Computational Methods for
Processing Ground Penetrating
Radar Data

by
Nurul Jihan Farhah Bostanudin

School of Engineering
This thesis is submitted in partial fulfilment of the requirements
for the award of the degree of Doctor of Philosophy of the
University of Portsmouth.

May 2013
Abstract

The aim of this work was to investigate signal processing and analysis techniques for Ground Penetrating Radar (GPR) and its use in civil engineering and construction industry. GPR is the general term applied to techniques which employ radio waves, typically in the Mega Hertz and Giga Hertz range, to map structures and features buried in the ground or in man-made structures. GPR measurements can suffer from large amount of noise. This is primarily caused by interference from other radio-wave-emitting devices (e.g., cell phones, radios, etc.) that are present in the surrounding area of the GPR system during data collection. In addition to noise, presence of clutter – reflections from other non-target objects buried underground in the vicinity of the target can make GPR measurement difficult to understand and interpret, even for the skilled human, GPR analysts.

This thesis is concerned with the improvements and processes that can be applied to GPR data in order to enhance target detection and characterisation process particularly with multivariate signal processing techniques. Those primarily include Principal Component Analysis (PCA) and Independent Component Analysis (ICA). Both techniques have been investigated, implemented and compared regarding their abilities to separate the target originating signals from the noise and clutter type signals present...
in the data. Combination of PCA and ICA (SVDPICA) and two-dimensional PCA (2DPCA) are the specific approaches adopted and further developed in this work. Ability of those methods to reduce the amount of clutter and unwanted signals present in GPR data have been investigated and reported in this thesis, suggesting that their use in automated analysis of GPR images is a possibility.

Further analysis carried out in this work concentrated on analysing the performance of developed multivariate signal processing techniques and at the same time investigating the possibility of identifying and characterising the features of interest in pre-processed GPR images. The driving idea behind this part of work was to extract the resonant modes present in the individual traces of each GPR image and to use properties of those poles to characterise target. Three related but different methods have been implemented and applied in this work – Extended Prony, Linear Prediction Singular Value Decomposition and Matrix Pencil methods. In addition to these approaches, PCA technique has been used to reduce dimensionality of extracted traces and to compare signals measured in various experimental setups. Performance analysis shows that Matrix Pencil offers the best results.
Contents

Abstract

List of Figures

List of Tables

List of Abbreviations

List of Symbols

Declaration

Acknowledgements

Dissemination

Chapter 1 Introduction ................................................................................ 1

Overview .................................................................................................... 1

1.1 Ground Penetrating Radar ................................................................ 1

1.2 Objectives ........................................................................................ 4

1.3 Summary of Previous Work ............................................................. 8

1.3.1 GPR as a Detection Tool ...................................................... 8

1.3.2 GPR Technology in Civil Infrastructures ............................ 9

1.3.3 Signal Processing for GPR ................................................... 9

1.4 Thesis Structure ............................................................................. 12

Chapter 2 Ground Penetrating Radar ..................................................... 15

Overview .................................................................................................. 15

2.1 Introduction ........................................................................................ 15
Chapter 3 Pre-processing using ReflexW ................................................ 61

Overview .................................................................................................. 61

3.1 Introduction.................................................................................... 61

3.2 ReflexW Software.......................................................................... 64

3.3 Basic GPR Data Processing Steps ................................................. 64

3.3.1 Data Editing ....................................................................... 66

3.3.2 Dewow Filtering ................................................................ 68

3.3.3 Topographic Corrections.................................................... 69

3.3.3.1 Elevation Static Correction ................................. 70

3.3.3.2 Maximum Phase Correction ............................... 72

3.3.4 Time-Zero Correction ........................................................ 73

3.3.5 Filtering .............................................................................. 74

3.3.5.1 Temporal Filters ....................................................... 75
3.3.5.2 Spatial Filters ................................................................. 78
3.3.6 Gain Function....................................................................... 80
3.3.7 Migration............................................................................. 84
3.4 ReflexW Processing of the Measured GPR Image in this Work... 85
3.5 Conclusion ............................................................................... 88

Chapter 4 Multivariate Pre-processing Techniques for GPR Images .. 90

Overview .............................................................................................. 90
4.1 Introduction.................................................................................. 90
4.2 Review of Prior Research ............................................................ 91
  4.2.1 Previous Work in Signal Processing................................. 91
  4.2.2 Multivariate Analysis in GPR Application ......................... 93
4.3 Multivariate Analysis................................................................... 95
  4.3.1 Principal Component Analysis......................................... 95
    4.3.1.1 One-Dimensional Principal Component Analysis
    (1DPCA).................................................................................. 99
      4.3.1.1.1 PCA via Eigenvalue Decomposition of
                  Covariance Matrix............................................... 99
      4.3.1.1.2 PCA via Singular Value Decomposition
                  of Data Matrix...................................................... 104
    4.3.1.2 Dimensionality Reduction using PCA........................... 106
    4.3.1.4 GPR Clutter Reduction using PCA............................. 109
4.3.1.2 Two-Dimensional Principal Component Analysis (2DPCA) .................................................................111

