may be required to improve the clarity of the data. However, it is important to be careful when applying these techniques, as they can also remove useful information. Additionally, understanding the limitations of GPR and the factors that can affect the data can help improve the accuracy and reliability of the interpretation.

The interpretation of GPR data is a subjective task that requires expertise and experience. Insufficient effort in data processing can lead to improper interpretations and inaccurate results. It is important for interpreters to have a thorough understanding of the data, the survey goals, and the limitations of the technology in order to produce reliable interpretations. (Goodman and Piro, 2013b). By exploring the various techniques and methods available for processing and interpreting GPR data, a better understanding of subsurface features and anomalies in the study area can be gained.

#### 2.3.4.1 Techniques in use for hyperbola detection in GPR radargrams

There are several techniques that are commonly used for processing GPR data and detecting hyperbolas in radargrams. These techniques can be broadly grouped into two categories, namely manual methods, and automated methods. Manual methods for hyperbola detection in GPR data include visual inspection, the use of cursor or measurement tools, drawing lines or curves on the data, and manual curve fitting. Manual methods can be time-consuming and may be prone to errors, but they can be useful for data that is difficult to analyse automatically or for verifying the results of automated methods. Automated methods, however, typically involve the use of techniques and algorithms that are designed to identify hyperbolas in the data automatically e.g. Hough transform (Wang and Su, 2013) and machine learning algorithms such as neural networks (Amaral et al., 2022). Automated methods have the potential to be efficient and accurate, but they may require significant processing power and may not be suitable for all types of GPR data.

Automated methods for detecting hyperbolas in GPR data are an active research area due to their potential to significantly improve the efficiency and accuracy of GPR data interpretation. There are several different approaches that have been developed for detecting hyperbolas in GPR data, including machine learning techniques, image processing techniques, and mathematical techniques. Some of the key challenges in this area include dealing with noise and interference in the data, accurately detecting and characterising hyperbolas in the presence of other features, and developing robust and efficient algorithms that can handle large amounts of data.

To further explore the various approaches for automated GPR data processing and hyperbola detection, the next section of the thesis will provide a critical review on these methods. The review will cover the different types of approaches that have been developed by researchers.

The Hough transform is often used to detect hyperbolas in Ground Penetrating Radar (GPR) images. Capineri et al. (1998) presented a modified version of the Hough transform that can detect both hyperbolas and straight lines in autonomous GPR B-scans for buried object detection. This method relies on the relationship between the electromagnetic wave round trip time and the position of a potential buried target, which requires the determination of position and time data. Borgioli et al. (2008) used a combination of the Hough transform and the time of flight of electromagnetic waves (TOF) with the least square algorithm to determine the spatial position of a buried target. However, Zhou et al. (2018) found that the least square method can only detect one hyperbola in an image and is unable to detect multiple hyperbolas in a B-scan. Shihab and Al-Nuaimy (2005) also criticised the least square method for its inability to sufficiently characterise the geometric parameters of hyperbolas, which results in insufficient information for buried object detection. Zhou et al. (2018) also argued that the Hough transform is only suitable as a parallel approach or for the detection of light-deflected shapes and may ignore other useful parameters when used as a standalone approach. Additionally, the Hough transform cannot determine the material or size of a buried object because it ignores these necessary parameters and only considers the apex of the hyperbola in practical applications. Both Shihab and Al-Nuaimy (2005) and Zhou et al. (2018) also noted the high computational time and cost associated with using the Hough transform method for hyperbola fitting. Dou et al. (2017) reported a 15-minute processing time for the generalised Hough transform method to process one 300x400 pixel GPR image and detect multiple hyperbolas using an Intel 3.6GHz processor. The computational time for the generalised Hough transform can vary and is directly related to the discretisation of the hyperbola parameters in the hyperbola equation.

The evolution of computational Artificial Intelligence (AI) techniques such as Artificial Neural Network (ANN) and Genetic Algorithm (GA) was a contributing factor for the enhancement of more intelligent evaluation techniques. ANN has been successfully applied as a complementary tool to different evaluation techniques such as GPR (Al-Nuaimy, 1999; Gamba and Lossani, 2000; Shihab et al., 2002; Szymczyk et al., 2015; Travassos et al., 2018). To sum up, the techniques proposed by researchers for locating and condition assessment of buried utilities can work efficiently for specific ground circumstances. However, the variation of ground properties in different situations (i.e. boundary conditions) can lead to inefficiency of these techniques due to the reasons explained above. Similar to above mentioned publications, Al-Nuaimy et al. (2000) and Shihab and Al-Nuaimy (2005) used a neural network to limit the image area with existing hyperbolas and used the Hough transformation to detect hyperbolic shapes (caused by pipe and land mines). Later, to detect buried pipes, Gamba and Belotti (2003) presented a neural network with the ability to detect hyperbolas. To finalise the result a large image database from different sized hyperbolas was created. The issue of this method was the need for an operator for manual evaluation of detected hyperbola (by ANN) using an appropriate image from the database. These methods above are mainly based on determining the relationship between trigonometrical parameters of the cylinder and the GPR hyperbola. Birkenfeld (2010) introduced an approach for training the ANN for automated detection of healthy and interfered hyperbolas from GPR images. The author suggested capturing the specific size of trapezoidal shaped images that included only the hyperbola shapes and avoiding the unnecessary background or surrounding information. Interfered training hyperbolas were synthetically produced by cutting a complete one and fed into the network. The small size of the images (639 pixels) led to a reduction of the size of the input matrix and accordingly, the size of the neural network and finally increasing the performance.

However, various antennas and frequencies need to be investigated as this experiment was carried out only using a 600 MHz antenna.

Szymczyk et al. (Szymczyk et al., 2015) introduced a neural network method for automatic classification of GPR traces using polynomial approximation coefficients as input into a neural network to classify sinkholes from other geological structures. The method showed high accuracy in distinguishing between sinkholes and other geological formations, but its accuracy is dependent on the quality of the input data. In another study, Liu et al. (2018) presented an inversion approach using a neural network to locate subsurface targets and evaluate their backscattering properties using GPR data. The method showed improved accuracy compared to traditional inversion methods but is computationally expensive and requires a large amount of training data.

In recent years, several methods have been proposed for automating the identification of buried cylindrical objects and estimating rebar properties using GPR. Giannakis et al. (2019) presented a machine learning-based numerical framework for estimating the depth and radius of metallic rebars inside concrete slabs using GPR, which reduces computational requirements compared to traditional algorithms and was found to have near real-time performance and accurate results. However, the method requires substantial computing power during the training set generation and tuning stages due to the large amount of input data required. Giannakis et al. (2021) also presented a machine learning framework for estimating rebar diameters in concrete structures using GPR, which combines neural networks and random forest regression and is trained with synthetic data. The method requires only one A-scan as input and was successfully evaluated with real data but may not yield accurate results when extrapolating bounds outside of those tested in the study.

Dewantara and Parnadi (2022) presented a method involving two modules: one using Faster-region based convolutional neural networks (RCNN) to extract boundary boxes around hyperbola segments and another estimating the coordinates of apex points through image processing. The method showed promising results in correctly detecting all hyperbolas present in the test set but has limitations such as a reliance on the training dataset used to train Faster-RCNN and difficulty in detecting interference hyperbolas due to noise in the search window. Amaral et al. (2022) conducted a systematic review of the literature on computer vision and machine learning algorithms for extracting information from GPR images. The review found that several techniques, including Hough transform, template matching, edge detection, thresholding, and machine learning methods such as YOLO V3, Viola-Jones, and AlexNet, have been successfully applied for automated detection and interpretation systems for GPR data. These techniques can reduce pre-processing effort while increasing accuracy, and the use of computer vision algorithms is increasing due to their ability to rapidly quantify buried targets without requiring high levels of expertise and training. However, the review noted that improvements are needed in terms of accuracy, speed, and efficiency when it comes to detecting objects using GPR, and that most studies only focus on two classes: object reflection or non-object reflections, with further work needed to explore other types of classifications such as landmines and pipes. Overall, these methods show potential in automating the detection of buried cylindrical objects and estimating rebar properties using GPR, but also have limitations that should be considered.

A list of reviewed works in the literature focusing on above methods and autonomy in GPR data interpretation is presented in Table 2-3 in a non-exhausting manner.

*Table 2-3. Overview of publications with focus on automated hyperbola recognition in GPR radargrams*

| | Author | Method |
|---|---|---|
| **1** | (Hart and Duda, 1972) | Hough Transform |
| **2** | (Capineri et al., 1998) | Hough Transform |
| **3** | (Borgioli et al., 2008) | Hough Transform, Least Square |
| **4** | (Duda and Hart, 2002) | Hough Transform |
| **5** | (Wang and Su, 2013) | Cross-Correlation and Hough Transform |
| **6** | (Ahn et al., 2001) | Least-squares orthogonal distances fitting |
| **7** | (Pilu et al., 2002) | Direct least-square fitting |
| **8** | (Ertöz et al., 2013) | Clustering method |
| **9** | (Kanungo et al., 2002) | k-means clustering |
| **10** | (Cao et al., 2013) | k-means clustering |
| **11** | (Gamba and Lossani, 2000) | Neural Network (ANN) |
| **12** | (Al-Nuaimy et al., 2000) | Neural Network (ANN) |
| **13** | (Belotti et al., 2003) | Neural Network and wavelet image processing |
| **14** | (Birkenfeld, 2010) | Neural Network |
| **15** | (Youn and Chen, 2002a) | Neural Network |
| **16** | (Lei et al., 2019) | Faster R-CNN |
| **17** | (Zhang et al., 2021) | Faster R-CNN |
| **18** | (Amaral et al., 2022) | ANN and CNN |
| **19** | (Pasolli et al., 2009) | Genetic optimizer |
| **20** | (Harkat et al., 2018) | Multi-Objective Genetic Algorithm |
| **21** | (Harkat et al., 2019) | Neural Network and Multi-Objective Genetic Algorithm |
| **22** | (Chen, 2010) | Probabilistic conic mixture model |
| **23** | (Rosin, 2001) | Unimodal thresholding |
| **24** | (Persico et al., 2010) | Microwave tomography |
| **25** | (Idi and Kamarudin, 2012) | Digital wavelets transform |
| **26** | (Maas and Schmalzl, 2013) | Viola-Jones algorithm |
| **27** | (Terrasse et al., 2015) | Curvelet transform |
| **28** | (Sagnard and Tarel, 2016) | Template-matching |

| | | |
|---|---|---|
| **29** | (Harkat et al., 2016) | Scale-Invariant Feature Transform |
| **30** | (Mertens et al., 2016) | Fitting analytical function |
| **31** | (Rajiv et al., 2017) | Hyperbola fitting method |
| **32** | (Bu et al., 2017) | Cluster-aware multiagent system |
| **33** | (García-Escudero et al., 2017) | Parabola fitting and Retinography |
| **34** | (Tzanis, 2017) | Curvelet-like directional filter |
| **35** | (Dou et al., 2017) | Column-connection clustering (C3) algorithm |
| **36** | (Zhou et al., 2018) | Parabolic fitting-based judgment (PFJ) |
| **37** | (Luo and Lai, 2020) | Pyramid method and Pattern recognition |
| **38** | (Ali et al., 2021) | Hybrid feature extraction |
| **30** | (Zhou et al., 2018) | Open-scan clustering algorithm (OSCA) |
| **40** | (Li et al., 2020) | TensorFlow |

These techniques are developed for specific applications of GPR. Thus, each suffer from limitations for cases other than their developed purpose, e.g. edge detection in Hough Transform, overfitting in ANN, high susceptibility to noise in Clustering method (Ali et al., 2021; Amaral et al., 2022; Ertöz et al., 2013). Several authors have used the Hough Transform to detect hyperbolas in GPR radargrams (Borgioli et al., 2008; Capineri et al., 1998; Duda and Hart, 2002; Hart and Duda, 1972). However, the Hough Transform has been found to have limitations including sensitivity to noise, computational intensity, and being limited to detecting specific types of features such as lines, circles, or other simple geometric shapes (Wang and Su, 2013). It is also robust to small deformations in the data, but may not be as robust to larger or more severe deformations (Borgioli et al., 2008). Additionally, the Hough Transform is dependent on the choice of threshold (Duda and Hart, 2002). Other authors have used least-squares methods, including least-squares orthogonal distances fitting (Ahn et al., 2001) and direct least-square fitting (Pilu et al., 2002), to detect hyperbolas in GPR data. These methods have been found to have limitations including potentially not providing the best fit

for the data and dependence on initial conditions (Ahn et al., 2001). Clustering methods have also been used to detect hyperbolas in GPR data, including the k-means clustering algorithm (Cao et al., 2013; Ertöz et al., 2013; Kanungo et al., 2002) and cluster-aware multiagent system (Bu et al., 2017). These methods have been found to have limitations including sensitivity to noise, sensitivity to initial conditions which may not always find the optimal solution, dependence on the choice of clustering algorithm, difficulty with overlapping clusters, and dependence on the similarity measure (Ertöz et al., 2013).

Neural networks, including both artificial neural networks (ANNs) and convolutional neural networks (CNNs), have been used by several authors to detect hyperbolas in GPR data (Al-Nuaimy et al., 2000; Amaral et al., 2022; Belotti et al., 2003; Birkenfeld, 2010; Gamba and Lossani, 2000; Youn and Chen, 2002a). These methods have been found to have limitations including sensitivity to noise, the need for a large dataset for training, difficulty with complex or nonlinear relationships, lack of interpretability, and computational intensity (Al-Nuaimy et al., 2000; Gamba and Lossani, 2000). Some authors have used template matching to detect hyperbolas in GPR data (Sagnard and Tarel, 2016), but this method has been found to have limitations including sensitivity to noise, issues with images having low similarities or significant variations, difficulty with variations in scale, orientation, or appearance, and being limited to matching known templates (Sagnard and Tarel, 2016). The Faster R-CNN algorithm has also been used to detect hyperbolas in GPR data (Lei et al., 2019; Zhang et al., 2021), but has been found to have limitations including the need for a large dataset for training and potentially not working well with data that has significant variations or deformations (Lei et al., 2019; Zhang et al., 2021).

Other techniques used to detect hyperbolas in GPR data include the probabilistic conic mixture model (Chen, 2010), which has been found to potentially not work well with data that

is highly nonlinear or has non-Gaussian noise (Chen, 2010), and unimodal thresholding (Rosin, 2001), which has been found to be sensitive to noise and potentially requiring pre-processing to smooth the image (Chen, 2010). Microwave tomography (Persico et al., 2010), digital wavelets transform (Idi and Kamarudin, 2012), the Viola-Jones algorithm (Maas and Schmalzl, 2013), the curvelet transform (Terrasse et al., 2015), the Scale-Invariant Feature Transform (Harkat et al., 2016), fitting analytical functions (Mertens et al., 2016), the hyperbola fitting method (Rajiv et al., 2017), the parabolic fitting-based judgment (PFJ) technique (Zhou et al., 2018), the pyramid method (Luo and Lai, 2020), and TensorFlow (Li et al., 2020) have also been used to detect hyperbolas in GPR data, each with their own limitations as described below.

Table 2-4 presents a summary of the abovementioned methods including a short description, advantages, and limitations.

*Table 2-4 Comparison of methods for hyperbola detection in GPR radargrams*

| Method | Description | Advantages | Limitations |
|---|---|---|---|
| Hough Transform(Dou et al., 2017) | A pattern recognition technique used to detect lines or curves, such as hyperbolas, in GPR data. | Can detect partially obscured or distorted features; robust to small deformations. | ▪ Sensitivity to noise<br>▪ Computational intensity<br>▪ Limited to detecting specific types of features (lines, circles, or other simple geometric shapes)<br>▪ Robust to small deformations in the data, but may not be as robust to larger or more severe deformations<br>▪ Dependence on the choice of threshold |

| | | | |
|---|---|---|---|
| Clustering<br>(Ertöz et al., 2013) | Grouping similar data points to identify patterns like hyperbolas in GPR data. | Can identify patterns and trends such as hyperbolas in large datasets (radargrams) | ▪ Sensitive to noise<br>▪ Sensitivity to initial conditions<br>▪ may not always find the optimal solution<br>▪ Dependence on the choice of clustering algorithm<br>▪ Difficulty with overlapping clusters<br>▪ Dependence on similarity measure |
| Template matching<br>(Sagnard and Tarel, 2016) | Comparing a template image to GPR data to identify known hyperbola shapes and characteristics. | It can be used to search for patterns in GPR data that match a known template for a hyperbola. This approach can be useful for identifying hyperbolas with known shapes and characteristics | ▪ Sensitivity to noise<br>▪ Issue with images having low similarities or significant variations<br>▪ Difficulty with variations in scale, orientation, or appearance<br>▪ Limited to matching known templates |
| Machine learning<br>(Harkat et al., 2019) | Training a machine learning algorithm to identify indicative patterns and hyperbolas in GPR data. | can be used to identify hyperbolas in GPR data by analysing the characteristics of the data | ▪ Sensitivity to noise<br>▪ Require a large dataset for training<br>▪ Difficulty with complex or non-linear relationships<br>▪ Lack of interpretability<br>▪ Computationally intensive |
| Neural networks<br>(Harkat et al., 2019) | Using artificial neural networks to learn and adapt to complex patterns in GPR data, like hyperbolas. | can be trained to recognise hyperbolas in GPR data. This can be done by using a large dataset of GPR data with labelled hyperbolas and training the neural network to recognise these patterns. | ▪ Sensitivity to noise<br>▪ Require a large dataset for training<br>▪ Difficulty with complex or non-linear relationships<br>▪ Lack of interpretability<br>▪ Computationally intensive |

| | | | |
|---|---|---|---|
| Curve fitting<br>(Terrasse et al., 2015) | Fitting a curve, like a hyperbola, to the data using mathematical functions or algorithms to accurately model complex trends in the data. | Useful for identifying hyperbolas in data that are not perfectly symmetrical or that have noise or other distortions (e.g., GPR radargrams). | ▪ Sensitivity to noise<br>▪ May not always provide the best fit for the data<br>▪ May require manual selection of the curve model<br>▪ Limited to fitting specific functional forms<br>▪ Dependence on initial conditions |
| Cross-Correlation<br>(Wang and Su, 2013) | A statistical measure to identify relationships between signals and detect patterns like hyperbolas in GPR data. | Can detect patterns in GPR data and identify hyperbolas | ▪ Sensitive to noise<br>▪ May not work well with data with low correlations<br>▪ Limited to identifying linear relationships<br>▪ Dependence on the choice of lag |
| k-means clustering<br>(Cao et al., 2013) | A clustering technique for grouping similar data points to identify patterns and trends in GPR data and detect hyperbolas. | Effective at identifying patterns in large datasets. | ▪ Sensitivity to noise<br>▪ Sensitivity to initial conditions<br>▪ May not always find the optimal solution<br>▪ Difficulty with overlapping clusters<br>▪ Assumes spherical cluster shape |
| Wavelet image processing<br>(Idi and Kamarudin, 2012) | Decomposing an image into different frequency bands using wavelets to enhance GPR data resolution and contrast for hyperbola detection. | Can enhance the resolution and contrast of GPR data and identify hyperbolas | ▪ Sensitivity to noise<br>▪ May require pre-processing to smooth the image<br>▪ Computational intensity |
| Probabilistic conic mixture model<br>(Chen, 2010) | A statistical model for fitting a mixture of conic distributions to data. It can be used to fit hyperbolas in GPR data. | Can fit hyperbola signatures in GPR data | ▪ Sensitivity to noise<br>▪ Sensitivity to initial conditions<br>▪ Limited to fitting a specific functional form<br>▪ May not work well with data that is highly nonlinear or has non-Gaussian noise<br>▪ Dependence on the choice of hyperparameters |

| | | | |
|---|---|---|---|
| Unimodal thresholding<br>(Rosin, 2001) | A technique for separating an image into foreground and background pixels based on intensity of the pixels. | Can identify hyperbolas in GPR data by thresholding the data to separate the hyperbolas from the background | ▪ Sensitive to noise<br>▪ May require pre-processing to smooth the image<br>▪ Limited to unimodal histograms<br>▪ Dependence on the choice of threshold |
| Microwave tomography<br>(Persico et al., 2010) | A technique for imaging the subsurface using microwave radiation | Can create high-resolution images of subsurface features, including hyperbolas | ▪ Sensitivity to noise<br>▪ Limited in-depth penetration<br>▪ May be affected by subsurface inhomogeneities<br>▪ Limited resolution<br>▪ Sensitivity to objects with high conductivity<br>▪ Dependence on the choice of excitation frequency |
| Viola-Jones algorithm<br>(Maas and Schmalzl, 2013) | A machine learning technique for detecting objects in images | Can detect hyperbolas in GPR data by training the algorithm on a dataset of labelled hyperbolas | ▪ Sensitivity to noise<br>▪ Dependence on the training data<br>▪ Issues with data with significant variations or deformations<br>▪ Sensitivity to variations in appearance<br>▪ Limited to specific types of objects |
| Scale-Invariant Feature Transform (SIFT)<br>(Harkat et al., 2016) | A computer vision technique to extract and match distinctive local features in images | Can detect hyperbolas in GPR data by identifying distinctive features in the data | ▪ Primarily designed for use with visible-light images, and it may not be directly applicable to GPR data<br>▪ Sensitivity to noise and changes in appearance<br>▪ May require pre-processing to smooth the image<br>▪ Limited to identifying specific types of features<br>▪ Computational intensity |
| Hyperbola fitting method | Fitting a hyperbolic curve to a set of data points to identify and | Effective for detecting hyperbolas | ▪ Sensitive to noise |

| | | | |
|---|---|---|---|
| (Rajiv et al., 2017) | analyse hyperbolas in GPR data | | ▪ May not work well with data that has low connectivity<br>▪ Limited to fitting specific functional forms<br>▪ Dependence on initial conditions<br>▪ May not always provide the best fit for the data<br>▪ May require manual selection of the curve model |
| Column-connection clustering (C3) algorithm<br>(Dou et al., 2017) | Clustering data points based on their connectivity to detect hyperbolas in GPR data | Can detect hyperbolas in GPR data by grouping connected data points into clusters | ▪ Sensitivity to noise<br>▪ Dependence on the choice of similarity measure<br>▪ Limited to small or medium-sized data sets<br>▪ Limited to specific types of data that can be represented as a graph (e.g., points and edges)<br>▪ Dependence on initial conditions<br>▪ Difficulty with overlapping clusters |
| Parabolic fitting-based judgment (PFJ)<br>(Zhou et al., 2018) | Fitting a parabolic curve to data points to identify hyperbolas in GPR data | Can identify hyperbolas in GPR data | ▪ Sensitivity to noise<br>▪ Dependence on initial conditions<br>▪ Limited to small or medium-sized data sets<br>▪ May not always provide the best fit for the data<br>▪ May require manual selection of the curve model (fitting algorithm)<br>▪ Limited to identifying linear trends |
| Pyramid method | Decomposing an image into a pyramid of progressively | Enhances resolution and | ▪ Sensitivity to noise |

| | | | |
|---|---|---|---|
| (Luo and Lai, 2020) | lower-resolution images to enhance resolution and contrast of GPR data for hyperbola detection | contrast of GPR data | ▪ May require pre-processing to smooth the image<br>▪ Limited to specific types of images<br>▪ Issues with data with low contrast<br>▪ Dependence on the choice of decomposition level<br>▪ Computational intensity |
| TensorFlow<br>(Li et al., 2020) | A machine learning framework for building and training neural networks to detect hyperbolas in GPR data | Adaptable to various data types and machine learning tasks | ▪ Sensitivity to noise<br>▪ May require a large dataset<br>▪ Complexity<br>▪ Limited to specific data types (numerical data, such as images or text)<br>▪ Dependence on the choice of model and hyperparameters<br>▪ Dependence on the quality and quantity of training data<br>▪ Limited to specific types of machine learning tasks (supervised learning tasks, such as classification and regression) |
| Fitting analytical function<br>(Mertens et al., 2016) | Finding the best curve that describes a set of data points to fit hyperbolic curves to GPR data | Effective at fitting specific functional forms like hyperbola | ▪ Sensitivity to noise<br>▪ Limited to fitting specific functional forms (e.g., such as polynomial, exponential, or sinusoidal curves)<br>▪ May not always provide the best fit for the data<br>▪ Limited to identifying linear or smooth trends<br>▪ May require manual selection of the curve model<br>▪ Dependence on initial conditions<br>▪ Limited to small or medium-sized data sets |

| | | | |
|---|---|---|---|
| Faster R-CNN<br>(Zhang et al., 2021) | A machine learning technique for detecting objects in images, used to detect hyperbolas in GPR data by training the algorithm on a dataset of labelled hyperbolas | Effective for object detection tasks | ▪ Sensitivity to noise<br>▪ Require a large training dataset<br>▪ Dependence on the quality and quantity of training data<br>▪ Computational intensity<br>▪ May not work well with data that has significant variations or deformations<br>▪ Limited to specific types of object detection tasks, such as identifying and localising objects in images or videos |

As shown above, there are many approaches for manual and automated GPR data processing and hyperbola detection because GPR data can be complex and varied, and different approaches may be better suited for different types of data and applications. Additionally, GPR data processing and hyperbola detection involve many trade-offs and challenges, such as dealing with noise, errors, and uncertainties in the data, and choosing the appropriate model or algorithm to fit the data. As a result, researchers and practitioners have developed a range of approaches to address these challenges and optimise the performance of the data processing and hyperbola detection tasks. However, none of these approaches can be classified as the best method, as different approaches may have different strengths and limitations, and may perform differently depending on the specific characteristics of the data and the goals of the analysis. For example, some approaches may be more efficient or accurate, while others may be more robust or flexible. Additionally, the performance of a given method may depend on the specific implementation or parameters used and may vary with the size and quality of the data. As a result, it is important to carefully evaluate and compare the performance of different approaches to determine which one is most suitable for a given application. (Solimene et al., 2014; Youn and Chen, 2002b).

The previous section of the literature review highlighted the various non-destructive geophysical methods that are used for subsurface exploration. However, the reliance on a single method or sensor has been shown to be problematic due to the inherent limitations and biases of each technique. These limitations may include sensitivity to specific subsurface materials or structures, depth of penetration, resolution, and accuracy. As a result, the use of a single method or sensor may not always provide a complete or accurate representation of the subsurface. To address these limitations and improve the accuracy and comprehensiveness of subsurface mapping, researchers have turned to multi-sensing data fusion techniques, which involve the combination and integration of data from multiple geophysical methods and sensors. The following section of this literature review will focus on the various approaches and techniques that have been developed for multi-sensing data fusion in subsurface mapping, as well as their advantages, limitations, and potential applications.

## 2.4 Data Fusion and multi remote sensing

### 2.4.1 Introduction

The motivation for using data fusion in non-intrusive survey techniques is to overcome the limitations of individual sensors and improve data quality and interpretation accuracy by combining multiple sources of data to obtain a more complete and accurate understanding of the phenomenon being studied. To achieve this goal, the data collected from multi-sensor surveys must be processed, fused, and interpreted in a meaningful, systematic, and objective manner (Gros, 1997b).

Fusing data from different non-destructive testing (NDT) sensors with similar or dissimilar data formats requires a comprehensive understanding of all relevant parameters. To reduce error and increase interpretation accuracy, it is important to use the most suitable fusion technique for the survey (Gros, 1997a; Liggins et al., 2009). To facilitate communication and data exchange within a chosen fusion algorithm, the participating sensors must be combined using a distribution scheme (Yang et al., 2019). This refers to the order and arrangement of sensors, which can be divided into two main categories: parallel and serial configurations, and a combination of these as a subcategory. These categories are based on the types of sensors, survey configuration, and boundaries (Gros, 1997b; Qiu et al., 2019).

Researchers have also introduced different categorisations to distinguish between different types and levels of NDT data fusion. Faundez-Zanuy (2009) and Gros (1997b) categorise NDT data fusion into three levels: low level, intermediate level, and high level data fusion. These levels will be discussed in the following section.

### 2.4.2 The low-level data fusion (data-oriented or signal level)

Low-level data fusion combines raw data from various sensors to generate a new dataset, enhance the signals, distinguish, and extract features from multisensory raw data. This level of data fusion does not necessarily specify objects from raw data, but rather, the specification, classification, and quantification of objects can be conducted through other processing and interpretation methods such as multiresolution analysis (MRA) (e.g., analysing different GPR images from various frequencies), optimisation methods (e.g., genetic algorithm), and heuristic methods (e.g., artificial neural network) (Carvalho et al., 2003). Heuristic methods are those that are based on or assisted by numerical models and simulations (Şenyiğgit et al., 2013).

### 2.4.3 Intermediate level or feature level data fusion

In intermediate-level data fusion, the aim is to combine different features such as edges, lines, and other existing features to generate a new thematic map for interpretation and detection. Instead of categorising the input data, it is implemented into specific classifiers such as neural networks (NNs), algorithms for data clustering and segmentation as pre-processing units. For instance, hyperbola detection using NN or Hough pattern recognition techniques (Faundez-Zanuy, 2009).

### 2.4.4 High level or decision level data fusion

In high-level data fusion, statistical and stochastic methods are used to model sensors and combine the opinions and decisions from different statistical theories, such as Bayesian theory and Dempster-Shafer paradigm. Calibrated supervised or unsupervised sensor models can be built through the application of statistical methods. The output of this level of data fusion is the probability of possible results through the use of statistical methods (Gros, 1997a). One advantage of this level of fusion is that all information can be applied separately to optimise performance, regardless of the limitations of each sensor (Abeynayake et al., 2003).

Liu et al. (2007) also categorised data fusion problems into two general categories: feature integration and feature classification. In feature integration, a completely new dataset is generated by combining data from different inspections at the pixel or signal level of data fusion. For example, combining two images (at the pixel level) with different features to create a new detailed image that includes all features from the other two images (Figure 2-9).

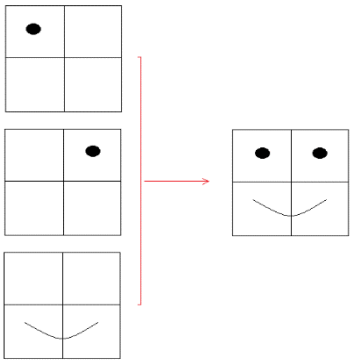


*Figure 2-9. Example of feature integration data fusion*

In *feature classification*, the new dataset is extracted from data integration of different inspections and interpretations at the processing and decision level. As illustrated in Figure 2-10, blue and grey coloured features are extracted individually from other existing features in the same image. After processing and fusing the data from both initial images, a new image including all features and requiring no further processing is produced.

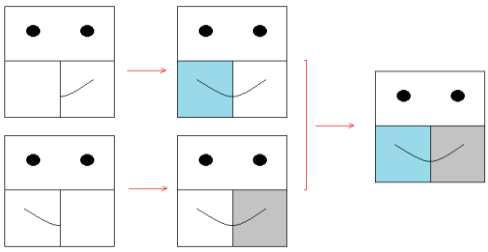


*Figure 2-10. Example of feature classification data fusion*

The acquired data from surveys can be similar in format or entirely different data types. This means that data from different or even similar sensors can have different resolutions, start and end points, and views. In other words, data fusion cannot proceed unless data registration is conducted to link the data from a multi-sensing survey to the physical or geometrical points of the surveyed scenario. Data registration is the process of converting data into a similar format. Physical and geometrical data registrations are conceptual meanings of data

registration and depend on the employed sensing technique and characteristics of its associated data (Liu and Forsyth, 2004). Data registration of different data with different characteristics is not always straightforward. For example, registering data from GPR and falling weight Deflectometry (FWD), which have different characteristics (e.g., type, format). GPR radargrams are characterised by the magnitude of signal amplitudes and two-way travel time of the signals, while FWD data is characterised by the magnitude of surface displacements at different positions.

## 2.4.5 Fusion of nonintrusive NDT geophysical methods

There are several works of literature on the fusion of GPR with other nonintrusive remote sensing methodologies for subsurface exploration. Table 2-5 presents a non-exhaustive list of reviewed works in this area.

*Table 2-5. List of works focused on remote sensing data fusion*

| | Author | Devices | Method | Purpose | Fusion Level |
|---|---|---|---|---|---|
| **1** | (Steinway et al., 1998) | GPR(SFCW) + Infrared + Metal Detection | Logic joint Fusion | Landmine detection | High |
| **2** | (Weisenseel et al., 1999) | GPR + EMIS[1] | Optimal maximum likelihood Method | Detection of plastic A/P mines | High |
| **3** | (Collins et al., 2002) | GPR + EMC[2] | Bayesian detection approach | Landmine detection | High |
| **4** | (Ligthart and Yarovoy, 2003) | GPR + GPR(SFCW[3]) | Inverse scattering method | Landmine detection | Low |
| **5** | (Scott et al., 2004) | GPR + EMC + Seismic | Logic joint Fusion | Landmine detection | High |

[1] Electro-Magnetic Induction Spectroscopy
[2] Electromagnetic Conductivity
[3] Stepped-Frequency continuous Wave

| | | | | | |
|---|---|---|---|---|---|
| **6** | (Moghadas et al., 2010) | GPR + EMC | Comparison of various fusion methods: Integrated full-waveform EMC and GPR inversion, Weighting methods, Bayesian data fusion, Sequential inversion | Estimating soil electrical properties | Low-High |
| **7** | (Lim and Cao, 2013) | GPR + IR[4] + UPV[5] | Logic joint Fusion | Sidewalk and concrete stairwell | High |
| **8** | (Dutta et al., 2013) | GPR + Vibro-acoustic | Dynamic Bayesian network | Buried plastic water pipe | Intermediate-High |
| **9** | (Li et al., 2015) | GPR + GPS[6] + GIS[7] | Hybrid GPR/GPS + 3D probabilistic uncertainty band | Mapping and visualisation of underground utilities | Low-High |
| **10** | (Dou et al., 2016) | GPR + Vibro-acoustic + PMF[8] + MG[9] + LFEM[10] | Marching-Cross-Section (MCS) algorithm | Locating buried utility segments | Low-Intermediate |
| **11** | (Capozzoli and Rizzo, 2017) | GPR + infrared + ERT[11] | Logic joint Fusion | Concrete Structure | High |
| **12** | (Knox et al., 2017) | GPR + EMI[12] | Logistic regression, Gradient boosting decision trees | Buried Explosive Threat Detection | Low-Intermediate |
| **13** | (Smith et al., 2017) | GPR + WEMI[13] | Aggregation of different Choquet integrals within and across sensors | Explosive hazard detection | High |
| **14** | (Hafsi et al., 2017) | GPR + EML (no data) | Merge using a linear combination | Detection and localisation of underground networks | Intermediate - High |
| **15** | (Xu et al., 2018b) | GPR + GPR (Multi-Frequency) | Fourier transform | Underground Disease Detection | Low |

[4] Impulse Response
[5] Ultrasonic Pulse Velocity
[6] Global Positioning System
[7] Geographical Information System
[8] Passive Magnetic Fields
[9] Magnetic Gradiometer
[10] Low Frequency Electromagnetic Fields
[11] Electrical Resistivity Tomography
[12] Electro-Magnetic Induction
[13] Wideband Electro-Magnetic Induction

| | | | | | |
|---|---|---|---|---|---|
| **16** | (Zhou et al., 2019) | GPR + Electric field device | Electric Field Locating + B-scan image interpreting | Plastic pipe detecting | Low - Intermediate |
| **17** | (Li et al., 2019) | GPR + Camera | 3D reconstruction method | Subsurface Pipeline Mapping | Intermediate |
| **18** | (Bianchini Ciampoli et al., 2020) | GPR + InSAR[14] | Data fusion logic | Assessment of linear transport infrastructures | High |
| **19** | (Lu et al., 2020) | GPR + GPR (Multi-Frequency) | Image fusion based on 2D wavelet transform | Multi-frequency and multi-attribute GPR | Intermediate |
| **20** | (Adamopoulos et al., 2021) | GPR + Laser scanning + imaging + thermography | Data fusion logic | Built heritage survey | High |
| **21** | (Massaro et al., 2021) | GPR + infrared | Long term short memory neural network algorithm | Road infrastructure inspection | High |

Different data fusion methods have their own strengths and limitations, and it is important to select a suitable fusion tool for the success of a project. As an example, the following is a brief description of three fusion methodologies:

1) Moghadas et al. (2010) analysed full-waveform information content in GPR and electromagnetic induction (EMI) synthetic data to reconstruct the electrical properties of multi-layered media. They used GPR and EMI systems operating in zero-offset, off-ground mode and designed using vector network analyser technology. They compared four approaches for GPR and EMI data fusion: two weighting methods, the naive Bayesian method, and the sequential inversions of GPR–EMI and EMI–GPR. They showed that fusing the two sources of information can improve the reconstruction of the electrical properties, but the formulation of an optimal objective function based on the weighting methods is not

[14] Interferometric synthetic aperture radar

straightforward. They also showed that the Bayesian method can be a relevant alternative for GPR and EMI data fusion, but it relies on assumptions. They found that sequential inversions can provide proper estimation of values in some cases, but not in others, depending on the conductivity or permittivity of the first layer. Figure 2-11 illustrates an example result presented by Moghadas et al. (2010), showcasing the variation in response surfaces of the objective function for different formulations. This includes the utilisation of separate and combined GPR and EMI data, as well as different parameter planes of their developed algorithm's complete 5-dimensional objective function.

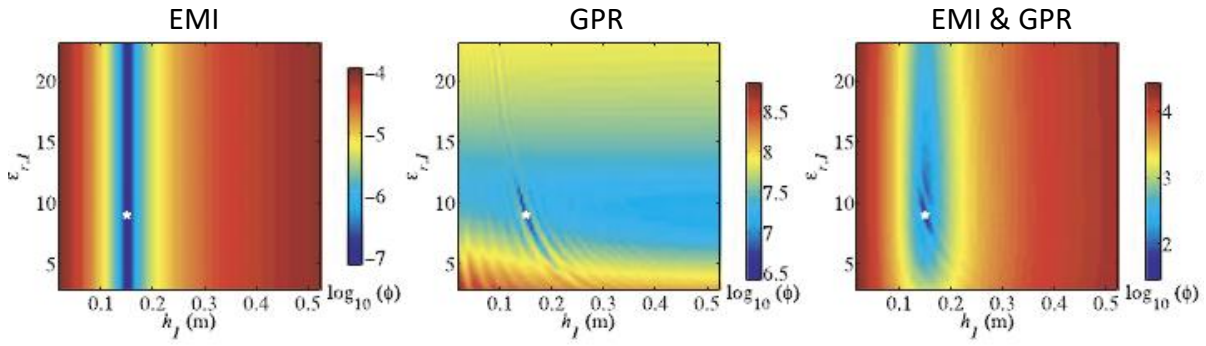



*Figure 2-11. Response surfaces of the objective function for EM, GPR, and EMI-GPR data fusion. The surface depends on the input data and fusion formulation, i.e. using GPR and EMI data separately or together. A white star marks the true parameter value (After Moghadas, 2010)*

2) Adamopoulos et al. (2021) integrated various close-range sensing techniques for the documentation of built heritage, using Castello del Valentino in Turin, Italy as a case study. They combined visible-spectrum and infrared imagery, photogrammetric techniques, terrestrial LiDAR, microwave measurements, near-infrared (NIR) imaging, thermal infrared imaging, and GPR techniques to record interdisciplinary data on the state of preservation of the historical facades. They showed that the integration of metric data with multi-wavelength

results can provide rich and valid information for historical architecture and facilitate integrated processing and interpretation.

3) Hafsi et al. (2018) proposed a new approach for the post-processing of multi-sensor detection based on data fusion provided by several technologies to improve the detection and localisation of underground networks. They suggested using a fusion technique that combines GPR imaging and electromagnetic sensors for detecting and locating buried pipes. They proposed to merge the results obtained with two detection methods, GPR and EM using a linear combination. They claimed that the combination and processing of GPR data have greatly improved the detection of underground networks. However, they did not provide specific quantitative results or metrics to evaluate the performance of the proposed approach. Figure 2-12 illustrates an example result presented by Hafsi et al. (2018), showing automatic detection using curvelet and adaptive detection methods, separately (Figure 2-12B and C) and together (Figure 2-12D), on a GPR radargram obtained from Savigny-Sur-Orge site, Paris (Figure 2-12A).

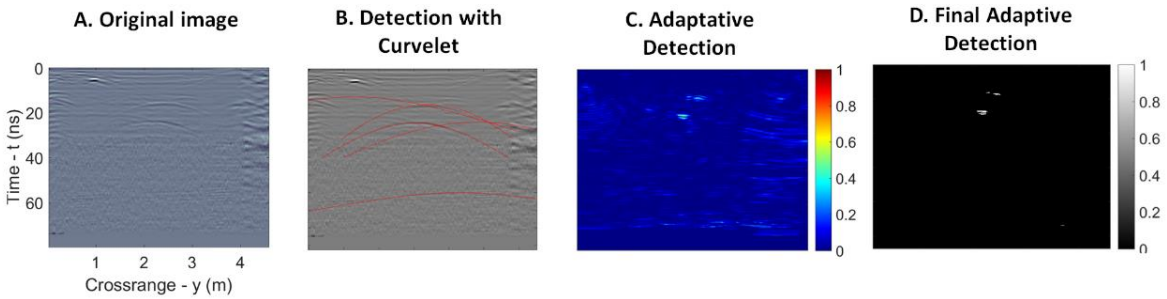



*Figure 2-12. Example of automatic detection using curvelet and adaptive detection methods on a GPR radargram obtained from Savigny-Sur-Orge site, Paris. D) The interpreted GPR radargram by both methods (After Hafsi et al. 2018)*

In general, as discussed by Gros (1997a), there are two important features that a selected data fusion method should offer: reducing the effect of sensor noise and low computational cost.

For example, it is important to reduce (or eliminate) noise and clutter from often noisy GPR data so that the desired information can be extracted.

In previous sections of this literature review, the various non-destructive geophysical methods that are commonly used for subsurface mapping have been examined, as well as the benefits of using multi-sensing data fusion techniques to combine and integrate data from multiple methods and sensors. To overcome the inherent limitations and biases of individual methods and sensors, it is often necessary to utilise advanced mathematical techniques to process and integrate the data from multiple sources. As mentioned above, automating the processing of GPR data can greatly improve the efficiency and accuracy of data analysis, particularly for large or complex data sets. A common way to achieve automation of systems in different research fields (e.g., auto-pilots, navigation system and self-driving cars) is using mathematical estimators such Kalman filter or particle filter. Such estimators have been widely used for state estimation and data fusion in dynamic systems.

In the context of subsurface mapping and exploration, these estimators have been the subject of several research studies to mitigate the effects of noise and uncertainty in GPR data and other geophysical measurements, as well as to process and integrate data from multiple geophysical methods and sensors, with the aim of improving the accuracy and reliability of the resulting maps (complete details, including advantages and disadvantages of these techniques are listed below in 2.6). Using such mathematical estimators to automate the processing of GPR data can help to reduce the need for manual data manipulation and interpretation and can improve the accuracy and reliability of the results. However, it is important to carefully select and tune the appropriate estimator for the specific goals and characteristics of the GPR data, as different estimators may have different strengths and

limitations. In the following section, the mathematical estimators that can be used as data fusion tools in GPR applications are critically reviewed.

## 2.5 Mathematical estimators for data fusion

Mathematical estimators are advanced techniques used to process and integrate data in dynamic systems. They can be divided into two categories: *stochastic estimators*, which are based on probabilistic models that incorporate uncertainty, and *statistical estimators*, which are based on statistical models of data distributions (Särkkä, 2013; Simon, 2006). Stochastic estimators, such as the Kalman filter and Markov chain monte carlo (MCMC) method, update predictions of a system's state based on sensory data and a system dynamics model (Bar-Shalom et al., 2001; Särkkä, 2013; Simon, 2006). Statistical estimators, like maximum likelihood estimator (MLE) and Bayesian estimators, estimate the parameters of a statistical model and are often used to estimate subsurface properties from GPR data (Bar-Shalom et al., 2001; Särkkä, 2013). While these estimators can improve the accuracy and reliability of GPR data processing, they also present challenges such as the need for computational resources and the risk of model misspecification or overfitting (Bar-Shalom et al., 2001; Särkkä, 2013). There are mathematical estimators that have been used in GPR applications:

1. Kalman filter: A recursive estimator that updates a prediction of the system's state based on sensory data and a model of the system's dynamics. Widely used in GPR applications for state estimation and data fusion, particularly when the data is noisy or uncertain (Simon, 2006). Carvic (1999) first introduced this technique in 1999, focusing on a series of Ground Penetrating Radar (GPR) surveys conducted over sandy soil with

low moisture content. The application of this technique was primarily for land mine detection in shallow depths of less than 10 cm. The key conclusion drawn from Carvic's work is that the Kalman Filter (KF) holds significant promise for enhancing object detection capabilities.

2. Extended Kalman filter (EKF): A variant of the Kalman filter used to estimate the state of a nonlinear system. Linearises the system around the current estimate of the state and applies the Kalman filter to the linearised system. Often used in GPR applications to handle nonlinear effects, such as dispersive waves (Särkkä, 2013; Simon, 2006). Eppstein and Dougherty (1998) proposed a method to construct a 3-D model of soil velocities at a Vermont site using GPR and an EKF approximation. The site, with stratified alluvial deposits, allowed automatic estimation of zonation values, geometry, and covariance using a "cluster tolerance criterion". The GPR estimates were consistent with cone penetrometer measurements and revealed the significant impact of soil heterogeneity on the location of infused salt water. This method also helped track 3-D soil moisture changes after saltwater injection, demonstrating potential for applications like leak detection in subsurface storage tanks. The study used geowells to facilitate GPR and electrical resistance tomography (ERT) data collection, offering a promising approach to estimate 3-D soil velocities and moisture changes.

3. Unscented Kalman filter (UKF): A variant of the Kalman filter that uses sigma points to represent the distribution of the state. More robust than the EKF when the system is highly nonlinear, but also more computationally expensive (Särkkä, 2013; Simon, 2006). Nguyen and Pyun (2015) claimed further improvement upon using UKF, however this was not quantified in their studies.

4. Markov Chain Monte Carlo (MCMC): A method that uses random sampling to estimate the parameters of a statistical model. Widely used in GPR applications to estimate the properties of the subsurface, such as depth or permittivity of layers (Gallagher et al., 2009). Wang et al. (2021) presented an innovative use of the MCMC method, combined with a double difference strategy, to achieve time-lapse GPR inversion. Utilising a priori information in the MCMC method, the researchers managed to implement local sampling in time-lapse inversion. This allowed sampling to be confined to the target area, hence reducing the impact of non-target areas and decreasing computational requirements. They used the Finite-Difference Time-Domain (FDTD) algorithm on a synthetic GPR data set, simulating a reference model (5.2 × 12 m in size and consisting of 6240 pixels, each 0.1 m × 0.1 m in size) generated from a multivariate Gaussian probability distribution, with a constant electrical conductivity of 3 mS/m and relative dielectric permittivity of 3.97. An extended Metropolis algorithm was utilised for local sampling, and double difference was introduced as a time-lapse inversion strategy. The results showed that, in comparison with full sampling, local sampling considerably enhanced the resolution of the target region, thus increasing the precision of the inversion results.

5. Particle filter: A Monte Carlo method that uses a set of particles to represent the distribution of the system's state. Similar to the Kalman filter but does not require a linear model and can handle non-Gaussian distributions. Often used in GPR applications to track the evolution of the subsurface structure over time (Simon, 2006). Ng et al. (2008) conducted a research study to compare the performance of KF with stochastic Particle filter-based methods in detecting buried landmines. The authors proposed a combination of reversible jump Markov chain Monte Carlo in conjunction

with Sequential Monte Carlo (SMC) for landmine detection, claiming that it offers improved performance. However, the quantitative analysis of their findings was lacking in their study.

6. Interacting multiple model (IMM) filter: A method that combines multiple Kalman filters to handle systems with multiple operating modes or models. Often used in GPR applications to track the evolution of the subsurface structure in the presence of multiple layers or interfaces (Mazor et al., 1998). Venkatasubramanian and Leung (2006) proposed an EM-IMM based abrupt change detector for landmine detection. Their method dynamically learns the parameters of different models in real-time without prior information. To validate their approach, they utilised real GPR data obtained from the VUB demining website, collected in a mixture of sand and small rocks. Each trace in the data represents an average of 10 consecutive returns, comprising 512 samples, sampled at 25 ps. The proposed method demonstrates promising results however, this was not precisely quantified in their studies.

7. Maximum Likelihood Estimator (MLE): A statistical estimator that estimates the parameters of a statistical model by maximising the likelihood function. Often used in GPR applications to estimate the properties of the subsurface from the measured data (Myung, 2003). Nadim (2008) presents an algorithm aimed at reducing clutter signal in GPR data, specifically caused by surface reflections. The proposed technique utilised a likelihood ratio test with estimated parameters. The algorithm is applied to GPR data for the detection of subsurface anti-personnel (AP) landmines. The results demonstrate that the proposed algorithm reduces clutter signal and enhances the detection of landmine objects in GPR data. However, the paper does not provide

information regarding ground conditions or soil properties, except for the size of buried landmines (12 x 5 x 7 cm).

8. Bayesian estimator: A statistical estimator that uses Bayesian inference to estimate the parameters of a statistical model. Combines the observed data with a prior distribution over the parameters to form a posterior distribution, which is used to estimate the parameters. Often used in GPR applications to incorporate prior knowledge about the subsurface into the estimation process (Särkkä, 2013). Luo and Fang (2005), utilised adaptive estimation and Bayesian approaches for the detection of landmines and clutter reduction. They achieved a detection rate of 90% and 95% for adaptive estimation and Bayesian approaches, respectively, in the domain of buried target detection. However, the report did not provide specific details regarding the characteristics of the buried target, or the ground conditions involved in their experiments.

Table 2-6 presents a comparison between the main characteristics of the above-mentioned methodologies for processing GPR data including their strengths and limitations.

*Table 2-6 Comparison of mathematical estimators in GPR application*

| Method | Advantages | Limitations |
|---|---|---|
| Kalman filter (Simon, 2006) | ▪ Provides optimal estimates for linear systems with Gaussian noise<br>▪ Is computationally efficient and can be implemented in real-time<br>▪ Can be extended to nonlinear systems using techniques such as the extended Kalman filter (EKF) or the unscented Kalman filter (UKF) | ▪ Limited to linear models: The Kalman filter is typically only applied to linear models, so it may not be suitable for nonlinear systems. Additionally, the linear models used in the Kalman filter may not accurately represent the dynamics of the system, leading to errors in the estimates.<br>▪ Can be sensitive to initialisation and parameter tuning: The KF relies on initial estimates of the parameters, and these estimates can greatly influence the result. If the initial estimates are far from the true values, it can be difficult for the Kalman filter to converge to the correct solution.<br>▪ Limited to systems with Gaussian noise: The Kalman filter is typically only applied to systems where the noise is Gaussian, so it may not be suitable for systems with non-Gaussian noise. Additionally, the assumption of Gaussian noise may not accurately represent the noise in the system, leading to errors in the estimates. |
| Particle filter (Simon, 2006) | ▪ Can handle nonlinear and non-Gaussian systems<br>▪ Can provide a probabilistic representation of the state of the system<br>▪ Can be implemented in real-time using resampling techniques | ▪ Computational complexity: Particle filters involve the use of a large number of particles to represent the probability distribution of the system, which can make them computationally intensive, especially for high-dimensional systems or complex models. This can make it challenging to implement particle filters in real-time applications.<br>▪ Sensitivity to model error: Particle filters rely on a model of the system to propagate the particles and update their weights. If the model is not |

| | | |
|---|---|---|
| | | accurately representing the true dynamics of the system, this can lead to errors in the estimates.<br>▪ Sensitivity to initial estimates: Particle filters rely on initial estimates of the parameters, and these estimates can greatly influence the final result. If the initial estimates are far from the true values, it can be difficult for the particle filter to converge to the correct solution.<br>▪ Limited to systems with additive noise: Particle filters are typically only applied to systems where the noise is additive, so they may not be suitable for systems with more complex noise structures. Additionally, the assumption of additive noise may not accurately represent the noise in the system, leading to errors in the estimates. |
| Markov Chain Monte Carlo (MCMC)<br>(Gallagher et al., 2009) | ▪ Can handle complex models with unknown or complex probability distributions<br>▪ Can provide a probabilistic representation of the posterior distribution of the parameters<br>▪ Can be used for a wide range of applications, including Bayesian inference, optimisation, and sampling from complex distributions | ▪ Computational complexity: MCMC is a computationally intensive method, especially for high-dimensional systems or complex models. This can make it challenging to implement MCMC in real-time applications.<br>▪ Sensitivity to initial estimates: MCMC relies on initial estimates of the parameters, and these estimates can greatly influence the final result. If the initial estimates are far from the true values, it can be difficult for MCMC to converge to the correct solution.<br>▪ Limited to models with known probability distributions: MCMC can only be applied to models where the probability distribution is known or can be accurately estimated. This means that it is not suitable for models with complex or unknown probability distributions. |

| | | |
|---|---|---|
| | | ▪ Convergence time: MCMC may require a large number of iterations to converge to the correct solution, especially for high-dimensional systems or complex models. This can make it time-consuming to use MCMC in practice, especially when the convergence time is not known a priori. |
| Extended Kalman filter (EKF)<br>(Särkkä, 2013) | ▪ Can be applied to nonlinear systems<br>▪ Is computationally efficient and can be implemented in real-time<br>▪ Can be used to estimate the state of a system in the presence of noise and uncertainty | ▪ Sensitivity to model error: The EKF relies on a linearised model of the system, which can lead to errors in the estimates if the model is not accurately representing the true dynamics of the system. This sensitivity to model error can be particularly problematic for nonlinear systems.<br>▪ Sensitivity to initial estimates: The EKF relies on initial estimates of the parameters, and these estimates can greatly influence the final result. If the initial estimates are far from the true values, it can be difficult for the EKF to converge to the correct solution.<br>▪ Limited to systems with Gaussian noise: The EKF is typically only applied to systems where the noise is Gaussian, so it may not be suitable for systems with non-Gaussian noise. Additionally, the assumption of Gaussian noise may not accurately represent the noise in the system, leading to errors in the estimates. |
| Unscented Kalman filter (UKF)<br>(Särkkä, 2013) | ▪ Can be applied to nonlinear systems<br>▪ Can provide accurate estimates even for systems with large nonlinearities | ▪ Complexity: The UKF involves the use of sigma points to represent the probability distribution of the system, which can make it more complex to implement compared to some other filtering methods.<br>▪ Sensitivity to model parameters: The performance of the UKF is highly dependent on |

| | | |
|---|---|---|
| | ▪ Is computationally efficient and can be implemented in real-time | the accuracy of the model parameters. If the model parameters are poorly estimated or changed significantly over time, the UKF may perform poorly.<br>▪ Limited to linear models: The UKF is typically only applied to linear models, so it may not be suitable for nonlinear systems. Additionally, the linear models used in the UKF may not accurately represent the dynamics of the system, leading to errors in the estimates. |
| Interacting multiple model (IMM) filter<br>(Mazor et al., 1998) | ▪ Can handle systems with multiple modes of operation<br>▪ Can provide a probabilistic representation of the state of the system<br>▪ Can improve the accuracy of the estimates by combining information from multiple models | ▪ Complexity: The IMM filter involves multiple models and requires the implementation of a switching mechanism to determine which model is most appropriate at each time step. This can make it more complex to implement compared to some other filtering methods.<br>▪ Sensitivity to model parameters: The performance of the IMM filter is highly dependent on the accuracy of the model parameters. If the model parameters are poorly estimated or changed significantly over time, the IMM filter may perform poorly.<br>▪ Limited to linear models: The IMM filter is typically only applied to linear models, so it may not be suitable for nonlinear systems. Additionally, the linear models used in the IMM filter may not accurately represent the dynamics of the system, leading to errors in the estimates<br>▪ Sensitivity to model switching: The IMM filter relies on accurate model switching to switch between the different models in the filter. If the model switching is not performed correctly, it can lead to errors in the estimates. |

| | | |
|---|---|---|
| | | ▪ Limited to systems with multiple modes: The IMM filter is typically only applied to systems with multiple modes, so it may not be suitable for systems with a single mode of operation.<br>▪ Limited to models with known probability distributions: The IMM filter requires the probability distributions of the different models to be known or accurately estimated. This means that it is not suitable for models with complex or unknown probability distributions. |
| Maximum Likelihood Estimator (MLE)<br>(Myung, 2003) | ▪ Can be robust to noise and outliers in the data<br>▪ Can be used to estimate the parameters of a statistical model<br>▪ Is widely used in a variety of fields, including statistics, machine learning, and engineering | ▪ Computational complexity: MLE can be computationally intensive, especially for high-dimensional systems or complex models. This can make it challenging to implement MLE in real-time applications.<br>▪ Sensitivity to initial estimates: MLE relies on initial estimates of the parameters, and these estimates can greatly influence the final result. If the initial estimates are far from the true values, it can be difficult for MLE to converge to the correct solution.<br>▪ Limited to models with known probability distributions: MLE can only be applied to models where the probability distribution is known or can be accurately estimated. This means that it is not suitable for models with complex or unknown probability distributions. |
| Bayesian estimator<br>(Särkkä, 2013) | ▪ Provides a probabilistic representation of the posterior distribution of the parameters | ▪ Requires a prior distribution: The Bayesian estimator relies on a prior distribution to estimate the posterior distribution of the parameters. This means that the prior distribution must be specified in advance, which can be challenging if it is not known a priori. |

| | |
|---|---|
| ▪ Can handle complex models with unknown or complex probability distributions<br>▪ Can incorporate prior knowledge or expert knowledge into the estimates | Additionally, the choice of the prior distribution can influence the final result, so it is important to choose a prior that is representative of the true distribution of the parameters.<br>▪ Limited to models with known likelihood functions: The Bayesian estimator requires the likelihood function, which is a measure of the probability of the observed data given the model and the parameters, to be known or accurately estimated. This means that it is not suitable for models with complex or unknown likelihood functions.<br>▪ Requires sufficient data: The Bayesian estimator relies on a large amount of data to accurately estimate the posterior distribution of the parameters. If the data is limited or noisy, it can be difficult for the Bayesian estimator to converge to the correct solution.<br>▪ Can be computationally intensive for high-dimensional systems or complex models<br>▪ Can be sensitive to initialisation and parameter tuning |

Among the various reviewed mathematical estimators that have been used for processing GPR data and fusion of multi-sensors, the Kalman filter stands out as a highly effective and efficient method. While the Kalman filter may have limitations such as its requirement for a linear model and sensitivity to initialisation and parameter tuning, these limitations are often less severe compared to those of other methods (Särkkä, 2013; Simon, 2006). For example, particle filters and Markov Chain Monte Carlo (MCMC) methods are computationally intensive and may not be able to handle high-dimensional systems, while the extended Kalman filter

and unscented Kalman filter can be less accurate for highly nonlinear systems (Särkkä, 2013; Simon, 2006). The interacting multiple model (IMM) filter can be complex to implement and may not be able to handle highly nonlinear systems or systems with rapidly changing operating modes (Mazor et al., 1998). The maximum likelihood estimator (MLE) may require more complex optimisation algorithms, while the Bayesian estimator can be computationally intensive for high-dimensional systems or complex models and may be sensitive to the choice of the prior distribution (Myung, 2003).

Zoubir et al. (2002), Ng et al. (2008), Luo and Fang (2005), Patel and Ferguson (2021), and Kaniewski and Kraszewski (2020) reviewed the use of KF in GPR data processing, comparing various techniques and exploring multiple applications, such as land mine detection, pipe detection, radar application, and general metal object detection with complex geometry. They concluded that KF is a powerful mathematical estimator that can be used for both processing GPR data and fusion of multi-sensors. Its combination of simplicity, efficiency, and robustness make it an attractive choice for many subsurface exploration applications, particularly when compared to the limitations of other methods.

### 2.5.1 Kalman Filter (KF)

Kalman filtering is a mathematical tool for estimation that has many practical applications, including in aerospace, radar, navigation systems, and autonomous vehicles (Auger et al., 2013; Li et al., 2016; Meinhold and Singpurwalla, 1983; Pearson and Stear, 1974). It was first introduced by Rudolf Emil Kalman in 1960 as a recursive algorithm for probabilistic state estimation and a solution for Bayesian filtering (Kalman, 1960). Recursive estimation is a technique that assumes that the posterior probability density function is based on previous measurements (Särkkä, 2013). Kalman filtering uses a system model, defined by a set of

ordinary differential equations, to determine the behaviour and probability density functions of a dynamic system over time, based on a series of observations including uncertainties (Simon, 2006). This system model can be used to measure the system's response and predict its future behaviour in order to make decisions (Brunton and Kutz, 2017). A dynamic system can be mathematically modelled using a probabilistic state-space model, which is a mathematical model of the fundamental physics of a system that characterises the system by describing how it will evolve through multiple states over time (Van Drongelen, 2010). The state of a dynamic system refers to unknown variables that provide information about the internal condition of the system at specific times and take into account previous observations (Simon, 2006).

It is important to determine the probabilistic state-space model as accurately as possible in order to avoid estimation drift (variation in estimated variables) and produce reliable results. Drift in a linear probabilistic state-space model can grow indefinitely with the development of the system and the acquisition of more measurements over time, leading to the divergence of the algorithm or the production of unreliable results (Särkkä, 2013).

Recursive Bayesian filtering is a Markov process (Marcus and Rosen, 2006) that operates under two assumptions: i) the Markov property, which states that the likelihood of a future state depends only on its current state (Kirkwood, 2015), and ii) the hidden Markov model (Figure 2-13), which states that although the system states are not directly visible, the system outputs, which depend on those invisible states, are visible (Kirkwood, 2015).

A brief overview of Bayesian filtering is provided here to gain a deeper understanding of the Kalman filter approach. Generally, Bayesian filtering consists of three recursive steps: initialisation, prediction, and correction (updating). The initialisation step is followed by the

prediction and correction steps. The prediction of a system state (a priori estimate) uses executed commands (if any) and the state-space model of the system. Prediction is essentially the convolution of two probability distributions: the system model and the previous estimate (belief) of the system. Correction (a posteriori estimate) is performed using sampled sensor data (observations) in order to update the previous estimate (Bar-Shalom et al., 2001; Gauss, 1857; Grewal and Andrews, 2008).

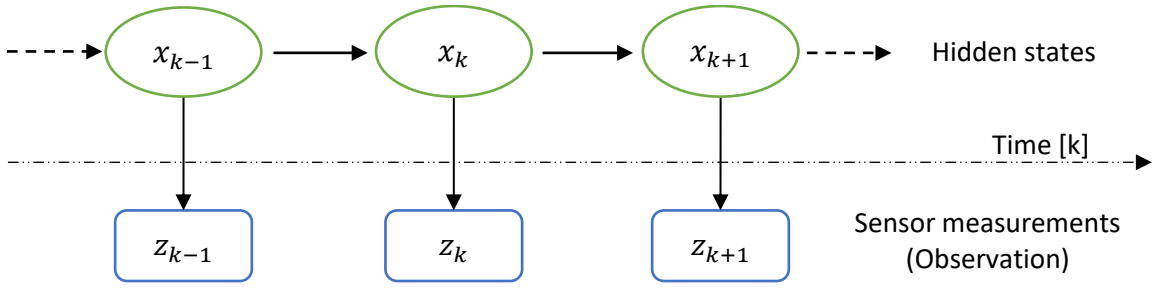



*Figure 2-13. Hidden Markov Model*

The Kalman filter is a implementation of the Bayesian filter that creates state estimates by propagating the mean of the state based on the probabilistic state-space model of the system and sensor measurements at each time frame in discrete-time, continuous-time, or sampled-data systems, but with two important assumptions: i) the system is linear (linear probabilistic state-space model), and ii) everything is Gaussian and characterised by mean and covariance values with a probability density function as shown in Eq. 2-3 (Maybeck, 1979; Simon, 2006).

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right] \quad (2\text{-}3)$$

Where $X$ is a Gaussian random variable [ $X \sim N(\mu, \sigma^2)$ ] with the normal distribution $N$, the mean value $\mu$, the standard deviation $\sigma$ and the covariance $\sigma^2$.

The principle of an optimal state estimate for a linear-Gaussian model is shown in Figure 2-14. The standard deviation of the initially predicted state estimate, which can include statistical or other types of uncertainties, is optimised in the optimal state estimate after updating the initial estimate by applying and taking a new measurement into account.

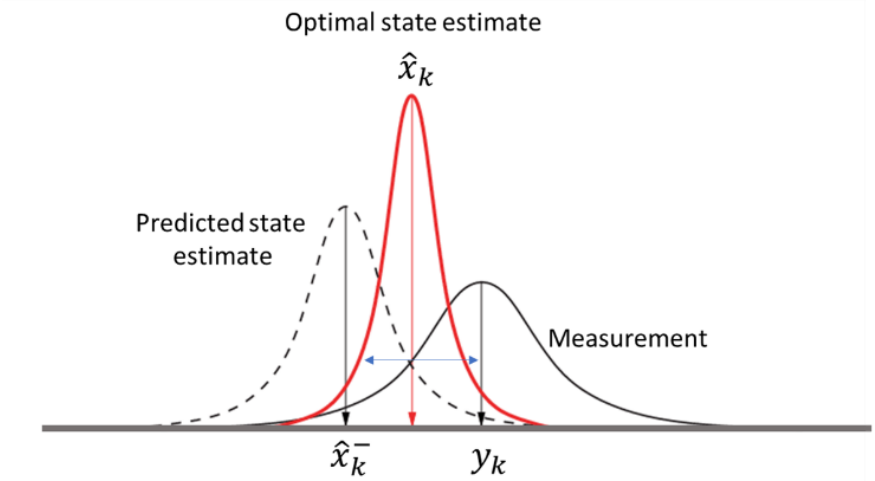



*Figure 2-14. Optimal estimate through prediction and correction steps in a Gaussian world*

The Gaussian random walk model is a simple Gaussian model that is suitable for systems with noisy measurements, such as GPR. It is a simple form of the Markov process and is represented in Eq. 2-4 and Eq. 2-5. It is commonly used in applications where the system exhibits random walk behaviour. (Lopes et al., 2015; Särkkä, 2013).

$$x_k = x_{k-1} + q_{k-1} \quad , \quad q_{k-1} \sim N(0, Q_{k-1}) \tag{2-4}$$

$$y_k = x_k + r_k \quad , \quad r_k \sim N(0, R_k) \tag{2-5}$$

Where $x_k$ is the signal ($x_k \in \mathbb{R}$) which is the hidden state (hidden Markov model) at instant time $k$, $y_k$ is the sensor measurement ($y_k \in \mathbb{R}$) at instant tine $k$, and the process and measurement (sensor) noises $q$ and $r$, respectively. Process and measurement noises are assumed to be Gaussian white noise with a normal distribution $N$ with zero-mean and error covariances $Q$ and $R$.

Taking the Gaussian random walk model into account, each noisy sensor measurement $y_k$ can be assumed as a function of signal $x_k$ and associated measurement noise $\upsilon_k$, as presented in Eq. 2-6.

$$y_k = g(x_k, \upsilon_k) \tag{2-6}$$

It is well-known that sensor measurements in a survey are not always accurate and reliable. Errors and uncertainties in real-world applications are inevitable but important variables that can significantly affect the accuracy of surveyed data. As Simon (2006) points out, it is always important to be aware of any information provided by each measurement, even if the level of confidence in the reliability of some of the measurements is low. However, inaccurate measurements in the Kalman filter can also lead to unstable posterior estimates and even divergence of the filter as a result of the propagation of errors over time. Therefore, it is important to define the level of confidence in both sensor measurements and Kalman filter estimates at each time frame after each measurement. The level of confidence can be established by determining a weighting parameter called the Kalman gain (KG), which controls the influence of the sensor measurement on updating future estimates based on the available information (including uncertainties). The general form of the Kalman gain is shown in Eq. 2-7.

$$\mathrm{KG} = \frac{U_e}{U_e + U_m} \qquad (0 \leq \mathrm{KG} \leq 1) \tag{2-7}$$

Where $U_e$ is the uncertainty in KF estimate, and $U_m$ is the uncertainty in sensor measurement (observations). As illustrated in Figure 2-15, if the sensor measurement is accurate, then the measurement uncertainty relative to the uncertainty in an estimate would be a very small value by which the value of the KG approximate to value one, which is the highest possible value of Kalman gain.

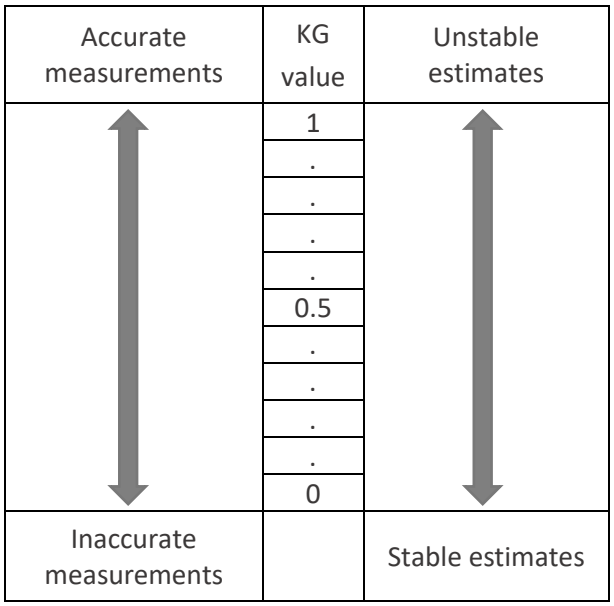



*Figure 2-15. Kalman Gain distribution*

The Kalman gain is updated every time a new measurement is added to the system at each time frame, based on the correction term $(y_k - \hat{x}_{k-1})$, which is the residual between a new sensor measurement $(y_k)$ and the priory estimate of the KF $(\hat{x}_{k-1})$. This allows the Kalman gain to adapt to the current measurement and update the estimate accordingly.

As previously mentioned, KF propagates the mean of the state of a system at each time frame. However, due to the unavoidable uncertainties in both process and sensor measurements, these mean values are not necessarily the absolute or true estimate of a system at each step. In other words, similar to process and sensor measurements, state estimates are also subject to deviations and uncertainties that need to be addressed. This problem can be addressed by determining and propagating a variable called the state estimation error covariance $(P_k)$, which indicates the uncertainty in KF estimates at each time frame. Determining the estimation error covariance is also essential for calculating the Kalman gain and defining the level of confidence and the influence of new measurements at each time frame.

An increase in the number of new measurements at each time frame (including less reliable measurements) will automatically increase the number of recursive steps required to take

these new measurements into account (Hartikainen and Särkkä, 2011; Simon, 2006).This can lead to a significant increase in computational costs in terms of processing and memory. However, KF prevents this increase by using the recursive Gauss least squares approach, developed by Carl Gauss (1857), to simply estimate the value of each new noisy measurement by minimising the sum of squares of the residual between the actual sensor measurements and the a priori estimates. In this way, KF does not rely on all previous measurements, but only considers new measurements to produce the next estimate at each time frame. The advantage of using this approach is that it has significantly lower computational costs due to fast processing and low memory usage, compared to other approaches such as the particle filter, which require complex calculations (Simon, 2006).

The KF process is depicted in Figure 2-16. It includes the following steps: prediction, sensor measurement, Kalman gain and error covariance calculation, estimate update, and new estimate generation.

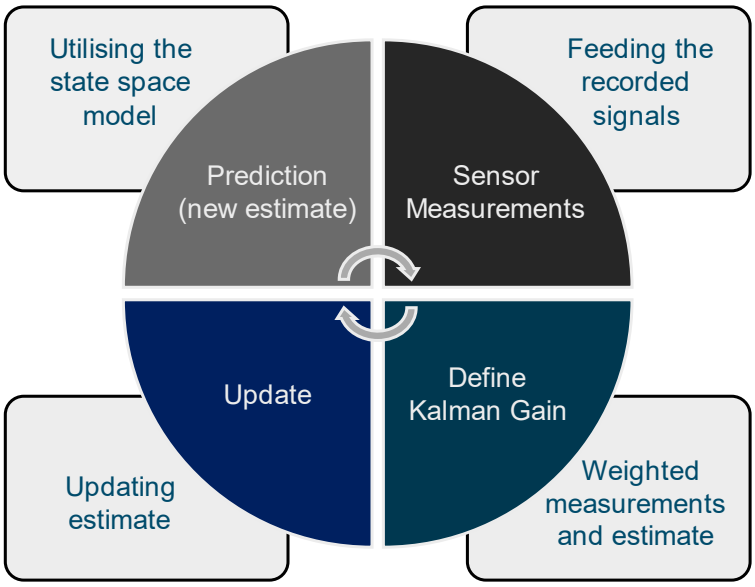



*Figure 2-16. The general framework of Kalman Filtering*

## 2.5.2 Kalman Filter in GPR application

The Kalman Filter (KF) has been used to estimate the background signal of GPR radargrams (Luo and Fang, 2005; Zoubir et al., 2002b). The KF works by iteratively combining noisy signals from a sensor (backscattered signals) with processed data (estimates generated by the KF) to estimate the target and background signals. The KF generates the best estimate by optimally separating signal and noise using a least square-based approach (Van Drongelen, 2010).

The KF-based algorithm for estimating and separating the background of GPR radargrams was first proposed by Carevic (1999). This approach involves a stepwise analysis of the backscattered signal to estimate the internal condition of the system at specific times, known as the system state (Simon, 2006). The algorithm uses a parameter called "innovation", which represents the residual between actual and estimated signal values, to determine whether a target signal is present. This decision is made using the chi-squared $\chi^2$ hypothesis test, a statistical method for comparing a model to observed data (Bar-Shalom et al., 2001; McHugh, 2012). The probability of target detection is based on the innovation values and the chi-squared test (Carevic, 1999; Luo and Fang, 2005; Ng et al., 2008; Zoubir et al., 2002b). However, the chi-squared method has some limitations in this application. First, there is no definite method for determining an optimal threshold value to reject the null hypothesis in the chi-squared test (Gurbuz, 2012). Second, the test statistics in GPR have an unknown distribution (Ng et al., 2008). Additionally, the Carevic (1999) algorithm cannot determine the width of buried targets (Ng et al., 2008; Zoubir et al., 2002b), which limits its usefulness for complex buried infrastructure.

The Carevic algorithm was implemented by Zoubir et al. (2002), Luo and Fang (2005), and Ng et al. (2008) for the detection of unexploded ordnances in a sandy soil environment using KF

and receiver operating characteristic (ROC) curves as a performance validation tool. ROC is a measure of the trade-off between the probability of true detections and false alarms in the context of target detection. It is defined as the area under the curve that relates the probability of false detection to the probability of true detection (Kerekes, 2008). Zoubir et al. (2002) compared the computational cost and false alarm ratio of Carevic's method to other background noise removal methods for landmine detection and found that it provided the best performance. Luo and Fang (2005) extended Carevic's work by proposing two additional approaches, adaptive estimation and Bayesian, and reported a buried target recognition rate of over 90%. KF has also been recently implemented by Smitha and Singh (2016) and Xu et al. (2018a) to reduce noise in GPR radargrams.

Figure 2-17 illustrates the performance of KF compared to other methods, using ROC curves, where KF has demonstrated the highest probability of detection (Zoubir et al., 2002a).

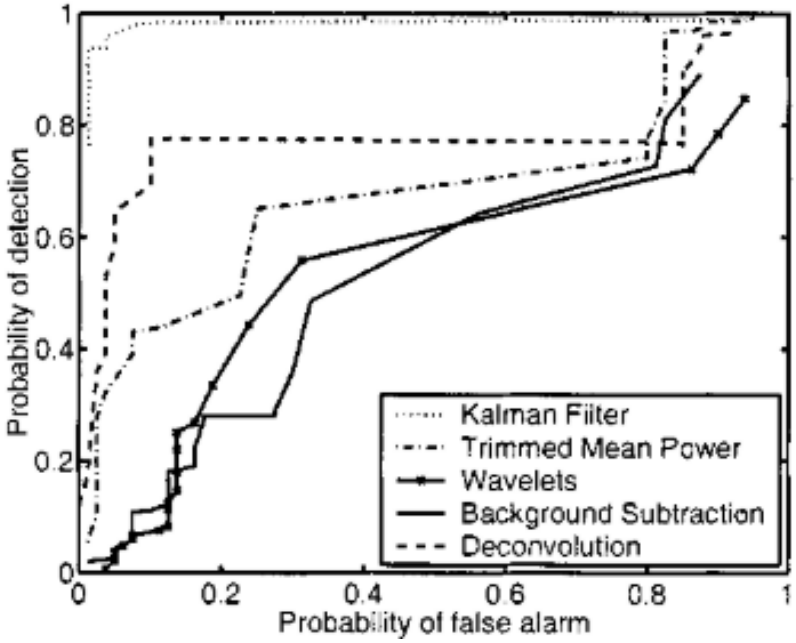



*Figure 2-17. Performance of the KF for detection of land mine in comparison to other methods using ROC curves (After Zoubir et al., 2002)*

Figure 2-18, presented by Smitha and Singh (2016), shows another example of using the KF for landmine detection. This figure compares a GPR radargram before and after applying the KF to reduce the clutter, specifically the air-ground bounce.

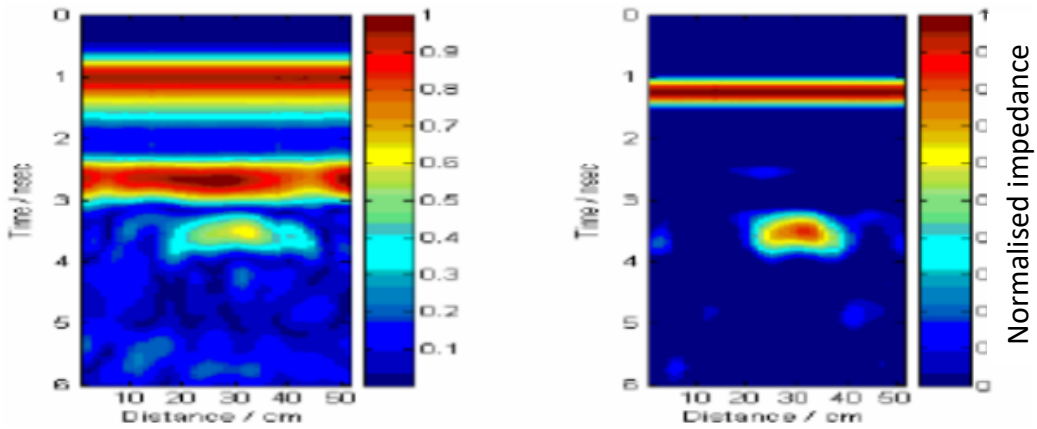



*Figure 2-18. GPR radargram before (left) and after applying the KF for reducing the clutter (right) (After Smitha and Singh, 2016)*

A summary of works in the GPR application in which KF is utilised, or its performance was compared with other methods, is presented in Table 2-7.

*Table 2-7. Summary of works on GPR in correlation with Kalman filtering*

| | Article | Approaches | Purpose of utilising the KF |
|---|---|---|---|
| **1** | (Carevic, 1999) | Kalman Filter<br>Chi-squared<br>Background Adaption | Target detection and locating in GPR radargrams |
| **2** | (Van Kempen and Sahli, 2001) | ARMA-Model<br>Standard Kalman Filter | Utilised the parameters of the ARMA model to estimate clutters |
| **3** | (Zoubir et al., 2002a) | Kalman Filter<br>Chi-squared<br>Background removal<br>Matched Filter Deconvolution<br>Wavelet Decomposition<br>Trimmed Average Power | Comparison of Carevic's KF performance in GPR application against other available methods for detection and locating buried targets |
| **4** | (Abeynayake et al., 2003) | standard Kalman filter<br>Metal detector | landmine detection using metal detector arrays |
| **5** | (Luo and Fang, 2005) | Improved Kalman Filter<br>Background removal<br>Improvement:<br>Adaptive estimation<br>Bayesian approach | Improvement of Carevic's KF for clutter reduction in GPR radargrams |
| **6** | (Torrione and Collins, 2006) | low- latency KF | Tracking of temporal position of the air/ground response in landmine detection |
| **7** | (Venkatasubramanian and Leung, 2006) | Interacting multiple model (IMM) filter<br>Kalman Filter | Performance comparison of both methods |
| **8** | (Ng et al., 2008) | Kalman Filter<br>Background removal<br>Particle filter:<br>Reversible jump Markov chain Monte Carlo<br>Sequential Monte Carlo | Comparison of particle filter, as a complex version of the KF, with Carevic's KF in landmine detection |
| **9** | (Smitha and Singh, 2016) | Kalman Filter<br>Likelihood ratio test<br>Singular Value Decomposition<br>Principal Component Analysis<br>Independent Component Analysis (ICA)<br>Wavelet Transform (WT) | Applied Carevic's KF to reduce noise in GPR radargrams |
| **10** | (Dou et al., 2016) | Extended KF | Marching utility tracks in B-scans |

| | | | |
|---|---|---|---|
| **11** | (Xu et al., 2018a) | Kalman filter<br>Geometric correction<br>Amplitude matching | Applied Carevic's KF to reduce noise in GPR radargrams |
| **12** | (Kaniewski and Kraszewski, 2020) | Kalman Filter<br>Pendulum motion model | Position estimation of GPR antenna on the surface |
| **13** | (Zou et al., 2020) | Kalman Filter<br>GPS | Tracking the position of the GPR antenna on the surface |

While there are several methods for processing and interpreting GPR data, each has shown to have various limitations in terms of accuracy, application, computational time etc. Kalman Filter is known to be an effective method for data analysis with advantages such as its ability to handle noisy and uncertain data, simplicity, and efficiency. However, relatively few studies, have focused on using the KF in GPR applications. The previous section of the literature review examined the capabilities and limitations of the KF in processing GPR data and fusing multi-sensing data. In particular, the literature review indicates there are some limitations associated with KF, including the need for a linear model, sensitivity to initialisation, and the requirement for parameter tuning. This suggests there are gaps in the literature that will be discussed in the following section.

Although KF has demonstrated significant advantages and outperformed other methods in the literature, the lack of a robust algorithm to identify the target width has remained a major issue. Additionally, there is no definite method to determine the optimal threshold value for the chi-squared hypothesis in GPR applications, resulting in the decision of whether the target signal is present or not being made through a trial-and-error approach or using threshold values defined from other fields. This lack of a robust KF algorithm for GPR applications is a major gap in the body of knowledge. Currently, the probability of identifying a target is estimated using externally defined benchmarks rather than actual data from a survey. Additionally, previous works have not investigated the use of KF in identifying the condition

of the ground supporting buried anomalies, which is another gap in the knowledge base. Furthermore, some input parameters for KF, such as the system and sensor noise covariances, are determined manually, which can be a time-consuming process and may cause further issues, which are discussed below as a technical note on the previous literature.

Improving such limitation by, e.g., automating the optimisation of KF parameters rather than manual selection by an expert can potentially increase the objectivity of KF performance.

#### 2.5.2.1 A technical note on KF in GPR application

All the reviewed works that have utilised KF for GPR data processing, initialised these parameters to zero (noise covariance matrix = 0). This initialisation, however, contradicts the theory of KF when dealing with the value of noise. As per the definition of KF, the noise covariance matrix, consisting of system and sensor noise covariances, should be initialised to zero only if an accurate level of noise can be determined. However, in the absence of a reliable and accurate level of noise, the noise covariance matrix must be initialised to a very large number ($\approx \infty$) (Bar-Shalom et al., 2001; Grewal and Andrews, 2008; Hartikainen and Särkkä, 2011; Kalman, 1960; Maybeck, 1979; Särkkä, 2013; Simon, 2006; Welch and Bishop, 1995).

It is well-known that in a GPR survey, the level of potential noise from both on the ground and in the subsurface sources of noise is unknown. Inappropriate initialisation of these parameters can potentially lead to unwanted errors and inaccuracy or even divergence of KF (Simon, 2006). Thus, rectifying the abovementioned systematic issue can improve the convergence and accuracy of KF performance. On this basis, the lack of a robust approach to determine optimised KF parameters, such as error covariance matrix, is an area deserving greater and immediate attention, i.e. gap in knowledge.

## 2.5.3 Optimisation of KF parameters in GPR application

In the above section, it was noted that some Kalman filter parameters need to be optimised for GPR applications and are currently initialised manually by experts. Non-linear optimisation techniques, such as evolutionary optimisation algorithms, can potentially automate this optimisation process (Zhang and Atia, 2020). Rezaei et al., (2009) and Jianguo et al.,(2008) have demonstrated the use of evolutionary optimisation algorithms, such as genetic algorithms, to optimise KF parameters for other applications, such as flight control systems and chemistry. These algorithms could potentially be used to optimise KF parameters in GPR applications as well.

### 2.5.3.1 Evolutionary optimisation

Evolutionary algorithms (EAs) are a branch of computational intelligence that use principles inspired by nature to optimise formulated problems, or to make something better. EAs are a process of identifying the best or optimal solution to a problem to improve the outcome. For example, in geophysics, the problem of converting analysed geophysical data into an image of the subsurface or an estimation of ground and buried anomaly properties (e.g. the location of a buried pipe or soil disturbance) can be formulated as an optimisation problem (Sen and Mallick, 2018; Yogi and Widodo, 2017). This formulation optimises the parameters of a numerical model to produce data similar to the analysed geophysical data (Maiti et al., 2013). Evolutionary algorithms are widely used in geophysical optimisation problems (Sen and Mallick, 2018).

The idea behind EAs is based on Darwin's theory of evolution and the survival of the fittest, as well as Mendel's principles of modern genetics (Allen, 2003; Oldroyd, 1986). EAs offer approximate solutions to complex problems. The problem is usually associated with an

objective function, or the problem's objective expressed as a function, that is solved based on user specifications and the desired accuracy as a decision variable (Gill et al., 2019). The fitness of a solution in an optimisation problem is determined by a fitness function that takes the provided solution as input and typically outputs a single value measuring how fit the solution is. Higher fitness values indicate a better solution (Deb and Deb, 2014; Gennert and Yuille, 1988; Oldroyd, 1986; Törn and Žilinskas, 1989). Some problems may involve multiple objective functions that need to be optimised simultaneously, called multi-objective optimisation (MOO) or Pareto optimisation, which satisfies multiple decision criteria (Chang, 2015; Cheng, 1999; Čuček et al., 2015). The solution to MOO problems is usually expressed as a set of Pareto solutions, or Pareto optima, which represent an optimal trade-off between the solution criteria of multiple objectives (Limleamthong and Guillén-Gosálbez, 2018).

Existing methods and tools for subsurface exploration typically involve the fusion of non-destructive testing (NDT) data through expert interpretation of the data either sequentially or in parallel. However, despite the effectiveness of Kalman filters in estimation and fusion applications and the growth and power of computational intelligence, there is a significant gap in current knowledge on the use of KFs in GPR applications. Specifically, there is a lack of autonomous computational solutions that utilise KFs in conjunction with evolutionary algorithms to autonomously fuse data from different sensors after objective post-processing and analysis, reducing the need for expert interaction.

In the following, two nonintrusive electromagnetic techniques are introduced as candidates, based on the discussions above in the literature, for data fusion with GPR. These two methods, namely electromagnetic location (EML) and electromagnetic conductivity (EMC) each were selected for different reasons. EMC data was used in previous studies for data fusion with GPR (Table 2-7) and was shown to possess complementary information to GPR data. Thus, EMC

was selected as a potential candidate to perform data fusion with GPR. For EML, on the other hand, there is no reported data fusion performed with GPR, to the best knowledge of the author. Furthermore, very limited systematic data acquisition has been performed for EML. Thus, studying the performance of fused GPR and EML data is an attractive gap to investigate.

## 2.6 EM-based techniques for data fusion with GPR

The following two EM-based techniques, i.e. EML and EMC, are among the widely used nonintrusive methods for subsurface exploration. The potential of these techniques for data fusion with GPR is further investigated during this thesis. Here, to better understand the principles, advantages, and limitations of these methods, a review of the reported studies in the literature is presented.

### 2.6.1 Electromagnetic locator (EML) – Cable avoidance tool (CAT)

For decades, electromagnetic induction (EMI) has been used in geophysical surveys to explore metallic minerals. Electromagnetic utility locating systems, known as EM-Locators (EMLs) or pipe and cable locators (PCLs), are the most common technique in the construction and utility detection sector for rapidly detecting metallic utilities in-situ (Bryakin and Bochkarev, 2019). EMLs come in a range of instruments, from simple cable avoidance tools (CATs) for locating buried service utilities to sophisticated condition assessment systems for detecting damage or deterioration in utilities and their insulation (Szyrowski, 2016). Unlike ground penetrating radar (GPR), which can detect and locate both metallic and non-metallic objects, EMLs can only detect conductive masses, or masses capable of carrying electric current (Chen et al., 1998). However, due to their simplicity, low training requirements, small size and cost, and rapid surveys with minimal data processing needs, EMLs and particularly CATs have become

widely used in the UK (Radiodetection, 2018). A comparison between GPR and EMLs, both of which are commonly used in subsurface exploration, is presented in Table 2-8. In this research, the term "conductor" refers to long conductive service utilities such as cables with metallic cores, metallic pipes, and tracing wires installed on non-metallic utilities.

*Table 2-8. Comparison of the relative advantages of EML and GPR*

| Subject | GPR | EML |
|---|---|---|
| **Equipment price** | High | Low |
| **Survey cost** | High | Low |
| **Equipment size** | Medium - Large | Small |
| **Equipment weight** | Light to heavy | Light |
| **Portable** | Yes | Yes |
| **Metallic utilities** | Yes | Yes |
| **Non-metallic utilities** | Yes | Yes*<br>*Only if equipped with tracing wire<br>or<br>*Accessible for running the sonde into it |
| **Condition assessment** | Yes*<br>*Pipe leakage<br>*Pavement condition<br>*Rebars in concrete<br>etc. | Yes*<br>*Some types of cable faults<br>*Coating condition |
| **Depth measurement** | Yes | Yes (under certain conditions) |
| **Penetration depth** | Near-surface to deep | Near-surface |
| **Need expert** | Yes | Yes (needs simple training) |
| **All soil conditions** | No | Yes |
| **Underwater survey** | Yes (conditional) | Yes |

### 2.6.1.1 Time-varying Electromagnetic Fields and Induction of Electric Current

The electromagnetic locating method (EML) is a quasi-static field (low-frequency field) method that uses time-varying electromagnetic fields (EMFs) as a key element and exploits coils of wire to emit and sense those EMFs (Huang et al., 2019). Cable avoidance tools (CATs),

the simplest form of this method, are time-domain EMLs composed of a signal emitting system and search coils (referred to as magnetic sensors). The emitting system actively induces a time-varying EMF into the ground-conductor system. The search coils then detect and measure changes in the intensity of the Earth's magnetic field during the survey. The concept of EMLs can be simply explained using Maxwell's equations, which are discussed in appendix I.

A time-varying electric current is applied to the transmitting coil to create a primary time-varying EMF. This primary EMF is induced into the ground where a buried conductor is located, which generates a time-varying current (alternating current (AC)) that flows along the conductor (Witten, 2014). The flowing AC in turn generates a secondary cylindrical-shaped time-varying EMF around the conductor (known as the signal) that causes changes in the Earth's magnetic field at that location (superposition of EMFs). This secondary time-varying EMF can be detected by a CAT's magnetic sensors as an anomaly in the Earth's magnetic field and used to identify and locate conductors (Wang et al., 2011). The secondary time-varying EMF generated around a conductor can be formed passively or actively from passively or actively generated AC.

Faraday's Law of electromagnetic induction explains the reason for the efficiency of AC signals compared to DC signals for generating passive or active EMFs. DC signals (single direction flow) generate a static magnetic field around the utility, which is difficult to detect, whereas AC signals are an alternating current that generates an oscillating magnetic field, which in turn generates a higher voltage in the conductor (Huray, 2010). This higher voltage results in a stronger signal that is necessary for an effective location survey (Doolittle and Brevik, 2014). The magnetic field strength (B) around a straight metallic conductor can be calculated using

Eq. 2-8, which references Faraday's law and applies Ampere's law (see Table 2-2 in Appendix I) (Lehner, 2008).

$$B = \frac{\mu_0 I}{2\pi r} \quad (2\text{-}8)$$

Where $\mu_0$ is the magnetic permeability of free space ( $\mu_0 = 4\pi E - 7\ [N/A^2]$ ), $I$ is the flowing current in the conductor, and $r$ is the radial distance from the conductor.

### 2.6.1.2 Locating subsurface conductors using EML

Buried conductors can be detected and located using quantified changes in the measured signals after measurements are taken at different locations. The target's characteristics, such as alignment, centre line, and potential depth in the ground, can be determined by the characteristics of the measured signal, i.e. the strength and position of the signal centre where it peaks and is assumed to be directly above the buried conductor (Fischer et al., 2002; Metje et al., 2007). Since the magnitude of a secondary time-varying EMF is proportional to the induced primary time-varying EMF, buried conductors further from the measurement point will be subject to a relatively weak electromagnetic induction and produce a relatively weak response, and vice versa. This concept is illustrated in Figure 2-19, based on Radiodetection (2018), which shows the changes in the strength of the measured signal at different locations, where the maximum signal strength is measured above the conductors and zero and weak measurements increase with distance from the centre of the object.

The sources of AC currents in EML methodology are not limited to active sources, i.e. the emitting systems integrated into EMLs, but can also originate from a variety of other active and passive sources.

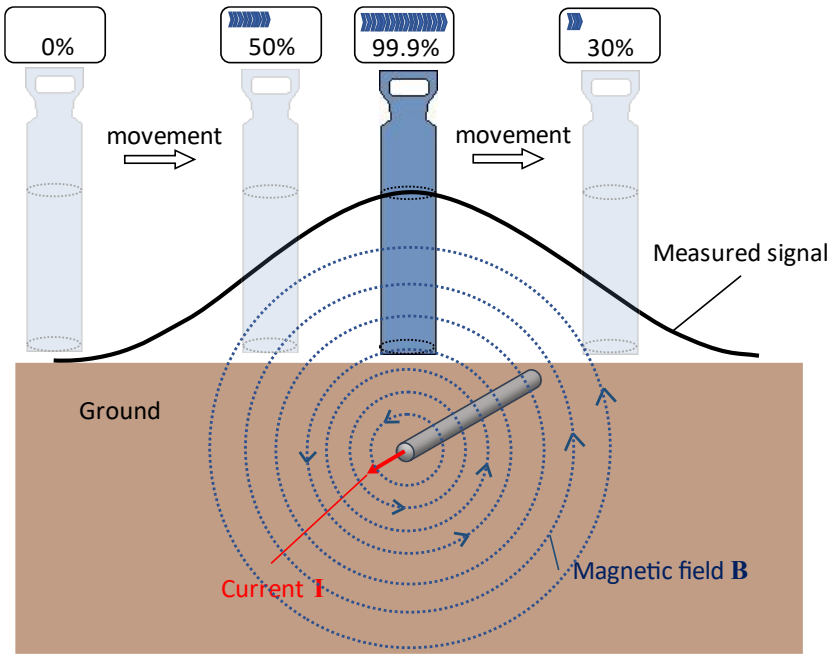



*Figure 2-19. Signal response from a conductor to EMF induction using cable avoidance tool (CAT) based on (After Radiodetection, 2008)*

### 2.6.1.3 Actively generated EMF

As an alternative to using the locator's integrated emitting system, external but compatible transmitters (known as Genny) can be employed to generate current at different dedicated frequencies (chosen based on the survey conditions) (Bryakin and Bochkarev, 2019). The current generated can be applied directly to the conductor using either a direct connection (Figure 2-20a) or induced using an indirect connection (clamping the signal) as shown in Figure 2-16b. Alternatively, the transmitter can act as an induction source by placing the device on the surface above the conductor. Another factor for implementing a successful active methodology (either type) is grounding the conductor (e.g., using ground stakes) and closing the circuit, which allows the current to return to the transmitter and create an EMF around the targeted conductor (Radiodetection, 2018)



*(a)* *(b)*

*Figure 2-20. a) Direct connection to induce AC, b) indirect connection by clamping the AC into the conductor (After Radiodetection ,2008)*

Grounding the conductor in EML applications has its limitations in certain circumstances that need to be addressed. In the case of existing adjacent conductors in the ground, it is expected that the electric current will flow through a lower resistance route to return to the transmitter, which could be provided by the adjacent conductor rather than the ground (soil) itself (VIVAX, 2011). However, the returning electric current in the adjacent conductor creates its own time-varying secondary EMF that can distort the primary EMF around the target conductor, leading to misinterpretation of sensor readings (Siu and Lai, 2019). Figure 2-21 illustrates the distortion and coupling effect of signals caused by an adjacent utility.

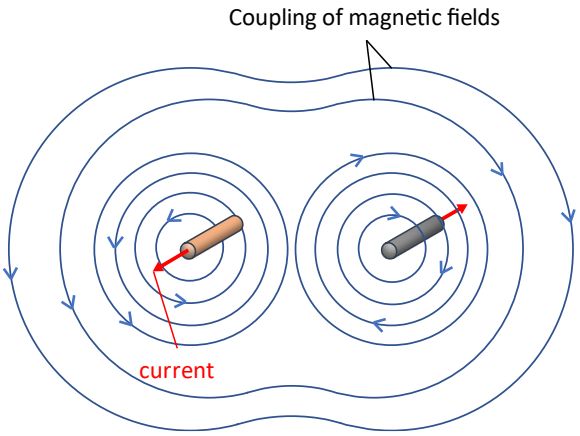



*Figure 2-21. Example of distorted electromagnetic field (coupling effect) caused by an adjacent utility (After Siu and Lai, 2019)*

#### 2.6.1.4 Passively generated EMF

Passively formed EMFs are self-potentials with various external sources without the need for active AC signal generation. As described by SanFilipo (2000) and Radiodetection (2018), passive EMFs are naturally present in a conductor and can be detected by EM-locators in their passive mode without activating the internal emitting system of the device to emit EML's signals. For example, passive EMFs can be formed by the electric current generated by very low-frequency (i.e., 15-27 kHz) electromagnetic radio signals emitted from distant radio transmitters (conductor should be grounded at both ends), or the EMFs produced around a live power cable (power distribution networks) carrying a 50-60 Hz AC signal.

#### 2.6.1.5 Calculation of depth using EML

The potential depth of buried conductors can be estimated using Eq. 2-9 (simplest form) and measurements from the top and bottom search coils of an EML device (e.g., CAT) as shown in Figure 2-22. However, it is important to properly connect the locator to the ground surface, as an incorrect connection will result in an inaccurate depth calculation based on the signal strength received at a particular distance from the ground surface rather than the actual signal strength at the surface level.

$$Depth = D = \frac{d\,A_1}{A_2 - A_1} \qquad (2\text{-}9)$$

Where $A_1$ and $A_2$ are the measurements from the top and bottom coils, respectively, *d* the coil-spacing, and *D* the estimated depth of the target.

The accuracy of a simplified depth estimate using Eq. 2-9 may be affected by external distortion and noise sources such as adjacent utilities or other local obstructions above the ground. However, this issue can be addressed using the triangulation depth estimation

method, which involves taking measurements from multiple locations to determine the depth of the buried conductor more accurately. This method has been found to be effective in overcoming issues with external distortion and noise sources (Radiodetection, 2018; VIVAX, 2011; Waite and Wellaratna, 2006).

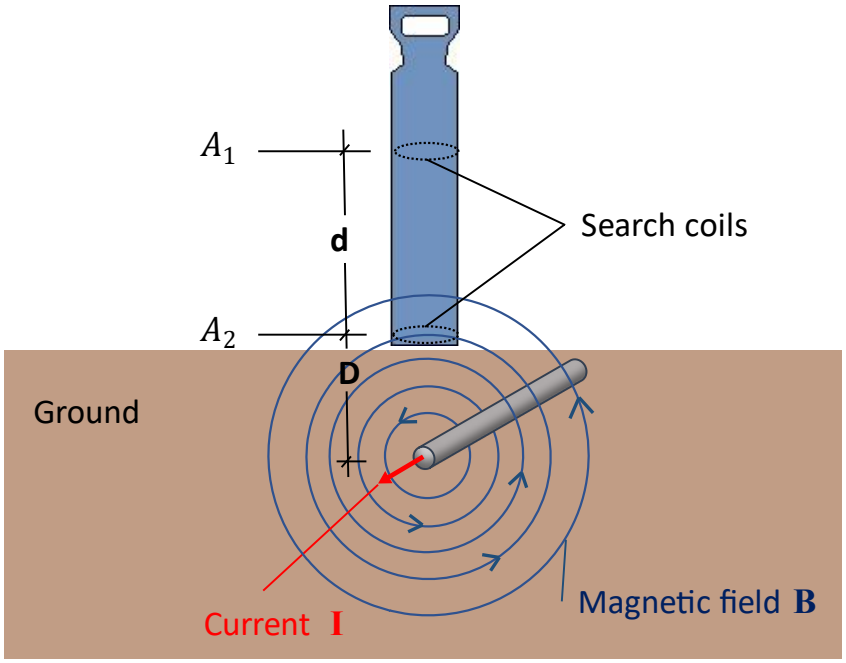



*Figure 2-22. The principle of utility detection using cable avoidance tool (CAT) (After Radiodetection, 2008)*

It is evident that EM surveys, including EML, that are conducted under controlled conditions have fewer unknown parameters, such as the induced signal frequency and strength. As a result, active EML surveys are expected to produce better results in terms of depth bias compared to passive surveys, which have no control over the parameters related to the induced signal and its source (Milsom and Eriksen, 2011).

### 2.6.1.6 EML method Limitations

The success and quality of electromagnetic (EM) surveys depends on various factors, such as soil conductivity, the size and conductivity of the conductor, the frequency and strength of

the time-varying electromagnetic field (EMF) proportional to the AC currents, the grounding, and potential surrounding noise sources (Waite and Wellaratna, 2006). Non-conductive utilities, such as concrete and plastic pipes and fibre-optic cables, cannot be detected using EM unless a tracing wire is installed or a sonde is run through them (Waite and Wellaratna, 2010). In highly conductive ground, such as soil with high water and mineral content (e.g., salt or sulphide ores), the AC in larger conductors decays faster due to scattering and loss of EMF energy (Jeong et al., 2003). The penetration depth of an emitted EMF into the ground depends on the frequency used and the properties of the ground. In general, higher frequencies result in shallower penetration depths due to rapid signal attenuation, while lower frequencies can travel further with a smaller leakage rate to adjacent structures (e.g., soil) (Akagi and Tamura, 2006; Dawalibi, 1986). However, using high frequencies for EM can also cause a coupling effect and make it difficult to distinguish between different conductive targets when they are close together (Siu and Lai, 2019). In the case of active EMF induction, a conductor can be detected if the frequency of the secondary EMF matches the frequency of the induced field.

EML systems, particularly CAT and Genny along with GPR, have become two of the most widely used non-intrusive methods for subsurface exploration and the detection and location of subsurface utilities. However, unlike GPR, which has been actively researched, there are few publications focused on processing EML data. This gap in the literature may be due to the previous limitations of older EML devices in recording data. According to communications with various providers such as SPX-Radiodetection, Leica, and VIVAX Metrotech, only a few of the newer generation CAT devices are capable of recording survey data. For example, the Leica "smart" series of instruments can record data internally, but only for company research purposes. Radiodetection, however, provides instruments such as the RD8200, RD8200G, and RD8100, which are equipped with a data logging system that allows the recorded data,

including signal strength, phase, estimated depth, and position (if equipped with GPS), to be transferred via USB or Bluetooth interfaces for further processing.

In an experimental trial, Siu and Lai (2019) used a Radiodetection RD8100 to investigate the coupling effect of electromagnetic fields (EMFs) in adjacent utilities (a 50 mm galvanised iron pipe and a 5A (amper) live power cable). However, the equipment's limitations prevented them from extracting the field strength of the adjacent utility from the recorded raw data.

Wang et al. (2010) and Goddard et al. (2012) employed data from a self-designed EM locator consisting of seven measurement coils to detect buried power cables. However, like Siu and Lai (2019), they reported that their conductive approach was limited to areas containing only one power cable. Goddard et al. (2012) attributed this issue to the static measurements (1.54 s for each) and suggested developing an analytical methodology to conduct continuous measurements to overcome this issue.

In another study, Bryakin and Bochkarev (2019) developed a new EML device that uses two harmonic transverse ferrite antennas to locate buried utilities and identify damaged areas. According to the authors, their design reduces inter-turn leakage and mitigates external interference on emitting coils. They did not mention any other drawbacks to their device, although it is likely that there are other issues that have not been fully validated through experimental data.

There are also patents for new devices that implement similar concepts for the detection of metallic and even non-metallic utility lines registered by Fischer et al. (2002) and Waite and Wellaratna (2006), respectively. However, a search of the literature did not reveal any practical evidence of these ideas beyond the patent registration. To the best knowledge of the author, except for Siu and Lai (2019), who attempted to process data from commercially

available devices, there are no other published studies that have specifically logged and processed CAT data.

Sudevan et al. (2019) recently used an EML device to locate buried oil and gas pipelines autonomously. They developed a robotic inspection system and a hierarchical tracking algorithm that used an EM pipe locator and sonar sensor to track a buried metallic wire, with an injected 50 mA current at 640 Hz, below an artificial wooden berm to simulate a pipeline. The main focus of their research was on tracking the route and preventing the robot from sliding off the artificial berm. They did not provide any information on how the data was sampled or processed.

As previously mentioned, EML, especially CAT, is widely used in the industry for locating service utilities and can be found on almost every construction site. Despite its limitations, EML can still provide valuable information about the subsurface, particularly about metallic service utilities, which can be used in conjunction with data from other sensing methods such as GPR. Based on the literature review (section 2.4.5), the available NDT data fusion methods are typically analysed one after the other or in parallel and are interpreted by experts rather than through a formal, objective process. This indicates that there is still a potential gap in the area of fused data. In particular, there is a lack of fusion algorithms in the literature that objectively process, and fuse data logged from EML with other geophysical techniques such as GPR.

### 2.6.2 Ground Conductivity Measurement (Train Conductivity)

Electromagnetic conductivity (EMC), also known as EM induction, EM survey, and train conductivity, is a commonly used active electromagnetic induction technique for measuring the conductivity of subsurface materials. It works by inducing eddy currents into the ground.

It's worth noting that conductivity is the reciprocal of resistivity. In the 1960s, this technique was developed primarily for the exploration of natural resources such as oil and minerals. However, it is now also used for detecting conductive materials and assets, as well as investigating ground lithography. Despite its low resolution, it has proved to be a valuable non-invasive tool for identifying deep underground resources (Cagniard, 1953; Christensen and Tølbøll, 2009; Madden and Swift, 1965; Milsom and Eriksen, 2011).

The EMC apparatus, also known as a ground conductivity meter, consists of a controller and two separate wire loops (coils) serving as the transmitter and receiver. These coils take advantage of the correlation between primary and secondary magnetic fields. When an alternating electrical current is applied to the transmitter coil, it produces a magnetic dipole field that is emitted into the ground-conductor system. This causes electrical currents to flow in shallow depth conductors, creating a secondary magnetic dipole field that can be measured by the receiver coil (Milsom and Eriksen, 2011). The coil spacing of the EMC apparatus can be either fixed, as with the EM31, EM38-MK2, DUALEM, and CMD explorer, or user-specified using loose coils, as with the EM34. The penetration depth of the EMC is affected by several factors: the orientation of the coils (horizontal or vertical), the coil spacing (larger spacing allows for deeper penetration), and the frequency of the induced magnetic field (higher frequencies are more easily attenuated). Ground conductivity meters typically operate at low frequencies (below 15 kHz) and measure the apparent conductivity, which is a composite of the ground conductivity responses to the induced horizontal currents in the horizontal ground layers (Beamish, 2011; McNeil, 1980; Milsom and Eriksen, 2011; Petersen, 2001).

Furthermore, as noted by McNeil (1980), these devices must be used under conditions of low induction number (LIN) in order to provide accurate conductivity measures. This is a limitation of using these devices. Operating under the LIN assumption results in a simple linear

relationship between the primary and secondary fields, such that the ratio of the measured secondary magnetic field ($H_s$) to the induced primary field ($H_p$) is linearly proportional to the conductivity ($\sigma$) of the subsurface, as shown in equation 2-10:

$$\frac{H_s}{H_p} = \sigma \quad (2\text{-}10)$$

Where $H_p$ is the primary field, $H_s$ is the secondary field, and $\sigma$ the conductivity.

### 2.6.2.1 Inversion algorithms for EMC in shallow depth investigation

The interpretation of EMC data requires inversion techniques (Madden and Swift, 1965; Ku, 1976; Middleton and Valkenburg, 2002; Peyton, 2015), which can be prone to non-uniqueness issues in the results (Ku, 1976; McNeil, 1980). To simplify the inversion problem, it is common to convert the real three-dimensional (3D) subsurface into one- or two-dimensional problems (Madden and Swift, 1965). When dealing with EM problems, it is common to consider the problem as two-dimensional if the target buried in the subsurface is significantly larger than the skin depth, which is the depth at which the propagated wave's amplitude decreases to 37% of its initial value over a large distance (Ku, 1976; Middleton and Valkenburg, 2002; Peyton, 2015). Sasaki (1989) proposed a 2D joint inversion algorithm that used both magnetotelluric (MT) data (Cagniard, 1953; Fournier, 1966; Lytle and Dines, 1980; Madden and Swift, 1965; Tikhonov, 1950) and dipole-dipole resistivity data to reduce the uncertainty in the 2D interpretation of the data. The algorithm was based on the finite element method (FEM) (Chaskalovic, 2008) and a nonlinear smoothness constraint least-square approach (Lytle and Dines, 1980). The method was only tested on one field dataset (Sasaki, 1989), so it is necessary to apply it to a larger number of datasets from both controlled and real-world surveys to increase confidence in the results (Sasaki, 1989). Additionally, the uncertainty in

the results remains an issue due to the lack of independent data from boreholes or geological data (Sasaki, 1989). Despite these drawbacks, the inversion method proposed by Sasaki (1989) remains a fundamental method that has been continuously updated.

Inspired by Sasaki's approach (Sasaki, 2000, 1989), Monteiro Santos (2004a) presented a one-dimensional (1D) inversion algorithm for constructing structural models of ground conductivity down to 20 meters below the ground surface using EM34 data. The forward modelling used by the author was based on the cumulative response of each layer to calculate the apparent conductivity response (McNeil, 1980). The algorithm used Sasaki's nonlinear smoothness-constrained inversion to calculate block conductivities of the discretised underground layers.

In the previous section, the inversion methods used for inverting electromagnetic conductivity (EMC) data in the literature were presented. Table 2-9 provided below contains a list of reviewed studies that have employed these methods in their research, offering an overview of the various ways in which these methods have been utilised in the literature.

*Table 2-9. List of published work that has utilised inverted EMC data*

| | Author | Application | Employed Device | Inversion method |
|---|---|---|---|---|
| **1** | (Sasaki, 1989) | Ground Resistivity | Radio Magnetotelluric and dipole-dipole resistivity | Smoothness-Constrained least-squares inversion + Full Solution |
| **2** | (Sasaki, 2000) | Ground Resistivity | Vertical magnetic dipole (VMD) | Smooth structure least-squares inversion + Full Solution |
| **3** | (Monteiro Santos, 2004b) | Hydrogeological study | EM34 | Modified 1D inversion with Laterally constrained inversion + Cumulative Functions |

| | | | | |
|---|---|---|---|---|
| **4** | (Viezzoli et al., 2008) | Airborne time-domain EM | SkyTEM | Spatially Constrained inversion + 1D forward modelling |
| **5** | (Christensen and Tølbøll, 2009) | Ground Resistivity | Radio magnetotelluric at four frequencies: 16.8<br>53.0, 126.8 and 183.0 kHz | Lateral parameter correlation inversion |
| **6** | (Santos et al., 2010) | Geoelectrical structures | DUALEM-421 | 1D inversion with Laterally constrained inversion + Cumulative Functions (CF) |
| **7** | (Monteiro Santos et al., 2010) | Geoelectrical structures | DUALEM-421<br>and<br>DUALEM-21 | 1D inversion with Laterally constrained inversion + Full Solution (FS) |
| **8** | (Triantafilis and Santos, 2011) | Hydro stratigraphic analysis | EM34<br>and<br>EM38-MK2 | Joint inversion with spatially constrained inversion + Cumulative Functions |
| **9** | (Saey et al., 2015a) | Resistivity imaging | DUALEM-21S | 1D laterally constrained smooth inversion & non-constrained 1D inversion |
| **10** | (Melo and Régis, 2017) | 2D geoelectrical structures | --- | 1D joint inversion with 1D laterally constrained inversion + Full Solution |
| **11** | (Holtham et al., 2017) | Airborne geophysical surveys | SkyTEM | Smoothness constrained + Methods as developed by UBC-GIF |
| **12** | (Guillemoteau et al., 2017) | Soil sciences, Hydrology, and Archaeology | DUALEM-21S | 1D nonlinear inversion |
| **13** | (Farzamian et al., 2019) | Soil salinity | EM38-MK2 | 1-D laterally constrained inversion algorithm + Full Solution & Cumulative Function |
| **14** | (Zare et al., 2020) | Mapping Soil Properties | DUALEM-421 | EM34soil software |

Despite the advantages that abovementioned inversion methods can offer for inverting EMC data, it is important to keep in mind that all of these methods have their own limitations that should be taken into consideration when selecting a method and interpreting the results (Reynolds, 2011). For example, joint inversion requires multiple data sources and can be

complex to implement and interpret, while smoothness-constrained least-squares inversion is limited in flexibility and may be computationally intensive (Monteiro Santos, 2004a). 1D nonlinear inversion has limited spatial coverage and is sensitive to noise, and smooth structure least-squares inversion may require additional data and may not be suitable for inverting data with complex or non-smooth subsurface structures (Guillemoteau et al., 2017; Sasaki, 1989). Cumulative functions (CF) have limited flexibility and may not be able to accurately represent more complex subsurface structures or conductivity distributions. They also assume that the variations in conductivity are smooth, which may not be the case in reality, and require additional data about lateral variations in conductivity. CF are sensitive to noise in the data (Moghadas, 2019; Moghadas et al., 2009). Laterally constrained inversion has limited lateral resolution and may not be able to accurately represent complex subsurface structures or conductivity distributions. It also assumes that the lateral variations in conductivity are smooth and requires additional data about lateral variations in conductivity (Saey et al., 2015a). Laterally constrained inversion is also sensitive to noise in the data. Both CF and laterally constrained inversion can be improved by using techniques such as smoothness-constrained inversion, but these techniques can be computationally intensive (Farzamian et al., 2019).

While EMC surveys are commonly used for detecting and locating conductive materials and assets in the subsurface, there have been few published studies that have attempted to use EMC data in combination with other remote sensing techniques. There is a lack of a robust algorithm for objectively processing, inverting, and autonomously combining conductivity data with other techniques such as GPR using multi-sensing estimators like the Kalman filter or particle filter approach. This suggests a potential gap in the use of fused data from EMC surveys and other remote sensing methods.

## 2.7 Gaps identified from the literature

Within the scope of geophysical methods for exploring the subsurface, nonintrusive non-destructive methods were systematically reviewed. Especial attention was paid to GPR as one of the most widely accepted and relied-upon techniques in industry as well as research domain. Other techniques such as EMC and EML were also critically reviewed. Furthermore, a comprehensive review of established geophysical techniques such as electrical resistivity tomography, infrared thermography, magnetometry, gravitational, as well as techniques that have focused on mechanical properties of the material (vibroacoustic, and those utilising motion sensors), and finally biologically inspired approaches were carried out and presented in appendix I of this thesis.

Based on the literature review, each of the reviewed methods can be beneficial to be utilised in a survey depending on the governing conditions, e.g. variations in type, size and depth of a buried target, and condition and the composition of the ground. Yet, a critical review of the literature shows that none of these techniques is a single best technique to be utilised under all circumstances due to specific limitations associated with each technique. For example, high noise sensitivity, imperfect methodologies or equipment, issue of penetration depth and non-uniqueness of their results, and many more are some of the limitations that the current methods face in different survey conditions.

Among all nonintrusive non-destructive techniques, the effectiveness of the GPR survey for the detection and location of buried assets, e.g. pipes, under different conditions is repeatedly demonstrated in the literature. Furthermore, a variety of methodologies and algorithms are proposed in the literature to improve accuracy and reliability of GPR surveys in terms of collecting, processing, pattern recognition, and inference of GPR data. To better understand

GPR data interpretation, the literature review continued by investigating both manual and automated noise reducing, decluttering, as well as hyperbola pattern recognition methods of GPR data. It was demonstrated that despite all efforts and the existence of many successful and useful approaches, the inference of GPR data remains a challenging task due to the presence of noise and clutters in acquired data, which is often conducted by experts in a time-consuming procedure.

To tackle the abovementioned limitations, it was demonstrated in the literature that data fusion of GPR with other geophysical techniques can potentially overcome the drawbacks of the currently utilised techniques and improve the survey quality. Thus, a critical review of the data fusion techniques, as well as mathematical estimators that can be used for the data fusion such as Kalman Filter (KF) and particle filter, was caried out. Among which, the superior performance and efficiency of the Kalman filter against similar data fusion methods was demonstrated in the literature. However, the true potential of this powerful estimator, i.e. KF, is still not thoroughly exploited.

It is well established in the literature that the level of uncertainty that occurs in the acquired GPR data, due to the type and level of noise and clutter and complexity of the physical and geometrical properties of the subsurface, puts a challenge on reliable processing and inference of the associated data. However, an optimised KF algorithm for reducing noise and decluttering in GPR radargrams, on one hand, and KF's data fusion capability, on the other hand, provides an opportunity for delivering a cost-effective yet optimal multi-sensing data fusion algorithm. This is a significant gap in knowledge which tackled could remarkably improve the GPR survey practice. The KF algorithm in GPR application for reducing noise and decluttering requires innovative enhancements in different areas. For example, optimisation of KF's input parameters, automating the procedure to minimise expert interaction, and

interpretation of data to identify buried targets such as buried pipes, voids, cavities, and soil disturbances.

As discussed previously, all geophysical techniques suffer from drawbacks, e.g. the subjective bias in the inversion of the EMC data, or have not received the due attention, for example there is close-to-none systematic data acquisition and processing for EML data. However, despite all the associated limitations, these techniques can be utilised as complementary information to be fused with other NDT techniques. Furthermore, there are new-generation EML devices equipped with data logging systems, e.g. radiodetection RD8200 CAT, that provide new data acquisition opportunities.

Data of EML, as a neglected technique in the literature, can be innovatively processed and utilised as a complementary sensor in a multi-sensing survey, e.g. GPR + EML, and add useful information about conductive utilities, such as metallic pipes and cables.

To conclude on objective 1 of this thesis, below are the major conclusions from the critical review carried out in this thesis on nonintrusive subsurface exploration techniques, focused on electromagnetic methods especially GPR:

- Although the variety of existing nonintrusive NDT methods opens many doors for the exploration of the shallow subsurface in different ways, performing a reliable survey for each of these methods is subject to various factors. For example, survey setup, boundary conditions, ground composition, physical properties, available information such as geological boreholes data, the judgement and interpretation of the expert, type, size and depth of buried assets, and type and level of noise. This has led to development of numerous approaches proposed in the literature aiming to enhance

these methods in terms of practicality, pre-and post-processing of data, performance, and efficiency.

- The electromagnetic based methods are the most commonly used methods for exploring the subsurface in terms of detection and locating buried utilities, yet the extraction of meaningful information and inference of associated data remains a challenging task, due to the complexity of subsurface space and the existing noise and clutters in sampled responses, and often relies on expert's knowledge and expertise.

- Electromagnetic methods are based on the propagation of electromagnetic waves in mediums such as soil. The propagated waves (signals) carry meaningful information about the subsurface that can be extracted from different components of signals, i.e. frequency, amplitude, and phase of signals using different analytical approaches. The attenuation of propagated signals due to physical properties of the buried target-ground system, and susceptibility to noise due to the lack of control over sources of noise in and above the subsurface are major concerns about electromagnetic methods.

- The corresponding data associated with GPR, as the most widely used electromagnetic method, in comparison to other electromagnetic methods is often very large (e.g. several gigabytes) that requires time-consuming pre-and-postprocessing analysis. In addition to such high computational costs, vital information can be suppressed in GPR data by noise and clutter or eliminated during the pre-and post-processing of data.

- Noise and clutter signals can simply suppress the weaker signals in GPR radargrams that carry useful information about the subsurface. This can mislead the expert and cause inaccuracy in the interpretation of a survey data. Signal processing is one of

many approaches proposed in the literature for processing and extracting such useful information from sampled GPR signals.

- All geophysical non-destructive techniques for subsurface exploration have their limitations. Therefore, none of these techniques can be nominated as the best stand-alone technique for all scenarios in subsurface exploration. However, depending on the governing physics and boundary conditions in a survey, these techniques can be used in parallel as complementary techniques to improve the outcome of a survey. Data fusion of multiple sensors can offer a remedy to the limitations associated with these techniques and improve the survey practice.

There are three further gaps which are also identified based on the review of the literature on other non-destructive techniques which are discussed in appendix I. The following chapter aims to develop methodologies to address the above gaps in knowledge and propose a series of computational solutions or improvements concerning each gap.

# Chapter 3 METHODOLOGY

## 3.1 Introduction

There are several geophysical nonintrusive techniques used for exploring and investigating underground spaces. These techniques rely on different physical properties of the materials in the ground and buried targets, such as electrical properties (e.g. dielectric permittivity and conductivity), magnetic properties (e.g. magnetic permeability), mechanical properties (e.g. strength and elasticity), thermal properties (e.g. thermal expansion), and scents, to detect and locate buried targets and anomalies in the ground. However, a critical review of the literature in chapter 2 of this thesis showed that none of these techniques can be considered the best single option for subsurface exploration due to various shortcomings associated with each of them. Within available techniques for subsurface exploration, GPR is one of the most widely used due to its ability to provide detailed, non-invasive information about the subsurface.

The analysis and interpretation of GPR data is a challenging task and an active area of research, as demonstrated in the literature review chapter of this thesis. Therefore, there have been numerous techniques and methods proposed by researchers to address the issue of processing and interpreting GPR data. Among these proposed methods, the Kalman filter (KF) is one approach that has been utilised for reducing noise and decluttering GPR data. The KF in GPR applications, as described in the previous chapter, needs improvements in various areas, such as determining optimal values for input parameters and using KF's internal parameters as benchmarks for detecting buried targets (objective 2). In addition, the potential of KF as a multi-sensing tool capable of fusing data from multiple sensors has not been fully explored (objective 2).

To achieve the objectives of this thesis, which include developing novel approaches for processing GPR data in both semi-autonomous and autonomous ways and identifying buried targets and ground anomalies using signal processing techniques and machine learning algorithms (objective 4, and 5), new optimised Kalman filter (KF) algorithms were needed. To understand the correlation between scattered GPR signals associated with buried targets and the parameters produced by the KF, a numerical model of a buried target-ground system was developed, and data was analysed. Based on a thorough understanding and characterisation of the relationship between GPR signals and data produced by the Kalman filter (KF), both semi-autonomous and autonomous algorithms were developed.

The literature review also showed that EM locators, particularly cable avoidance tools (CAT), are the most widely used technique in practical applications for detecting and locating buried conductive utilities such as metallic pipes and cables. However, these devices have not been extensively studied due to the limitations of older generation devices in terms of data sampling and storage, or if data is stored, the inability of users to access it. Additionally, it was shown that new-generation EM locators, capable of logging data, can be used not only as a standalone technique for detecting buried conductive targets, but also as a complementary technique, where EM locator data can be fused with data from other nonintrusive methods. Based on these findings, the developed autonomous algorithm was used to contribute to the development of a novel algorithm based on KF for fusing processed GPR data with inverted data from other complementary nonintrusive sensors, specifically EML and electromagnetic conductivity (EMC) (objective 6).

Finally, conducting EM locator (EML) surveys can be challenging in more complex scenarios with multiple conductive utilities, such as cables laid next to each other, which can cause coupling effects due to induced magnetic fields. To address this issue, a fusion-inversion

algorithm was developed that uses finite element modelling and optimisation techniques (objective 7).

This chapter provides a detailed description of the computational solutions and methodologies used in the research, including the processing and analysis of data from GPR, EML, and EMC, and the conclusions drawn from this methodology.

## 3.2 The journey of this research project

To investigate the identified gaps in knowledge through the literature review and to define the precise scope of this research, the researcher has taken different paths including experimental tests and numerical modelling. An overview of this journey which resulted in the current computational solutions of this research, outlined below, is provided in Appendix II.

This research aimed to develop a number of novel, robust autonomous computational solutions using a combination of classic and artificial intelligence (AI) based processing algorithms to process ground penetrating radar (GPR) data as well as effectively fusing complementary data from non-destructive geophysical techniques, namely GPR, EML, and EMC, to improve detecting and locating of shallow buried utilities (e.g. pipes and cables) and ground anomalies (e.g. ground disturbance). To support fulfilling this aim, the Kalman Filter (KF) was selected as a powerful stochastic based estimator for the development of robust computational solutions in this research, which will be discussed in detail in the following sections.

### 3.2.1 Why Kalman Filter?

As discussed in section 2.5 of the literature review, KF is a powerful, robust, and dynamic modelling technique that recursively updates an initial estimate in a stochastic process based on a sequence of measurements (observations). It only requires measurements from one previous step (timeframe) to update the estimate, making it fast and low in computational cost in terms of processing time and memory usage (Simon, 2006). Because of this, the KF is suitable for both offline analysis of estimation and control system problems and real-time analysis of various applications, such as self-driving cars, auto-pilots in flight, and radar systems (Grewal and Andrews, 2008; Särkkä, 2013; Simon, 2006).

In addition, reviewing the literature (sections 2.5.1 and 2.5.2) also revealed that the KF in GPR applications is capable of outperforming other approaches in terms of denoising and decluttering GPR radargrams. Furthermore, KF is not limited to data provided by one sensor, but can also be used for estimating system states using parallel or sequentially provided data (measurements) from multiple sensors (statistically independent sensors) (Willner et al., 2008). It is worth mentioning that particle filtering, which is another stochastic approach based on the KF that can be utilised to solve non-linear non-gaussian problems, is capable of delivering smoother results compared to the KF but at a significantly higher computational cost (processing time and memory usage) due to very complex calculations (Ko and Kim, 2012; Ng et al., 2008). On this basis, the KF remained the best approach for this thesis for the GPR application for reducing noise, decluttering, and fusion of multiple sensor data for the detection and locating buried pipes and anomalies in the surrounding soil.

## 3.3 Semi-autonomous algorithm for GPR data processing using Kalman filter

This section outlines the development of a semi-autonomous algorithm for processing raw GPR data. It begins by explaining the structure of the Discrete Kalman Filter (DKF) as applied to GPR data. Next, the integration of pre-processing filters is discussed. The section then covers numerical modelling and preliminary analysis of a synthetic model and real GPR survey data. Finally, the process of detecting targets (or anomalies) and the optimisation approach using machine learning are described.

### 3.3.1 Overview of the semi-autonomous algorithm

The proposed semi-autonomous GPR processing algorithm has three main steps, which are illustrated in a flowchart in Figure 3-1.

First, the above-mentioned signal processing filters (i.e. time zero correction, gain adjustment, DC shift, horizontal banding removal, and band pass filters) are integrated into the KF algorithm to improve the data quality. Next, a multi-objective optimisation algorithm using evolutionary computing is adopted and applied to the problem. This method was chosen over conventional optimisation strategies due to the complexity of the problem, and because it has the ability to generate Pareto solutions (PSs), which are solutions associated with different cost functions of the problem (i.e. mean of innovations and the area under the ROC curves), as discussed in section 3.4.1. In the context of target detection, receiver operating characteristic (ROC) curves can be used to represent the trade-off between the probability of true detection and false alarms. Finally, new Kalman Filter algorithms are developed to analyse

GPR radargrams (B-scans), differentiate noise from target and anomaly signals, reduce noise and clutter, and generate improved quality radargrams for final visual inspection.

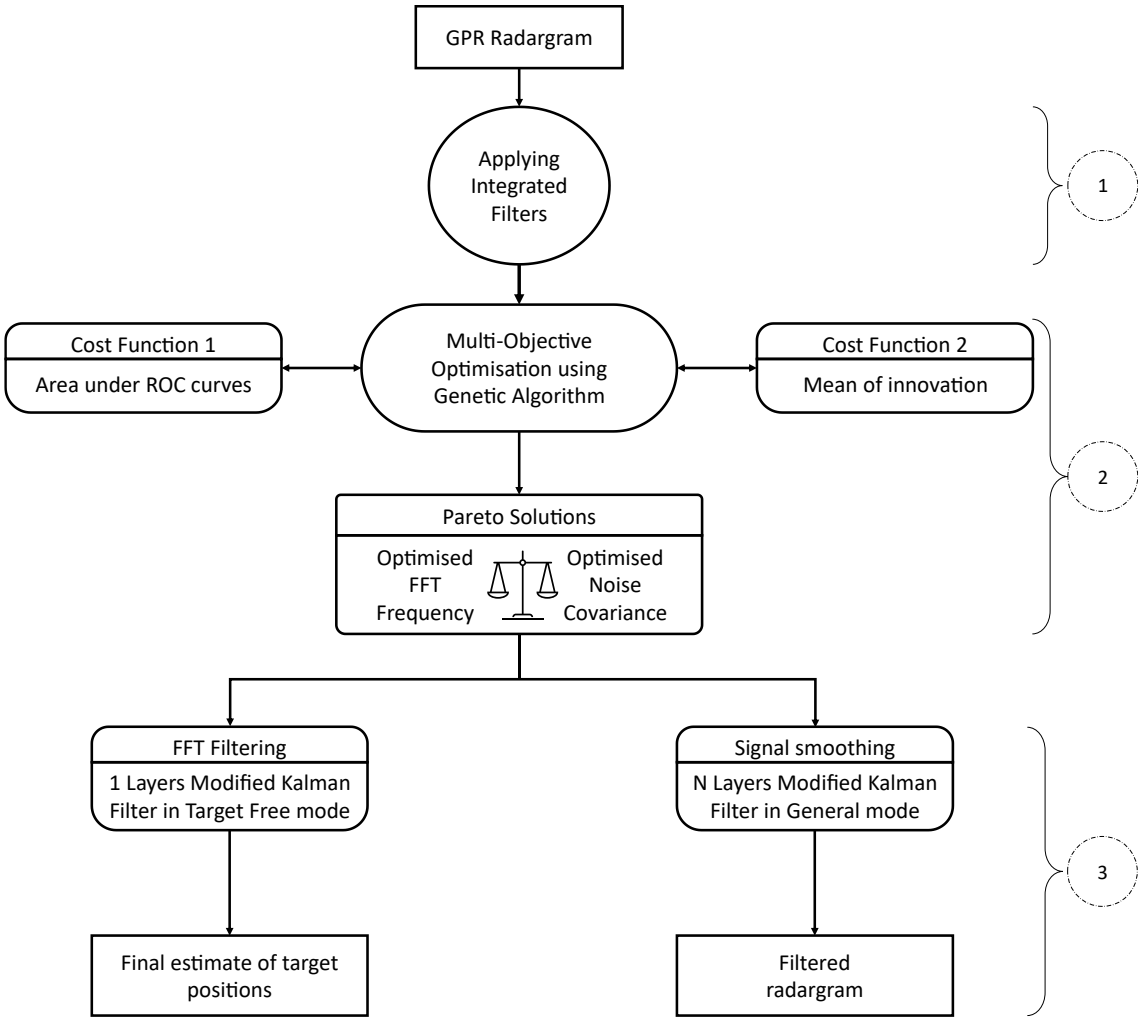



*Figure 3-1. Flowchart of the semi-autonomous algorithm for processing GPR data*

To implement the proposed semi-autonomous GPR processing algorithm, a new version of the Linear Kalman Filter algorithm is developed for GPR applications. The details of this modified algorithm are discussed in the following sections of the chapter.

### 3.3.2 Kalman Filter in GPR applications

As discussed in the previous chapter (section 2.5), the KF is statistically an optimal state estimator for the analysis of linear models with Gaussian probability distributions (Särkkä, 2013). However, as noted by Simon (2006), real-world systems often have nonlinearities, resulting in nonlinear probabilistic state-space models. To address this, it is common to approximate nonlinear systems as linear ones and use linear system theory for estimation and control purposes (Simon, 2006). In the context of GPR applications, it is typical to assume that the ground's EM properties, and therefore the background signal, change slowly with increasing depth (Carevic, 1999; Zoubir et al., 1999). Under this assumption, the probabilistic state-space model of a system can be considered a linear model with Gaussian probability distributions, allowing for the implementation of a standard KF, known as the Linear Discrete Kalman Filter (LDKF).

If the nonlinearity of a system cannot be ignored, the extended Kalman Filter (EKF) can be used to address it. The EKF introduces a local linearisation step through the use of Taylor approximations to linearise the model at each time step and create a new linearisation point for the next KF estimate (Jentzen and Kloeden, 2011). This research project assumes that the EM properties of the ground, and therefore the background signal, change slowly with respect to the spatial position. In other words, the ground is treated as a linear medium for which linear calculations are applicable.

The design of a KF algorithm depends on the specific problem at hand (Simon, 2006). There are many different types of KFs that have been proposed in the literature for different tasks and problems. The KF proposed in this research was designed based on the assumed linearity of the ground, the presence of a sampled-data system, and the discontinuity of GPR impulse

signals. These considerations led to the decision to implement a random walk model and develop a LDKF, discussed by Hartikainen and Särkkä (2011) and Simon (2006). In this research, a new LDKF was developed based on the work of Carevic (1999) for use in the automation of GPR data processing, and later for the fusion of GPR data with other geophysical methods such as EML and EMC. The LDKF was chosen as the appropriate model and algorithm for these tasks due to the assumptions of linearity of the ground, the presence of a sampled-data system, and the discontinuity of GPR impulse signals.

#### 3.3.2.1 Formulation of the estimation problem

As mentioned in section 2.3.2, noise and clutter can obscure important information about buried anomalies and make radargrams difficult to interpret (Daniels, 2004). Therefore, generating an accurate estimate of the scattered GPR signal using a mathematical model, known as the state-space model of GPR systems, could be an ideal approach to identifying and distinguishing between different types of signals and overcoming the shortcomings of data interpretation (Van Drongelen, 2010).

#### 3.3.2.2 Probabilistic discrete-time state-space model of GPR system

The probabilistic state-space model of the GPR system is presented below in the form of a series of mathematical equations (Eq.3-1 to Eq.3-10). For a detailed description of the process and the derivation of the equations, please refer to Carevic, (1999). The mathematical description of the basic model for sampled scattered GPR signals, known as the measurement equation $y(n,k)$, can be depicted as the summation of the signal components (Eq.3-1), i.e. the background signal $s^b(n,k)$, target signal $s^t(n,k)$, and the associated noise $\omega(n,k)$

$$y(n,k) = s^t(n,k) + s^b(n,k) + \omega(n,k) \quad (3\text{-}1)$$

where indexes $n = \{1, \dots, N\}$ and $k = \{1, 2, 3, \dots\}$, as illustrated in Figure 3-2, refer to the sample points on the vertical axis and distance from the origin (trace numbers) on the horizontal axis of a radargram, respectively. It is worth mentioning, that the vertical axis of a radargram in typical processing software indicates the two-way travel times of the emitted signal into the ground and is typically in nanoseconds [ns].

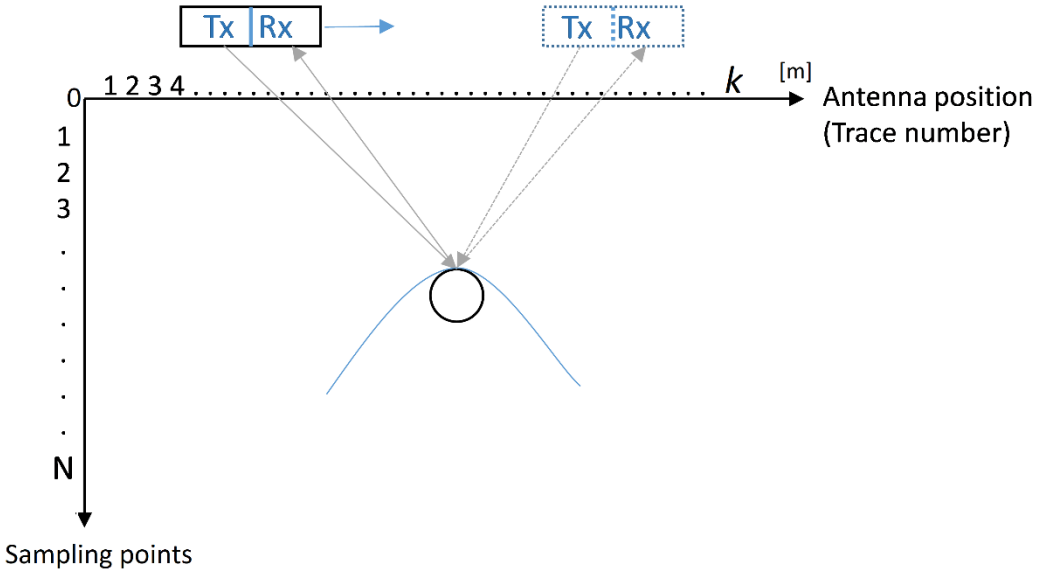



*Figure 3-2. GPR survey model*

The vector of measurement equation $u(n,k)$ (Eq.1) in its matrix form for scattered signals for $N$ sampling points can be defined as $(N \times 1)$ matrices as shown in Eq.3-2.

$$\begin{pmatrix} y(0,k) \\ y(1,k) \\ \vdots \\ y(N,k) \end{pmatrix} = \begin{pmatrix} s^t(0,k) \\ s^t(1,k) \\ \vdots \\ s^t(N,k) \end{pmatrix} + \begin{pmatrix} s^b(0,k) \\ s^b(1,k) \\ \vdots \\ s^b(N,k) \end{pmatrix} + \begin{pmatrix} \omega(0,k) \\ \omega(1,k) \\ \vdots \\ \omega(N,k) \end{pmatrix} \tag{3-2}$$

Figure 3-3 shows an example of the signal components (target and background signals) associated with a selected signal (trace). Typically, radargrams in a perfect world can be divided into two main regions: i) Target-free regions consisting of the ground itself (generally the host medium) with a constant distributed characteristic over the entire surveyed area with

associated ground signals (background signal); ii) Target present regions where targets are buried, and target signals are usually detectable within the background signal.

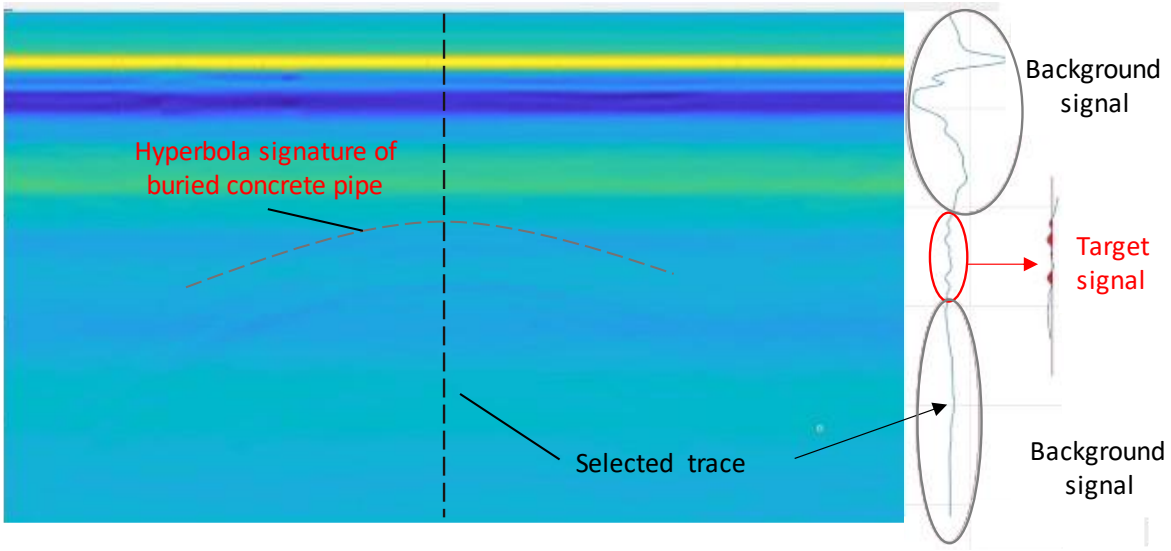



*Figure 3-3. Sample radargram from an experiment at NBIF at university of Birmingham consisting of a concrete pipe (d=30 cm) buried 30 cm deep in compacted sand*

### 3.3.2.3 Overview of measurement equations

Considering the backscattered signal components $s^t$, $s^b$ and $\omega$, to distinguish and separate the noise and the background signal from the target signals, the KF algorithm can be operated under two general state assumptions, discussed below. Switching between both state assumptions offers the possibility of estimating the target and ground signals based on the results associated with the state assumption under which KF is operating.

1) The *target-free state-space model*, in which the return signal only consists of scattered ground signals (background signal). The result of the operation of the KF under the target free state assumption is the KF's estimate of the background signal as shown in Eq.3-3:

$$x = s^b(k) \tag{3-3}$$

Accordingly, the sensor measurement equations $u(k)$ and the target free state model $s^b(k)$ in a random walk model can be defined using linear equations as shown in Eq.3-4 and Eq.3-5.

The background signal $s^b(k)$ at each timeframe is considered to have a constant characteristic over the surveyed area (as discussed above) and is equal to its previous timeframe plus some additional noise.

$$y(k) = s^b(k) + \omega_1(k) \quad (3\text{-}4)$$

$$s^b(k) = s^b(k-1) + \upsilon_1(k) \quad (3\text{-}5)$$

Where $\upsilon_1$ and $\omega_1$ are target free state model noise and measurement noise, respectively. Noise in signal processing is described as a gaussian distributed random variable with a zero-mean covariance. The zero mean characteristic guarantees the contribution of all possible noise frequencies in the analysis (Van Drongelen, 2010). The noise is to be assumed as a zero-mean Gaussian white noise with variances of $\sigma^2_{\upsilon_1}$ and $\sigma^2_{\omega_1}$ , respectively.

2) The *target-present state-space model*, in which the scattered signal composition consists of all signals namely, the ground, target, and noise signals, as shown in Eq.3-6:

$$x = [\ (s^b(k))^T, \quad (s^t(k))^T, \quad (b(k))^T\ ]^T \quad (3\text{-}6)$$

Where $b(k)$ is a random bias that is added to the system to deploy the coloured noises and realise a more realistic scenario. Whereas white noise represents the ideal state of a system and can ease the procedure, the coloured noise is physically more realistic. In other words, systems driven by coloured noise will produce more unusual characteristics, which represent better with reality (typical in real-world applications) (Zhang et al., 2009). If the measurement noise at each time frame is a function of its previous value, then noise should be considered as coloured noise (Simon, 2006). The issue of implementing coloured noise for KF in GPR application is addressed using added bias (coloured noise) to the system as shown in Eq.3-6. The quality of the ground and target signal estimates produced by the KF can be improved by

estimating the system noise as a separate variable consisting of a combination of coloured and white noise.

The sensor measurement equation $u(k)$ in the target-present state can, therefore, be defined as represented in Eq.3-7 to Eq.3-10. It has also been assumed that the ground in the target presence state has a very small and negligible alteration.

$$y(k) = s^b(k) + s^t(k) + \omega_2(k) \quad (3\text{-}7)$$

$$s^b(k) = s^b(k-1) \quad (3\text{-}8)$$

$$s^t(k) = s^t(k-1) + b(k) \quad (3\text{-}9)$$

$$b(k) = b(k-1) + \upsilon_2(k) \quad (3\text{-}10)$$

Where $\upsilon_2$ and $\omega_2$ are the model and measured noises respectively, assuming the zero-mean Gaussian white noise with variances of $\sigma^2_{\upsilon_2}$ and $\sigma^2_{\omega_2}$.

As previously mentioned, the Linear Discrete Kalman Filter (LDKF) was chosen and developed as a suitable variant of the Kalman Filter for use in this research project for processing GPR data. The next section provides a general overview of the LDKF, introducing the parameters, variables, and equations used in this research. Detailed discussions and the derivation of equations are beyond the scope of this research, and interested readers are referred to the literature for further information, such, Bar-Shalom et al., (2001), Grewal and Andrews, (2008), Hartikainen and Särkkä, (2011), and Simon, (2006).

Table 3-1 presents the fundamental variables of LDKF, including estimates and uncertainties (covariances). The superscript notations "-" and "+" indicate "a priori" and "a posteriori" values of the associated variables, respectively.

*Table 3-1. Relationships between Kalman Filter parameters i.e., estimates and uncertainties. (After Simon, 2006)*

| | |
|---|---|
| $\widehat{\boldsymbol{x}}_{\boldsymbol{k}}^{-}$ | a priori estimate |
| $\widehat{\boldsymbol{P}}_{\boldsymbol{k}}^{-}$ | a priori state estimation error covariance |
| $\widehat{\boldsymbol{x}}_{\boldsymbol{k}}^{+}$ | a posteriori estimate |
| $\widehat{\boldsymbol{P}}_{\boldsymbol{k}}^{+}$ | a posteriori estimation error covariance |

### 3.3.3 Overview of the linear discrete Kalman filter algorithm (LDKF)

Once the system model parameters have been determined based on the governing physics of the system, the variables of the system model can be estimated in two steps (1 and 2) as follows:

1) Producing predictions also referred to as *a priori estimates* $\hat{x}_k^-$ of the system state, creating the associated estimate error covariances $\mathbf{P}_k^-$, and Kalman gains $\mathbf{K}_k$ using measurement equations Eq.3-11, Eq.3-12, and Eq.3-13, respectively:

A priori estimate:

$$\hat{x}_{k+1}^- = \mathbf{A}\hat{x}_{k-1}^+ + \mathbf{B}u \tag{3-11}$$

A priori error covariance matrix:

$$\mathbf{P}_k^- = \mathbf{A}\mathbf{P}_{k-1}^+\mathbf{A}^T + \mathbf{Q} \tag{3-12}$$

Where $u$ is system commands (if exists) and variables, $\mathbf{A}$ is the '*dynamic model's transition matrix*', and $\mathbf{B}$, $\mathbf{Q}$, $\mathbf{P}$, are the 'input', '*system noise covariance*', and '*estimation-error covariance*', respectively, in their matrix form.

The importance of proper initialisation of the estimation error matrix **P** was discussed in the technical note in section 2.2.2.1 of the previous chapter. It is worth mentioning that there is published literature that employed an evolutionary algorithm to determine optimised values for the initial estimation error matrix $\mathbf{P_0}$ (Jianguo et al., 2008). However, in this thesis due to a high level of uncertainty associated with the unknown level of noise and clutter in GPR data, diagonal elements of the $\mathbf{P_0}$ matrix can be initialised as large values. This will comply with the theory of the KF (Simon, 2006). Subsequently, the $\mathbf{P_0}$ is initialised to 1E6 and associated results will be presented in the next chapter (results and discussion).

Kalman gain:

$$\mathbf{K}_k = \mathbf{P}_k^- \mathbf{H}_k^T (\mathbf{H}_k \mathbf{P}_k^- \mathbf{H}_k^T + \mathbf{R}_k)^{-1} \quad (3\text{-}13)$$

Where $\mathbf{H}, \text{and } \mathbf{R}$ are 'the *measurement model matrix*', and '*the sensor measurement noise covariance matrix*', respectively.

2) Delivering a posteriori state estimate $\hat{x}_k^+$ by updating the a priori state estimate variable determined in the previous step (Eq.3-14) and updating the state estimate error covariance $\mathbf{P}_k^+$ for the next recursive step (Eq.3-15):

A posteriori (updating estimate)

$$\hat{x}_k^+ = \mathbf{A}\hat{x}_k^- + \mathbf{K}_k(y_k - \mathbf{H}_k \mathbf{A}\hat{x}_k^-) \quad (3\text{-}14)$$

A posteriori estimation error covariance (updating $\mathbf{P}$):

$$\mathbf{P}_k^+ = (\mathbf{I} - \mathbf{K}_k \mathbf{H}_k)\mathbf{P}_k^- (\mathbf{I} - \mathbf{K}_k \mathbf{H}_k) + \mathbf{K}_k \mathbf{R}_k \mathbf{K}_k^T \quad (3\text{-}15)$$

The estimation-error covariance matrix is an important variable that provides information about the expected uncertainty in the procedure, which is also an indicator of the accuracy of the state estimate $(x_k)$ and hence, the performance of the KF. If the value of the covariance

matrix $\mathbf{P}$ at each timeframe is high, then the uncertainty in estimated values through KF will increase and vice versa. As stated by Simon (2006), there are two primary causes for divergence and failure of the Kalman filter. One of these is due to the use of finite precision arithmetic, which involves representing real numbers using a limited number of bits. This can lead to errors in calculations and can cause the filter to diverge. Finite precision arithmetic can be caused by factors such as limitations on memory usage or coding choices, such as using integer data types instead of more precise options like double precision floating point numbers (Simon, 2006). The second cause of KF failure is the modelling errors including defining an inappropriate value for estimation error covariance $\mathbf{P}$ (Simon, 2006).

Generally, the level of knowledge and confidence about system and sensor is a decisive factor for updating the values of $\mathbf{P}$ during the KF procedure. The value of $\mathbf{P}$ can vary in accordance with the variation of system and sensor measurement noise covariances $\mathbf{Q}$ and $\mathbf{R}$, respectively. In other words, whereas accurate knowledge about the system and the sensor can minimise the value of error covariance $(\mathbf{P} \sim 0)$, insufficient or zero knowledge will produce a large value of error covariance $(\mathbf{P} \sim \infty)$ (Simon, 2006). To process the noisy GPR signals in this research, $\mathbf{P}$, which is a diagonal matrix, is initialised by setting the value of diagonal elements to a large number (e.g., 1E6) because of the high degree of uncertainty in the initial state. The input values of each of the matrices $\mathbf{A}, \mathbf{H}, \mathbf{Q}$ and $\mathbf{R}$ specified for each system state are represented in Table 3-2 (Carevic, 1999).

*Table 3-2. Input variables of KF in GPR application*

| Target-free | Target-present |
|---|---|
| $\mathbf{A} = \mathbf{I}_{N\times N}$ | $\mathrm{A} = \begin{pmatrix} \mathrm{I} & 0 & 0 \\ 0 & \mathrm{I} & \mathrm{I} \\ 0 & 0 & \mathrm{I} \end{pmatrix}_{3N\times 3N}$ |
| $\mathbf{H} = \mathbf{I}_{N\times N}$ | $\mathrm{H} = (\mathrm{I} \quad \mathrm{I} \quad 0)_{N\times 3N}$ |
| $\mathbf{Q} = \boldsymbol{\sigma}_{\upsilon_1}^2 \times \mathbf{I}_{N\times N}$ | $\mathrm{Q} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & \sigma_{\upsilon_2}^2 \mathrm{I} \end{pmatrix}_{3N\times 3N}$ |
| $\mathbf{R} = \boldsymbol{\sigma}_{\omega_1}^2 \times \mathbf{I}_{N\times N}$ | $\mathrm{R} = \sigma_{\omega_2}^2 \times \mathrm{I}_{N\times N}$ |

Incorporating and implementing optional and adjustable signal processing filters into the Kalman filter (KF) required modifying Equation 3-14 to include a new parameter and the integration of signal gain, an averaging filter, and signal drift removal. The following section provides a detailed description of the modifications made to the LDKF to process noisy GPR signals with these adjustable signal processing filters (Figure 3-4).

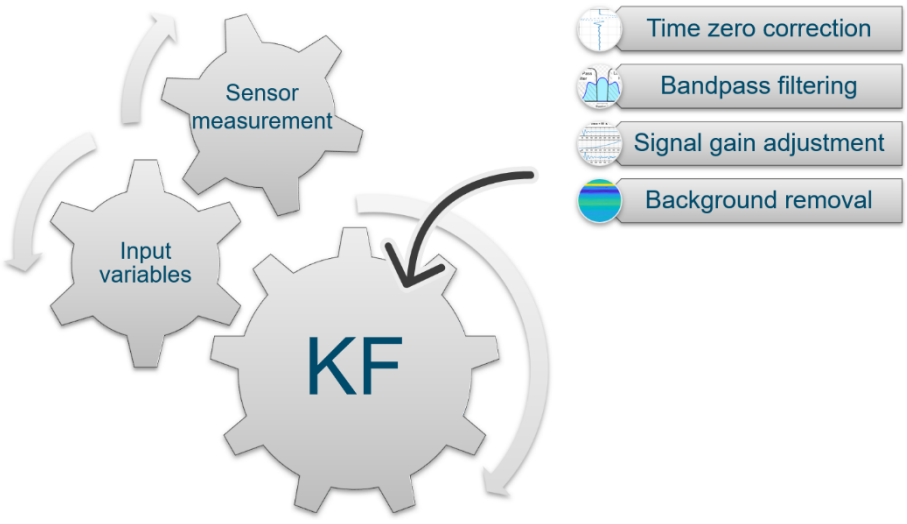



*Figure 3-4. Integration of basic, but fundamental, GPR signal processing filters*

### 3.3.4 Incorporation of basic signal pre-processing filters

As discussed in the previous chapter (section 2.3.2) the acquired raw GPR data can be displayed as high-resolution, often noisy, two-dimensional cross-sectional radargrams. Noise and clutter in GPR radargrams can suppress important information about buried targets and anomalies. Interpreting GPR radargrams is often a challenging process that requires knowledge and expertise (Milsom and Eriksen, 2011). Figure 3-5 and Figure 3-6 show noisy GPR radargrams consisting of a buried tunnel with adjacent buried utilities, and five buried pipes in shallow depths, respectively.

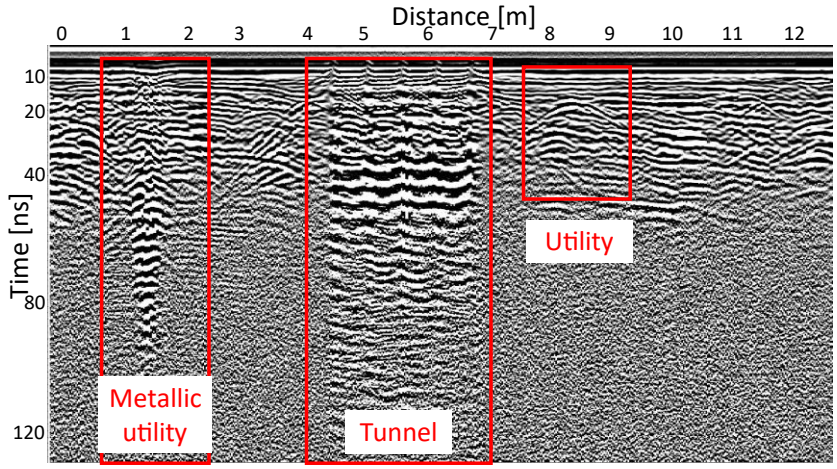



*Figure 3-5. Example of a noisy GPR radargram consisting of a tunnel and adjacent utilities*

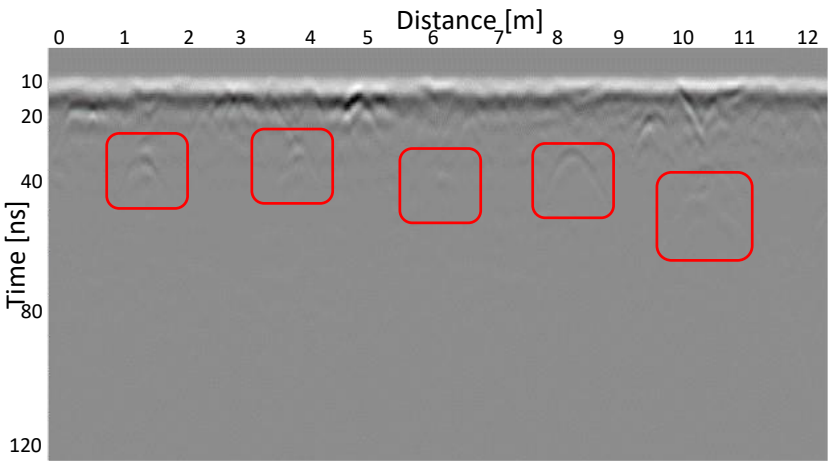



*Figure 3-6. Example of a noisy GPR radargram consisting of five buried pipes in shallow depth*

Radargrams can be used to provide an initial understanding of subsurface structures, but in order to accurately characterise and interpret them, it is necessary to analyse, adjust, and process the collected data to improve its quality and readability (Goodman and Piro, 2013b). Some common processing steps for radargrams include: i) Time zero correction (aligning the first signal reflection event from the ground surface at the zero position to accurately measure depth of targets or anomalies) (Zadhoush, 2021), ii) DC-shift (shifting the deviated signal from zero with increasing depth) (Goodman and Piro, 2013b), iii) Background removal (using an averaging filter to remove horizontal bandings), iv) Bandpass filtering (focusing on a specific frequency range related to the antenna frequency) (Liu et al., 2017), v) Signal gain adjustments (adjusting the signal amplitude to improve visibility of deteriorated signals due to increased penetration depth) (Daniels, 2004), and vi) Wave velocity estimation (determining the accurate depth of targets or anomalies) (Huber and Hans, 2018).

All of these basic but essential processing steps have been the focus of numerous research papers, which have proposed various algorithms to address the challenges involved in developing and implementing effective algorithms (Goodman and Piro, 2013a; Huber and Hans, 2018). In general, these processing steps are used to improve the interpretability of the data and enhance the accuracy of GPR surveys in detecting and locating buried targets and anomalies. There are also advanced processing techniques, such as deconvolution and migration, that can be applied when needed to handle more complex radargrams (Jol, 2009). However, these goals cannot be achieved without using suitable pre-processing methods. Automating GPR data pre-processing is a difficult task due to the uniqueness of the collected data.

On this basis, five of the above-mentioned filters (i - v), as the most common pre-processing filters in use for processing GPR data, were integrated into the developed algorithms in this

thesis as optional and adjustable filters for the expert user to be utilised depending on the conditions and quality of the sampled data. The integration of these filters offered a series of basic solutions to resolve the above challenges toward automating the GPR data processing in terms of reducing the noise and decluttering to emphasising the hyperbola signatures for the ease of interpretation and accurately detecting and locating targets and anomalies in the subsurface.

#### 3.3.4.1 Time zero correction (zero offset removal)

Time zero correction is an important step in the GPR data processing process, as it allows for accurate shallow depth estimates by determining the true two-way travel time of propagating signals. The prediction of time zero position in scattered GPR signals, which corresponds to the true ground surface, has always been a challenge for researchers, who seek to achieve high accuracy and robust error management (Strange and Yelf, 2012; Yelf, 2004; Zadhoush, 2021). However, as thoroughly discussed by Zadhoush (2021), determining an appropriate time zero correction method for numerical processing of GPR data is a complex topic that goes beyond the scope of this study. Therefore, in this thesis, a visual approach using the polarity of response signals and in accordance with conventional GPR data processing software (e.g. Reflexw) is implemented using the MATLAB 'ginput' command [(x, y) = ginput(n)]. This command allows users to identify the coordinates of desired points and satisfies the minimum requirements for predicting the true ground surface (Figure 3-7). It is worth noting that a more advanced approach can be implemented to replace the current time zero correction approach at any time in the future.

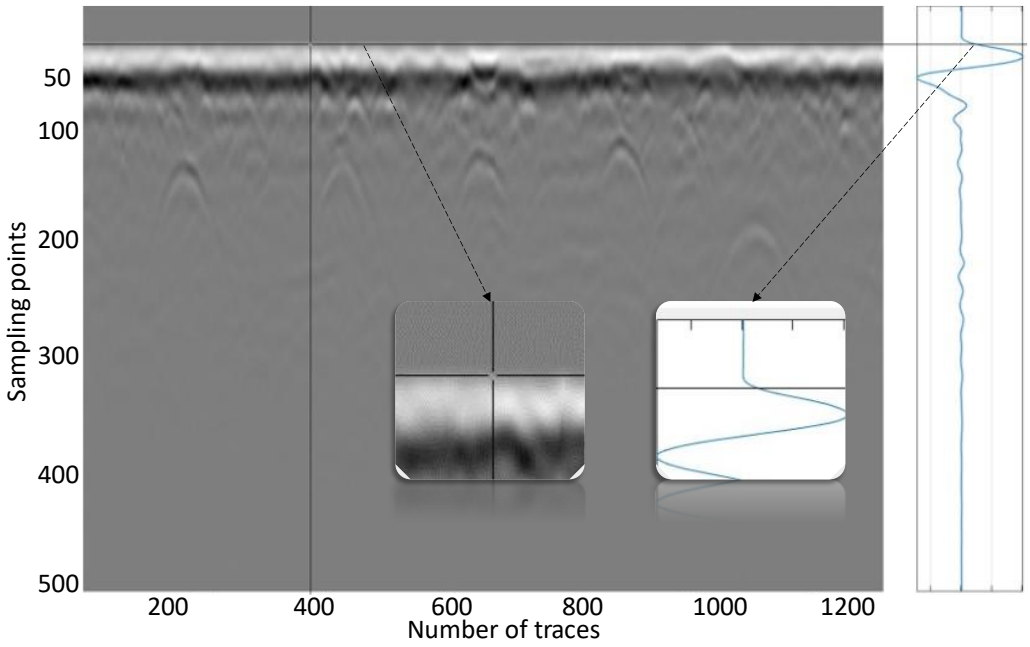



*Figure 3-7. Overview of prediction of time zero position*

After determining the desired time zero position (coordinates), the row containing the selected coordinates and all rows above it will be selected and deleted. The remaining rows will then be shifted up to earlier sampling points, and the first row will be set as the predicted time zero position. Figure 3-8 shows the result of the radargram after applying the time zero correction.

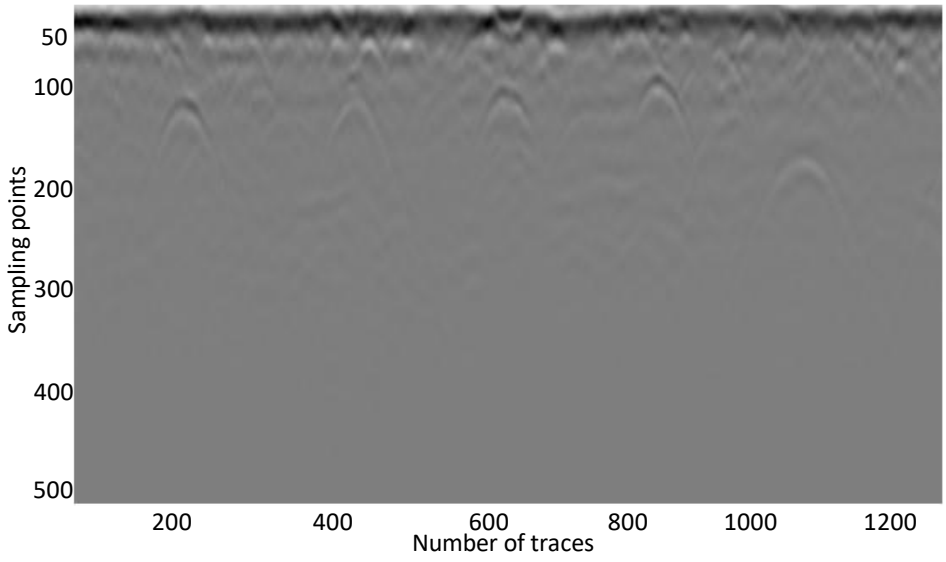



*Figure 3-8. Radargram with a new time zero position after utilising the time zero correction*

### 3.3.4.2 Bandpass filtering

Bandpass filtering is a simple, yet effective, method of noise reduction that can be used to eliminate unwanted signals outside the frequency range of a GPR system (Utsi, 2017b). These unwanted frequencies may be generated by external sources and recorded during a survey. In addition, signal gain adjustment can also lead to amplified noise in a radargram (Utsi, 2017b). The frequency range of different GPR systems can vary between manufacturers, but if it is not provided, a suitable frequency range can be suggested by determining a range around the central frequency (e.g. central frequency ± 50%). For example, a frequency range of 200 to 600 MHz could be considered for data from a GPR antenna with a 400 MHz central frequency. In this research, an adjustable, optional bandpass filter has been integrated into the proposed algorithm that can be used by the expert user depending on the antenna's specifications, boundary conditions, and the quality of the data of interest.

Figure 3-9 shows a radargram before applying the bandpass filter, and Figure 3-10 shows the result of the radargram after the bandpass filter has been applied.

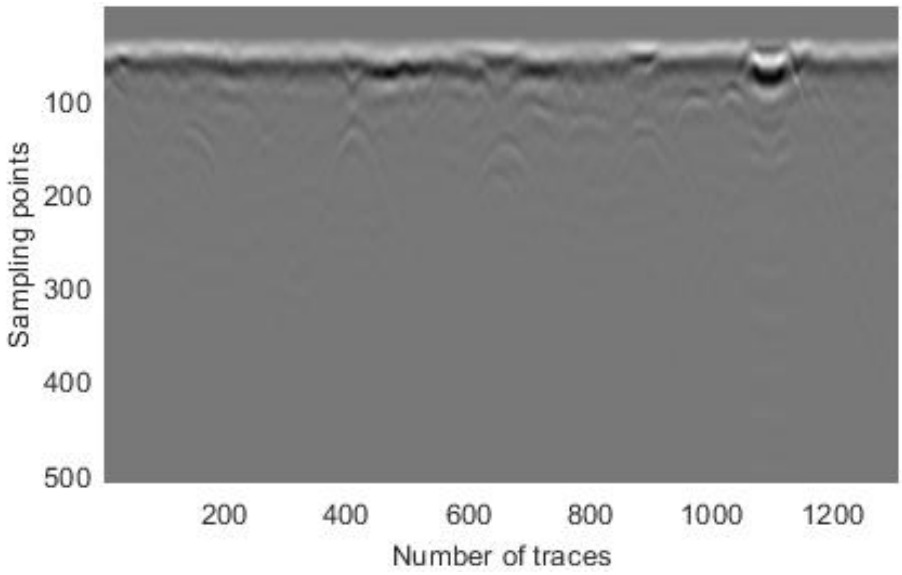



*Figure 3-9. Unprocessed radargram without applying bandpass filter.*

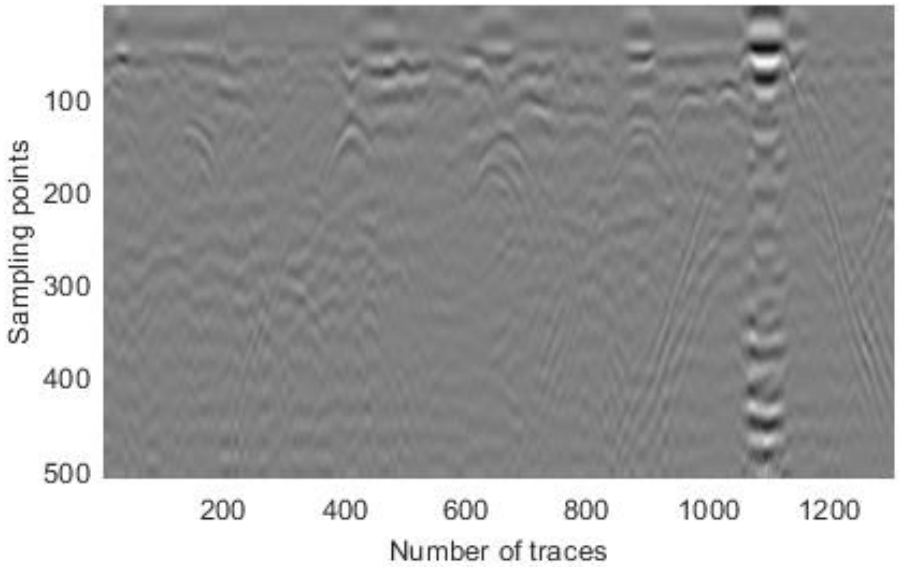



*Figure 3-10. Radargram with applied bandpass filter*

### 3.3.4.3 Background filter (horizontal banding removal)

To retrieve suppressed signals in a radargram caused by horizontal banding (HB), it is necessary to remove these HBs carefully, while avoiding affecting other sought-after signals that have a similar structure (e.g. scattered signals from soil layers or a pipe parallel to the survey direction). Figure 3-11 shows various HBs associated with different soil layers in a radargram containing a concrete pipe buried in sand (compacted in layers). It can be seen that the compacted sand layers have a similar structure to the HBs associated with the antenna and signal saturation above the ground surface. As noted by Goodman and Piro(2013b), a common and simple approach to remove HBs is to use a windowed moving average filter and subtract the calculated average values from the raw data. This approach is also implemented in this thesis.

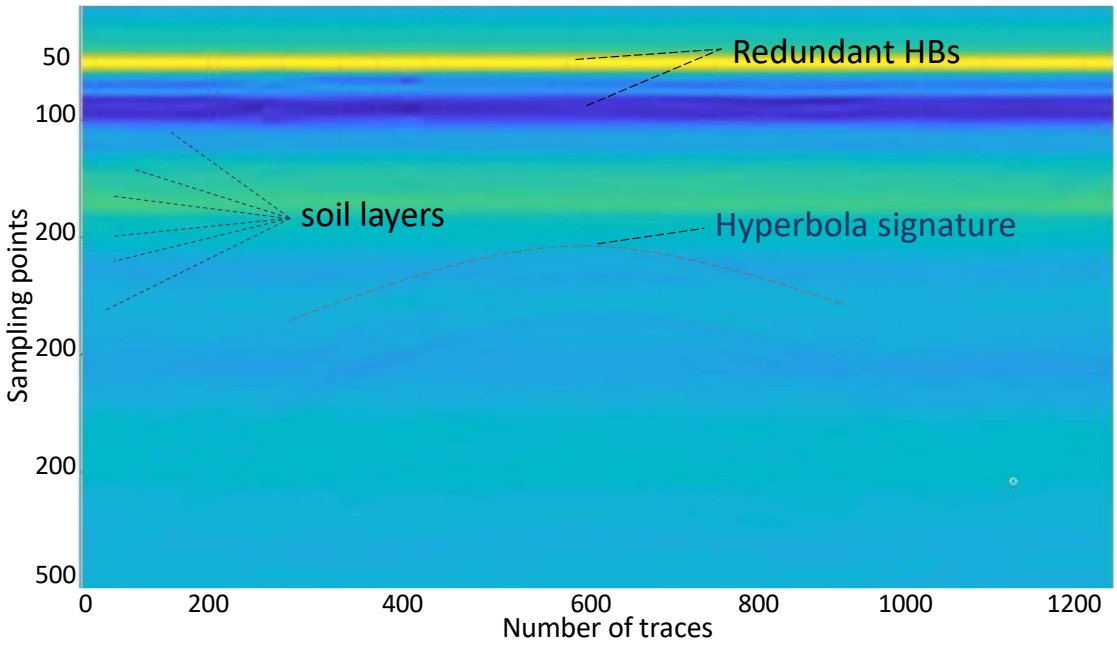



*Figure 3-11. Example of a radargram containing horizontal bandings associated with both informative (soil layers) and redundant HB (antenna itself and signal saturation between the antenna and the ground surface)*

Figure 3-12 shows the result of the radargram in Figure 3-9 after the HB removal filter has been applied.

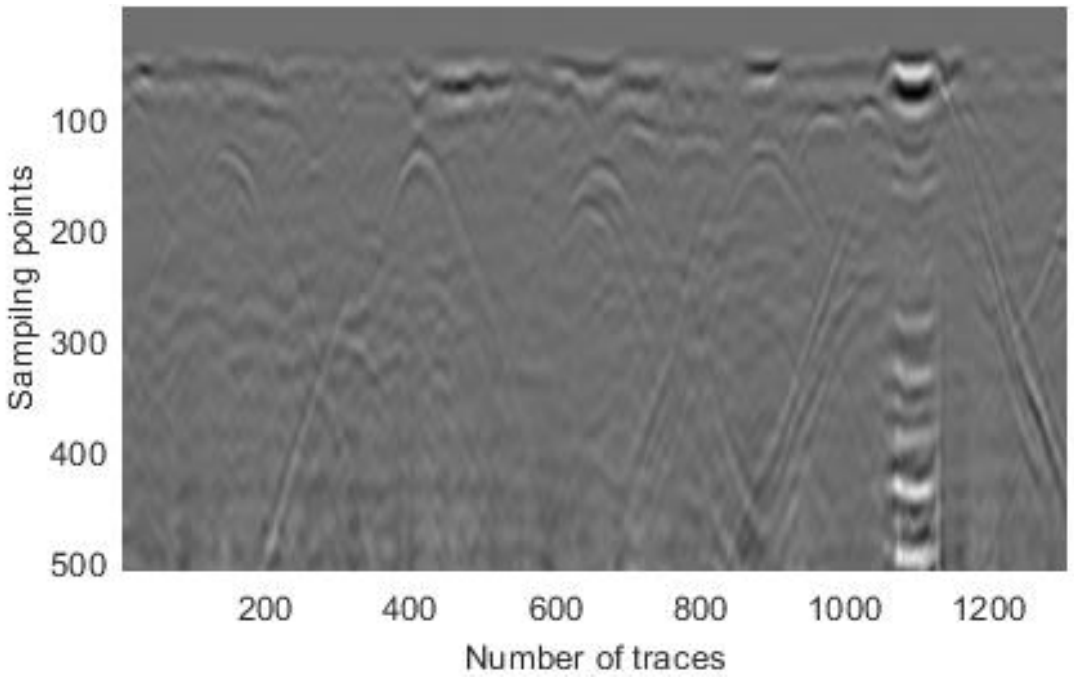



*Figure 3-12. Radargram with applied moving average filter*

#### 3.3.4.4 Signal gain adjustments

The attenuation of the signal amplitude as it travels through the ground can cause a weak reflection of the signal after it encounters a potential target in the subsurface. As previously mentioned, low-energy scattered signals containing information about a potential target can be suppressed by stronger signals with higher amplitudes. The amplitude of the GPR signals can be adjusted (gain) based on the quality of the initial signal by multiplying a time-varying gain function with the initial signal (Daniels, 2004). However, a common issue with gaining GPR signals is a phenomenon called DC drift noise, in which the gained signal starts to deviate from its zero mark (zero axis) as the depth increases. This is due to the amplification of noise during the gaining process (Goodman and Piro, 2013c). For example, as shown in Figure 3-13a, an unprocessed original trace from the radargram of Figure 3-6 does not show any deviation from the zero axis as the depth increases (along 510 sampling points on the X-axis). However, after applying an exponential gain function (Figure 3-13b), the signal starts drifting from the zero axis (Figure 3-13c).

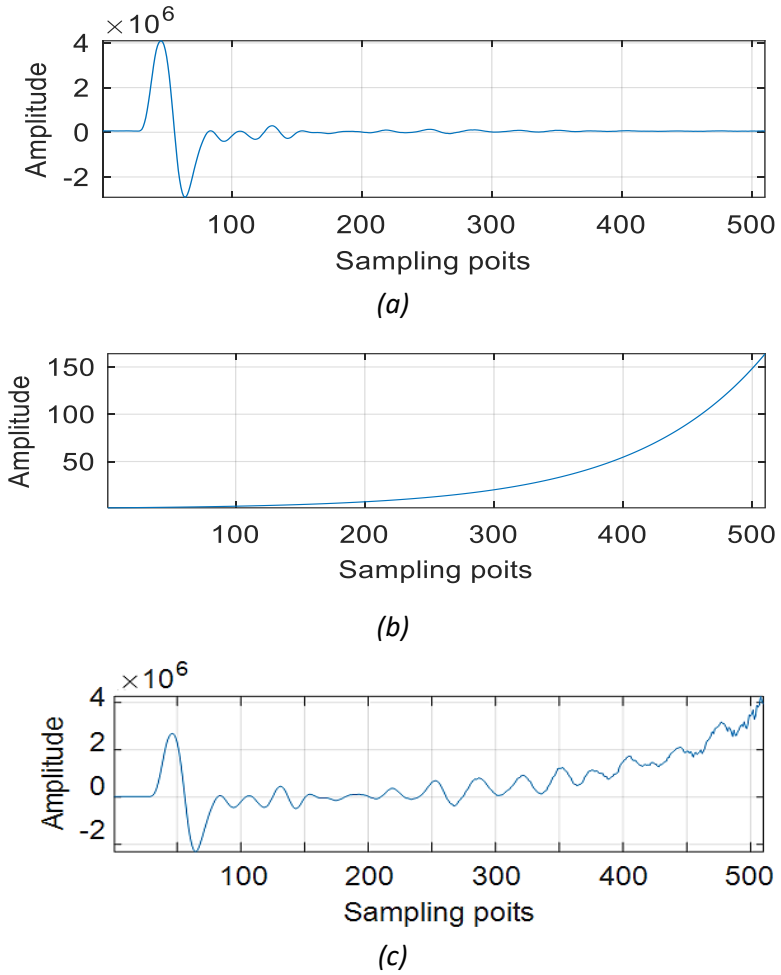



*Figure 3-13. Deviation of signals from zero by increasing the depth after utilising a gain function, a) original signal (randomly selected trace from radargram in Figure 3-6, b) graph of the applied gain function (exponential), c) Signal deviation (drift) after utilising the gain function*

As discussed by Goodman and Piro (2013a), the problem of DC-drift noise can be solved in two ways; i) implementation of a moving average filter on the gained trace, which is practically a background filter, ii) implementation of a bandpass filter to exclude the frequencies outside the GPR antenna's frequency range.

The implementation and integration of the moving average filter in the proposed algorithm are discussed in detail below.

For example, the attenuation of scattered signals in the unprocessed radargram in Figure 3-14 is such that almost no information (or very little) can be extracted below the 250$^{th}$ sampling

point. Without implementing an appropriate signal gain adjustment on this radargram, useful information about the subsurface may be suppressed and ignored.

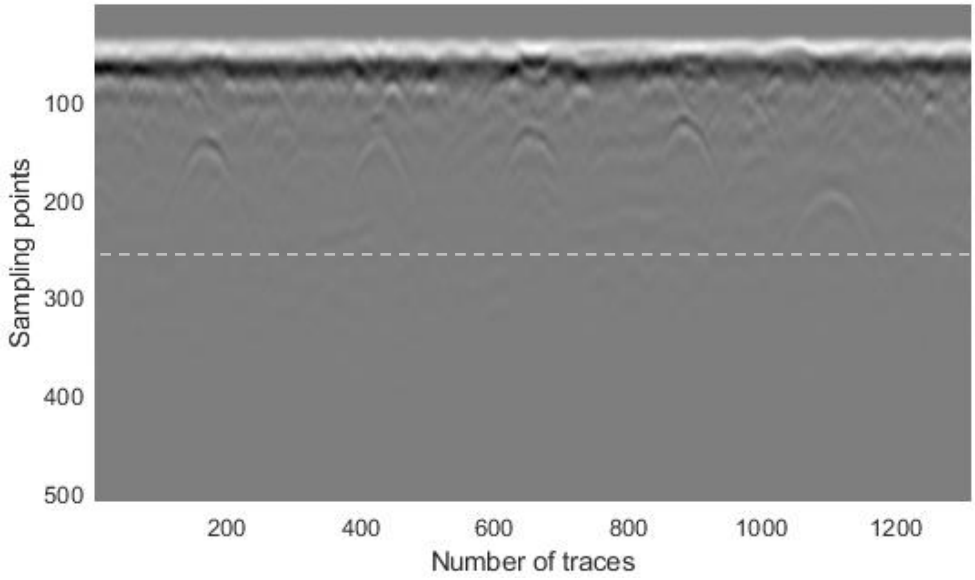



*Figure 3-14. Unprocessed radargram selected from the dataset*

In this study, the power and exponential gain functions (Eq.3-16 and Eq.3-17) provided by Huber and Hans (2018), which are commonly used gain functions for GPR data processing, have been selected and combined (Eq.3-18) for use.

$$G_{power}(n) = n^{\alpha} \quad (3\text{-}16)$$

$$G_{exponetial}(n) = e^{\alpha n} \quad (3\text{-}17)$$

$$G_{combi} = n^{\alpha} \times e^{\alpha n} \quad (3\text{-}18)$$

Where $(G)$ is the gain function, $(n)$ is the sampling point, and $(\alpha)$ specifies the intensity of the signal amplification or through which the gaining procedure can be opt-out by the user when $(\alpha = 0)$.

In addition, to eliminate the signal drift, a moving average subtraction function (Eq.3-19) was multiplied by the gain function (Eq.3-21) (Benedetto et al., 2013; Goodman and Piro, 2013b):

$$f_{sub} = y(n,k) - \bar{y}(n,k) \tag{3-19}$$

Where:

$$\bar{y}(n,k) = \frac{1}{n}\sum_{i=1}^{n} y_i \tag{3-20}$$

The parameter $\bar{y}(n,k)$ is the average value of all traces in each radargram, which will be subtracted from individual traces at each of the N successive steps within the algorithm. Eq.3-21 shows the equation for the gained sensor measurements $y_G(n,k)$ at sampling point $(n)$ and instant antenna position $k$ (trace number).

$$y_G(n,k) = G_{combi} \times f_{sub} \tag{3-21}$$

Finally, to integrate the desired filters into the algorithm, a new function $\xi(n,k)$ was developed, as expressed in Eq.3-22, which replaces the classical system input parameter $(y_k)$ from Eq.3-14 with the new parameter $(\xi_k)$, as expressed in Eq.3-23, when the gain sensitivity $\alpha = 1$. This allows for the modification of Eq.3-14 as follows:

$$\xi(n,k) = n \times e^n \times \big(y(n,k) - \bar{y}(n,k)\big) \tag{3-22}$$

$$\hat{x}_k^+ = \boldsymbol{A}\hat{x}_k^- + \boldsymbol{K}_k(\xi_k - \boldsymbol{H}_k\boldsymbol{A}\hat{x}_k^-) \tag{3-23}$$

An example demonstrating the performance of the applied linear-exponential gain adjustment is shown below in conjunction with the radargram in Figure 3-14. There is significant attenuation of signal amplitudes below the 250$^{th}$ sampling point (horizontal dashed line) and throughout the radargram. A selected unprocessed trace (40$^{th}$ trace) of the radargram (Figure 3-15a) illustrates this attenuation in detail. The signal amplitude approaches zero after almost the 150$^{th}$ sampling time. The performance of the applied linear-exponential gain functions (Figure 3-15b) can be seen in Figure 3-15c, which generates a gain-

adjusted trace and the associated radargram shown in Figure 3-16 with gradually amplified signal amplitudes as the depth increases.