4.3.1.3 Order Selection .................................................................................114

4.3.2 Independent Component Analysis (ICA) ..............................................114

4.3.2.1 Independent Component Analysis Algorithms .........................119

4.3.3 Proposed Novel Algorithm ..........................................................121

4.4 Results ..........................................................................................123

4.4.1 Algorithm Evaluation Parameters ...............................................124

4.4.1.1 Peak-Signal-to-Noise Ratio (PSNR) ..............................................125

4.4.1.2 Mean Structural Similarity (M-SSIM) Index .........................126

4.4.2 Simulated Data ..........................................................................130

4.4.2.1 Clutter Reduction .................................................................131

4.4.2.2 Unwanted Signals Reduction .................................................136

4.4.3 Measured Data ...........................................................................136

4.4.3.1 Clutter Reduction .................................................................137

4.4.3.2 Unwanted Signals Reduction .................................................145

4.5 Conclusion ..................................................................................151

Chapter 5 Evaluation of Developed Pre-processing Techniques and Characterisation of Targets using Signal Modelling ................................153

Overview ................................................................................................153

5.1 Introduction ..................................................................................154

5.2 Techniques ...................................................................................155
5.2.1 Prony Analysis ................................................................. 155
5.2.2 Extended Prony Method.................................................. 158
5.2.3 Linear Prediction Singular Value Decomposition
(Kumaresan-Tufts) Method.................................................... 159
5.2.4 Matrix Pencil Method ...................................................... 160
5.3 Results .................................................................................. 164
5.3.1 Synthesised Signals.......................................................... 164
5.3.2 GPR Signals and Model Order Estimation ...................... 168
  5.3.2.1 Model Order Estimation Problem ......................... 168
  5.3.2.2 Analysis of Measured GPR Signals .................... 171
5.3.3 Towards Target Characterisation – Analysis of GPRMax
  Simulated Signals............................................................... 183
5.3.4 Target Characterisation using Principal Component
  Analysis on Simulated Data............................................... 187
5.4 Conclusion ........................................................................... 193

Chapter 6 Findings and Conclusions ........................................... 195

Overview ................................................................................. 195

6.1 Extended Summary of the Main Ideas and Work Done on the
  Project .................................................................................. 195
6.2 Author’s Contributions ......................................................... 203
6.3 Recommendations for Further Work .................................... 205
  6.3.1 GPR Simulations......................................................... 205
6.3.2 GPR Measurements .......................................................... 205
6.3.3 Pre-processing Algorithms ............................................... 206
6.3.4 Analysis Algorithms ........................................................ 206

Bibliography .................................................................................. 208

Appendix A: Chapter 2 ....................................................................... 227
  A.1 Input Files of GPRMax Simulation on Joints, Holes and Cracks
    Detection .................................................................................. 227
  A.1.1 GPR 4 GHz Frequency for Joints and Holes ....................... 227
  A.1.2 GPR 4 GHz Frequency for Crack ...................................... 228
  A.2 Input Files of GPRMax Simulation on Pipes Detection .......... 229
    A.2.1 GPR 4 GHz Frequency for Three Pipes in a Box ............. 229
  A.3 Media File of the GPRMax Data ............................................. 230
  A.4 Matlab Code for GPRMax ..................................................... 231

Appendix B: Chapter 4 ....................................................................... 233
  B.1 Clutter Reduction Algorithms Code ...................................... 233
  B.2 Unwanted Signals Reduction Algorithms Code .................... 238
  B.3 Built-in Functions Code ....................................................... 244
    B.3.1 mrem.m ........................................................................ 244
    B.3.2 imscale.m .................................................................. 244

Appendix C: Chapter 5 ....................................................................... 246
  C.1 One-dimensional Analysis Code .......................................... 246
  C.2 Selecting Model Order Code ............................................... 251
C.3 Built-in Functions Code............................................................... 255

C.3.1 myprincomp.m ........................................................................... 255

C.3.2 spread.m .............................................................................. 256

C.3.3 sortpoles.m ........................................................................ 257
List of Figures

Figure 1.1: Research system implementation. ......................................................... 5

Figure 2.1: Typical GPR system. ........................................................................... 16

Figure 2.2: Basic functional principle of a GPR device........................................ 18

Figure 2.3: Principles of GPR in CMP mode which the transmitting and receiving antennas are moved away from each other [44]. .......... 19

Figure 2.4: Linescan and oscilloscope displays (right) show the (a) direct coupling and (b) reflections and diffraction hyperbola of GPR data scanned over the plywood top panel of the pipe in a box. 31

Figure 2.5: Collecting the position of cracks in practical approach with measuring tape. ................................................................. 34

Figure 2.6: GPR measurement taken across a pipe in a box of (a) one-dimensional and (b) two-dimensional GPR data. The one dimensional data is taken at the distance of 0.4m in two-dimensional image. (c) Three-dimensional GPR data represents an empty box. ................................................................. 37

Figure 2.7: GPR Groundvue 5. ........................................................................... 38

Figure 2.8: GPR Groundvue 3 with 400 MHz, 1 GHz, and Crack Detection Head antenna systems. ................................................................. 39

Figure 2.9: The two-dimensional GPR forward problem and its GPRMax2D domain bounded by ABC. ......................................................... 43
Figure 2.10: The two-dimensional FDTD cell [58]. ..............................................45

Figure 2.11: Example of GPRMax2D input file in simulating concrete slab
with joints using 1.5 GHz frequency radar. Red arrows
represent lines that were deactivated and ignored by the
program as the first character of each command starts without
the hash character (#). ........................................................................48

Figure 2.12: Example of GPRMax2D model.................................................49

Figure 2.13: (a) Experimental setup used in collecting GPR data for pipes
detection, (b) sand was placed in the small container box, (c)
top panel with grid drawn on it, and (d) position of three pipes
in the container box. ........................................................................51

Figure 2.14: (a) Completed container boxes and (b) sand preparation for dry
condition tests......................................................................................54

Figure 2.15: (a) Preparation of concrete slabs and (b) two concrete slabs
were ready with joints attached.........................................................55

Figure 2.16: Introduction of holes and cracks on the concrete slabs. .......55

Figure 2.17: (a) Depiction of joints, holes, and cracks position and (b) the
approximate positions of the radar traverses during the
experiments. .......................................................................................56

Figure 2.18: Data collection over a cracked concrete metal fibre mixtures
slab. .....................................................................................................57

Figure 3.1: Overview of GPR data processing flow [37], [62]..................63
Figure 3.2: GPR processing flow using ReflexW................................. 65

Figure 3.3: Example of ‘desaturation' function is used in correcting the
clipped form of saturated traces [37]. ............................................. 67

Figure 3.4: Dewow filter correction on a raw GPR trace [37]................. 69

Figure 3.5: Datum statistics terminology [64]. ........................................ 70

Figure 3.6: The position of Sender/Receiver distance. ......................... 71

Figure 3.7: The application of maximum phase correction step to measured
GPR data (a) before and (b) after. ................................................... 72

Figure 3.8: Time-zero variations in a selected 1.5 GHz GPR data. (a) Time-
zero position on the original GPR data (white dashed-line) and
(b) the white dashed-line change in time-zero position. ............. 74

Figure 3.9: Principle of a simple band-pass filter in the frequency domain
and the form of two common filter functions: a notch and
Butterworth filters [37]. ............................................................. 77

Figure 3.10: Realistic example of the use band-pass Butterworth,
background removal, and average subtraction filtering on a 4
GHz GPR section collected over a concrete slab with crack
structure at 0.18 m. ................................................................. 80

Figure 3.11: Concept of time-varying gain where signal amplification varies
with time [81]. ........................................................................... 81

Figure 3.12: The effect of inordinate acquisition gain selection [7]......... 82

Figure 3.13: Acquisition gain removal during pre-processing [7]........... 83
Figure 3.14: The problem – the reflected energy of the object is spread out over the different receiver positions leading to the measurement hyperbola. The idea of migration is to recombine this energy in one position [79].

Figure 3.15: Example of ReflexW process on GPR data to detect leakages from joints.

Figure 4.1: Graphical model for Principal Component Analysis.

Figure 4.2: (a) GPR image contained in data matrix $X$ and (b) single GPR trace $x_i (i = 120)$ from $X$ – A-scan.

Figure 4.3: Graphical depiction of SVD of data matrix $X$.

Figure 4.4: One-dimensional Principal Component Analysis (1DPCA).

Figure 4.5: Two-dimensional Principal Component Analysis (2DPCA).

Figure 4.6: Blind Source Separation (BSS) block diagram. $s(t)$ are the sources, $x(t)$ are the recordings, $\hat{s}(t)$ are the estimated sources, $A$ is a mixing matrix and $W$ is an un-mixing matrix [111].

Figure 4.7: ICA method in GPR application.

Figure 4.8: Three stages of SVPICA method – SVD decomposition of original data matrix $X$, discarding of the first $l$ and last $n-k$ PCs, ICA un-mixing the reduced matrix of eigenvectors $\hat{U}_R$ to obtain matrix $\tilde{U}_R$.

Figure 4.9: Diagram of the structural similarity (SSIM) measurement system [122].
Figure 4.10: Plot of eigenvalues against space sample (component number) of simulated GPR data represent joints (a) with PCA via SVD and (b) selected 30 PCs via SVDPICA method. .................... 132

Figure 4.11: Results of GPR image decluttering on simulated data represent joints using different techniques (a) original raw image, (b) mean removal, (c) PCA via SVD, (d) SVDPICA and (e) “Ideal” data which is obtained by subtracting the image of concrete from the image of concrete containing joints. ........ 133

Figure 4.12: Evaluation of clutter reduction methods in terms of PSNR for simulated data to detect leakages from joints, holes and cracks......................................................... 134

Figure 4.13: Evaluation of clutter reduction methods in terms of PSNR for simulated data in detecting pipes................................. 135

Figure 4.14: Plot of eigenvalues against space sample (component number) of measured GPR data in detecting pipes – big size (a) with PCA via SVD and (b) selected 30 PCs via SVDPICA method......................................................... 138