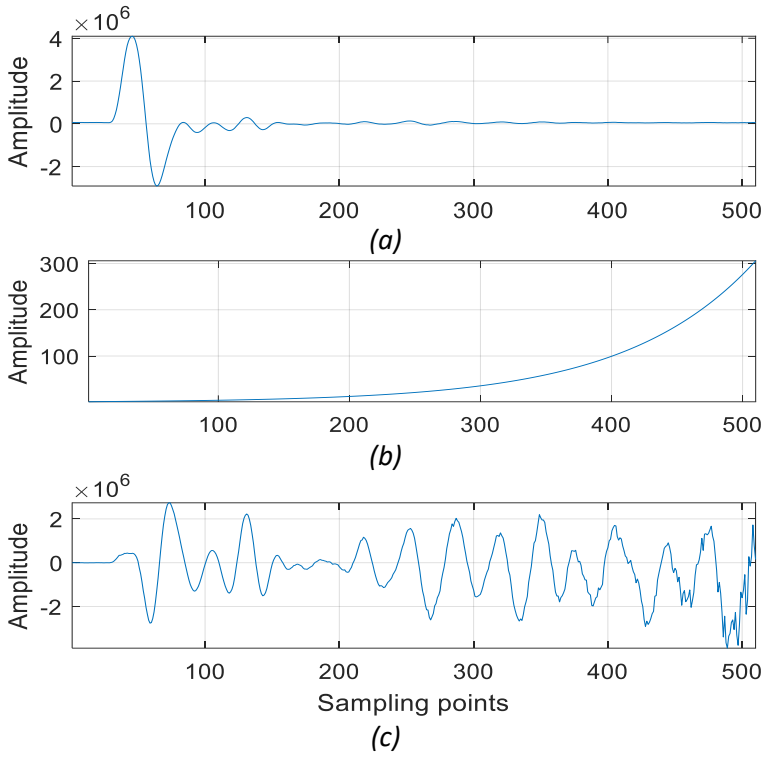



*Figure 3-15. Gain adjustment of a 40th trace of the radargram. a) unprocessed trace, b) applied gain functions, c) gain adjusted trace*

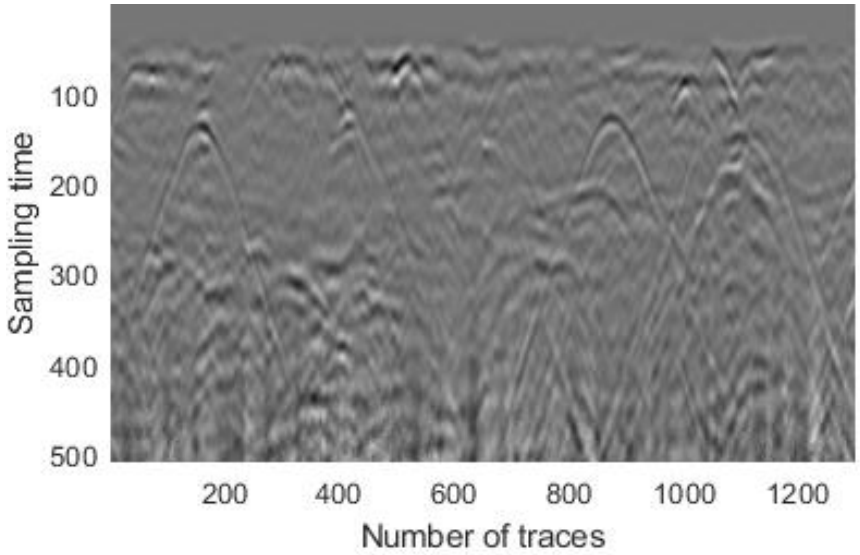



*Figure 3-16. The selected radargram after utilising the linear-exponential signal gain adjustment.*

In addition, the issue of drifting signals from the zero position (Figure 3-17) caused by the signal gain adjustment is addressed by utilising the integrated windowed moving average filter as shown in Figure 3-18.

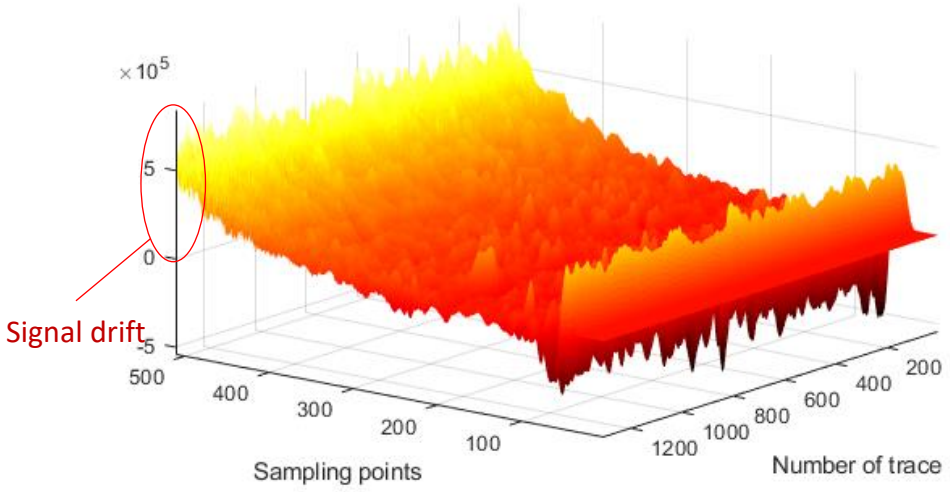



*Figure 3-17. Signal drift after utilising gain function*

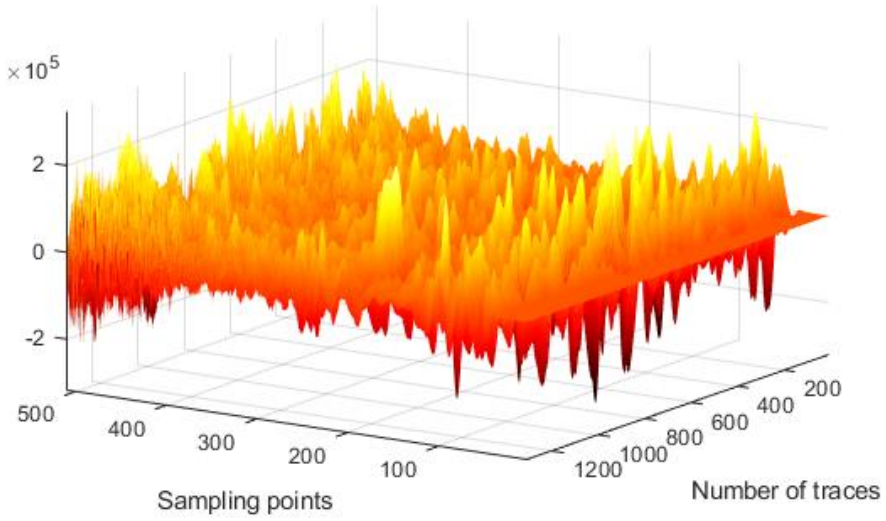



*Figure 3-18. Solved signal drift issue using windowed moving average*

Similarly, the performance of other applied gain functions including linear, exponential, and polynomial gain functions on the same radargram is shown in Figures 3-19 to Figure 3-21

respectively. To demonstrate the effect of the applied gain functions in detail, a single trace (the 40th trace) of the radargram was selected, and the gain functions were used in conjunction with a windowed moving average filter. The issue of signal attenuation was addressed by integrating an adjustable gaining mechanism (consisting of linear, exponential, and polynomial gaining functions) and the moving average filter into the proposed algorithms.

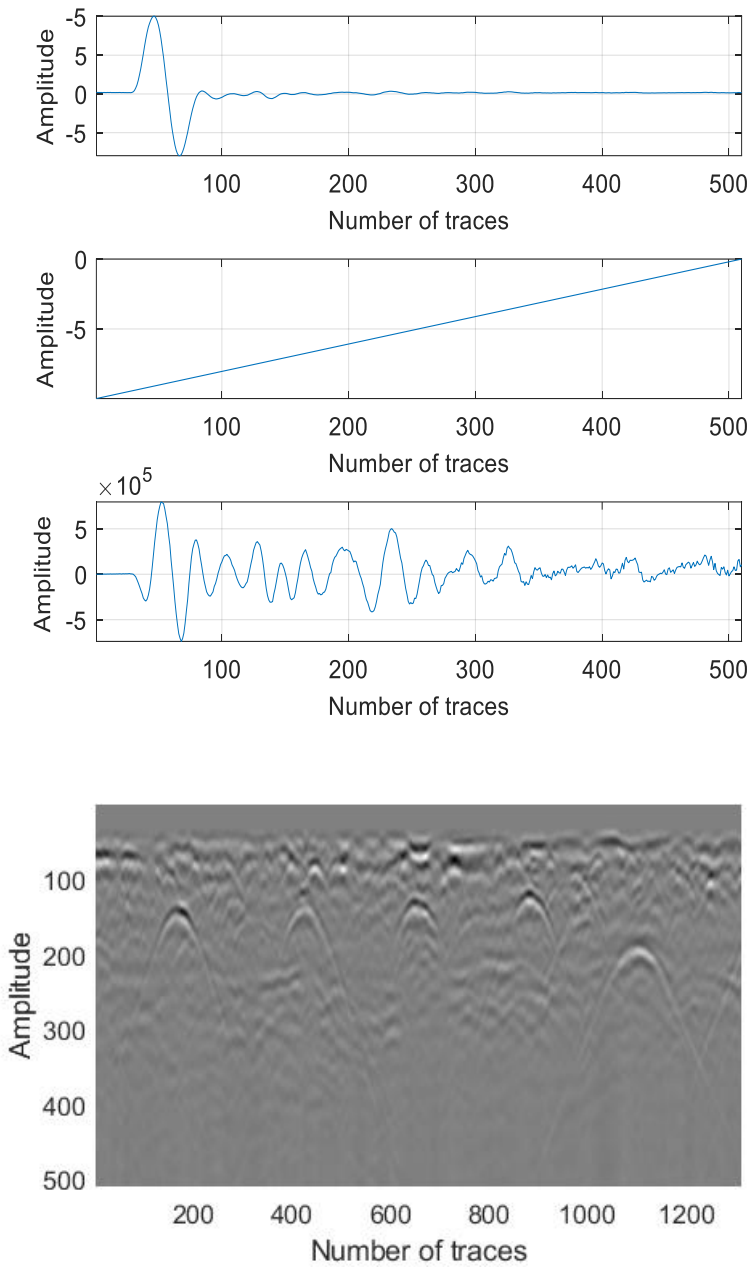



*Figure 3-19. Applying linear gain function, a) unprocessed trace, b) applied gain functions, c) gain adjusted trace, d) linear gain function applied on a selected radargram*

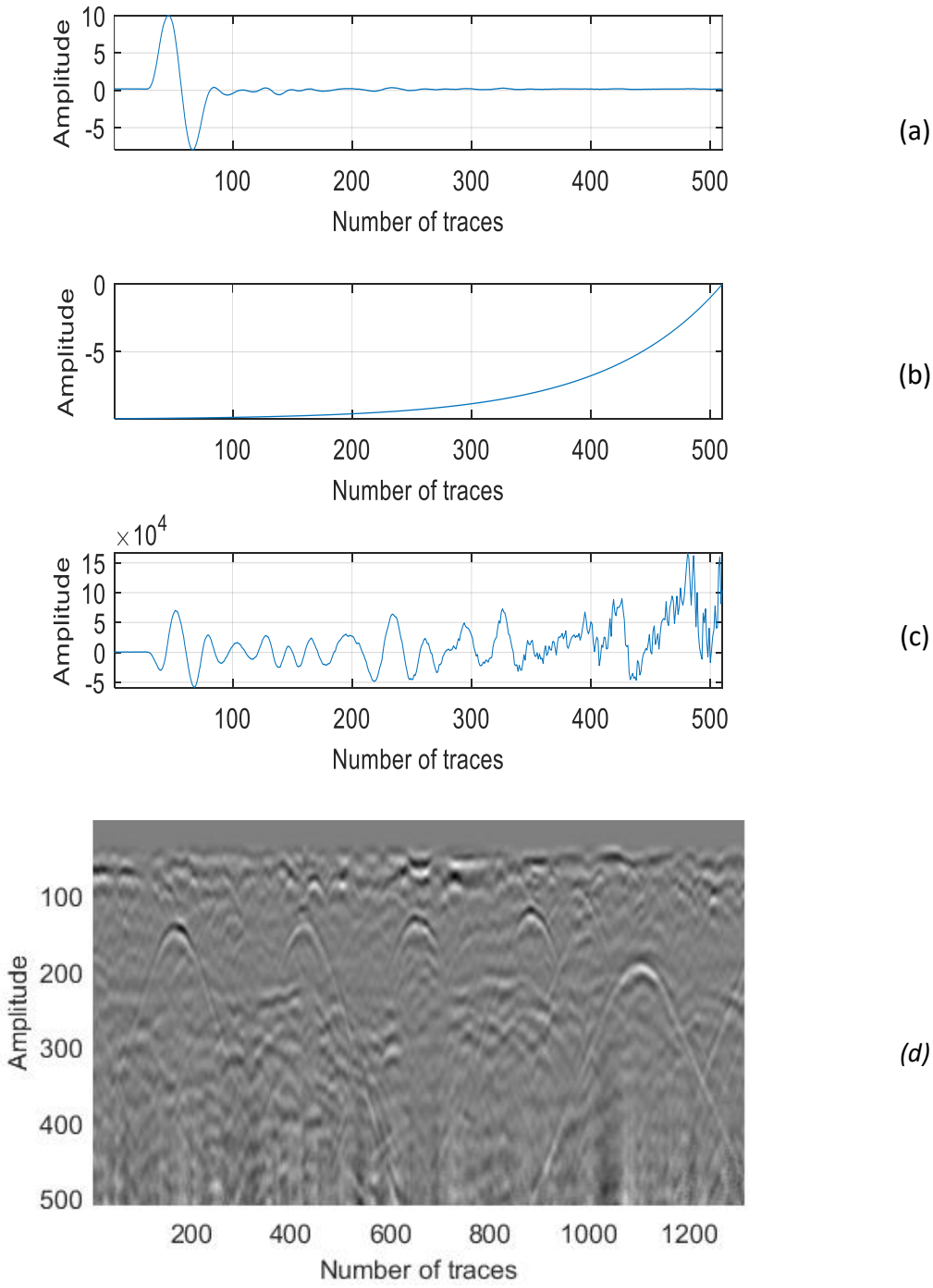



*Figure 3-20. Applying exponential gain function, a) unprocessed trace, b) applied gain functions, c) gain adjusted trace, d) exponential gain function applied on a selected radargram*

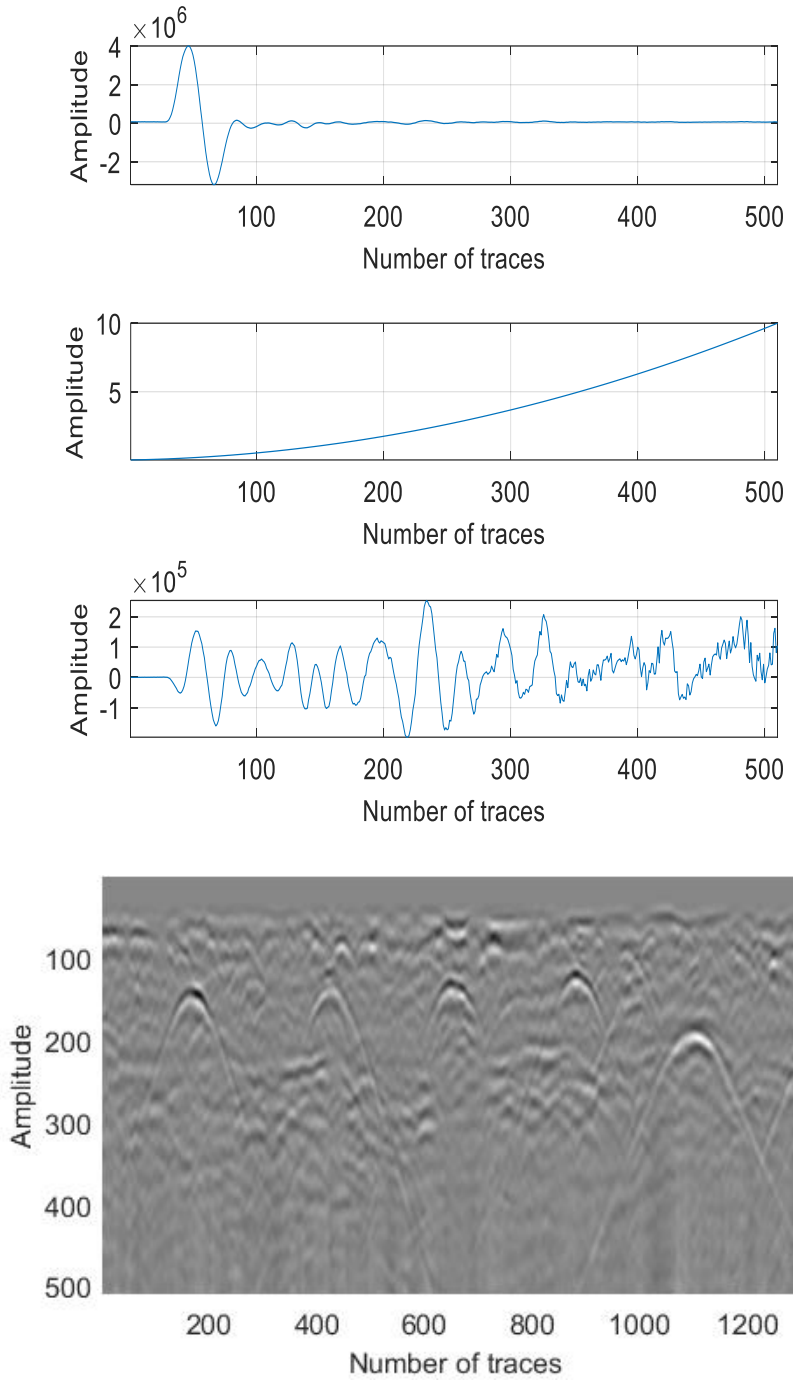



*Figure 3-21. Applying polynomial gain function, a) unprocessed trace, b) applied gain functions, c) gain adjusted trace, d) polynomial gain function applied on a selected radargram.*

Each of the aforementioned basic pre-processing filters can improve the quality of the data and, as a result, the quality of survey interpretations in terms of accuracy in detecting and locating targets and anomalies in the subsurface. However, further post-processing and analysis are often needed to reduce noise, declutter, and extract important information from

sampled scattered signals. Typically, the post-processing of GPR data is performed manually by an expert depending on their level of knowledge and expertise.

The general overview of the KF algorithm after integrating the signal processing filters and using the modified equation (Eq.3-23) can be depicted as a flowchart (Figure 3-22). The procedure for identifying target signals using the proposed KF differs from the previous work discussed in the literature review chapter and will be discussed in the next section.

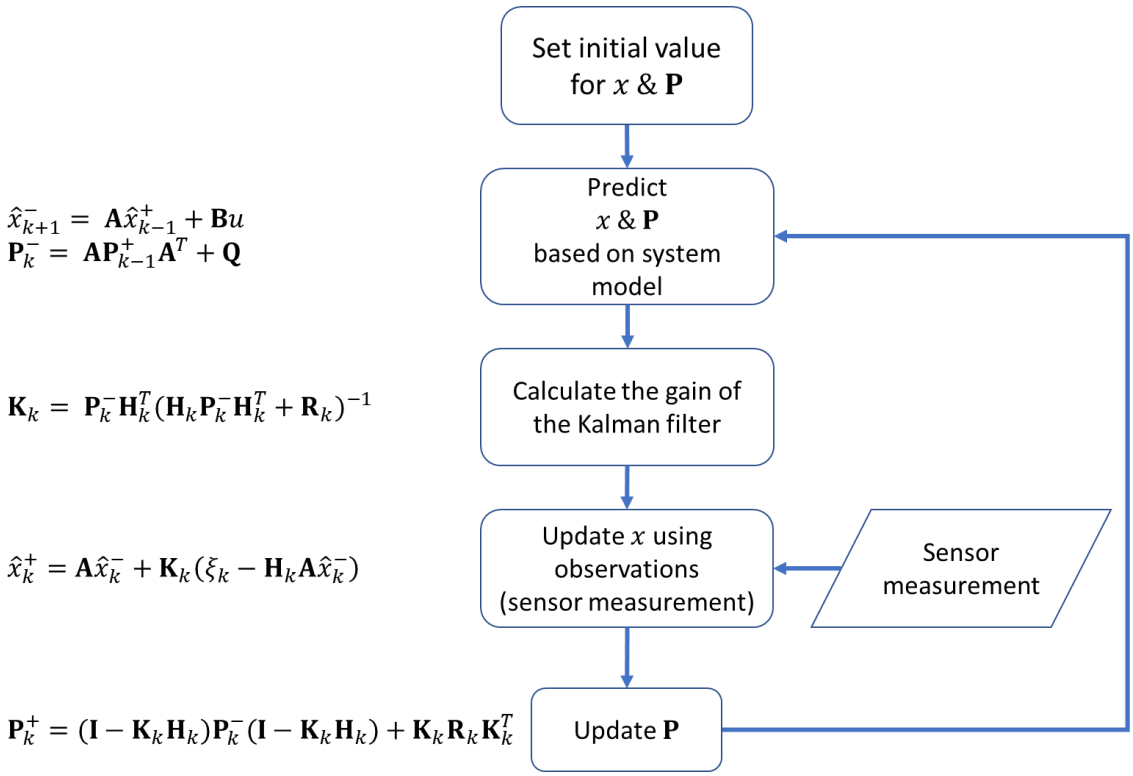



*Figure 3-22. Linear discrete Kalman filter flowchart. In the above equations, $\widehat{\boldsymbol{x}}_{\boldsymbol{k}}^{-}$ is the a priori estimate, $\widehat{\boldsymbol{P}}_{\boldsymbol{k}}^{-}$ is the a priori state estimation error covariance, $\widehat{\boldsymbol{x}}_{\boldsymbol{k}}^{+}$ is the a posteriori estimate, and $\widehat{\boldsymbol{P}}_{\boldsymbol{k}}^{+}$ is the a posteriori estimation error covariance. $\boldsymbol{K}_k$ is the Kalman gains, $\boldsymbol{A}$ is the 'dynamic model's transition matrix', $\boldsymbol{B}$, $\boldsymbol{Q}$, are the 'input', 'system noise covariance', $\boldsymbol{P}$ is the 'estimation-error covariance', $\boldsymbol{H}$ is the 'measurement model matrix', and $\boldsymbol{R}$ is the 'sensor measurement noise covariance matrix'*

### 3.3.5 A novel target detection method using Normalised Innovation Squared values (NIS)

As previously mentioned, the accuracy and performance of the KF can be evaluated using different parameters (Simon, 2006). One of these parameters is the *innovation* ($\mathbf{\nu}$), which is the residual between the estimated return signal value by the KF and its actual measured value from the receiver antenna, presented in its general form in Eq.3-24.

$$\mathbf{\nu} = y(k) - \mathbf{HA}\hat{x}_k^- \tag{3-24}$$

Another KF parameter produced during the filtering process is the normalised innovation squared (NIS), which is a measure used to evaluate the performance of the KF and update the KF estimates accurately. The NIS is a scalar derivative of the *innovation* parameter ($\mathbf{\nu}$) and can be expressed using Eq.3-25.

$$NIS(k) = \nu(k)^T\, \mathbf{S}^{-1}\, \nu(k) \tag{3-25}$$

Where $S$ is the *innovation covariance* (Eq.3-26)

$$\mathbf{S} = \mathbf{H}(k)\mathbf{P}(k)^{-1}\mathbf{H}(k)^T + \mathbf{R}(k) \tag{3-26}$$

The recorded innovation values of the radargram are presented in Figure 3-23, which shows overlapped vectors associated with 510 sampling points from 1200 traces (A-scans) of the radargram. These 1200 overlapped vectors create a complex and overwhelming signal that can be difficult and time-consuming to process and analyse, leading to high computational costs.

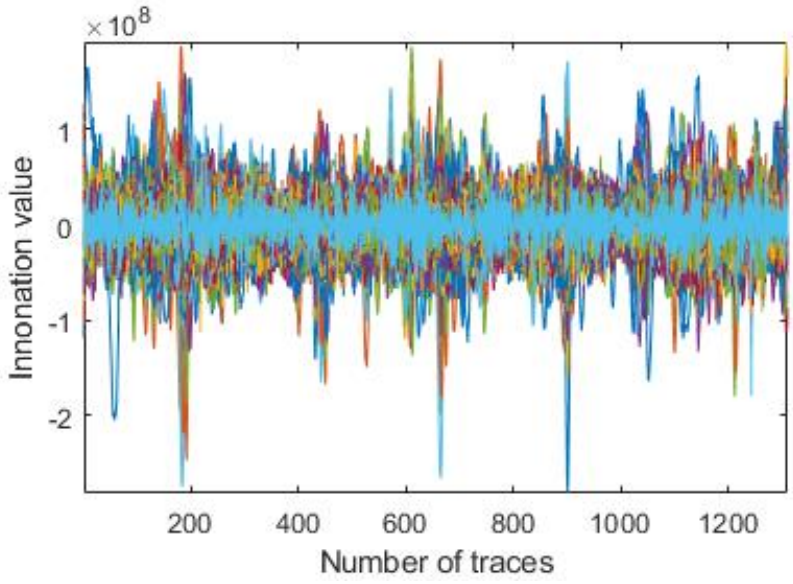



*Figure 3-23. Example of innovation values in their vector form*

In the previous chapter (section 2.5), the detection of target signals using the Kalman filter (KF) relied on the innovation parameter (vector) and the implementation of the chi-squared hypothesis testing method. However, this approach has several drawbacks, including a lack of accuracy, reliance on subjective user opinions, and a dependence on predefined databases that may not be relevant. Additionally, as discussed in section 2.5.1 of the literature review chapter, the chi-squared testing method cannot address the issue of unknown target widths of buried targets.

To address these problems, this study proposes a novel approach that utilises only scalar NIS values, without the need for the chi-squared hypothesis testing method or any other external parameters. As shown in Figure 3-24, the scalar NIS values can be collected and combined into a new vector called the NIS signal function in this thesis, which can be processed using conventional signal processing tools to accurately detect and locate buried targets and ground anomalies where impedance exists. To clarify, the NIS signal is created by recording the NIS values at each step of the Kalman filtering process and concatenating them into a single

function. The Y-axis values of this function represent the recorded NIS values at specific locations, which are associated with the actual trace numbers where the NIS values were generated and recorded.

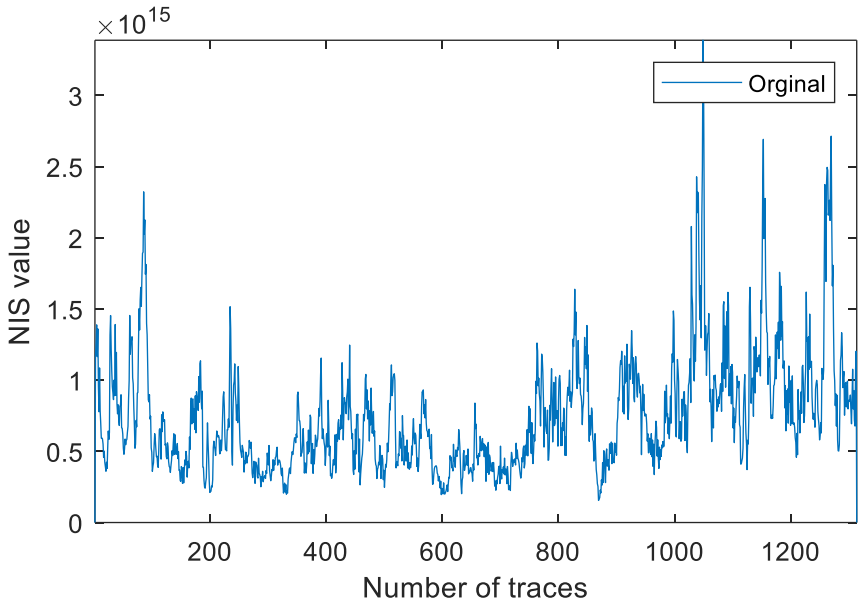



*Figure 3-24. Example of created NIS signal using scalar NIS values by collecting them and putting them together in a new matrix.*

Like the innovation parameter, the NIS can also be used as a benchmark for KF performance, but it is a scalar value. This means that the NIS values are updated iteratively at each step of the estimation process and indicate the accuracy of the KF estimates. A significant increase in the NIS value, such as a sudden increase compared to the previous adjacent point, indicates a growing difference between the KF estimate and sensor measurements, which may indicate a faulty estimate by the KF. For example, a sudden increase in the NIS value from ~ 0.6 x $10^{15}$ at the 100th trace on the X-axis to ~ 2.4 x $10^{15}$ would be an indicator of a faulty estimate by the KF. In other words, as shown in Figure 3-25, the higher the value of NIS at each estimation step, the more significant the difference between the KF estimate and the measured signal (error in the estimate).

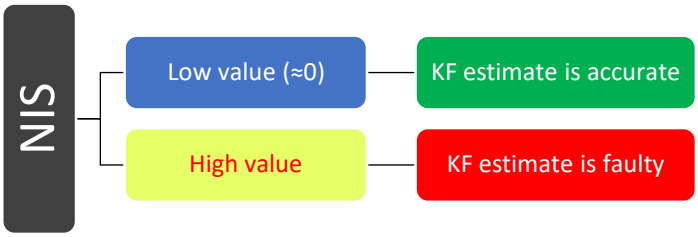



*Figure 3-25. Normalised innovation (NIS) evaluation chart*

By analysing the NIS signal, the KF algorithm can make decisions about target positions or ground anomalies based on actual measured data, rather than relying on an external testing approach or subjective threshold chosen by an expert. This allows the analysis to be performed within the KF algorithm without the potential for bias from external sources.

To comply with the provided KF in the previous section (Eq.3-23), the parameter $y(k)$ of the *innovation* (Eq.3-24) was replaced with the parameter $\xi(k)$ (Eq.3-22) to generate the new *innovation* (Eq.3-27) from which a new NIS equation was specified (Eq.3-28) and adapted into the developed LDKF.

To conform to the Kalman filter equation provided in the previous section (Eq. 3-23), the parameter $y(k)$ of the innovation (Eq. 3-24) was replaced with the parameter $\xi(k)$ (Eq.3-22) to create a new innovation equation (Eq.3-27). From this equation, a new NIS equation (Eq. 3-28) was specified and incorporated into the developed LDKF.

$$\boldsymbol{\nu}' = \xi(k) - \mathbf{H}\mathbf{A}\hat{x}_k^- \tag{3-27}$$

$$NIS(k) = \boldsymbol{\nu}'(k)^T\, \mathbf{S}^{-1}\, \boldsymbol{\nu}'(k) \tag{3-28}$$

An accurate investigation of the NIS signal obtained from real data needs a good overview of its characteristics, which could be obtained from a noise-free radargram.

### 3.3.5.1 Generating noise-free NIS signal using numerical modelling

To generate a noise-free radargram, numerical modelling was used to simulate a survey configuration with zero noise contribution (Figure 3-26). This was done using data from a trial site (the BT test site) set up in conjunction with the FINDIT project (Metje et al., 2022). The trial site covered an area of 11 m x 13 m, with 23 scan lines and five linear pipes installed at shallow depths. The GPR data was recorded in DZT GGSI format using a single-channel antenna with a centre frequency of 400 MHz, with scan line spacing of 0.5 m and lines perpendicular to the pipes. The ground conditions at the trial site were confirmed through BGS borehole testing as 6.1 m of glacial sand and gravel over red crag sand. The weather at the time of the survey was dry, with substantial rain the night before the survey (< 25 mm). Hyperbola matching in the GPR data indicated an approximate dielectric constant of 6.5 for the sand.

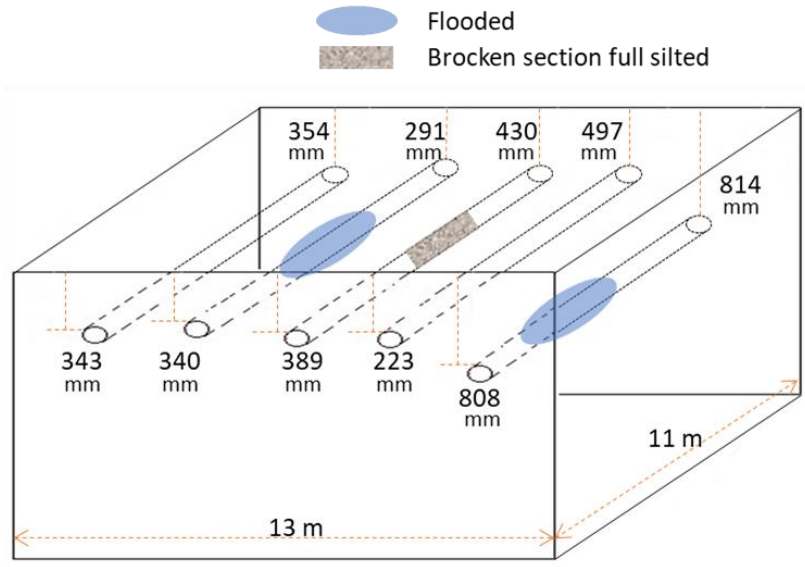



*Figure 3-26. Detailed configuration of the buried pipes on a trial site*

A numerical noise-free radargram (Figure 3-27a) was created using a numerical model of the trial site and its associated radargram (Figure 3-27b) with the GPR modelling software matGPR (Tzanis and Kafetsis, 2004). This numerical radargram was used to compare the characteristics

of the numerically generated NIS signal (Figure 3-27c) with the NIS signal generated from real data (Figure 3-27d).

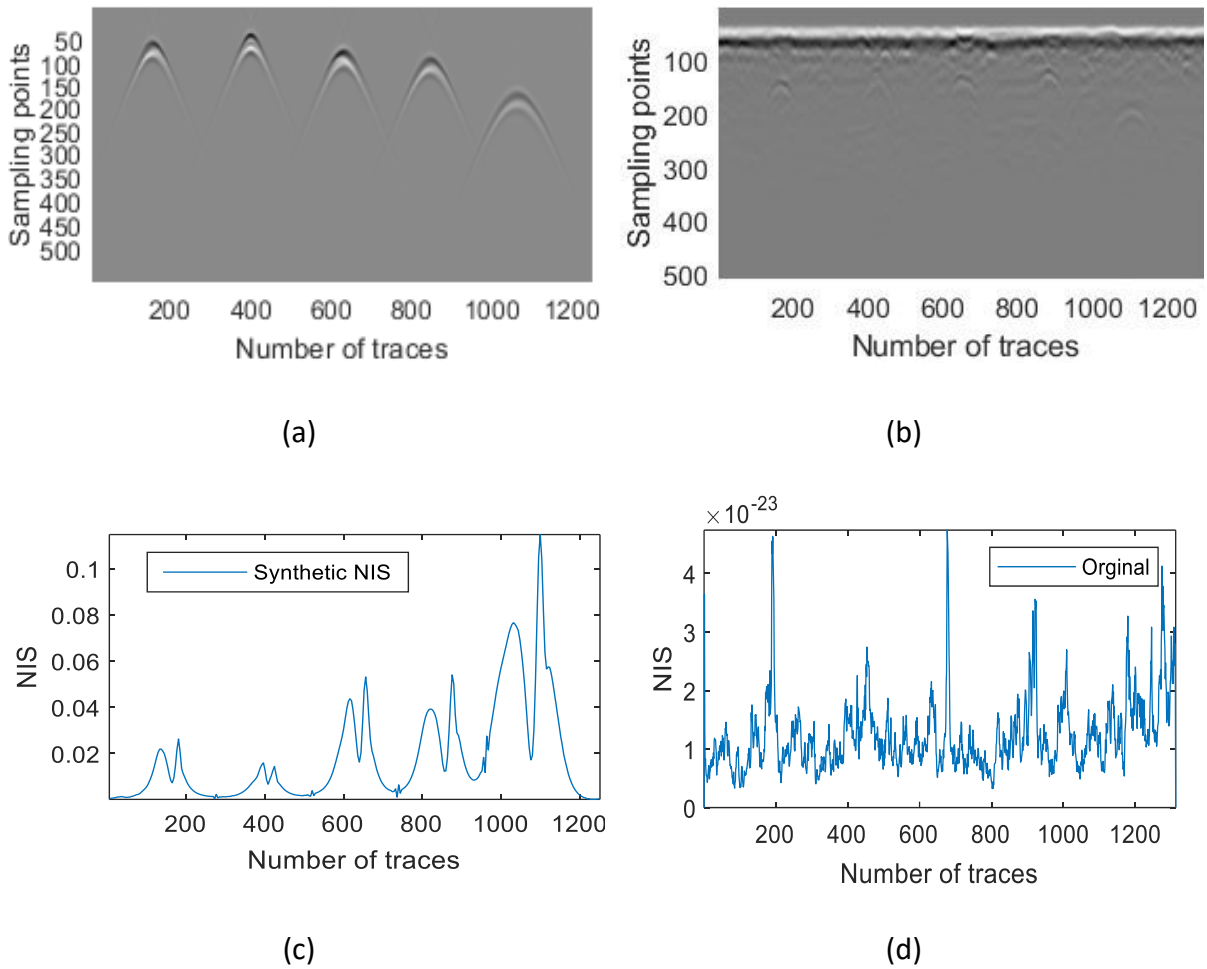



*Figure 3-27. a) Numerical GPR model of the trial site, b) Experimentally obtained GPR radargram from the trial site, c) Noise-free NIS-signal from numerical radargram, d) Noisy NIS-signal from the trial site*

An abrupt increase in the NIS (Figure 3-27c) can be observed in the vicinity of trace number 600 where KF enters the zones of radargram containing hyperbolas (Figure 3-27a), which is an indicator of a faulty KF estimate. A double signal-peak trend can be observed on the hyperbola tails (left, and right of the vertex) instead of a single peak at the vertex. The logical explanation for these trends is KF's attempt to correct its faulty estimate at these positions. After entering the KF into a target-present section, the corresponding trace containing target information

will be added to the system. Hence, the discrepancy between the predicted value and the sensor's output will be increased. Therefore, KF attempts to reduce the discrepancy in subsequent updates. After an abrupt increase of the NIS signal, more weight will be put on sensor data instead of the KF estimate. Consequently, a new estimate will be approximated with regard to the sensor data by which the amplitude of the NIS signal at the vertex will be reduced. The same trend occurs when the KF leaves the hyperbola zone and moves into a target-free section by which a second peak will be created.

As shown in Figure 3-27c, the NIS signal indicates significantly higher amplitudes in sections with targets or anomalies (impedance) compared to target-free sections. Similar characteristics can be observed in the experimentally obtained radargram (Figure 3-27b) and its generated NIS signal (Figure 3-27d). In other words, changes in NIS values in sections with targets (or anomalies) exhibit higher amplitudes and lower frequencies compared to background noise signals with higher frequencies and lower amplitudes distributed throughout the entire signal. Therefore, sampling and analysis of NIS signals can be a powerful tool for locating shallow pipes and ground anomalies. The position of identified peaks can be indicated using dashed lines generated by the Kalman filter (Figure 3-28).

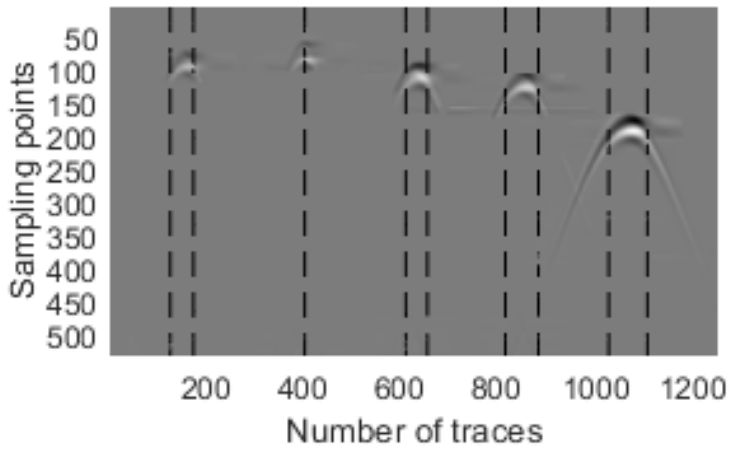



*Figure 3-28. Result radargram with allocated dashed lines produced by the KF after analysing the numerical model generated by the matGPR software.*

Analysis of the numerically generated NIS signal confirmed the accuracy of the proposed algorithm at this stage. However, real GPR data may contain significant ambient noise and other perturbations that result in a noisier and more ambiguous NIS signal. To address this issue and improve the performance of the proposed detection algorithm, signal processing techniques were implemented in combination with an optimisation technique using machine learning. A step-by-step description of this method is discussed in the following sections.

#### 3.3.5.2 The algorithm for detection of targets using NIS signal

Under the assumption of being in the target-free state of the KF, the NIS values for each sample point (the entire row) will be calculated and recorded individually in a new vector. As discussed above, these individual scalar NIS values generate a NIS signal associated with each of these sampling points and are called *local* NIS *signals* in this thesis. The mean value of the local NIS signal associated with the selected sampling point, which is called *global* NIS (GNIS) in this thesis, is then calculated to be utilised as a decision criterion (threshold) to decide whether the KF should remain and operate in the target-free state, or the KF should switch to the target-present state. Subsequently, the KF starts operating in the general mode (standard mode), i.e. under no governance assumptions. The values of the local NIS signal at each sampling point will be compared against the associated GNIS value. If the value of a local NIS exceeds the value of the GNIS threshold, this section will be marked as target-present and consequently the KF switches from the general mode into the target-present mode. Similarly, any section associated with a local NIS value below the GNIS threshold will be marked as a target-free section and consequently, the KF switches from general mode into the target-free mode.

It is worth noting that, to the best of the author's knowledge, there are no publications in the literature that implement this novel approach. Therefore, the mean of the local NIS signals, defined as GNIS (threshold), was determined through a trial-and-error process by analysing various GPR datasets available for this project. It should be noted that the GPR data was provided by third parties in the industry. The mean value of the GNIS has proven to be a reasonable threshold. Analysing various coefficients for the GNIS value (e.g., 0.1GNIS, …, 0.9GNIS, …, 1.1GNIS, …) and comparing the outcome showed that defining a threshold above and below the mean value can lead to overlooking (ignoring) the target associated information or overestimated (hyped) the unwanted information in radargrams, respectively. On the one hand, as illustrated in Figure 3-29, if the threshold is selected above the mean value (e.g. dashed line 1), most of the signal will be placed in the target-free area (area below the threshold) and will be ignored, albeit consisting of real target information (missing targets). On the other hand, if the threshold is selected below the mean value (i.e. dashed line 2), then most of the signal will be placed in the target-present area (area above the threshold) and consequently, part of the signal with no target's associated information will be processed as a target-present state (causing false alarms). Nevertheless, if needed, the GNIS threshold can be defined independently according to project needs and by changing the coefficient.

As discussed above, anomalies in the dielectric properties of the ground can be detected and located by evaluating the NIS signal at each sampling point of the radargram by identifying the peaks of the signal. However, like noisy EM GPR signals, the NIS signal can also contain high-frequency signal components (noises). This issue, however, can be addressed by processing the NIS signal by making use of signal processing techniques. Generally, signal processing is a tool to extract the information (e.g. specific frequency components) by some sort of manipulating the signal (e.g. capturing, digitising, and filtering) (Abraham, 2017). Converting

an analogue signal – which is a signal with different frequency and amplitude at various times – to a digital signal (which consists of only 0 and 1 values) offers the opportunity to manipulate the signal using mathematical techniques to improve, adjust or visualise the information in a specific way (Brice, 2001).

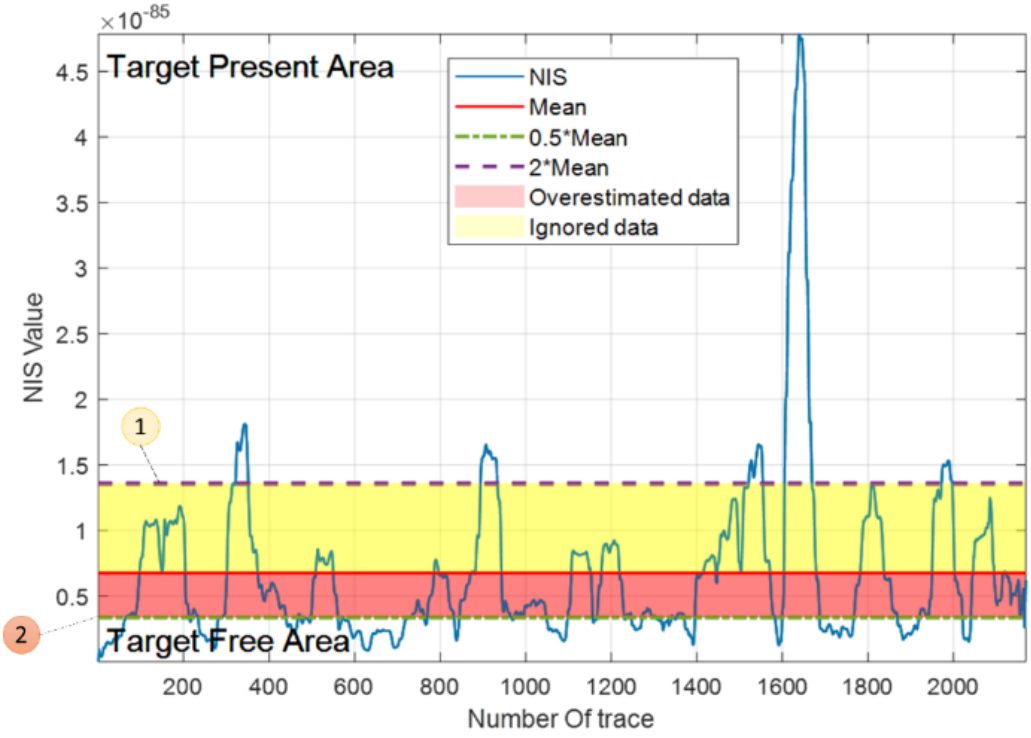



*Figure 3-29. Experimental determination of the global NIS*

The generated NIS signal in this research is an analogue signal (similar to noisy GPR signals) consisting of various frequencies and amplitudes at different times. As such, the captured NIS signal can be mathematically analysed and processed (e.g., denoised) to extract useful information (e.g., frequency components). A commonly used technique for determining the frequency components of a signal (e.g., noise) is the Fast Fourier Transform (FFT), an algorithm for fast and efficient computation of the discrete Fourier Transform (DFT), which is a well-established signal processing tool for manipulating signals by breaking them down into frequency components (sums of sines and cosines) and vice versa (Brunton and Kutz, 2017). A detailed explanation of the Fourier Transform and the FFT is beyond the scope of this

research, and interested readers are referred to other literature on the subject (Brunton and Kutz, 2017; Rader and Mailing, 1967).

The analysis of the numerically created NIS signal in section 3.3.5.1 has characterised the NIS signal function as a signal with non-extreme, varying frequencies over time (low noise characteristic). Based on this, a frequency threshold can be determined for the Fourier transform so that frequencies above this threshold (i.e., noise frequencies) can be filtered (eliminated). This allows for the creation of a new NIS signal function using all sin and cosine frequency components of the initial NIS signal function, except for the eliminated frequency components. This reconstructed NIS signal function can provide more detailed and specific information about buried targets and ground anomalies (e.g., spatial location), which will be discussed in detail in the following sections.

The performance of the FFT as a denoising filter is often highly dependent on the selected range of frequencies that are allowed to pass through the filter. The optimal range of Fourier frequencies in this research was determined by using the correlation between receiver operating characteristic curves (ROC) - which represent the trade-off between the probability of true detections and false alarms - and the FFT frequencies, with the help of an optimisation technique. For a detailed description of the ROC, interested readers are referred to the literature (Carter, 1977; Chang, 2010; Dawood and Narayanan, 2001; Kerekes, 2008).

Initially, the area under the ROC curves (AuC) and the mean of innovation were set as the cost functions of the optimisation problem. Because these two cost functions are associated with different objectives, and no single solution can satisfy both objectives at the same time (non-dominated solution), the problem becomes a non-trivial problem for which a multi-objective optimisation approach was implemented. Furthermore, due to the complexity of the problem,

traditional optimisation approaches are often not efficient at finding global solutions. Therefore, a multi-objective genetic algorithm (MOGA) was used to solve the optimisation problem, which will be discussed in the next section.

It is worth noting that determining the optimal FFT frequency range is typically a case-dependent task. Therefore, a general optimal frequency range cannot be determined at this stage of research unless a comprehensive test that includes all soil and pipe type combinations is carried out.

## 3.4 Overview of the proposed optimisation algorithm

Before the Kalman filter algorithm is executed, the input variables including the process and sensor noise covariance matrices **Q** and **R**, as well as the FFT frequency, must be determined appropriately. Two different cost functions, the mean of innovation and the area under the ROC curves (AuC), were considered for the optimisation process due to the differences in the characteristics of these input parameters. The overview of the MOGA developed in this research is shown in Figure 3-30.




*Figure 3-30. Overview of the multi-objective genetic algorithm (MOGA) optimisation procedure of the Kalman filter*

### 3.4.1 Multi-objective genetic algorithm (MOGA) and Pareto solutions

As mentioned previously, in this thesis, a MOGA was used to solve the optimisation problem of the proposed semi-autonomous algorithm due to its ability to handle multiple objectives and find global solutions. In general, a Pareto solution (PS) curve associated with a MOGA represents the objective values of solutions to a multi-objective problem, using which the dominance of one or more PSs over others (best solution) can be identified. For example, Figure 3-31 shows a PS curve associated with a radargram, offering nine PSs for the optimisation problem. It can be observed that except for one PS, i.e. P7, which is dominated by P8 (i.e. P8 offers better results in terms of satisfying both cost functions $f_1$ and $f_2$), none of the other PSs dominate the others. In other words, none of the eight remaining PSs can simply be selected as the best solution for the problem, and further investigation will be required to select the outperforming PS.

For a detailed description of the MOGA and PSs, the reader is referred to study the available literature (Deb, 2011; Fonseca and Fleming, 1993).

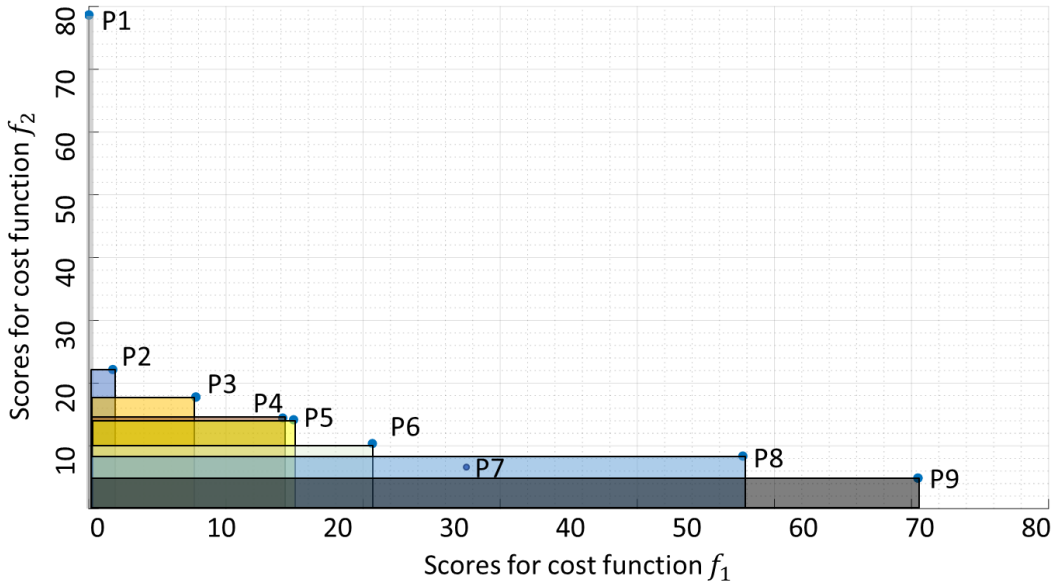



*Figure 3-31. An example of a Preto solution curve for two cost functions $f_1$ and $f_2$*

To optimise the performance of the proposed KF algorithm, a MOGA was developed using the global optimisation toolbox of MATLAB® to determine the optimal input parameters for the algorithm, which can accurately detect and locate buried targets (such as pipes) and ground anomalies. This approach is in line with objective 4 of this thesis.

MATLAB® is a widely used platform for programming and numeric computing that is commonly used by researchers and engineers to analyse data, design models and systems, and develop algorithms[15]. It has been employed by many researchers to address various geophysical problems, such as forward modelling and inverting electromagnetic-based geophysical problems (Elwaseif et al., 2017; Neyamadpour et al., 2009; Pidlisecky and Knight, 2008; Stocco et al., 2009). Additionally, MATLAB® offers a variety of apps and toolboxes that focus on specific problems, such as optimisation. For example, the global optimisation toolbox of MATLAB® includes functions for creating a genetic algorithm to search for global solutions to complex problems with multiple minima and maxima[16]. One advantage of using this toolbox is the ability to access and modify the source code to adjust and customise the search function based on the requirements of the problem. Therefore, in this project, MATLAB® was used as the platform for developing computational solutions and the global optimisation toolbox was used to create the SOGA and MOGA optimisation algorithms.

### 3.4.2 Optimisation of noise covariances Q and R

Improperly selecting noise covariances **Q** and **R** can significantly degrade the performance of the KF (e.g., by manipulating the NIS values). Noise is a random variable that can have various sources and can be mathematically described as a probability density function (pdf) (Van

[15] www.uk.mathworks.com
[16] www.uk.mathworks.com/products/global-optimization.html

Drongelen, 2010). Determining whether the chosen values for these parameters are appropriate can be challenging, especially for the model noise covariance matrix (**Q**) for which no information is available. According to Simon (2006), the innovation parameter in a KF should have a mean of zero ($\mu = 0$) to ensure the accuracy and high performance of the KF. This characteristic is utilised in this research to optimise the values of the noise covariances matrices **Q** and **R** using MOGA-based optimisation, ensuring that the necessary zero-mean condition (Eq. 3-29) is satisfied (Simon, 2006).

$$J = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{1}{k}\sum_{j=1}^{k}\boldsymbol{\nu}_{ij}^{2}\right) = 0 \qquad (3\text{-}29)$$

It should be noted that the values of innovations can be both positive and negative. To avoid negative values and simplify calculations, the values of the innovation were squared. The number of variables involved in the optimisation process depends on the number of sampling points in the radargrams. For example, for the radargrams shown in Figure 3-27b, which contain 510 sampling points, 1020 variables were used for the optimisation of **Q** and **R** (510 each).

## 3.4.3 Optimisation of Fourier frequencies (FFreq)

Reducing signal noise to improve the signal-to-noise ratio (SNR) must be done in a way that does not eliminate vital information that is suppressed by noisy signals. To determine the optimal value for the Fourier frequency, MOGA was used to maximise the area under the ROC curves (AuC) while minimising the means of innovation. AuC of ROC is a summary of ROC curves by scoring them on a scale from 0.5 (random level) to 1 (best ranking) (Yang and Liu, 2013).

It is worth noting that using AuC of ROC curves in MOGA as a trade-off between the probability of true target detection and the probability of false alarms, in conjunction with the mean of the KF's innovation and the FFT frequencies, can be challenging and requires compromise. In other words, the principle of ROC, i.e. the higher the AuC, the better the performance, should be approached with caution, as shown in Table 3-3 below. The definitions of terms used in this table are presented separately in Table 3-4. Theoretically, an AuC of ROC of 100% should result in the highest detection rate with no false alarms. However, since two objectives are required to be satisfied in the proposed MOGA optimisation algorithm in this study, defining a high sensitivity setup, despite generating a low AuC (<50%), can lead to a high ratio of false alarms. On the other hand, if the specificity of the system is set to a high level, despite generating a healthy AuC (>50%), the probability of missing targets can significantly increase. Therefore, it is important to find a suitable compromise and select a Pareto solution that provides an appropriate FFR frequency and AuC of ROC based on the quality of the GPR data and the selected goals. This will be discussed in detail in the following sections.

*Table 3-3. Overview of ROC curves and the area under these curves (AuC) for GPR application in terms of potential false alarms and missing targets associated with high sensitivity and high specificity setup of the system, respectively*

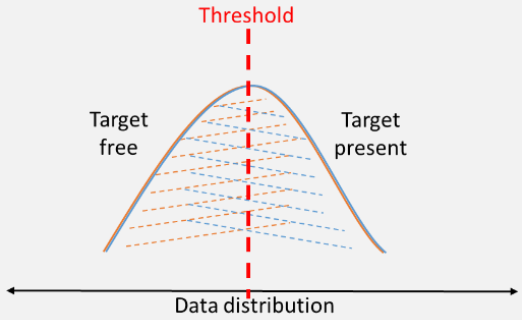
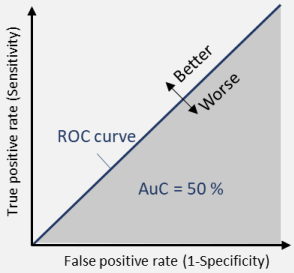
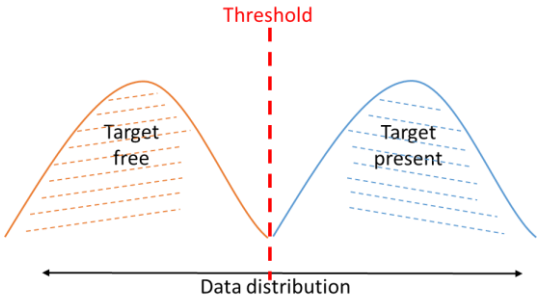
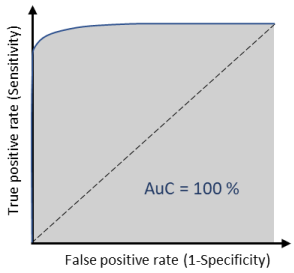
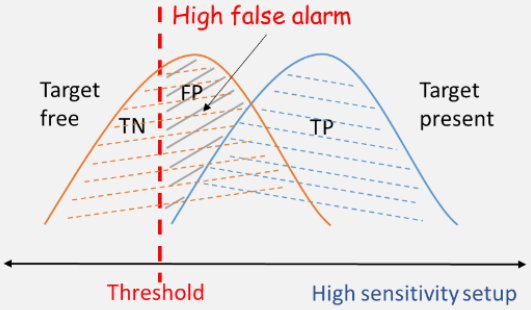
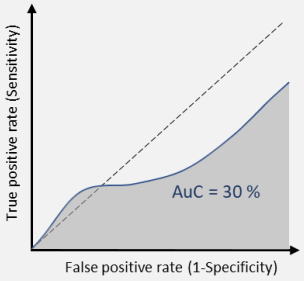
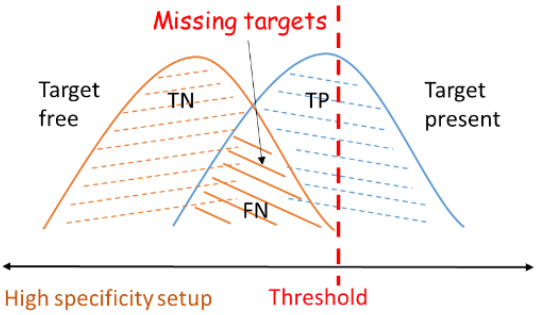
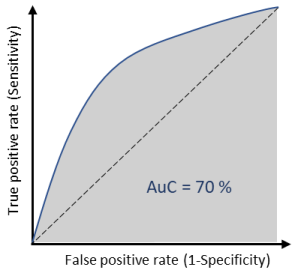
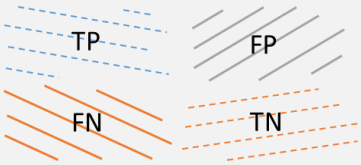
| | | |
|---|---|---|
|  |  | Random probability due to 50% value of the area under the ROC curve (AuC) |
|  |  | Theoretically, the highest detection rate (100% detection) with no false alarms |
|  |  | Probability of high ratio of false alarm due to high sensitivity setup yielding low AuC (<50%) |
|  |  | Probability of missing targets due to high specificity setup despite healthy AuC (>50%) |
| TP<br>FP<br>TP<br>TN | True Positive<br>False Positive<br>True positive<br>True Negative |  | |

*Table 3-4. Definition of terms used for generating the ROC curves (Carter et al., 2016)*

| Term | Definition |
|---|---|
| True positive | Correct result identified |
| False positive | Identifying an incorrect result as correct results |
| True negative | Correctly identified an incorrect result |
| False negative | Incorrectly identified a correct result as incorrect |
| Sensitivity (true positive rate) | probability of positive results |
| Specificity (true negative rate) | probability of negative results |
| False positive rate | The proportion of false positive results (= 1 – Specificity) |
| False negative rate | The proportion of false negative results (= 1 – Sensitivity) |

Configuration of the MOGA parameters including population size, number of generations, probability of mutations and crossover were done through a series of trial and error. A population size of 30 and a generation size of 100 as well as GA's default mutation rate of 0.01 were chosen for the model amongst standard parameter values defined in MATLAB. The computational time of the process using the BlueBEAR, 32 CPUs in two nodes, and 13 GB of memory was 09:31 hours. A typical result of the MOGA analysis is shown in Figure 3-32, similar results were obtained for all the other radargrams of the trial site. Nevertheless, the analysis of PSs showed that not all PSs can guarantee an optimal result in terms of detection rate, and false alarm ratio. This will be further discussed in the results and discussion section.

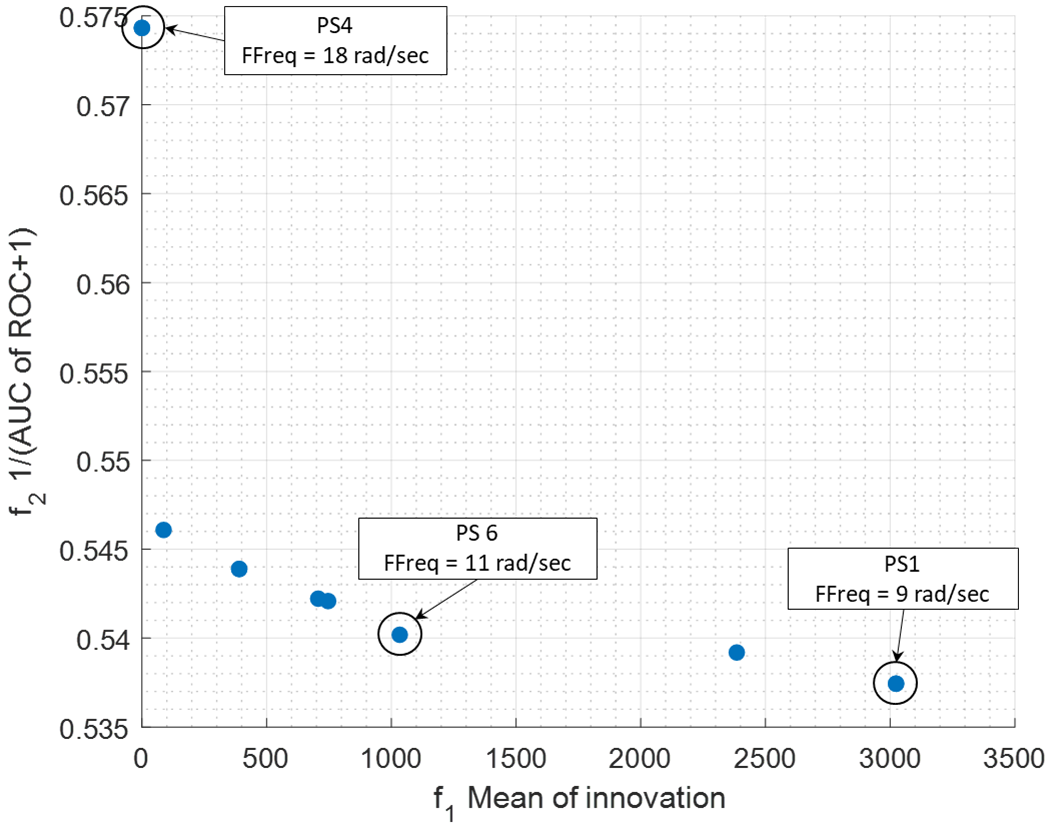



*Figure 3-32. Pareto solutions of 6th radargram based on AuC of ROC curves and Fourier frequency (FFreq)*

## 3.4.4 The solution to 'identify the target width'

The estimated radargram from a noisy backscattered signal after implementation of the KF algorithm, presented above (section 3.4), can be generally treated as the final radargram used by the expert for visual confirmation. In addition, as discussed in the previous chapter of this thesis (section 2.5.2), the main issue of previous studies is the incapability of those algorithms in identifying the target width. This issue was addressed in this study by implementing the above-discussed approach. In other words, instead of a segment-by-segment processing of individual traces, $N$ independent KFs ($N$ sampling points) were implemented to process $N$ rows of a radargram. In other words, since each trace contains $N$ sampling points, $N$ independent KFs can be implemented simultaneously to estimate the information of each

row, individually. Therefore, the mean value of NIS from the target-free state can be compared with the actual NIS value of each point on a row. Subsequently, the information of every single point of a trace can be evaluated to detect the endpoint of single or multi buried targets. The pseudo-code of the step-by-step above-discussed procedure is demonstrated below in Figure 3-33.

```
1.  PROGRAM KF_GPR(Radargram, A, H, Q, R, FFreq)
2.  .     FOR Width = [ FullLength, 1 ] DO
3.  .           Calculate number of layers ;
4.  .           FOR p = 1 to final layer DO
5.  .                 Select u ;
6.  .                 Calculate u' ;
7.  .                 FOR k = 1 to last trace DO
8.  .                       KalmanFilter(A, H, Q, R) in target free mode;
9.  .                       Saving NIS for each trace;
10. .                       Go to next trace ;
11. .                 ENDFOR
12. .                 IF Width == FullLength THEN
13. .                       FourierFilter(NIS, Ffreq);
14. .                       Calculate the position of peaks in filtered NIS ;
15. .                       Go to line 2 ;
16. .                 ELSE
17. .                       Calculate the average of NIS ;
18. .                 ENDIF
19. .                 FOR k = 1 to last trace DO
20. .                       IF NIS > average of NIS THEN
21. .                             KalmanFilter(A, H, Q, R) in target present mode;
22. .                       ELSE
23. .                             KalmanFilter(A, H, Q, R) in target free mode;
24. .                       ENDIF
25. .                       Go to next trace;
26. .                 ENDFOR
27. .           ENDFOR
28. .     ENDFOR
29. END.
```

*Figure 3-33. Pseudocode for detection of target width*

In the previous sections of this chapter, the semi-autonomous algorithm based on an optimised Kalman filter was presented for processing GPR radargrams making use of a multi-objective genetic algorithm optimisation technique in correlation with Fast Fourier transform and the receiver operator characteristics (ROC) for estimation and reduction of noise and clutter as well as accurate detection and locating buried targets and ground anomalies. The proposed algorithm was successfully utilised to process GPR data from different datasets for

which the findings will be discussed in detail in the next chapter of this thesis. A limitation associated with this algorithm was the interaction of the user to approximate locations of interest on one or two selected radargrams from a survey that might contain potential targets. However, this issue could be addressed by developing a novel autonomous algorithm based on the discrete Kalman filter (DKF) with integrated signal processing filters (developed in section 3.4) but utilising discrete wavelet transforms (DWT) also referred to as discrete wavelet decomposition. The proposed autonomous algorithm not only tackled the issue of the semi-autonomy of GPR data processing but rather significantly reduced the computational cost which is discussed in detail below.

## 3.5 Autonomous algorithm using Kalman filter and wavelet transform

This section of the chapter focuses on the autonomy of processing GPR data (i.e. objective 5 of this thesis) and accurate detecting and locating targets and ground anomalies making use of the discrete Kalman filter (DKF) including all integrated signal pre-processing filters discussed in the above sections. Furthermore, another most widely used signal processing technique, namely the wavelet transform, was adapted into the algorithm in correlation with a single objective optimisation technique making use of a genetic algorithm (GA) that replaced the MOGA-based optimisation approach which is used for the proposed semi-autonomous algorithm.

### 3.5.1 Overview of the autonomous algorithm

The proposed algorithm for autonomous processing of GPR data involves three main sections which are demonstrated as a flowchart in Figure 3-34.

Initially, utilising the developed Kalman filter and applying the desired filters. Secondly, adapting and utilising a single-objective optimisation algorithm, making use of evolutionary computing, which is chosen over conventional optimisation strategies given the complexity of the problem and the ability to generate the optimal solutions for the KF input parameters i.e. sensor and system noise covariances by minimising the associated cost function (mean of KF innovation). Thirdly, developing a sequentially running modified Kalman filter algorithms with integrated wavelet transform, as one of the most widely used signal processing tools, to investigate the GPR radargrams (B-scans) and distinguish noise, target, and anomaly signals, reducing noise and decluttering, and finally generating improved quality radargrams for a visual confirmation through the expert.

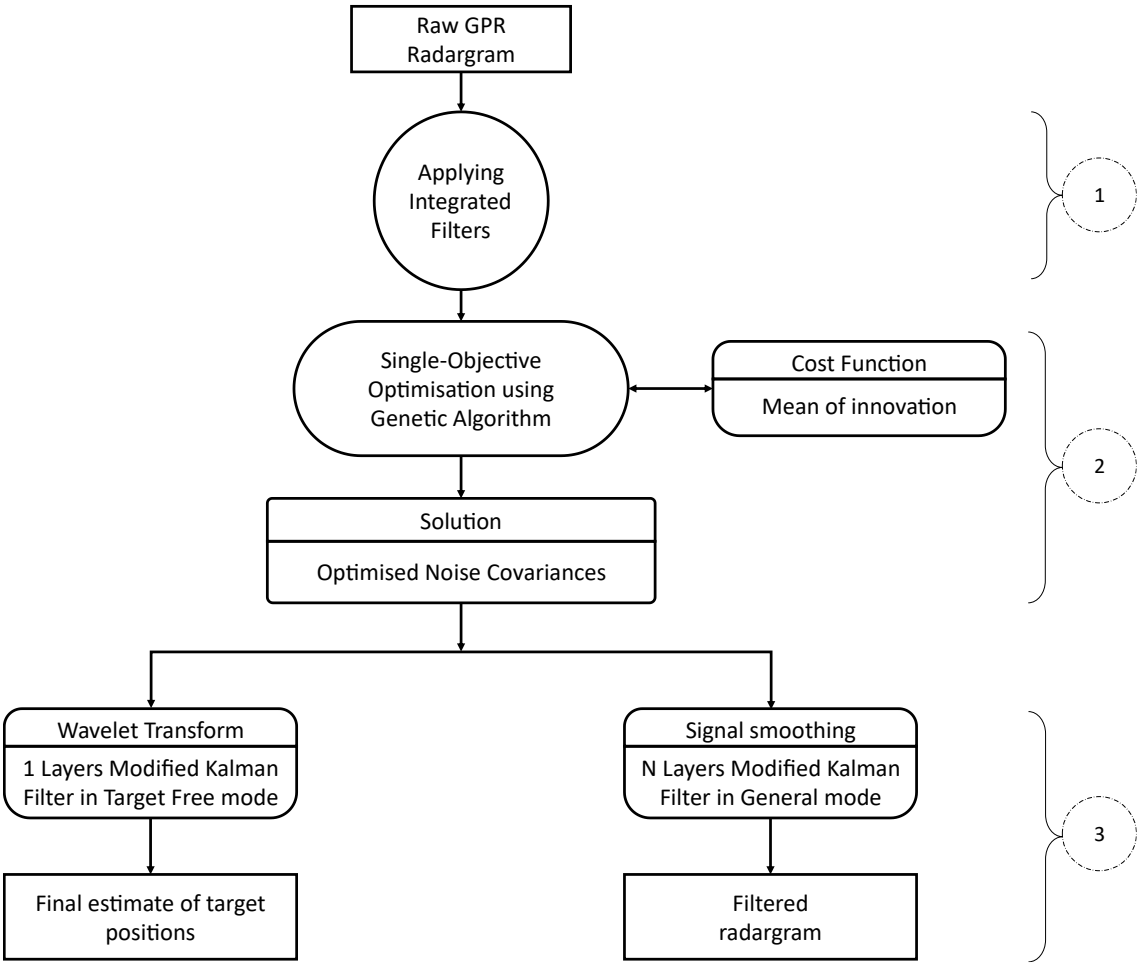



*Figure 3-34. Flowchart of the autonomous algorithm for processing GPR data*

To avoid repetition, the development of the proposed DKF for processing GPR data as well as the algorithm for detecting buried targets and ground anomalies will not be discussed in this section. If desired, readers can refer to section 3.4 of this chapter for more information. Instead, the following section will focus on discussing the autonomy of the proposed DKF algorithm by incorporating wavelet transforms, as well as optimising the proposed algorithm to enhance its performance by using a single-objective optimisation technique, specifically genetic algorithm.

### 3.5.2 Wavelet transform as a multi-resolution signal analysis

As discussed in the previous chapter (section 2.3.3), wavelet transform, also referred to as wavelet decomposition, is an adaptive method to analyse and characterise non-stationary signals, such as the signals with changing characteristics (e.g. frequencies) over time shown in Figure 3-35a or the highly non-stationary modulated time-domain NIS signal function produced in this study (Baili et al., 2009; Ni et al., 2010). The motivation for implementing wavelet transform in this thesis is to quantify and visualise the temporal dynamics of features and extract the frequency-specific (spectral-specific) information associated with the frequency-domain (Figure 3-35b) of a non-stationary signal (e.g. NIS signal) over time. In other words, by using wavelet transform, it provides a time-frequency analysis of the NIS signal by combining the information from both time-domain and frequency-domain investigations. In addition, performing time-frequency analysis, instead of frequency analysis, means performing the frequency analysis on short windows of time frames along the signal. This helps to increase the interpretability of the results due to the preservation of stationarity within those short windowed time frames (Cohen, 2014).

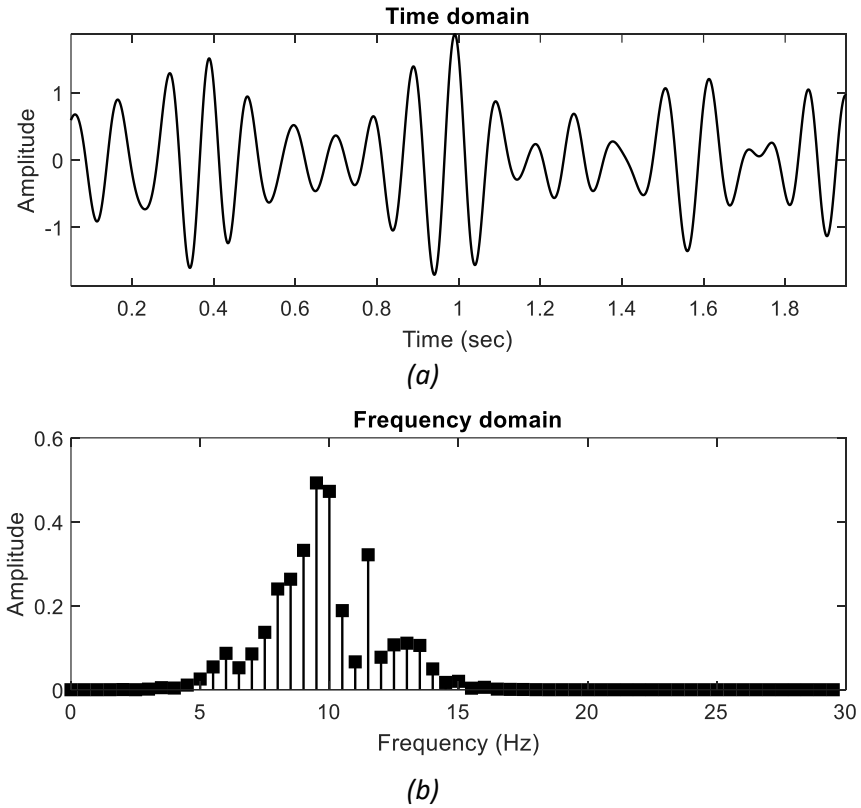



*Figure 3-35. An example of a) Time-domain non-stationary signal and b) the associated frequency-domain of the same signal*

As stated by Cohen (2014), a key difference between the Fourier transform (used for processing the NIS signal of the semi-autonomous algorithm) and the wavelet transform is the temporal specificity of the wavelet at each step. In other words, the Fourier transform at each step involve computing the dot product between the signal of interest (e.g. NIS signal) (Figure 3-37) and the Fourier transform kernel (e.g. a stationary sine wave with a constant amplitude over time) (Figure 3-36), which is also referred to as convolution. With this method, the entire signal is constantly weighted (Figure 3-38), and its output corresponds to frequencies. Therefore, when performing the Fourier transform, no temporal specificity can be expected.

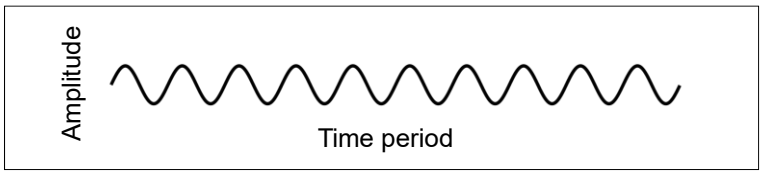



*Figure 3-36. Stationary sine wave with constant amplitude over time associated with a Fourier transform*

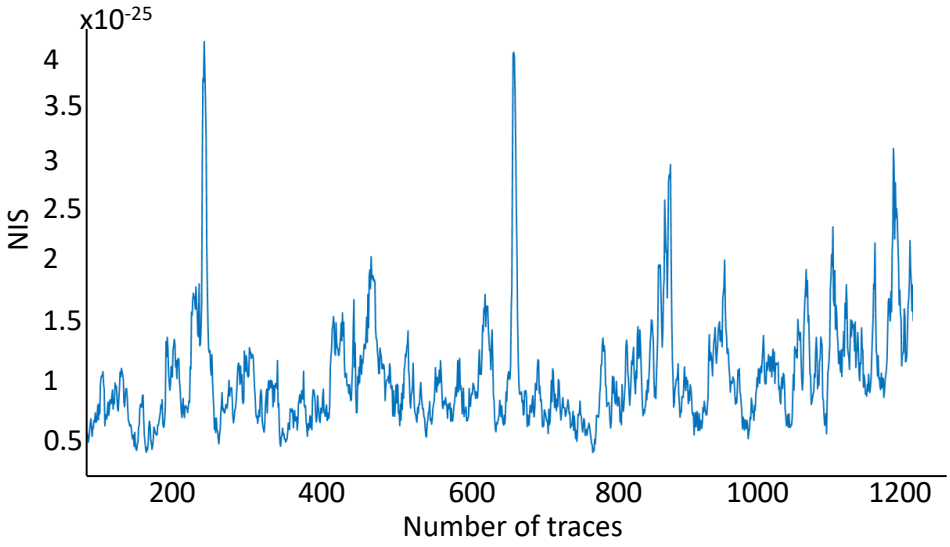



*Figure 3-37. Example of a non-stationary NIS signal (signal of interest) to be processed by Fourier and wavelet transforms*

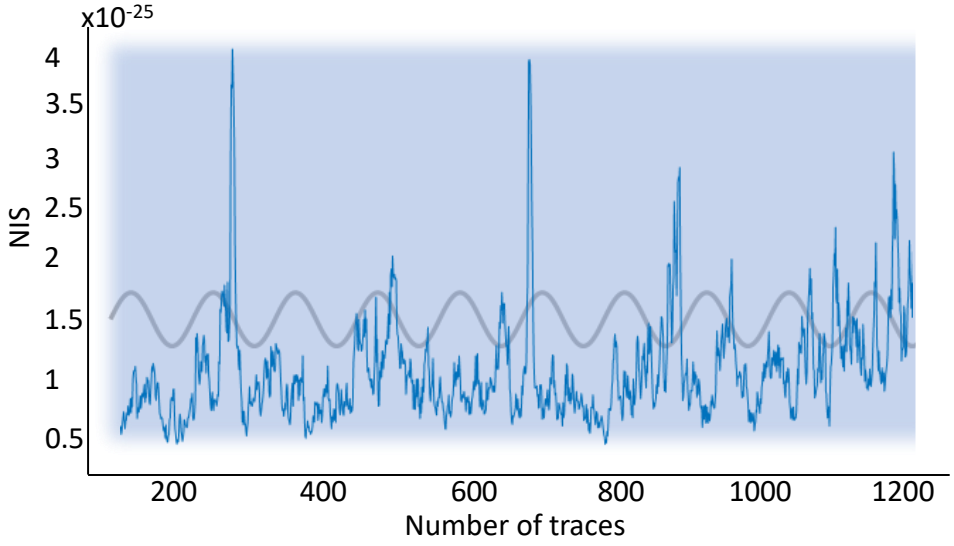



*Figure 3-38. The expected Fourier transform's weighing function (weighted area) for the NIS signal after computing the dot product between the signal and the sine wave*

While the convolution of the same signal of interest (Figure 3-37) and a wavelet kernel (mother wavelet), which is a wavelike function and typically tapered to zero at both ends by a Gaussian function, provides a temporal weighting function. The output corresponds to wavelet coefficients (2x2 matrix) identified by dilation (or scaling) and translation (or shifting) of the wavelet (three-dimensional amplitude as a function of dilation and translation). This will be discussed in more detail below. Figure 3-39 and Figure 3-40 illustrate an example of a Morlet wavelet, which is generally a sine wave tapered to zero by a Gaussian function (to introduce non-stationarity to the wavelet), and the expected temporal weighting function after convolution between the wavelet and the signal.

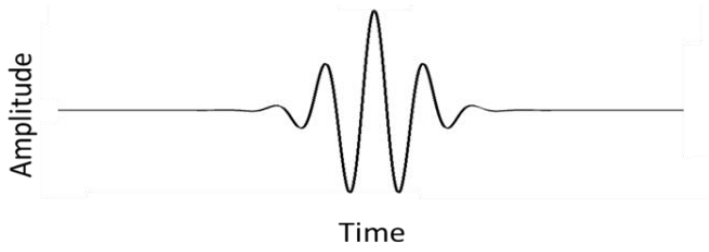



*Figure 3-39. Wavelet example (Morlet wavelet) with zero tapering at both ends*

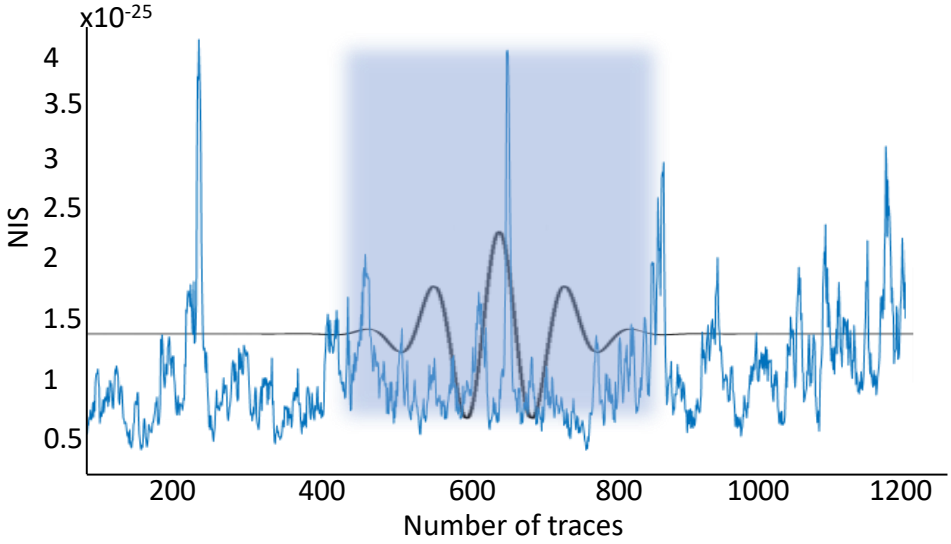



*Figure 3-40. The expected weighting function after computing the dot product between the NIS signal and a real-valued wavelet (e.g. Morlet wavelet here) yields spectrally specific information at the position of the wavelet.*

Wavelets are typically characterised by their compact support, which is the transient waveform of finite length between these zero points at both ends (Dondurur, 2018). Another important characteristic of wavelets is their zero mean oscillations, i.e. the area underneath the signal curve is zero (Foufoula-Georgiou and Kumar, 1994). This is a mandatory factor for performing a successful wavelet transform and cannot be compromised (Addison, 2017). For instance, the expected value after a wavelet transform of a flat signal (x = 1) is zero. Translating and scaling a selected wavelet kernel (mother wavelet) along a signal and simultaneously performing the convolution at each step results in extracting spectrally specific (localised) information about the signal at that point. It is worth mentioning that the wavelets are not limited to real-valued wavelets but can also be complex-valued (e.g. Figure 3-41) consisting of real and imaginary parts corresponding to sine and cosine waves, respectively.

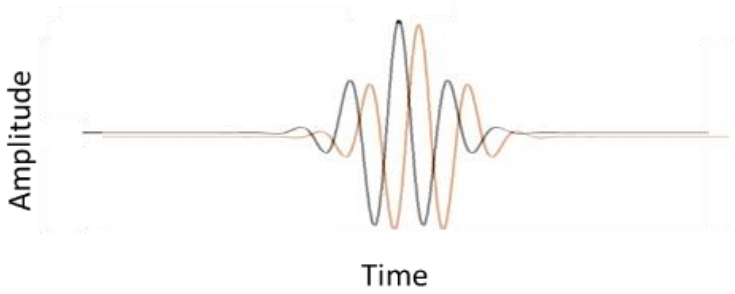



*Figure 3-41. An example of a complex wavelet consisting of zero tapered sine and cosine waves corresponded to real and imaginary parts*

The convolution of a signal and a complex-valued wavelet (Figure 3-42) will be a complex number consisting of real and imaginary parts (Figure 3-43). As a result, the convolution of complex-valued wavelets and signals at each point yields a time series of complex values corresponding to the translation and the scale of a wavelet at those points.

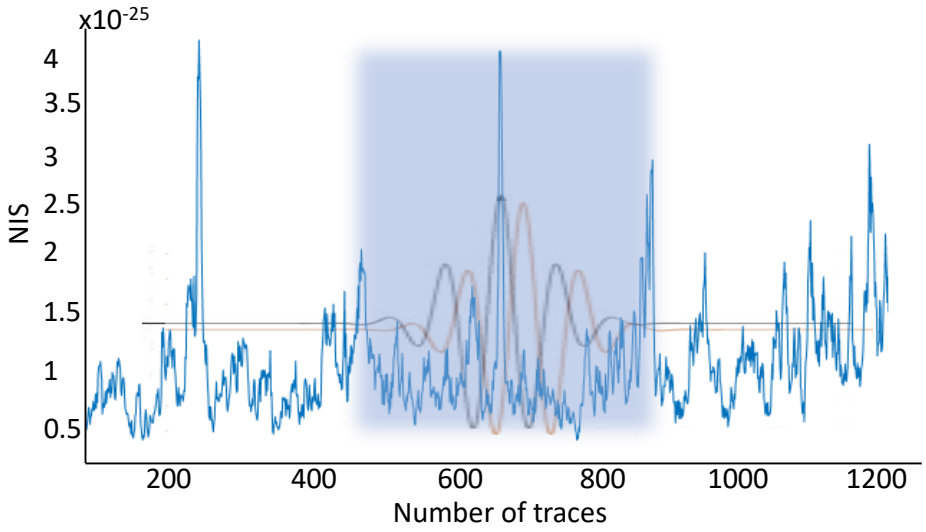



*Figure 3-42. The expected weighting function after computing the dot product between the NIS signal and a complex wavelet (e.g. Morlet wavelet here) yields spectrally specific information at the position of the wavelet.*

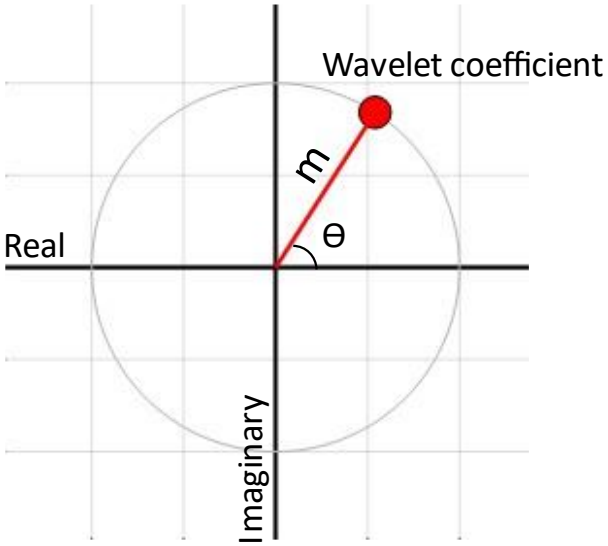



*Figure 3-43. An example of a wavelet coefficient consisting of real and imaginary parts*

This complex number is referred to as the wavelet coefficient, which is theoretically the convolution of the wavelet and a segment of the signal of interest (e.g. NIS signal) that represents the Euler notation ( $me^{i\theta}$ ) where $m$ is the amplitude, $\theta$ the phase angle of the signal (i.e. decoupling of both components) (Cohen, 2014). In other words, the amplitude, and the phase angle of the signal at each step can be extracted from wavelet coefficients

determined at these points. The values of the amplitudes and phase angles vary in correlation with translation (position) as well as the dilation (scaling value) of a wavelet along a signal of interest. The larger the amplitude of the wavelet coefficient, the closer the amplitude of the wavelet (centre point) to the amplitude of the signal of interest at that point and vice versa. In other words, the larger amplitude of wavelet coefficients the greater the similarity of the scaled wavelet to the signal of interest at that point and vice versa. For more detailed information and the theory of wavelet transform, the reader can refer to (Addison, 2017; Chui and Heil, 1992; Cohen, 2014; Daubechies, 1990; Mallat, 2009b; Percival and Walden, 2000b).

### 3.5.3 Analysing NIS signal using wavelet transform

The wavelet transform of a continuous signal (e.g. NIS-signal) in association with a problem-dependent chosen mother wavelet $\psi(t)$ –which is a localised basic wavelike function (kernel)– can be represented mathematically using (Eq.3-30 and Eq.3–31) (Addison, 2017).

Continuous Wavelet Transform function (CWT):

$$T_{a,b} = \frac{1}{\sqrt{a}} \int_{-\infty}^{+\infty} x(t) \psi \left( \frac{t-b}{a} \right) dt \qquad (3\text{-}30)$$

CWT Inverse function to reconstructing the decomposed signal (Ni et al., 2010):

$$x(t) = \frac{1}{C_g} \int_{-\infty}^{+\infty} \int_{0}^{+\infty} T_{a,b} \psi_{a,b}(t) \frac{d_a d_b}{a^2} = \sum_{a=-\infty}^{+\infty} \sum_{b=-\infty}^{+\infty} T_{a,b}\, \psi_{a,b} \qquad (3\text{-}31)$$

Where *a* and *b* are dilation (or scaling) and translation (or shifting) parameters respectively, $x(t)$ is the input signal which (i.e. NIS-signal in this study), and $\psi(t)$ is the mother wavelet that can be chosen depending on the problem from a variety of existing mother wavelets that will be discussed below in section 3.5.4 of this chapter.

The discrete wavelet transform (DWT), on the other hand, is generally defined as a discretised version of CWT which only differs by the values of dilation and translation parameters $a$ and $b$ (*Eq.3-30*), where both can take any real number in CWT ($a, b \in \mathbb{R}$) but limited to only $2^l$ and $k2^l$ values, respectively when implementing the DWT (Baili et al., 2009; Chui and Heil, 1992; Ni et al., 2010). Where $k \in \mathbb{Z}$ , and $l$ indicates the selected level of signal decomposition

In this thesis, the discrete wavelet transform (DWT) was utilised to process and characterise the NIS signal due to its efficiency in signal denoising problems and offering a shorter processing time and hence lower computational cost (Baili et al., 2009).

The coefficient $T(a, b)$ at a specific decomposition level ($l$) can be expressed hierarchically as shown in Figure 3-44 and is referred to as multiresolution analysis through separated coefficients namely approximation coefficient $A_l$, and detail coefficient $D_l$ as shown in Eq.3-32 to Eq.3-34 (Addison, 2017; Baili et al., 2009).