Figure 4.15: Results of GPR image decluttering on measured data – big pipe using different techniques (a) original raw image, (b) mean removal, (c) PCA via SVD, (d) SVDPICA and (e) ReflexW software................................. 139
Figure 4.16: Evaluation of clutter reduction methods in terms of PSNR for pipes detection of measured data................................. 141

Figure 4.17: Evaluation of clutter reduction methods in terms of PSNR for measured data to detect leakages from joints, holes and cracks (GPR GV3)........................................................................... 142

Figure 4.18: Evaluation of clutter reduction methods in terms of PSNR for measured data to detect leakages from joints, holes and cracks (GPR GV5)................................................................. 143

Figure 4.19: Evaluation of clutter reduction methods in terms of PSNR for measured data to detect cracked concrete. ...................... 144

Figure 4.20: GPR images with removed mean trace (a) empty box and (b) box with three plastic pipes.............................................. 145

Figure 4.21: (a) Eigenvalues spectrum of PCs obtained using 2DPCA procedure and (b) reconstructed side reflection using 2DPCA eigenimages 1 – 6. ................................................................. 146

Figure 4.22: Independent images obtained using ICA in reconstructing unwanted features (side reflections)............................... 147

Figure 4.23: Unwanted signals (side reflections) reduction results via (a) 2DPCA and (b) ICA algorithms. ............................................. 148

Figure 4.24: The "difference" image is obtained by subtracting the empty box image from the image with three pipes. ......... 149
Figure 4.25: Evaluation results of unwanted signals reduction techniques for pipes detection of measured data. ........................................ 150

Figure 5.1: Performance of studied algorithms for SNR = 10 dB case, (a) original signal, (b), (c) and (d) reconstructed signals and residual using (b) EP, (c) LPSVD and (d) MP. ....................... 167

Figure 5.2: (a) Singular Entropy Increment and (b) Differential of Singular Entropy Increment characteristics for simulated signal (SNR = 10 dB) depicting correct signal model order at \( L = 6 \). .......... 171

Figure 5.3: Big pipe image after pre-processing via (a) mean removal, (b) 2DPCA, (c) ICA and (d) difference techniques. The white line in (a) represents the position of analysed trace – column 83. ............................................................................................ 172

Figure 5.4: Extracted GPR traces of pre-processed data via (a) mean removal, (b) 2DPCA, (c) ICA and (d) difference techniques. 173

Figure 5.5: (a) SEI and (b) DSEI for each of extracted traces. ............... 174

Figure 5.6: The signal reconstructions using selected model order of a single trace. ................................................................. 176

Figure 5.7: Results of Matrix Pencil on a single trace. ....................... 177

Figure 5.8: (a) Power and (b) maximal amplitude of the residual signal for different model orders \( L \). ......................................................... 178
Figure 5.9: Comparative results of MP analysis obtained for a single pipe in the box (a) frequency-amplitude and (b) frequency-damping characteristics.......................................................... 180

Figure 5.10: Image of all three pipes in the box after pre-processing via (a) mean removal, (b) 2DPCA, (c) ICA and (d) difference techniques. The white lines in (a) represent the positions of analysed traces – column 25, 83 and 138.......................... 181

Figure 5.11: Comparative results of MP analysis obtained for all three pipes in the box (a) frequency-amplitude and (b) frequency-damping characteristics (for traces 25, 83 and 138 from Figure 5.10). 182

Figure 5.12: Simulated GPRMax images with indicated traces (white lines) used in further analysis (a) joint, (b) crack, (c) crack and healthy concrete and (d) hole at the joint position. ............... 185

Figure 5.13: Comparative results of MP analysis obtained from simulated images containing various targets in concrete slab (a) frequency-amplitude and (b) frequency-damping characteristics................................................................. 186

Figure 5.14: PCs represent different target types on simulated data in (a) two-dimensional and (b) three-dimensional...................... 191
List of Tables

Table 2.1: Pipe dimensions used in the experiment performed for pipes detection. .............................................................................................................. 52

Table 3.1: Sender/Receiver distance values. .................................................. 71

Table 3.2: Band-pass Butterworth filter cut-off frequencies. ...................... 76

Table 4.1: Two ways of performing PCA including common dimensionality reduction and data reconstruction steps implemented in Matlab; left column – Eigenvalue Decomposition approach and right column – SVD approach. .......................................................... 108

Table 4.2: PSNR and M-SSIM index indicators for described clutter reduction algorithms of simulated data experiment to detect leakages from joints, holes and cracks. ............................................. 134

Table 4.3: PSNR and M-SSIM index indicators for described clutter reduction algorithms of simulated data experiment to detect pipes. ........................................................................................................ 135

Table 4.4: PSNR and M-SSIM index indicators for described clutter reduction algorithms of measured data experiment to detect pipes. ........................................................................................................ 141

Table 4.5: PSNR and M-SSIM index indicators for described clutter reduction algorithms of measured data experiment to detect leakages from joints, holes and cracks (GPR GV3). ............... 142
Table 4.6: PSNR and M-SSIM index indicators for described clutter reduction methods of measured experiment to detect leakages from joints, holes and cracks (GPR GV5). ........................................ 143