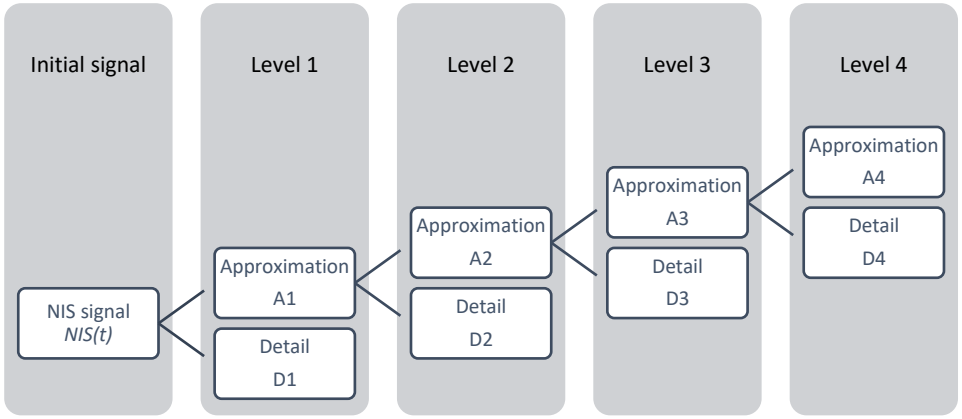



*Figure 3-44. Example of multiresolution analysis of NIS signal (level=4)*

Approximation coefficient in general form for input signal $x(t)$:

$$A_l(t) = \sum_{m=0}^{+\infty} x(m)\, \phi_{l,k}(m) = \sum_{m=0}^{+\infty} x(m) \frac{1}{\sqrt{2^l}} \phi\left(\frac{i-k2^l}{2^l}\right) \quad (3\text{-}32)$$

Where $\phi$ is the scaling function corresponding to the selected mother wavelet.

Detail coefficient in general form for input signal $x(t)$:

$$D_l = \sum_{m=0}^{+\infty} x(m) \frac{1}{\sqrt{2^l}} \psi\left(\frac{i-k2^l}{2^l}\right) \quad (3\text{-}33)$$

Where $x(m)$ is the input signal i.e. the generated NIS signal within the KF in this study $[NIS(m)]$, $\psi$ is the mother wavelet, $l$ is the chosen signal decomposition level and $k \in \mathbb{Z}$

The NIS signal can be represented using approximation and detail coefficients as (Eq.3-34):

$$NIS(t) = A_l(t) + \sum_{m=0}^{+\infty} D_l(t) \quad (3\text{-}34)$$

As mentioned above, to characterise and extract the spectrally specific information about a signal (time-frequency analysis), there are infinite number of various mother wavelets (kernels) that can be selected for performing the convolution procedure. In addition, each of these mother wavelets is characterised by its compact support.

## 3.5.4 Selection of mother wavelets

Nominating a mother wavelet $\psi(t)$ for implementation in the signal analysis is an arbitrary task on one hand, but problem dependant and critical on the other hand (Percival and Walden, 2000a). Wavelets can be chosen based on their characteristics in correlation with the problem's objective (Bishop et al., 2020). In other words, wavelets have to be chosen based on the characteristics of the measured (or created) signal of interest (e.g. NIS signal), the type of required information about the signal, as well as the expected correlation quality of the selected wavelet and the signal of interest (Addison, 2017). It is worth mentioning that the complexity of a wavelet increases by increasing the number of vanishing moments. Such an increase in complexity can be advantageous for accurately representing a complex signal. Similarly, the regularity of wavelets (smoothness) is proportional to the number of vanishing

moments i.e. the higher the number of vanishing moments the higher the regularity of a wavelet (Dondurur, 2018).

One aspect of selecting a suitable wavelet is the selectivity in frequency of the wavelet. It was shown that the semi-autonomous algorithm in the previous section of this thesis was a highly selective approach in terms of frequencies (frequency range) of the performed Fourier transform and typically, one frequency will be identified for each analysing function. By the wavelet transforms, however, typically a range of various frequencies are encompassed in these mother wavelets (kernels) which is referred to as the selectivity of wavelets. Higher selectivity is proportional with a lower frequency range and a higher scale for wavelet (less compact support) and lower resolution in time and vice versa. Comparably, utilising a high-frequency wavelet (low selectivity) with very compact support (low scale) can provide high resolutions in time but a low resolution in frequency and vice versa. This principle is referred to as Heisenberg's uncertainty principle (Cohen, 2014).

On this basis, due to the uniqueness of the proposed approach in this thesis and hence, the lack of information about the performance of each of the existing wavelets in correlation with the Kalman filter, the wavelet transforms for processing the KF's NIS signal in this study was performed comprehensively using various DWT associated mother wavelets including the well-known mother wavelets as such the Daubechies (Figure 3-45a), Symlet (Figure 3-45b), and Discrete Meyer (Figure 3-45c) by utilising different vanishing moments for each of those mother wavelets. It is worth mentioning that the scaling functions associated with each of these mother wavelets, also referred to as father wavelets (in association with the decomposition level of the approximated part of the signal), are used to determine lower frequency components of a signal (i.e. low pass filter), while mother-wavelets are dealing with

higher frequency components (i.e. a series of high pass filters in association with the decomposition level and detail parts of the signal).

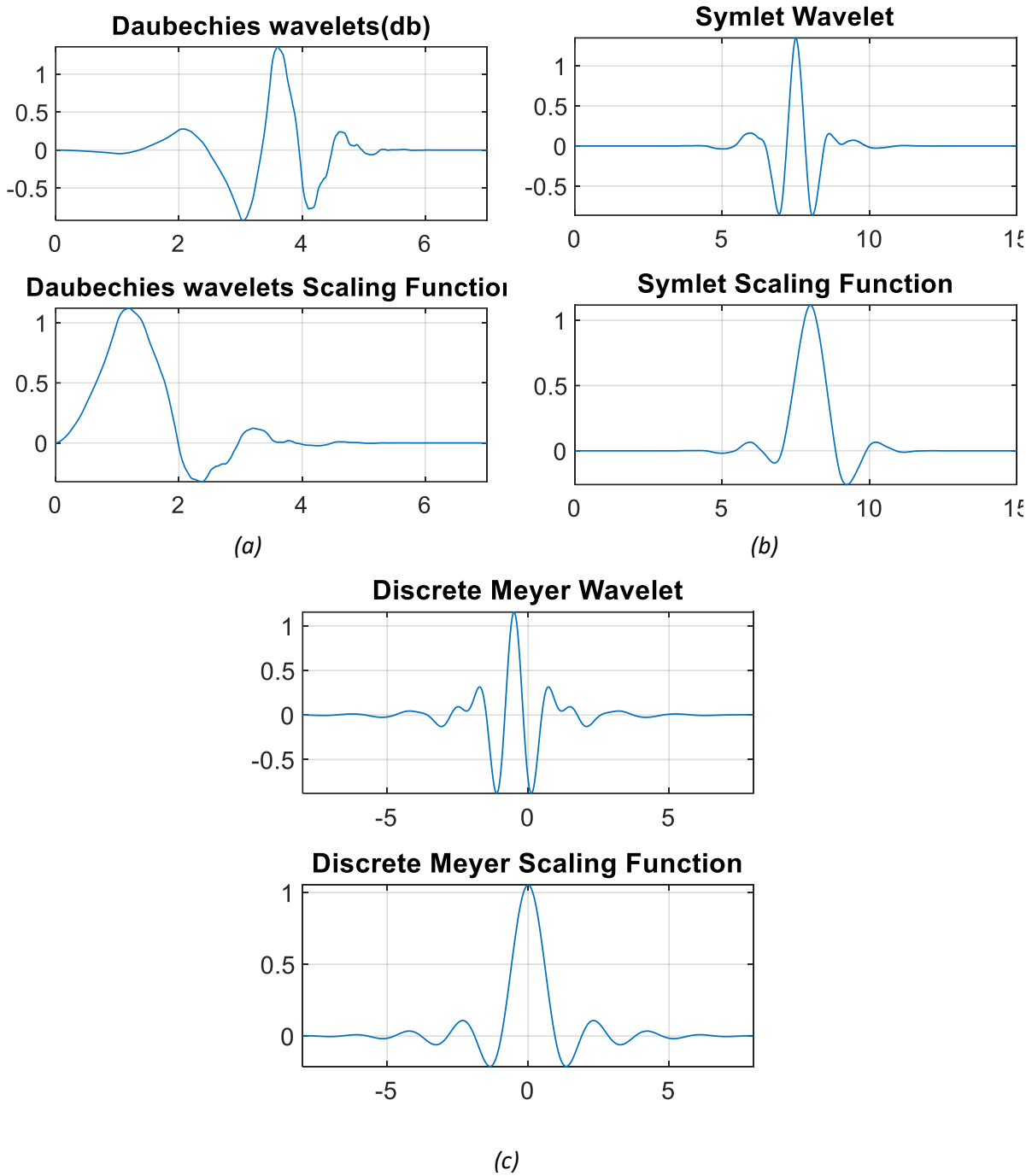



*Figure 3-45. Examples of evaluated wavelets and scaling functions: a) Daubechies, b) Symlet, c) Discrete Meyer*

In addition, to achieve a satisfactory separation level of the target signal and the background signal, different decomposition levels were used for processing the radargrams. A decomposition level range of ( $l \in [3 \;\; 7]$ ) was found out to be suitable as the results obtained by other decomposition levels outside this range were either not tangible when $l > 7$ with minimum (or no) changes in the processed NIS signal and a very high false alarm ratio, or were exaggerated in letting the frequencies passing through when $l < 3$ and missing all anomalies in the NIS signal. The performance of utilised mother wavelets is evaluated by the accuracy in identified target positions and achieving a good correlation between nominated wavelets and the NIS signal by satisfying the admissibility condition $C_g$ presented in Eq.3-35 that guarantees a zero-mean mother wavelet as a mandatory factor for the success of wavelet transform, as discussed earlier (Addison, 2017).

$$C_g = \int_0^{\infty} \frac{\left|\hat{f}(NIS(t))\right|^2}{NIS(t)} dNIS < \infty \qquad (3\text{-}35)$$

Where, $\hat{f}(NIS(t))$ is the Fourier transform of the input signal $NIS(t)$.

To determine the performance of each utilised DTW mother wavelets (i.e. db, Sym, and dmey) in this research, statistical analysis was performed. The area under the ROC curves (AuC) is utilised as a benchmark for statistical analysis to determine the performance of each of those mother wavelets (different vanishing moments and decomposition levels). Utilising AuC of ROC as a performance evaluation benchmark is the subject of different published literature (Carter et al., 2016; González-Huici et al., 2014; Oberti et al., 1999; Yang and Liu, 2013). The true positions of buried pipes on selected GPR datasets were manually identified to be fed into the ROC analysis. On this basis, the above-mentioned mother wavelets were utilised individually and the corresponding AUCs of ROC curves were calculated. As per the definition of AuC of ROC, higher valued ROC indicated mother wavelets with higher performance.

The novel incorporation and utilising the wavelet transform in the proposed discrete Kalman filter algorithm for processing the GPR data has eliminated the necessity to satisfy multiple objectives (associated with the FFT and ROC curves) and hence perform a MOGA optimisation procedure to determine an optimal value for the KF's input parameters. On this basis, a single objective optimisation approach is introduced resulting in a significantly lower computational cost, which will be discussed below.

### 3.5.5 Optimisation algorithm using single-objective optimisation technique

To determine an optimised initial value of both, system, and sensor noise covariances (**Q** and **R**), as the only unknown input variables in this KF problem, an optimisation technique was conducted making use of machine learning. Particularly, the genetic algorithm (GA) was selected for this study due to its efficiency in finding global solutions to such complex problems. The cost function associated with the problem in this study was defined based on a fundamental characteristic of the KF, namely the necessity of zero mean *innovations* for delivering accurate estimates and high performance of the KF (Bar-Shalom et al., 2001; Simon, 2006). Thereby, an optimised value for system and sensor noise covariances (**Q** and **R**) can be determined in correlation with a guaranteed minimised mean of *innovation.* To fulfil this aim, the KF initially operates under the assumption of a target-free state due to a significantly lower number of signals associated with background signals in GPR datasets used for this study (a typical scenario for most of the radargrams). However, this assumption can vary under different circumstances if the target signals dominate the background signal i.e. the ratio of target signals to background signal is remarkably high. For instance, a radargram consists of a long metallic plate with associated ringing that is largely distributed below the plate.

The number of parameters that need to be optimised for minimising the mean of *innovation* vectors is proportional to the number of sampling points ($N$) in a radargram. In other words, for a radargram consisting of $N$ sampling points (e.g. 510 in this study), $N + N$ parameters associated with **Q** and **R**, individually, need to be optimised (e.g. 510+510 parameters in this study) by employing $N$ optimisation algorithms, simultaneously. This can cause a high computational cost. This issue, however, can be addressed by computing a single preference cost function (i.e. a linear weighted sum of these $N$ parameters) using a mathematical scalarisation method called the weighted sum method (WSM) (Alizadeh et al., 2019).

On this basis, for the ease of calculation and reduction of computational cost, the *innovation* vectors (Eq.3-36) were scalarised by executing the WSM. The corresponding cost function $(J)$ can be expressed as in Eq.3-37 that needs to be minimised:

$$\boldsymbol{\nu} = \xi_k - \mathbf{H}\mathbf{A}\hat{x}_k^- \tag{3-36}$$

$$J = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{1}{k}\sum_{j=1}^{k}\boldsymbol{\nu}_{ij}^2\right) = 0 \tag{3-37}$$

Where $(\boldsymbol{\nu})$ is the innovation, $(\xi_k)$ the input signal, $(\hat{x}_k^-)$ a priori estimate, $(N)$ sampling point, and $(k)$ the trace number. The number of optimised parameters for this study was 1020 (i.e. 510 parameters for Q and R, individually).

The optimal configuration of the genetic algorithm (GA) parameters, including population size, number of generations, and probability of mutations, were determined through a series of trials and errors. Based on these findings, it is suggested that a population size of 30 and a generation size of 100 individuals be utilised, along with the standard parameter values for MATLAB (v2021a), specifically a mutation rate of 0.01 and a crossover ratio of 0.07, for conducting the optimisation. The computational time for the optimisation process, performed on a personal computer equipped with an Intel(R) Core (TM) i7-3770 CPU @ 3.40GHz and 32

GB of RAM, was found to be 00:27 hours. This is significantly lower in comparison to that of the Multi-Objective Genetic Algorithm (MOGA) optimisation for the semi-autonomous algorithm.

As demonstrated in this section, the proposed algorithm was successful in achieving autonomy (i.e. objective 5 of this thesis) by incorporating the DWT into the proposed DKF algorithm in GPR applications. This was done to reduce noise, declutter, and accurately detect and locate buried targets (e.g. pipes) and potential ground anomalies (e.g. soil discontinuities) on GPR radargrams. As previously discussed in section 2.4 of the previous chapter, each of the existing non-destructive sensing techniques employed for exploring the subsurface, including GPR, are associated with limitations that need to be considered before conducting a survey. However, it was also discussed that the fusion of data from these techniques, when used in isolation, is a smart way to tackle and overcome such limitations in order to increase data quality and subsequently enhance interpretation accuracy.

On this basis, in correlation with objective 6 of this thesis, the proposed autonomous Kalman filter (KF)-based algorithm was utilised to develop a novel fusion algorithm for integrating data obtained from multiple sensors, specifically GPR, Electromagnetic Locating (EML), and Electromagnetic Conductivity (EMC) data. This will be discussed in further detail in the upcoming sections. This approach was chosen due to the multi-sensing capabilities of the Kalman filter, which were discussed in section 2.5 of the previous chapter.

## 3.6 Experiments and data collection

To assess the performance of the proposed computational solutions and to satisfy objectives 6 and 7 of this research, customised experiments were set up in the National Buried

Infrastructure Facility (NBIF) at the University of Birmingham. The experiments replicated scenarios where the spatial location of different buried pipes were to be determined.

The test pit in NBIF comprised a concrete lined pit of 10 m by 5.8 m in plan and 5 m deep in which pipes of different sizes and materials were buried in compacted sand. The details of the material parameters, i.e. average dry density, average moisture content, density, diameter, thickness, Young's modulus, and Poisson's ratio, are presented below in Table 3-5.

*Table 3-5. Parameters of the soil and buried pipes in the NBIF experiments*

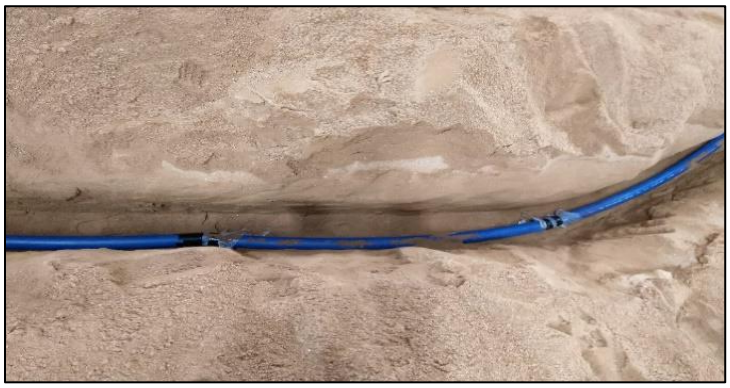
| Material | Parameter | Value | Unit |
|---|---|---|---|
| **Soil (Sand)** | Average Dry Density | 1680 | $kg/m^3$ |
| | Average Moisture Content | 5.27 | % |
| | Young's Modulus | 80 | MPa |
| | Poisson's ratio | 0.35 | - |
| | Friction angle | 35 | $^o$ |
| **Large Pipe (PE 63 HDPE)** | Diameter | 2 | m |
| | Density | 955 | $kg/m^3$ |
| | Young's Modulus | 1 | GPa |
| | Poisson's ratio | 0.42 | - |
| **Void (Gym Ball)** | Diameter | 1.2 | m |
| **Small Water Pipe (MDPE)** | Diameter | 0.1 | m |
| | Density | 1380 | $kg/m^3$ |
| | Young's Modulus | 3 | GPa |
| | Poisson's ratio | 0.42 | - |
| | Thickness | 0.01 | m |
| **Small Metallic Pipe (galvanised steel scaffold)** | Diameter | 0.048.3 | m |
| | Density | 7750 | $kg/m^3$ |
| | Young's Modulus | 210000 | $N/mm^2$ |
| | Thickness | 0.04 | m |

A high-density polyethylene (HDPE) pipe of 2 m diameter, i.e. the large pipe, was positioned roughly at the centre of the pit (4.7 m from the south edge of the pit and 5.3 m from the north edge of the pit) and the void (a gym ball with 1 m diameter) was buried adjacent to the HDPE

pipe with cover depths of 0.3 m and 0.17 m, respectively. It is worth mentioning that the HDPE pipe and the void were initially buried in connection with other research projects in NBIF, which was afterwards used for the experiment in this research. Figure 3- 46 shows the test pit partially filled with compacted sand and the HDPE pipe. The MDPE water pipe (Figure 3-47a) and the galvanised steel pipe (Figure 3-47b) were positioned to the left- and right-hand sides of the HDPE pipe at a 2 m and 2.5 m distance and burial depths (from the surface to the invert of pipes) of 1 m and 0.3 m, respectively.



*Figure 3-46. The test pit at NBIF and buried HDPE pipe*



*(a)*



*(b)*

*Figure 3-47. Buried pipes in the pit (NBIF), a) PVC water pipe buried at 1 m depth, b) Galvanised steel pipe buried at 0.3 m depth*

Figure 3-48 shows a cross-section through the NBIF pit illustrating the configuration of buried pipes and a void.

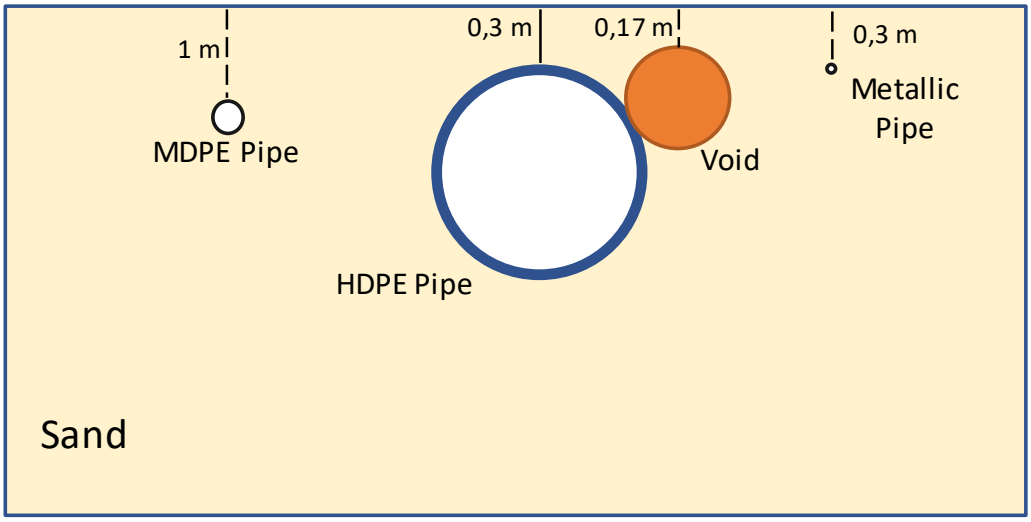



*Figure 3-48. Cross- section showing the configuration of buried pipes and void in the test pit in NBIF*

### 3.6.1 Experiment setup: EML and GPR

To increase the accuracy of the outcome, the exact position of the buried pipes and the void was surveyed using a Leica TS15 total station device. Figure 3-49 shows the surveyed points consisting of the pipe centre points (crown), the pit corners, and the gridline points for EML and GPR surveys. This survey datapoints were measured relative to a local stationary point in NBIF (red rectangle).

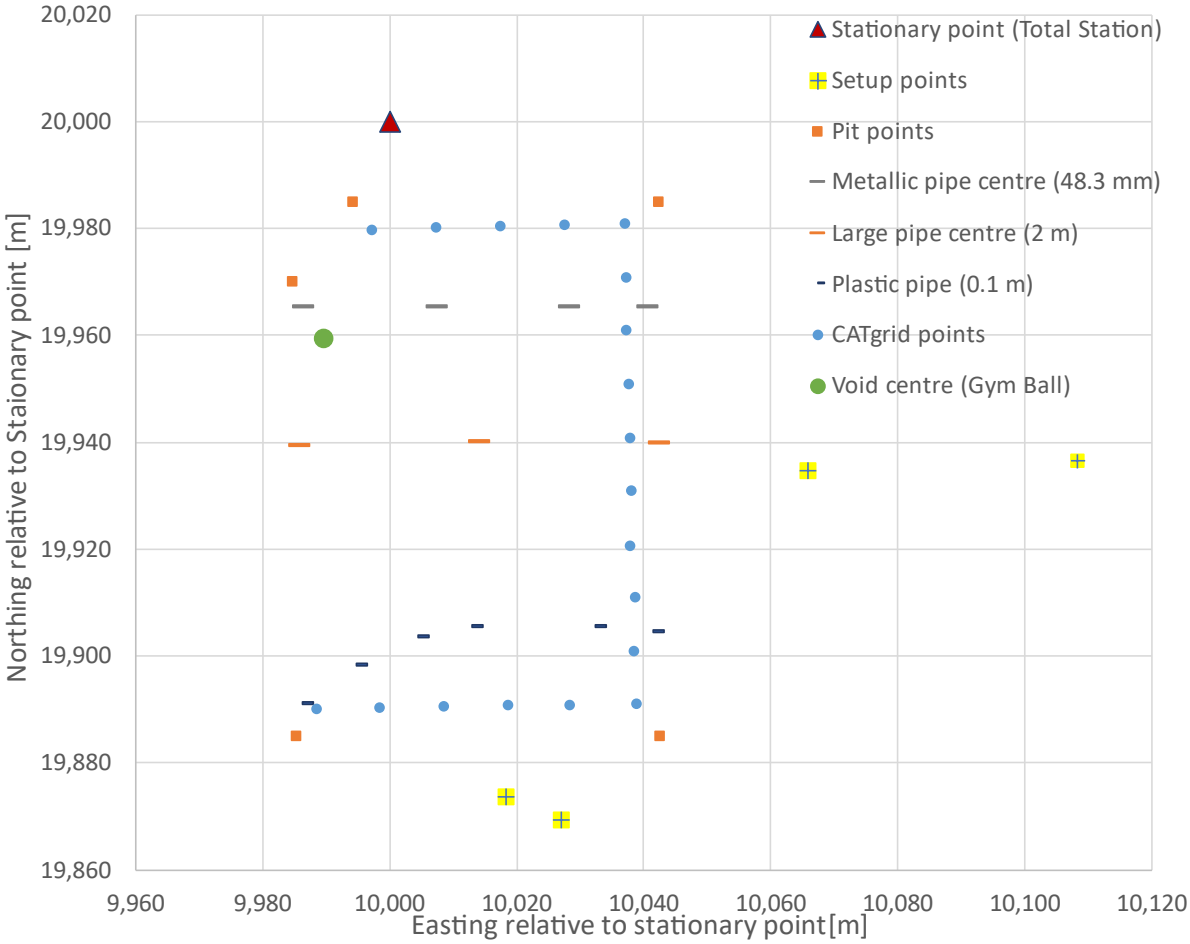



*Figure 3-49 Experiment setup in NBIF – surveyed points.*

### 3.6.2 CAT and Genny survey

An EML survey was conducted using an RD8200G EM-locator (Radiodetection) with the capability of logging data such as signal strength, estimated depth, current, CD phase, and

locator position. Measurement points were defined at specific distances shown in Table 3-6 by installing gridlines as shown in Figure 3-50.

It is worth mentioning that the appropriate gridline spacing for a GPR survey will depend on the specific application and the resolution required for the survey. In general, a closer gridline spacing will result in higher resolution data, but it will also require more time and resources (Daniels, 2000). Conversely, a wider gridline spacing will result in lower resolution data but will be faster and require fewer resources. Specification for underground utility detection, verification and location, PAS128, (2014) provides suggestions for gridline spacing pertinent to the application. For most applications, a gridline spacing of between 0.5 m and 2 m is appropriate. However, in some cases, a spacing of as close as 0.1 m may be required (Bigman, 2012). On the other hand, if the objective is to cover large areas or detect deep targets, a wider spacing of 5 m or 10 m may be appropriate. It is also important to note that the frequency of the radar antenna and the dielectric constant of the subsurface materials also plays an important role in determining the appropriate gridline spacing. In general, for high-resolution imaging of shallow subsurface features, a high-frequency antenna should be used with a close gridline spacing. Whereas for deep penetration and imaging of buried objects, a low-frequency antenna should be used in combination with wider gridline spacings (Bigman, 2012; Daniels, 2000; Utsi, 2017a).

Given these considerations, a gridline spacing of 0.5m has been chosen for the survey based on the above-mentioned ground conditions and utilising the IDS Stream C and IDS DUO GPR devices. The IDS Stream C with a 600 MHz, and the IDS DUO with dual 250 MHz and 700 MHz central frequencies. Furthermore, the scan width of the IDS stream C is 82.5 cm which allows a sufficient overlap of scans. This is also in line with the 40 x 50 cm antenna footprint of the IDS DUO device, which was the second GPR device used for this research.

*Table 3-6. Gridline set up and measurement point distances*

| Gridlines | Survey Direction | Point distance |
|---|---|---|
| Long Gridline | North to South | every 0.5 m |
| Short Gridline | East to West | every 1 m |
| Short gridline over the metallic pipe (additional) | East to West | every 1 m |




*Figure 3-50. Schematic showing the gridlines for the EML and GPR surveys*

Measurements were taken using both passive and active modes of the EM-locator. Initially, attempts were made to take measurements using the passive mode (without the use of a Genny) but were unsuccessful. Subsequently, measurements were taken using the active mode in which a Genny was connected to the metallic pipe through a direct connection (as shown in Figure 3-51), which resulted in successful measurements. The measurements were

taken at specified points, with a frequency rate of 33 kHz and a gain value of 96, and the resulting signal strengths were recorded and transmitted via Bluetooth for further processing.



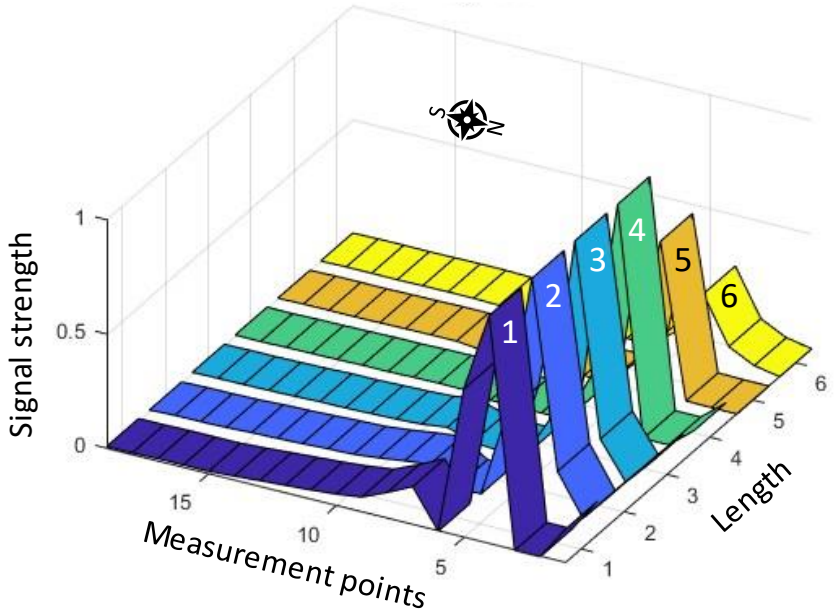
*Figure 3-51. Direct connection of the Genny with the buried metallic pipe*

182 measurements were performed using the active mode with a direct connection of the Genny, at a spacing of every 0.5 m on the longer gridlines. Figure 3-52 shows an example of a measurement taken directly over the metallic pipe, with a detection rate of 99.9%, indicating the direction of the buried pipe and an estimated depth of 0.3 m.

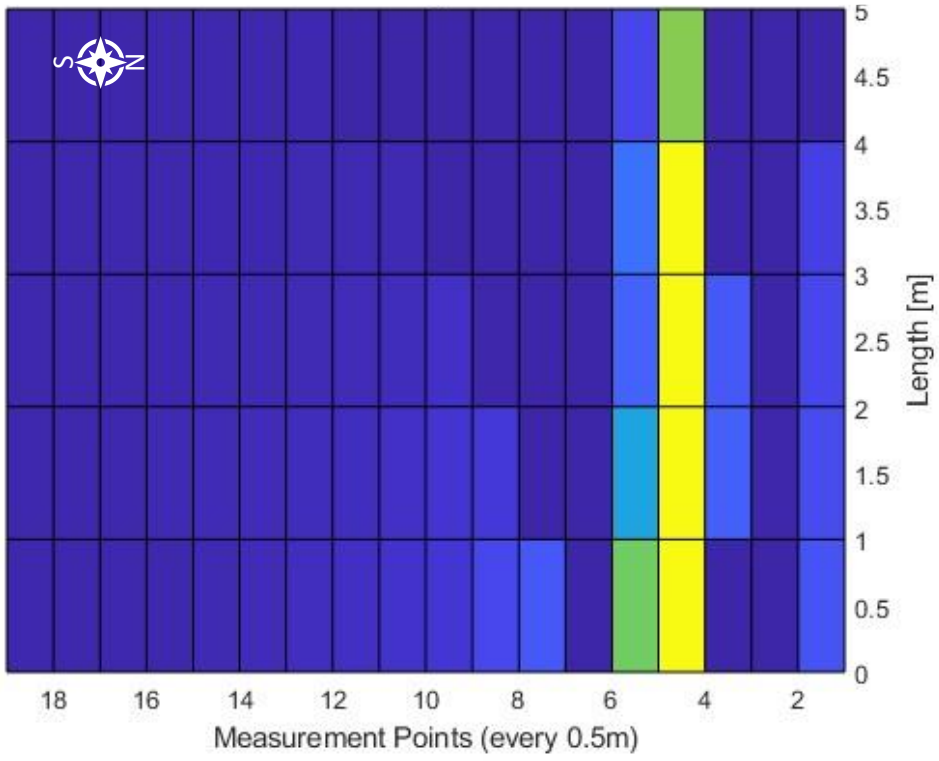



*Figure 3-52. Example of an EM locator measurement using the RD8200G device with direct Genny connection to the metallic pipe*

The resulting signals reconstructed from the measured signal strengths at each measurement point of the surveyed area are presented in Figures 3-53(a) and (b).




*(a)*




*(b)*

*Figure 3-53. Visualised signal of the EM locating survey as: a) strips to replicate the measurement direction, b) top view in which the buried pipe can be identified.*

The initial observation of the reconstructed signals (Figure 3-53a) and the overlayed interpolated signals in Figure 3-54 showed an increase in the signal amplitudes at two marked positions (1 and 2 in Figure 3-54) adjacent to the signal amplitudes associated with the position of the metallic pipe (between 1 and 2 in Figure 3-54). Furthermore, a descending trend can be observed among the six signals (1 to 6 in Figure 3-53a) at the level of the highest measured signal strengths (peaks) with increasing distance from the origin (first gridline). It was found that the cause of this issue was the proximity of the Genny device to the pipe location. Through trial-and-error experiments, it was determined that keeping a distance of approximately 3 meters between the survey area and the Genny device can solve this problem. Additionally, another amplitude decay can be observed for the signals with increasing length of the metallic pipe. As discussed in the literature review chapter, this is a general issue associated with the leakage of flowing electrical currents from a conductor into the ground as it travels farther away from the transmitter (Genny).

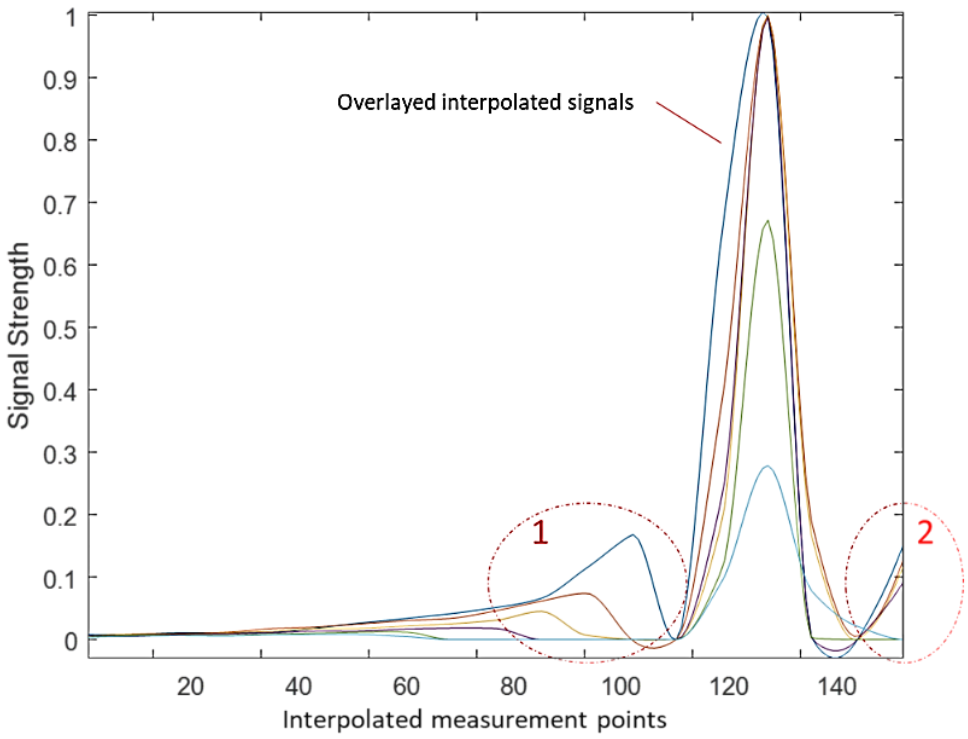



*Figure 3-54. Overlayed interpolated signals reconstructed from EML measurements*

The depth estimated by the EM-locator, as discussed in section 2.6.1.5 of the literature review chapter, is proportional to the signal strength measured by the top and bottom search coils of the device. The strength of a signal, on the other hand, is also proportional to the position of the locator relative to the pipe centre. It was mentioned that external distortion factors such as adjacent utilities, cars, and generators might affect the quality of signal measurements (Radiodetection, 2018; Siu and Lai, 2019; VIVAX, 2011). Therefore, the measured signals in this experiment were considered as noise-free signals, as the test was conducted under laboratory conditions, and with the absence of external distortion factors.

### 3.6.3 GPR surveys

The GPR data used in this research were obtained from a series of measurements taken using two georadar GPR antennas, i.e. an IDS DOU (dual central frequencies of 250 and 700 MHz), and a Multi-channel IDS Stream-C (central frequency of 600 MHz), shown in Figure 3-55. The scan spacing was set to 0.5 m and 1 m for long gridlines and short ones, respectively. The sampling time window was set to 120 ns distributed over 512 sampling points for each trace. Hyperbola fitting using the GPR antennas indicated an approximate value of 8 for the dielectric permittivity of the ground.



*(a)* *(b)*

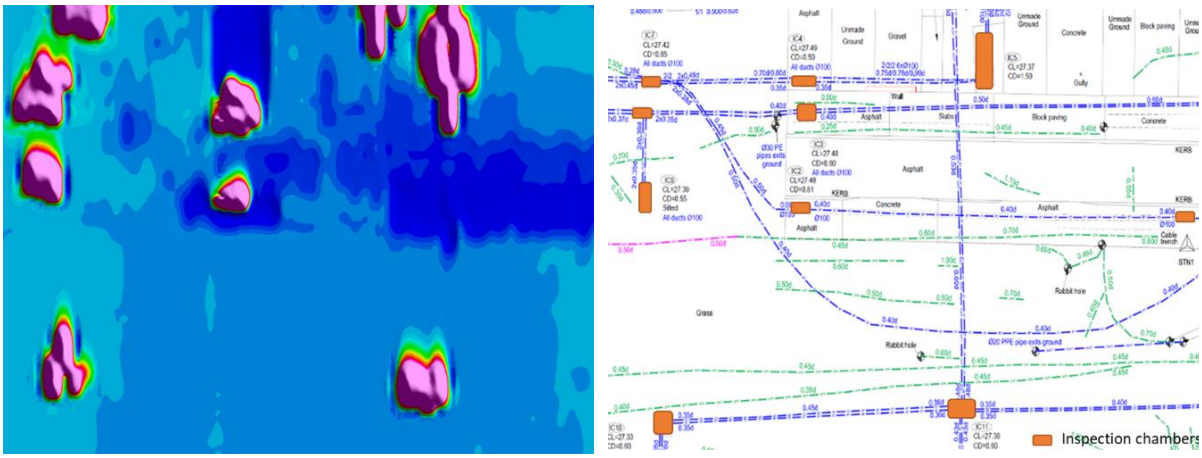
*Figure 3-55. Utilised GPR antennas for collecting data; a) IDS duo, b) Multi-channel Stream C.*

## 3.6.4 EMC and GPR data provided by a third party

### 3.6.4.1 EMC data

Concerning the EMC data, it is worth mentioning that after meeting with the Estates Department of the University of Birmingham, due to COVID-19 restrictions and national lockdown during this PhD project and concerns about infection risks, it was decided to use existing data provided by third parties instead of conducting experiments. On this basis, data was provided by the RSK group. The provided datasets came from GPR and EM conductivity surveys from a different trial site associated with a research project called FINDIT (Metje et al., 2022). Moreover, EMC data could not be collected in NBIF due to a high volume of conductive materials (e.g. concrete reinforcing bars of the pit walls, metallic structures, and metallic equipment around the pit). Figures 3-56(a) and (b) show the conductivity map of the surveyed area of the FINDIT trial site and the plan of the site, respectively, for the data provided by the third party. The dominance of the existing inspection chambers with

significantly higher apparent conductivities compared to the buried utilities at close distances can be observed.




(a) (b)

*Figure 3-56. a) visualisation of the EMC survey indicating the dominance of inspection chambers with relatively higher apparent conductivities compared to buried utilities. b) the plan of the site showing all the buried utilities some of which run through inspection chambers (coloured blocks)*

### 3.6.4.2 GPR data

The GPR data provided by the third party used in this study were obtained from a series of measurements taken from an accurately constructed trial site, which covered 23 scan lines across an 11 m x 13 m area with five linear pipes installed at shallow depths (Figure 3-26 on page 139). GPR data were recorded in DZT GGSI format using a single channel antenna with a centre frequency of 400 MHz. Scan line spacing was 0.5 m with lines located perpendicular to the pipes. The ground condition was tested and confirmed as a 6.1 m thick glacial sand and gravel over red crag sand. Hyperbola matching in the GPR indicated an approximate value of 6.5 for the dielectric constant of the sand.

## 3.7 Fusion of GPR, EML, and EMC data

In this section, the procedure of Kalman Filtering based fusion of NDT techniques i.e. GPR, EML, and EMC data will be discussed which satisfies the Objective 6 of this thesis. The following subsections include the alignment and correlation of associated data, inversion of EMC data, Fusion of GPR and EMC data, Fusion of experimentally obtained GPR and EML data, and finally fusion of EML and GPR data for a more complicated scenario where these techniques reach their limitations making use of numerical modelling and employing the genetic algorithm.

### 3.7.1 Multiple sensors data alignment and data correlation

Typically, multi-sensing data (i.e. data from multiple sensors) reveal variations in properties e.g. geometric property due to variations in measurement types and sometimes data collection at different points during a survey. Therefore, it is vital to match such properties temporally and spatially as the initial step of data fusion through alignment and correlating steps (Figure 3-57) referred to as co-registration of data to ensure a proper connection between such data is achieved (Schmitt and Zhu, 2016). This aim can be fulfilled by adjustments of units and transformation of coordinates (i.e. data alignment) and then associating the measurements from multiple sensors (i.e. data correlation) concerning the object (targets) of interest (Schmitt and Zhu, 2016).

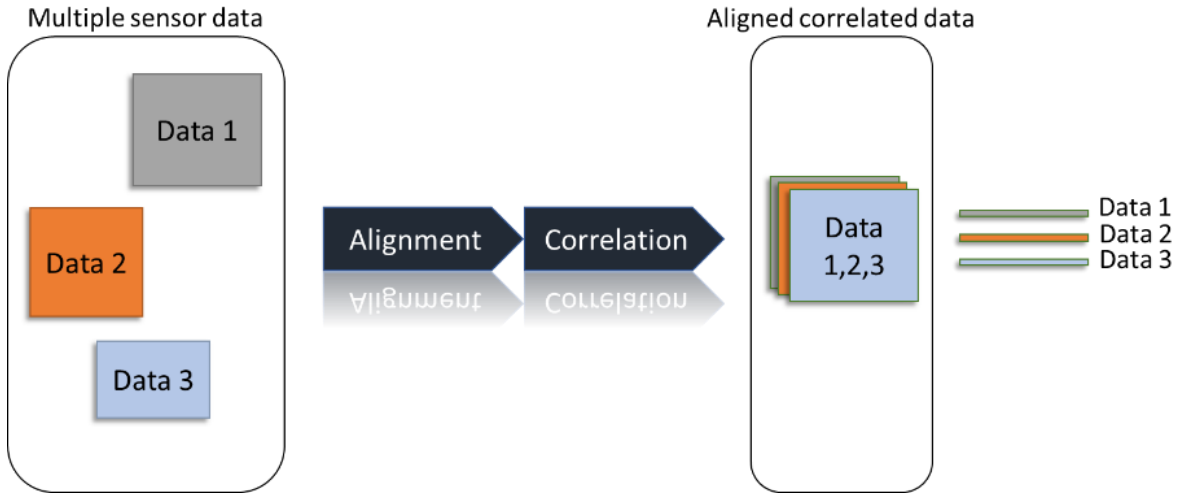



*Figure 3-57. Overview of data alignment and correlation in this project*

### 3.7.1.1 Alignment and correlation of GPR and EMC data

Assuming that the data collection of both GPR and EMC techniques is conducted using the same gridlines, i.e. the start and the end points of collected data (X and Y coordinates) are identical, only the third dimension (Z-direction, i.e. depth) needs to be considered for the data alignment and correlation. Under such circumstances, characteristics of the measured data in X and Y direction do not affect the alignment in the X-Y direction, e.g. GPR provides a continuous measurement over the measured length whereas EMC consists of discrete measurement points over the same length.

As discussed in the literature review chapter (sections 2.6.1 and 2.6.2), the characteristics of recorded GPR and EMC signals is different in the Z-direction, depending on the type, model, and specifications of the employed devices (e.g. GPR antenna and EMC conductivity meters), and their penetration depth. For instance, as shown in Figure 3-58, a simplified synthetic replica of a GPR radargram that was used in the experiment in this thesis consists of 510 sampling points distributed over a depth of approximately 5 metres whereas the provided

EMC data associated with the EM-38MK2 conductivity meter offers a maximum cover depth of 0.75 m and 1.5 m based on the antenna polarisation. Figure 3-58 illustrates the alignment of a GPR radargram and an EMC cross-section map in an exaggerated scale for illustration purposes. In this study, to align the EMC data with GPR radargrams in the Z-direction, a set of measurement values below the cover depth of EMC, aligned with the sampling points of the GPR radargrams, were generated and set to zero. In this way, the covered depth in both techniques were equalised.

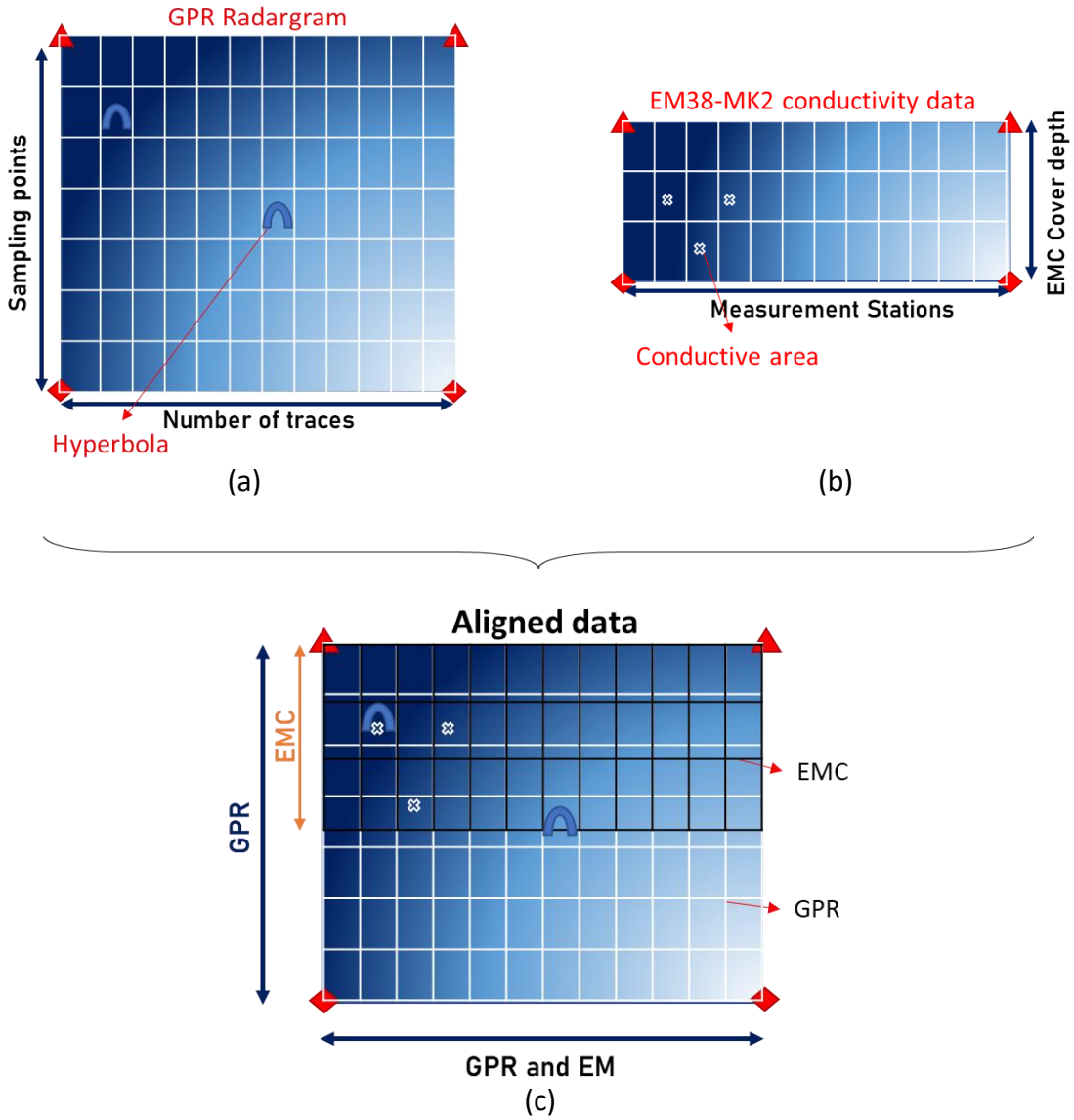



*Figure 3-58. Schematic of data alignment of an (a) example GPR radargrams (510 sampling point and 1300 traces) with a signal penetration depth of e.g. 5 m and (b) EM38-MK2 conductivity meter data with significantly lower cover depth (0.75 m or 1.5 m), (c) the aligned data*

## 3.7.2 Kalman-based fusion of data (sensor level)

In order to be able to integrate different sensors in the process of estimating variables (state estimation) with the KF, it is vital to develop a mathematical model (Eq. 3-38) that correlates the sensor output $y_{sensor}$ with the estimation variables $X$ :

$$y_{sensor} = g(X, u) + \omega_{sensor} \quad (3\text{-}38)$$

As discussed in section 3.3.2, the developed mathematical model in this research associated with GPR consists of three parameters: the estimation variables, i.e. the target signal $s^t$, the background signal $s^b$, and the bias noise $\omega_i$. These variables are defined by three length vectors in connection with the length of associated traces in the GPR radargram (e.g. traces with 510 sampling points). However, the other two sensors used in this study, namely EMC and EML, are different from the GPR in terms of sensor output and measurement data characteristics. In other words, the EMC and EML sensors each provide only one scalar measurement at each station which represents the ground condition in that position. This causes difficulties in using the defined GPR mathematical model in section 3.3.2 for EMC and EML data. In other words, creating a mathematical model that can correlate the EMC and the EML (e.g. CAT) data with the estimation variables (e.g. target, background, and noise signals) is a challenging task and beyond the scope of this research. To solve this problem, data associated with both sensors (i.e. EMC, and EML) were pre-processed separately (discussed in the next subsection) to replicate a new output similar to a GPR radargram. This pre-processing creates a representation of EMC and EML data which consists of vectors similar to the GPR model in section 3.3.2 with normalised amplitudes of the recorded signals. In this manner, the developed GPR mathematical model can be applied to EMC and EML data. This mathematical

model of the system which incorporates all GPR, EMC, and EML data is presented below in Table 3-7, for both the target-free and the target-present states of the KF.

*Table 3-7. Mathematical model of the system consisting of three sensors (GPR, EM, and EML)*

| | System Model | Sensor Model | Sensor Matrix [H] |
|---|---|---|---|
| **Target Free** | $S^b(k) = S^b(k-1) + v_1(k)$ | $u_{GPR}(k) = S^b(k) + \omega_1(k)$<br>$u_{CAT}(k) = S^b(k) + \omega_2(k)$<br>$u_{EM}(k) = S^b(k) + \omega_3(k)$ | $H = \begin{bmatrix} I \\ I \\ I \end{bmatrix}_{3N \times N}$ |
| **Target present** | $S^b(k) = S^b(k-1)$<br>$S^t(k) = S^t(k-1) + b_p(k)$<br>$b_p(k) = b_p(k-1) + v_2$ | $u_{GPR}(k) = S^b(k) + S^t(k) + \omega_1(k)$<br>$u_{CAT}(k) = S^b(k) + S^t(k) + \omega_2(k)$<br>$u_{EM}(k) = S^b(k) + S^t(k) + \omega_3(k)$ | $H = \begin{bmatrix} I & I & 0 \\ I & I & 0 \\ I & I & 0 \end{bmatrix}_{3N \times 3N}$ |

It can be observed that compared to the system model presented in section 3.3.2, the above new mathematical model of the system (i.e. system model) is not changed. Therefore, the system matrix **A** and the system covariance matrix **Q** remained unchanged. The sensor model, however, was changed due to the added EMC and EML sensor data. Such changes, on the other hand, changed the dimension of the sensor matrix **H** of the initial GPR model in section 3.3.2 from $N \times N$ to $3N \times 3N$ consisting of all three sensors. In addition, the dimension of the sensor covariance matrix **R** changed in the same way (i.e. a $3N \times 3N$ matrix). In other words, the covariance-specific characteristics of the EMC and EML sensors were added to the initial diagonal of the matrix **R** of the GPR model in section 3.3.2. Assuming that $r_{GPR}$, $r_{EMC}$, and $r_{CAT}$ are vectors associated with the covariance values of the utilised sensors, the matrix **R** can be presented using Eq. 3-39:

$$\mathbf{R} = diag([r_{GPR} \quad r_{EMC} \quad r_{CAT}]) \tag{3-39}$$

A key factor in determining the sensor covariance matrix **R** is assigning the noise covariance values of the EMC and the EML sensors. Since such noise covariances are approximated values obtained from pre-processing of associated data, a lower accuracy compared to the GPR's sensor covariance can be expected. The lower accuracy of these values is the result of the following: i) The suggested mathematical model in this section for all models was designed based on the developed model for the GPR, ii) The GPR data are the real output of the sensor and not the results of any pre-processing operations. Therefore, it was assumed that within these three techniques the GPR is the more accurate one.

On this basis, significantly larger values can be defined for the noise covariance of the EMC and the EML compared to the equivalent GPR noise covariance values so that the KF is more reliant on GPR data and is less affected by the output of the EMC and EML sensors. In the following, the pre-processing of the EMC and EML data is discussed.

#### 3.7.2.1 Pre-processing of EMC using the EMagPy

As discussed in the literature review chapter, a common approach for pre-processing EMC data is to create a two-dimensional image of the subsurface (similar to GPR radargrams) by inverting the EMC data at each measurement station, which is referred to as conductivity map. To this effect, in accordance with section 2.6.2.1 of the literature review chapter, an inversion algorithm was developed to invert the apparent conductivities based on the cumulative responses of the ground. However, as discussed in the literature review (section 2.6.2.1), this method is limited to surveys with a low induction number (LIN). To address this limitation, instead of utilising the developed inversion algorithm, another open-source Python-based software called EMagPy (McLachlan et al., 2021) was utilised. This is a standalone software that can be used to process, forward model, and invert EM conductivity data (McLachlan et

al., 2021). Other different inversion methods proposed in the literature are discussed in section 2.6.2.1. Figure 3-59 illustrates an example result of the EMagPy software, i.e. the conductivity map of the subsurface obtained for a selected line from the provided dataset discussed in section 3.6.4.

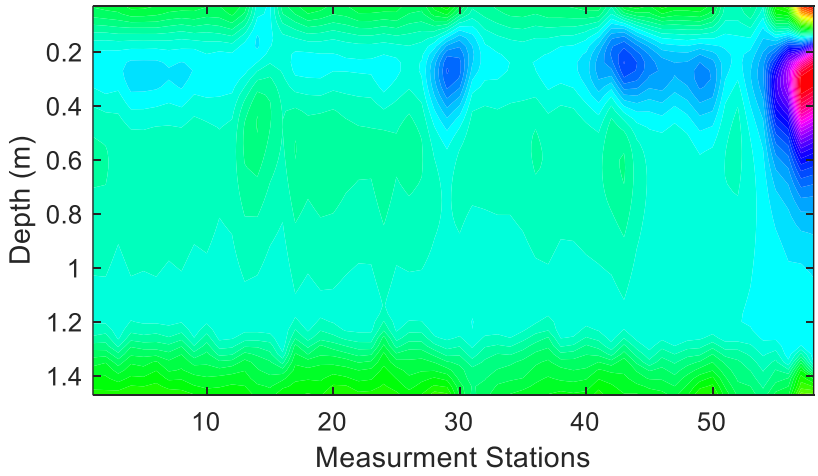



*Figure 3-59. Example conductivity map from EMagPy software for a selected subset of the EMC data used in this thesis, showing the conductive utilities including the dominating inspection chamber on the right-hand side*

It should be noted that the conductive material (or area) in the subsurface is identified by high-amplitude zones (peaks) in the measured signals which bears no resemblance to the hyperbolas in a GPR radargram.

Although the conductivity map provided by the EMagPy is in matrix form (similar to a GPR radargram), the dimension of the output matrix is in accordance with the signal penetration depth of the conductivity meter (e.g. 0.75 m and 1.5 m for utilised EM38-MK2 conductivity meter in this research). Therefore, the EMC and GPR matrices need to be aligned and correlated. To solve this problem, bilinear interpolation i.e. interpolation of two variables at specific distances using linear interpolation (Kirkland, 2010), is used to estimate the required rows and columns to match GPR and EMC matrices.

### 3.7.2.2 Pre-processing of EM locator data

As discussed in the literature review chapter, EML is one of the most common techniques in use in the construction industry for the detection and location of buried conductors, e.g. metallic pipes, or cables. In addition, if adequate information about the sensor coils is available, the depth of buried conductors can be estimated with this technique. To convert the recorded EML data (discussed in section 3.6.2) to a cross-section (like a GPR radargram), the recorded signal peaks in each traverse were identified (the position of potential targets). Then, the estimated depth was extracted. It is worth mentioning that, as discussed in the literature review, some of the new generation devices (e.g. RD8200G used in this research) can automatically estimate the approximate depth of the buried conductors target in the ground. By considering the approximate position of the potential targets on the x-axis ($x_0$), and their calculated depth ($y_0$) estimated by the device, a 2D Gaussian function (Eq. 3-40) is used to create a EML cross-section with the same dimensions as a GPR radargram.

$$u_{CAT} = Aexp\left(-\left(\frac{(x-x_0)^2}{2\sigma^2}+\frac{(y-y_0)^2}{2\sigma^2}\right)\right) \quad (3\text{-}40)$$

Where, A is the signal amplitude ∈ [0 1], $x_0$ and $y_0$ the approximate coordinate of the potential target, and $\sigma$ is the standard deviation.

The remaining unknown parameter in this problem is the standard deviation σ, which is defined by the expert based on the available information about the survey and potential targets. For example, if it is known that the potential target has a large dimension (e.g. a large metallic pipe), the value of σ should be a large number so that the Gaussian function covers a larger area of the conductivity map. Whether the estimated value of σ can be determined from the output of the EML signal is beyond the scope of this study and has not been addressed at this stage. In this study, the value of σ was selected based on the experimental

information. The results and discussion associated with the proposed above data alignment and correlation, and in turn data fusion algorithm, are presented in the results and discussion chapter of this thesis.

## 3.8 Fused GPR-EML data inversion using FEM and GA

The principals of the proposed methodology for the fusion of EML and GPR data was outlined in the previous section. However, real-world scenarios are often more complex and detailed, and often various utilities, e.g. different types of pipes, and high- and low-voltage live cables of various diameters are laid next to each other. On this basis, fusion of survey data for such scenarios requires a more detailed algorithm to address the complexities of the problem. In the following, such a complicated scenario with adjacent utilities is presented. This scenario was developed using numerical modelling, i.e. finite element method (FEM). FEM is a powerful numerical technique for analysing complex engineering systems. The method is based on the concept of breaking down a physical system into a collection of small, interconnected elements, each of which can be individually analysed to understand the behaviour of the entire system (Chaskalovic, 2008). The method is particularly well-suited for solving problems involving partial differential equations (PDEs), which are widely used in engineering and physics to describe the behaviour of physical systems. The finite element method provides a way to numerically approximate the solutions to these PDEs by replacing the continuous system with a discrete one, composed of a large number of small elements. Once the system has been discretised in this way, the solution can be approximated by solving the equations that govern the behaviour of the elements (Chaskalovic, 2008; Liu and Quek, 2003). This process is typically carried out using specialised software, such as COMSOL Multiphysics,

which can take care of the complex computations involved in solving the equations (Dickinson et al., 2014).

Under the developed scenario, the employed techniques, namely EML and GPR, reach their limits in terms of their capability in the detection of both adjacent buried utilities, i.e. each can identify one of the buried utilities. To solve this problem, in the next subchapters a novel methodology in form of an inversion framework for the fusion of GPR and EML data is presented, in which numerical modelling and an evolutionary algorithm are incorporated. The desired unknown parameters of the subsurface scenario, i.e. depth, and level of current flowing through the adjacent utilities, were estimated with the proposed inversion framework.

## 3.8.1 Forward model of EM field due to multiple buried utilities

In this section of the thesis, the undertaken process for developing a forward model for a scenario of a system consisting of adjacent buried utilities and the ground is described (aligned with objectives 6 and 7 of this thesis). The developed forward model is being utilised for the inversion of the unknown parameters of the model by fusing the measured data (i.e. measured magnetic field) using EML and the complementary information obtained from the GPR. The term forward model refers to the role of this model in the proposed inversion framework. The forward model was developed based on governing theory and the equation of the electromagnetic response (i.e. magnetic field) of the buried utilities

### 3.8.1.1 The theory for developing the forward model

The forward model developed for the proposed inversion framework is a simulation of the electromagnetic response of adjacent buried utilities in the ground, e.g. live electricity, and

telecommunication cables. The governing equations for developing the forward model are Maxwell's equations (Equation 3-41 to 3-47) and the constitutive equation (Equation 3-44 to 3-46) which were discussed in detail in Appendix I of this thesis.

$$\nabla \times D = \rho_v \quad (3\text{-}41)$$

$$\nabla \times H = J \quad (3\text{-}42)$$

$$\nabla \times E = -\frac{\partial B}{\partial t} \quad (3\text{-}43)$$

$$\nabla \times \mathrm{H} = J + \frac{\partial D}{\partial t} \quad (3\text{-}44)$$

$$J = \sigma E \quad (3\text{-}45)$$

$$D = \varepsilon E \quad (3\text{-}46)$$

$$B = \mu H \quad (3\text{-}47)$$

Where, $D$ is the density of electric flux in the buried cables, $E$ is the electric field, $H$ is the intensity of the generated magnetic field by the buried cables, $J$ is the electric current density in the cables, B is the magnetic field, $v$ is the electric charge, $\sigma$ is the dielectric conductivity, $\varepsilon$ is the electric permittivity, and $\mu$ is the magnetic permeability. The parameter ∇ is the 'curl operator' i.e. the second derivative of interested parameters.

These equations simulate the electromagnetic response of buried utilities by correlating all the above-listed parameters with the material properties of the buried utility-ground system. In this forward model, electric current flows through two buried adjacent live cables in the ground. Each of these cables carries a different voltage, i.e. one is a high-voltage, and the other is a low voltage cable. Depending on the distance between adjacent cables, two individual distinct MFs or a coupled single MF (which was discussed in section 2.6.1.6 of the literature review) can be generated. The detectability of individual MFs using EML in their simplest form

was demonstrated experimentally in section 3.6.2 of this chapter. Both individual and coupled MFs were simulated using the developed forward model of the ground and buried adjacent utilities.

The above-listed equations were solved using the finite element method (FEM) which is a numerical approach for solving partial differential equations. FEM is widely used in different fields for modelling and simulating complex geometries of advanced engineering problems which offers more flexibility compared to similar approaches such as finite difference method (Liu and Quek, 2003). The software package used for developing the model was COMSOL Multiphysics which provides a basis for developing the geometry, formulation of the FEM by discretising, and providing the results by solving the differential equations based on the selected governing physics of the problem. One advantage of COMSOL Multiphysics is access to utilised equations and the source code of the software. In addition, the capability of COMSOL in parametric modelling provides high flexibility for researchers in tuning model parameters. Another advantage of COMSOL Multiphysics is a live cooperation capability with MATLAB® and its toolboxes (e.g. global optimisation toolbox) through its LiveLink module. These are valuable advantages for researchers to handle COMSOL using MATLAB® to control and modify model parameters, running specifically developed codes based on projects requirements and applying external functions to solve inversion and optimisation problems (González-Quirós and Comte, 2021; Li et al., 2010; Nor et al., 2015; Ozana et al., 2016).

#### 3.8.1.2 Dimensions of the forward models

The focus of this research was on shallow depth utilities in the ground such as buried pipes and cables that are typically extended in one direction. On this basis, a two-dimensional (2D) model of the ground-adjacent buried utilities was developed. A two-dimensional model of a

system with one dimension significantly larger than the other, such as in buried parallel cables, provides a low computational cost of the simulation while maintaining the required modelling accuracy. Nevertheless, there are other anomalous bodies in the ground such as voids (e.g. installed void in section 3.6) or soil disturbances (e.g. fully silted broken pipe in section 3.3.5.1) that are not two-dimensional.

To reduce the computational cost and as a common practice in the literature, such anomalies are assumed to be two-dimensional (Liang et al., 2021; Van Vorst et al., 2014; Wang, 2017). However, such assumptions are problem dependent and can be invalid based on the governing conditions, e.g. scale, geometry, soil property, and model parameters. Under such circumstances, a three-dimensional (3D) model will be the alternative solution to simulate the system more realistically. It should be noted that the accuracy of the established 2D forward model in this thesis was validated against real scale experiments presented in section 3.6.2, laboratory-based experiments provided by K.L. Siu and Lai, (2019), and the national grid's technical note (Nationalgrid, 2015).

#### 3.8.1.3 Geometry of the FE forward model

The origin coordinate for the created forward model was defined to be in the centre of the model, labelled '*O*' in Figure 3-60. The minimum cover depths for directly buried and in duct buried high voltage (HV) and low voltage (LV) cables ($D_1$ and $D_2$, respectively), and the separation distance $D_3$ were defined in accordance with the National Joint Utilities Group Guidelines (NJUG, 2013). The diameters of buried cables, i.e. $D_{p1}$ and $D_{p2}$, are defined based on British Telecommunications openreach guideline (BT opernreach, 2019). The positions of both cables were defined using polar coordinates ($r_{p1}$ and $\theta_{p1}$) and ($r_{p2}$ and $\theta_{p2}$) within the main coordinate system. The depth and length of model components are represented by the

vertical and horizontal axis of the coordinate system, respectively. Finally, the properties of dry sand were assigned to the uniform ground.

Because the proposed forward model in this thesis is developed for measuring the magnetic fields on the surface due to the buried utilities, a set of uniformly distributed measurement points were considered on the ground surface. Infinite boundaries (i.e. physical infinity) were applied to the model so that the effect of boundaries on the analysis was eliminated. Details of the numerical model such as discretisation is provided in the next sections. Details of the numerical model, including the discretisation method, are provided in the following sections.

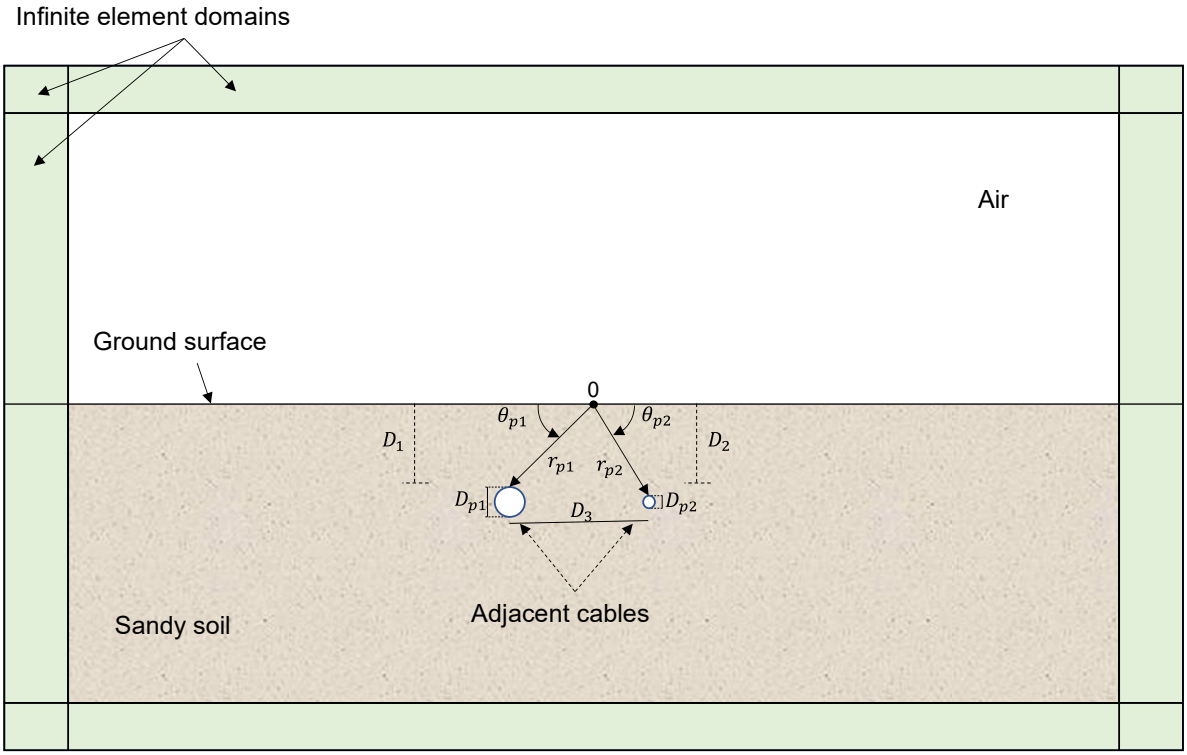



*Figure 3-60. Schematic of the forward model of directly buried utilities utilised in the proposed inversion framework. $D_1$ and $D_2$ represent the cover depths of buried cables, $D_3$ represents the separation distance between them. The diameters of the buried cables are represented by $D_{p1}$ and $D_{p2}$, respectively. $r_{p1}$ and $r_{p2}$ represent the radial distances of each cable from the origin of the coordinate system, and $\theta_{p2}$ and $\theta_{p1}$ are the angular coordinates of each cable with respect to the main coordinate system.*

#### 3.8.1.4 Discretisation of the FE forward model

To discretise the developed FE model, a sensitivity analysis was carried out to identify the efficient discretisation, i.e. meshing, of the model. A computationally optimal and accurate solution of the magnetic field due to the buried cables was achieved by the following meshing. The ground area was meshed using triangular meshes due to the geometrical complexity of the FE model in this area where utilities are buried. Then the uniformly distributed squared mesh elements were used for meshing the air and the infinite boundaries. The reason for the complexity of the ground area is due to the changing position of the buried utilities during the GA operation. Defining a parametric meshing area (variable) could solve the issue of such complexity to have a fully controlled discretisation of the FE model. In addition, the automatic meshing by COMSOL Multiphysics, referred to as physics-controlled meshing, was employed for comparison purposes. Figure 3-61 presents examples of discretised FE models using user-controlled meshing method, and Figure 3-62 presents examples of discretised FE models using physics-controlled meshing method.

The statistics associated with the quality analysis of meshing elements for both user- and physics-controlled discretisation are presented in Figures 3-63 and 3-64. An average element quality of 0.8734 was achieved for the user-controlled discretisation whereas the average element quality of purely physics-controlled discretisation was 0.8294. It is stated that a good element quality is deemed necessary to obtain of accurate and stable solutions in finite element modelling (COMSOL Multiphysics, 2017). The quality of the elements is determined through the average element quality metric, which ranges between 0.4 and 1.0, with a higher value indicating a better quality (COMSOL Multiphysics, 2017). In this context, an average element quality of 0.8734 and 0.8294 are both considered to be good as they fall within this range. The proper selection of the type of elements and the adjustment of the discretisation

of the model are factors that contribute to the achievement of a good element quality. As such, it can be inferred that effective discretisation of the model has been performed, resulting in well-shaped elements, leading to a reliable solution.

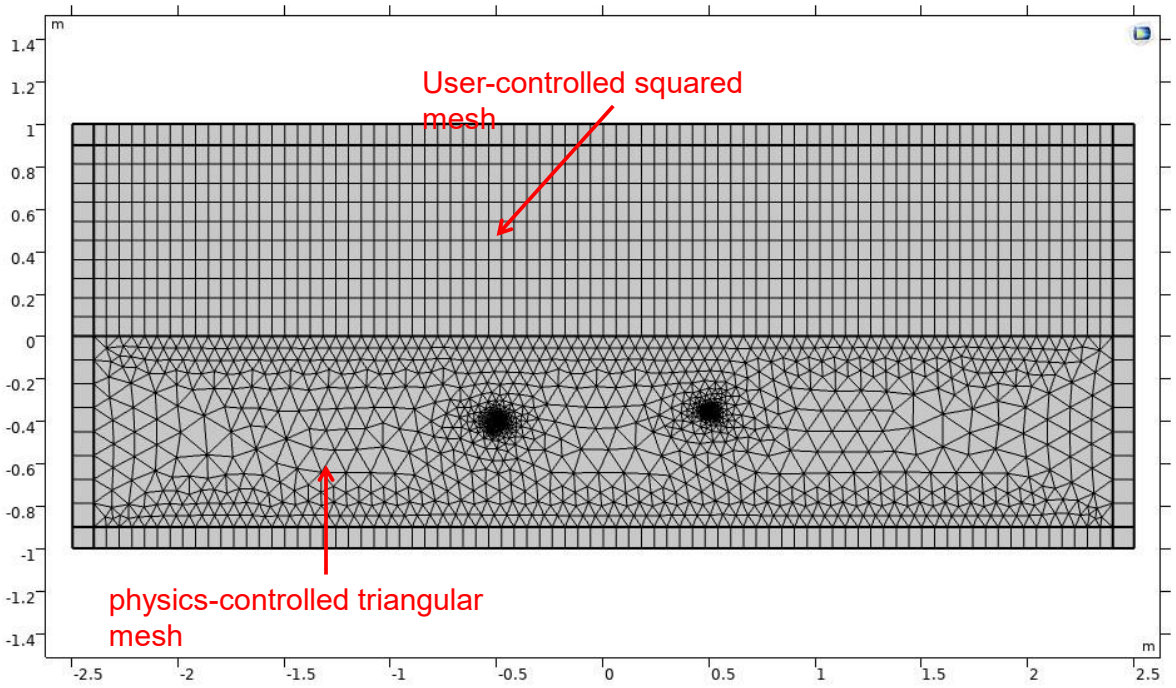



*Figure 3-61. Schematic of discretisation of the model using user-controlled meshing with a fine element size*

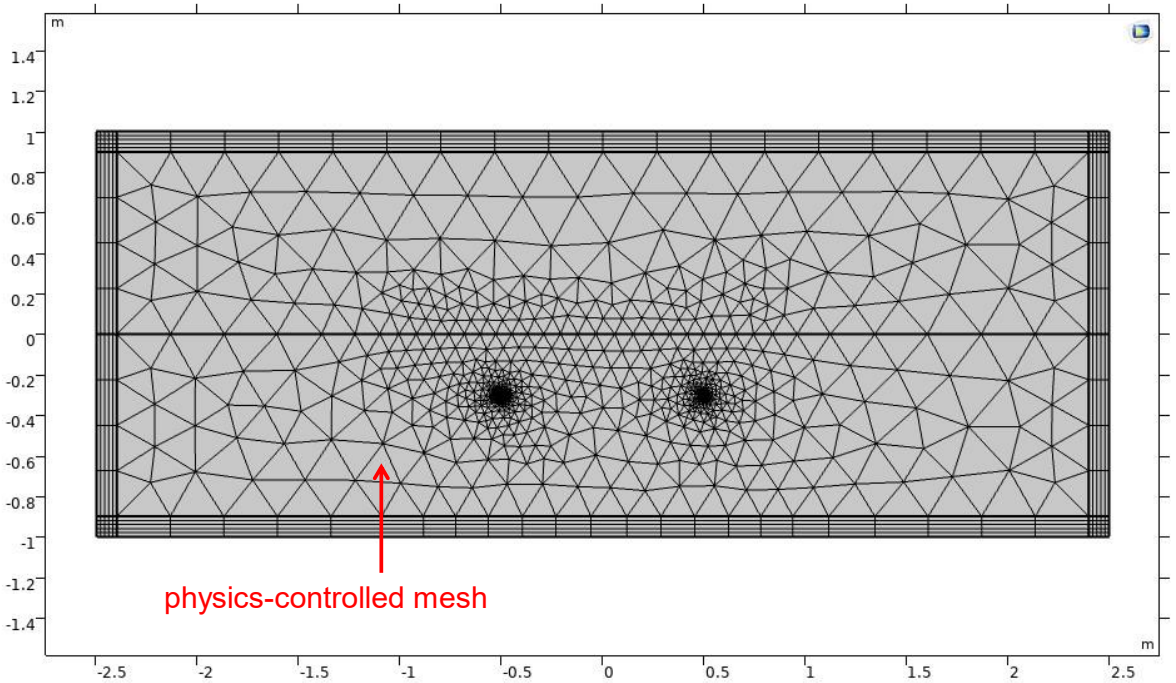



*Figure 3-62. Schematic of discretisation of the model using physics-controlled meshing with a fine element size*




*Figure 3-63. Statistic results produced by COMSOL associated with the user-controlled discretisation mode*

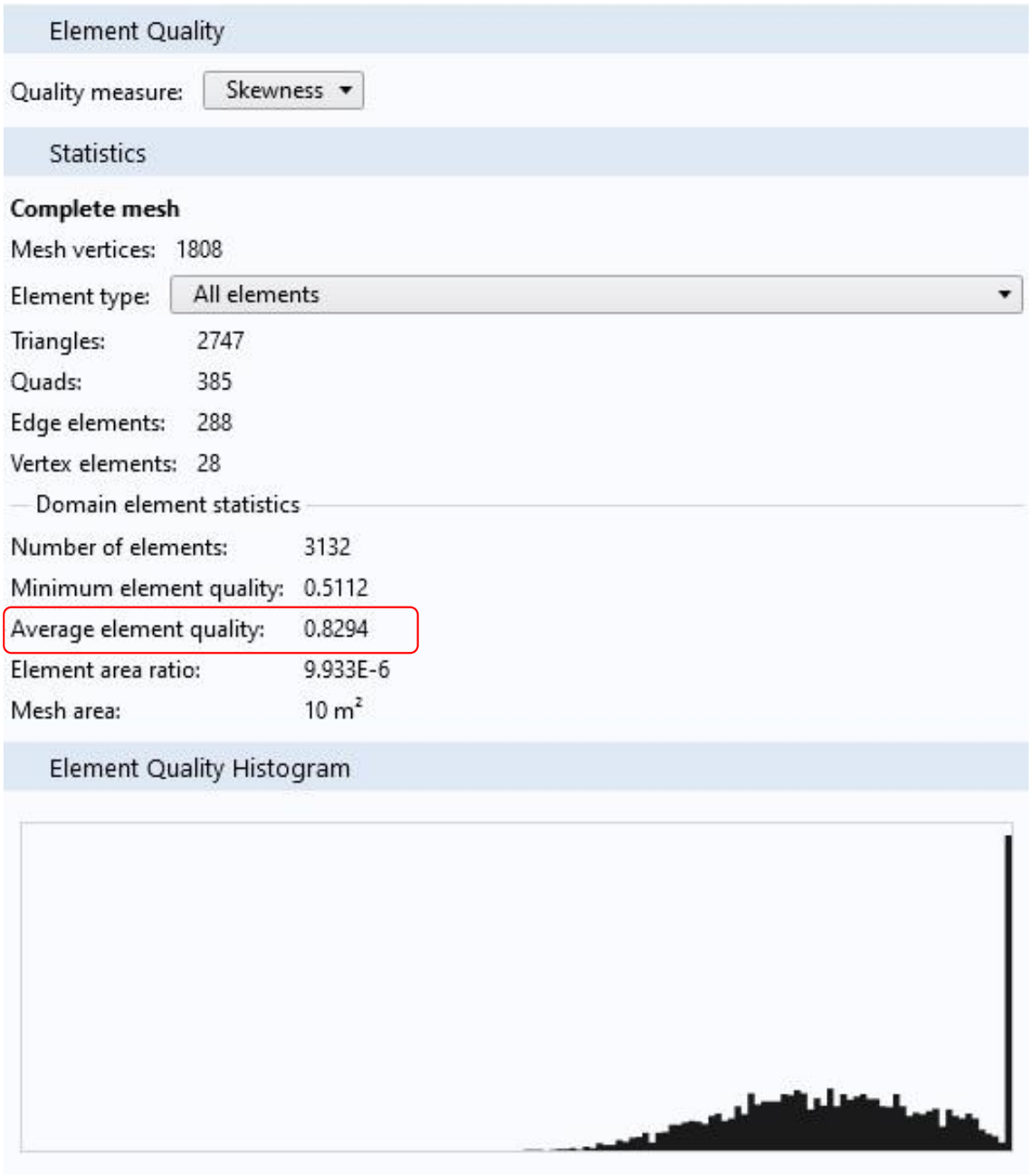



*Figure 3-64. Statistic results produced by COMSOL associated with the Physics-controlled discretisation mode*

In addition, the effect of the element size on the simulated magnetic field on the ground surface was carried out through a sensitivity analysis (Table 3-8). The variation in the shape and size of the mesh elements did not improve the accuracy of the results while the computational cost was increased proportionally by reducing the size of the elements.

*Table 3-8. Results associated with a sensitivity analysis for determining the computational cost of different element sizes*

| Meshing method | Element size | Domain elements | Boundary elements | Degrees of freedom | Solution time [sec] |
|---|---|---|---|---|---|
| physics-controlled | Normal | 3030 | 269 | 15298 | 2 s |
| | Fine | 3132 | 288 | 15973 | 2 s |
| | Finer | 4130 | 340 | 20896 | 2 s |
| | Extra Fine | 7198 | 470 | 36070 | 3 s |
| | Extremely Fine | 21688 | 802 | 104443 | 4 s |
| User-controlled | Normal | 3956 | 514 | 22609 | 2 s |
| | Fine | 3956 | 514 | 22609 | 2 s |
| | Finer | 4506 | 514 | 25084 | 2s |
| | Extra Fine | 5432 | 514 | 29251 | 3 s |
| | Extremely Fine | 6966 | 514 | 62050 | 3 s |

The sensitivity analysis results, summarised in Table 3-8, showed that the element size (and consequently the degrees of freedom) did not have significant impact on the computational time of the problem. Thus, fine element size with physics-controlled meshing method was used to maintain the accuracy of the calculations.

#### 3.8.1.5 Validating the FE forward model

The purpose of the developed FE forward model in this study was to correlate the spatial position and the intensity of the flowing current in the buried conductive utilities with their associated magnetic fields (i.e. objective 6). The developed FE model in this research was validated using an experiment carried out by Siu and Lai (2019). In their experiment, CAT and Genny were utilised (RD8100) for a buried utilities-ground scenario consisting of a galvanised pipe of 50 mm diameter and a single-phase live cable with 5 A current in sandy soil (Figure 3-

65). The measurements were conducted using different current densities and signal frequencies.

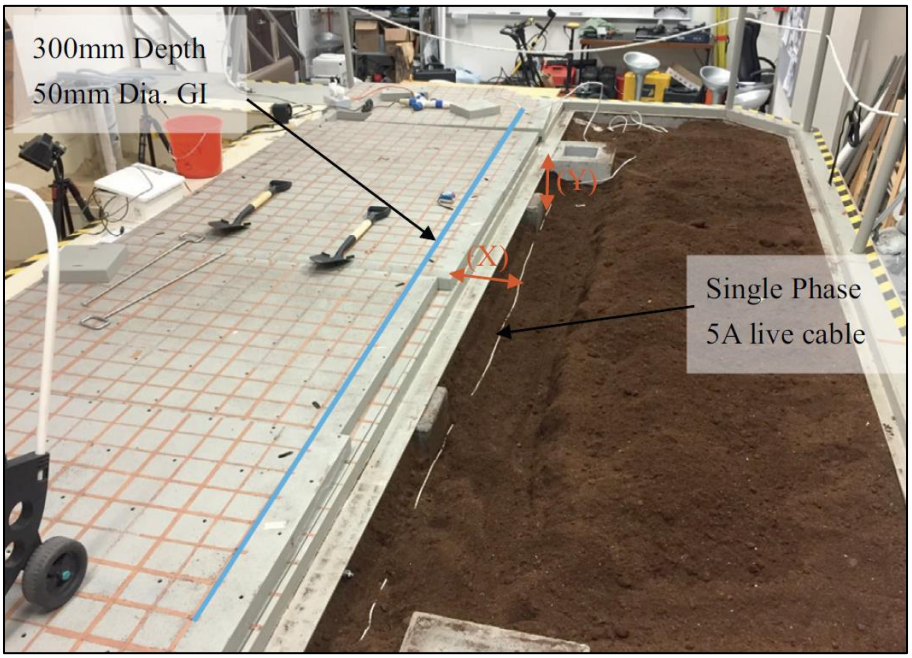
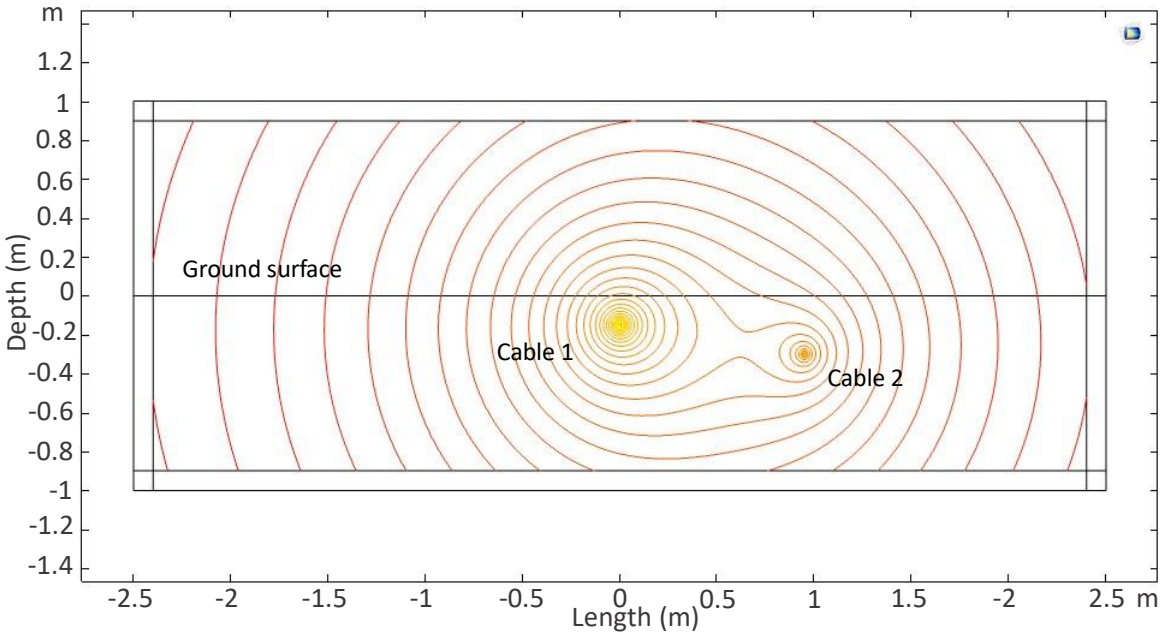
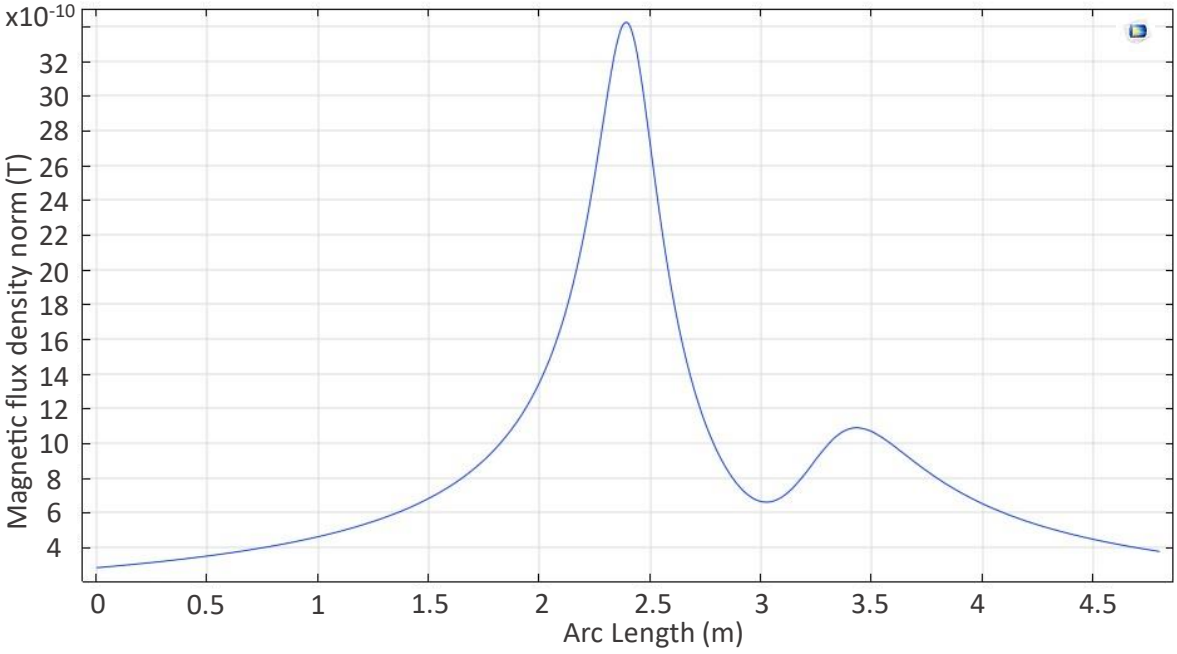



*Figure 3-65. The experiment scenario by Siu and Lai (2019) for investigation of the electromagnetic field pattern caused by the coupling effect of adjacent buried conductive utilities.*

Their experiment was simulated using the FE model in this study and their results were utilised to validate the functionality of the FE model. Figures 3-66(a) and (b) show example results obtained from the developed FE forward model in this study that simulates the electromagnetic field pattern as well as the magnetic flux density in and above the ground.




*(a)*




*(b)*

*Figure 3-66. Example results obtained from the FE model of the experiment reported by Siu and Lai, a) Electromagnetic field pattern, b) The magnetic flux density measured on the ground surface*

## 3.8.2 The inversion algorithm using the FE forward model and a GA

The inversion algorithm in this thesis consists of three main components, i.e. the FE forward model, the GPR radargram and complementary EMC or EML data, and the genetic algorithm (GA) optimisation. The inversion, in line with objectives 6 and 7 of this thesis, aims to identify the position and the intensity of the flowing current in buried conductive utilities in shallow depth (e.g. buried HV and LV cables, and metallic pipes) based on their MFs. To fulfil this, the FE forward model (developed using COMSOL Multiphysics) was defined as GA's objective function. Depending on the problem variables, e.g. position and dimension of buried utilities, the intensity of the flowing current, the associated voltage, MF, and the dielectric property of materials, GA generates a set of random estimates for the variables as an initial estimation. Thereafter, the forward model is utilised by the GA to calculate the MFs around buried utilities (individual or coupled) for each of the initially estimated variables. The GA ranks the estimated variables based on the closeness of the estimated MFs to the experimentally measured MFs. In addition, the information extracted from the GPR radargram, e.g. spatial position of the potential target(s) in the ground, is used by the GA to constrain the unknown problem variables and hence reduce the search space for the best solution. It is worth mentioning that the data fusion in the proposed inversion in this research can be considered as an intermediate level (decision-based) data fusion as discussed in the literature review (section 2.4.3).

### The GA optimisation

As described by Yang (2014), the GA consists of three main operators, i.e. crossover, mutation, and selection, that operate as follows. Crossover operators swap parts of the high-ranked solutions with other high-ranked solutions to produce new solutions. The mutation operators change parts of a solution randomly so that the diversity of the estimated population is

increased. The mutation provides a mechanism for the GA to escape from a local optimum and explore a broader space for further solutions. Finally, the best estimates (fittest solutions) are selected to create the next generation (i.e. set) of solutions. This procedure will be repeated so that the divergence of the GA estimates from both, experimentally measured MFs, and the constrained information (fused data) provided by the GPR radargram (i.e. the estimated spatial position of the detectable buried utility by the GPR) is minimised. In other words, the estimated MF by GA matches the experimentally measured MFs in correlation with the information obtained from the GPR radargram. It should be noted that the obtained results are subject to a tolerated error. The error tolerance for the optimisation problem in this research can be defined based on the type and characteristics of recorded MFs using an EML device (i.e. a signal with a strength indicator between 0 to 100).

The benchmark for identifying the minimised divergence is a mean squared error (MSE) function called misfit which is calculated using Eq. 3-48.

$$MSE = \sqrt{\frac{1}{n}(\sum_{i=1}^{n} MF_{i-estimated} - MF_{i-experiment})^2} \quad (3\text{-}48)$$

Where $MF_{estimated}$ is the estimated MF by the FE model, the $MF_{experimrnt}$ is the experimentally measured MF, and $n$ is the number of estimates in each iteration (also referred to as generation size of the GA).

The MSE serves as the termination criteria of the optimisation procedure which decides whether the optimisation process has reached the optimal values for the unknown variables. Alternatively, the error tolerance can be utilised in correlation with the MSE as a benchmark for stopping the optimisation procedure (i.e. when MSE < error tolerance). For example, a 10% error tolerance terminates the optimisation if the estimated values are above 90% of the

recorded signal strength. However, to obtain the most accurate results in this research, an error tolerance equal to zero was utilised.

A secondary termination criterion for the GA optimisation procedure was defined based on the gradient of the solutions over the iterations. When the optimisation results stop improving, i.e. the gradient of the solutions becomes zero, for successive generations, i.e. 10 generations in the current inversion, the algorithm selects the latest estimated values as the best estimates and terminates the optimisation procedure.

### The GA optimisation parameters

Robust performance of the GA is dependent on the selection of suitable options for the GA parameters which is a problem-dependent task (Sumathi et al., 2008). Determining (pre-adjustment) suitable GA parameters is vital for the GA algorithm to obtain good quality results (Haines et al., 2014). To determine a suitable set-up for the GA algorithm in this research, a series of pre-analysis based on available suggestions in the literature was carried out. For example, a broad suggestion about the number of populations is to have a population size of ten times the problem variables, e.g. a population size of 60 for the current inversion problem with six variables of $X_1, Y_1, X_2, Y_2, C_1, C_2$ associated with the positions and flowing current in the buried utilities. Table 3-9 presents the values that were determined as suitable control options for solving the problem in this research.

*Table 3-9. Example of suitable GA control options determined for the optimisation problem in this research*

| Control Options | Value |
|---|---|
| Population Size | 60 |
| Max Generations | 60 |
| Function Tolerance | 0 |
| Fitness Limit | 0 |
| Constraint Tolerance | 0 |
| Crossover Heuristic | 1.2 |
| Crossover Fraction | 0.75 |

The flowchart of the proposed inversion algorithm in this research to reconstruct the underground, in terms of the position of the buried conductive utilities, is shown below in Figure 3-67.

The performance of the proposed inversion algorithm was investigated, and the associated results are presented and discussed in the following results and discussion chapter of this thesis.

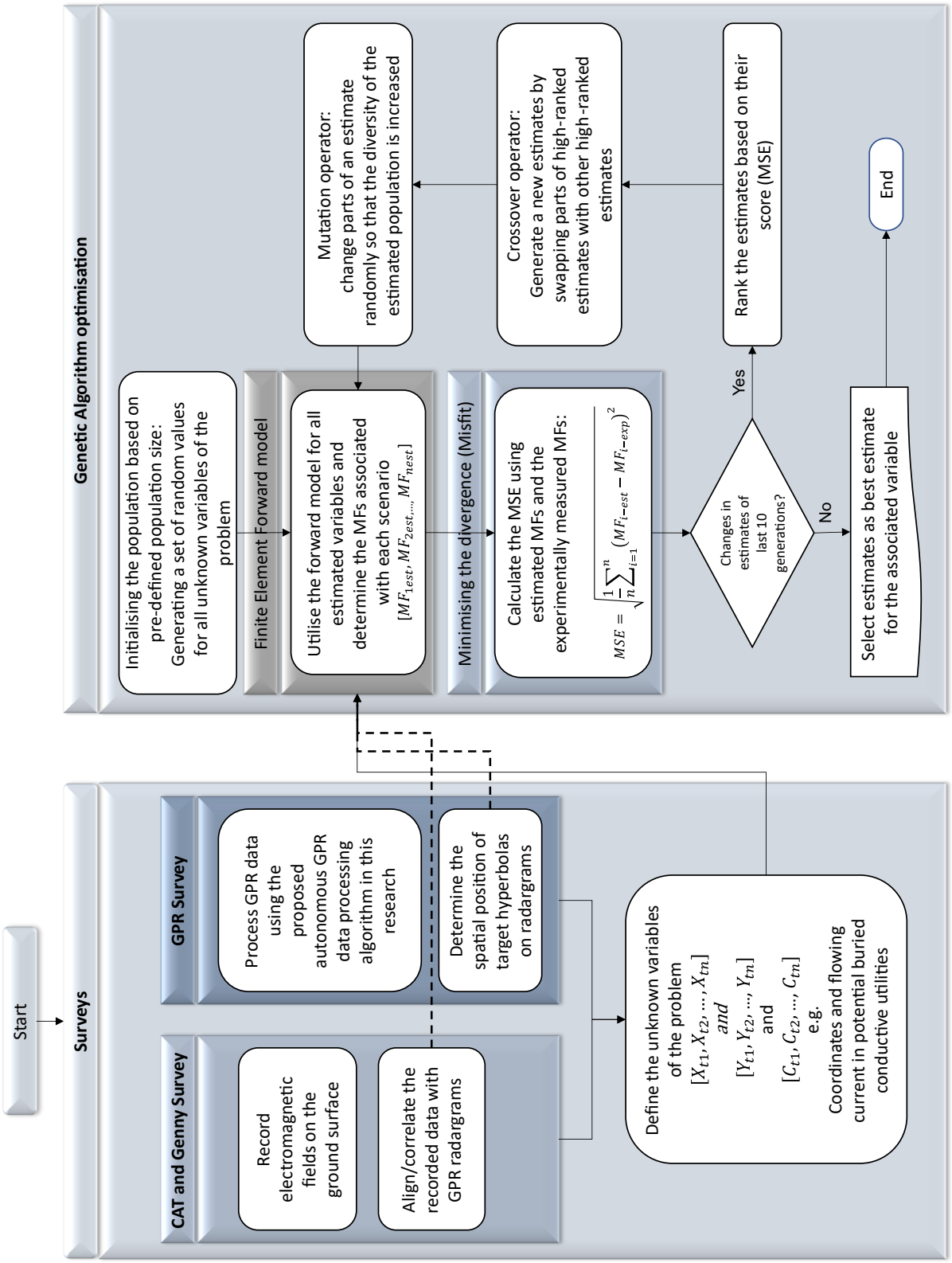



*Figure 3-67. Flowchart of the proposed inversion algorithm consisting of surveys, numerical simulation using a FE forward model, constraining the variables using GPR information (data fusion), and the genetic algorithm optimisation*

## 3.9 Summary

In this chapter, four novel algorithms were developed in line with objectives 3, 4, 5, 6, and 7 of the thesis for the following application:

i) pre- and post-processing the GPR data, reduction of noise and clutter.

ii) detection and location of buried targets and ground anomalies in the subsurface (e.g. pipes and soil disturbances).

iii) fuse the information obtained from processed GPR data with other NDT techniques (EML and EMC).

A summary of these developed novel algorithms is presented below.

1) A semi-autonomous algorithm for pre-and post-processing of GPR data, reduction of noise and clutter, and detection and locating of buried anomalies in shallow subsurface (e.g. buried pipes and ground anomalies such as soil disturbance)

This novel semi-autonomous algorithm was developed based on the KF and by incorporating the following filters for processing the GPR data: time zero correction, signal gain adjustment, bandpass filter, horizontal banding removal, and DC-shift. The normalised innovation squared parameter of the KF (a benchmark for KF's performance) was utilised to produce a novel signal (named the NIS signal in this thesis) which was processed and employed for the detection and locating of the ground anomalies, e.g. buried pipes and soil disturbances. In the proposed algorithm, the optimal values for input parameters of the KF (i.e. system, sensor, and error covariance matrices) were determined. This goal was achieved using a multi-objective optimisation algorithm (MOGA) by making use of a genetic algorithm incorporated with the ROC curves, and the FFT for processing the NIS signal.

2) An autonomous algorithm for pre- and post-processing of GPR data, reduction of noise and clutter, and detection and locating of buried anomalies in shallow subsurface (e.g. buried pipes and ground anomalies such as soil disturbance)

The second novel algorithm utilised the above-mentioned developed KF in a more efficient way by incorporating the wavelet transform as a powerful tool for processing signals and images. The NIS signal was simply processed using a discrete wavelet transform by utilising various mother wavelets (e.g. db, Sym, dmey) with no further need for utilising the ROC curves or the FFT. On this basis, the optimal values for KF's input parameters were determined using a single objective optimisation algorithm (GA).

3) An algorithm for reconstructing the shallow subsurface by fusion of the GPR, EML, and EMC data.

The suggested novel algorithm is a multi-sensing KF based algorithm using which experimentally obtained data of GPR, EMC, and EML are fused/integrated. Among these three techniques, the GPR was considered as the more reliable technique (in terms of accuracy) which highlights the spatial positions of potential targets by their associated hyperbolas in radargrams. The data obtained from the EMC and the EML sensors, on the other hand, were utilised to improve the survey interpretations by adding complementary information to the system (e.g. material type). In other words, the EMC and EML data were fused into the system to determine the type of material of a potential buried target or the area of interest. For example, identifying a conductive buried target (e.g. metallic pipe or cable), or a ground area with salinity in the subsurface (highly conductive).

4) An inversion algorithm for reconstructing the shallow subsurface in terms of the location of buried conductive utilities (e.g. live cables and metallic pipes).

This novel inversion algorithm consisted of a FE forward model, GA optimisation, and the fusion of data (information) obtained from the processed GPR data into the algorithm. The location of potential buried conductive utilities as well as the intensity of the flowing current in such utilities were inverted by simulating the MF(s) around them. Such MF simulation was carried out using a FE forward model developed using COMSOL Multiphysics which was controlled by the GA optimisation algorithm. Experimentally obtained GPR data were processed and utilised for the fusion of data by constraining the unknown variables of the inversion algorithm. In other words, the spatial positions of target hyperbolas were determined (using the above-mentioned autonomous algorithm) and added to the algorithm to reduce the number of unknown variables.

# Chapter 4 RESULTS AND DISCUSSION

## 4.1 Introduction

This chapter presents and discusses the results obtained in numerical and experimental examples of buried utility detection using the proposed algorithms presented in Chapter 3. To preserve consistency, the results in this chapter are presented in the following same order as the proposed algorithms in Chapter 3:

1. The results from the semi-autonomous and autonomous algorithms for processing GPR data in an experimental example.
2. The validation of the performance of the semi-autonomous and autonomous algorithms using three sets of experimental data.
3. The evaluation of the performance of both algorithms against one method proposed in the literature and one conventional method used by the industry.
4. The fusion of the GPR and EMC field data to identify buried metallic pipes in an experimental test.
5. The fusion of GPR and EML field data to identify a buried metallic pipe in an experimental test.
6. The inversion of fused GPR-EML data using FE modelling and GA to identify two adjacent buried live cables in a numerical example.

Finally, a summary of the key outcomes from the above results is presented at the end of this chapter.

## 4.2 GPR survey data

The main GPR data used in this study were obtained from a series of measurements taken from an accurately constructed trial site, which covered 23 scan lines across an 11 m x 13 m area with five linear pipes installed at shallow depths. Furthermore, as shown in the site layout in Figure 3-26 in Chapter 3 (page 139), there was a pipe which had a broken section with a blockage over 2 m (filled with soil) between the 12th and 15th traverses (Figure 4-1).

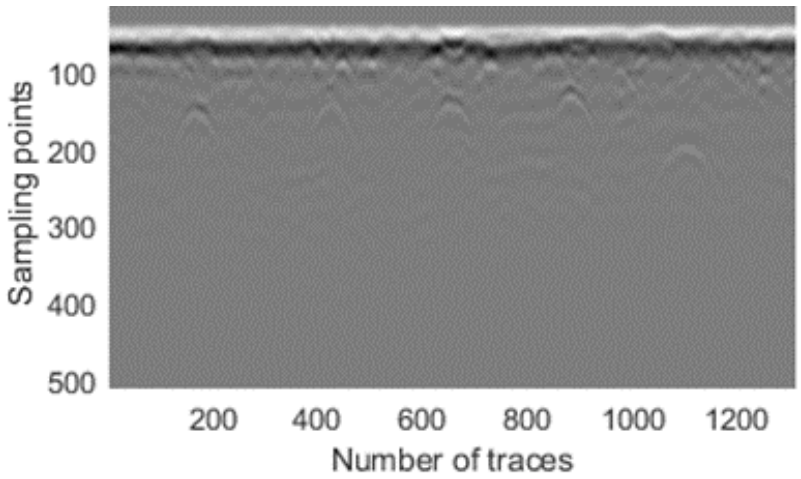
It is worth noting that the experimental tests in this thesis were carried out under semi-controlled conditions and the results are specific to the current properties such as soil type, water content, and environmental noises in the lab. These are influential parameters in the experimental setup which limit the results in this study to the current testing conditions.



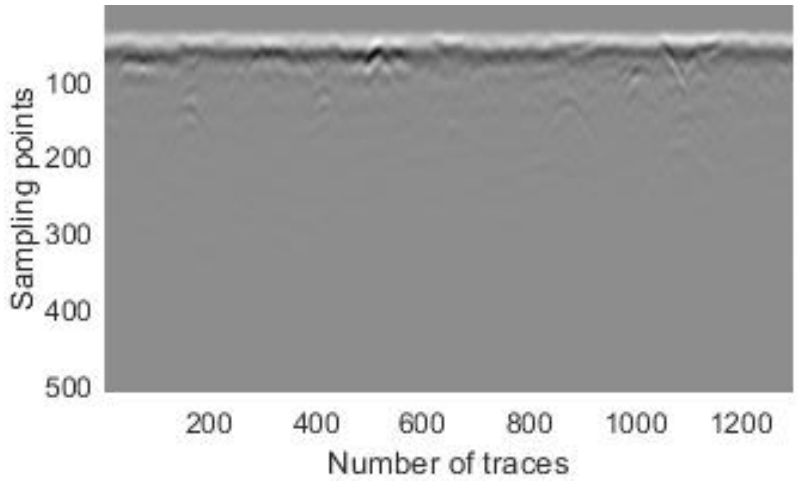
*Figure 4-1. Brocken pipe section (image: RSK)*

All the radargrams associated with this dataset were processed using both of the proposed algorithms in this research (i.e. the semi-autonomous and the autonomous algorithms). To demonstrate the performance of both algorithms in this thesis, two radargrams were selected such that one contained five fully intact pipes (Figure 4-2a) and the other included four fully intact and the broken pipe section filled with soil at approximately trace 650 (Figure 4-2b). In

addition, the performance of both proposed algorithms was validated using four additional GPR datasets from different independent surveys which will be discussed in section 4.6.




*(a)*




*(b)*

*Figure 4-2. Selected radargrams: a) Radargram 9 with no defective pipes (five fully intact pipes), b) Radargram 13 with a broken pipe section at approximately trace 650 (four fully intact pipes and a broken soil filled pipe).*

## 4.3 The semi-autonomous algorithm

### 4.3.1 Optimisation of the GPR outcome using the MOGA analysis

The optimal Pareto solutions (PSs) discussed in the previous chapter were utilised to compare the performance of the algorithm by considering the position of the peaks of the normalised innovation squared (NIS) signal function, which is visualised through dashed lines in Figures 4-3 and 4-4. The detection outcome of the algorithm using two PSs (PS1 and PS4) for two different radargrams (009 and 013) are presented in Figures 4-3 to 4-6. As previously discussed, reflected GPR signals due to a contrast in the dielectric permittivity of the materials are indicators for existing targets or ground anomalies. The position of such contrasts was detected by the proposed algorithm and indicated automatically using dashed lines on the original radargrams (Figures 4-3a to 4-6a). The processed radargrams generated by the algorithm with estimated and removed backgrounds are presented in Figures 4-3b to 4-6b.

The comparison of PSs has indicated that the solutions are optimised in diverse ways. In other words, using a PS with different optimisation criteria can deliver different results in terms of detected anomalies (dashed lines) which is an indicator of the variation of system characteristics in terms of the receiver operating characteristic (ROC) and Fourier frequency. For example, the evaluation of PS4 results shows that the corresponding mean value of the innovation tends to be minimised and producing high number of dashed lines (Figure 4-3a) compared to the PS1 results which tend to maximise the area under the ROC curves and producing minimum number of dashed lines (Figure 4-4a). However, the minimum number of dashed lines in Figure 4-3a resulted in missing the fifth target hyperbola around 1100th trace in Figure 4-3b. A similar trend can be observed for Figures 4-5a and 4-6a. Dashed lines can also

be considered as false alarms if an inappropriate PS is selected. The number of dashed lines has a direct relationship to the sensitivity of the system which can be set by the expert based on the PS.

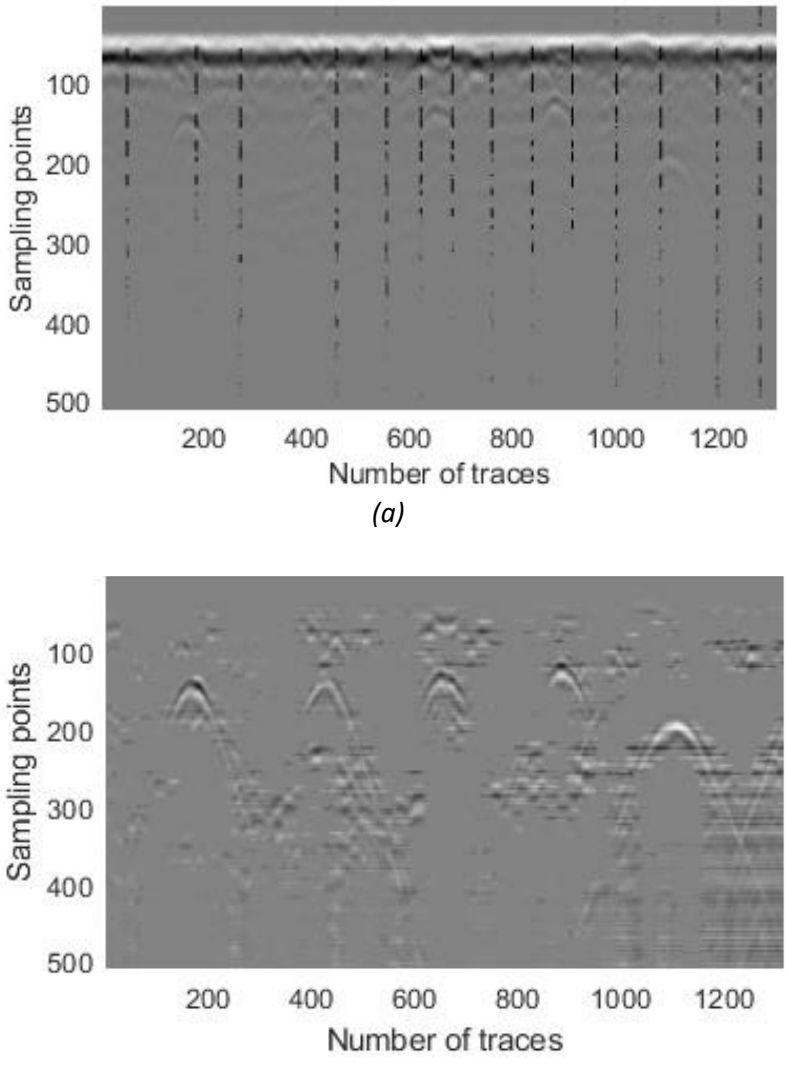
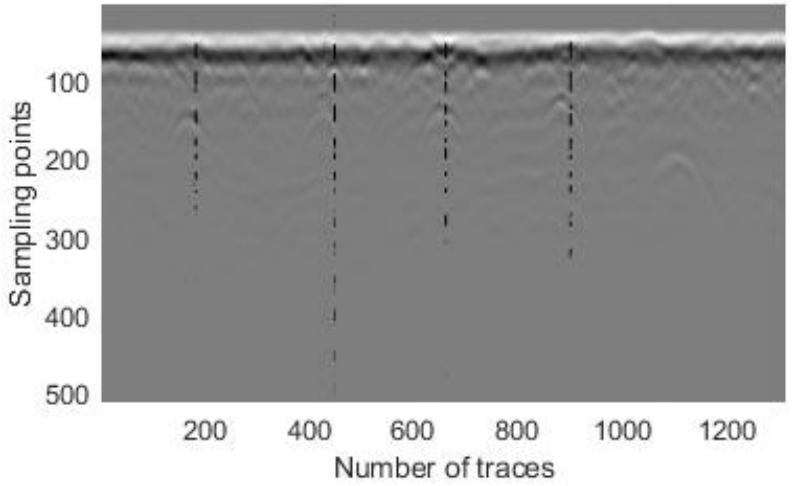
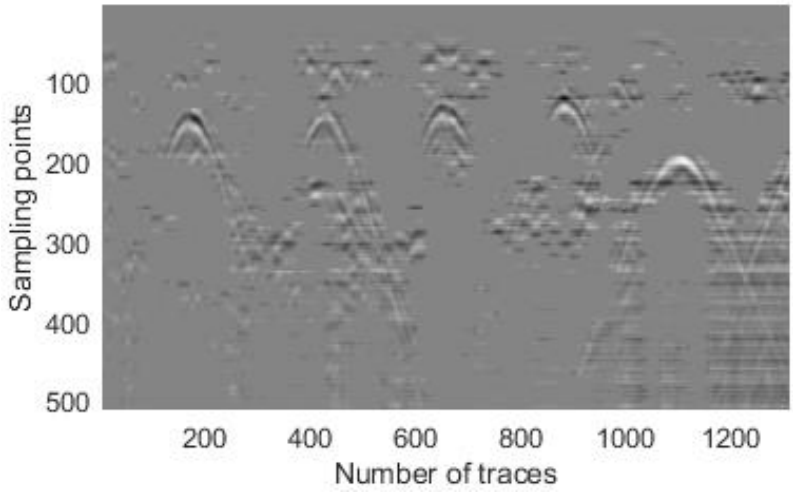



*Figure 4-3. a) Original radargram 009 with allocated dashed lines using PS4, b) By the KF processed radargram 009 using PS4*




*(a)*

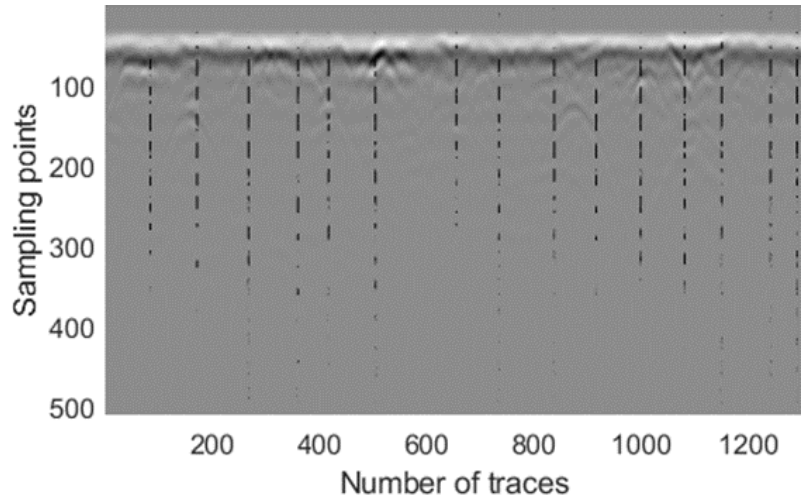



*(b)*

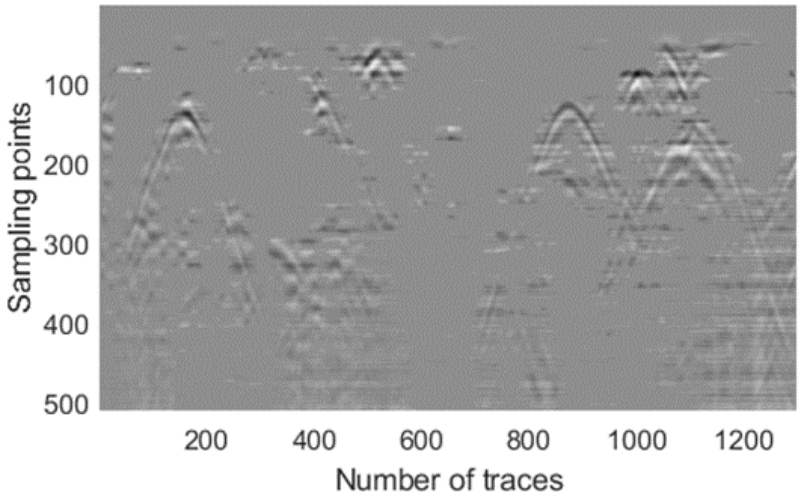
*Figure 4-4. a) Original radargram 009 with allocated dashed lines using PS1, b) By the KF processed radargram 009 using PS1*

As shown in Figure 3-21 in the previous section (page 109), there is a pipe close to the 700th trace in the radargram (Figure 4-2b) that has a broken section with a blockage between the 12th and 15th traverses. The processed radargrams in Figure 4-5b and Figure 4-6b do not show any hyperbola signature related to this pipe. However, the existence of the dashed lines generated by the algorithm, clearly indicates a potential anomaly in this region. This indicates that the proposed algorithm is also capable of detecting a defective target section (e.g., a broken pipe section) and corresponding ground anomalies using dashed lines.

In any case, the benefit of the proposed algorithm is evident by comparing the processed radargrams on the right-hand side with those on the left. The likely positions of the anomalies are highlighted on the original radargrams by a series of dashed lines that were not obvious otherwise. Additionally, the algorithm also detected additional near-surface hyperbolas alongside the actual five hyperbola signatures. These additional near-surface anomalies such as pebbles, tree roots, and rabbit/mole holes could be the reason for these additional hyperbola features.

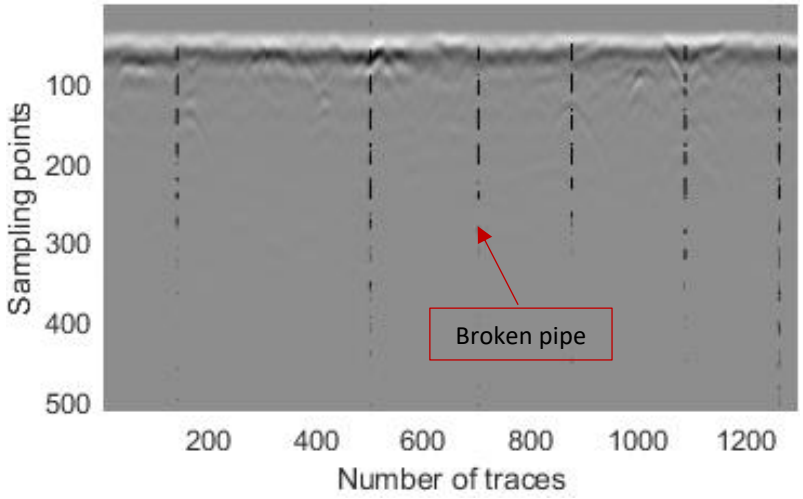



*(a)*

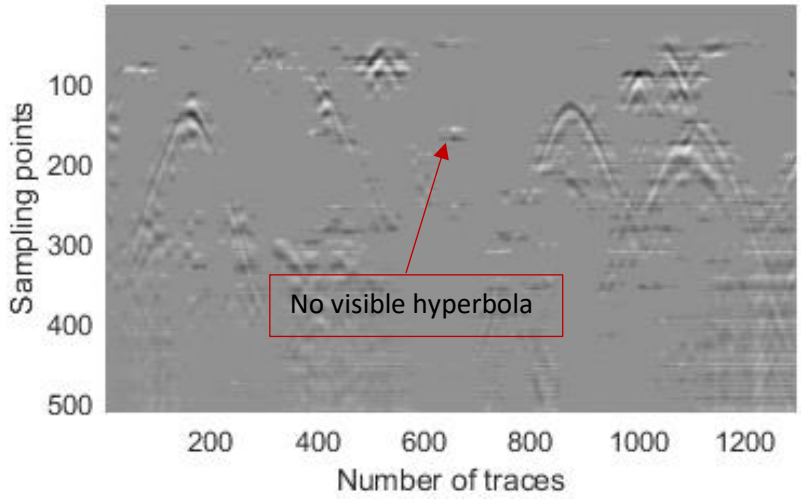



*(b)*

*Figure 4-5. (a) Original radargram 013 with allocated dashed lines using PS4, (b) By the KF processed radargram 013 using PS4*




*(a)*




*(b)*

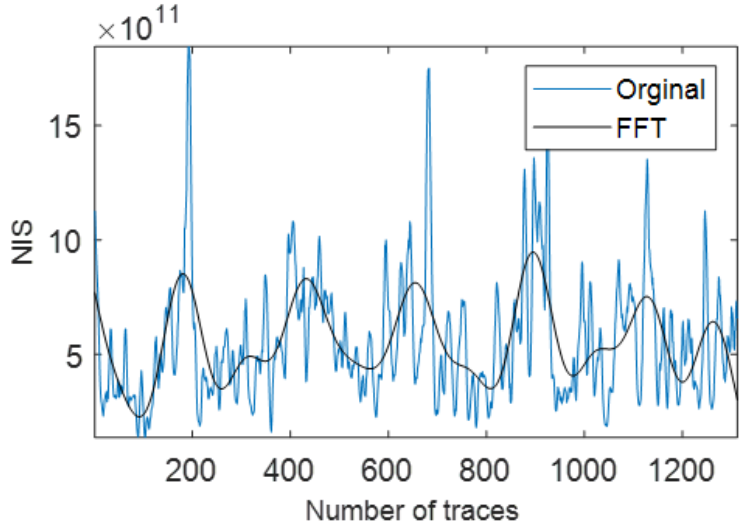
*Figure 4-6. (a) Original radargram 013 with allocated dashed lines and detected ground anomaly by KF algorithm using PS1, (b) By the KF processed radargram 013 and detected ground anomaly using PS1*

## 4.3.2 Evaluating the performance of the MOGA optimisation of KF

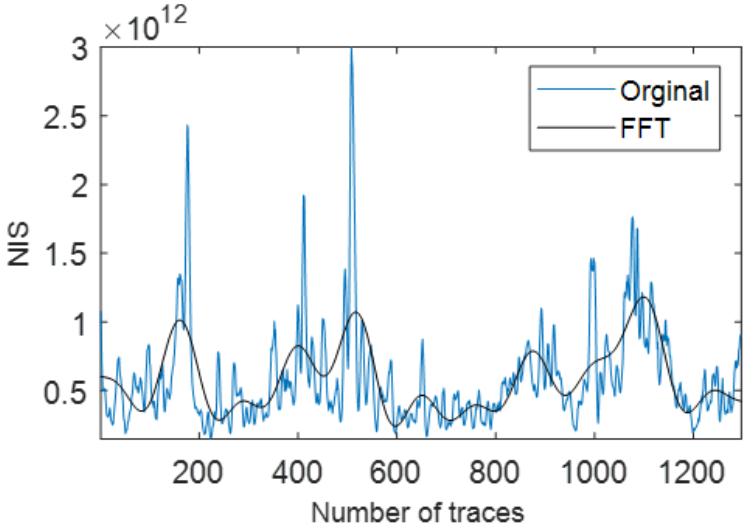
To determine a benchmark for comparison, the results of the MOGA optimisations (the values of covariances Q and R) were compared against those obtained from a trial-and-error approach. The significant difference between MOGA-based generated normalised innovation squared (NIS) signals and experimental-based NIS signals is illustrated in Figures 4-7 and 4-8. As shown on the vertical axis of each plot (Figure 4-7a and 4-7b), the experimental-based NIS

signals had values in the order of $\times 10^{11}$, and $\times 10^{12}$ for radargrams 009 and 013, respectively, compared to MOGA-based NIS signals using PS4 which had values in the order of $\times 10^{-23}$ , and $\times 10^{-22}$ for radargrams 009 and 013 respectively (Figure 4-8a and 4-8b).

The high performance and accuracy of the proposed KF algorithm in this study is evident from the almost zero order of MOGA-based NIS signals, which fully comply with the theory of the Kalman filter. As indicated by Simon (2006), smaller values of NIS indicate a higher performance of the algorithm.




*(a)*




*(b)*

*Figure 4-7. (a) NIS signal for experimentally obtained Q-R from radargram 009, (b) NIS signal for experimentally obtained Q-R from radargram 013*

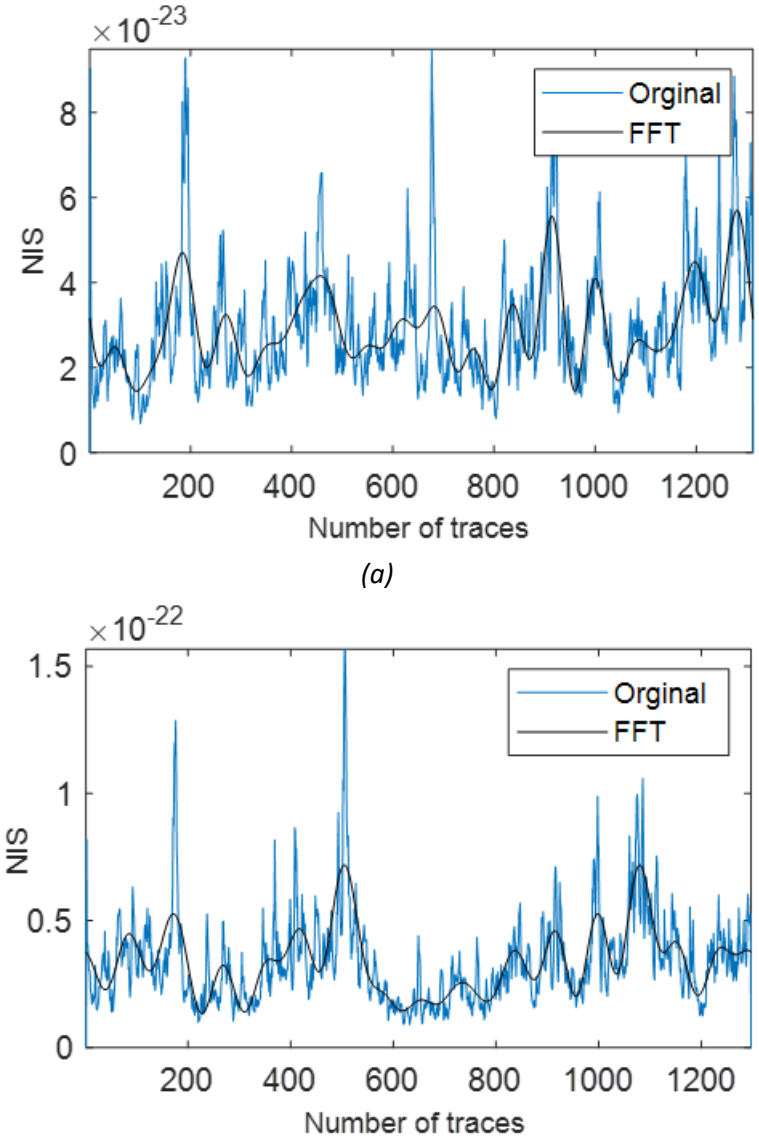



*Figure 4-8. a) NIS signal for optimised Q-R from radargram 009 using PS4, b) NIS signal for optimised Q-R from radargram 013 using PS4*

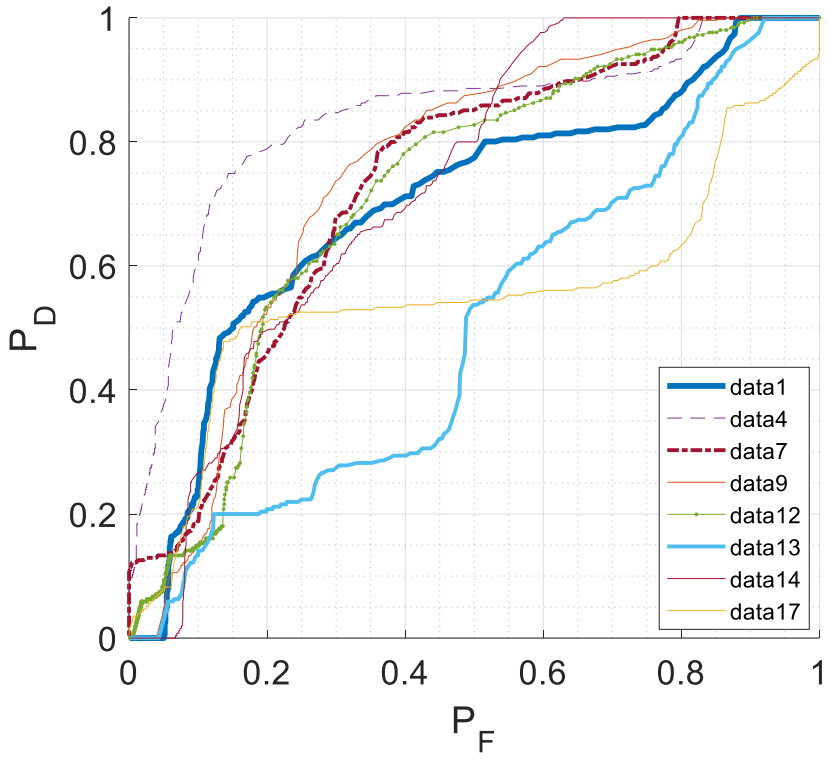
## 4.3.3 Generated ROC curves from Pareto solutions (PSs)

In this study, the area under the ROC curve for each radargram was calculated for different Fourier frequencies for two points on the Pareto solution (PS) curve (PS4 and PS1). The corresponding covariance matrices, R and Q, were also determined. The ROC curves for all

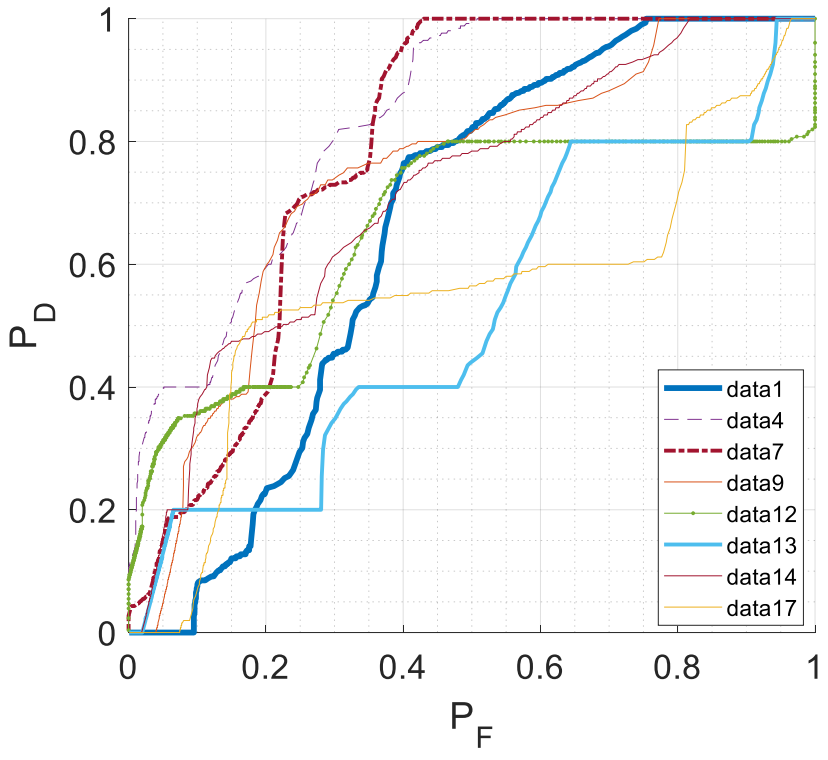
radargrams in the dataset are illustrated in Figures 4-9a and 4-9b. A higher general average value for the ROC curves was observed for PS1 (Figure 4-9b) compared to the ROC curves for PS4 (Figure 4-9a). As previously mentioned, a higher value on the Y-axis of the ROC curve indicates a higher probability of target detection versus the probability of false alarms on the X-axis.

It was observed that the mean value of the area under the ROC curves changed significantly when Fourier frequencies (FFreq) were selected randomly outside of the PS. Weaker mean values were observed when utilising randomly selected FFreq, compared to the optimised values associated with the PS. For example, randomly selecting FFreq such as 20, 35, and 40 rad/sec, instead of those provided by PS4, yielded ROC mean values of 68.27%, 65.35%, and 65.45% respectively, whereas PS4 optimal FFreq such as 9, 10, and 18 rad/sec yielded higher ROC mean values of 71.29%, 73.05%, and 68.89%, respectively. The same pattern was observed for the PS1 results compared to PS4, but generally yielded higher mean values in a range of 69.78% to 74.03%. Two tables consisting of the associated values for different cases are presented in Appendix III of this thesis.

It is important to note that PS4 was the point on the Pareto solution curve where the weighted parameter was the innovation that was optimised in terms of its zero-mean characteristic. In other words, the optimisation algorithm generated PS4 in a way that guaranteed the zero-mean condition. On the other hand, PS1 was the point on the curve where the weighted parameter was the optimised area under the ROC curves, in which the maximum area under the curves was guaranteed. The results of this study indicate that the performance of the fast Fourier transform (FFT) is highly dependent on the selected frequencies threshold (frequency range) that can pass through the filter




*(a)*




*(b)*

*Figure 4-9. (a) Generate ROC curves for all radargrams in dataset using PS4, (b) Generate ROC curves for all radargrams in dataset using PS1*

As mentioned above, to verify the performance of the algorithm in terms of functionality and biased detections, three additional independent datasets were analysed using the proposed algorithm in this study, which will be discussed in section 4.5.

## 4.4 The autonomous algorithm

In this study, a novel autonomous algorithm based on Kalman filtering (KF) was proposed for the processing of ground-penetrating radar (GPR) data. The performance of the algorithm was validated using similar field data, as outlined in Section 4.2. The algorithm was able to process all 23 measured radargrams automatically, without the need for user interaction.

The algorithm's performance was evaluated through a comparison of the original noisy radargrams (such as those depicted in Figure 4-10a and Figure 4-10b) with the processed radargrams, which had the background signals removed (depicted in Figure 4-10c and Figure 4-10d). The results indicated that the algorithm effectively removed noise and clutter while maintaining the presence of shallow pipes, as indicated by the preservation of hyperbolas in the radargrams. However, additional signals were observed in the processed radargrams, which were determined to be related to changing permittivity contrasts (impedance) at specific locations. For example, a broken pipe at approximately trace 650 in Figure 4-10b, which was located in disturbed soil, is identified with a red circle in Figure 4-10d.

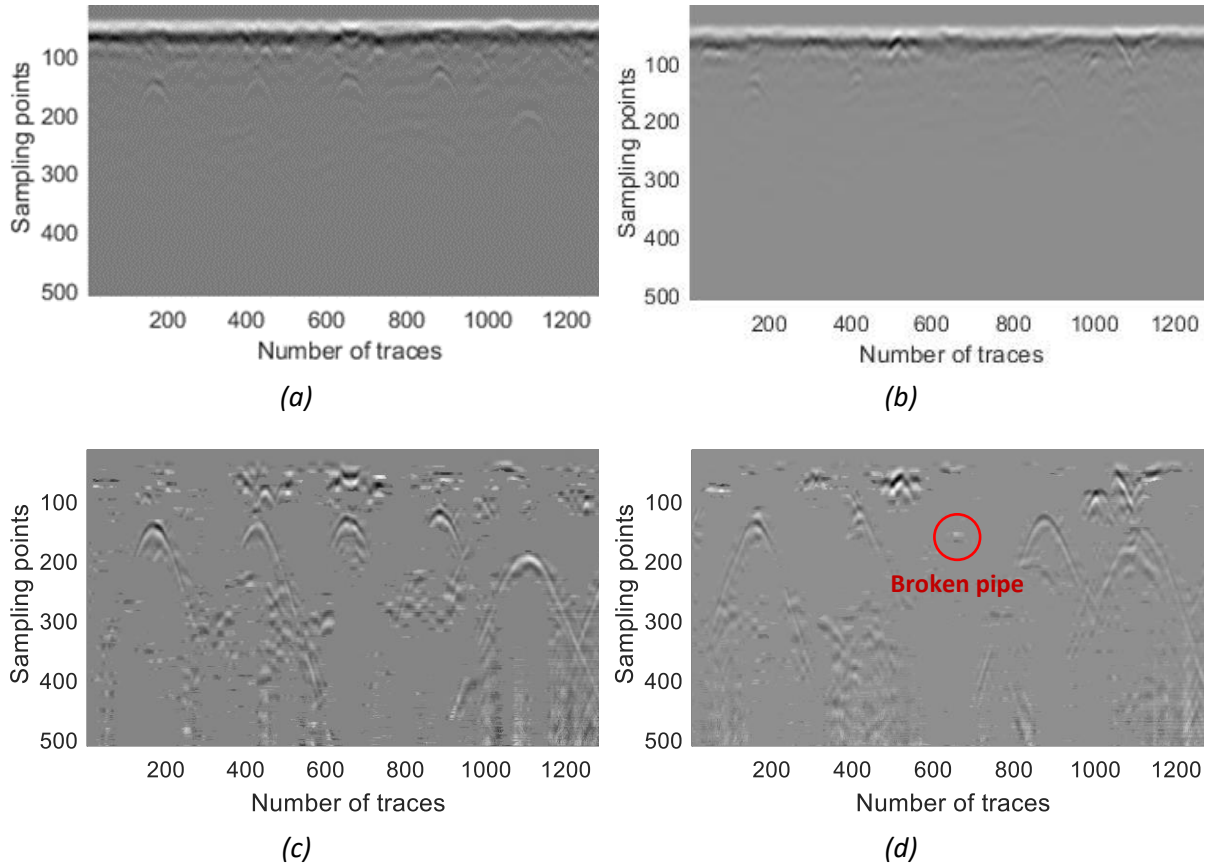



*Figure 4-10. a) Radargram 9 with no defective pipes (five fully intact pipes), b) Radargram 13 with a broken pipe section at approximately trace 650 (four fully intact pipes and a broken soil-filled pipe), c) Radargram 9 with the removed background. d) Radargram 13 with removed background*

Both the shallow pipes and ground anomalies were identified and automatically marked by the algorithm using dashed lines that were allocated at the position of the identified signal anomalies. A noticeable variation in the number of dashed lines, associated with true-positive and false-positive detections, in the processed radargrams was observed (Figures 4-11 to 4-13). This indicates a direct relationship between the level of sensitivity of the proposed algorithm and the decomposition level of the wavelet transform for selected mother wavelets (i.e. db, Sym, and dmey wavelets). Comparison of the results of a processed radargram under different decomposition levels i.e. $l = 3\ to\ 7$ and utilising selected mother wavelets (i.e. db, Sym, and dmey wavelets) illustrated significant changes in sensitivity of the algorithm in terms

of the identified anomalies. For example, Figure 4-11 illustrates the results associated with selected radargram 9 (consisting of five fully intact pipes) from the dataset used in this study, after utilising the db10 mother wavelet under different decomposition levels, i.e. $l = 3, 6, \text{and } 7$. The visible sensitivity difference in results (Figure 4-11) can be explained using the analysis of the corresponding NIS signals (Figure 4-11(a), (b), and (c) left hand side figures) and assigned dashed lines associated to signal anomalies (Figure 4-11(a), (b), and (c) right hand side figures). It can be observed that the number of dashed lines is in accordance with the smoothness of the transformed NIS signal (red signals) generated at different decomposition levels. In other words, the higher the decomposition level the smoother the transformed (filtered) NIS signal with a reduced number of dashed lines, and vice versa.

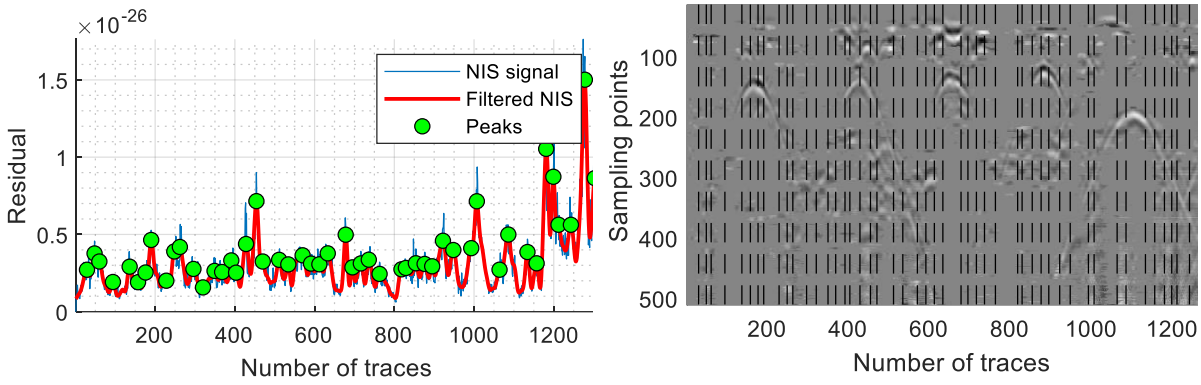
On this basis, a high sensitivity level can help to detect all detectable anomalies (impedance) in a radargram, but at the same time can yield a higher false-positive detection rate (false alarms). Nevertheless, the rate of true- and false-positive detections can be optimised by the user and based on a range of factors such as the project requirements, the boundary conditions, and the quality of the acquired data.

A comparison of the results was conducted for all three mother wavelets under similar and different decomposition levels. This was done to determine the performance and the difference in the sensitivity of the proposed algorithm in terms of the number and location of detected signal anomalies. Further examples of the results are shown in Figure 4-12, which shows the performance of three selected mother wavelets (db20, Sym20, and dmey) under the same decomposition level (i.e. level 6). Figures 4-12(a), (b), and (c) (left hand side figures) show the generated and transformed (filtered) NIS signals where Figures 4-12(a), (b), and (c) (right hand side figures) show the associated processed radargram with removed background

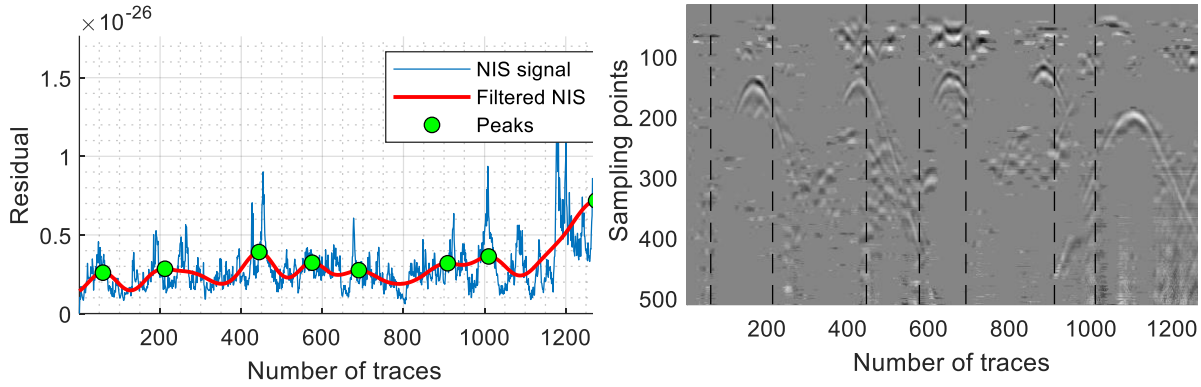
signals and allocated dashed lines. A comparison of the results shows that in this specific example the dmey wavelet (Figure 4-12c) performed better by not missing any of the existing hyperbolas. However, the hyperbola around the 900$^{th}$ trace (Figure 4-12b) could not be detected by the Sym20 wavelet.

Another example result is presented in Figure 4-13 associated with another radargram (13) consisting of four fully intact pipes and one pipe with a broken section (filled with soil) using the same three mother wavelets (db20, Sym20, and dmey) under a similar decomposition level (level 6). A better performance, in terms of the detected anomalies including both target and ground anomalies associated with the broken pipe section, can be observed for the dmey wavelet. This can be explained due to the properties of the dmey wavelet in the frequency domain, which were found to be better suited to the characteristics of the NIS function signal being analysed. The time localisation and frequency localisation properties of the dmey wavelet were found to be better for this specific GPR dataset, resulting in a more accurate representation of the NIS signal function. Furthermore, the shape of dmey wavelet is designed to have high frequency resolution for representing abrupt changes which is suitable for some type of signals such as the NIS signal function in this experiment. This also contributes to its improved performance in comparison to the db and Symlet wavelets in this test. However, it should be noted that the above explanations are limited to the examined data and cannot not be generalised without conducting comprehensive analysis using more datasets.

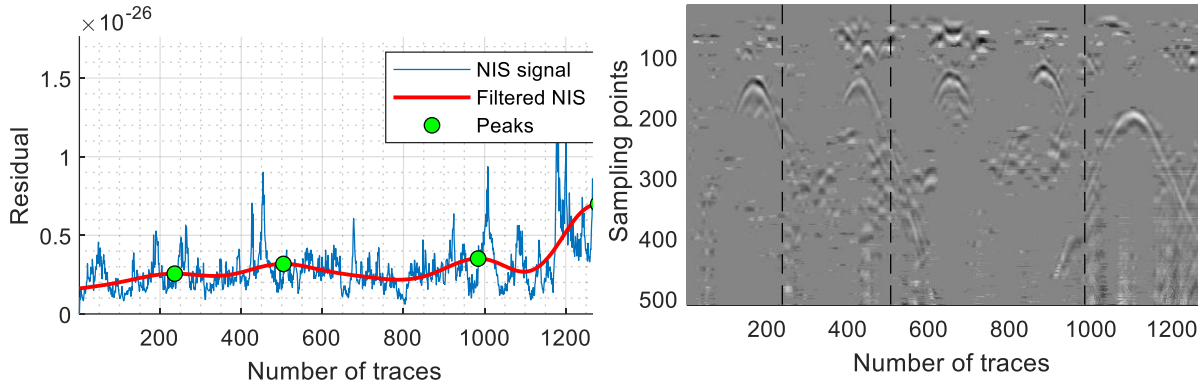
As shown in Figures 4-13(a), (b), and (c) (right hand side figures), except for the Sym20 wavelet, the broken section was successfully detected by utilising dmey (accurate detection) and db20 (low divergence) under decomposition level $l = 6$ . The dashed lines are assigned at approximately trace 650 where the ground anomaly (soil-filled pipe section) is located.




*(a) Decomposition level (l) = 3*

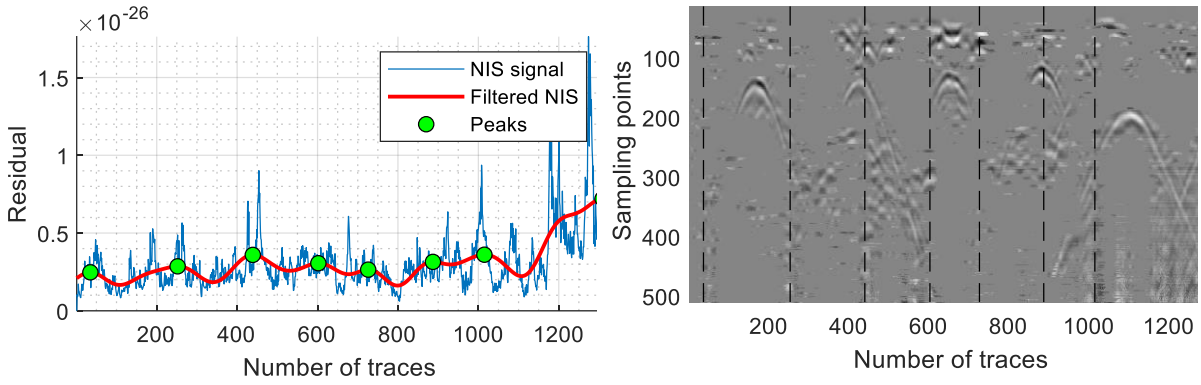



*(b) Decomposition level (l) = 6*

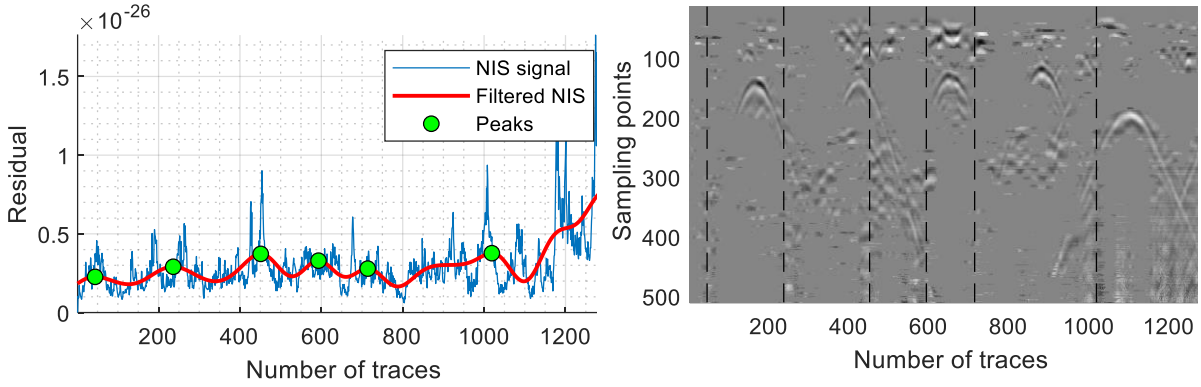



*(c) Decomposition level (l) = 7*

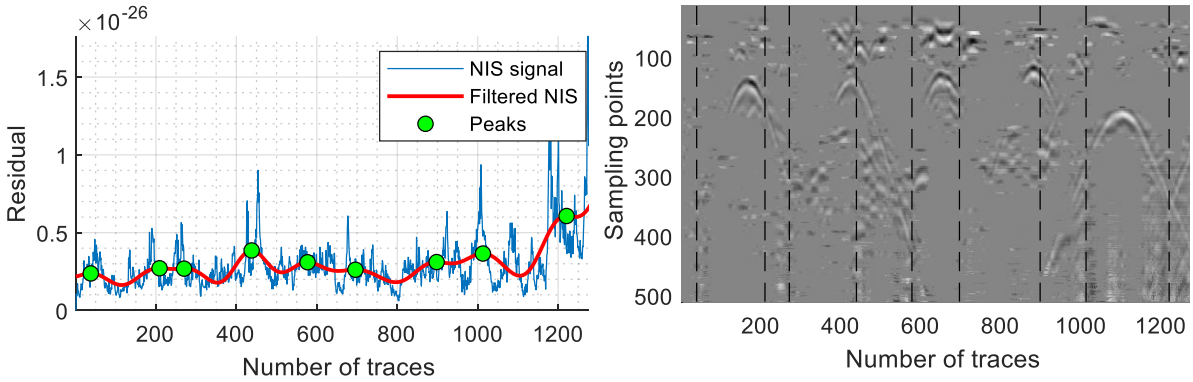
*Figure 4-11. Processed NIS signals (left) and associated identified anomalies (right) for radargram 9 with five fully intact pipes, using db10 Daubechies wavelet under decomposition levels of a) l = 3, b) l = 6, and c) l = 7, showing the effect of wavelet decomposition level on the sensitivity of the algorithm*




*(a) Wavelet: db20*

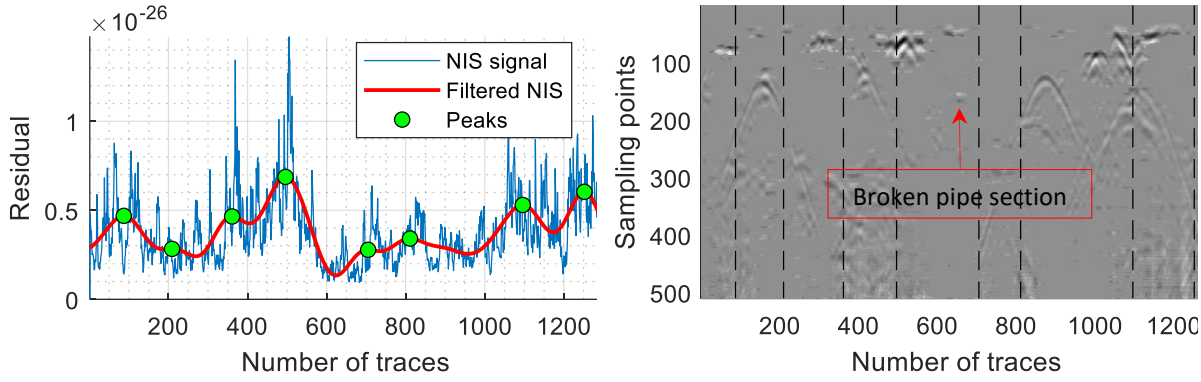



*(b) Wavelet: Sym20*

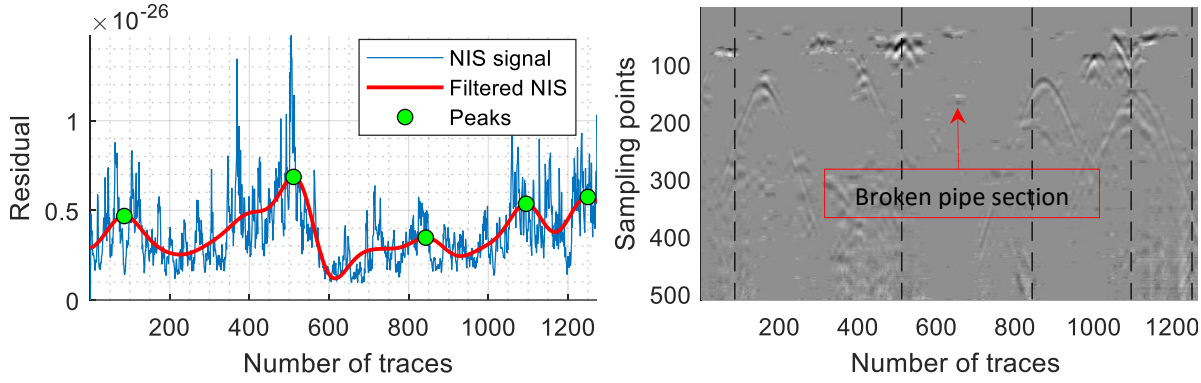



*(c) Wavelet: dmey*

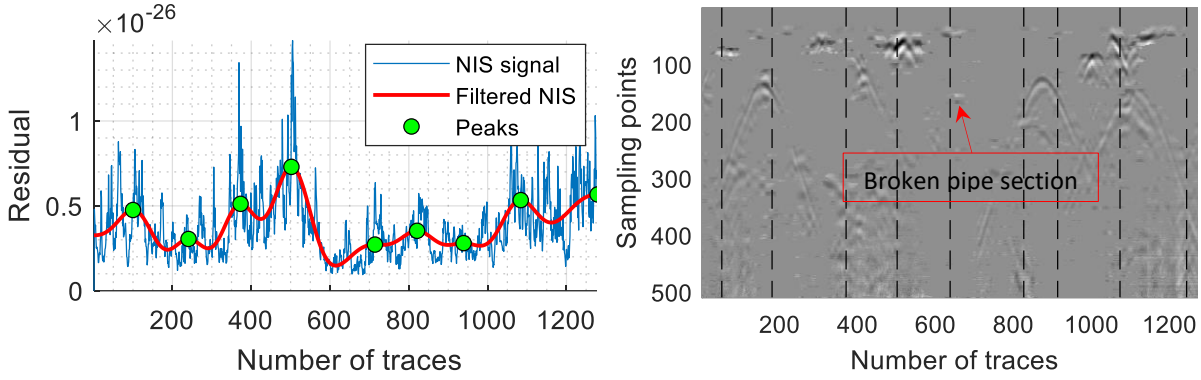
*Figure 4-12. Processed NIS signals (left) and associated identified anomalies (right) for denoised radargram 9 (FINDIT data) using the proposed autonomous algorithm with three different mother wavelets under the decomposition level (l) = 6, a) db20 wavelet, b) Sym20 wavelet, c) dmey wavelet*




*(a) Wavelet: db20*




*(b) Wavelet: Sym20*




*(c) Wavelet: dmey*

*Figure 4-13. Processed NIS signals (left) and associated identified anomalies (right) for denoised radargram 13 with 4 intact pipes and 1 broken pipe at approximately trace 650 (FINDIT data) using the proposed autonomous algorithm with three different mother wavelets under the decomposition level (l) = 6, a) db20 wavelet, b) Sym20 wavelet, c) dmey wavelet*

A statistical sensitivity analysis of the results was conducted for all the different variants of all three mother wavelets, i.e. mother wavelets with every possible vanishing moment under different decomposition levels (i.e. 3 to 7).

An example result of the statistical sensitivity analysis is shown in Figure 4-14 associated with radargram ten presented in the form of various heatmaps associated with each of the utilised mother wavelets. The values in the presented heatmaps are the scores based on the calculated AuC of ROCs. The higher the value of each box, the higher the score for the vanishing moment of the mother wavelet and the decomposition level of the wavelet transfer. Analysis of all the available radargrams in the provided dataset has shown that the decomposition levels $l = 5$ and $l = 6$ are the most appropriate levels for analysing the NIS signal in this study. However, no certain vanishing moment(s) can be determined for each of these mother wavelets at this stage. The analysis of all the results has also shown that the dmey as a single variant mother wavelet (i.e. no variation in the vanishing moments) can be a suitable candidate for analysing the NIS signal. In other words, the utilised dmey under the appropriate decomposition level (e.g. 5 and 6) was able to identify all buried targets (and ground anomalies) and the only required factor was to determine the desired decomposition level. It is important to mention that the aforementioned statements are only limited to the dataset used in this study and cannot be generalised unless a large number of datasets are tested and analysed using the proposed algorithm in this study.

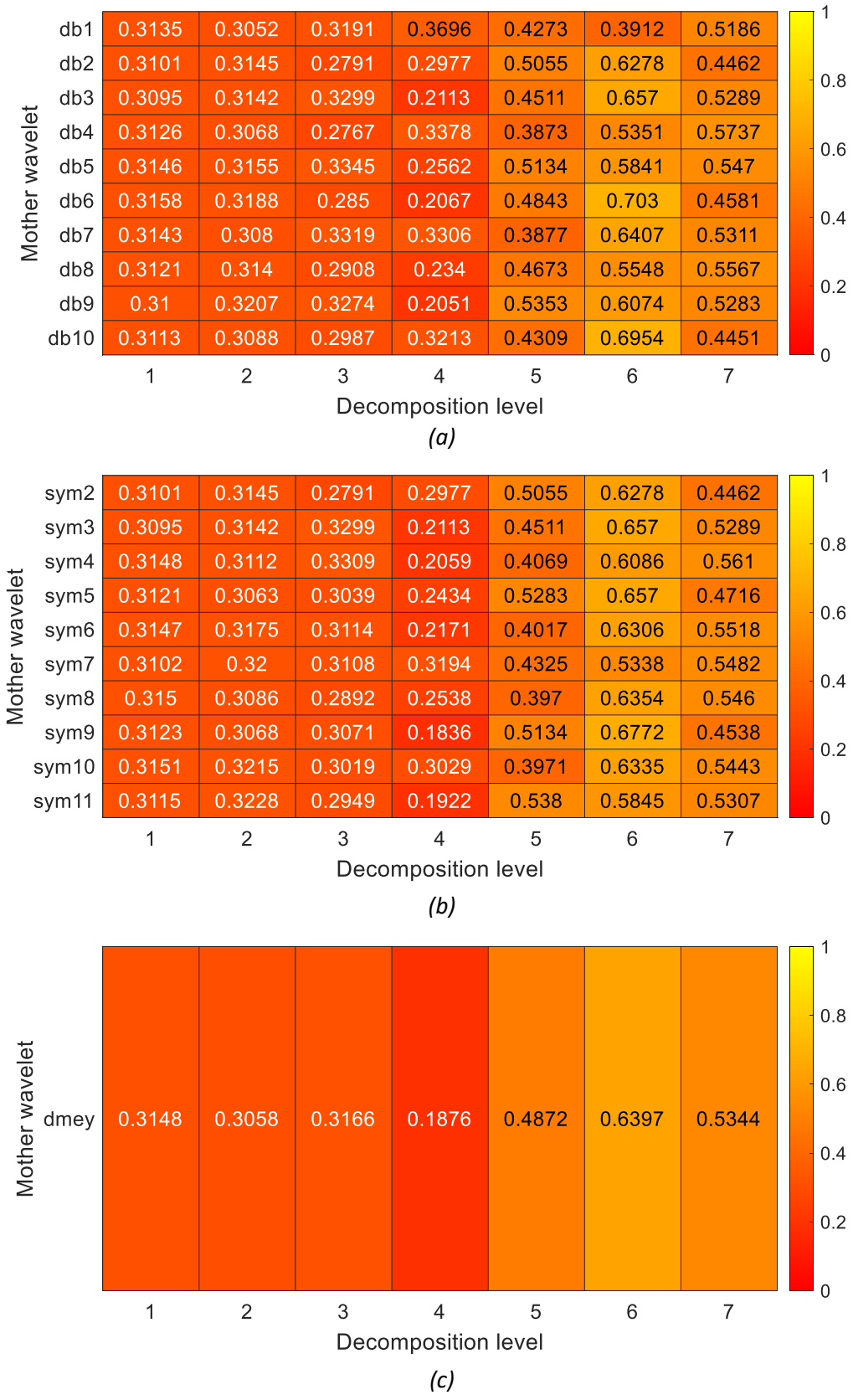



*Figure 4-14. Heatmaps are associated with the scoring function for identifying the best mother wavelet for analysing the NIS signal. a) Heatmap for evaluation of the Daubechies wavelet, b) Heatmap for evaluation of the Symlet wavelet, c) Heatmap for evaluation of the Discrete Meyer wavelet*

## 4.5 Verification of the proposed algorithms

In this section, three different and independent GPR datasets were employed to validate the performance of both the proposed semi-autonomous and autonomous algorithms in this research (i.e. Objective 8). The results for each of the proposed algorithms are presented separately.

To validate the performance of the proposed algorithms, a series of field survey data interpreted by an expert surveyor were compared to generated results by the developed algorithm. The comparison of the results showed that all buried utilities which were identified by the expert surveyor could also be identified by the proposed algorithm of this study. In addition to targets identified by the expert surveyor, the proposed algorithm placed warning dashed lines where potential anomalies could be identified.

### 4.5.1 Case study one

#### The semi-autonomous algorithm

The first data set was obtained from the north section of the same test site presented in section 4.2, i.e. first survey site, with the same ground conditions but different pre-defined buried utilities. The associated PSs (Figure 4-15) showed healthy values for ROC curves above 91.4% (Y-axis) among an almost zero mean of innovation values in the order of $\times 10^{-4}$ (X-axis).

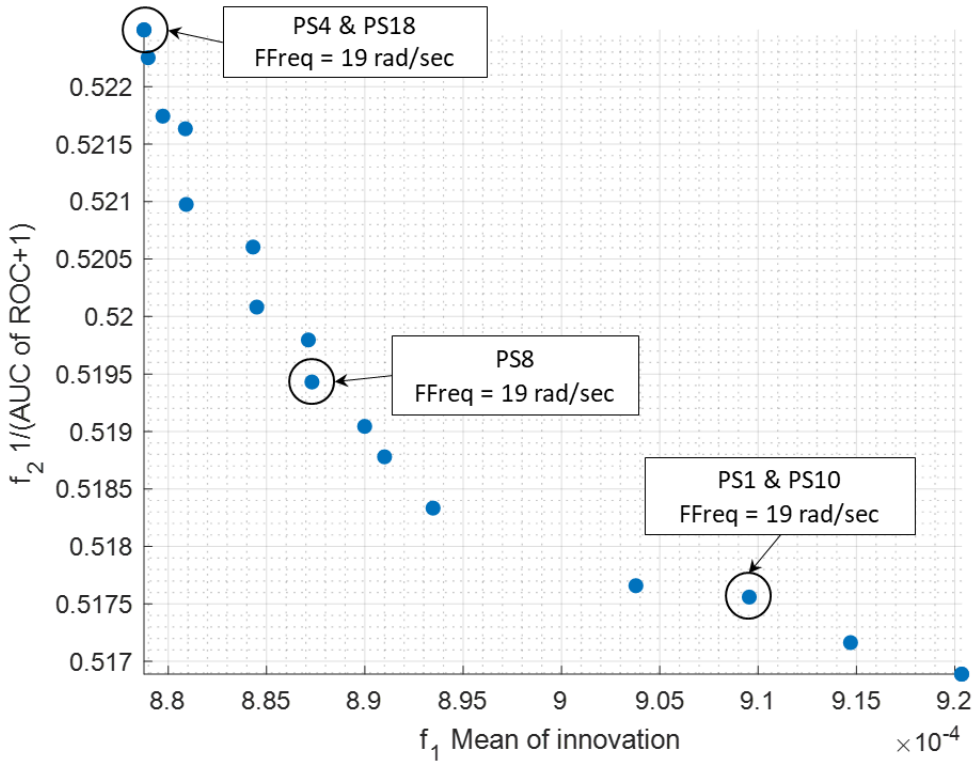



*Figure 4-15. Pareto solutions for the north section of the first survey site with different buried utilities*

The outcome of the algorithm for the two survey radargrams, presented in Figure 4-16(a) and (c), using PS4 (or PS18 with similar characteristics) from the associated PS curve are illustrated in Figures 4-16 (b) and 4-16 (d), respectively. The estimated positions of possible anomalies, including the target positions, were detected by the algorithm, and are shown on Figures 4-16(a) to (d) using vertical dashed lines on both original and processed radargrams. The results show that all the utilities were successfully identified. There are some additional warning lines that can be considered as false alarms. However, as mentioned earlier, such warning lines are generated by the presence of impedance contrast in the ground. In other words, the warning lines signify anomalies in the composition and material properties of the ground.

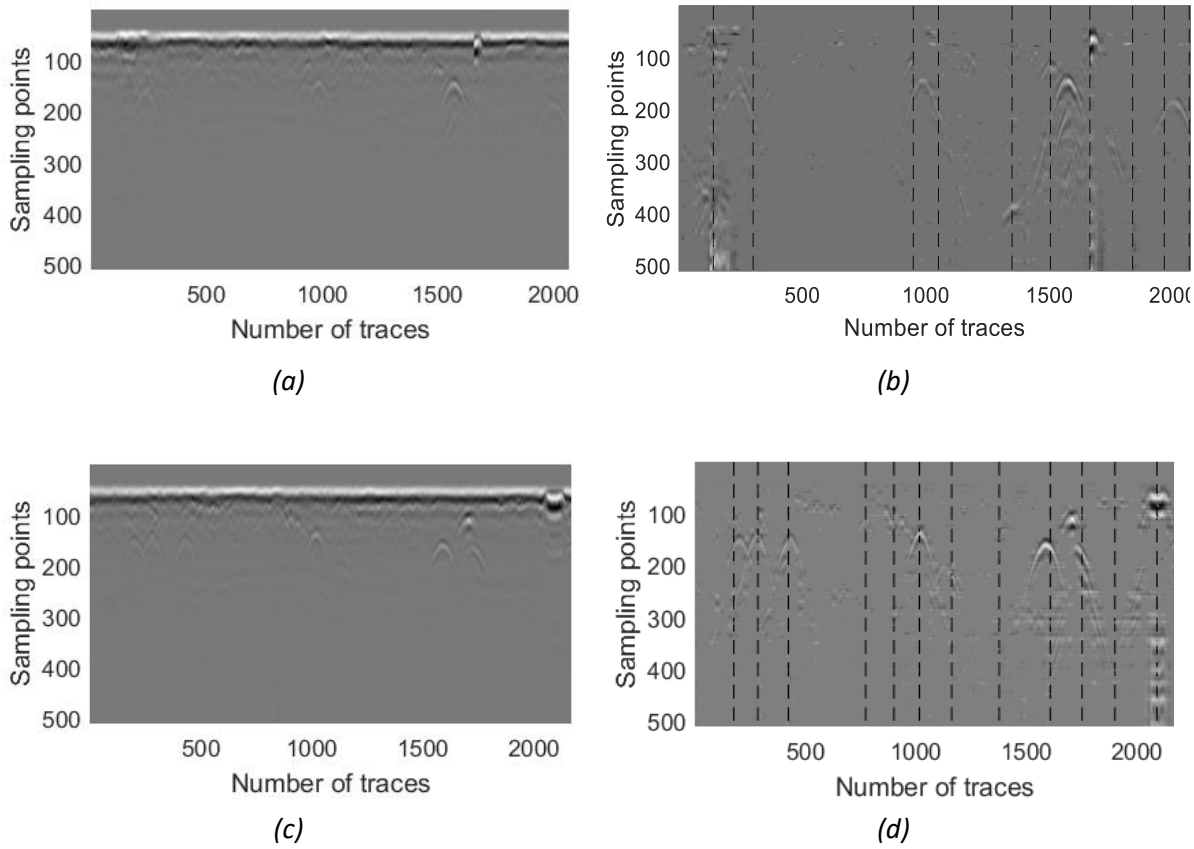



*Figure 4-16. Radargrams from the north section of the first survey site for two different cases of buried utilities. a) Original radargram 7, b) processed radargram 7, c) original radargram 10, d) processed radargram 10*

## The autonomous algorithm

Figure 4-17 and 4-18 show the processed radargrams 7 and 10 from the previous subsection (Figure 4-16(a) and Figure 4-16(b), respectively), which are produced by the proposed autonomous algorithm in this research. To keep the comparison consistent, the results associated with the same three mother wavelets in section 4.4 (i.e. db20, Sym20, and dmey) and under the decomposition level of $l = 6$ are presented in these figures.

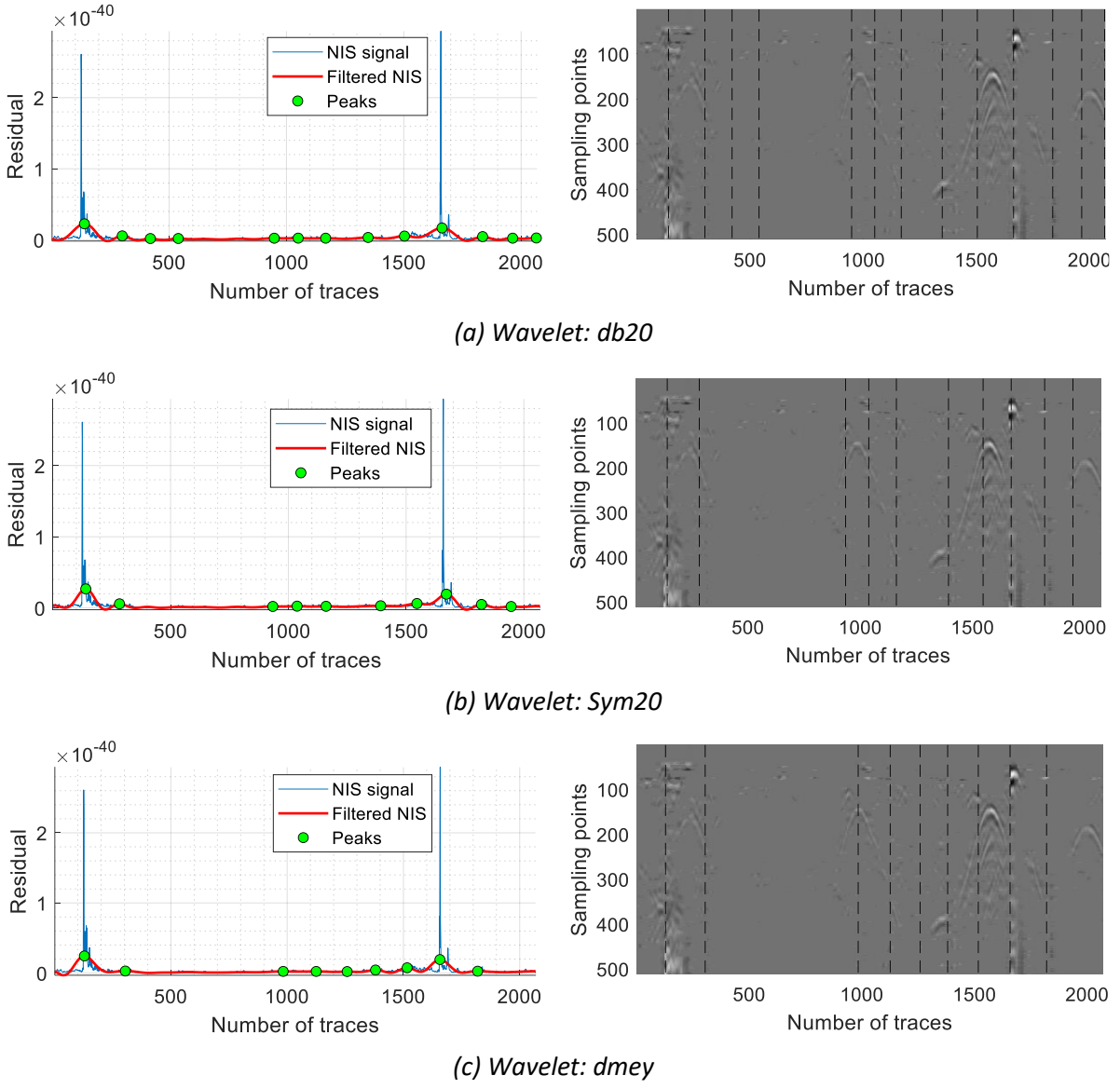
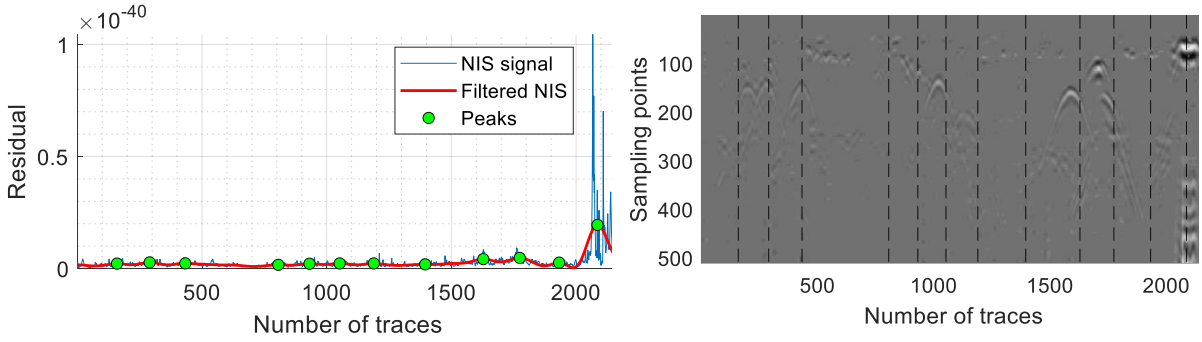


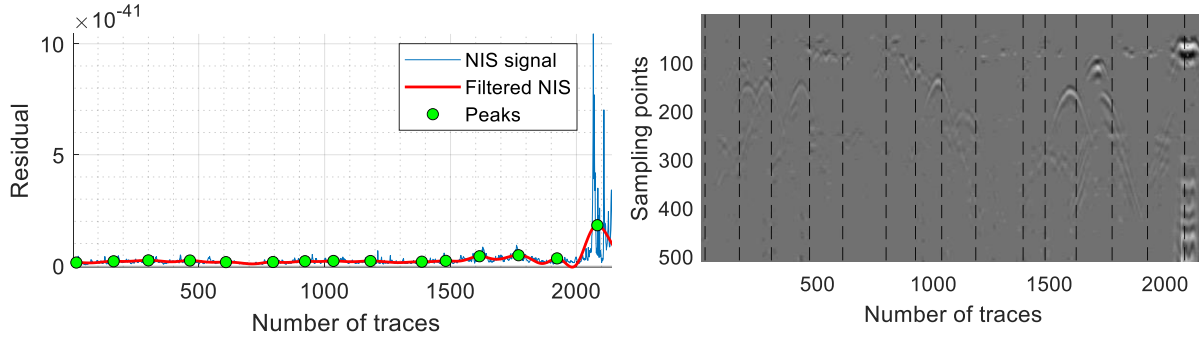

*Figure 4-17. Processed NIS signals (left) and associated identified anomalies (right) for denoised radargram 7 (case study one) using the proposed autonomous algorithm with three different mother wavelets under the decomposition level (l) = 6, a) db20 wavelet, b) Sym20 wavelet, c) dmey wavelet*

It can be observed in Figure 4-17 and Figure 4-18 that regardless of the selected mother wavelets, the noise and clutter are successfully eliminated by the algorithm for all the radargrams. In terms of identified targets and anomalies, the results in Figure 4-17 show that the selection of mother wavelets for achieving high performance of the algorithm is a problem-dependent task, which confirms the discussions in section 4.4. In other words, it can be observed that, apart from the detected ground anomalies, db20 and Sym20 wavelets could

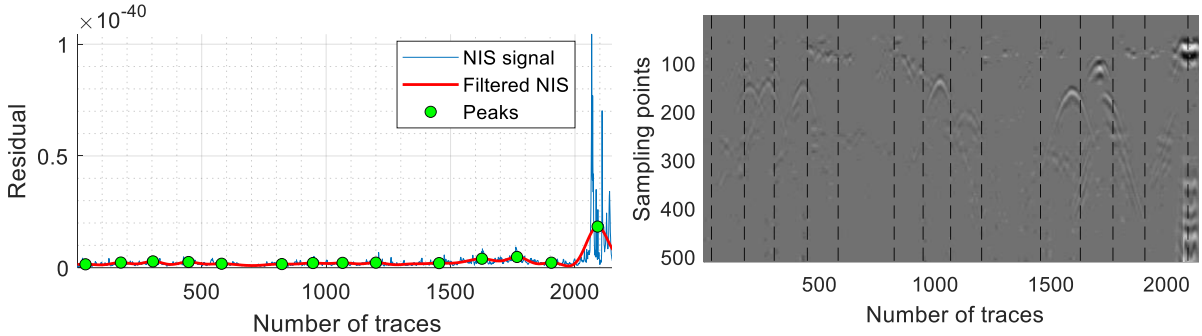
detect all the hyperbola signatures whereas dmey failed to detect the hyperbola at the ca. 2000$^{th}$ trace. Nevertheless, all the mother wavelets provided high performance for processing the radargrams in Figure 4-18 (aligned with the statistical sensitivity analysis in the previous section).




*(a) Wavelet: db20*




*(b) Wavelet: Sym20*




*(c) Wavelet: dmey*

*Figure 4-18. Processed NIS signals (left) and associated identified anomalies (right) for denoised radargram 10 (case study one) using the proposed autonomous algorithm with three different mother wavelets under the decomposition level (l) = 6, a) db20 wavelet, b) Sym20 wavelet, c) dmey wavelet*

## 4.5.2 Case study two

In this section, the results associated with the data from the NBIF experiment (section 3.6) are presented and discussed. The experiment replicated a scenario with three pipes with different materials and diameters (2 m HDPE, 0.1 m PVC, 0.048 m galvanised steel) buried in sand. In addition, a gym ball of 1.2 m diameter replicated a void next to the large pipe. Figure 4-19 shows the original radargram 4 from the dataset selected for this case study. Example results for both proposed algorithms are presented below.

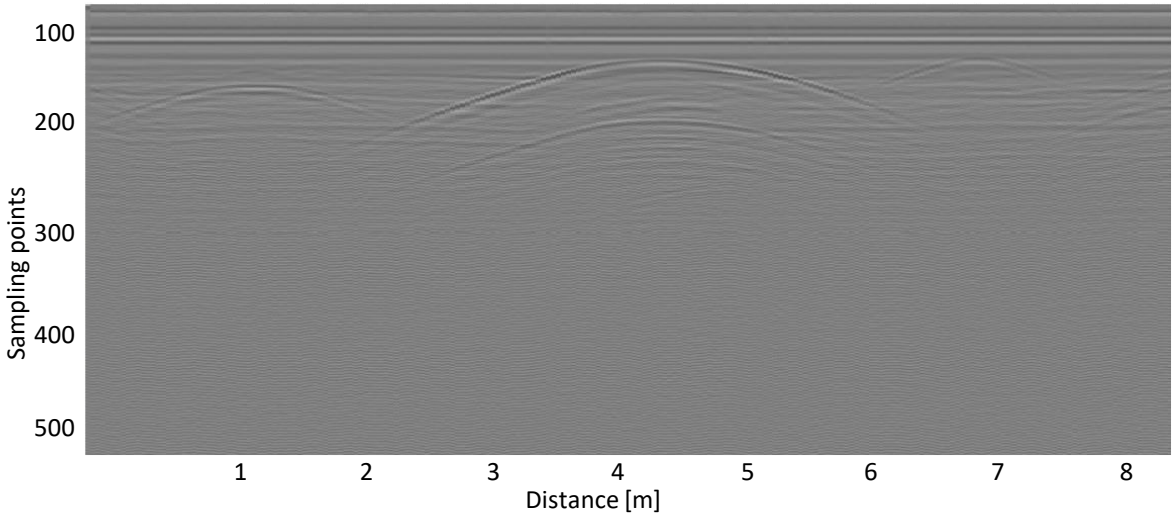



*Figure 4-19. Original radargram 4 from the experiment in NBIF consists of three pipes and a void buried in sand*

### The semi-autonomous algorithm

All radargrams in the dataset were processed using the proposed semi-autonomous algorithm for which the Pareto solutions in Figure 4-20 were generated. It can be observed that four PSs are generated by the MOGA where three of them have the same values (i.e. PS1, PS2, and PS4). The PSs are differentiated by their Fourier frequencies (FFreq), i.e. 44 rad/sec which

indicates a lower AuC of ROC curves but an optimal mean of innovation values, whereas the 3 rad/sec indicates a higher AuC of ROC curves but a greater mean of innovation values.

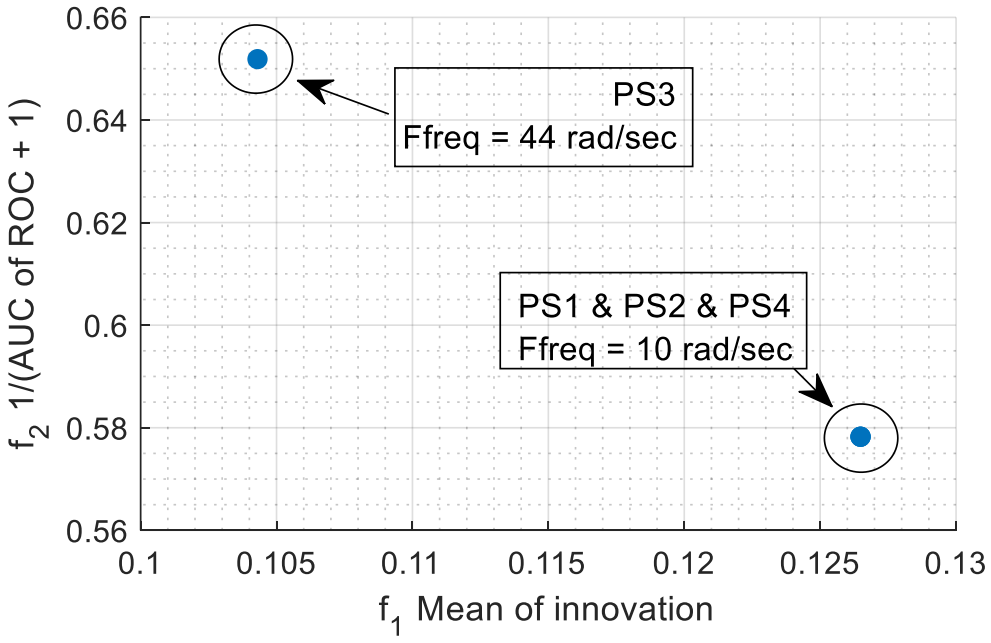
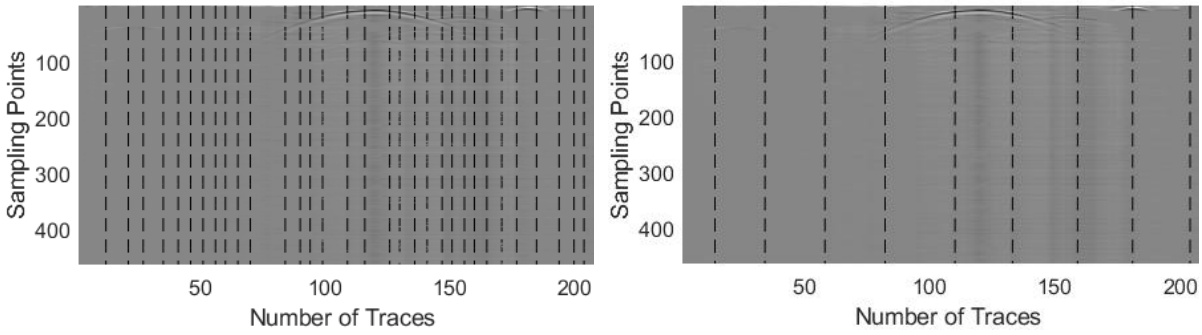



*Figure 4-20. Pareto solutions for the NBIF GPR experiment*

As discussed in section 4.3, all the buried utilities must be detected even if few false alarms exist in the processed radargrams by the algorithm. On this basis, the PSs PS1, PS2, and PS4 (FFreq = 10 rad/sec) were initially selected for processing the radargrams in this dataset. However, as shown in Figure 4-21a, the number of false positive warning lines is very high. Based on the previous analysis in section 4.3, it is expected that PSs with FFreq = 44 rad/sec delivers an optimal result with a relatively low number of false alarms, which is not the case here. On the other hand, the number of false alarms in Figure 4-21b, which is the result of utilising the PS3 (FFreq = 44 rad/sec), is very low and makes the PS3 the optimal solution for processing the NBIF GPR dataset. This issue can be the result of the homogeneity of the ground and therefore a very low level of environmental noise in this dataset. In other words, the generated NIS signal by the algorithm is significantly less affected by such low-level

environmental noise. As a result, the highest AuC of ROC provides results with lower false positives for such a scenario. It should be noted that such a statement cannot be generalised until an inclusive database of different GPR datasets are analysed, which is out of the scope of this study.