Table 4.7: PSNR and M-SSIM index indicators for described clutter reduction methods of measured experiment to detect cracked concrete. ................................................................. 144

Table 4.8: Comparison of proposed algorithms in reducing unwanted signals using PSNR measure of pipes detection experiment. ............... 150

Table 5.1: Performance of three signal modelling algorithms using simulated signals with different SNRs............................................. 166

Table 5.2: Extracted traces of GPR simulated data on different target types. ......................................................................................... 188

Table 5.3: Calculated distance of each target depicted in Figure 5.14 (a). 192
List of Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1D</td>
<td>One-dimensional</td>
</tr>
<tr>
<td>2D</td>
<td>Two-dimensional</td>
</tr>
<tr>
<td>3D</td>
<td>Three-dimensional</td>
</tr>
<tr>
<td>1DPCA</td>
<td>One-dimensional Principal Component Analysis</td>
</tr>
<tr>
<td>2DPCA</td>
<td>Two-dimensional Principal Component Analysis</td>
</tr>
<tr>
<td>ABC</td>
<td>Absorbing Boundary Condition</td>
</tr>
<tr>
<td>AGC</td>
<td>Automatic Gain Control</td>
</tr>
<tr>
<td>AM</td>
<td>Amplitude Modulation</td>
</tr>
<tr>
<td>AP</td>
<td>Anti-Personal</td>
</tr>
<tr>
<td>ASCII</td>
<td>American Standard Code for Information Interchange</td>
</tr>
<tr>
<td>BSS</td>
<td>Blind Signal Separation</td>
</tr>
<tr>
<td>CDH</td>
<td>Crack Detection Head</td>
</tr>
<tr>
<td>CFL</td>
<td>Courant, Freidrichs, and Lewy</td>
</tr>
<tr>
<td>cm</td>
<td>Centimetre</td>
</tr>
<tr>
<td>CMP</td>
<td>Common-Mid-Point</td>
</tr>
<tr>
<td>dB</td>
<td>Decibel</td>
</tr>
<tr>
<td>DC</td>
<td>Direct Current</td>
</tr>
<tr>
<td>DSEI</td>
<td>Differential of Singular Entropy Increment</td>
</tr>
<tr>
<td>EM</td>
<td>Electromagnetic</td>
</tr>
<tr>
<td>EP</td>
<td>Extended Prony</td>
</tr>
<tr>
<td>Symbol</td>
<td>Definition</td>
</tr>
<tr>
<td>--------</td>
<td>------------</td>
</tr>
<tr>
<td>F</td>
<td>Farads</td>
</tr>
<tr>
<td>FA</td>
<td>Factor Analysis</td>
</tr>
<tr>
<td>FastICA</td>
<td>Fixed-Point ICA</td>
</tr>
<tr>
<td>FDTD</td>
<td>Finite-Difference Time-Domain</td>
</tr>
<tr>
<td>FFT</td>
<td>Fast Fourier Transform</td>
</tr>
<tr>
<td>GHz</td>
<td>Giga Hertz</td>
</tr>
<tr>
<td>GPR</td>
<td>Ground Penetrating Radar</td>
</tr>
<tr>
<td>GPS</td>
<td>Global Positioning System</td>
</tr>
<tr>
<td>GV3</td>
<td>Groundvue 3</td>
</tr>
<tr>
<td>GV5</td>
<td>Groundvue 5</td>
</tr>
<tr>
<td>H</td>
<td>Henries</td>
</tr>
<tr>
<td>HVS</td>
<td>Human Visual System</td>
</tr>
<tr>
<td>Hz</td>
<td>Hertz, cycle per second</td>
</tr>
<tr>
<td>IC</td>
<td>Independent Component</td>
</tr>
<tr>
<td>ICA</td>
<td>Independent Component Analysis</td>
</tr>
<tr>
<td>Infomax</td>
<td>Information Maximisation</td>
</tr>
<tr>
<td>JADE</td>
<td>Joint Approximate Diagonalisation of Eigen-matrices</td>
</tr>
<tr>
<td>kg</td>
<td>Kilogram</td>
</tr>
<tr>
<td>KT</td>
<td>Kumaresan-Tufts</td>
</tr>
<tr>
<td>LDA</td>
<td>Linear Discriminant Analysis</td>
</tr>
<tr>
<td>LPSVD</td>
<td>Linear Prediction Singular Value Decomposition</td>
</tr>
<tr>
<td>LS</td>
<td>Least-Squares</td>
</tr>
<tr>
<td>Symbol</td>
<td>Abbreviation</td>
</tr>
<tr>
<td>--------</td>
<td>--------------</td>
</tr>
<tr>
<td>m</td>
<td>Metre</td>
</tr>
<tr>
<td>MHz</td>
<td>Mega Hertz</td>
</tr>
<tr>
<td>mm</td>
<td>Millimetre</td>
</tr>
<tr>
<td>MP</td>
<td>Matrix Pencil</td>
</tr>
<tr>
<td>MR</td>
<td>Mean Removal</td>
</tr>
<tr>
<td>ms</td>
<td>Millisecond</td>
</tr>
<tr>
<td>MSE</td>
<td>Mean Square Error</td>
</tr>
<tr>
<td>M-SSIM</td>
<td>Mean Structural Similarity</td>
</tr>
<tr>
<td>m/ns</td>
<td>Metre per nanosecond</td>
</tr>
<tr>
<td>m/s</td>
<td>Metre per second</td>
</tr>
<tr>
<td>NDCD</td>
<td>Non-Destructive Crack Depth</td>
</tr>
<tr>
<td>ns</td>
<td>Nanosecond</td>
</tr>
<tr>
<td>PA</td>
<td>Prony Analysis</td>
</tr>
<tr>
<td>PC</td>
<td>Principal Component</td>
</tr>
<tr>
<td>PCA</td>
<td>Principal Component Analysis</td>
</tr>
<tr>
<td>PML</td>
<td>Perfectly Matched Layer</td>
</tr>
<tr>
<td>ps</td>
<td>Picosecond</td>
</tr>
<tr>
<td>PSNR</td>
<td>Peak-Signal-to-Noise Ratio</td>
</tr>
<tr>
<td>rads</td>
<td>Radians</td>
</tr>
<tr>
<td>RMS</td>
<td>Root Mean Square</td>
</tr>
<tr>
<td>s</td>
<td>Second</td>
</tr>
<tr>
<td>S</td>
<td>Siemens</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
</tr>
<tr>
<td>--------------</td>
<td>-------------------------------------------</td>
</tr>
<tr>
<td>SE</td>
<td>Singular Entropy</td>
</tr>
<tr>
<td>SEI</td>
<td>Singular Entropy Increment</td>
</tr>
<tr>
<td>SNR</td>
<td>Signal-to-Noise Ratio</td>
</tr>
<tr>
<td>S/R</td>
<td>Sender/Receiver</td>
</tr>
<tr>
<td>SSA</td>
<td>Singular Spectrum Analysis</td>
</tr>
<tr>
<td>SSIM</td>
<td>Structural Similarity</td>
</tr>
<tr>
<td>SVD</td>
<td>Singular Value Decomposition</td>
</tr>
<tr>
<td>SVDPICA</td>
<td>Singular Value Decomposition of PCA and ICA</td>
</tr>
<tr>
<td>TDSEP</td>
<td>Temporal Decorrelation Source Separation</td>
</tr>
<tr>
<td>V</td>
<td>Voltage</td>
</tr>
</tbody>
</table>
## List of Symbols