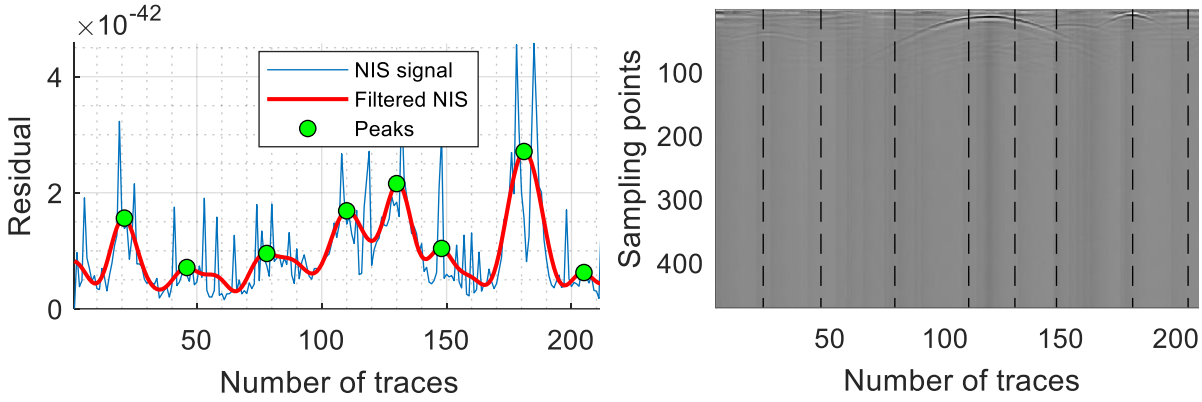



*(a)* *(b)*

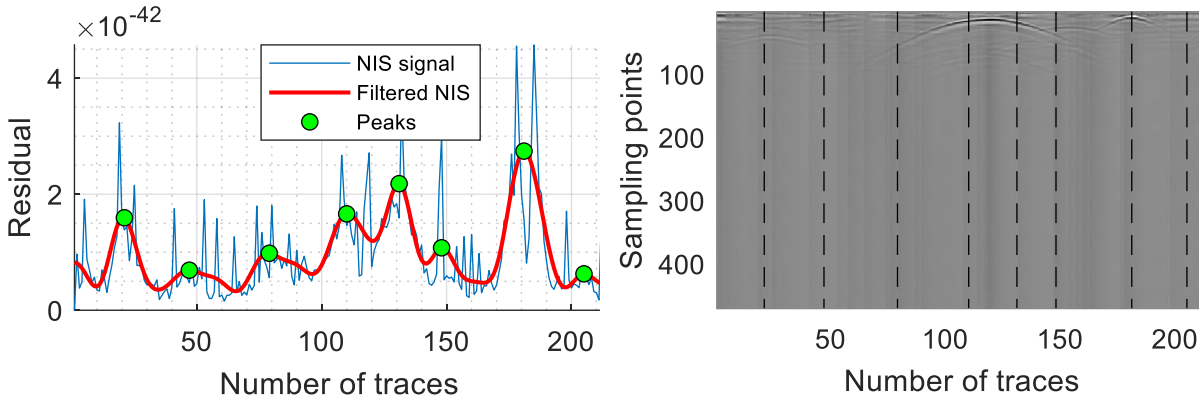
*Figure 4-21. Processed radargram using the semi-autonomous algorithm in case study 2*

### The autonomous algorithm

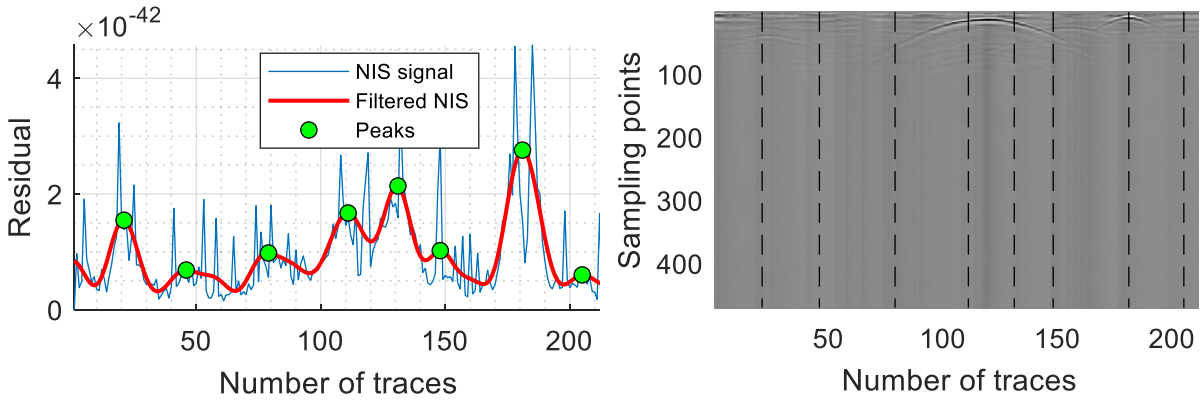
Analysing the processed radargrams using the proposed autonomous algorithm in this section showed similar results to those obtained from the semi-autonomous algorithm. In other words, as shown in Figure 4-22, the low level of environmental noise in the current dataset required a lower decomposition level e.g., $l = 3$ (instead of 6 for a higher environmental noise in other datasets). It was also observed that different mother wavelets produced similar results, in terms of the detected targets and ground anomalies, without insignificant differences.




*(a) Wavelet: db20*




*(b) Wavelet: Sym20*




*(c) Wavelet: dmey*

*Figure 4-22. Processed NIS signals (left) and associated identified anomalies (right) for denoised radargram 4 from the NBIF experiment using the proposed autonomous algorithm with three different mother wavelets under the decomposition level (l) = 3, a) db20 wavelet, b) Sym20 wavelet, c) dmey wavelet*

## 4.5.3 Case study three

Data from a third site, an industrial unit geographically located far from the first two, were used to further validate the extent to which the proposed algorithm can detect buried anomalies. Unlike the previous two GPR datasets, for this third dataset there was no initial information available for the ground condition or the buried objects, except for the type of the surface pavement, which was asphalt. The GPR data consisted of 52 radargrams. It is worth noting that, similar to the previous case studies, the data presented in this section were also processed by an expert surveyor as a benchmark for validating the performance of both the proposed semi-autonomous and autonomous algorithms. Both algorithms not only identified all the targets identified by the expert, but also the locations of the potential anomalies with changes in permittivity.

### The semi-autonomous algorithm

The proposed semi-autonomous algorithm was utilised to process all radargrams in the dataset. The associated Pareto curve was generated (Figure 4-23) and example results for the three selected radargrams from the dataset are presented in Figure 4-24.

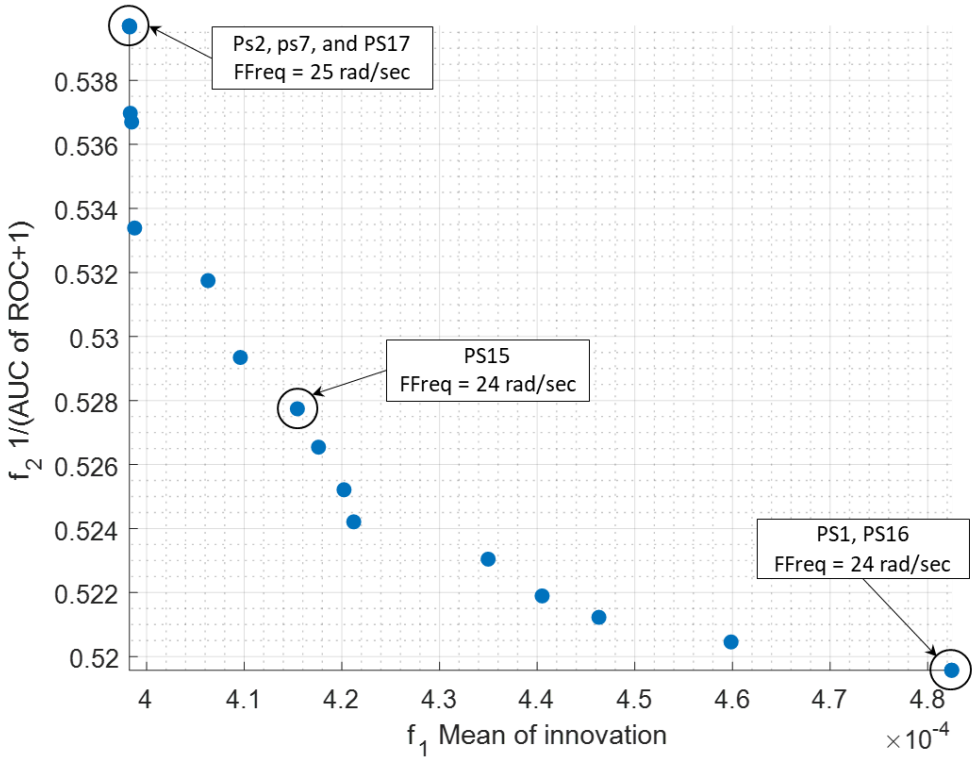



*Figure 4-23. Pareto solutions for the additional validation, case study 3*

After analysing the generated Pareto solutions, PS17 was selected as the most suitable solution as the proposed semi-autonomous algorithm was able to successfully identify all buried targets by utilising it without missing any targets. PS17 was then used to process all 52 radargrams in the dataset. The outcome of the algorithm for three radargrams of Figure 4-24 (a), (c), and (e) are presented in Figures 4-24 (b), (d), and (f), respectively. Observation of Figure 4-24b and Figure 4-24d shows that background signals are estimated and eliminated by the semi-autonomous algorithm, and the target related hyperbolas are successfully extracted. In addition, the potential target hyperbolas, and the ground anomalies are automatically indicated by the algorithms using vertical warning dashed lines.

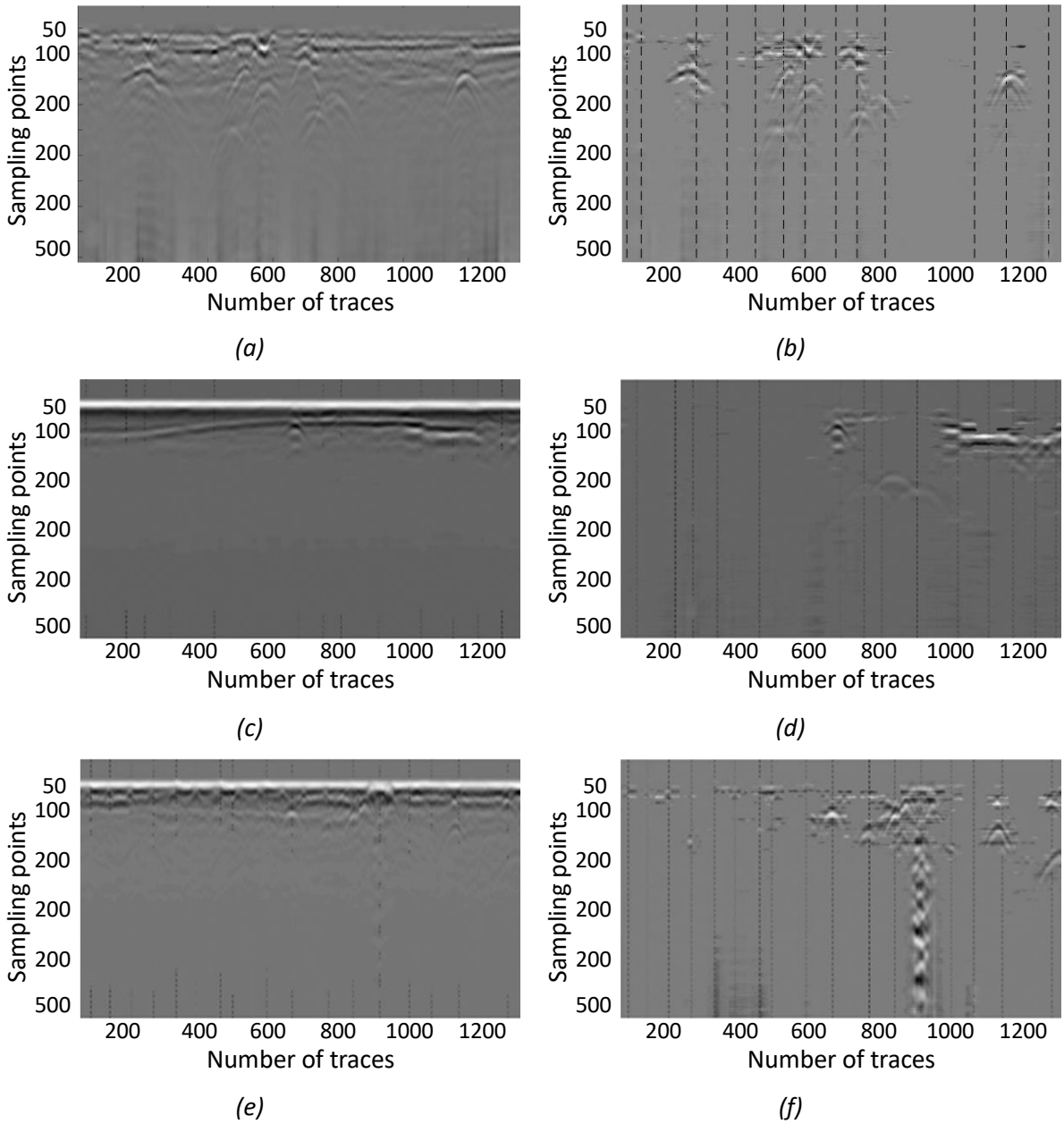



*Figure 4-24. Additional validation of the performance of the proposed semi-autonomous algorithm using three different radargrams from the third site survey. All three processed radargrams are showing a successfully removed background, identifiable hyperbola signatures, and allocated dashed lines at the position of potential targets or ground anomalies. (a) original radargram 60, (b) processed radargram 60 (c) original radargram 19, (d) processed radargram 19, (e) original radargram 39, (f) processed radargram 39.*

## The autonomous algorithm

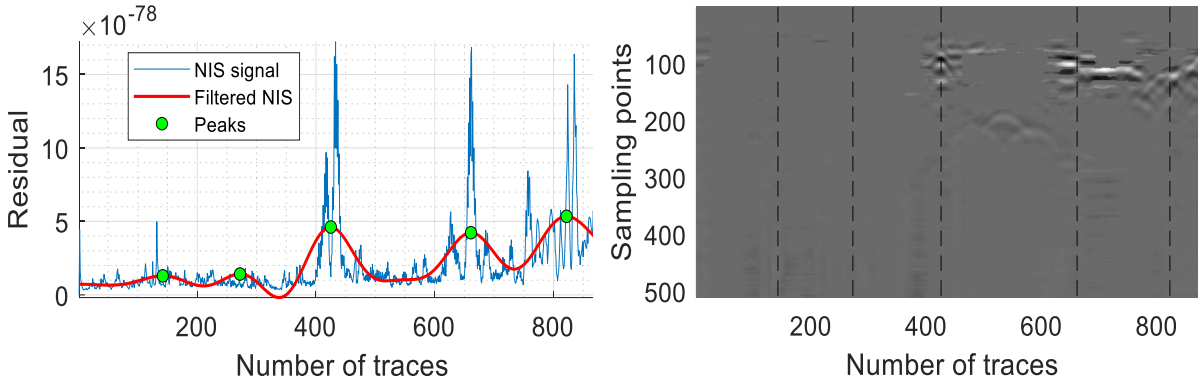
In this subsection, the proposed autonomous algorithm was utilised to process the 52 radargrams of the third site. Example results associated with three mother wavelets, i.e. db20,

Sym20, and dmey, using the decomposition levels of $l = 5$ and $l = 6$ for the same radargrams in Figure 4-24 are presented in Figures 4-25 to 4-27. The high performance of the algorithm can be confirmed through NIS values for all radargrams (i.e., $\times 10^{-77}$and $\times 10^{-78}$), which are in full agreement with the theory of the KF (necessity for zero mean innovations).

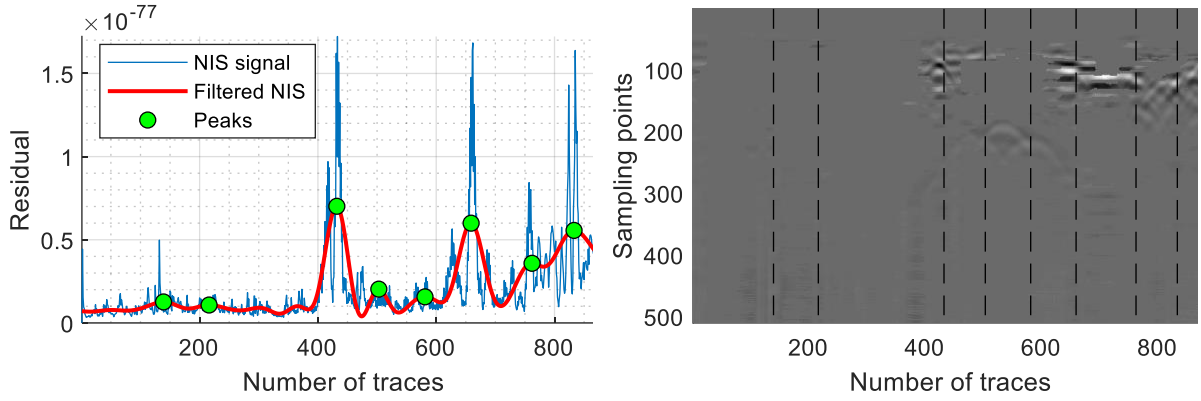
Apart from successful noise and clutter elimination, a comparison of the processed radargrams in Figure 4-25 shows that in terms of target detection the Sym20 mother wavelet performed significantly better than the db20 and the dmey mother wavelets for processing this radargram. Again, this confirms the problem dependency of the algorithm on the selection of suitable mother wavelets. In addition, Figure 4-25 shows that the decomposition level $of\ l = 5$ was a suitable level for processing the radargrams in Figures 4-24(c) and (d), whereas the decomposition level of $l = 6$ provided higher performance for processing the radargrams on in Figures 4-24(a), (b), (e), and (f), as shown in Figure 4-26 and Figure 4-27, respectively.

A comparison of results in both figures, on the other hand, shows that all three mother wavelets performed well compared to the actual 'truth' on the site with almost identical results in terms of the target and ground anomaly detection. It should be noted that, as discussed earlier, the performance of the wavelets utilised in signal analysis can vary depending on the properties of the wavelet in the frequency domain and the characteristics of the signal being analysed. Wavelets that have better time localisation properties are often more effective at representing signals that have high energy at specific time intervals, whereas wavelets that have better frequency localisation properties are often more effective at representing signals that have high energy at specific frequency intervals. Additionally, the smoothness of a wavelet's scaling function can also affect its performance in the frequency domain. This can be seen as a logical explanation for the variations in the performance of

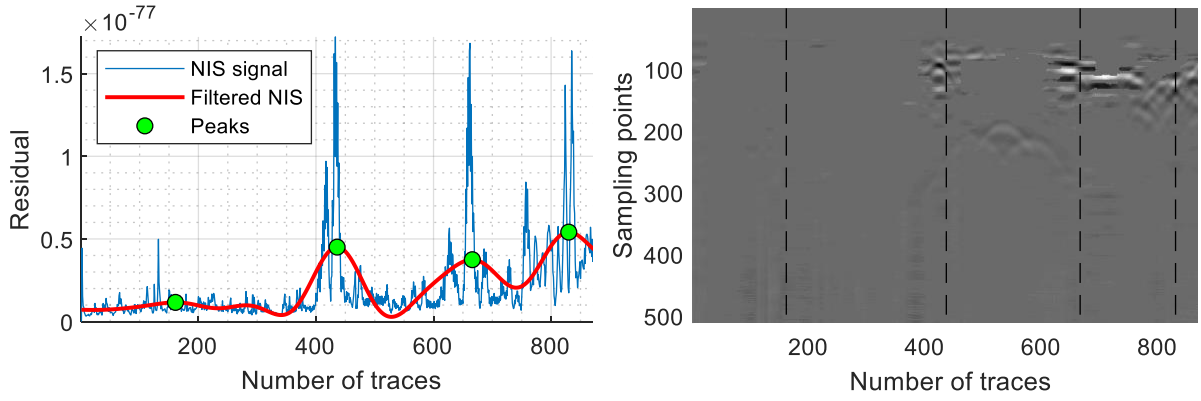
utilised wavelets i.e. db, Symlet, and dmey when analysing the generated NIS signal functions associated with different GPR datasets analysed in this thesis.




*(a) Wavelet: db20*

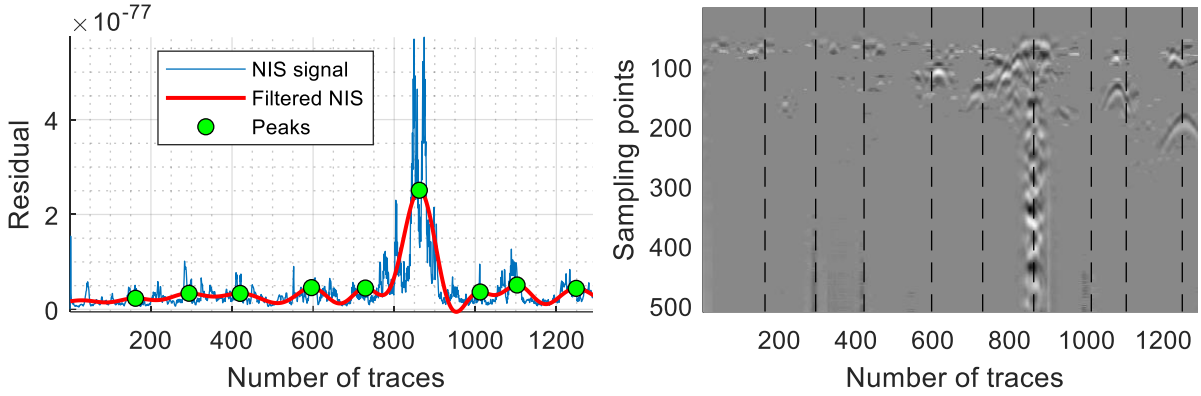



*(b) Wavelet: Sym20*

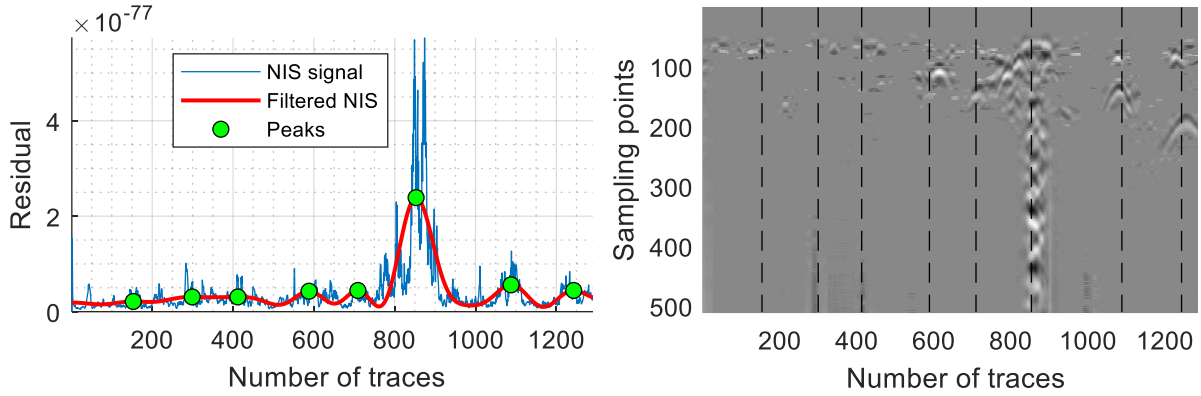



*(c) Wavelet: dmey*

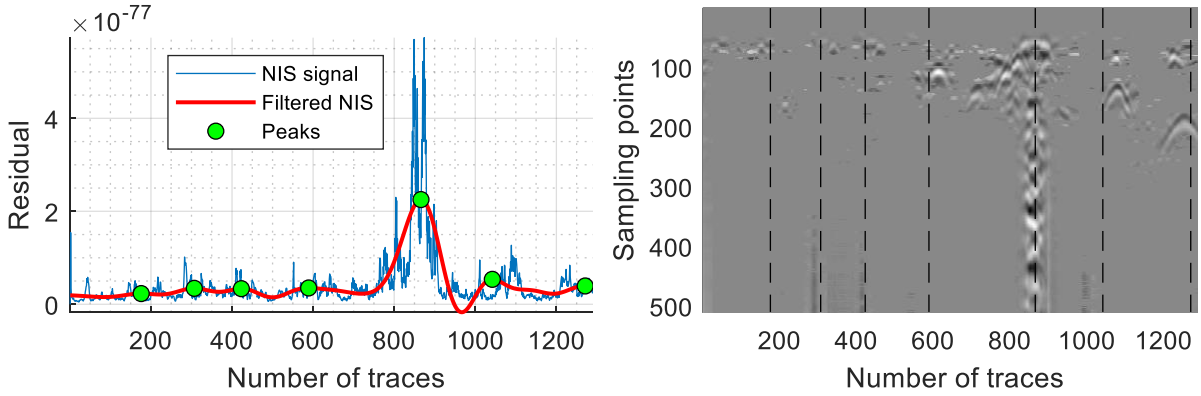
*Figure 4-25. Processed NIS signals (left) and associated identified anomalies (right) for denoised radargram 19 (case study three) using the proposed autonomous algorithm with three different mother wavelets under the decomposition level (l) = 5, a) db20 wavelet, b) Sym20 wavelet, c) dmey wavelet*




*(a) Wavelet: db20*

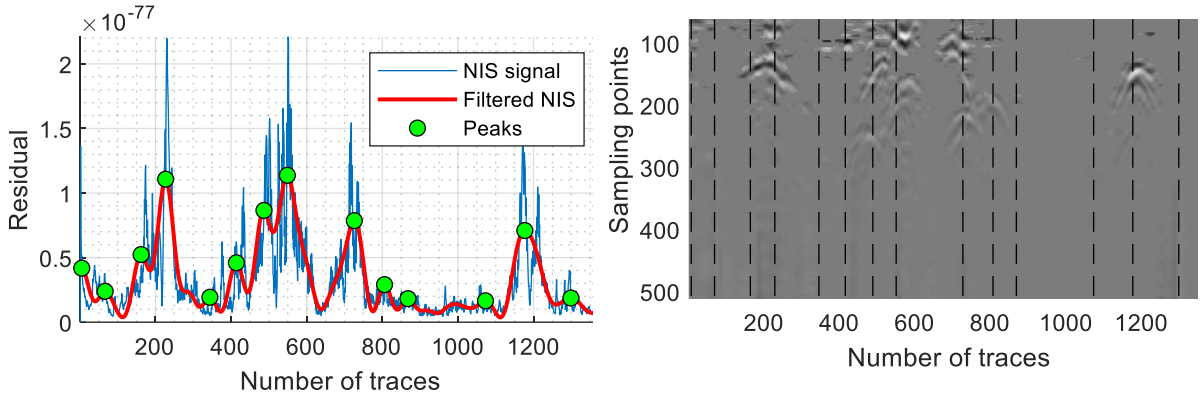



*(b) Wavelet: Sym20*

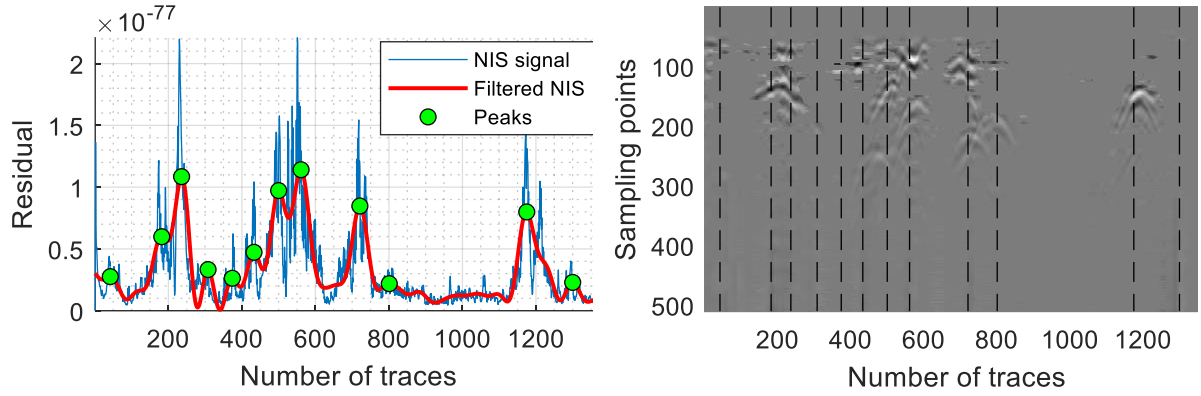



*(c) Wavelet: dmey*

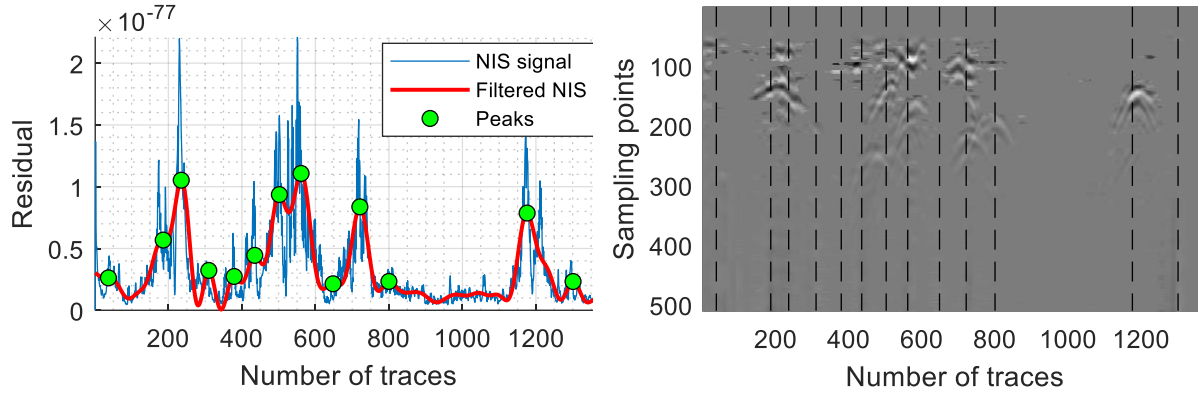
*Figure 4-26. Processed NIS signals (left) and associated identified anomalies (right) for denoised radargram 39 (case study three) using the proposed autonomous algorithm with three different mother wavelets under the decomposition level (l) = 6, a) db20 wavelet, b) Sym20 wavelet, c) dmey wavelet*




*(a) Wavelet: db20*




*(b) Wavelet: Sym20*




*(c) Wavelet: dmey*

*Figure 4-27. Processed NIS signals (left) and associated identified anomalies (right) for denoised radargram 60 (case study three) using the proposed autonomous algorithm with three different mother wavelets under the decomposition level (l) = 5, a) db20 wavelet, b) Sym20 wavelet, c) dmey wavelet*

The functionality and the performance of both novel semi-autonomous and autonomous algorithms were validated using a limited number but different GPR datasets from different surveys.

In addition, the performance of these algorithms was evaluated against other methodologies, such as the method provided by Carevic (1999), i.e., the chi-squared hypothesis testing method, and a conventional method in use by the industry, namely the REFLEX software, which will be discussed below.

## 4.6 Evaluation of the proposed semi-and-autonomous algorithms

To evaluate the performance of the novel semi-autonomous and autonomous algorithms proposed in this research for processing GPR data (Objective 8), the results obtained from these two algorithms were compared to two different methods of GPR interpretation: i) a proposed algorithm in the literature (Carevic, 1999) that uses the chi-squared hypothesis testing method as a benchmark for identifying targets, and ii) a conventional method used in the industry based on the REFLEX software which was conducted manually by an expert.

### 4.6.1 Evaluation using the chi-squared hypothesis testing method

The zero-mean property of KF's performance benchmark parameters (i.e. the innovation and NIS) was utilised in this research for determining the optimal values for **Q** and **R** covariances. As stated by Grewal and Andrews (2008), assuming that sources of errors have a Gaussian distribution, then the NIS parameter will have a chi-square distribution with **m** degree of freedom, where **m** is the length of the innovation vectors. This property (i.e. the chi-square

distribution of the NIS parameter) was used by Carevic 1999 within their proposed algorithms (discussed in the literature review chapter) to detect buried targets. In other words, there are five parameters, i.e. the threshold, null hypothesis rejection parameters ($K_0$ and $K_1$), width of the targets, and the number of layers that need to be specified by the user in advance. Due to variations in the quality and specifications of different GPR datasets, such parameters are not unique and might vary for different datasets. On the other hand, the performance of Carevic's algorithm is highly dependent on the correct choice of these parameters. Figure 4-28 demonstrates a comparison between results obtained from the algorithm proposed by Carevic, (1999), Figures 4-28(a), (b), and (c), with the result obtained from the proposed algorithm in this research for which no predefined parameter values were needed, Figure 4-28d.

The dependency of Carevic's algorithm on the user's knowledge and expertise for selecting the required parameters, e.g. size of the window, and the consequent effect on the results can be observed in Figures 4-28(a), (b), (c). As discussed in section 2.5.2, this approach has considerable drawbacks and limitations. One of the shortcomings of this algorithm is that the width of the targets cannot be identified. Defining the width of the potential targets (by the user) can change the performance of the algorithm significantly. This can lead to problems such as missing targets, see Figure 4-28a, or a high level of false positives, see Figure 4-28c. In this study, the need to determine such parameters through the user was eliminated by replacing the chi-squared hypothesis testing with processing the NIS signal and employing the mean of NIS values as a benchmark for detecting the targets and ground anomalies.

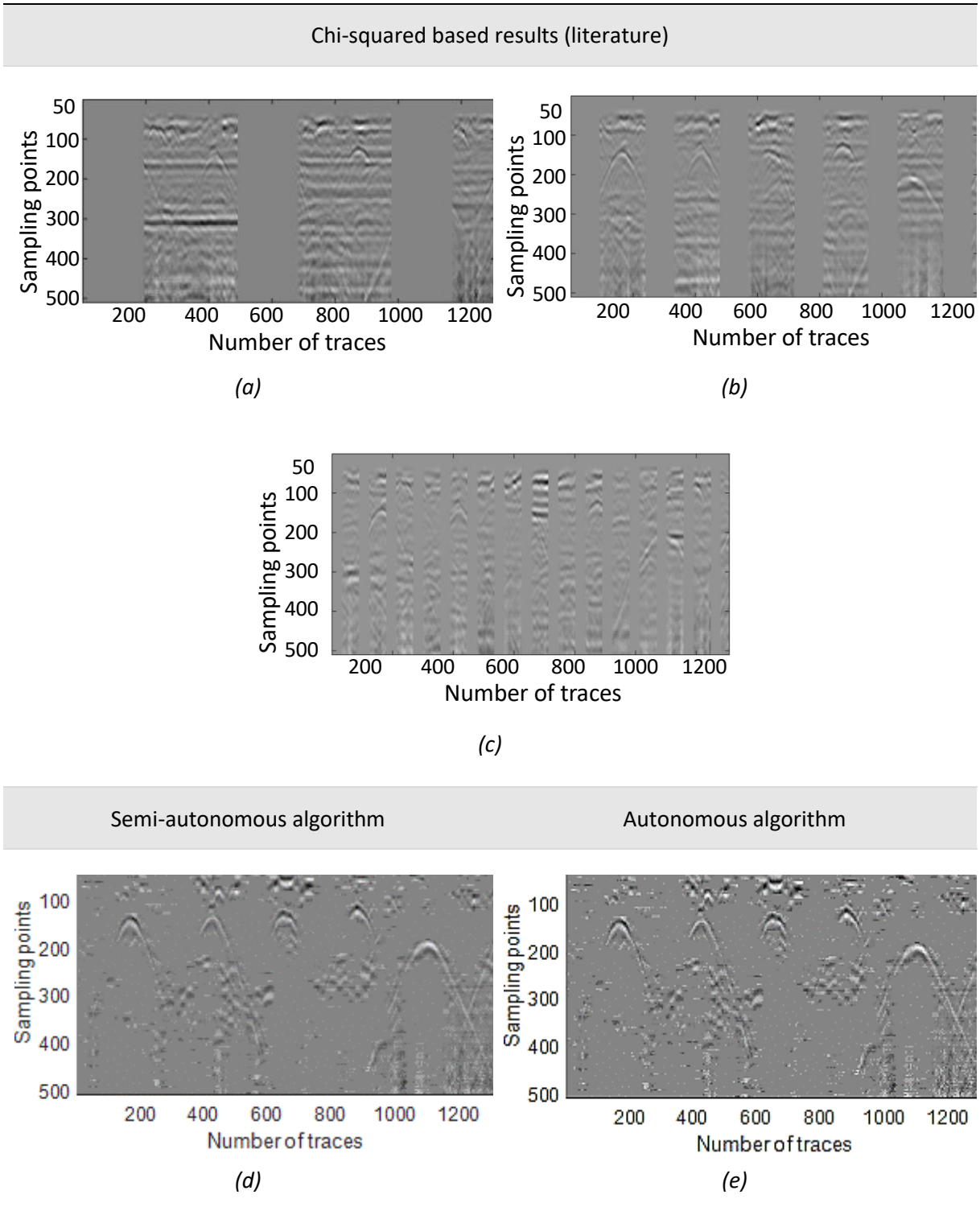
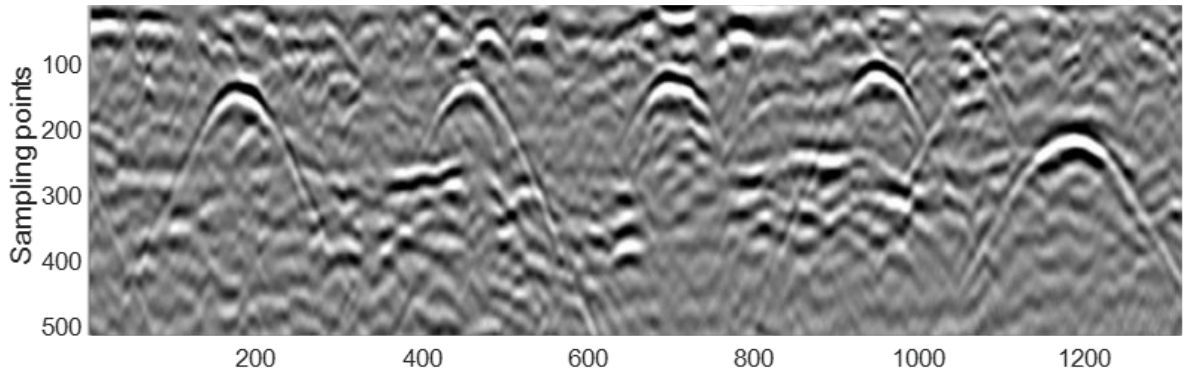



*Figure 4-28. Comparison of the Carevic (1999) algorithm based on chi-squared hypothesis testing with the proposed algorithm in this research; a) user-defined window =300, b) user-defined window =150, c) user-defined window =75, d) result from semi-autonomous algorithm, e) result from autonomous algorithm.*

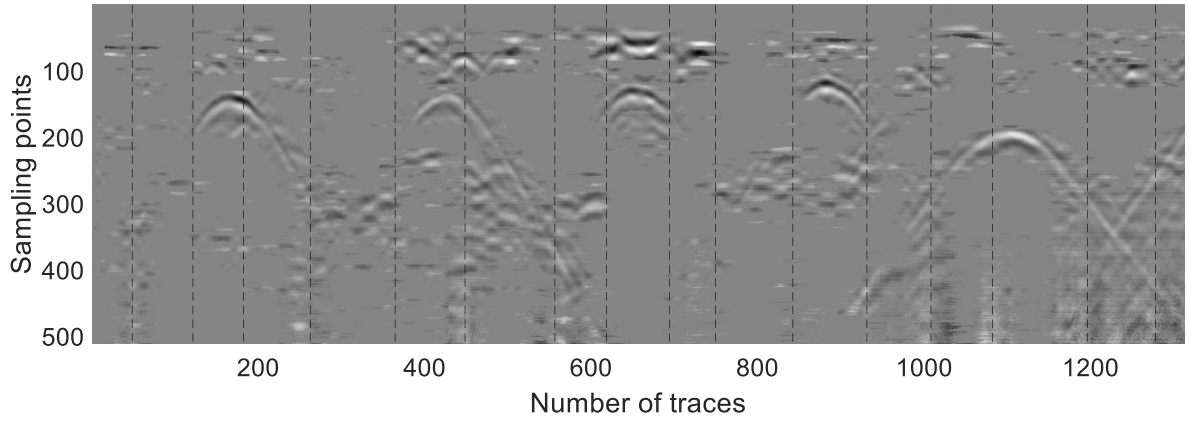
## 4.6.2 Evaluation using REFLEX software

It is known that pre-and post-processing of GPR data using conventional methods, e.g. using the REFLEX software, is a complex and time-consuming and, hence, costly task that is typically conducted by an expert manually (Ernenwein and Kvamme, 2008). The performance of the proposed algorithms in this thesis was evaluated in an additional step by utilising the REFLEX software as a conventional method used in the industry. Figure 4-29 shows the output obtained from REFLEX, which was pre- and post-processed by an expert from the RSK Geophysics[17] team for the FINDIT project[18]. Figure 4-29a shows the result from REFLEX with all the necessary filters applied to the radargram manually and in a step-by-step procedure. In this figure, the target associated hyperbolas are visible, however the background signals consisting of noise and clutter remained. Figure 4-29b, on the other hand, shows the result of the autonomous algorithm where the interaction of the expert is limited to only a visual inspection of the final output. It can be observed that, in addition to the removal of background signals, the position of both potential targets and ground anomalies with high impedance are indicated by the algorithm with vertical dashed lines.

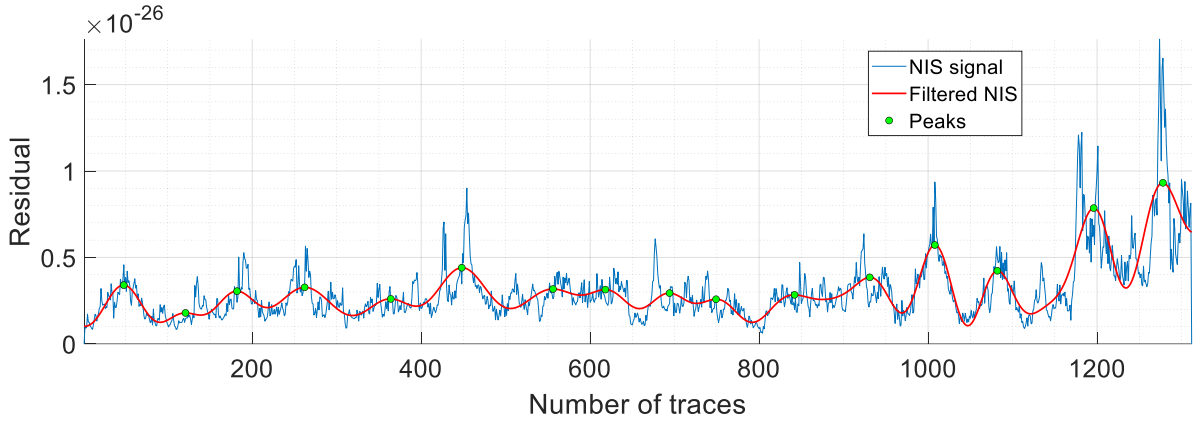
[17] https://www.environmental-geophysics.co.uk/
[18] https://www.findit-research.org/




*(a)*




*(b)*




*(c)*

*Figure 4-29. Comparison of the results between (a) the proposed autonomous algorithm in this research and (b) the REFLEX software for a sample case study, and (c) the associated NIS signal.*

Considering the time-consuming manual steps that an expert requires to undertake for processing GPR radargrams, the proposed novel semi-autonomous and autonomous

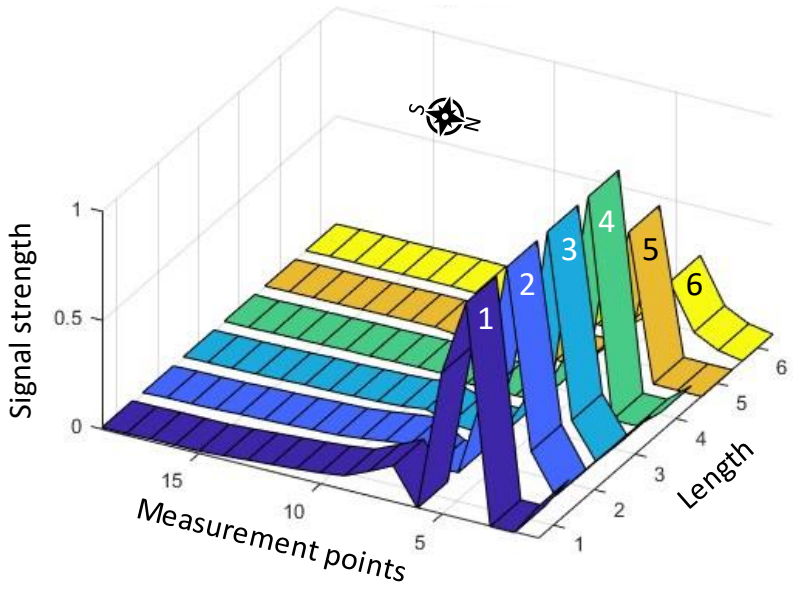
algorithms in this study offer a computationally cost-efficient and accurate solution for GPR data interpretation which minimise required expert intervention.

## 4.7 Fusion of data

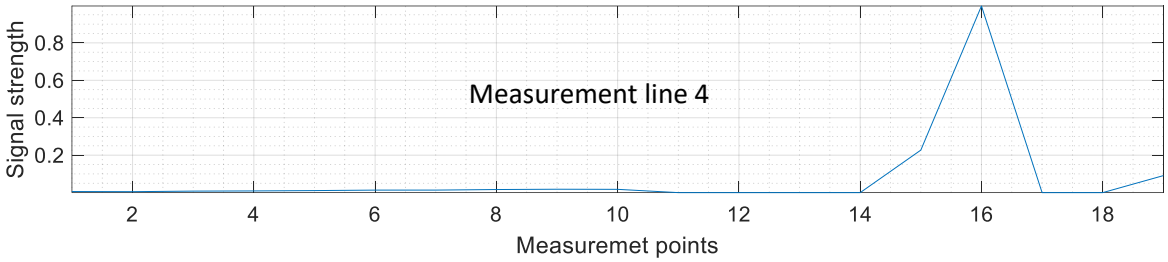
This section of the chapter is dedicated to presenting and discussing the results of the proposed fusion algorithms for both experimentally and numerically obtained data. First, results of fusion of both experimentally obtained GPR and EML data and experimentally obtained GPR and EMC data techniques are presented. Second, the results of a numerical inversion of EML data fused with complementary information provided by GPR data are presented and discussed.

### 4.7.1 Fusion of experimentally obtained GPR and EML data

As discussed in section 3.6.2 of the methodology chapter, Figure 4-30a shows the visualised MF measurements around a metallic pipe buried in sand (30 cm depth). The data was collected using a Radiodetection RD8200G EML device in active mode by direct connection of the Genny device to the pipe. Figure 4-30b, on the other hand, is the recorded signal strength correlated with the measurement line 4 of the Figure 4-30a. It can be observed that all 6 measurement lines indicate a signal peak (maximum MF strength) where the EML device measures the MFs above the metallic pipe. Under ideal conditions, i.e. low environmental noise, the device is able to calculate the burial dept of the metallic pipe using the measured values of its top and bottom coils. Due to the ideal governing laboratory conditions for this experiment, the calculated depth could be employed as a known parameter for the fusion of corresponding data. It is worth mentioning that there are other methods for calculating the depth for more complex scenarios, which are out of the scope of this research.




*(a)*




*(b)*

*Figure 4-30. a) Cable avoidance tool measurements (6 measurement lines), b) the measured signal strength associated with the measurement line 4*

Two radargrams from the experiment dataset were selected to outline two slightly different scenarios of the experiment. Figure 4-31(a) and (b) outlines a scenario in which the void (i.e. a buried gym ball next to the 2 m pipe) is detectable along with the other three buried pipes. Whereas the scenario presented in Figure 4-32(a) and (b) only the three buried pipes are detectable and not the void.

The outcome of the proposed KF-based fusion algorithm are presented in Figures 4-31c and 4-32c. As discussed in section 3.7 of the previous chapter, a 2D Gaussian function was applied to the EML measurements and the outcome was fed into the proposed fusion algorithm to be

processed by the KF alongside with the GPR data. The results of the fusion are shown in Figure 4-31(c) and (d) in which the algorithm was able to accurately detect the metallic pipe along with other buried pipes and the void. This fusion can be considered as a feature level fusion as discussed in Chapter 2. It can also be observed that the background signal is not completely removed by the KF, due to the low quality of the recorded data using the Stream-C GPR device. In other words, the large number of vertical bandings caused by the antenna affects the accuracy of the KF's output. These vertical horizontal bandings were removed to a large extent by applying a vertical moving average filter on all the radargrams. However, the effect of such vertical bandings could not be completely eliminated by the KF.

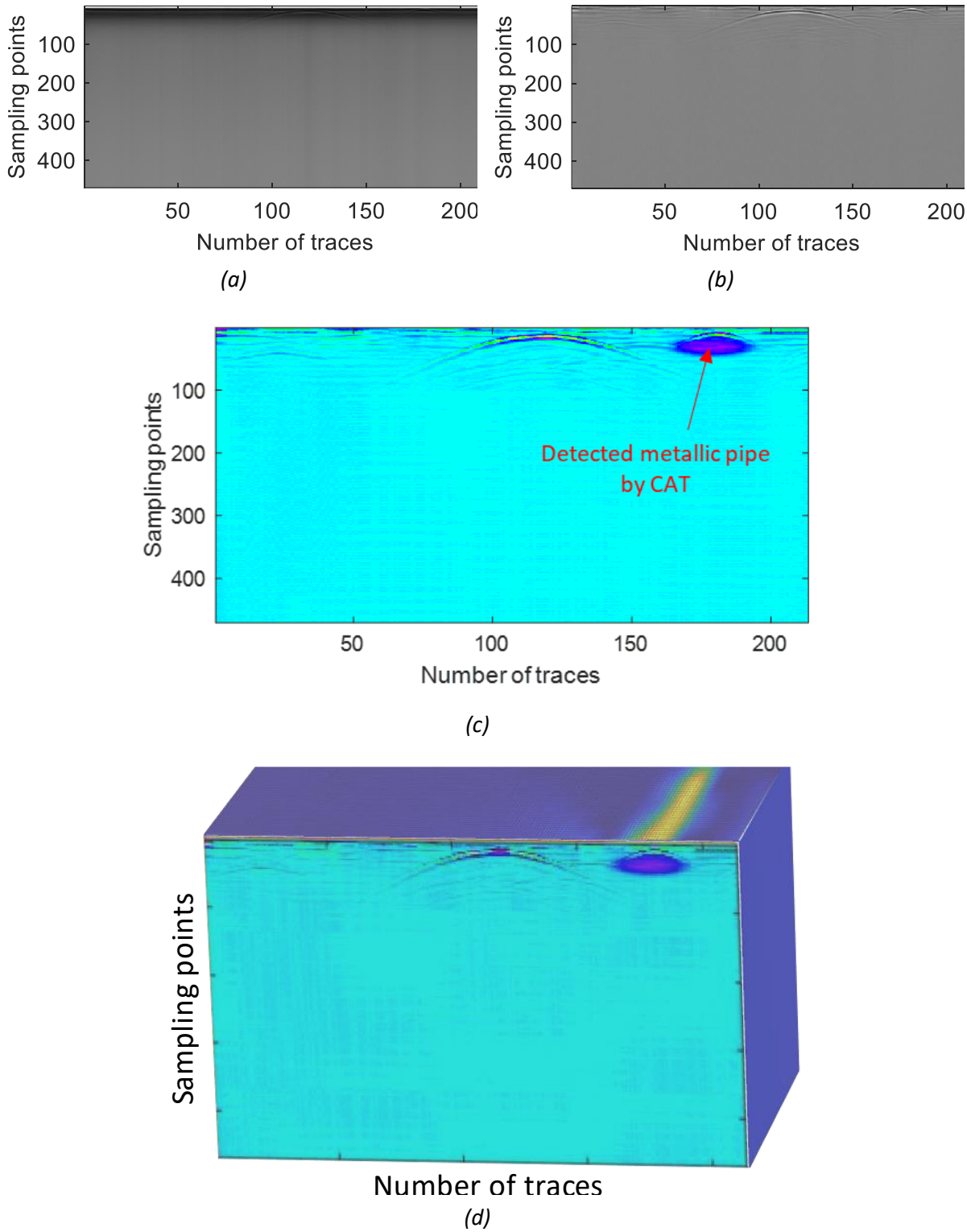



*Figure 4-31. a) unprocessed radargram 4 from NBIF experiment consisting of three buried pipes and a void, b) the radargram 4 with applied zero correction, moving average filters (gain adjustment was not required due to the quality of data), c) result of the fusion of EML measurements, d) 3D view of the fusion of GPR and EML data*

Figure 4-32c presents the result of the fusion of GPR and EML data for the second radargram, i.e. radargram 5. Similar to the result in Figure 4-31c, the metallic pipe is accurately detected by the algorithm, while the other three buried pipes were not identified and remained as potential anomalies in the ground. The hyperbola associated with the buried plastic water pipe is not a full hyperbola signature, which complies with the corresponding hyperbola signature in the unprocessed radargram. This is due to the bend in the water pipe at this location.

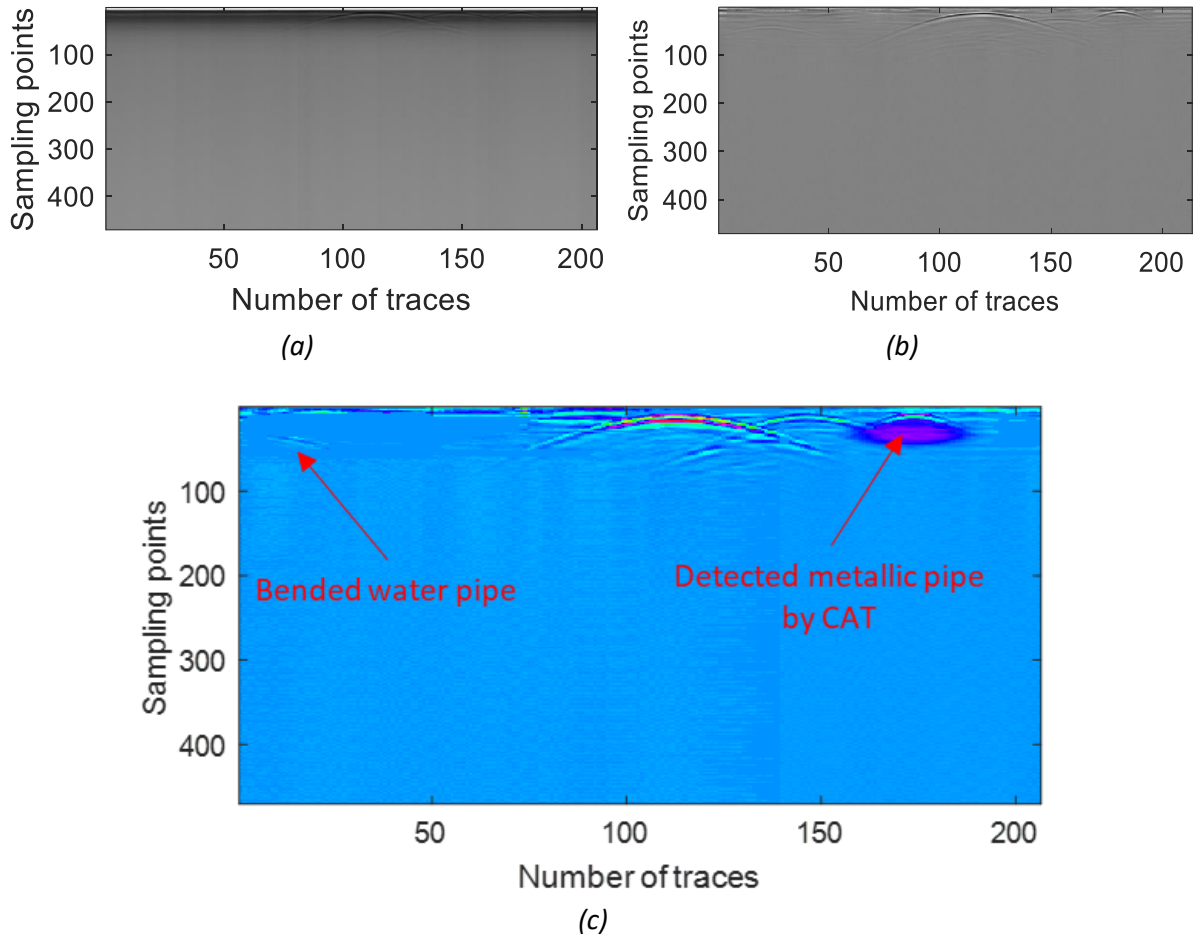



*Figure 4-32. a) unprocessed radargram 5 from NBIF experiment consisting of three buried pipes, b) the radargram 5 with applied zero correction, moving average filters (gain adjustment was not required due to the quality of data), c) result of the fusion of EML measurements in Figure 4-30 (measurement line 5) and the radargram 5.*

## 4.7.2 Fusion of experimentally obtained GPR and EMC data

In this section, the results from the proposed fusion algorithm in section 3.7.2 for the fusion of the GPR and EMC data are presented and discussed. The presented results are associated with two sample GPR radargrams, i.e. Figures 4-33(a) and 4-34(a), and their correlated EMC measurements, i.e. Figures 4-33(b) and 4-34(b), from the datasets. The results of the fusion process are shown in Figures 4-33(c) and 4-34(c), respectively. The corresponding subsurface conductivity cross-sections are inverted by the EMagPy software (section 3.7.2.1). There are eight targets associated hyperbola signatures identifiable in radargram 19 in Figure 4-33(a). However, the type of these potential targets is unknown. On the other hand, the generated conductivity map of the correlated data (Figure 4-33b) indicates three conductive regions in the subsurface which are marked using red rectangles in Figure 4-33b.

Both data were fed into the proposed KF-based algorithm and processed based on the pre-defined error covariances matrices (**R**) discussed in section 3.7.2 of this thesis. It was mentioned that due to higher level of confidence in GPR technique a higher value of error covariance were designated to the EMC data. Figure 4-33c shows the result of the data fusion in which all eight hyperbolas are remaining and identified by the algorithm as potential buried targets (e.g. pipes or cables). At the same time, it can be observed that the conductive regions from the EMC map are also kept by the KF as areas with high conductive values which can be related to the type of buried targets in that area.

The radargram 10 in Figure 4-34a indicates seven targets associated hyperbola signatures and one of which is a known metallic inspection chamber around 2100$^{th}$ trace. On the other hand, the conductive regions on the reconstructed EMC map are marked using red rectangles on Figure 4-34b. Due to a dominant high conductivity measured from the metallic inspection

chamber, the variation of the conductivity in the marked regions (2 and 3) is almost invisible. In other words, these conductive areas are damped. This issue can be clearly observed on the non-inverted conductivity measurements shown in Figure 4-34c. Although there are variations in measured conductivity values associated with the marked regions (2 and 3), the high conductivity values (ECa) associated with the inspection chamber (e.g. ECa with 600 mS/m) are dominating the rest of the measurement with significantly lower values. On this basis, the inverted EMC data as an input data for the fusion is an inaccurate data which causes inaccuracy in the fusion process and thus in result shown in Figure 4-34d. It was expected that the conductive regions 2 and 3 are identifiable on the result in Figure 4-34d. However, due to the discussed dominance of the conductivity of the inspection chamber, these regions remained invisible.

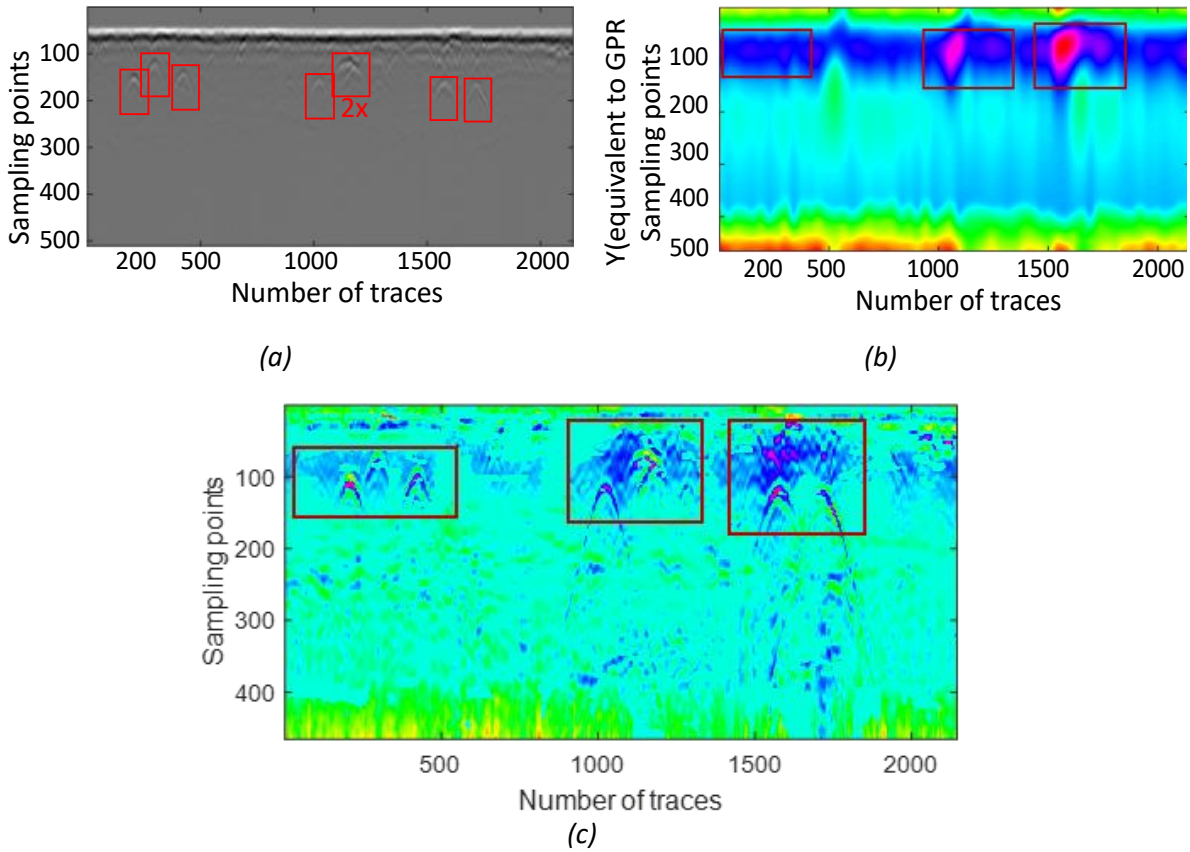
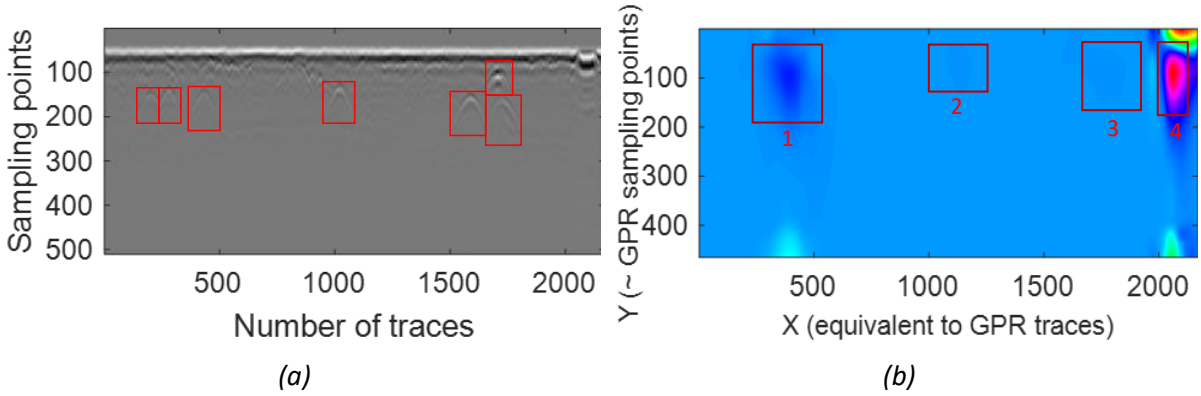
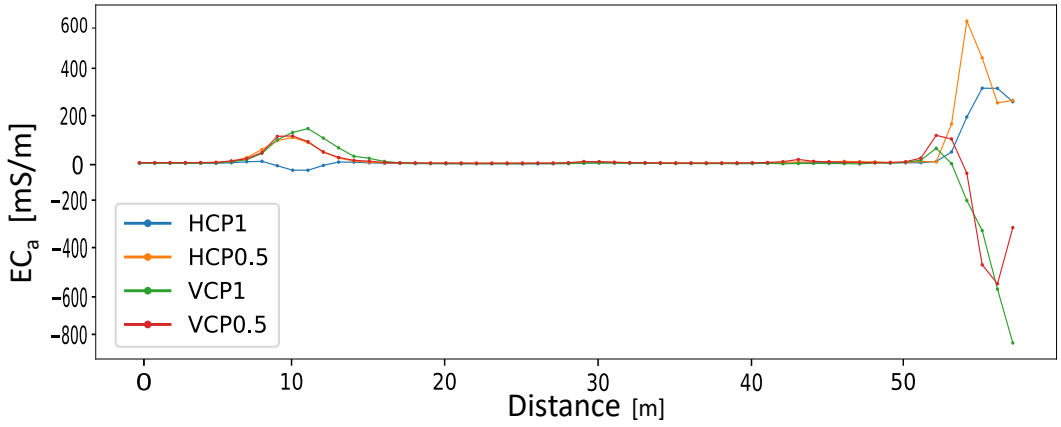


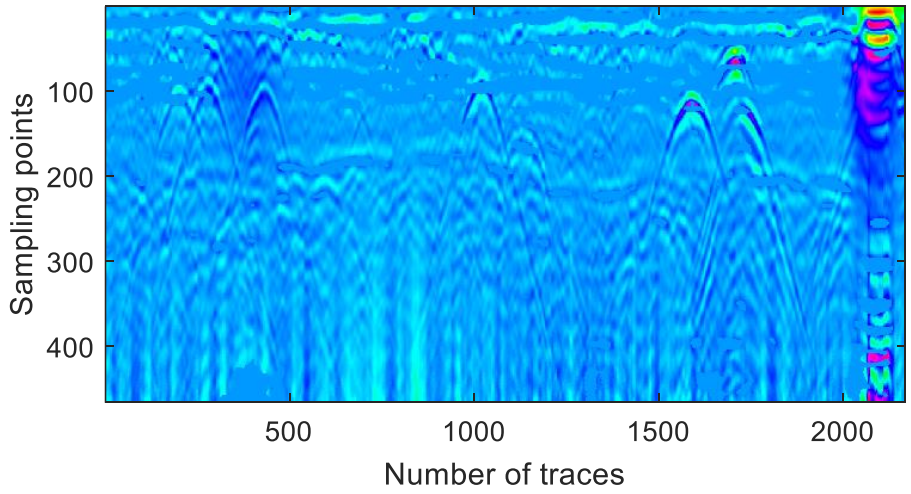

*Figure 4-33. Fusion result. a) unprocessed radargram 19 from the trial site, b) EMC map inverted by the EMagPy software showing two conductive areas around 1100th and 1600th points of the X-axis, c) fused GPR and EM data*




*(a)* *(b)*




*(c)*




*(d)*

*Figure 4-34. Fusion result. a) unprocessed radargram 10 from the trial site, b) EMC map inverted by the EMagPy software showing two conductive areas around 400$^{th}$ and 2100$^{th}$ points of the X-axis, c) non-inverted EMC measurement associated with the radargram 10, d) fused GPR and EM*

### 4.7.3 Inversion of fused GPR and EML using FE forward model and GA

The developed and validated FE forward model in section 3.8.5 of the methodology chapter was used in correlation with the GA for the inversion of the unknown parameters of a ground model, i.e. location of buried utilities in the ground. The inversion was enhanced by fusing the recorded MFs using EML (e.g. CAT) and the GPR as complementary information. This is aligned with objective 7 of this thesis. The position of the buried utilities (e.g. live cables) and the intensity of the following current (amperage) in these cables are the unknown variables of this scenario that are estimated by the GA. The complementary information provided by the GPR radargram (i.e. the spatial position of detectable utility by the GPR) reduces the number of unknown variables by constraining them. In other words, by estimating the spatial position of detectable cable by GPR (i.e. two of the unknown parameters), the number of unknown parameters were reduced from six to four (i.e. the spatial position of undetected cable by the GPR and amperage of both cables).

To demonstrate the performance of the proposed algorithm, the FE forward model presented in section 3.8.1.3 (see Figure 3-57) was used to produce four different numerical scenarios. Each of the scenarios consists of a high and a low voltage live cable (different diameters) carrying 15 and 5 amper currents, respectively. The cables were positioned at various locations which caused different coupling effects of MFs. Figure 4-35a demonstrates two adjacent cables (short separation) buried at the same depth level, i.e. scenario 1. Figure 4-35b demonstrates the same cables but buried at different depth levels, i.e. scenario 2. Figure 4-35c demonstrates the same cables but with a larger horizontal compared to scenario 1, and hence no (or minimal) MFs coupling effect, i.e. scenario 3. Finally, Figure 4-35d demonstrates a scenario with similar cables with almost zero separation distance, i.e. scenario 4.

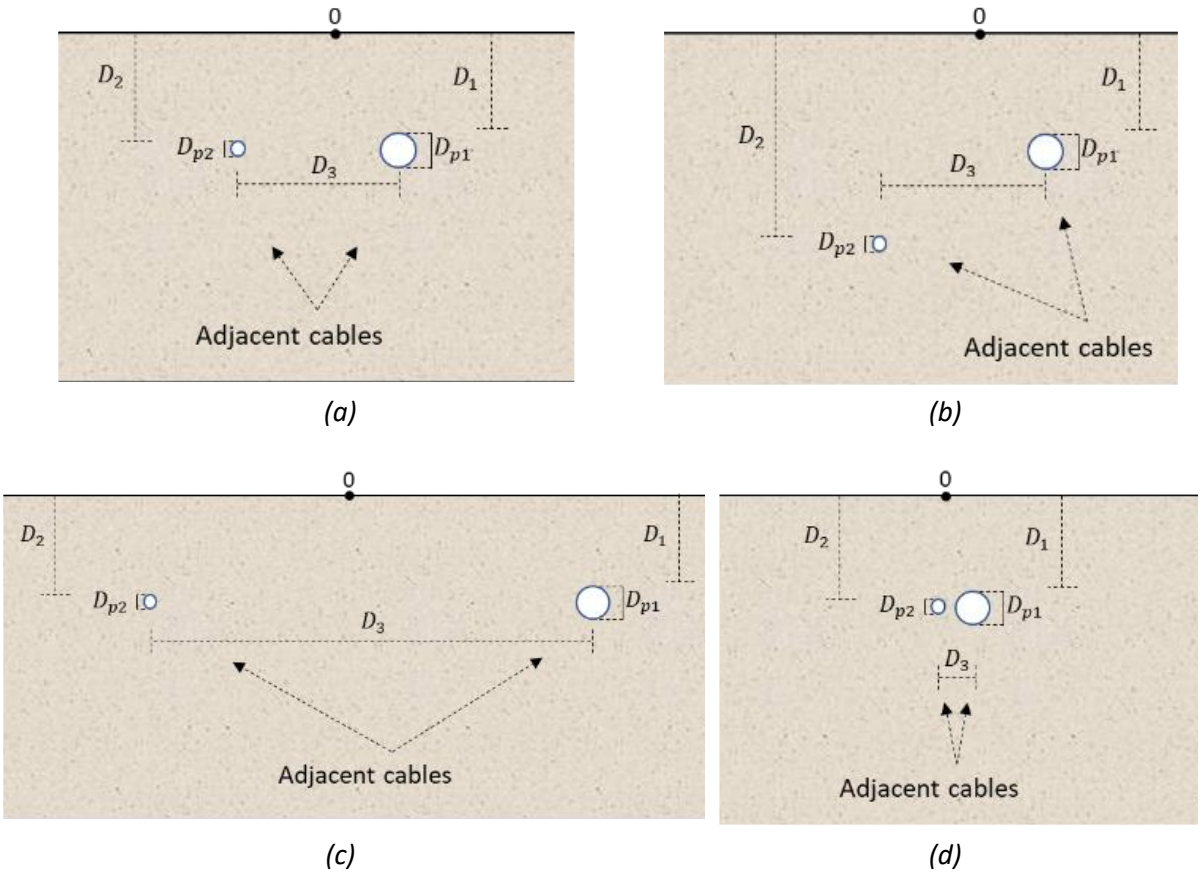



*Figure 4-35. Four different scenarios of two buried cables modelled using the developed FE forward model in section 3.8.1.3 (see Figure 3-57) to demonstrate different types of MF coupling effects, a) 30 cm horizontal separation between the two cables with the same burial depth, b) different burial depts with 30 cm separation, c) large separation (1 m) to minimise coupling effects, d) very short separation (1 cm) to demonstrate the maximum coupling effect.*

The GPR radargrams associated with all four scenarios using the matGPR software and a 400 MHz centre frequency antenna were numerically generated. An example radargram associated with Figure 4-35a is presented in Figure 4-36. It can be observed that the GPR was not able to detect the smaller cable and only the hyperbola signature associated with the larger cable is existing on this radargram (and on other radargrams for other scenarios). The spatial positions of the larger cable for all scenarios are determined and fed into the inversion algorithm as complementary information to constrain two of six existing unknown parameters.

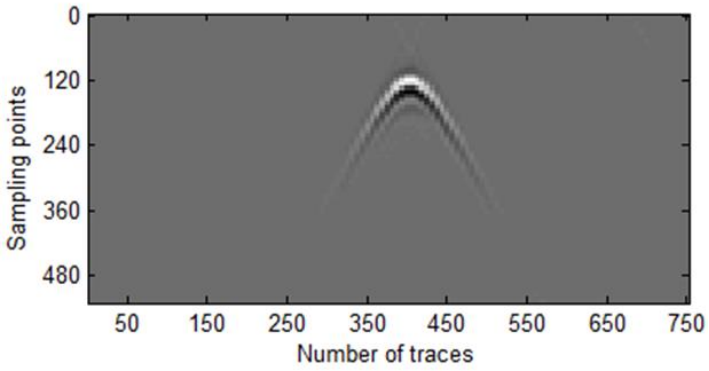



*Figure 4-36. Example of numerically generated GPR radargrams using matGPR software associated with the corresponding forward model in Figure 4-35a*

The minimal available information about the subsurface (i.e. the measured coupled MFs, and estimated position of the larger cable) turns such a scenario into a complex problem to be solved by the GA. The estimated scenarios after the application of the inversion algorithm for all four scenarios are presented in Figure 4-37.

Table 4-1 presents the expected values, estimated values, the discrepancy between the expected and estimated values, and computational times for all four scenarios. It can be seen that the inversion algorithm was able to generate very close estimations to the expected values with a discrepancy range between 0% to 22.26%. It can be observed that the maximum discrepancies are from the scenario where two cables are at different depths. However, it should be noted that while the 22.26% discrepancy was observed in the horizontal location of $C_2$ pipe the discrepancy in its vertical location was 11.80%. In other words, the distance between the estimated and actual location of the $C_2$ pipe is less than 7 cm.

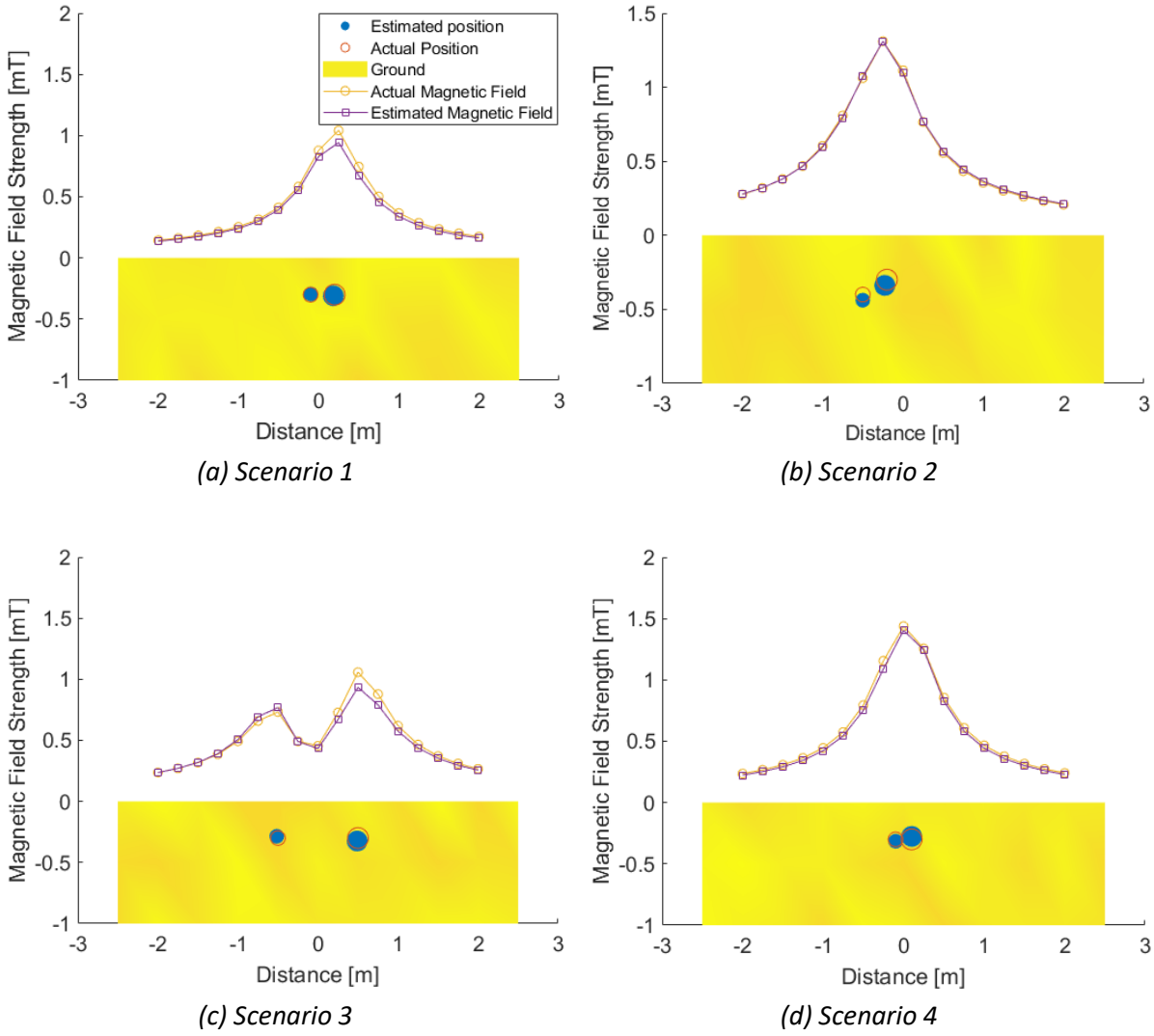



*Figure 4-37. Expected values versus estimated values (i.e. estimated positions and magnetic fields) by the numerical inversion-fusion algorithm associated for four different numerical scenarios*

The presented results indicate the robustness of the proposed inversion algorithm at this stage for the estimation of the position and current of buried live cables in the subsurface. It should be noted that these results are limited to the discussed numerical examples and does not account for environmental noise in the data. So that, further analysis using experimental and field data is required to confirm the efficiency of this algorithm for real-world scenarios.

*Table 4-1. Parametric results associated with FE forward models in Figure 4-35 and Figure 4-37*

| Scenario | | Parameter | Expected value [m] | Estimated value [m] | Discrepancy [%] | Time [s] |
|---|---|---|---|---|---|---|
| 1 | Position | $C_{1x}$ | -0.1 | -0.0967 | 3.35 | 1.417e+04 |
| | | $C_{2X}$ | 0.2 | 0.1898 | 5.11 | |
| | | $C_{1z}$ | -0.3 | -0.2911 | 2.97 | |
| | | $C_{2z}$ | -0.3 | -0.3091 | -3.05 | |
| | Current (Amperage) | $C_{1J}$ | 5 | 4.9295 | 1.41 | |
| | | $C_{2J}$ | 15 | 14.0655 | 6.23 | |
| 2 | Position | $C_{1x}$ | -0.5 | -0.5000 | 0 | 1.369e+04 |
| | | $C_{2X}$ | -0.2 | -0.1764 | 11.80 | |
| | | $C_{1z}$ | -0.4 | -0.3765 | 5.87 | |
| | | $C_{2z}$ | -0.3 | -0.2332 | 22.26 | |
| | Current (Amperage) | $C_{1J}$ | 5 | 4.7171 | 5.65 | |
| | | $C_{2J}$ | 15 | 14.9428 | 0.38 | |
| 3 | Position | $C_{1x}$ | -0.5 | -0.4977 | 0.46 | 1.426e+04 |
| | | $C_{2X}$ | -0.5 | 0.4931 | 1.38 | |
| | | $C_{1z}$ | -0.3 | -0.3022 | -0.74 | |
| | | $C_{2z}$ | -0.3 | -0.2918 | 2.75 | |
| | Current (Amperage) | $C_{1J}$ | 5 | 4.8950 | 2.10 | |
| | | $C_{2J}$ | 15 | 14.8365 | 1.09 | |
| 4 | Position | $C_{1x}$ | -0.1 | -0.0939 | 6.10 | 1.716e+03 |
| | | $C_{2X}$ | 0.1 | 0.1056 | -5.61 | |
| | | $C_{1z}$ | -0.3 | -0.3075 | -2.50 | |
| | | $C_{2z}$ | -0.3 | -0.2964 | 1.19 | |
| | Current (Amperage) | $C_{1J}$ | 5 | 5.1400 | -2.80 | |
| | | $C_{2J}$ | 15 | 14.3220 | 4.52 | |

## 4.8 Summary

In this chapter, the results of four proposed algorithms were systematically presented and investigated to meet objectives 4, 5, 6, 7, and 8 of this thesis. Various experimental and numerical scenarios were considered that outlined different buried utility setups. Depending on the employed algorithm, results were presented as both images and tables consisting of numerical results. To demonstrate the performance of both semi- autonomous and autonomous KF-based algorithms for post-processing of GPR data, the unprocessed radargrams were compared against the processed radargrams by each algorithm to show how it removes the noise and clutter and clarifies the hyperbolas.

The first group of results were from the proposed novel semi-autonomous algorithm for processing the noisy GPR radargrams and detecting and locating shallow buried targets (e.g. pipes) and ground anomalies (e.g. soil disturbance). The generated and processed NIS signals, by the algorithm, associated with each of the sample radargrams were presented as 2D visual images. Moreover, examples of ROC curves that were utilised to evaluate the performance of the algorithm were presented and compared against each other. Real GPR datasets were processed using this novel algorithm. It was discovered that the reduction of the area under the ROC curves is proportional with the rate of false positives in terms of the detected utilities and ground anomalies. For example, for the data set in section 4.2, the lowest optimal area under the ROC curve (> 74%) was selected for the best Pareto solution from which 5 of 5 targets were detected, and the false positive rate was kept to a minimum. In other words, within the optimised Pareto solutions, the lower the optimal area under the ROC curves, the higher the optimal detection rate of the algorithm with no missing targets. Additionally, a broken/blocked pipe section, which was invisible within a normal radargram, was detected by warning lines generated by the algorithm.

The second group of presented results were from the second proposed novel algorithm for the autonomous processing of noisy GPR radargrams and detecting and locating shallow buried targets (e.g. pipes) and ground anomalies (e.g. soil disturbance). The novel autonomous algorithm is based on the development of an optimised discrete Kalman Filter (DKF) by making use of a single objective genetic algorithm (SOGA) and wavelet transform. Three different mother wavelets, (i.e. db, Sym, and dmey), with all their vanishing moments at different wavelet decomposition levels ($l = 3$ to $l = 7$) were utilised for the decomposition and analysis of the NIS signal for the detection of utilities and ground anomalies. The processed radargrams and the associated NIS signals were compared against the unprocessed version of the presented radargrams to show how it removes the noise and clutter and clarifies the hyperbolas. The performance of the above-mentioned mother wavelets was compared through a statistical sensitivity analysis using the ROC curves. Processing various radargrams has shown that, the level of true and false positive target detections is proportional to the level of wavelet decomposition and chosen mother wavelets. The optimal detection rate for all three mother wavelets was obtained at signal decomposition levels of 5 and 6 under which all buried utilities and targeted ground anomalies, were successfully identified with a minimum number of false positives. The proposed novel autonomous algorithm showed a significant improvement in processing the GPR data compared to the proposed novel semi-autonomous algorithm in this thesis. Noise and clutter cancelation was improved and locating of both buried utilities and the ground anomalies were performed more with a higher accuracy. In addition, the computational time was significantly reduced, e.g. reduced from 06:31 hours to only 27 minutes for one dataset.

Within the third group of presented results, the effectiveness of both semi- and autonomous algorithms was verified and validated using three different additional GPR datasets for which

no initial information was available for one of these regarding the number, position, type of buried targets and the ground conditions. Additionally, to evaluate the performance of both novel semi- autonomous and autonomous algorithms, their results were compared against the result from two methods, namely i) proposed algorithm in the literature (Carevic,1999) that utilises the chi-squared hypothesis testing method as a benchmark for identifying targets, and ii) a convectional software widely used in the industry (i.e. REFLEX) for processing GPR radargrams. Both proposed algorithms outperformed those methods by proving accurate outputs and minimising the interaction of the expert for either determining the chi-squared hypothesis testing parameters by the user or conducting complex time-consuming processing steps using REFLEX on every single GPR radargram in a dataset.

The fourth group of the presented results were associated with the novel fusion of the experimentally obtained GPR, EMC, and EML data. These results were separated into two sections, i.e. the fusion of GPR and EML data, and the fusion of GPR and EMC data. The results of the fusion of GPR and EML data showed that all the buried targets including the metallic pipe and the buried void could successfully be identified with a high accuracy. In other words, the type of the pipe i.e. metallic pipe, which is challenging to define by the GPR techniques alone, could successfully be defined. It is worth mentioning that the presented results were based on a scenario with controlled laboratory conditions. The same approach as the above was utilised for the fusion of GPR and EMC data, except from applying a 2D Gaussian function to the EML measurements. Similar results to the GPR-EML fusion were provided by the proposed algorithm. The results showed that the performance of GPR-EMC fusion is highly dependent on the quality of EMC data inversion. This inversion was conducted using EMagPy software that utilises the most common inversion methodologies in the literature. On the

other hand, the analysis of the EMC data in this research has shown that extreme variations in conductivity values can lead to inaccuracy in inverted EMC maps of the subsurface.

Finally, the results of a novel inversion-fusion framework were presented as the fourth group of results in this chapter. The proposed algorithm was developed based on a FE forward model, a genetic algorithm, and data fusion of recorded MF (using EML) and GPR for the inversion of the unknown parameters of a ground-buried utility model, i.e. location of buried cables. Both 2D visual images and a table of numerical results were presented for different numerical scenarios. These scenarios outlined coupled MFs with different levels of complexity to assess the performance of the novel inversion-fusion framework for locating buried cables. All the unknown parameters of the problem, i.e. the position and current of both buried live cables in the subsurface were estimated with acceptable discrepancies. It should be noted that these results are limited to the discussed numerical examples and does not account for environmental noise in the data.

# Chapter 5 CONCLUSIONS AND RECOMMENDATIONS FOR FUTURE WORK

In this research, a critically reviewed comprehensive literature on current geophysical non-destructive techniques for detection, locating, and condition assessment of buried utilities and their surrounding ground revealed specific gaps in knowledge and areas in this field that require improvement. These gaps, summarised below, were discussed in detail in Chapter 2 (Objective 1).

- The lack of a robust automated and optimised Kalman Filter algorithm for identifying the presence and width of underground objects.
- Limited data and fusion algorithm for EML.
- The lack of a robust autonomous data fusion algorithm based on Kalman filter and evolutionary algorithm.

Various methods for processing and denoising of GPR data were investigated in Chapter 3, among which the Kalman Filter (KF) was selected as one of the well-established techniques for both estimation and fusion problems with a strong history in practical applications such as those associated with aerospace systems, radars, navigation systems, autonomous driving cars (Objective 2). The required GPR and EMC data was acquired from third party collaborators from industry and published research. Furthermore, a series of EML and GPR test were carried out at NBIF, and numerical models were developed to produce synthetic data (Objective 3).

To tackle the identified gaps in knowledge and limitations of the existing methods, the KF, evolutionary algorithms (i.e. GA), and signal processing techniques (i.e. Fourier transform, and wavelet transform) were used in this research to develop novel algorithms for i) improving the current GPR data processing using the Kalman Filtering (KF) (Objective 4) and ii) developing a

novel computational solution for automation of the GPR data processing. A semi-autonomous and an autonomous algorithms were developed in this research (Objective 5). Then, the developed algorithms were employed for the detection and location of buried targets (e.g. pipes) and ground anomalies (e.g. soil disturbances) in the shallow subsurface. In addition, the autonomous algorithm was used to develop another algorithm for data fusion of multiple geophysical sensors (i.e. GPR, EML, and EMC) to improve the quality of survey interpretations (Objective 6). The proposed algorithms above were validated using different GPR datasets as per mentioned in Objective 3. Then, the efficiency of these algorithms were evaluated against the conventional software in industry (Objective 8). Furthermore, a numerical model of EML test for multiple buried live cables was developed and coupled with GPR data using a genetic algorithm to develop a novel multi-sensing fusion inversion algorithm (Objective 7). This algorithm was validated against experimental test results as well as published data (Objective 8).

In the following sections, the key findings, achievement of the objectives and the aim of the project, conclusions regarding the proposed algorithms and their applications are systematically presented.

## 5.1 The semi-autonomous algorithm

The following specific conclusions can be drawn from the results of this research on the semi-autonomous algorithm:

- It was shown that the proposed novel NIS signal function analysis can successfully replace the traditional chi-squared hypothesis testing method for identifying buried

utilities in a GPR radargram, by eliminating the subjectivity of the analysis and providing reliable estimation of the width of the target.

- The proposed NIS signal function analysis identifies target locations in a GPR radargram by sampling data from the same radargram without the need for assuming a null hypothesis threshold and target width by the expert, as in in chi-squared hypothesis testing.

- Using the proposed computational solution in this study, the complexity of GPR data post-processing and consequently the associated computational cost is significantly reduced, compared to existing conventional techniques (e.g. REFLEX software and chi-squared hypothesis testing method), owing to the scalar format of NIS values within the developed novel NIS signal function in this study and the semi-autonomous nature of the proposed algorithm.

- Adapting pre-processing filters into the proposed KF based semi-autonomous algorithm in this study effectively reduced the pre-processing effort and enhanced the quality of the results.

- The multi-objective optimisation of the proposed semi-autonomous algorithm can optimise the target detection rate by determining the best fit for Fourier frequency and the covariances values of noise, i.e. **Q** and **R**.

- This research identified a systematic error in the literature (see technical note in section 2.5.2.1 of the literature review) on initialising the matrix **P** which is significant regarding the reliability of the KF performance and hence the quality of noise reduction process in a radargram. In the current publications, **P** is initialised from zero which indicates an absolute confidence in initial estimates of the system state. However, this

does not comply with real world systems due to existence of unavoidable noise and clutter. Conversely, initialising matrix **P** with a large value is in line with the initial uncertainty in a GPR survey of the subsurface.

- The reduction of the area under the ROC curves is proportional to the rate of false positives in terms of the detected utilities (e.g. buried pipes) and ground anomalies (e.g. ground disturbances). In other words, within the optimised Pareto solutions, the lower the optimal area under the ROC curves, the higher the optimal detection rate of the algorithm with no missing targets.

- Ground disturbance, which is commonly invisible within a normal radargram, can be detected by inferring the peak impedance zones from the proposed NIS signal function using signal processing techniques such as FFT. This approach was proposed and successfully implemented in this research.

## 5.2 The autonomous algorithms

In addition to the above-mentioned conclusions for the semi-autonomous algorithm (that are valid for the autonomous algorithm too), the following specific conclusions can be drawn from the results of the developed autonomous algorithm:

- The performance of the algorithm was significantly enhanced in terms of the computational cost compared to the proposed semi-autonomous algorithm (e.g., from 06:31 hours to only 27 minutes for the same dataset on an Intel(R) Core (TM) i7-3770 CPU @ 3.40GHz and 32 GB of RAM machine).
- The new algorithm reduces the need for expert intervention even further compared with the first approach to a visual inspection of the results (Objective 8).

- The convergence and performance of the Kalman Filter was improved by employing SOGA instead of the current trial-and-error approach to produce optimal values of KF input variables, i.e. system, sensor covariances.
- The proposed autonomous algorithm was shown to be capable of separating buried anomalies with different impedance, i.e. different levels of true and false positive target detections, using different levels of wavelet decomposition, mother wavelet types, and mother wavelet vanishing moments. This offers better object detection compared with the semi-autonomous algorithm. Additionally, the algorithm offers better agility when working with different site conditions (e.g., ground water content, soil types, anomaly size and type).

## 5.3 The fusion of multiple geophysical data

In addition to the above, to achieve objective 6, a novel KF-based algorithm was developed for the fusion of data from multiple geophysical sensors i.e. GPR-EML, and GPR-EMC.

### Fusion of GPR-EML

The key findings are presented below.

- It was demonstrated that the proposed KF-based autonomous algorithm is an effective tool to fuse GPR and EML data for buried utility detection.
- The novel fusion of GPR and EML data carried out in this research showed that the recorded EML and GPR data include complementary information which improves buried utility detection.
- Both depth and material of the buried utilities were successfully identified using GPR-EML data fusion with the proposed autonomous algorithm, while each of the above techniques alone are incapable of identifying both characteristics.

## 5.4 The inversion of multi-sensing data in complex buried utility scenarios

To achieve objectives 7 of the thesis, a FE forward model was developed for complex EML survey scenarios, i.e. two adjacent buried cables. The developed FE model and GA were utilised to develop an inversion algorithm to determine the spatial location of both cables, and the flowing current (amperage) in each cable. The location and amperages of buried live cables were successfully determined with discrepancies between 0.38% to 22.26%. The 22.26% discrepancy between the estimated and actual location of the second cable corresponds to less than a 7 cm distance between the estimated and actual locations (Objective 8).

The robustness of the proposed novel inversion algorithm was confirmed by numerical examples. However, such an outcome does not account for environmental noise in the data. Therefore, further analysis using experimental and field data is required to confirm the efficiency of this algorithm for real-world scenarios.

Through meeting the objectives of the thesis, as demonstrated above, novel robust autonomous computational solutions were developed using a combination of classic and AI-based processing algorithms to interpret GPR data. Furthermore, data fusion algorithms were developed to fuse complementary data from GPR, EML and EMC which resulted in improved detection and locating of shallow buried utilities (e.g. pipes and cables) and ground anomalies (e.g. ground disturbance). Hence, the aim of the thesis was achieved.

## 5.5 Recommendations for future work

### The semi-and autonomous algorithm for processing GPR data

- The random walk model, used in the development of the semi-and autonomous algorithms, has simplified assumptions about the ground model. Thus, employing a more representative model of the system which can capture its complexities would potentially improve the performance of the developed algorithms.
- In addition to the GPR filters used in this research, namely time zero correction, signal gain adjustment, bandpass filter, moving average and DC-shift removal, it is proposed to investigate other available signal processing filters to further improve GPR signal interpretation.
- Integrating AI approaches such as Artificial Neural Networks as a pattern recognition technique for identifying the hyperbola signatures, which unlike the proposed algorithms in this research does not rely on the electrical impedance contrast in the ground, may strengthen the identification of the hyperbola patterns in different ground and noise conditions.
- Despite the capability of KF that was demonstrated in this research, it is a linear estimator which limits its accuracy for modelling non-linear problems. It is suggested to employ improved versions of KF which can incorporate non-linear modelling. This should potentially improve the accuracy of GPR data interpretation in the framework of the proposed algorithms.
- The developed NIS-signal function in this research is a crucial element of the proposed algorithms to distinguish noise from actual target signals. This function was calibrated for the available data using trial-and-error approach. However, the threshold of the NIS-signal function is case dependent and may vary for different GPR survey data sets. Thus, it is suggested to develop a systematic approach for estimating the thresholds for processing the NIS-signal functions.

### The fusion algorithm for GPR, EMC, and EML

- In the current research, EMC data was inverted using EMagPy which is a multipurpose inversion tool. Then the output result as an image was manually fed to the developed algorithms. It is recommended to develop an EMC inversion tool which is compatible with the developed algorithms. This will automate the fusion process and eliminate the manual interference which, in turn, improve the time efficiency of the algorithms.

### Numerical inversion-fusion algorithm for EML and GPR

- In the current study, the EML test was carried out in active mode with one specific frequency. However, activating a buried conductor with different frequencies result in different EML outputs which in turn influences the result of the proposed fusion-inversion algorithm. Thus, it is recommended to investigate different active EML modes, as well as passive EML mode, to understand its influence on the performance of the fusion-inversion algorithm.

- Unlike the 2D test cases presented in this study, in many conditions the orientation of the buried cables can only be represented in 3D modelling. So that in is recommended to expand the proposed fusion-inversion algorithm to 3-dimentional to be able to incorporate more realistic scenarios.
- As a general recommendation, further data acquisition through field and laboratory GPR, EML and EMC tests in different soil and buried target scenarios are required to address the above recommendations for improving the proposed GPR data processing as well as data fusion algorithms.

# References


Abdelgwad, A.H., Nashat, A.A., 2017. Investigation of Utilizing L-Band Horn Antenna in Landmine Detection.

Abdelnour, A., Lazaro, A., Villarino, R., Kaddour, D., Tedjini, S., Girbau, D., 2018. Passive Harmonic RFID System for Buried Assets Localization. Sensors 18, 3635.

Abeynayake, C.G., Chant, I.J., Nash, G., 2003. Kalman detection of landmines in metal detector array data. Detect. Remediat. Technol. Mines Minelike Targets VIII 5089, 1243.

Abraham, D.A., 2017. Signal Processing. In: Applied Underwater Acoustics: Leif Bjørnø. Elsevier, pp. 743–807.

Abujarad, F., 2007. Ground penetrating radar signal processing for landmine detection. Work.

Adamopoulos, E., Colombero, C., Comina, C., Rinaudo, F., Volinia, M., Girotto, M., Ardissono, L., 2021. Integrating multiband photogrammetry, scanning, and GPR for built heritage surveys: The façades of Castello del Valentino. ISPRS Ann. Photogramm. Remote Sens. Spat. Inf. Sci. 8, 1–8.

Addison, P.S., 2017. The illustrated wavelet transform handbook : introductory theory and applications in science, engineering, medicine and finance, Second Edi. ed, CRC Press. CRC Press, Edinburgh.

Addison, P.S., 2018. Introduction to redundancy rules: The continuous wavelet transform comes of age. Philos. Trans. R. Soc. A Math. Phys. Eng. Sci.

Ahn, S.J., Rauh, W., Warnecke, H.-J., 2001. Least-squares orthogonal distances fitting of circle, sphere, ellipse, hyperbola, and parabola. Pattern Recognit. 34, 2283–2303.

Akagi, H., Tamura, S., 2006. A passive EMI filter for eliminating both bearing current and ground leakage current from an inverter-driven motor. IEEE Trans. Power Electron. 21, 1459–1468.

Al-Habaibeh, A., Anderson, G., Damji, K., Jones, G., Lam, K., 2010. Using infrared thermography for monitoring thermal efficiency of buildings - case studies from Nottingham Trent University.

Al-Nuaimy, W., 1999. Automatic feature detection and interpretation in ground-penetrating radar data. The University of Liverpool.

Al-Nuaimy, W., Huang, Y., Nakhkash, M., Fang, M.T.C., Nguyen, V.T., Eriksen, A., 2000. Automatic detection of buried utilities and solid objects with GPR using neural networks and pattern recognition, Journal of Applied Geophysics.

Ali, H., Ahmad Zaidi, A.F., Wan Ahmad, W.K., Zanar Azalan, M.S., Tengku Amran, T.S., Ahmad, M.R., Elshaikh, M., 2021. A cascade hyperbolic recognition of buried objects using hybrid feature extraction in ground penetrating radar images. In: Journal of Physics: Conference Series. IOP Publishing, p. 012018.

Alizadeh, V., Fehri, H., Kessentini, M., 2019. Less is more: From multi-objective to mono-

objective refactoring via developer's knowledge extraction. In: Proceedings - 19th IEEE International Working Conference on Source Code Analysis and Manipulation, SCAM 2019. Institute of Electrical and Electronics Engineers Inc., pp. 181–192.

Allen, G.E., 2003. Mendel and modern genetics: The legacy for today. Endeavour.

Allred, B.J., Daniels, J.J., Fausey, N.R., Chen, C., Peters, L., Youn, H., 2013. Important considerations for locating buried agricultural drainage pipe using ground penetrating radar. Appl. Eng. Agric. 21, 71–87.

Amaral, L.C.M., Roshan, A., Bayat, A., 2022. Review of Machine Learning Algorithms for Automatic Detection of Underground Objects in GPR Images. J. Pipeline Syst. Eng. Pract. 13, 04021082.

American Society of Civil Engineers, 2022. Standard Guideline for Investigating and Documenting Existing Utilities, Standard Guideline for Investigating and Documenting Existing Utilities.

Anindya, C., Fuada, S., Firmansyah, S., Lestari, D., 2014. Prototyping of the Pipe Tracking Robot Using Electronic Nose Technology as Gas Leakage Detector 10, 1–5.

Annan, P., 2009. Electromagnetic principles of ground penetrating radar. Gr. Penetrating Radar 1–40.

Atef, A., Zayed, T., Hawari, A., Khader, M., Moselhi, O., 2016. Multi-tier method using infrared photography and GPR to detect and locate water leaks. Autom. Constr. 61, 162–170.

Auger, F., Hilairet, M., Guerrero, J.M., Monmasson, E., Orlowska-Kowalska, T., Katsura, S., 2013. Industrial applications of the kalman filter: A review. IEEE Trans. Ind. Electron. 60, 5458–5471.

Baek, J., Yoon, J.S., Lee, C.M., Choi, Y., 2018. A Case Study on Detection of Subsurface Cavities of Urban Roads Using Ground-coupled GPR. In: 2018 17th International Conference on Ground Penetrating Radar, GPR 2018. Institute of Electrical and Electronics Engineers Inc.

Bai, H., Sinfield, J. V., 2020. Improved background and clutter reduction for pipe detection under pavement using Ground Penetrating Radar (GPR). J. Appl. Geophys. 172, 103918.

Bailey, D.H., Swarztrauber, P.N., 1994. A Fast Method for the Numerical Evaluation of Continuous Fourier and Laplace Transforms. SIAM J. Sci. Comput. 15, 1105–1110.

Baili, J., Lahouar, S., Hergli, M., Al-Qadi, I.L., Besbes, K., 2009. GPR signal de-noising by discrete wavelet transform. NDT E Int. 42, 696–703.

Baker, G.S., Jordan, T.E., Pardy, J., 2007. An introduction to ground penetrating radar (GPR). Spec. Pap. Soc. Am. 432, 1.

Bar-Shalom, Y., Li, X.-R., Kirubarajan, T., 2001. Estimation with Applications to Tracking and Navigation, Estimation with Applications to Tracking and Navigation.

Beamish, D., 2011. Low induction number, ground conductivity meters: A correction procedure in the absence of magnetic effects. J. Appl. Geophys. 75, 244–253.

Belotti, V., Dell'Acqua, F., Gamba, P., 2003. Denoising and Hyperbola Recognition in GPR data.

In: Serpico, S.B. (Ed.), Image and Signal Processing for Remote Sensing VII. pp. 70–80.

Benedetto, A., Benedetto, F., 2014. Application Field-Specific Synthesizing of Sensing Technology: Civil Engineering Application of Ground-Penetrating Radar Sensing Technology. In: Comprehensive Materials Processing. Elsevier, pp. 393–425.

Benedetto, A., Tosti, F., Bianchini Ciampoli, L., D'Amico, F., 2017. An overview of ground-penetrating radar signal processing techniques for road inspections. Signal Processing 132, 201–209.

Benedetto, A., Tosti, F., Luca Bianchini Ciampoli, F.D., 2013. An overview of ground-penetrating radar signal processing techniques for road inspections. J. Chem. Inf. Model. 53, 1689–1699.

Bentley, L.R., Trenholm, N.M., 2002. The Accuracy of Water Table Elevation Estimates Determined from Ground Penetrating Radar Data. J. Environ. Eng. Geophys. 7, 37–53.

Bi, W., Zhao, Y., An, C., Hu, S., 2018. Clutter elimination and random-noise denoising of GPR signals using an SVD method based on the Hankel matrix in the local frequency domain. Sensors (Switzerland) 18, 3422.

Bianchini Ciampoli, L., Gagliardi, V., Clementini, C., Latini, D., Del Frate, F., Benedetto, A., 2020. Transport Infrastructure Monitoring by InSAR and GPR Data Fusion. Surv. Geophys.

Bigman, D.P., 2012. The use of electromagnetic induction in locating graves and mapping cemeteries: An Example from native North America. Archaeol. Prospect. 19, 31–39.

Billings, S.D., 2004. Discrimination and classification of buried unexploded ordnance using magnetometry. IEEE Trans. Geosci. Remote Sens. 42, 1241–1251.

Bimpas, M., Amditis, A., Uzunoglu, N., 2010. Detection of water leaks in supply pipes using continuous wave sensor operating at 2.45 GHz. J. Appl. Geophys. 70, 226–236.

Binder, D., Bru¨cklbru¨bru¨ckl, E., Roch, K.H., Behm, M., Scho¨ner, W., Scho¨ner, S., Hynek, B., 2020. Determination of total ice volume and ice-thickness distribution of two glaciers in the Hohe Tauern region, Eastern Alps, from GPR data.

Birkenfeld, S., 2010. Automatic detection of reflexion hyperbolas in gpr data with neural networks. World Autom. Congr. (WAC), 2010 1–6.

Bishop, M.P., Young, B.W., Huo, D., Chi, Z., 2020. Spatial Analysis and Modeling in Geomorphology. In: Reference Module in Earth Systems and Environmental Sciences. Elsevier.

Blizkovsky, M., 1979. Processing and applications in microgravity surveys. Geophys. Prospect. 27, 848–861.

Boddice, D., Atkins, P., Rodgers, A., Metje, N., Goncharenko, Y., Chapman, D., 2018. A novel approach to reduce environmental noise in microgravity measurements using a Scintrex CG5. J. Appl. Geophys. 152, 221–235.

Boddice, D., Metje, N., Tuckwell, G., 2019. Quantifying the effects of near surface density variation on quantum technology gravity and gravity gradient instruments. J. Appl. Geophys. 164, 160–178.

Bongs, K., Atkins, P.R., Baldwin, E., Boddice, D., 2016. GG-top gravity gradient sensors: archaeological prospecting, subsea mapping, underground navigation, urban mapping.

Borgioli, G., Capineri, L., Falorni, P., Matucci, S., Windsor, C.G., 2008. The detection of buried pipes from time-of-flight radar data. IEEE Trans. Geosci. Remote Sens. 46, 2254–2266.

Bostanudin, N., 2013. Computational methods for processing ground penetrating radar data.

Brice, R., 2001. Music Engineering. In: Music Engineering. Newnes, pp. 464–474.

Brunton, S.L., Kutz, J.N., 2017. Data Driven Science & Engineering - Machine Learning, Dynamical Systems, and Control.

Bryakin, I. V, Bochkarev, I. V, 2019. Cable Avoidance Tool.

BT opernreach, 2019. Cable Identification Booklet V4. British Telecommunications plc.

Bu, Z., Gao, G., Li, H.J., Cao, J., 2017. CAMAS: A cluster-aware multiagent system for attributed graph clustering. Inf. Fusion 37, 10–21.

Burger, H.R., Sheehan, A.F., JONES, C.H., 2006. Introduction to applied geophysics : exploring shallow subsurface. WW Norton New York.

Busby, J.P., Merritt, J.W., 1999. Quaternary deformation mapping with ground penetrating radar. J. Appl. Geophys. 41, 75–91.

Cagniard, L., 1953. Basic theory of the magneto-telluric method of geophysical prospecting. Geophysics 18, 605–635.

Cao, J., Wu, Z., Wu, J., Liu, W., 2013. Towards information-theoretic K-means clustering for image indexing. Signal Processing 93, 2026–2037.

Capineri, L., Grande, P., Temple, J.A.G., 1998. Advanced image-processing technique for real-time interpretation of ground-penetrating radar images. Int. J. Imaging Syst. Technol. 9, 51–59.

Capozzoli, L., Rizzo, E., 2017. Combined NDT techniques in civil engineering applications: Laboratory and real test. Constr. Build. Mater. 154, 1139–1150.

Carevic, D., 1999. Kalman filter-based approach to target detection and target-background separtion in ground-penetrating radar data. In: Detection and Remediation Technologies for Mines and Minelike Targets IV. p. 1284.

Carreño-Alvarado, E.P., Ayala-Cabrera, D., Pérez-García, R., Izquierdo, J., 2014. Identification of buried pipes using thermal images and data mining. Procedia Eng. 89, 1445–1451.

Carrick Utsi, E., 2017. Multichannel and Single Channel Systems. Gr. Penetrating Radar 153–160.

Carter, G.C., 1977. Receiver Operating Characteristics for a Linearly Thresholded Coherence Estimation Detector. IEEE Trans. Acoust. 25, 90–92.

Carter, J. V., Pan, J., Rai, S.N., Galandiuk, S., 2016. ROC-ing along: Evaluation and interpretation of receiver operating characteristic curves. Surg. (United States) 159, 1638–1645.

Carvalho, H.S., Heinzelman, W.B., Murphy, A.L., Coelho, C.J.N., 2003. A general data fusion architecture. Proc. 6th Int. Conf. Inf. Fusion, FUSION 2003 2, 1465–1472.

Cassidy, N., 2009a. Ground penetrating radar data processing, modelling and analysis, First Edit. ed, Ground Penetrating Radar. Elsevier.

Cassidy, N., 2009b. Electrical and magnetic properties of rocks, soils and fluids. Gr. Penetrating Radar 41–72.

Chang, C.I., 2010. Multiparameter receiver operating characteristic analysis for signal detection and classification. IEEE Sens. J. 10, 423–442.

Chang, K.-H., 2015. Multiobjective Optimization and Advanced Topics. e-Design 1105–1173.

Chaskalovic, J., 2008. Finite element methods for engineering sciences: Theoretical approach and problem solving techniques, Finite Element Methods for Engineering Sciences: Theoretical Approach and Problem Solving Techniques. Springer Berlin Heidelberg.

Che Ku Melor, C.K.N.A.H., Joret, A., Razali, M., Ponniran, A., Sulong, M.S., Omar, R., 2021. Frequency based signal processing technique for pulse modulation ground penetrating radar system. Int. J. Electr. Comput. Eng. 11, 4104–4112.

Chen, H., 2010. Probabilistic conic mixture model and its applications to mining spatial ground penetrating radar data. SIAM Conf. Data Min. (.

Chen, S.W., Chen, C.M., Liu, W.C., 1998. Electric current effects upon the Sn\Cu and Sn\Ni interfacial reactions. J. Electron. Mater. 27, 1193–1199.

Cheng, F.Y., 1999. Multiobjective optimum design of structures with genetic algorithm and game theory: application to life-cycle cost design. Comput. Mech. Struct. Eng. 1–16.

Cheng, L., Pamukcu, S., 2017. Underground Sensing Monitoring and Hazard Detection for Environment and Infrastructure, Elsevier Science. Elsevier Inc.

Christensen, N.B., Tølbøll, R.J., 2009. A lateral model parameter correlation procedure for one-dimensional inverse modelling. Geophys. Prospect. 57, 919–929.

Chui, C.K., Heil, C., 1992. An Introduction to Wavelets, Computers in Physics.

Ciampoli, L.B., Tosti, F., Economou, N., Benedetto, F., 2019. Signal Processing of GPR Data for Road Surveys. Geosciences 9, 96.

Clair, A., Sinha, S., 2012. State-of-the-technology review on water pipe condition, deterioration and failure rate prediction models! Urban Water J. 9, 85–112.

Cohen, M.X., 2014. Fundamentals of Time-Frequency Analyses in Matlab/Octave 238.

Collin, F., Warnant, R., 1995. Application of the Wavelet Transform for GPS Cycle Slip Correction and Comparison with Kalman Filter. Manuscripta Geod. 20, 161–172.

Collins, L.M., Huettel, L.G., Simpson, W.A., Tantum, S.L., 2002. Sensor fusion of EMI and GPR data for improved land mine detection. In: Detection and Remediation Technologies for Mines and Minelike Targets VII. pp. 872–879.

COMSOL Multiphysics, 2017. Detailed Explanation of the Finite Element Method (FEM).

COMSOL Multiphysics.

Conyers, L.B., 2013. Ground-Penetrating Radar for Archaeology, Third. ed. AltaMira Press.

Conyers, L.B., Lucius, J.E., 1996. Velocity Analysis in Archaeological Ground-Penetrating Radar Studies. Archaeol. Prospect. 3, 25–38.

Costello, S.B., Chapman, D.N., Rogers, C.D.F., Metje, N., 2007. Underground asset location and condition assessment technologies. Tunn. Undergr. Sp. Technol. 22, 524–542.

Čuček, L., Klemeš, J.J., Kravanja, Z., 2015. Overview of environmental footprints. Assess. Meas. Environ. Impact Sustain. 131–193.

Daniels, D.J., 2004. Ground Penetrating Radar-2nd edition.

Daniels, J.J., 2000. Fundamentals of Ground Penetrating Radar. In: Symposium on the Application of Geophysics to Engineering and Environmental Problems. Society of Exploration Geophysicists, pp. 62–142.

Dashti, S., Bray, J.D., Reilly, J., Glaser, S., Bayen, A., Mari, E., 2014. Evaluating the reliability of phones as seismic monitoring instruments. Earthq. Spectra 30, 721–742.

Dashti, S., Reilly, J., Bray, J.D., Bayen, A., Glaser, S., Mari, E., 2011. iShake: Using Personal Devices to Deliver Rapid Semi-Qualitative Earthquake Shaking Information. US Geol. Surv. Final Tech. Rep. 51.

Daubechies, I., 1990. The Wavelet Transform, Time-Frequency Localization and Signal Analysis. IEEE Trans. Inf. Theory 36, 961–1005.

Davis, P., De Silva, D., Marlow, D., Moglia, M., Gould, S., Burn, S., 2008. Failure prediction and optimal scheduling of replacements in asbestos cement water pipes. J. Water Supply Res. Technol. - AQUA 57, 239–252.

Dawalibi, F., 1986. Electromagnetic fields generated by overhead and buried short conductors part 2 - ground networks. IEEE Trans. Power Deliv. 1, 112–119.

Dawood, M., Narayanan, R.M., 2001. Receiver operating characteristics for the coherent UWB random noise radar. IEEE Trans. Aerosp. Electron. Syst. 37, 586–594.

De Coster, A., Pérez Medina, J.L., Nottebaere, M., Alkhalifeh, K., Neyt, X., Vanderdonckt, J., Lambot, S., 2019. Towards an improvement of GPR-based detection of pipes and leaks in water distribution networks. J. Appl. Geophys. 162, 138–151.

De Coster, A., Van der Wielen, A., Grégoire, C., Lambot, S., 2018. Evaluation of pavement layer thicknesses using GPR: A comparison between full-wave inversion and the straight-ray method. Constr. Build. Mater. 168, 91–104.

Deb, K., 2011. Multi-objective Optimisation Using Evolutionary Algorithms: An Introduction. In: Multi-Objective Evolutionary Optimisation for Product Design and Manufacturing. Springer, London, pp. 3–34.

Deb, K., Deb, K., 2014. Multi-objective Optimization. Search Methodol. Introd. Tutorials Optim. Decis. Support Tech. Second Ed. 403–449.

Denisov, A.Y., Denisova, O. V., Sapunov, V.A., Khomutov, S.Y., 2006. Measurement quality estimation of proton-precession magnetometers. Earth, Planets Sp. 58, 707–710.

Dewantara, D., Parnadi, W.W., 2022. Automatic Hyperbola Detection and Apex Extraction Using Convolutional Neural Network on GPR Data. In: Journal of Physics: Conference Series. IOP Publishing, p. 012027.

Dey, A., Morrison, H.F., 1979. Resistivity modelling for arbitrarily shaped two-dimensional structures. Geophys. Prospect. 27, 106–136.

Dickinson, E.J.F., Ekström, H., Fontes, E., 2014. COMSOL Multiphysics®: Finite element software for electrochemical analysis. A mini-review. Electrochem. commun.

Difrancesco, D., Grierson, A., Kaputa, D., Meyer, T., 2009. Gravity gradiometer systems - advances and challenges. In: Geophysical Prospecting. Blackwell Publishing Ltd, pp. 615–623.

Dojack, L., 2012. Ground Penetrating Radar Theory, Data Collection, Processing, and Interpretation: A Guide for Archaeologists. Proj. Report, Univ. Br. Columbia.

Dondurur, D., 2018. Acquisition and Processing of Marine Seismic Data, Acquisition and Processing of Marine Seismic Data. Elsevier.

Doolittle, J.A., Brevik, E.C., 2014. The use of electromagnetic induction techniques in soils studies. Geoderma.

Dossi, M., Forte, E., Pipan, M., 2015. Automated reflection picking and polarity assessment through attribute analysis: Theory and application to synthetic and real ground-penetrating radar data. Geophysics 80, H23–H35.

Dou, Q., Wei, L., Magee, D.R., Atkins, P., Chapman, D.N., Curioni, G., Goddard, K.F., Hayati, F., Jenks, H., Metje, N., Muggleton, J., Pennock, S.R., Rustighi, E., Swingler, S.G., Rogers, C.D.F., Cohn, A.G., 2016. 3D buried utility location using a marching-cross-section algorithm for multi-sensor data fusion. Sensors 16, 1827.

Dou, Q., Wei, L., Magee, D.R., Cohn, A.G., 2017. Real-Time Hyperbola Recognition and Fitting in GPR Data. IEEE Trans. Geosci. Remote Sens. 55, 51–62.

Downs, C., Jazayeri, S., 2021. Resolution enhancement of deconvolved ground penetrating radar images using singular value decomposition. J. Appl. Geophys. 193, 104401.

Duda, R.O., Hart, P.E., 2002. Use of the Hough transformation to detect lines and curves in pictures. Commun. ACM 15, 11–15.

Dutta, R., Cohn, A.G., Muggleton, J.M., 2013. 3D mapping of buried underworld infrastructure using dynamic Bayesian network based multi-sensory image data fusion. J. Appl. Geophys. 92, 8–19.

Economou, N., Benedetto, F., Bano, M., Tzanis, A., Nyquist, J., Sandmeier, K.J., Cassidy, N., 2017. Advanced Ground Penetrating Radar Signal Processing Techniques. Signal Processing 132, 197–200.

El Afi, M., Belattar, S., 2014. Application of the Infrared Thermography in Detection of Pipes in Concrete Structures.

Elwaseif, M., Robinson, J., Day-Lewis, F.D., Ntarlagiannis, D., Slater, L.D., Lane, J.W., Minsley, B.J., Schultz, G., 2017. A matlab-based frequency-domain electromagnetic inversion code (FEMIC) with graphical user interface. Comput. Geosci. 99, 61–71.

Eppstein, M.J., Dougherty, D.E., 1998. and moisture monitoring from cross-well ground-penetrating radar test site Vermont phase liquid ( DNAPL ) plume month ; Water Resour. Res. 34, 1889–1900.

Ernenwein, E.G., Kvamme, K.L., 2008. Data processing issues in large-area GPR surveys: Correcting trace misalignments, edge discontinuities and striping. Archaeol. Prospect. 15, 133–149.

Ertöz, L., Steinbach, M., Kumar, V., 2013. Finding Clusters of Different Sizes, Shapes, and Densities in Noisy, High Dimensional Data 47–58.

Ervasti, M., Dashti, S., Reilly, J., Bray, J.D., Bayen, A., Glaser, S., 2011. iShake. Proc. 10th Int. Conf. Mob. Ubiquitous Multimed. - MUM '11 43–52.

Esfetanaj, N.N., Nojavan, S., 2018. The Use of Hybrid Neural Networks, Wavelet Transform and Heuristic Algorithm of WIPSO in Smart Grids to Improve Short-Term Prediction of Load, Solar Power, and Wind Energy. In: Operation of Distributed Energy Resources in Smart Distribution Networks. pp. 75–100.

Everett, M.E., 2013. Near-Surface Applied Geophysics. Cambridge University Press.

Fahmy, M., Moselhi, O., 2017. Detecting and Locating Leaks in Underground Water Mains Using Thermography. Proc. 2009 Int. Symp. Autom. Robot. Constr. (ISARC 2009) 61–67.

Farzamian, M., Paz, M.C., Paz, A.M., Castanheira, N.L., Gonçalves, M.C., Monteiro Santos, F.A., Triantafilis, J., 2019. Mapping soil salinity using electromagnetic conductivity imaging—A comparison of regional and location-specific calibrations. L. Degrad. Dev. 30, 1393–1406.

Faundez-Zanuy, M., 2009. Data fusion at different levels. In: Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). Springer, Berlin, Heidelberg, pp. 94–103.

Feng, M., Fukuda, Y., Mizuta, M., Ozer, E., 2015. Citizen sensors for SHM: Use of accelerometer data from smartphones. Sensors 15, 2980–2998.

Feng, S., Liu, D., Cheng, X., Fang, H., Li, C., 2017. A new segmentation strategy for processing magnetic anomaly detection data of shallow depth ferromagnetic pipeline. J. Appl. Geophys. 139, 65–72.

Fischer, G., Fraedrich, V., Wendel, C., 2002. Method and device for locating a metal line. US 6,411,073 B1 (accessed 30/01/2020).

Fleming, P.R., Frost, M.W., Lambert, J.P., 2008. Review of Lightweight Deflectometer for Routine in Situ Assessment of Pavement Material Stiffness. Transp. Res. Rec. J. Transp. Res. Board 2004, 80–87.

Fonseca, C., Fleming, P., 1993. Genetic Algorithms for Multiobjective Optimization: Formulation, Discussion and Generalization. Proc. 5th Int. Conf. Genet. Algorithms 624.

Foufoula-Georgiou, E., Kumar, P., 1994. Wavelet Analysis in Geophysics: An Introduction,

Wavelet Analysis and Its Applications.

Fourier, J.B.J., 1829. La theorie analytique de la chaleur. Mem Acad.R.Sci. 8.

Fournier, H.G., 1966. Essai d'un historique des connasissances magnéto-telluriques. Institut de physique du globe.

Fox, P.J., 2011. Sonar imaging of flooded subsurface voids phase I : proof of concept. Ohio State University.

Gallagher, K., Charvin, K., Nielsen, S., Sambridge, M., Stephenson, J., 2009. Markov chain Monte Carlo (MCMC) sampling methods to determine optimal models, model resolution and model choice for Earth Science problems. Mar. Pet. Geol. 26, 525–535.

Gamba, P., Belotti, V., 2003. Two fast buried pipe detection schemes in Ground Penetrating Radar images. Int. J. Remote Sens. 24, 2467–2484.

Gamba, P., Lossani, S., 2000. Neural detection of pipe signatures in ground penetrating radar images. IEEE Trans. Geosci. Remote Sens. 38, 790–797.

Gao, Y., Brennan, M.J., Joseph, P.F., Muggleton, J.M., Hunaidi, O., 2005. On the selection of acoustic/vibration sensors for leak detection in plastic water pipes. J. Sound Vib. 283, 927–941.

García-Escudero, L.A., Mayo-Iscar, A., Sánchez-Gutiérrez, C.I., 2017. Fitting parabolas in noisy images. Comput. Stat. Data Anal. 112, 80–87.

Gauss, K.F., 1857. Theory of the Motion of the Heavenly Bodies Moving about the Sun in Conic Sections. Math. Comput. 18, 525.

Gebremichael, T.T., 2013. Signal processing techniques for landmine detection using impulse ground penetrating radar. TU Darmstadt. TU Darmstadt.

Gennert, M.A., Yuille, A.L., 1988. Determining the optimal weights in multiple objective function optimization 87–89.

Giannakis, I., Giannopoulos, A., Warren, C., 2019. A Machine Learning-Based Fast-Forward Solver for Ground Penetrating Radar with Application to Full-Waveform Inversion. IEEE Trans. Geosci. Remote Sens. 57, 4417–4426.

Giannakis, I., Giannopoulos, A., Warren, C., 2021. A Machine Learning Scheme for Estimating the Diameter of Reinforcing Bars Using Ground Penetrating Radar. IEEE Geosci. Remote Sens. Lett. 18, 461–465.

GICHD, G.I.C. for H.D., 2003. Mine Detection Dogs: Training, Operations and Odour Detection. Geneva International Centre for Humanitarian Demining GICHD.

Gill, P.E., Murray, W., Wright, M.H., 2019. Practical Optimization. Society for Industrial and Applied Mathematics, Philadelphia, PA.

Goddard, K.F., Wang, P., Lewin, P.L., Swingler, S.G., 2012. Detection and location of underground cables using magnetic field measurements. Meas. Sci. Technol. 23, 055002.

Gonzalez-DIaz, M., Garcia-Fernandez, M., Alvarez-Lopez, Y., Las-Heras, F., 2020. Improvement

of GPR SAR-Based Techniques for Accurate Detection and Imaging of Buried Objects. IEEE Trans. Instrum. Meas. 69, 3126–3138.

González-Huici, M.A., Catapano, I., Soldovieri, F., 2014. A comparative study of GPR reconstruction approaches for landmine detection. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 7, 4869–4878.

González-Quirós, A., Comte, J.C., 2021. Hydrogeophysical model calibration and uncertainty analysis via full integration of PEST/PEST++ and COMSOL. Environ. Model. Softw. 145, 105183.

Goodman, D., Piro, S., 2013a. GPR Remote Sensing in Archaeology, Geotechnologies and the Environment. Springer-Verlag Berlin Heidelberg.

Goodman, D., Piro, S., 2013b. GPR Remote Sensing in Archaeology. Geotechnol. Environ. Springer-Verlag Berlin Heidelberg.

Goodman, D., Piro, S., 2013c. GPR remote sensing in archaeology -Chapter2. In: GPR Remote Sensing in Archaeology. pp. 1–233.

Goschnick, J., Harms, M., 2002. Landmine Detection with an Electronic Nose Mounted on an Airship. In: Detection of Explosives and Landmines. Springer, Dordrecht, pp. 83–91.

Greaves, R.J., Lesmes, D.P., Lee, J.M., Toksöz, M.N., 1996. Velocity variations and water content estimated from multi-offset, ground-penetrating radar. Geophysics 61, 683–695.

Greggio, N., Giambastiani, B.M.S., Balugani, E., Amaini, C., Antonellini, M., 2018. High-resolution electrical resistivity tomography (ERT) to characterize the spatial extension of freshwater lenses in a salinized coastal aquifer. Water (Switzerland) 10, 1067.

Grewal, M., Andrews, A., 2008. Kalman filtering : Theory and Practice Using MATLAB.

Gros, 1997a. NDT data fusion. Arnold, London.

Gros, 1997b. Data Fusion – A Review. Butterworth-Heinemann, Oxford, pp. 5–42.

Guillemoteau, J., Christensen, N.B., Jacobsen, B.H., Tronicke, J., 2017. Fast 3D multichannel deconvolution of electromagnetic induction loop-loop apparent conductivity data sets acquired at low induction numbers. Geophysics 82, E357–E369.

Gunatilake, A., Piyathilaka, L., Tran, A., Vishwanathan, V.K., Thiyagarajan, K., Kodagoda, S., 2021. Stereo vision combined with laser profiling for mapping of pipeline internal defects. IEEE Sens. J. 21, 11926–11934.

Guo, Z.-Y., Liu, D.-J., Pan, Q., Zhang, Y.-Y., 2015. Forward modeling of total magnetic anomaly over a pseudo-2D underground ferromagnetic pipeline. J. Appl. Geophys. 113, 14–30.

Gurbuz, A.C., 2012. Determination of background distribution for ground-penetrating radar data. IEEE Geosci. Remote Sens. Lett. 9, 544–548.

Habib, M.K., 2007. Controlled biological and biomimetic systems for landmine detection. Biosens. Bioelectron. 23, 1–18.

Hadagali, M.D., Suan, C.L., 2017. Advancement of sensitive sniffer bee technology. TrAC -

Trends Anal. Chem. 97, 153–158.

Hafsi, M., Bolon, P., Dapoigny, R., 2017. Detection and localization of underground networks by fusion of electromagnetic signal and GPR images. In: Nagahara, H., Umeda, K., Yamashita, A. (Eds.), Thirteenth International Conference on Quality Control by Artificial Vision 2017. SPIE, p. 1033803.

Haines, A.L., Mills, K.L., Filliben, J.J., 2014. Determining Relative Importance and Best Settings for Genetic Algorithm Control Parameters. Evol. Comput. 1–34.

Hamran, S.E., Paige, D.A., Amundsen, H.E.F., Berger, T., Brovoll, S., Carter, L., Damsgård, L., Dypvik, H., Eide, J., Eide, S., Ghent, R., Helleren, Ø., Kohler, J., Mellon, M., Nunes, D.C., Plettemeier, D., Rowe, K., Russell, P., Øyan, M.J., 2020. Radar Imager for Mars' Subsurface Experiment—RIMFAX. Space Sci. Rev.

Hao, T., Rogers, C.D.F., Metje, N., Chapman, D.N., Muggleton, J.M., Foo, K.Y., Wang, P., Pennock, S.R., Atkins, P., Swingler, S.G., Parker, J., Costello, S.B., Burrow, M.P.N., Anspach, J.H., Armitage, R.J., Cohn, A.G., Goddard, K., Lewin, P.L., Orlando, G., Redfern, M.A., Royal, A.C.D., Saul, A.J., 2012. Condition assessment of the buried utility service infrastructure. Tunn. Undergr. Sp. Technol. 28, 331–344.

Harkat, H., Elfakir, Y., Bennani, S.D., Khaissidi, G., Mrabti, M., 2016. Ground penetrating radar hyperbola detection using Scale-Invariant Feature Transform. In: Proceedings of 2016 International Conference on Electrical and Information Technologies, ICEIT 2016. IEEE, pp. 392–397.

Harkat, H., Ruano, A., Ruano, M.G., Bennani, S.D., 2018. Classifier Design by a Multi-Objective Genetic Algorithm Approach for GPR Automatic Target Detection. IFAC-PapersOnLine 51, 187–192.

Harkat, H., Ruano, A.E., Ruano, M.G., Bennani, S.D., 2019. GPR target detection using a neural network classifier designed by a multi-objective genetic algorithm. Appl. Soft Comput. J. 79, 310–325.

Hart, P., Duda, R.O., 1972. Use of the hough trasformtion to detect lines and curves in pictures. Commun. ACM 15, 11–15.

Hartikainen, J., Särkkä, S., 2011. Optimal filtering with Kalman filters and smoothers – a Manual for Matlab toolbox EKF / UKF. Dep. Biomed. Eng. Comput. Sci. Aalto Univ. Sch. Sci.

Haurum, J.B., Moeslund, T.B., 2020. A Survey on Image-Based Automation of CCTV and SSET Sewer Inspections. Autom. Constr. 111, 103061.

Herter, C., 2019. The electromagnetic spectrum: A critical natural resource. In: Transboundary Resources Law. pp. 89–101.

Hildebrandt, C., Raschner, C., Ammer, K., 2010. An overview of recent application of medical infrared thermography in sports medicine in Austria. Sensors 10, 4700–4715.

Ho, K.C., Gader, P.D., 2002. A linear prediction land mine detection algorithm for hand held ground penetrating radar. IEEE Trans. Geosci. Remote Sens. 40, 1374–1384.

Ho, K.C., Gader, P.D., Wilson, J.N., 2004. Improving landmine detection using frequency domain features from ground penetrating radar. In: International Geoscience and Remote Sensing Symposium (IGARSS). pp. 1617–1620.

Holtham, E., Marchant, D., McMillan, M., Haber, E., 2017. Advanced Modelling of Electromagnetic Data. Proc. Explor. 17 Sixth Decenn. Int. Conf. Miner. Explor. 209–219.

Huang, X., Wang, W., Lv, Z., Meng, Y., Yuan, C., Feng, J., 2019. Precise Positioning of the Underground Power Cable by Magnetic Field Detectio. In: Annual Report - Conference on Electrical Insulation and Dielectric Phenomena, CEIDP. Institute of Electrical and Electronics Engineers Inc., pp. 702–705.

Huber, E., Hans, G., 2018. RGPR - An open-source package to process and visualize GPR data. In: 2018 17th International Conference on Ground Penetrating Radar, GPR 2018. Institute of Electrical and Electronics Engineers Inc.

Huray, P.G., 2010. Maxwell ' s Equations, IEEE press. Wiley.

Idi, B.Y., Kamarudin, M.N., 2012. Interpretation of ground penetrating radar image using digital wavelet transform. Asian J. Appl. Sci. 5, 174–182.

Institution of Civil Engineers, 2014. Specification for underground utility detection, verification and location (PAS128:2014). BSI.

Iskander, M., 2018. Geotechnical Underground Sensing and Monitoring. In: Underground Sensing: Monitoring and Hazard Detection for Environment and Infrastructure. Academic Press, pp. 141–202.

Iswandy, A., Serma, A., Setan, H., 2009. Ground penetrating radar (GPR) for subsurface mapping: preliminary result. Geoinf. Sci. J. 9, 45–62.

James, J.F., 1969. A Student ' s Guide to Fourier Transforms, Cambridge University Press. Cambridge University Press.

Jentzen, A., Kloeden, P.E., 2011. Taylor approximations for stochastic partial differential equations, CBMS-NSF Regional Conference Series in Applied Mathematics. SIAM.

Jeong, H.S., Arboleda, C.A., Abraham, D.M., Halpin, D.W., Bernold, L.E., 2003. Imaging and Locating Buried Utilities. Technol. Transf. Proj. Implement. Inf. 7, 65,66.

Jianguo, Y., Dongli, Y., Xiaojun, X., Qiuling, J., 2008. Kalman filtering parameter optimization techniques based on genetic algorithm. Institute of Electrical and Electronics Engineers (IEEE), pp. 1717–1720.

Jianguo Yan, Dongli Yuan, Xiaojun Xing, Qiuling Jia, 2008. Kalman filtering parameter optimization techniques based on genetic algorithm. Institute of Electrical and Electronics Engineers (IEEE), pp. 1717–1720.

Jiles, D.C., 1990. Review of magnetic methods for nondestructive evaluation (Part 2). NDT Int. 23, 83–92.

Jol, H.M., 2009. Ground Penetrating Radar Theory and Application, Ground Penetrating Radar. Elsevier.

Kadioglu, S., 2018. GPR Data Visualization Enhancement by Deconvolution and Some Image Attributes. In: 2018 41st International Conference on Telecommunications and Signal Processing, TSP 2018. Institute of Electrical and Electronics Engineers Inc.

Kalkowski, M.K., Muggleton, J.M., Rustighi, E., 2018. Axisymmetric semi-analytical finite elements for modelling waves in buried/submerged fluid-filled waveguides. Comput. Struct. 196, 327–340.

Kalman, R.E., 1960. A new approach to linear filtering and prediction problems. J. Fluids Eng. Trans. ASME 82, 35–45.

Kang, J.M., Song, S., Park, D., Choi, C., 2017. Detection of cavities around concrete sewage pipelines using impact-echo method. Tunn. Undergr. Sp. Technol. 65, 1–11.

Kaniewski, P., Kraszewski, T., 2020. Novel algorithm for position estimation of handheld ground-penetrating radar antenna. In: Proceedings International Radar Symposium. IEEE Computer Society, pp. 100–102.

Kanungo, T., Mount, D.M., Netanyahu, N.S., Piatko, C.D., Silverman, R., Wu, A.Y., 2002. An efficient k-means clustering algorithms: Analysis and implementation. IEEE Trans. Pattern Anal. Mach. Intell. 24, 881–892.

Karamitrou, A., Bogiatzis, P., Tsokas, G.N., 2020. Fusion of geophysical images in the study of archaeological sites. Archaeol. Prospect. 27, 119–133.

Kemna, A., Kulessa, B., Vereecken, H., 2002. Imaging and characterisation of subsurface solute transport using electrical resistivity tomography (ERT) and equivalent transport models. J. Hydrol. 267, 125–146.

Kerekes, J., 2008. Receiver operating characteristic curve confidence intervals and regions. IEEE Geosci. Remote Sens. Lett. 5, 251–255.

Khamzin, A.K., Varnavina, A. V., Torgashov, E. V., Anderson, N.L., Sneed, L.H., 2017. Utilization of air-launched ground penetrating radar (GPR) for pavement condition assessment. Constr. Build. Mater. 141, 130–139.

Khan, U.S., Al-Nuaimy, W., 2010. Background removal from GPR data using eigenvalues. In: Proceedings of the 13th Internarional Conference on Ground Penetrating Radar, GPR 2010.

Kim, J.H., Cho, S.J., Yi, M.J., 2007. Removal of ringing noise in GPR data by signal processing. Geosci. J. 11, 75–81.

Kirkland, E.J., 2010. Bilinear Interpolation. In: Advanced Computing in Electron Microscopy. Springer, Boston, MA, pp. 261–263.

Kirkwood, J.R., 2015. Markov processes, 1st ed. CRC Press.

Knox, M., Rundel, C., Collins, L., 2017. Sensor fusion for buried explosive threat detection for handheld data. In: Detection and Sensing of Mines, Explosive Objects, and Obscured Targets XXII. p. 101820D.

Ko, N.Y., Kim, T.G., 2012. Comparison of Kalman filter and particle filter used for localization of an underwater vehicle. In: 2012 9th International Conference on Ubiquitous Robots

and Ambient Intelligence, URAI 2012. pp. 350–352.

Kobayashi, M., Uchikado, T., Nakano, K., 2012. Wavelet-based position detection of buried pipes from GPR signals by use of angle information. In: International Conference on Wavelet Analysis and Pattern Recognition. pp. 390–396.

Koganti, T., Van De Vijver, E., Allred, B.J., Greve, M.H., Ringgaard, J., Iversen, B. V., 2020. Mapping of agricultural subsurface drainage systems using a frequency-domain ground penetrating radar and evaluating its performance using a single-frequency multi-receiver electromagnetic induction instrument. Sensors 20, 1–26.

Koppenjan, S., 2009. Ground penetrating radar systems and design, First Edit. ed, Ground Penetrating Radar. Elsevier.

Kovalenko, V., Masalov, S.A., 2000. 2D matrix filtering of ground penetrating radar data. In: International Conference on Mathematical Methods in Electromagnetic Theory, MMET. IEEE Computer Society, pp. 236–238.

Ku, C.C., 1976. Numerical Inverse Magnetotelluric Problems. Geophysics 41, 276–286.

Kumar, D., 2012. Efficacy of electrical resistivity tomography technique in mapping shallow subsurface anomaly. J. Geol. Soc. India 80, 304–307.

Kvamme, K.L., Ernenwein, E.G., Menzer, J.G., 2019. Putting it all together: Geophysical data integration. In: Innovation in Near-Surface Geophysics. Elsevier, pp. 287–339.

Lai, W.W.L., Chang, R.K.W., Sham, J.F.C., Pang, K., 2016. Perturbation mapping of water leak in buried water pipes via laboratory validation experiments with high-frequency ground penetrating radar (GPR).

Lehner, G., 2008. Electromagnetic field theory for engineers and physicists, Electromagnetic Field Theory for Engineers and Physicists.

Lei, W., Hou, F., Xi, J., Tan, Q., Xu, M., Jiang, X., Liu, G., Gu, Q., 2019. Automatic hyperbola detection and fitting in GPR B-scan image. Autom. Constr. 106, 102839.

Lenham, J.W., McDonald, R., Miller, S., Reynolds, J.M., 2005. Integrated seismic investigations across the Mersey Estuary, Halton District, UK. Q. J. Eng. Geol. Hydrogeol. 38, 7–21.

Lenzerini, M., 2002. Data integration: A theoretical perspective. ACM SIGACT-SIGMOD-SIGART 233–246.

Lesparre, N., Robert, T., Nguyen, F., Boyle, A., Hermans, T., 2019. 4D electrical resistivity tomography (ERT) for aquifer thermal energy storage monitoring. Geothermics 77, 368–382.

Li, H., Chou, C., Fan, L., Li, B., Wang, D., Song, D., 2019. Toward Automatic Subsurface Pipeline Mapping by Fusing a Ground-Penetrating Radar and a Camera. IEEE Trans. Autom. Sci. Eng. 17, 722–734.

Li, Q., Li, R., Ji, K., Dai, W., 2016. Kalman filter and its application. Proc. - 8th Int. Conf. Intell. Networks Intell. Syst. ICINIS 2015 74–77.

Li, R., Fang, L., Deng, Y., Liu, S., 2010. Multi-parameter inverse analysis research based on

Comsol Multiphysics and Matlab. 2010 Int. Conf. Comput. Des. Appl. ICCDA 2010 2.

Li, S., Cai, H., Kamat, V.R., 2015. Uncertainty-aware geospatial system for mapping and visualizing underground utilities. Autom. Constr. 53, 105–119.

Li, Y., 2015. Detection and Identification of Subsurface Corrosion in Ferromagnetic Double-casing Pipes via Pulsed Remote Field Eddy Current Technique. In: Studies in Applied Electromagnetics and Mechanics. IOS Press, pp. 228–235.

Li, Y., Zhao, Z., Luo, Y., Qiu, Z., 2020. Real-time pattern-recognition of GPR images with YOLO V3 implemented by tensorflow. Sensors (Switzerland) 20, 1–17.

Liang, J., Tong, D., Tan, F., Jiao, Y.Y., Yan, C., 2021. Two-Dimensional magnetotelluric modelling based on the numerical manifold method. Eng. Anal. Bound. Elem. 124, 87–97.

Liggins, M.E., Hall, D.L., Llinas, J., 2009. Handbook of Multisensor Data Fusion: Theory and Practice, CRC Press.

Ligthart, L., Yarovoy, A., 2003. STW project “Advanced relocatable multi-sensor system for buried landmine detection.” In: Proceedings of the 2nd International Workshop on Advanced Ground Penetrating Radar.

Lim, M.K., Cao, H., 2013. Combining multiple NDT methods to improve testing effectiveness. Constr. Build. Mater. 38, 1310–1315.

Limleamthong, P., Guillén-Gosálbez, G., 2018. Combined Use of Bilevel Programming and Multi-objective Optimization for Rigorous Analysis of Pareto Fronts in Sustainability Studies: Application to the Redesign of the UK Electricity Mix. In: Computer Aided Chemical Engineering. Elsevier, pp. 1099–1104.

Liu, C., Song, C., Lu, Q., 2017. Random noise de-noising and direct wave eliminating based on SVD method for ground penetrating radar signals. J. Appl. Geophys. 144, 125–133.

Liu, G.R., Quek, S.S., 2003. Finite Element Method: A Practical Course, Finite Element Method: A Practical Course.

Liu, T., Su, Y., Huang, C., 2018. Inversion of ground penetrating radar data based on neural networks. Remote Sens. 10.

Liu, Z., Forsyth, D.S., 2004. Registration of multi-modal NDI images for aging aircraft. Res. Nondestruct. Eval. 15, 1–17.

Liu, Z., Forsyth, D.S., Komorowski, J.P., Hanasaki, K., Kirubarajan, T., 2007. Survey: State of the art in NDE data fusion techniques. IEEE Trans. Instrum. Meas. 56, 2435–2451.

Lopes, P.A.C., Gerald, J.A.B., Piedade, M.S., 2015. The Random Walk Model Kalman Filter in Multichannel Active Noise Control. IEEE Signal Process. Lett. 22, 2244–2248.

Lu, G., Zhao, W., Forte, E., Tian, G., Li, Y., Pipan, M., 2020. Multi-frequency and multi-attribute GPR data fusion based on 2-D wavelet transform. Meas. J. Int. Meas. Confed. 166, 108243.

Luo, T.X.H., Lai, W.W.L., 2020. GPR pattern recognition of shallow subsurface air voids. Tunn. Undergr. Sp. Technol. 99, 103355.

Luo, Y., Fang, G.Y., 2005. GPR clutter reduction and buried target detection by improved Kalman filter technique. In: 2005 International Conference on Machine Learning and Cybernetics, ICMLC 2005. IEEE, pp. 5432–5436.

Lytle, R.J., Dines, K.A., 1980. Iterative Ray Tracing Between Boreholes for Underground Image Reconstruction. IEEE Trans. Geosci. Remote Sens. 18, 234–240.

Maas, C., Schmalzl, J., 2013. Using pattern recognition to automatically localize reflection hyperbolas in data from ground penetrating radar. Comput. Geosci. 58, 116–125.

Madden, T.R., Swift, C.M., 1965. Magnetotelluric Studies of the Electrical Conductivity Structure of the Crust and Upper Mantle. Earth's Crust Up. Mantle 9, 317–345.

Mahmoudzadeh Ardekani, M.R., 2013. Off- and on-ground GPR techniques for field-scale soil moisture mapping. Geoderma 200–201, 55–66.

Maiti, S., Patra, S.K., Bhattacharya, A., 2013. Modelling GPR for characterization of subsurface em properties. In: 2013 IEEE MTT-S International Microwave and RF Conference, IMaRC 2013. IEEE Computer Society.

Makana, L.O., Metje, N., Jefferson, I., Sackey, M., Rogers, C.D., 2018. Cost Estimation of Utility Strikes: Towards Proactive Management of Street Works. Infrastruct. Asset Manag. 1–34.

Makar, J., Chagnon, N., 1999. Inspecting systems for leaks, pits, and corrosion. J. / Am. Water Work. Assoc. 91, 36–46.

Mallat, S., 2009a. A Wavelet Tour of Signal Processing, A Wavelet Tour of Signal Processing. Elsevier Inc.

Mallat, S., 2009b. A theory for multiresolution signal decomposition: The wavelet representation. In: Fundamental Papers in Wavelet Theory. pp. 494–513.

March, H.W., 1953. The field of a magnetic dipole in the presence of a conducting sphere. Geophysics 18, 671–684.

Marcus, M.B., Rosen, J., 2006. Markov processes and local times, 1st ed, Cambridge University Press. Cambridge University Press.

Marecos, V., Solla, M., Fontul, S., Antunes, V., 2017. Assessing the pavement subgrade by combining different non-destructive methods. Constr. Build. Mater. 135, 76–85.

Massaro, A., Savino, N., Selicato, S., Panarese, A., Galiano, A., Dipierro, G., 2021. Thermal IR and GPR UAV and vehicle embedded sensor non-invasive systems for road and bridge inspections. In: 2021 IEEE International Workshop on Metrology for Industry 4.0 and IoT, MetroInd 4.0 and IoT 2021 - Proceedings. Institute of Electrical and Electronics Engineers Inc., pp. 248–253.

Maybeck, P.S., 1979. Stochastic models, estimation, and control. Volume 1 Chapter 1. In: Stochastic Models, Estimation, and Control.

Mazor, E., Averbuch, A., Bar-Shalom, Y., Dayan, J., 1998. Interacting multiple model methods in target tracking: A survey. IEEE Trans. Aerosp. Electron. Syst. 34, 103–123.

McHugh, M.L., 2012. The Chi-square test of independence. Biochem. Medica 23, 143–149.

McLachlan, P., Blanchy, G., Binley, A., 2021. EMagPy: Open-source standalone software for processing, forward modeling and inversion of electromagnetic induction data. Comput. Geosci. 146.

McNeil, J.D., 1980. Electromagnetic Terrain Conductivity Measurement at Low Induction Numbers. Tech. note TN.

Meinhold, R.J., Singpurwalla, N.D., 1983. Understanding the kalman filter. Am. Stat. 37, 123–127.

Melo, D.C., Régis, C.R.T., 2017. 1D laterally constrained inversion of 2D MT data. pp. 154–159.

Meola, C., Carlomagno, G.M., 2004. Recent advances in the use of infrared thermography. Meas. Sci. Technol. 15, 27.

Mertens, L., Persico, R., Matera, L., Lambot, S., 2016. Automated Detection of Reflection Hyperbolas in Complex GPR Images with No A Priori Knowledge on the Medium. IEEE Trans. Geosci. Remote Sens. 54, 580–596.

Metje, N., Atkins, P., Brennan, M.J., Chapman, D.N., Lim, H.M., Machell, J., Muggleton, J.M., Pennock, S., Ratcliffe, J., Redfern, M., Rogers, C.D.F., Saul, A.J., Shan, Q., Swingler, S., Thomas, A.M., 2007. Mapping the Underworld - State-of-the-art review. Tunn. Undergr. Sp. Technol. 22, 568–586.

Metje, N., Chapman, D., Stringfellow, M., Rebegea, S., Tuckwell, G., Guy, M., Parkin, N., Roberts, D., Leech, C., Lord, A., 2022. Locating blockages in buried (telecoms) ducts: a new approach. Proc. Inst. Civ. Eng. - Smart Infrastruct. Constr. 175, 59–72.

Metje, N., Hojjati, A., Beck, A., Rogers, C.D.F., 2019. Improved Underground Utilities Asset Management – Assessing the impact of PAS128: Specification for Underground Utility Detection, Verification and Location in practice. Proc. Inst. Civ. Eng. - Munic. Eng. 1–31.

Middlemiss, R.P., Samarelli, A., Paul, D.J., Hough, J., Rowan, S., Hammond, G.D., 2016. Measurement of the Earth tides with a MEMS gravimeter. Nature 531, 614–617.

Middleton, W.M., Valkenburg, M.E. van, 2002. Reference data for engineers, 9th ed, Reference data for engineers. Elsevier.

Milsom, J., Eriksen, A., 2011. Field geophysics, 4th ed, John Wiley & Sons Ltd. John Wiley & Sons Ltd.

Ministère de la Transition Écologique et Solidaire, 2019. Guide d'application de la réglementation relative aux travaux à proximité des réseaux Fascicule 1, Version 2 76.

Moghadas, D., 2019. Probabilistic Inversion of Multiconfiguration Electromagnetic Induction Data Using Dimensionality Reduction Technique: A Numerical Study. Vadose Zo. J. 18, 1–16.

Moghadas, D., André, F., Slob, E.C., Vereecken, H., Lambot, S., 2010. Joint full-waveform analysis of off-ground zero-offset ground penetrating radar and electromagnetic induction synthetic data for estimating soil electrical properties. Geophys. J. Int. 182, 1267–1278.

Moghadas, D., André, F., Vereecken, H., Lambot, S., 2009. Full-waveform modeling and joint

inversion of electromagnetic induction and ground penetrating radar data. In: Near Surface 2009 - 15th European Meeting of Environmental and Engineering Geophysics. European Association of Geoscientists and Engineers, EAGE.

Mogharehabed, T., 2014. Experimental investigation of cast iron corrosion. University of Birmingham.

Mohd Shahrul Azmi, M.Y., 2019. Defect Detection in Pipeline Using Acoustic Pulse Reflectometry School of Mechanical, Aerospace and Civil Engineering.

Monteiro Santos, F.A., 2004a. 1-D laterally constrained inversion of EM34 profiling data. J. Appl. Geophys. 56, 123–134.

Monteiro Santos, F.A., 2004b. 1D laterally constrained inversion of 2D MT data. J. Appl. Geophys. 56, 123–134.

Monteiro Santos, F.A., Triantafilis, J., Bruzgulis, K., 2011. A spatially constrained 1D inversion algorithm for quasi-3D conductivity imaging: Application to DUALEM-421 data collected in a riverine plain. Geophysics 76.

Monteiro Santos, F.A., Triantafilis, J., Taylor, R.S., Holladay, S., Bruzgulis, K.E., 2010. Inversion of conductivity profiles from EM using full solution and a 1-D laterally constrained algorithm. J. Environ. Eng. Geophys. 15, 163–174.

Muggleton, J., Papandreou, B., Brennan, M., 2012. Detection of buried pipes using a shear wave ground surface vibration technique. Scopus.

Muggleton, J.M., Brennan, M.J., Gao, Y., 2011. Determining the location of buried plastic water pipes from measurements of ground surface vibration. J. Appl. Geophys. 75, 54–61.

Muggleton, J.M., Hunt, R., Rustighi, E., Lees, G., Pearce, A., 2020. Gas pipeline leak noise measurements using optical fibre distributed acoustic sensing. J. Nat. Gas Sci. Eng. 78, 103293.

Muggleton, J.M., Papandreou, B., 2014. A shear wave ground surface vibration technique for the detection of buried pipes. J. Appl. Geophys. 106, 164–172.

Muggleton, J.M., Rustighi, E., 2013. 'Mapping the Underworld': recent developments in vibro-acoustic techniques to locate buried infrastructure. Géotechnique Lett. 3, 137–141.

Myung, I.J., 2003. Tutorial on maximum likelihood estimation. J. Math. Psychol. 47, 90–100.

Nabavi, S. mohammed, 2019. Inversion and Depth Range of Dipole Electromagnetic Induction Measurements Measurements in Geophysics.

Nadim, G., 2008. Clutter reduction and detection of landmine objects in ground penetrating radar data using likelihood method. 2008 3rd Int. Symp. Commun. Control. Signal Process. ISCCSP 2008 98–106.

Nationalgrid, 2015. Undergrounding high voltage electricity transmission lines: The technical issues. Networks 18.

Nayana, D., Mamatha, K.H., Dinesh, S. V., Lokesh, T.R., 2021. Pavement evaluation using falling weight deflectometer (FWD). Springer, Singapore, pp. 445–456.

Nega, A., Nikraz, H., Al-Qadi, I.L., 2016. Dynamic analysis of falling weight deflectometer. J. Traffic Transp. Eng. (English Ed. 3, 427–437.

Neyamadpour, A., Taib, S., Wan Abdullah, W.A.T., 2009. Using artificial neural networks to invert 2D DC resistivity imaging data for high resistivity contrast regions: A MATLAB application. Comput. Geosci. 35, 2268–2274.

Ng, W., Chan, T.C.T., So, H.C., Ho, K.C., 2008. Particle filtering based approach for landmine detection using ground penetrating radar. IEEE Trans. Geosci. Remote Sens. 46, 3739–3755.

Nguyen, V.H., Pyun, J.Y., 2015. Location detection and tracking of moving targets by a 2D IR-UWB radar system. Sensors 15, 6740–6762.

Ni, S.H., Huang, Y.H., Lo, K.F., Lin, D.C., 2010. Buried pipe detection by ground penetrating radar using the discrete wavelet transform. Comput. Geotech. 37, 440–448.

NJUG, 2013. NJUG Guidelines on the Positioning and Colour Coding of Underground Utilities' Apparatus, NJUG Publication.

Nor, K.A.M., Muthalif, A.G.A., Wahid, A.N., 2015. Optimization in active vibration control: Virtual experimentation using COMSOL multiphysics - MATLAB integration. Proc. - Int. Conf. Intell. Syst. Model. Simulation, ISMS 2015-Septe, 385–389.

Núñez-Nieto, X., Solla, M., Gómez-Pérez, P., Lorenzo, H., 2015. Signal-to-noise ratio dependence on ground penetrating radar antenna frequency in the field of landmine and UXO detection. Meas. J. Int. Meas. Confed. 73, 24–32.

Oberti, F., Teschioni, A., Regazzoni, C.S., 1999. ROC curves for performance evaluation of video sequences processing systems for surveillance applications. In: IEEE International Conference on Image Processing. Institute of Electrical and Electronics Engineers, pp. 949–953.

Oldroyd, D.R., 1986. Charles Darwin's theory of evolution: A review of our present understanding. Biol. Philos. 1, 133–168.

Olhoeft, G.R., 2000. Maximizing the information return from ground penetrating radar. In: Journal of Applied Geophysics. Elsevier, pp. 175–187.

Oshana, R., 2006. Overview of Digital Signal Processing Algorithms. In: DSP Software Development Techniques for Embedded and Real-Time Systems. Newnes, pp. 59–121.

Ozana, S., Pies, M., Docekal, T., 2016. Case studies on optimization problems in MATLAB and COMSOL multiphysics by means of the livelink. In: AIP Conference Proceedings. AIP Publishing LLCAIP Publishing, p. 370006.

Özdemir, C., Demirci, Ş., Yiğit, E., Yilmaz, B., 2014. A review on migration methods in b-scan ground penetrating radar imaging. Math. Probl. Eng. 2014.

Pal, M., Dixon, N., Flint, J., 2010. Detecting & locating leaks in water distribution polyethylene pipes. In: WCE 2010 - World Congress on Engineering 2010. London, pp. 889–894.

Pan, Q., Liu, D.-J., Guo, Z.-Y., Fang, H.-F., Feng, M.-Q., 2016. Magnetic anomaly inversion using magnetic dipole reconstruction based on the pipeline section segmentation method. J.

Geophys. Eng. 13, 242–258.

Park, Sehwan, Kim, J., Jeong, S., Park, Seunghee, 2019. GPR Data-Based Computer Vision for the Detection of Material Buried Underground. In: Proceedings - 2019 3rd International Conference on Smart Grid and Smart Cities, ICSGSC 2019. Institute of Electrical and Electronics Engineers Inc., pp. 41–44.

Pasolli, E., Melgani, F., Donelli, M., 2009. Automatic analysis of GPR images: A pattern-recognition approach. IEEE Trans. Geosci. Remote Sens. 47, 2206–2217.

Pašteka, R., Kušnirák, D., Wilken, D., Putiška, R., Papčo, J., Godová, D., Zvara, I., Nogová, E., Ondrášová, L., 2019. Effective combination of microgravimetry and geoelectrical methods in the detection of subsurface cavities in archaeological prospection - Selected case-studies from Slovakia. Contrib. to Geophys. Geod. 49, 479–496.

Pasternak, M., Kaczmarek, P., 2019. Continuous wave ground penetrating radars: state of the art. In: SPIE 11055, XII Conference on Reconnaissance and Electronic Warfare Systems. SPIE, p. 10.

Pearson, J.B., Stear, E.B., 1974. Kalman Filter Applications in Airborne Radar Tracking. IEEE Trans. Aerosp. Electron. Syst. AES-10, 319–329.

Pennock, S.R., Abedt, T.M., Curionit, G., Chapmant, D.N., John, U.E., Jenks, R.J., 2014. Investigation of Soil Contamination by Iron Pipe Corrosion and its Influence on GPR Detection 381–386.

Percival, D.B., Walden, A.T., 2000a. Wavelet Methods for Time Series Analysis, Cambridge University Press. Cambridge University Press.

Percival, D.B., Walden, A.T., 2000b. Wavelet methods for time series analysis, Cambridge University Press. Cambridge University Press.

Perrone, A., Lapenna, V., Piscitelli, S., 2014. Electrical resistivity tomography technique for landslide investigation: A review. Earth-Science Rev.

Persico, R., Soldovieri, F., Utsi, E., 2010. Microwave tomography for processing of GPR data at Ballachulish. J. Geophys. Eng. 7, 164–173.

Petersen, M.L., 2001. Use of Electromagnetic Induction Tools in Salinity Assessment/Appraisals in Eastern Colorado ERP1–ERP1.

Peyton, A.J., 2015. Electromagnetic induction tomography. In: Industrial Tomography: Systems and Applications. Woodhead Publishing, pp. 61–107.

Pidlisecky, A., Knight, R., 2008. FW2_5D: A MATLAB 2.5-D electrical resistivity modeling code. Comput. Geosci. 34, 1645–1654.

Pieraccini, M., 2018. Noise Performance Comparison Between Continuous Wave and Stroboscopic Pulse Ground Penetrating Radar. IEEE Geosci. Remote Sens. Lett. 15, 222–226.

Pilu, M., Fitzgibbon, A.W., Fisher, R.B., 2002. Ellipse-specific direct least-square fitting. In: Proceedings of 3rd IEEE International Conference on Image Processing. IEEE, pp. 599–602.

Plati, C., Loizos, A., Gkyrtis, K., 2020. Integration of non-destructive testing methods to assess asphalt pavement thickness. NDT E Int. 115, 102292.

Prego, F.J., Solla, M., Puente, I., Arias, P., 2017. Efficient GPR data acquisition to detect underground pipes. NDT E Int. 91, 22–31.

Pringle, J.K., Holland, C., Szkornik, K., Harrison, M., 2012. Establishing forensic search methodologies and geophysical surveying for the detection of clandestine graves in coastal beach environments. Forensic Sci. Int. 219, e29–e36.

Pringle, J.K., Stimpson, I.G., Wisniewski, K.D., Heaton, V., Davenward, B., Mirosch, N., Spencer, F., Jervis, J.R., 2020. Geophysical monitoring of simulated homicide burials for forensic investigations. Sci. Rep. 10, 1–13.

Pringle, J.K., Stimpson, I.G., Wisniewski, K.D., Heaton, V., Davenward, B., Mirosch, N., Spencer, F., Jervis, J.R., 2021. Geophysical monitoring of simulated clandestine burials of murder victims to aid forensic investigators. Geol. Today 37, 63–65.

Professional Surveyors Canada, 2019. CSA Standard Review – S250 Mapping of Underground Utility Infrastructure Professional Surveyors Canada.

Qiu, H., Qiu, M., Lu, Z., Gerard, M., 2019. An efficient key distribution system for data fusion in V2X heterogeneous networks. Inf. Fusion 50, 212–220.

Rader, C.M., Mailing, G.C., 1967. What Is the Fast Fourier Transform? Proc. IEEE 55, 1664–1674.

Radiodetection, 2018. The theory of buried cable and pipe location. Bristol.

Rajiv, K., Chandra, G.R., Rao, B.B., 2017. GPR Objects Hyperbola Region Feature Extraction.

Rao, K.R., Kim, D.N., Hwang, J.-J., 2010. Fast Fourier Transform - Algorithms and Applications. Signals and Communication Technology.

Rasol, M., Pérez-Gracia, V., Assunção, S.S., 2018. Analysis and Calibration of Ground Penetrating Radar Shielded Antennas. In: 2018 17th International Conference on Ground Penetrating Radar, GPR 2018. Institute of Electrical and Electronics Engineers Inc.

Ratliff, A., 2003. An Overview of Current and Developing Technologies for Pipe Condition Assessment. In: Proceedings of the ASCE International Conference on Pipeline Engineering and Construction: New Pipeline Technologies, Security, and Safety. American Society of Civil Engineers, pp. 103–114.

Reynolds, J.M., 2011. An Introduction to Applied and Environmental Geophysics, 2nd ed, European Space Agency, (Special Publication) ESA SP. Wiley.

Rezaei, N., Kordabadi, H., Elkamel, A., Jahanmiri, A., 2009. An optimal extended Kalman filter designed by genetic algorithms. Chem. Eng. Commun. 196, 602–615.

Rhee, J.-Y., Park, K.-T., Cho, J.-W., Lee, S.-Y., 2021. A Study of the Application and the Limitations of GPR Investigation on Underground Survey of the Korean Expressways. Remote Sens. 2021, Vol. 13, Page 1805 13, 1805.

Riley, C., 2017. Dynamic Evaluation of Transportation Structures with iPod-Based Data

Acquisition. Transp. Res. Educ. Center(TREC).

Riley, C., 2018. Teaching Modal Analysis with Mobile Devices. Am. Soc. Eng. Educ.

Robbes, D., 2006. Highly sensitive magnetometers-a review. Sensors Actuators, A Phys. 129, 86–93.

Rodrigues, D., Barraca, N., Costa, A., Borges, J., Almeida, F., Fernandes, L., Moura, R., Madureira-Carvalho, Á., 2021. Multi-technique detection of buried inert explosive devices in urban context: Metal detection, magnetometer and ground-penetrating radar. In: Symposium on the Application of Geophysics to Engineering and Environmental Problems. Society of Exploration Geophysicists, pp. 218–219.

Rogers, C.D.F., Hao, T., Costello, S.B., Burrow, M.P.N., Metje, N., Chapman, D.N., Parker, J., Armitage, R.J., Anspach, J.H., Muggleton, J.M., Foo, K.Y., Wang, P., Pennock, S.R., Atkins, P., Swingler, S.G., Cohn, A.G., Goddard, K., Lewin, P.L., Orlando, G., Redfern, M.A., Royal, A.C.D., Saul, A.J., 2012. Condition assessment of the surface and buried infrastructure - A proposal for integration. Tunn. Undergr. Sp. Technol. 28, 202–211.

Rogers, C.D.F., Overton, C., 2018. A brief history of the Underworld. Assess. Underworld An Integr. Perform. Model City Infrastructures.

Rosin, P.L., 2001. Unimodal thresholding. Pattern Recognit. 34, 2083–2096.

Roubal, M., 2018. Non-destructive testing of ground & structural integrity of underground assets. In: ISRM International Symposium 2000, IS 2000. OnePetro.

Ryden, N., Mooney, M.A., 2009. Analysis of surface waves from the light weight deflectometer. Soil Dyn. Earthq. Eng. 29, 1134–1142.

Saey, T., De Smedt, P., Delefortrie, S., Van De Vijver, E., Van Meirvenne, M., 2015a. Comparing one- and two-dimensional EMI conductivity inverse modeling procedures for characterizing a two-layered soil. Geoderma 241–242, 12–23.

Saey, T., Van Meirvenne, M., De Smedt, P., Stichelbaut, B., Delefortrie, S., Baldwin, E., Gaffney, V., 2015b. Combining EMI and GPR for non-invasive soil sensing at the stonehenge world heritage site: The reconstruction of a WW1 practice trench. Eur. J. Soil Sci. 66, 166–178.

Sagnard, F., Tarel, J.P., 2016. Template-matching based detection of hyperbolas in ground-penetrating radargrams for buried utilities. J. Geophys. Eng. 13, 491–504.

SanFilipo, B., 2000. Final report on passive and active low-frequency electromagnetic spectroscopy for airborne detection of underground facilities.

Santos, F.A.M., Triantafilis, J., Bruzgulis, K.E., Roe, J.A.E., 2010. Inversion of Multiconfiguration Electromagnetic (DUALEM-421) Profiling Data Using a One-Dimensional Laterally Constrained Algorithm. Vadose Zo. J. 9, 117.

Särkkä, S., 2013. Bayesian filtering and smoothing, 1st ed, Cambridge University Press. Cambridge University Press.

Sasaki, Y., 1989. Two-dimensional joint inversion of magnetotelluric and dipole data. Soc. Explor. Geophys. 54, 55–57.

Sasaki, Y., 2000. Full 3-D inversion of electromagnetic data on PC. J. Appl. Geophys. 46, 45–54.

Sbartaï, Z.M., Breysse, D., Larget, M., Balayssac, J.P., 2012. Combining NDT techniques for improved evaluation of concrete properties. Cem. Concr. Compos. 34, 725–733.

Schlinger, C.M., 1990. Magnetometer and gradiometer surveys for detection of underground storage tanks. Bull. - Assoc. Eng. Geol. 27, 37–50.

Schmidt, A., Chapman, D., Rogers, C., 2006. Physiochemcial changes in London Clay adjacent to cast iron pipes. Proc. Int. Assoc. Eng. Geol. Environ. Conf. 2006 12.

Schmitt, M., Zhu, X.X., 2016. Data Fusion and Remote Sensing: An ever-growing relationship. IEEE Geosci. Remote Sens. Mag. 4, 6–23.

Scott, W.R., Kim, K., Larson, G.D., Gurbuz, A.C., McClellan, J.H., 2004. Combined seismic, radar, and induction sensor for landmine detection. J. Acoust. Soc. Am. 123, 3042–3042.

Sen, M.K., Mallick, S., 2018. Genetic Algorithm with Applications in Geophysics. Springer, Cham, pp. 487–533.

Şenyiğgit, E., Düğenci, M., Aydin, M.E., Zeydan, M., 2013. Heuristic-based neural networks for stochastic dynamic lot sizing problem. Appl. Soft Comput. J. 13, 1332–1339.

Shakmak, B., Al-Habaibeh, A., 2015. Detection of Water Leakage in Buried Pipes Using Infrared Technology.

Shihab, S., Al-Nuaimy, W., 2005. Radius estimation for cylindrical objects detected by ground penetrating radar. Subsurf. Sens. Technol. Appl. 6, 151–166.

Shihab, S., Al-Nuaimy, W., Huang, Y., Eriksen, A., 2002. Neural network target identifier based on statistical features of GPR signals. In: Koppenjan, S., Lee, H. (Eds.), Ninth International Conference on Ground Penetrating Radar. International Society for Optics and Photonics, pp. 135–138.

Simon, D., 2006. Optimal State Estimation: Kalman, H∞, and Nonlinear Approaches, Optimal State Estimation: Kalman, H∞, and Nonlinear Approaches.

Siu, K.L., Lai, W.W.L., 2019. A lab study of coupling effects of electromagnetic induction on underground utilities. J. Appl. Geophys. 164, 26–39.

Skartados, E., Kargakos, A., Tsiogas, E., Kostavelis, I., Giakoumis, D., Tzovaras, D., 2019. GPR antenna localization based on a-scans. Eur. Signal Process. Conf. 2019-Septe, 1–5.

Smith, R.E., Anderson, D.T., Ball, J.E., Zare, A., Alvey, B., 2017. Aggregation of Choquet integrals in GPR and EMI for handheld platform-based explosive hazard detection. In: Detection and Sensing of Mines, Explosive Objects, and Obscured Targets XXII. p. 1018217.

Smitha, N., Singh, V., 2016. Clutter reduction techniques of ground penetrating radar for detecting subsurface explosive objects. 2016 Int. Conf. Inf. Commun. Embed. Syst. ICICES 2016 1–8.

Solimene, R., Cuccaro, A., Dell'Aversano, A., Catapano, I., Soldovieri, F., 2014. Ground clutter removal in GPR surveys. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 7, 792–798.

Soltanova, D., Baranov, P., Baranova, V., Chudinova, A., 2015. Simulating magnetic field of a ferromagnetic pipe underwater in COMSOL Multiphysics. IOP Conf. Ser. Mater. Sci. Eng. 93, 012023.

Stein, E.M., Shakarchi, R., 2011. Fourier analysis: An introduction, Fourier Analysis: An Introduction.

Steinway, W.J., Duvoisin, H.A., Tomassi, M.S., Thomas, J.E., Betts, G., Morris, C., Kahn, B., Stern, P., Krywick, S., Johnson, K., Dennis, K., Simoneaux, W., Blood, B., 1998. Multi-sensor system for mine detection. In: IGARSS '98. Sensing and Managing the Environment. 1998 IEEE International Geoscience and Remote Sensing. Symposium Proceedings. (Cat. No.98CH36174). pp. 228–230 vol.1.

Stern, W., 1929. Versuch einer elektrodynamischen Dickemessung von Gletschereis. Beiträge zur Geophys. 23, 292–333.

Stocco, S., Godio, A., Sambuelli, L., 2009. Modelling and compact inversion of magnetic data: A Matlab code. Comput. Geosci. 35, 2111–2118.

Strange, A.D., Yelf, R.J., 2012. What is the true time range of a GPR system? In: 2012 14th International Conference on Ground Penetrating Radar, GPR 2012. IEEE Computer Society, pp. 436–440.

Sudevan, V., Shukla, A., Karki, H., 2019. Hierarchical Controller for Autonomous Tracking of Buried Oil and Gas Pipelines and Geotagging of Buried Pipeline Structure. In: Proceedings of the IEEE 2019 9th International Conference on Cybernetics and Intelligent Systems and Robotics, Automation and Mechatronics, CIS and RAM 2019. Institute of Electrical and Electronics Engineers Inc., pp. 338–343.

SUEZ, 2021. Water leak detection using helium - SUEZ in UK [WWW Document]. URL https://www.suez.co.uk/en-gb/our-offering/businesses/what-are-you-looking-for/water-management/water-network-management/leak-detection/water-leak-detection-using-helium (accessed 12.30.21).

Sumathi, S., Hamsapriya, T., Surekha, P., 2008. Evolutionary Intelligence: An Introduction to Theory nd Applications with Matlab, Control. Springer.

Sun, H., Zhu, J., 2020. Nondestructive evaluation of steel-concrete composite structure using high-frequency ultrasonic guided wave. Ultrasonics 103, 106096.

Szymczyk, M., Szymczyk, P., 2013. Preprocessing of GPR data. ipc 18, 83–90.

Szymczyk, P., Tomecka-Suchoń, S., Szymczyk, M., 2015. Neural networks as a tool for georadar data processing. Int. J. Appl. Math. Comput. Sci. 25, 955–960.

Szyrowski, T., 2016. An intelligent, fast-acquisition remote sensing system for locating and measuring burial of subsea power and telecommunication cables.

Tabbagh, A., Panissod, C., 2000. 1D complete calculation for electrostatic soundings interpretation. Geophys. Prospect. 48, 511–520.

Terrasse, G., Nicolas, J.M., Trouve, E., Drouet, E., 2015. Application of the curvelet transform for pipe detection in GPR images. In: International Geoscience and Remote Sensing

Symposium (IGARSS). pp. 4308–4311.

Tesfamariam, T.G., 2013a. Signal Processing Techniques for Landmine Detection Using Impulse Ground Penetrating Radar (ImGPR). IEEE Sens. J.

Tesfamariam, T.G., 2013b. Signal Processing Techniques for Landmine Detection Using Impulse Ground Penetrating Radar ( ImGPR ). Technischen Universitaet Darmstadt.

Tikhonov, A.N., 1950. Determination of the electrical characteristics of the deep strata of the earth's crust: Doklady Akadamia Nauk, 73, 295–297. dak DAKNEQ 0869-5652 Dokl. Akad. Nauk.

Tlemçani, M.K., Ait Brahim, L., El Mahsani, A., Labrihmi, A., 2018. Contribution of geophysical methods to the detection of underground cavities in salt series. Development in different projects in Morocco (roads, rail, etc…). MATEC Web Conf. 149, 02083.

Torbaghan, M.E., Li, W., Metje, N., Michael Burrow, Chapman, D.N., Rogers, C.D.F., 2020. Automated detection of cracks in roads using ground penetrating radar. J. Appl. Geophys. 104118.

Törn, A., Žilinskas, A., 1989. Global Optimization, Lecture Notes in Computer Science. Springer Berlin Heidelberg, Berlin, Heidelberg.

Torrione, P., Collins, L., 2006. Ground response tracking for improved landmine detection in ground penetrating radar data. In: International Geoscience and Remote Sensing Symposium (IGARSS). pp. 153–156.

Travassos, X.L., Avila, S.L., Ida, N., 2018. Artificial Neural Networks and Machine Learning techniques applied to Ground Penetrating Radar: A review.

Triantafilis, J., Santos, F.A.M., 2011. Hydrostratigraphic analysis of the Darling River valley (Australia) using electromagnetic induction data and a spatially constrained algorithm for quasi-three-dimensional electrical conductivity imaging. Hydrogeol. J. 19, 1053–1063.

Tso, C.H.M., Kuras, O., Wilkinson, P.B., Uhlemann, S., Chambers, J.E., Meldrum, P.I., Graham, J., Sherlock, E.F., Binley, A., 2017. Improved characterisation and modelling of measurement errors in electrical resistivity tomography (ERT) surveys. J. Appl. Geophys. 146, 103–119.

Tzanis, A., 2017. A versatile tuneable curvelet-like directional filter with application to fracture detection in two-dimensional GPR data. Signal Processing 132, 243–260.

Tzanis, A., Kafetsis, G., 2004. A freeware package for the analysis and interpretation of common-offset ground probing radar data, based on general purpose computing engines. Bull. Geol. Soc. Greece 36, 1347.

USAG, 2019. 2017-2018 Utility Strike Damages Report [WWW Document]. URL https://www.utilitystrikeavoidancegroup.org/strike-damages-report.html (accessed 3.28.21).

USAG, 2020. 2019 Utility Strike Damages Report [WWW Document]. URL https://www.utilitystrikeavoidancegroup.org/strike-damages-report.html (accessed 3.28.21).

Utsi, E.C., 2017a. Ground penetrating radar: Theory and practice, Ground Penetrating Radar: Theory and Practice. Elsevier.

Utsi, E.C., 2017b. Ground Penetrating Radar-Theory and Practice, Journal of Chemical Information and Modeling. Joe Hayton.

Van de Vegte, J., 2002. Crash Course In Processing Digital Signal Processing. In: Fundamentals of Digital Signal Processing. Prentice-Hall International, pp. 1–28.

Van Drongelen, W., 2010. Signal Processing for Neuroscientists, A Companion Volume, Signal Processing for Neuroscientists, A Companion Volume.

Van Drongelen, W., 2018. Wavelet Analysis - an overview | ScienceDirect Topics [WWW Document]. Signal Process. Neurosci. (second Ed. URL https://www.sciencedirect.com/topics/earth-and-planetary-sciences/wavelet-analysis (accessed 1.21.21).

Van Kempen, L., Sahli, H., 2001. Signal processing techniques for clutter parameters estimation and clutter removal in GPR data for landmine detection. In: IEEE Workshop on Statistical Signal Processing Proceedings. pp. 158–161.

Van Merwe, A. Der, Gupta, I.J., 2000. A Novel signal processing technique for clutter reduction in gpr measurements of small, shallow land mines. IEEE Trans. Geosci. Remote Sens. 38, 2627–2637.

Van Vorst, D.G., Yedlin, M.J., Virieux, J., Krebes, E.S., 2014. Three-dimensional to two-dimensional data conversion for electromagnetic wave propagation using an acoustic transfer function: Application to cross-hole GPR data. Geophys. J. Int. 198, 474–483.

Venkatasubramanian, V., Leung, H., 2006. An EM-IMM based abrupt change detector for landmine detection. Detect. Remediat. Technol. Mines Minelike Targets XI 6217, 62172Z.

Viezzoli, A., Christiansen, A.V., Auken, E., Sørensen, K., 2008. Quasi-3D modeling of airborne TEM data by spatially constrained inversion. Geophysics 73.

VIVAX, M., 2011. VM-480B User Handbook.

Vo, C.K., Staples, S.G.H., Cowell, D.M.J., Varcoe, B.T.H., Freear, S., 2020. Determining the Depth and Location of Buried Pipeline by Magnetometer Survey. J. Pipeline Syst. Eng. Pract. 11, 04020001.

Wahab, W.A., Jaafar, J., Yassin, I.M., Ibrahim, M.R., 2013. Interpretation of Ground Penetrating Radar (GPR) image for detecting and estimating buried pipes and cables. In: Proceedings - 2013 IEEE International Conference on Control System, Computing and Engineering, ICCSCE 2013. pp. 361–364.

Wai-Lok Lai, W., Dérobert, X., Annan, P., 2018. A review of Ground Penetrating Radar application in civil engineering: A 30-year journey from Locating and Testing to Imaging and Diagnosis. NDT E Int. 96, 58–78.

Wai-Lok Lai, W., Kind, T., Kruschwitz, S., Wöstmann, J., Wiggenhauser, H., 2014. Spectral absorption of spatial and temporal ground penetrating radar signals by water in construction materials. NDT E Int. 67, 55–63.

Waite, J.W., Wellaratna, R., 2006. Centerline and depth locating method for non-metallic buried utility lines.

Waite, J.W., Wellaratna, R., 2010. Sensor fusion for model-based detection in pipe and cable locator systems.

Wang, D., 2017. Application and Evaluation of Non-Destructive Testing Methods for Buried Pipes 80–117.

Wang, J., Su, Y., 2013. Fast detection of GPR objects with cross correlation and hough transform. Prog. Electromagn. Res. C 38, 229–239.

Wang, M.L., Lynch, J.P., Sohn, H., 2014. Sensor Technologies for Civil Infrastructures, Sensor Technologies for Civil Infrastructures.

Wang, P., Goddard, K., Lewin, P., Swingler, S., 2011. Electromagnetic field application to underground power cable detection. In: XVII International Symposium on High Voltage Engineering. ISH, Hannover.

Wang, P., Lewin, P., Goddard, K., Swingler, S., 2010. Design and testing of an induction coil for measuring the magnetic fields of underground power cables. In: Conference Record of IEEE International Symposium on Electrical Insulation.

Wang, S., Han, L., Gong, X., Zhang, S., Huang, X., Zhang, P., 2021. Full-waveform inversion of time-lapse crosshole gpr data using markov chain monte carlo method. Remote Sens. 13.

Warren, C., Giannopoulos, A., Giannakis, I., 2016. gprMax: Open source software to simulate electromagnetic wave propagation for Ground Penetrating Radar. j 209.

Weik, M.H., 2006. constitutive relation. In: Computer Science and Communications Dictionary. Springer US, pp. 291–291.

Weisenseel, R.A., Karl, W.C., Castanon, D.A., Miller, E.L., Rappaport, C.M., DiMarzio, C.A., 1999. Statistical fusion of GPR and EMI data. In: Dubey, A.C., Harvey, J.F., Broach, J.T., Dugan, R.E. (Eds.), Detection and Remediation Technologies for Mines and Minelike Targets IV. SPIE, p. 1179.

Welch, G., Bishop, G., 1995. An Introduction to the Kalman Filter. In Pract. 7, 1–16.

Willner, D., Chang, C., Dunn, K., 2008. Kalman filter algorithms for a multi-sensor system. Institute of Electrical and Electronics Engineers (IEEE), pp. 570–574.

Wilson, A.D., Baietto, M., 2009. Applications and advances in electronic-nose technologies. Sensors.

Wirahadikusumah, R., Abraham, D.M., Iseley, T., Prasanth, R.K., 1998. Assessment technologies for sewer system rehabilitation. Autom. Constr. 7, 259–270.

Witten, A.J., 2014. Handbook of Geophysics and Archaeology, Taylor & Francis. Taylor & Francis.

Wolfe, W., Zissis, G., 1985. The Infrared Handbook, The Infrared Handbook, Revised Edition. SPIE.