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta$</td>
<td>Wave number</td>
</tr>
<tr>
<td>$c$</td>
<td>Wave velocity of free space</td>
</tr>
<tr>
<td>$d$</td>
<td>Depth</td>
</tr>
<tr>
<td>$E$</td>
<td>Electric</td>
</tr>
<tr>
<td>$e$</td>
<td>Reference image</td>
</tr>
<tr>
<td>$\varepsilon$</td>
<td>Permittivity</td>
</tr>
<tr>
<td>$\varepsilon_0$</td>
<td>Permittivity of free space</td>
</tr>
<tr>
<td>$\varepsilon_r$</td>
<td>Relative permittivity</td>
</tr>
<tr>
<td>$\varepsilon_s$</td>
<td>DC permittivity</td>
</tr>
<tr>
<td>$\varepsilon_\infty$</td>
<td>Permittivity at light frequencies</td>
</tr>
<tr>
<td>$f$</td>
<td>Frequency</td>
</tr>
<tr>
<td>$g$</td>
<td>Test image</td>
</tr>
<tr>
<td>$H$</td>
<td>Magnetic</td>
</tr>
<tr>
<td>$\lambda$</td>
<td>Wavelength</td>
</tr>
<tr>
<td>$m$</td>
<td>Number of data samples</td>
</tr>
<tr>
<td>$\mu$</td>
<td>Permeability</td>
</tr>
<tr>
<td>$\mu_0$</td>
<td>Permeability of free space</td>
</tr>
<tr>
<td>$\mu_r$</td>
<td>Relative permeability</td>
</tr>
</tbody>
</table>
\( n \)  
Number of GPR traces

\( \sigma \)  
Conductivity

\( \sigma_0 \)  
Conductivity of free space

\( \sigma_i \)  
Damping

\( R_s \)  
Receiver

\( T_s \)  
Transmitter

\( t \)  
Time

\( \nu \)  
Velocity

\( \nu_r \)  
Relative propagation velocity

\( \omega \)  
Angular frequency

\( \Delta l \)  
Discretisation step

\( \Delta t \)  
Time step

\( \Delta x \)  
Spatial step in \( x \) direction

\( \Delta y \)  
Spatial step in \( y \) direction

\( \Delta z \)  
Spatial step in \( z \) direction
Declaration

‘Whilst registered as a candidate for the above degree, I have not been registered for any other research award. The results and conclusions embodied in this thesis are the work of the named candidate and have not been submitted for any other academic award.’