Wong, P.T.W., Lai, W.W.L., Sato, M., 2016. Time-frequency spectral analysis of step frequency continuous wave and impulse ground penetrating radar. In: Proceedings of 2016 16th International Conference of Ground Penetrating Radar, GPR 2016. Institute of Electrical and Electronics Engineers Inc.

Wu, R., Clement, A., Li, J., Larsson, E.G., Bradley, M., Habersat, J., Maksymonko, G., 2001. Adaptive ground bounce removal. Electron. Lett. 37, 1250–1252.

Xie, X., Qin, H., Yu, C., Liu, L., 2013a. An automatic recognition algorithm for GPR images of RC structure voids. J. Appl. Geophys. 99, 125–134.

Xie, X., Zeng, C., Wang, Z., 2013b. GPR signal enhancement using band-pass and K-L filtering: A case study for the evaluation of grout in a shielded tunnel. J. Geophys. Eng. 10, 1–10.

Xu, X., Gao, W., Zhang, D., Wang, Y., 2018a. Research on the Difference Detection Method Based on GPR Data. 2018 17th Int. Conf. Gr. Penetrating Radar, GPR 2018 1–4.

Xu, X., Li, J., Ma, Z., 2018b. A Multi - Frequency GPR Data Fusion Method for Underground Disease Detection. In: 2018 17th International Conference on Ground Penetrating Radar, GPR 2018.

Xu, X., Li, J., Qiao, X., Fang, G., 2019. Fusion of multiple time-domain GPR datasets of different center frequencies. Near Surf. Geophys. 17, 141–150.

Yang, J., Wang, S., Na, X., Yang, Y., 2019. A weighted data fusion method in distributed multi-sensors measurement and control system. In: Proceedings - 2019 6th International Conference on Information Science and Control Engineering, ICISCE 2019. Institute of Electrical and Electronics Engineers Inc., pp. 654–658.

Yang, X.S., 2014. Nature-Inspired Optimization Algorithms, Nature-Inspired Optimization Algorithms. Elsevier Inc.

Yang, Y.H., Liu, J.Y., 2013. Quantitative study of music listening behavior in a social and affective context. IEEE Trans. Multimed. 15, 1304–1315.

Yelf, R., 2004. Where is true time zero? In: Proceedings of the Tenth International Conference Ground Penetrating Radar, GPR 2004. pp. 279–282.

Yogi, I.B.S., Widodo, 2017. Time domain electromagnetic 1D inversion using genetic algorithm and particle swarm optimization. AIP Conf. Proc. 1861, 030014.

Youn, H.-S., Chen, C.-C., 2002a. Automatic GPR target detection and clutter reduction using neural network. In: Ninth International Conference on Ground Penetrating Radar. SPIE, pp. 579–582.

Youn, H.-S., Chen, C.-C., 2002b. Automatic GPR target detection and clutter reduction using neural network. In: Koppenjan, S., Lee, H. (Eds.), Ninth International Conference on Ground Penetrating Radar. SPIE, pp. 579–582.

Zadhoush, H., 2021. Numerical modelling of ground penetrating radar for optimization of the time-zero adjustment and complex refractive index model. The University of Edinburgh.

Zaitseva, K.A., Korolev, V.I., Akhi, A. V., Akhi, A.A., 2021. Adaptation of dolphins' (Tursiops truncatus) location signals when searching for and identifying objects hidden by sea

sediments. Bioacoustics.

Zare, E., Li, N., Khongnawang, T., Farzamian, M., Triantafilis, J., 2020. Identifying potential leakage zones in an irrigation supply channel by mapping soil properties using electromagnetic induction, inversion modelling and a support vector machine. Soil Syst. 4, 1–18.

Zhang, A., Atia, M.M., 2020. An efficient tuning framework for Kalman filter parameter optimization using design of experiments and genetic algorithms. Navig. J. Inst. Navig. 67, 775–793.

Zhang, W.Y., Hao, T., Chang, Y., Zhao, Y.H., 2017. Time-frequency analysis of enhanced GPR detection of RF tagged buried plastic pipes. NDT E Int. 92, 88–96.

Zhang, X., Xu, W., Zhou, B., 2009. Mean first-passage time in a bistable system driven by multiplicative and additive colored noises with colored cross-correlation. Commun. Nonlinear Sci. Numer. Simul. 14, 4220–4225.

Zhang, X., Xue, F., Wang, Z., Wen, J., Guan, C., Wang, F., Han, L., Ying, N., 2021. A novel method of hyperbola recognition in ground penetrating radar (GPR) B-Scan image for tree roots detection. Forests 12, 1019.

Zhang, Y., Collins, L., Yu, H., Baum, C.E., Carin, L., 2003. Sensing of unexploded ordnance with magnetometer and induction data: Theory and signal processing. IEEE Trans. Geosci. Remote Sens. 41, 1005–1015.

Zhao, S., Al-Qadi, I., 2017. Pavement drainage pipe condition assessment by GPR image reconstruction using FDTD modeling. j 154.

Zheng, J., Peng, S. ping, Yang, F., 2014. A novel edge detection for buried target extraction after SVD-2D wavelet processing. J. Appl. Geophys. 106, 106–113.

Zheng, X., Liu, S., 2011. The design of broadband electromagnetic method system with GPS simultaneous localization. In: Lecture Notes in Electrical Engineering. Springer, Berlin, Heidelberg, pp. 611–617.

Zhou, X., Chen, H., Hao, T., 2019. Efficient detection of buried plastic pipes by combining GPR and electric field methods. IEEE Trans. Geosci. Remote Sens. 57, 3967–3979.

Zhou, X., Chen, H., Li, J., 2018. An Automatic GPR B-Scan Image Interpreting Model. IEEE Trans. Geosci. Remote Sens. 56, 3398–3412.

Zhu, J., Xue, Y., Zhang, N., Li, Z., Tao, Y., Qiu, D., 2017. A noise reduction method for Ground Penetrating Radar signal based on wavelet transform and application in tunnel lining. In: IOP Conference Series: Earth and Environmental Science. Institute of Physics Publishing, p. 12088.

Zou, L., Wang, Y., Giannakis, I., Tosti, F., Alani, A.M., Sato, M., 2020. Mapping and assessment of tree roots using ground penetrating radar with low-cost GPS. Remote Sens. 12, 1300.

Zoubir, A.M., Chant, I.J., Brown, C.L., Barkat, B., Abeynayake, C., 2002a. Signal processing techniques for landmine detection using impulse ground penetrating radar. IEEE Sens. J. 2, 41–51.

Zoubir, A.M., Chant, I.J., Brown, C.L., Barkat, B., Abeynayake, C., 2002b. Signal processing techniques for landmine detection using impulse ground penetrating radar. IEEE Sens. J. 2, 41–51.

Zoubir, A.M., Iskander, D.R., Chant, I.J., Carevic, D., 1999. Detection of landmines using ground-penetrating radar. In: Detection and Remediation Technologies for Mines and Minelike Targets IV. SPIE, p. 1301.

Zwinkels, J., 2015. Light, Electromagnetic Spectrum. In: Encyclopedia of Color Science and Technology. pp. 1–8.

# Appendix I LITERATURE REVIEW

## 2.1 Background and Theory of Electromagnetic

Electromagnetic (EM) based methods are common techniques in use for locating buried service utilities such as shallow pipes and also the main focus of this research. Among all the existing geophysical methods, the design and implementation of remote electromagnetic based methods have attracted the interest of many researchers and is rising rapidly (Cheng and Pamukcu, 2017). Employing electromagnetic induction-based methods as nonintrusive methods for subsurface explorations goes back to the nineteenth century after Michel Faraday's findings in this field (Radiodetection, 2018). Detection of buried objects using electromagnetics is a feasible approach by emitting and propagating energy using electromagnetic waves into the ground and then sampling and analysing the reflected signals (Reynolds, 2011). It is worth noting that EM-based methods suffer from the issue of penetration depth due to the variations in the parameters such as the emitted frequencies or the electromagnetic properties of the ground (Everett, 2013; Milsom and Eriksen, 2011; Witten, 2014).

The subsequent sections look at in more detail at the underlying principles of EM techniques.

### 2.1.1 Active and passive EM detection methods

Existing electromagnetic based methods for subsurface explorations are broadly divided into two main categories of passive and active modes (Reynolds, 2011). Surveys in passive mode are those in which naturally generated magnetic fields are associated with natural or fabricated sources. For instance, the Earth's magnetic field is a natural source, while a magnetic field generated by radio signals, transmitted by a radio antenna, is a fabricated source. Nevertheless, such methods in passive mode are not capable of measuring the level

of the induced energy into the ground and hence the intensity of passively generated magnetic fields. This is due to the lack of control over the transmitting source (Everett, 2013).

Active methods are those by which signals, emitted into the ground, are generated in a controlled manner by applying a specific level of energy. Furthermore, depending on the employed method, the responses of the ground are either reflected signals that are sampled by a receiver (e.g., GPR) or in the form of a measurable secondary field, generated by a buried anomaly, which will be superimposed on a primary field generated by a transmitter (Milsom and Eriksen, 2011). The level of reflected energy for both reflected signals and for the secondary fields is proportional to the soil's electromagnetic properties (Huray, 2010).

### 2.1.2 Electromagnetic Fields and Maxwell's equations

Maxwell's equations shown in Table 2-2 describe how electromagnetic field components (electric field $E$ and magnetic fields $B$) are generated, propagate, and interact in a time-domain system within a medium. Maxwell's equations are a set of four fundamental equations (laws) unified by James Clerk Maxwell (Huray, 2010). Gauss' law for the electric field (Eq. 2-1) states that the existence of electric flux $(D)$ is always due to the existence of electric charges $(v)$. Furthermore, the volume charge density $(\rho_v)$ is the total enclosed charge $(Q)$ distributed throughout the volume. Faraday's law of induction (Eq. 2-2) states that an electric field $(E)$ can simply be produced by changing magnetic fields $(B)$. The Amper's circuital law (Eq. 2-3) on the other hand, indicates that a magnetic field can be generated by passing an electric current $(I)$ through a close loop and the intensity of the generated magnetic field $(H)$ is proportional to the density of the electric flux or displacement $(D)$ plus the electric current density $(J)$. Finally, the Gauss' law for magnetic field (Eq. 2-4) states that due to absence of magnetic charge, magnetic fields have no beginning and end point (Huray, 2010). Parameter $\overline{\nabla}$ is the 'curl operator', which is the second derivative of the associated parameter.

Table 2-2. Maxwell's Equations (Huray, 2010)

| Differential Form | Integral Form | Governing Law | |
|---|---|---|---|
| $\overline{\nabla}\times\overline{\boldsymbol{D}}=\boldsymbol{\rho_v}$ | $\boldsymbol{Q(t)=\oiint_s \vec{D}\,(t).ds=\iiint_v \rho_v(t)dv}$ | Gauss' law for electric field | (2-1) |
| $\overline{\nabla}\times\overline{\boldsymbol{E}}=-\frac{\partial\overline{\boldsymbol{B}}}{\partial t}$ | $\boldsymbol{V=\oint_L \vec{E}\,(t).dl=-\iint_s\left[\frac{\partial\overline{B}(t)}{\partial t}\right].ds}$ | Faraday's Law of induction | (2-2) |
| $\overline{\nabla}\times\overline{\boldsymbol{H}}=\bar{J}+\frac{\partial\overline{D}}{\partial t}$ | $\boldsymbol{I_{(t)}=\oint_L \vec{H}\,(t).dl=\iint_s\left[\bar{J}(t)+\frac{\partial\overline{D}(t)}{\partial t}\right].ds}$ | Ampere's circuital law | (2-3) |
| $\overline{\nabla}\times\overline{\boldsymbol{B}}=0$ | $\boldsymbol{\oiint \vec{B}(t)\,.ds=0}$ | Gauss' law for magnetic field | (2-4) |

The material response and interrelationship to the induced magnetic field or electromagnetic waves, based on their electrical dielectric and magnetic properties (i.e., dielectric permittivity $(\varepsilon)$, dielectric conductivity $(\sigma)$, and magnetic permeability $(\mu)$), in conjunction with Maxwell's equations, are described by a set of linear equations called constitutive equations (Eq. 2-5 to Eq. 2-7) represented in Table 2-3 (Weik, 2006).

Table 2-3. Constitutive Equations (Weik, 2006)

| | |
|---|---|
| $\boldsymbol{J=\sigma.E}$ | (2-5) |
| $\boldsymbol{D=\varepsilon.E}$ | (2-6) |
| $\boldsymbol{B=\mu.H}$ | (2-7) |

### 2.1.3 The electromagnetic properties of a material

The ground is composed of different layers and constituents, which have specific electrical properties. In this section, the key electrical dielectric and magnetic properties of ground that affect the emitted EM-forces into the ground are defined. These electrical properties are a subset of properties including dielectric permittivity $(\varepsilon)$, dielectric conductivity $(\sigma)$, and magnetic permeability $(\mu)$ which determine the electrical susceptibility of materials.

#### *2.1.3.1 Electrical dielectric permittivity* $(\varepsilon)$

Permittivity, also known as "dielectric constant", is the degree of electrical polarisation of material or the capability of storing and releasing electromagnetic energy under the influence of an external electric field. Permittivity can be seen as a primary factor for the reflection of EM waves in soil composition. As soon as a propagating electromagnetic wave in the ground encounters any contrast in dielectric permittivity (this phenomenon is called impedance), part of the signal energy will be reflected. The amount of reflected energy is proportional to the degree of encountered contrast (impedance) (Koganti et al., 2020). The higher the difference between permittivity values of adjacent layers the stronger the impedance and signal reflection (Daniels, 2004).

The velocity of propagating waves in the material is also a function of its dielectric permittivity in its simplest form. The wave velocity $(v)$ of a material can be determined as shown in Eq. 2-8 as an inverse proportion of light velocity in air $(C_0)$ to the square root of its dielectric permittivity $(\varepsilon)$ (Goodman and Piro, 2013b):

$$v = \frac{C_0}{\sqrt{\varepsilon}} \qquad (2\text{-}8)$$

#### *2.1.3.2 Electrical dielectric conductivity* ($\sigma$)

The conductivity is the degree of capability in the movement of electric charges (conduct of electric charges) through the material in the presence of an electric field. The conductivity of the material defines the value of signal attenuation of a propagating wave in the host medium. In general, the higher the conductivity, the higher the EM signal attenuation and consequently, the penetration depth of the GPR signal will be decreased (Cassidy, 2009b). For instance, the propagation of GPR waves in the ground is strongly dependent on the electrical conductivity of soil at each survey location. High soil conductivity yields high attenuation of the radar signal in the soil which can highly restrict the performance of the radar and maximum possible penetration depth of waves. As stated by Wai-Lok Lai et al. (2018), the groundwater content acts as a dominant factor for variation of conductivity of materials. In the same manner, the existence of mineral components such as salt (e.g., seawater in soil) increases both the conductivity and polarisation factor associated with the dielectric permittivity of the material (Daniels, 2004).

#### *2.1.3.3 Electrical magnetic permeability* ($\mu$)

Magnetic permeability is the capability of the material to form a magnetic field under the influence of an externally induced magnetic field (Baker et al., 2007). As stated by Cassidy (2009b), the effect of the magnetic permeability on EM-signals is in most cases negligible unless the medium is a magnetic mixture (contains ferromagnetic materials such as magnetic minerals).

## 2.2 Nonintrusive electromagnetic methodologies

### 2.2.1 Electrical Resistivity Tomography (ERT)

Electrical Resistivity Tomography (ERT) is an established non-destructive subsurface exploration and electric current based method that exploits the EM properties of the ground and materials in the ground, including buried utilities, to reconstruct a cross-section (CS) image of the subsurface. In contrast to transient (time-varying) methods (e.g., EML), the ERT relies on a static measurement, i.e. the measured responses will be identical at the measurement points independent of when those measurements were taken (Dey and Morrison, 1979; Witten, 2014). Moreover, ERT is a technique that uses electric current with reversing directions but at intervals of one or two seconds, which can be considered as a direct current (DC) (Milsom and Eriksen, 2011).

In the ERT method, the electric potentials (voltages) are measured between pairs of electrodes, installed in arrays, and attached to the ground using a high-impedance receiver (voltmeter). Due to the homogeneity of the ground with a constant resistivity, the measured voltages between electrode pairs (in arrays) can be interpreted as apparent resistivity and inverted as an electrical resistivity map (CS image) of the subsurface (electrical resistivity imaging (ERI)) (Greggio et al., 2018; Kemna et al., 2002; Lesparre et al., 2019; Tso et al., 2017). The ERI correlates the distributed apparent resistivity of the ground with the measured voltage at electrode pairs by utilising the system parameters, i.e. as electrode separation, electrode array configuration, and electrical current (Perrone et al., 2014). The penetration depth of currents increases as the electrode spacing increases. Thus, the resolution of the data can be simply specified to satisfy the objectives of a survey (Reynolds, 2011). It is worth noting that, although the electrodes must be attached (inserted) into the ground and this contradicts with the definition of a nonintrusive method, ERT is categorised as a nonintrusive method due

to the dependency of the survey depth on inter-electrode spacing (Kemna et al., 2002). In addition, inserting the electrode pairs in the ground is a time-consuming task that can be considered as a disadvantage of the ERT, but the survey can entirely be performed by one person (cost-effective) and the data acquisition is fully controlled by the software, which are considerable advantages (Kumar, 2012).

### 2.2.2 Inversion of EMC data for exploring the deep subsurface

Interpretation of EMC data relies on inversion procedures which can be subject to the non-uniqueness issue of the results. To create a less complex inversion problem to solve, the real three-dimensional (3D) subsurface is usually converted into one- or two-dimensional problems (Madden and Swift, 1965). In general, when dealing with electromagnetic problems, this is common to consider the problems as a two-dimensional problem if a target buried in the subsurface remains steady large compared to skin depth, i.e. depth by which the amplitude of the propagated wave decreases to 0.37 of its initial amplitude within a large distance (Ku, 1976; Middleton and Valkenburg, 2002; Peyton, 2015). As an inherent fact, it is important to understand that inversions of geophysical problems might not converge if the parameters of the algorithm were poorly selected and initialised. Such divergence issues can be identified along the inversion process if the algorithm generates unreasonable outcomes that cannot exist in the real world. For instance, generating negative values for conductivity or a negative thickness for soil layers (Ku, 1976; McNeil, 1980). The inversion of EMC data goes back to the mid-1960s, and often focused on magnetotelluric (MT) method, i.e. based on measurements of the horizontal components of both the electric and magnetic fluctuations (Cagniard, 1953; Fournier, 1966; Lytle and Dines, 1980; Madden and Swift, 1965; Tikhonov,

1950). However, an accurate conclusion and interpretation were not possible due to the lack of proper data.

Later, Sasaki (1989) presented a 2D joint inversion algorithm in which MT data and diploe-dipole resistivity data were employed simultaneously to reduce the uncertainty in the 2D interpretation of associated data. Sasaki's algorithm was based on the finite element method (FEM), discussed by Chaskalovic (2008), in correlation with a nonlinear smoothness constraint least-square adopted approach discussed by Lytle and Dines (1980). Compared to previous work in the literature, in which all receiver stations were computed in a single mesh with many nodes, the conductivity response in Sasaki's work was computed independently using individual discretised meshes for each receiver. Regardless of both the receiver location and the induced frequency, the number of nodes was kept fixed instead of changing it with a variation in frequency. As a result, the issue of non-uniqueness of the results was reduced by stabilising the results and considering the relationship between both neighbouring and non-neighbouring ground mesh cells. This relationship is determined by utilising the roughness and Jacobian matrices under the assumption of constant conductivity in each cell. The schematic of such a cell relationship is presented in Figure 2, showing how the effect of changes in characteristics of each cell can be translated into the neighbouring and non-neighbouring mesh cells for both 2D and 3D cases.

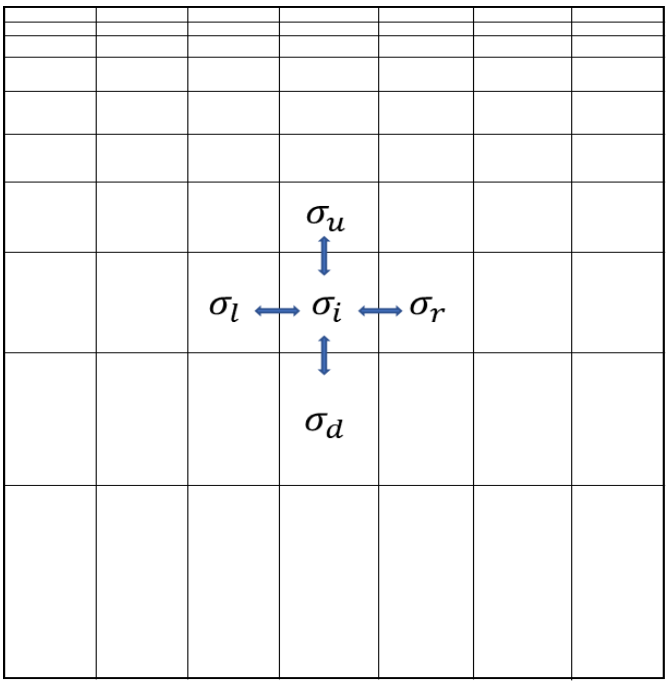



Figure 2. Correlation of neighbouring and non-neighbouring cells in a quadrilateral ground mesh

The results associated with field data presented by the author significantly reduced the non-uniqueness issue compared to previous work in the literature, although, the level of uncertainty remained relatively high. The outcome of the algorithm was limited to one field dataset. To increase the level of confidence, it is necessary to perform the method on a larger number of datasets from both controlled and real-world surveys. Furthermore, due to the lack of independent data obtained through boreholes or geological data, the uncertainty in the results remains an issue. The author emphasised a concern of the similarity between the presented results and reality as trial-and-error modelling was an inevitable part of the process. Nevertheless, despite all aforementioned drawbacks, the inversion method proposed by Sasaki (1989), remains one of the fundamental methods which have been continuously updated.

Various implemented methodologies for the inversion of conductivity data proposed by various researchers in recent years are presented in Figure 3. A list of associated references is presented in Table 2-5.

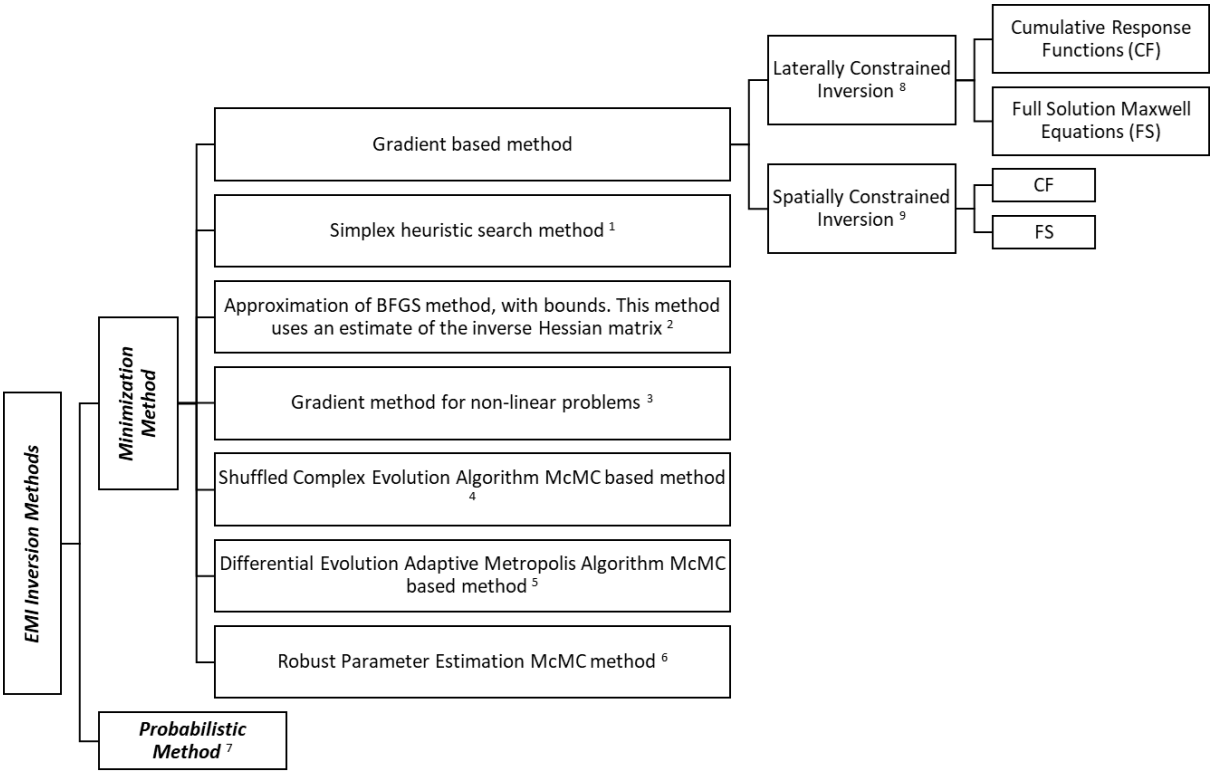



Figure 3. Inversion methodologies for inverting EMC data

Table 2-5. List of references for the Figure2-11

| | | | | | |
|---|---|---|---|---|---|
| 1 | (Nelder and Mead, 1965) | 2 | (Byrd et al., 1995) | 3 | (Fletcher and Reeves, 1964) |
| 4 | (Duan et al., 1994) | 5 | (Vrugt and Ter Braak, 2011) | 6 | (Bardossy et 381 al., 2008) |
| 7 | (Moghadas, 2019) | 8 | (Farzamian et al., 2019) | 9 | (Monteiro Santos et al., 2011) |

### 2.2.3 Infrared Thermography (IRT)

IRT is a well-known non-invasive tool in use for different research applications from sustainability applications and energy efficiency of buildings (Al-Habaibeh et al., 2010) to medical purposes (Hildebrandt et al., 2010). Infrared waves are part of the invisible spectral region of the EM spectrum, which includes a full range of fluctuating electrical and EM waves passing through space (Herter, 2019; Zwinkels, 2015). Invisible infrared waves compromise thermal energy, which can be transformed into visible images using the IRT (Meola and Carlomagno, 2004). IRT was utilised by Carreño-Alvarado et al., (2014) in the infrastructure sector, using infrared (IR) cameras, to investigate the possibility of locating buried services such as pipes in urban areas by taking images during cold seasons of the year. The authors stated that the visibility of buried pipes under real-world conditions depends on the contrast between the outside temperature and the buried pipes. Such contrasts, and correspondingly the visibility of pipes, can significantly vary between day and night times or through different seasons of the year. It is easier to locate a pipe during cold days, with a high temperature contrast between the ambient conditions and the buried pipe, compared to an almost zero contrast during a spring day with moderate temperature for both the environment and the pipes. In a similar approach, El Afi and Belattar,(2014) conducted numerical simulations for locating pipes for different materials (PVC, copper and iron), different diameters (16 to 36 mm), and at different shallow depths (50 to 70 mm), in concrete structures. Authors have concluded a better detectability of metallic pipes, and those buried closer to the surface (without an in-depth discussion).

In a step further, Fahmy and Moselhi, (Fahmy and Moselhi, 2017) and Shakmak and Al-Habaibeh, (2015) and Atef et al., (2016) implemented thermal infrared cameras in independent research to detect and locate leakage in water mains beneath the ground

surface. Shakmak and Al-Habaibeh, (2015) investigated the water leakage from water pipes buried in dry sand in relatively dry countries by conducting experiments under controlled laboratory conditions. In constrast, Fahmy and Moselhi, (Fahmy and Moselhi, 2017) investigated water main leakage beneath asphalt roads in different urban areas of Montreal in Canada over a two year period in different weather conditions. 42 buried water mains (150 to 200mm diameter) of which 25 experienced leakages, were scanned using an IR camera mounted at different heights (1.20 -12 m) on a moving vehicle (5-20 km/hr). Atef et al., (2016), carried out field experiments for detecting and locating water leakage over two years in Doha, Qatar. An IR camera was utilised as a complementary technique to identify leaks after locating the pipes using GPR. The outcome of these studies indicated that water leaks can successfully be located using IR cameras with a margin of error of 2.9% to 5.6%, which indicates a good rate if thermal contrast between surface and pipe is substantial. The findings of both studies are in good agreement with Carreño-Alvarado et al., (2014). Researchers have shown that infrared thermography is able to detect buried pipes under certain circumstances. However, its limitations with respect to the seasons and strong dependency on the ambient temperature (relative to the temperature of the pipe), soil type, and road construction can cause considerable limitations in utilising this method.

It is worth noting that although several publications have focussed on employing IR cameras for the detection and location of buried pipes, to the best knowledge of the author, there is still clear gaps in knowledge in this field. A gap is more clearly defining the conditions under which the technique is likely to work. Another gap is a lack of a robust fusion algorithm in the literature in which IRT data is objectively processed and fused with other geophysical methods for the detection of subsurface assets.

### 2.2.4 Magnetometry

Detection of buried ferromagnetic objects by measurement of the Earth's total magnetic field anomalies (distortions) goes back to almost a century ago (March, 1953). The applied principle behind magnetometry is to measure any fluctuation (superposition) of the Earth's magnetic field (MF) caused by ferromagnetic objects, i.e. objects made of material with high electric conductivity and high magnetic permeability (e.g., metallic pipes and cables) (Xie et al., 2013a). The Earth's magnetic field, at the survey location, is a constant value in the near-surface measurements and is considered as an external MF. Other types of external MFs are those generated by ferromagnetic objects above the surface no matter how close or how far they are from the measurement sensor. Distinguishing between external MFs and those generated by the target object is the key principle under which magnetometry works (Everett, 2013; Robbes, 2006; Schlinger, 1990). Different types of equipment are available for magnetic measurements. Magnetometer is the general term associated with equipment used to measure MFs and there are several techniques available: The *Fluxgate magnetometer* measures the MF caused by magnetic flux leakage (MFL); *Atomic Proton Precession Magnetometers (PPM)* measure the magnitude of the MF vector known as the total field (Denisov et al., 2006); *Atomic Alkali Vapor Magnetometers,* also referred to as *cesium magnetometers,* work in a similar way as PPMs but are not dependent on the induced electric current in a coil, rather they exploit the variations in optical transparency of alkali vapours by utilising a certain type of atoms, namely caesium (Robbes, 2006), and finally, *Gradiometers,* which consist of two vertically placed magnetometers with a small separation both of which measure the magnetic field at a particular location and the contribution of measured external fields can be eliminated by the gradient measurements (Witten, 2014).

Magnetometry is a popular technique for detecting remaining unexploded ordnance (UXO) in ex-conflict or war regions around the globe. There are various methodologies in magnetometry for detecting ferromagnetic objects such as UXOs and metallic pipes. These include, using signal processing techniques for detecting UXO (Zhang et al., 2003), defining a library of ordnance as a comparison point for the classification of the UXO (Billings, 2004), employing a multi-sensing approach to enhance the survey result of UXO detection using a metal detector, magnetometer and GPR (Rodrigues et al., 2021). Numerical modelling using an FEM approach was used by Churchill et al. (2012), who suggested creating more realistic but complex numerical bodies compared to the typically simple shape models used in the literature for a significant increase (ca. 50%) in the dipole moment response of the UXO model. FEM was also employed for the detection of ferrous buried pipes by Guo et al. (2015) who investigated the anomalies in shape, amplitude, and magnetic width due to changing model parameters (e.g., diameter, depth, thickness, azimuth, the inclination of the pipe and the geomagnetic intensity). A linear change in the magnetic field was reported for variations in the geomagnetic intensity, thickness, and diameter. Whereas changes in other parameters led to non-linear changes. Another conclusion of this report was that thin and slender pipes are not good candidates for the magnetometry technique as they could not be detected using this approach. However, it seems that this conclusion cannot be generalised. Similarly, Pan et al. (2016) investigated magnetic anomalies which are influence by the model parameters and the applied section segmentation. The outcome was very similar, yet the computational cost was reduced. In another piece of research, Soltanova et al. (2015) investigated the influence of pipe diameter on the Earth's magnetic field. The outcome matched previous reports and the authors concluded that increasing pipe diameter increases the magnetic field anomalies. The outcome of this research can be exploited as a good start point. However, to generalise such

conclusions, further investigations need to be conducted for more complex and realistic scenarios.

Critical shortcomings of the magnetometry technique as a standalone NDT survey were discussed by Holt (2014). This pointed out that the magnetometer cannot detect the ferrous material directly, rather the anomalies in the Earth's magnetic field. Accordingly, distinguishing if the buried object is a pipe, cable or any other ferrous-based mineral is impossible. Moreover, because of the omnidirectional measurement property of the magnetometry technique, there is no need for the buried utility to be directly below the magnetometer to be detected. Another issue associated with the detection of magnetic anomalies is determining the burial depth. Feng et al. (2017) tackled this issue and introduced a new segmentation approach for the detection of magnetic anomalies caused by ferromagnetic pipes. The authors suggested employing the magnetic dipole reconstruction (MDR) technique, i.e. segmentation of the pipe, which leads to cutting the pipe into smaller dipoles and at the same time increasing the stability of the analysis and reducing the calculation time. Nevertheless, this strategy needs to be validated by changing model parameters as it was already shown by Guo et al. (2015), Pan et al. (2016) and Guo et al. (2016) that variations in model parameters has a significant impact on the results.

## 2.5 Quantum technology based gravitational methodologies

Gravity is a force of reciprocal attraction between two masses that can be utilised to detect and locate buried objects in shallow depth. Gravitational based methods, known as microgravity surveys, are passive methodologies for exploring the subsurface. They were initially used to identify large geological anomalies by identifying weaker spots in the measured gravity fields (Blizkovsky, 1979; Tlemçani et al., 2018). Nowadays, the evolution of technology over the last few decades offers new opportunities in identifying near-surface

hollow spaces (e.g., tunnels, cavities, sinkholes, large graves) (Pašteka et al., 2019). The quantum technology based gravity sensors aim to illustrate the density distribution of the underground space of the surveyed area (Bongs et al., 2016). In addition, the principle under which these sensors can fulfil this aim is to identify the smallest variations in a measured gravity field (in micro-scale) that are proportional to the scale of density variation caused by mass anomalies in the ground (Pašteka et al., 2019). As their name suggests, atomic quantum technology indicates a high accuracy in gravity measurements, yet such accurate sensors are susceptible to both internal and external noises generated from either the instrument itself or environmental sources, respectively (Boddice et al., 2018). Noise vulnerability and challenges in the measurement of weak gravity fields along with slow and time-consuming surveys are the limiting factors that can affect the applicability of quantum technology gravity sensors.

Reviewing the literature indicates that, there is a lack of a robust fusion algorithm in which an autonomous (or with a low level of expert interaction) objective data processing is conducted for both gravity and magnetometry methods and utilise powerful mathematical estimation tools such as Kalman filtering or particle filtering to fuse/integrate processed data with other commonly used methods such as GPR.

## 2.6 Methods Based on the Mechanical Properties of Materials

The key principle under which geophysical methods are acting is identifying tangible contrasts (inhomogeneities) in the physical properties of associated material involved in surveyed systems (Milsom and Eriksen, 2011). Methods associated with the motion sensors and acoustic both focus on the mechanical properties of materials.

### 2.6.1 Vibro-acoustic

Vibro-acoustic for instance, as a method for locating buried pipes works via listening and recording the acoustic response of propagated waves which are generated by impulse or vibration. The main concern associated with these techniques which are emphasised by authors is that the accuracy of the results reduces considerably when used for disturbed ground (e.g. soils in urban areas) or influenced by external vibration sources (Gao et al., 2005; Kalkowski et al., 2018; Muggleton et al., 2012, 2011; Muggleton and Papandreou, 2014; Muggleton and Rustighi, 2013).

### 2.6.2 Motion sensors – Geophones and Accelerometers

*Geophones*

The geophone is a sensor to monitor and invert the velocity into the voltage and is utilised in a wide range of applications (Iskander, 2018). One of the well-known nonintrusive tools that utilise geophones, for the geotechnical monitoring purposes, are falling weight deflectometer (FWD) and lightweight deflectometer (LWD). Both tools are commonly used to evaluate the structural condition of subgrades of pavements and railways tracks to predict their stiffness using back-calculation of the layer's elastic modulus (Fleming et al., 2008; Nayana et al., 2021; Nega et al., 2016; Ryden and Mooney, 2009). Both systems typically consist of a loading system (controlled dropping weights) and an array of geophones to measure the velocity of surface displacements, at different points, after applying the load (Marecos et al., 2017; Plati et al., 2020).

Although there are many works available in the literature that have utilised these tools in correlation with various structural parameters of the subgrade structures, yet their potential and capabilities in terms of subsurface asset detection and location are not fully utilised.

*Accelerometers*

Rapid growth and improvement of smartphones along with simplicity in operation, attracted many researchers' attention to use smartphone's built-in sensors for in-situ data collection for different scientific activities. For instance, investigation of the Earth's seismic activities in a research project called iShake developed by Dashti et al., (2011), Ervasti et al.,(2011), and Dashti et al.,(2014). Smartphones were used to record and collect the Earth's seismic activities around the world in a simple manner and without being an expert. A mobile application made it possible for any individual in possession of a smartphone to record the Earth's seismic activities. The sampled data addressed with smartphone location (using GPS) was subsequently sent to a data hub for further processing and analysis. Smartphones were also used to study structural dynamics by Riley, (2017) in experimental research to detect a self-made crack on a timber cantilever, and frames made of concrete and steel in a controlled laboratory environment. Subsequently, Riley (2018) implemented his approach on a three-span pedestrian walkway. Both experiments conducted by Riley were based on acceleration measurement, by which iPod® devices were used with a maximum frequency-sampling rate of 100Hz to determine the seismic response (frequency modal shapes). In a similar approach, Feng et al., (Feng et al., 2015) analysed the accuracy of smartphone's accelerometer sensors for monitoring the health of structures under seismic vibrations. Unlike Riley, (2017) this research was not limited to laboratory experiments. Rather, laboratory experiments were conducted under controlled seismic and white noise excitations. Furthermore, a field test was also conducted on a concrete pedestrian bridge in Princeton (NJ, USA). Vibration frequencies using smartphones (sensitivity rates between 1 to 18 mg/digit), indicated a satisfactory average error of 1% to 5% compared to measured frequencies by industrial reference sensor (piezoelectric 393B04) with a significantly higher sensitivity rate (1000 mV/g).

As shown here, smartphones with all integrated sensors can be utilised as data collecting, and a processing tool using a suitably designed application (App). However, knowledge gap can simply be identified in the literature, so that such potential is utilised (despite the limitations) for the detection of buried assents in a combination with other available methods such as above mentioned LWD.

## 2.7 Biologically Inspired Methods (Scents and Gas-based Methods)

Another, but still not very common, technique for detecting and condition assessment of buried target and service utilities is to employ biologically inspired sensors such as electronic noses to detect exposed scents and/or gases. There are various biologically inspired methods in use to fulfil different aims of the detection and condition assessment field. For instance, employing trained animals such as dogs (GICHD, 2003) and rats with highly sensitive biological sensors, i.e. noses for sniffing out the UXOs (Hadagali and Suan, 2017), using genetically engineered biosystems such as growing genetically engineered colour changing plants in contaminated areas due to leakage of landmines with time and exposing explosive material into the surrounding ground (Habib, 2007), using trained dolphins to locate underwater mines using their bio sonar system (Zaitseva et al., 2021).

The limitation of many nonintrusive methods in their sensitivity level along with the complexity of operations is always a motivation for researchers to develop innovative bio-inspired methods to overcome such issues (Wilson and Baietto, 2009). Bio-inspired electronic sensors such as artificial noses are developed to fulfil various aims. As such the detection of UXOs and utilities as well as the condition assessment of buried utilities (e.g., water mains). Electronic noses mounted on a radio-controlled airship was used by Goschnick and Harms (2002), to locate and classify objects with gas exposure characteristic (e.g., landmines). This method does not offer the accuracy for detection of single buried objects, rather it can be

used to investigate over terrain distributed objects. Anindya et al.,(2014) introduced a tracking robot to investigate the condition of buried utilities using an electronic nose as a gas leakage detector using which presence of the gas (other than clean air) was detected, yet due to lack of desired sensitivity, a desirable result was achieved in one-minute intervals after each measurement, and at a maximum distance of 5 cm between the sensor and the ground surface. Using the same concept, electronic noses are utilised in the UK (SUEZ, 2021) to localise the leakage in water mains beneath the road surface. Non-hazardous gas such as helium (not toxic, inert, non-flammable, and non-condensable) is injected into the running system so that it can be detected where it diffuses through the road surface. The main concern associated with this method is the true location of the leakage as the helium will travel through a path with lowest resistance and not necessarily where the actual leak exists.

Although biological systems can be highly sensitive in different ways, e.g. sensitivity to various scents, yet this property has proven to be very challenging, if not impossible, to be replicated. As shown above, despite advancements in technologies, there is still a gap in the development of highly effective artificial noses close to the biological versions, in terms of scent sensitivity, and analysis time. Furthermore, there are only very few publications with a focus on sniffing sensors for detection and locating buried assets. Sniffing sensors were mostly employed to detect UXOs and leakages in pipe networks as a standalone method in practical applications. To the best knowledge of the author, there is also a lack of fusion algorithms in the literature in which data collected from sniffing sensors were objectively processed and fused with other geophysical methods, such as GPR.

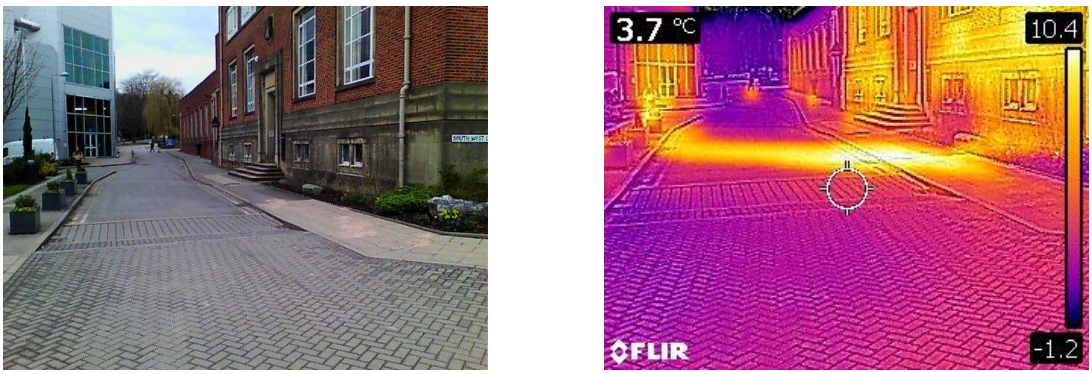
# Appendix II THE JOURNEY OF THIS RESEARCH PROJECT

The pursuit of a PhD degree is a challenging and rigorous endeavour that requires significant dedication and effort. Along with the general challenges of time management, balancing various tasks and responsibilities, and dealing with uncertainty and self-doubt, specific challenges have been encountered during the course of the research on the detection, location and condition assessment of buried utilities such as pipes. In order to investigate the identified gaps in knowledge through literature review and to define the precise scope of this research, the researcher has undertaken a multifaceted approach that includes both experimental tests and numerical modelling. The research aims to address identified gaps in knowledge by investigating various techniques including the utilisation of phone accelerometer sensors in combination with falling weight deflectometry, infrared thermography and artificial noses, either as standalone techniques or through the fusion of these techniques. However, limitations such as low sampling rate of accelerometer, low signal-to-noise ratio, high cost and non-readiness of equipment have presented difficulties. The current research endeavours to find innovative solutions to enhance the accuracy and efficiency of buried utility detection and assessment through developing new methods, improving existing ones and finding new ways to combine them.

This section presents a brief overview of the journey of this research and the path that was taken concerning the gaps identified in the literature review, and before the proposed computational solutions and associated methodologies in this thesis could be established.

## Infrared thermography (IRT)

The review on IRT (Appendix I) showed there is a gap in the literature for utilising IRT in conjunction with other techniques (e.g., GPR). Generally, objects with a temperature above -273.15 degrees of Celsius radiate infrared radiation, i.e. part of the electromagnetic spectrum that is invisible to human eyes due to its longer wavelengths compared to the visible part of the EM spectrum, but this radiation can be detected (Wolfe and Zissis, 1985). Infrared sensors (e.g. IRT cameras) can invert the emitted infrared radiation to determine the temperature of the object. This property can potentially be exploited to identify, and potentially determine the condition assessment of, buried pipes in the shallow subsurface. On this basis, research was conducted to develop an innovative algorithm for the fusion of data (images) captured by IRT cameras (e.g. Figure 1 taken using a FLIR C2 thermal camera) with other commonly used techniques in subsurface exploration (e.g. GPR). Various experiments under different conditions were conducted in association with numerical modelling of the experimentally obtained data and making use of COMSOL geophysics which is a software to create a numerical model by utilising the physics of the system and finite element method (FEM) for the back-calculation of the size and depth of the buried pipes.




*(a)*




*(b)*

Figure 1. Infrared images taken using a FLIR C2 thermal camera , a) warm water heating duct with an approximate depth of 0.75 m at the University of Birmingham, c) sewer (typical diameter between 225 mm to 2100 mm) beneath a road with an approximate depth of 2 m from invert of the pipe.

Initial numerical analysis of simplified scenarios under the following controlled idealised conditions was carried out. Water main pipe (50 mm MDPE) buried 750 mm beneath the surface (pipe centreline) in dry sand carrying water at approximately 17°C and an ambient temperature of 20°C. On the one hand, the outcome of the initial analysis in terms of detectability of both buried pipe, and potential damages in surrounding soil due to water leakage were promising (Figure 2).

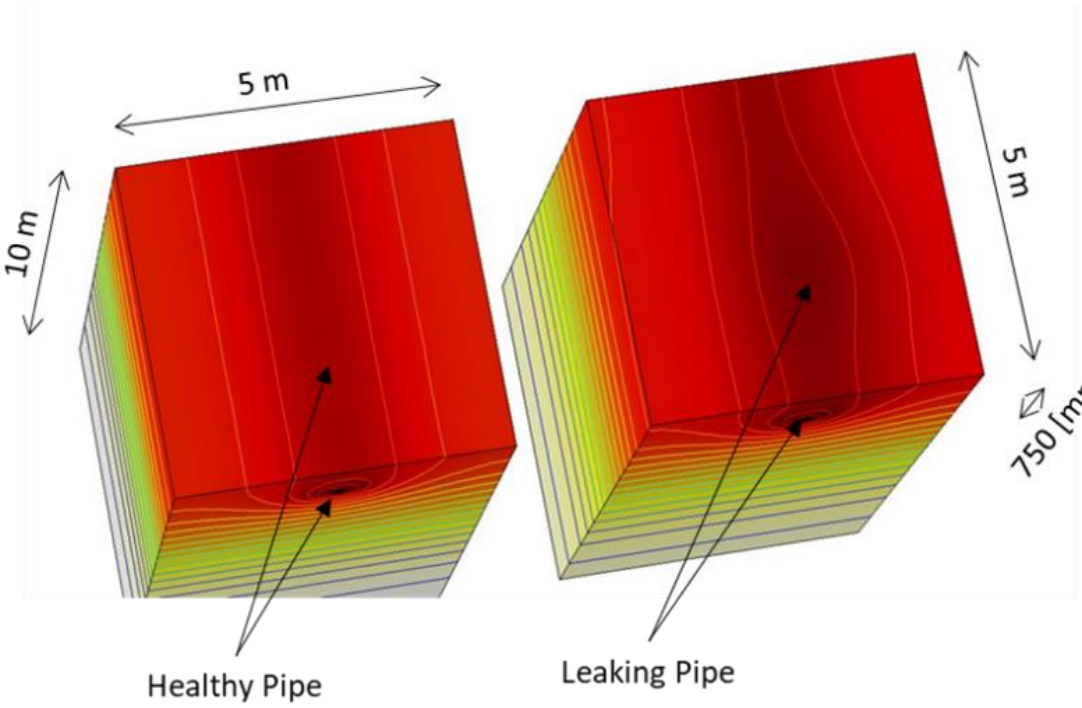



Figure 2. Numerical model using COMSOL showing the detectability of both a healthy and a leaking buried water main in sandy soil under laboratory conditions, a) healthy pipe, b) leaking pipe with thermal expansion in the surrounding soil

On the other hand, the numerical analysis revealed a very slow temperature propagation from the buried pipe to the surface, i.e. on average 50 - 100 hrs for a 1 °C temperature change depending on the temperature of the water in the pipe. Such a slow temperature propagation can cause limitations in performing surveys if there is no significant long term temperature difference (ca. 20°C) between the pipe and the surrounding environment. Such limitations can be the necessity of longer waiting times for a survey to reach a noticeable temperature on the surface (higher cost), the limitation in detecting certain types of pipes (or cables) with or without content, but with significant temperature difference compared to its surrounding environment. On this basis, it was decided to search for a more suitable technique in terms of the practicality and the cost of the associated survey.

The initial idea behind this PhD research project was utilising built-in sensors in smartphones such as accelerometers, and magnetometers to locate buried infrastructure utilities (e.g., pipe and cables), and underground anomalies such as air voids. In fact, the mechanical response of associated objects and materials in ground compositions needed to be measured after seismic loadings. The principle behind this idea was to apply an impact loading (seismic loading) to the ground surface and record the ground response using geophones, i.e., measuring the acceleration and displacement of the ground surface. Established techniques such as the falling weight deflectometer (FWD) (Nega et al., 2016) or the smaller version, the light weight deflectometer (LWD)(Fleming et al., 2008), in which the response of the ground surface are commonly in use to estimate the stiffness of the subsurface layers with the possibility of detecting subsurface anomalies such as voids using this principle.

## Use of data from built-in sensors in smartphones

Rapid growth and improvement of smartphones along with simplicity in operation, attracted many researcher's attention to use smartphone's built-in sensors for in-situ data collection for different scientific activities. For instance, investigation of the Earth's seismic activities in a research project called iShake developed by Dashti et al., (2011), Ervasti et al.,(2011), and Dashti et al.,(2014) . Smartphones were used to record and collect the Earth's seismic activities around the world in a simple manner and without being an expert. A mobile application made it possible for any individual in possess of a smartphone to record the Earth's seismic activities. The sampled data addressed with smartphone location (using GPS) was subsequently send to a data-hub for further processing and analysis. Smartphone were also used to study structural dynamics by Riley, (2017) in an experimental research to detect a self-made crack on a timber cantilever, and frames made of concrete and steel in a controlled laboratory environment.

Subsequently, Riley (2018) implemented his approach on a three-span pedestrian walkway. Both experiments conducted by Riley were based on acceleration measurement, by which iPod devices were used with a maximum frequency-sampling rate of 100Hz to determine the seismic response (frequency modal shapes). In a similar approach, Feng et al., (Feng et al., 2015) analysed the accuracy of smartphone's accelerometer sensors for monitoring the health of structures under seismic vibrations. Unlike the Riley, (2017) this research was not limited to laboratory experiments. Rather, laboratory experiments were conducted under controlled seismic and white noise excitations. Furthermore, a field test was also conducted on a concrete pedestrian bridge in Princeton (NJ, USA). Vibration frequencies using smartphones (sensitivity rates between 1 to 18 mg/digit), indicated a satisfactory average error of 1% to 5% compared to measured frequencies by industrial reference sensor (piezoelectric 393B04) with a significantly higher sensitivity rate (1000 mV/g).

It is evident from aforementioned works that smartphones are capable to measure the vibration frequencies of various open structures under seismic vibrations. However, to achieve initial aims of this PhD research project, the capability of smartphones in measuring vibration frequencies of buried structures such as buried infrastructure under dynamic impact loadings needed to be tested. Implementing the above-mentioned approaches on a buried infrastructure like pipes even under controlled laboratory condition would be still challenging and entirely different. Compared to an open structure (steel frame), a buried infrastructure is surrounded by soil with diverse physical or mechanical properties (e.g., density, elasticity, water content and porosity etc.) that control the behaviour of that buried infrastructure. Hence Various set of LWD tests were undertaken to investigate the health condition of a buried concrete pipe (30 cm diameter) embedded in dry sand under different conditions (healthy, void, crack). Tests were carried out using iPhone 5 (frequency sampling rate of

100Hz) and Industrial SM-24 geophone (frequency sampling rate of 240 Hz). Data collected from both iPhone and SM-24 geophone were analysed and compared. The results indicated discontinuities and missing readings at different points of measured data using smartphone close to the position of impact loading. It was found out that the limited frequency sampling rate (100 Hz) of the smartphone yielded to incapability in measuring the displacements during the bouncing effect after each impact loading. The incapability of to date smartphone built-in sensors, with low sampling rate, in dealing with seismic impact loadings changed the direction of this research project to fusion or integration of data collected from different NDT techniques.

## Smartphone sensors and light-weight deflectometer (LWD)

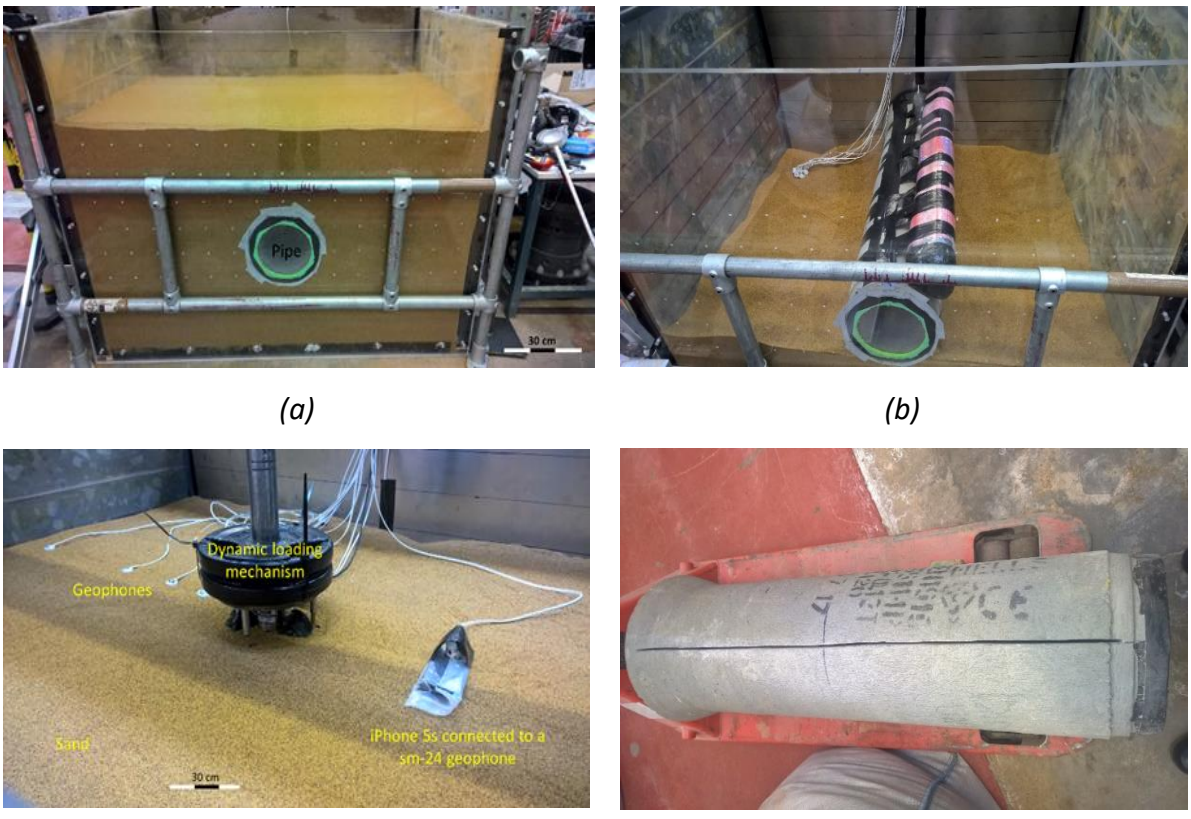
The review of vibroacoustic methods in Appendix I revealed the potential of using integrated sensors, such as accelerometers in smartphones, in correlation with techniques that utilise motion sensors, such as the Lightweight Deflectometer (LWD) and the Falling Weight Deflectometer (FWD), for detecting and locating buried utilities. These techniques are conventional methods for investigating the performance and condition of road structures, such as the road surface or the interlock performance between adjacent layers (load transfer efficiency "LTE") (Schmalzer, 2006). The principle of the LWD and FWD techniques is to simulate the deflection of the surface due to moving trucks on the road (Schmalzer, 2006).

The expected shape of the surface deflection is a bowl-shaped pattern that is formed by generating loading impulses by dropping a weight from a certain height onto a loading plate placed on the road surface. Additionally, it was demonstrated in the literature review that smartphones have the potential to function as portable geophysics laboratories, due to their various integrated sensors (e.g. accelerometers). Various tests were conducted using different smartphones (e.g. iPhone 5s with integrated Bosch sensortec BMA220 triaxial acceleration

sensor with a frequency range of 32Hz to 1kHz) under controlled laboratory conditions to locate and assess the condition of a buried concrete pipe buried in dry Leighton Buzzard sand (Figure 3). In addition, industrial geophones (SM-24 with 240 Hz) were used simultaneously as a performance benchmark. The goal was to check the applicability of smartphones in conjunction with the LWD technique for locating and determining the condition of buried utilities, such as a pipe, by analysing the generated bowl-shaped deflection pattern of the surrounding soil. Extracting anomalies from the usually symmetric bowl shapes can be an indicator of the existence of potential buried objects such as pipes, their possible orientation, or defects such as broken utilities or voids.

A series of LWD tests (5 kg, 10 kg mass and falling height of 0.75 m) were undertaken by using National Instrument hardware (data logger NI 9218 and NI 9228, Column Load cell F218 and USB cDAQ NI-9174) to investigate the condition of a buried concrete pipe (30 cm diameter) embedded in dry sand (Leighton Buzzard Fine Sharp Sand) under different conditions (healthy, void, crack). Sensors used for this experiment were iPhone 5 and Industrial SM-24 geophone (Figure 3c). Tests were carried out in the following conditions:

- Fully intact concrete pipe surrounded by a single layer of dry sand (Figure 3a).
- Demonstrating void by placing an air-void (using swim noodles to simulate a void with a 30 cm) next to the pipe buried in a single layer of dry sand (Figure 3b).
- Single Longitudinal split crack without void by cutting the concrete pipe along the long axis on the crown (Figure 3d).
- Double Longitudinal split crack without void (pipe was halved along the long axis).




*(a)* *(b)*

*(c)* *(d)*

Figure 3. Smartphone experiment setup in a, a) 30 cm concrete pipe buried 30 cm beneath the surface in a 1.21 m x 1.54 m x 1.2 m (d x w x h) steel sand box with plexiglass front panel, b) iPhone 5s connected to an sm-24 geophone (placed at different distances similar to the positioned geophone) and dynamic loading mechanism (10 Kg weight dropped from different heights to adjust the load) above the pipe centre line., (c) Void demonstration, (d) Longitudinal split crack on the pipe crown

For all the aforementioned scenarios (except the double longitudinal scenario which is still ongoing), weight drops were repeated 100 times for each set to acquire accurate data and to minimise the presence of inaccurate data. A comparison of the results of the first two stages for the healthy and voided pipe showed slightly visible differences in the bowl-shaped graph precisely at the area where the void was implemented. However, the results of the longitudinal crack (without the void) did not show any obvious changes in the graphs. Analysis of the results demonstrates a gap and discontinuity in the graph for both devices, and the possible interpretation for this is that the bouncing effect of the traveling wave caused by the impact might push the device up so that there was no contact between the phone and the

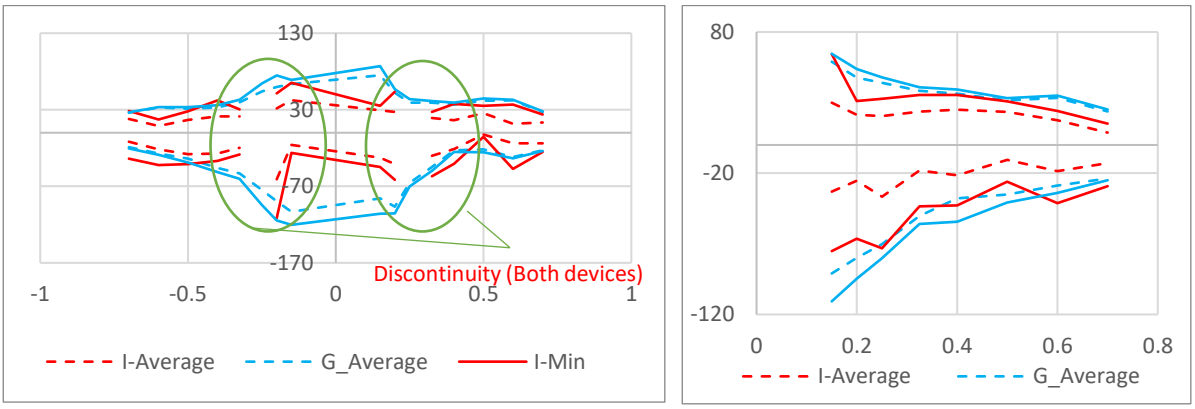
sand surface for a concise period. These points should be connected in a standard case. However, this problem can be solved by designing a better fixing mechanism so that the phone could be placed on the sand surface and to avoid undesired movement because of the bouncing effect.

A comparison of the data illustrates another issue with the iPhone data, which is probably because of the low sampling rate (limited by software) of the sensor. The 100 Hz sampling rate of the iPhone accelerometer sensor could not capture all particle movements in particular distances from the pipe centre and shows a noisy result (Figure 5). The most considerable irregularity of the results is related to the closest measurement points between 0.15 to 0.325 m from the centre (on both sides) where the mass impact is more significant than anywhere else, and the bouncing causes the maximum effect on the devices.

Compared to the iPhone, the industrial geophone with the sampling rate of max 240 Hz (two times more than any current iPhone or Android phones) delivered more reliable data. After data post-processing, a relatively clear bowl-shaped graph was produced for the SM-24 geophone compared to the iPhone. The maximum displacement of the sand particles was measured close to the impact area, and the minimum displacement at the box edges with the maximum distance to the pipe centre (Figure 4, 5, and 6).

At the end of the experiment, a decision should be made on whether phone accelerometers are capable of obtaining reliable data (as these sensors are limited by the operating system and it is not possible to increase the sampling rate manually) through improving the mechanical dropping system. Analysis of the measured data showed that this novel technique has the potential to be employed in isolation or as a parallel approach to other techniques (e.g. GPR and CAT) for locating and condition assessment of shallow pipes. However, it was found that the sampling rate of the accelerometers in the phone is generally limited (limited

by the operating system) to approximately 100 Hz, i.e. far below the capability of these embedded sensors. In other words, such a limitation in the sampling rate can cause discontinuities in the sampled data. Nevertheless, this issue can be solved by increasing the sampling rate of the sensor manually (applying changes in the core of the operating system), which was out of the scope of this research. Example reports associated with two loading weights (10KN and 20KN) are presented below in Figures 7 and 8.




(a) (b)

Figure 4. (a) Discontinuity of the results (both devices), (b) Right side graph with the same trend for both devices-Voided pip

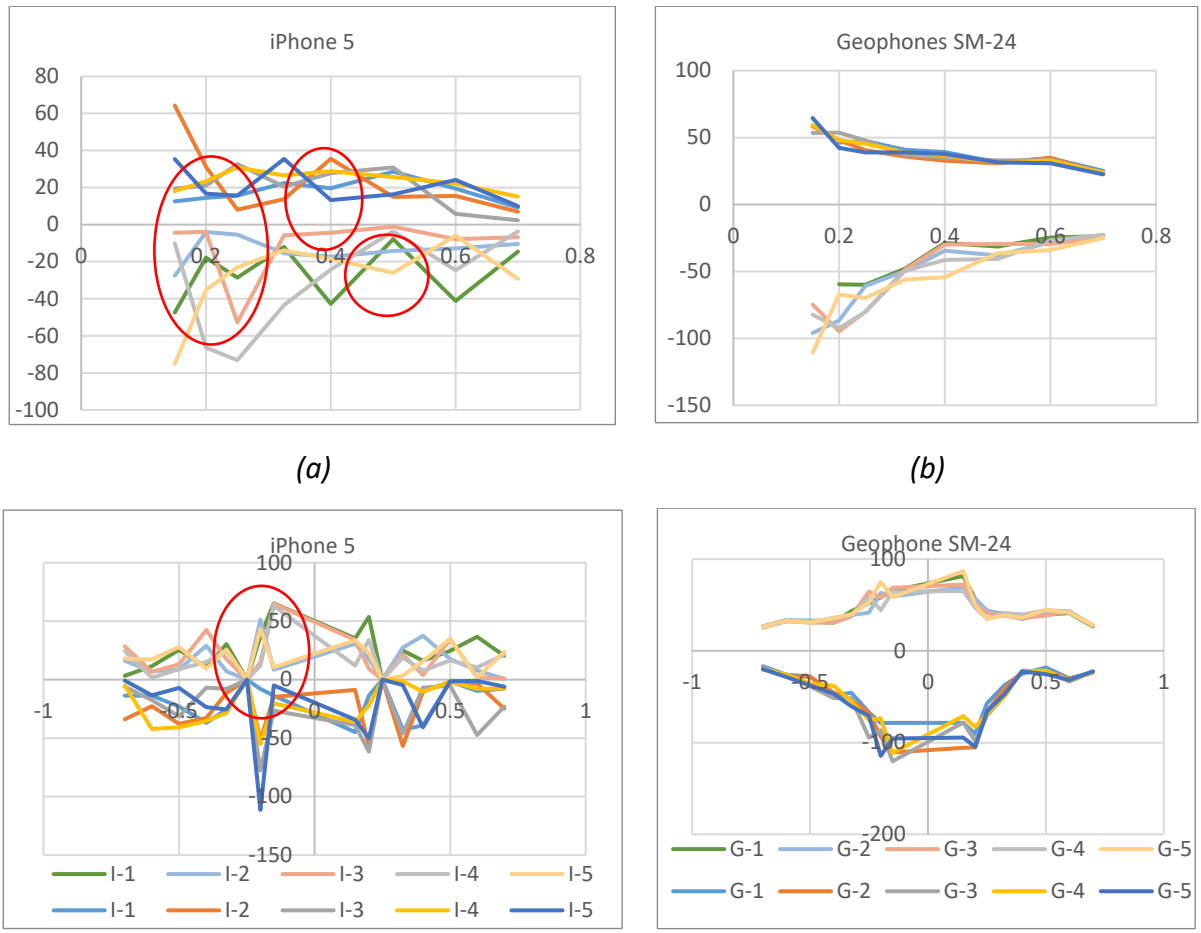
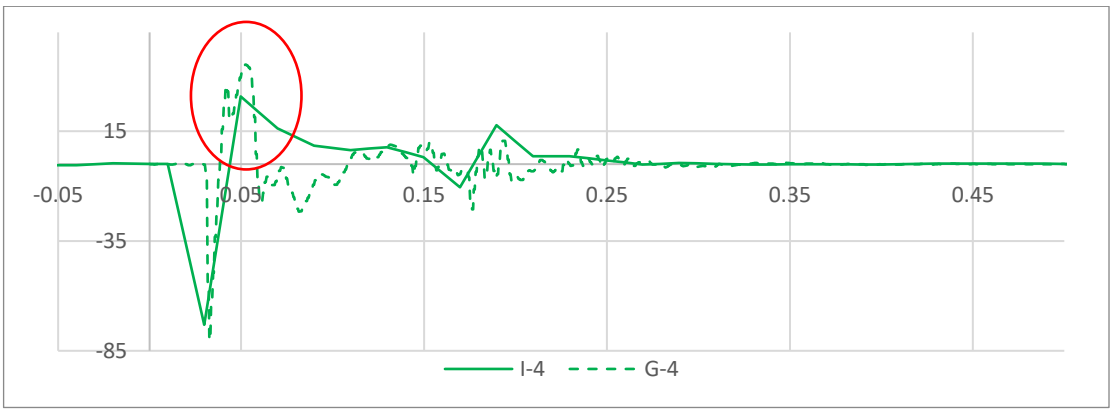
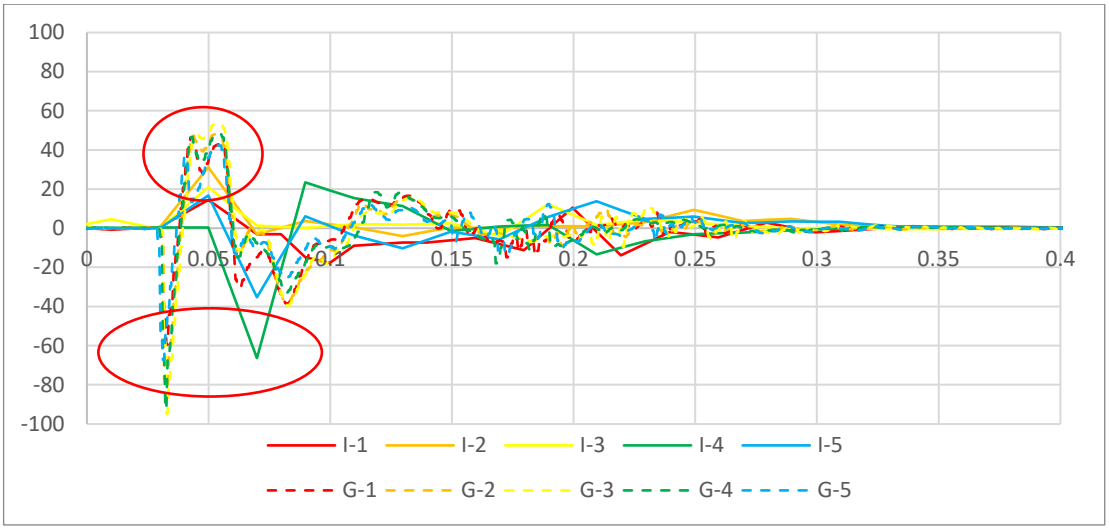


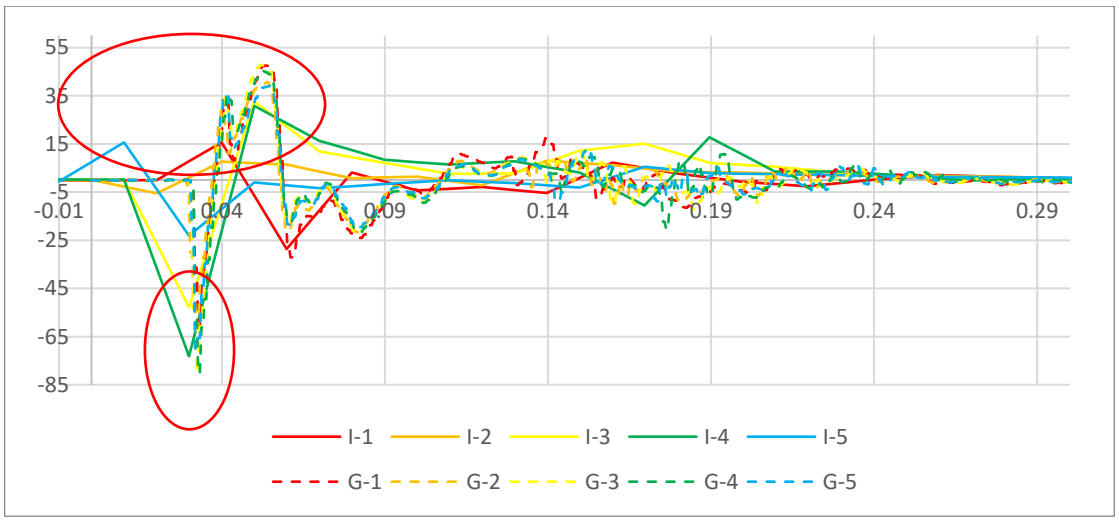

Figure 5. (a) iPhone- only right side of the Min-Max, (b) Geophones only right side of Min-Max, (c) iPhone Min-Max data (both sides), (d) Geophone Min-Max data create bowl shape




*(a)*




*(b)*




*(c)*

Figure 6. (a) Difference in iPhone and Geophone results (single drop for simplicity), (b) Results of displacements from 25 cm distance to the centre of the pipe with void, (c) Matching results from 20 cm distance to the centre of the pipe with void

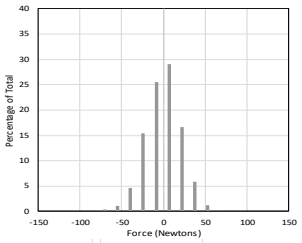
## TEST CONDITIONS

| | | | | GEOPHONE LOCATIONS | | | |
|---|---|---|---|---|---|---|---|
| TEST ID | H_10kN_GA1_1.01 | INITIAL STATIC LOAD | 111 | Geophone ID | Radial Offset (m) | Geophone ID | Radial Offset (m) |
| DATE | 08/02/2018 | TARGET IMPACT LOAD | 10000 | 1 | -0.15 | 7 | 0.4 |
| TIME | 11:38:00 | MEASURED IMPACT LOAD (N) | 5054 | 2 | -0.25 | 8 | 0.7 |
| OPERATOR | J Pring | SAND ABSORBED LOAD (N) | 4946 | 3 | -0.4 | Geophone ID | |
| BLOW NUMBER | 1 | OFFSET TEST FREQUENCY (Hz) | 2048 | 4 | -0.7 | 9 | -0.06 |
| PIPE CONDITION | 0.3m | OFFSET TEST NUMBER OF READINGS | 2048 | 5 | 0.15 | 10 | 0.06 |
| PIPE DEPTH | Healthy | IMPACT TEST FREQUENCY (Hz) | 2048 | 6 | 0.25 | 11 | Pipe |
| SAND DENSITY | 1.67 | IMPACT TEST NUMBER OF READINGS | 10240 | | | | |

## NULL OFFSET AND ERROR DATA

| | NULL OFFSET | STANDARD DEVIATION | MAXIMUM | MINIMUM |
|---|---|---|---|---|
| GEOPHONE 1 (mm/s) | 0.016878 | 0.007226 | 0.021664 | -0.025061 |
| GEOPHONE 2 (mm/s) | 0.022019 | 0.007353 | 0.02365 | -0.026507 |
| GEOPHONE 3 (mm/s) | 0.020469 | 0.007001 | 0.022296 | -0.022581 |
| GEOPHONE 4 (mm/s) | 0.017491 | 0.007111 | 0.019994 | -0.023035 |
| GEOPHONE 5 (mm/s) | 0.0167 | 0.007126 | 0.02237 | -0.020923 |
| GEOPHONE 6 (mm/s) | 0.020437 | 0.007126 | 0.024704 | -0.024924 |
| GEOPHONE 7 (mm/s) | 0.024354 | 0.007243 | 0.021051 | -0.023034 |
| GEOPHONE 8 (mm/s) | 0.022248 | 0.006752 | 0.021309 | -0.024888 |
| GEOPHONE 9 (mm/s) | 0.019373 | 0.008559 | 0.047942 | -0.046827 |
| GEOPHONE 10 (mm/s) | 0.029916 | 0.008975 | 0.049806 | -0.062914 |
| GEOPHONE 11 (mm/s) | 0.020575 | 0.008348 | 0.023774 | -0.02559 |
| LOAD CELL (N) | 351.225648 | 20.932684 | 78.158696 | -75.253186 |

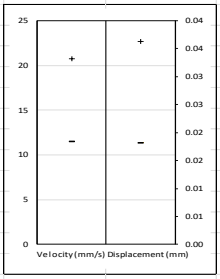
LOAD CELL DATA SPREAD



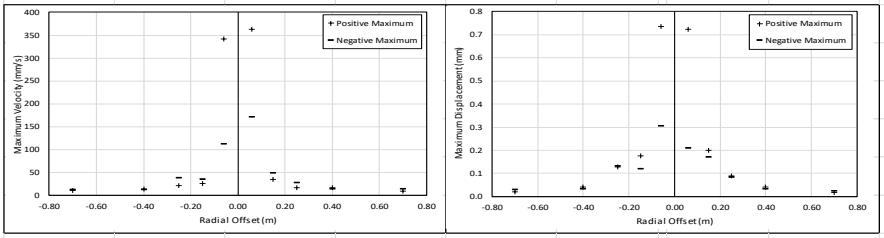

## TEST DATA

| GEOPHONE ID | LOCATION | MAX POSITIVE VELOCITY (mm/s) | MAX NEGATIVE VELOCITY (mm/s) | MAX POSITIVE DISPLACEMENT (mm) | MAX NEGATIVE DISPLACEMENT (mm) |
|---|---|---|---|---|---|
| 1 | -0.15 | 25.35 | 35.05 | 0.18 | 0.12 |
| 2 | -0.25 | 20.82 | 37.96 | 0.13 | 0.13 |
| 3 | -0.40 | 13.92 | 12.90 | 0.04 | 0.03 |
| 4 | -0.70 | 10.15 | 12.94 | 0.02 | 0.03 |
| 5 | 0.15 | 34.25 | 49.20 | 0.20 | 0.17 |
| 6 | 0.25 | 16.62 | 27.68 | 0.09 | 0.08 |
| 7 | 0.40 | 16.74 | 13.74 | 0.04 | 0.03 |
| 8 | 0.70 | 9.46 | 14.04 | 0.02 | 0.02 |
| 9 | -0.06 | 342.25 | 111.63 | 0.73 | 0.30 |
| 10 | 0.06 | 363.60 | 171.26 | 0.72 | 0.21 |
| 11 | Pipe | 20.74 | 11.46 | 0.04 | 0.02 |

## PIPE MOVEMENT




## SURFACE MOVEMENT



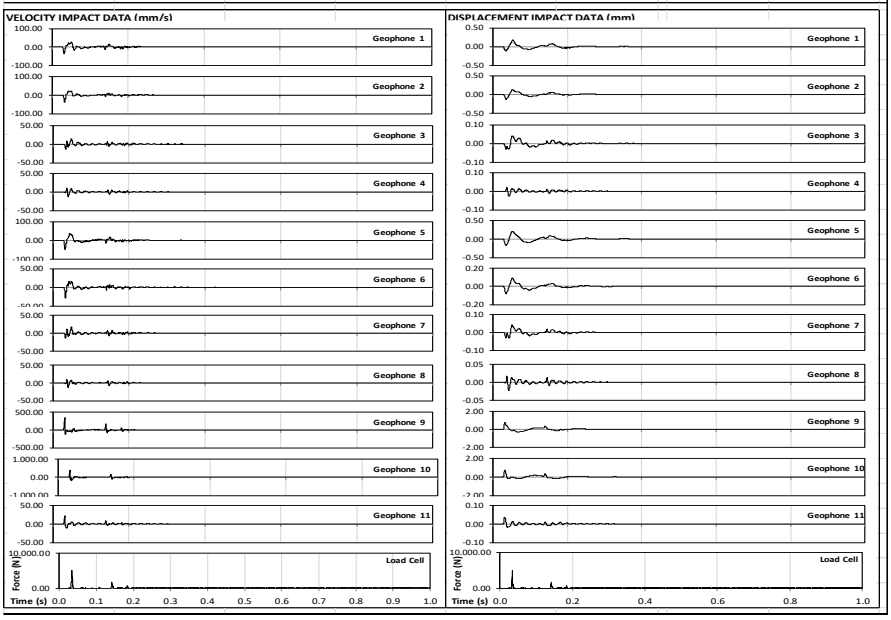




Figure 7. Report for the loading weight 10KN

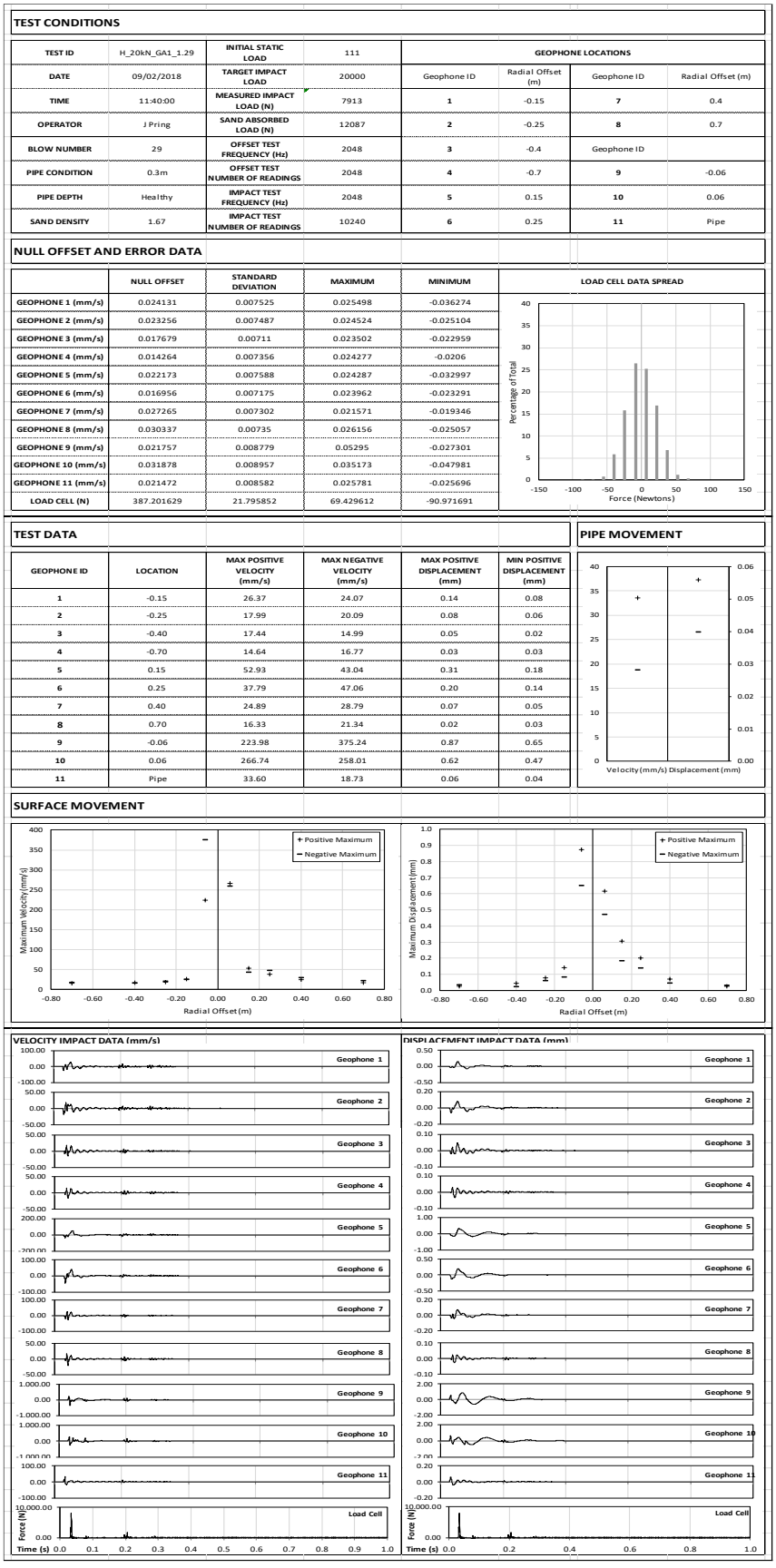



Figure 8. Report for the loading weight

## Scent and gas detection sensors

The discussion in section 2.6 of the literature review (Appendix I) in the previous chapter showed that artificial noses, i.e. sent and gas detection sensors, can potentially be utilised to develop a novel approach for the detection and location of buried pipes and objects in conjunction with other geophysical subsurface exploration techniques. However, despite the uniqueness of such an approach, as demonstrated in section 2.6, existing sensors do not provide the desired performance in terms of analysis time (not even close to the sensitivity of biological noses of, for instance, dogs, pigs, or rats), which again can cause delays and very long survey times (high cost). The development of new sensors was out of the scope of this research.

# Appendix III AUC TABLES OF ROC CURVES

This appendix contains the two AuC of ROC curves associated with the results presented section 4.3.3 of the results and discussion chapter of the thesis.

Table 2. AUC corresponded to Pareto solution 4

| | | Fourier Frequency [rad/sec] | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | FFreq | 7.00 | 9.00 | 10.00 | 11.00 | 12.00 | 13.00 | 15.00 | 18.00 | 20.00 | 25.00 | 30.00 | 35.00 | 40.00 |
| B-scan number | 4 | 55.98 | 64.91 | 65.95 | 68.95 | 65.47 | 64.33 | 66.38 | 69.85 | 70.00 | 67.79 | 68.10 | 68.43 | 70.50 |
| | 5 | 84.05 | 83.74 | 83.27 | 78.96 | 80.17 | 80.83 | 79.29 | 74.32 | 73.08 | 72.24 | 70.95 | 71.05 | 70.68 |
| | 6 | 82.19 | 79.76 | 80.43 | 81.92 | 79.73 | 79.83 | 76.86 | 74.12 | 71.26 | 69.58 | 67.71 | 67.78 | 69.13 |
| | 7 | 89.33 | 78.20 | 80.89 | 83.74 | 79.61 | 83.74 | 82.85 | 83.74 | 84.49 | 80.37 | 77.71 | 77.65 | 77.14 |
| | 8 | 81.90 | 80.27 | 80.44 | 76.41 | 73.47 | 73.25 | 73.19 | 71.44 | 70.92 | 67.97 | 66.97 | 67.48 | 66.76 |
| | 9 | 70.50 | 68.86 | 68.73 | 69.39 | 68.65 | 69.00 | 74.56 | 67.69 | 67.75 | 63.09 | 62.86 | 63.17 | 62.25 |
| | 10 | 79.65 | 78.56 | 79.09 | 81.10 | 76.88 | 77.12 | 75.84 | 73.17 | 71.60 | 67.68 | 68.05 | 69.22 | 69.33 |
| | 11 | 75.52 | 72.11 | 73.18 | 74.89 | 75.24 | 72.71 | 70.14 | 64.18 | 64.74 | 61.65 | 61.63 | 63.08 | 63.36 |
| | 12 | 72.31 | 74.19 | 75.12 | 77.69 | 77.33 | 78.11 | 72.54 | 74.79 | 73.13 | 69.78 | 70.32 | 72.00 | 70.63 |
| | 13 | 67.37 | 64.31 | 64.45 | 70.90 | 65.84 | 66.75 | 68.61 | 67.93 | 67.04 | 66.08 | 64.17 | 63.00 | 62.82 |
| | 14 | 75.24 | 76.03 | 80.43 | 73.73 | 73.36 | 75.09 | 76.16 | 72.61 | 70.92 | 67.22 | 67.15 | 66.55 | 67.45 |
| | 15 | 65.82 | 68.08 | 74.18 | 69.77 | 71.22 | 71.19 | 72.07 | 71.73 | 70.38 | 67.66 | 66.55 | 66.72 | 66.43 |
| | 16 | 55.33 | 53.08 | 56.28 | 53.73 | 53.82 | 53.95 | 52.35 | 49.99 | 49.47 | 49.17 | 48.53 | 49.06 | 48.76 |
| | 17 | 78.96 | 77.04 | 80.21 | 77.50 | 74.94 | 74.59 | 74.09 | 72.66 | 71.78 | 69.45 | 69.66 | 68.89 | 69.28 |
| | 18 | 69.06 | 69.87 | 72.95 | 74.46 | 71.22 | 71.25 | 68.94 | 67.75 | 68.03 | 63.80 | 63.82 | 62.05 | 63.39 |
| | 19 | 61.78 | 66.32 | 69.69 | 63.35 | 59.95 | 60.15 | 61.50 | 60.12 | 59.74 | 59.43 | 59.35 | 58.78 | 59.03 |
| | 20 | 56.75 | 56.54 | 56.49 | 56.59 | 54.07 | 53.68 | 56.32 | 55.11 | 56.20 | 55.42 | 55.62 | 56.10 | 55.79 |
| | Mean | 71.87 | 71.29 | 73.05 | 72.53 | 70.65 | 70.92 | 70.69 | 68.89 | 68.27 | 65.79 | 65.24 | 65.35 | 65.45 |

AUC of ROC [%]

Table 3. AUC corresponded to Pareto solution 1

| | | Fourier Frequency [rad/sec] | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | FFreq | 7.00 | 9.00 | 10.00 | 11.00 | 12.00 | 13.00 | 15.00 | 18.00 | 20.00 | 25.00 | 30.00 | 35.00 | 40.00 |
| B-scan number | 4 | 59.45 | 65.89 | 67.31 | 70.48 | 68.50 | 67.31 | 68.91 | 71.55 | 71.86 | 70.54 | 70.94 | 71.64 | 73.16 |
| | 5 | 84.22 | 83.64 | 83.96 | 80.25 | 80.53 | 81.52 | 79.62 | 74.73 | 74.83 | 73.86 | 73.32 | 73.35 | 73.22 |
| | 6 | 84.31 | 85.46 | 84.93 | 84.71 | 82.92 | 82.43 | 79.38 | 76.46 | 73.52 | 72.58 | 71.24 | 71.40 | 72.71 |
| | 7 | 90.31 | 83.36 | 85.80 | 89.50 | 86.40 | 89.91 | 86.76 | 86.86 | 86.74 | 85.61 | 82.91 | 83.48 | 83.40 |
| | 8 | 81.29 | 80.00 | 80.22 | 78.19 | 75.47 | 75.36 | 75.66 | 74.74 | 74.29 | 70.81 | 70.20 | 70.41 | 70.13 |
| | 9 | 72.43 | 71.63 | 70.92 | 72.47 | 71.90 | 72.85 | 77.78 | 71.61 | 70.72 | 66.27 | 65.98 | 65.76 | 65.76 |
| | 10 | 80.64 | 78.90 | 80.22 | 79.11 | 75.84 | 76.05 | 75.55 | 74.31 | 73.43 | 70.87 | 70.00 | 70.98 | 71.12 |
| | 11 | 80.54 | 76.69 | 77.68 | 77.58 | 77.50 | 76.58 | 72.61 | 70.19 | 71.52 | 68.77 | 67.71 | 68.92 | 68.25 |
| | 12 | 72.11 | 74.01 | 74.42 | 77.24 | 77.07 | 78.18 | 72.23 | 75.02 | 73.14 | 70.59 | 70.01 | 72.48 | 71.22 |
| | 13 | 67.06 | 65.54 | 65.27 | 69.81 | 65.60 | 67.19 | 69.08 | 69.23 | 68.81 | 66.53 | 63.82 | 62.75 | 63.28 |
| | 14 | 79.15 | 81.65 | 85.25 | 78.26 | 78.74 | 79.29 | 81.23 | 76.98 | 76.50 | 73.79 | 74.69 | 73.28 | 74.79 |
| | 15 | 62.70 | 65.50 | 70.36 | 66.55 | 67.24 | 67.50 | 68.08 | 67.78 | 66.00 | 63.44 | 63.09 | 63.34 | 63.43 |
| | 16 | 54.07 | 52.00 | 56.99 | 51.52 | 51.25 | 51.61 | 50.21 | 48.91 | 47.96 | 48.64 | 47.79 | 48.61 | 48.20 |
| | 17 | 72.70 | 71.63 | 75.77 | 72.84 | 69.50 | 70.19 | 70.23 | 69.64 | 69.34 | 68.07 | 67.77 | 67.69 | 68.22 |
| | 18 | 68.74 | 69.21 | 73.73 | 73.03 | 67.88 | 68.16 | 66.71 | 67.13 | 66.26 | 63.57 | 62.62 | 61.11 | 62.23 |
| | 19 | 62.05 | 65.20 | 69.55 | 64.04 | 60.03 | 61.00 | 61.01 | 59.25 | 59.08 | 59.09 | 59.89 | 58.72 | 59.00 |
| | 20 | 56.44 | 55.60 | 56.10 | 55.71 | 51.25 | 51.41 | 53.36 | 51.81 | 52.38 | 52.12 | 52.10 | 52.63 | 52.30 |
| | Mean | 72.25 | 72.11 | 74.03 | 73.02 | 71.04 | 71.56 | 71.08 | 69.78 | 69.20 | 67.36 | 66.71 | 66.86 | 67.08 |

AUC of ROC [%]