This thesis has 30876 words.
Acknowledgements

This dissertation would not have been possible without the guidance and help of several individuals who in one way or another contributed and extended their valuable assistance in the preparation and completion of this study. First and foremost I offer my deepest gratitude to my supervisor Dr. Branislav Vuksanovic, whose encouragement, guidance, and support from the initial to the final level enabled me to develop and understanding of the subject. I also would like to show my gratitude to Dr. Alan Hewitt, my second supervisor, for his technical guidance on Image and Signal Processing, and also his support throughout this study. My sincere gratitude goes to Dr. Boris Gremont and Mr. Hassan Parchizadeh for the time and expertise contributed to the enhancement of this research.

I sincerely thank Mr. Andrew Farrar, whose enormous contributions to the measurement lab greatly facilitated the progress of my work. I would like to express my gratitude to the members of the School of Engineering of the University of Portsmouth, for giving continuous support in various ways, including the opportunities for attending relevant training workshops, seminars, and presenting my work at conferences.

Special thanks go to Prof. Dr. Amir Alani from University of Greenwich for his assistance in collecting GPR measurements, and Utsi Electronics Ltd, Erica and Vincent for their great help in training and conducting GPR
system. I offer my regards and blessing to all of those who supported me in any respect during the completion of the thesis. I am indebted to my friend Dr. Dragana Nikolic for her support and motivation.

Sincere gratitude goes to my family especially to my parents, Bostanudin and Nasria, my sister, Nurul A, my brothers, Fikri, Fauzi, and Fawwaz for their love and encouragement when it was most required. Their pride in my accomplishments guided me throughout the best and the worst moments of my work.

But, most of all to my fiancé, Sham for his love, belief in me, and constant support. His guidance, patience and understanding helped me finalise this research.
Dissemination

- **Publication:**

- **Conference Proceedings:**

- **Poster:**
Chapter 1

Introduction

Overview

This chapter describes the problems associated with GPR data and the significance of the research undertaken in analysing and improving the performance of GPR to detect and characterise objects in civil engineering structures. It outlines the objectives of the research reported in this thesis. A summary of the work that has been conducted to date concerning detection using GPR and signal processing is provided. Finally an overview of the rest of the thesis is presented.

1.1 Ground Penetrating Radar

Radar is used to detect an object based on radio waves. The theory of electromagnetism on which radar operation is based was described in detail by the Scottish physicist, James Clerk Maxwell in [1]. It is governed by four equations referred to as Maxwell’s equations. The principle of radar was found to be applicable for detecting airborne objects in the early 1900s
Introduction

which consists of transmitting electromagnetic (EM) waves and receiving the reflected signals from any object in the path of the beam [2].

Ground Penetrating Radar (GPR) is a geophysical method that uses EM wave designed primarily to locate an object that are buried beneath the earth’s surface or image the subsurface. In the recent 30 years, the technique of GPR has become an effective method for detecting shallowly buried targets [3]. It is a non-destructive and non-invasive geophysical technique with a wide variety of applications in site and ground investigations [4].

A GPR system uses the difference in the permittivity of both the target and the surrounding medium to detect a target [5]. A target is declared present if the GPR detect a local change (or discontinuity) in the media’s (such as soil) dielectric [6]. During data collection, radar antennas do not need to be in contact with the surface of the medium. This gives benefit to GPR for enabling rapid surveying as well as time consuming [7].

It is expected that the interest in, and demand for, GPR surveys will grow even more in the near future. However, there are a number of limitations which exist that are related mainly to interpretation of the results of GPR test [2]:

- the obtained images produced by the reflected signals are not photographs of the subsurface features. These images, which are plotted in colours using user-defined colour codes, represent the amplitude of the radar-reflected signals from the interfaces with
different dielectric properties. A considerable amount of experience and operator skills are therefore required to interpret subsurface radar results correctly;

- extensive amount of data;
- a prior knowledge of the dielectric properties of the material is necessary to determine the exact location of a reflecting feature beneath the surface;
- change of dielectric constant with depth mainly because of the presence of moisture;
- losses in the materials, especially due to the presence of moisture or conducting subgrade soils;
- reflections from thin layers may overlap depending on the GPR system resolution;
- insufficient dielectric contrast between layers may hinder the detection of the layers underneath.

GPR is very sensitive to changes in the EM characteristics of a medium, which gives it the ability to detect both metallic and non-metallic targets [8]. However, in a large number of situations, the signals reflected by the target are very weak and usually overlap with the clutter and noise – signals that are uncorrelated to the target scattering characteristics but occupy the same frequency band as the target [6], [8]. This causes difficulties in
distinguishing the signal from noise without applying proper signal processing techniques.

Background clutter and noise consist of three main components [9]:

(i) the cross-talk between the transmitting and receiving antennas;
(ii) the reflection from the air-ground interface;
(iii) the scattered signals from other objects within the soil.

Due to constant distance between both antennas, the cross-talk signal will arrive at the antenna receiver earlier than any other signal which makes it easier to distinguish and reject the noise component in (i) [10]. The clutter signal due to the reflection from the ground surface is usually very strong and represents a major problem in processing GPR signals. Removing signals scattered from other objects within the soil is a difficult problem as well since it cannot be effectively tackled with clutter reduction methods. Signal processing techniques have recently been applied to GPR signals in order to improve the quality of obtained data and enable easier understanding and classification of buried objects and surveyed area.

1.2 Objectives

This thesis addresses the problem of analysing and improving the performance of GPR in detecting and characterising objects in civil engineering infrastructures. Significant amount of works in this area has been attempted recently including detection of pipes, cables and
reinforcement [11], locating subsurface cavities and fractures in bedrock [12], as well as ground water and moisture [13].

This work tackles some of the issues outlined above, by developing and adapting a range of image and signal processing techniques from other disciplines into a comprehensive automated detection system that can be used effectively by untrained radar practitioners in site survey situations. The goal is to develop and apply the necessary techniques without the requirement for extensive off-site processing by a radar expert. The topics and tasks tackled in this work are outlined in Figure 1.1.

Figure 1.1: Research system implementation.
The objectives of the thesis can be summarised as follows:

1. study the process of acquiring raw GPR data to establish the properties and problems introduced during the GPR survey and the measurement process;
2. design the simulated data using GPRMax simulation program to replicate the laboratory or on-site conditions during the measurement procedure;
3. study GPR software package (ReflexW) to pre-process measured GPR data;
4. investigate subspace and data decomposition techniques to reduce clutter and unwanted signals presented in GPR data;
5. compare and evaluate the performance of the subspace and data decomposition techniques on measured data against the pre-processed data using ReflexW;
6. investigate signal modelling analysis to evaluate the performance of the subspace and data decomposition techniques;
7. identify and characterise features.

These objectives have been achieved and the author’s specific contributions can be summarised as follows:

- design and implementation of simulated data for the comparison and evaluation against measured data in reducing the amount of clutter and unwanted signals presented in GPR data;
• analysis and interpretation of pre-processing GPR data using ReflexW to be used in the evaluation of subspace and data decomposition techniques applied to GPR data;

• analysis of subspace and data decomposition techniques – Principal Component Analysis (PCA) and Independent Component Analysis (ICA) on GPR data;

• implementation and modification of traditional PCA and ICA techniques to improve their performance in image processing and GPR clutter removal tasks – combination of Singular Value Decomposition of PCA and ICA (SVDPCA) and 2DPCA, not previously suggested in GPR field;

• analysis and implementation of signal modelling techniques – Extended Prony (EP), Linear Prediction Singular Value Decomposition (LPSVD) and Matrix Pencil (MP) to evaluate the subspace and data decomposition algorithms as well as characterisation of surveyed targets;

• implementation of Singular Entropy based model order estimation method to be used with MP analysis on GPR signals;

• analysis of subspace method – PCA to characterise features.
1.3 Summary of Previous Work

1.3.1 GPR as a Detection Tool

Research in the field of GPR can broadly be divided into the categories of radar system and hardware design, geophysical inversion and modelling, and image and signal processing and analysis [7]. The application of GPR to detect and monitor cracks induced on a historical building by a landslide is discussed in [14] using 2 GHz bipolar antenna, which allows the acquisition of multi-component data. The interpretation of processed data on its profiles is in good correlation with the structure of the floor. The profiles and time slices managed to detect two different kinds of anomalies, only a few which can be due to utilities and the metallic mesh, whilst others were associated with cracks induced by the landslide.

A processing chain for the spatial analysis of the data recorded by GPR is presented in [15]. In particular, the detection and localisation of pipes is implemented by exploiting the a priori knowledge that a buried cylinder gives rise to a hyperbolic signature in GPR images. The image interpretation is performed by a suitably trained simple neural detector after some pre-processing steps aiming towards the enhancement of the buried objects’ signatures.
1.3.2 GPR Technology in Civil Infrastructures

Interest in radar as a possible non-destructive evaluation tool in civil engineering has developed and various researchers have shown interest in pavement structure. Non-Destructive Crack Depth (NDCD) detector developed by Utsi Electronics in [16], which is based on a slow speed GPR, was used to determine crack depth in flexible pavements and road condition monitoring. However, false readings can be made if there are other linear objects/features in the layer, thus it is advisable to take more than one run over a crack to guarantee the detection.

Measurement of reinforcement bar depths and diameters in concrete is discussed in [17] where the results of controlled measurements using 4 GHz GPR system are compared to the simulated data by GPRMax3D of a range of rebar sizes at different depths. In-build rebar measurements were done to investigate the depth, position, and diameter of rebars in a series of ceiling slabs.

1.3.3 Signal Processing for GPR

The difficulty in reducing clutter presented in GPR data is an important problem in many areas. A number of clutter reduction approaches suggested in the literature cover simple mean scan subtraction that was discussed in [18], two-dimensional digital filtering [19], wavelet packet decomposition [18], likelihood ratio test [10], [18], [20], parametric system identification
Introduction

[18], [21], and Kalman filter [20], [22]. Most of these methods depend on the background signal estimation [10] by taking the mean value of the unprocessed ensemble collected GPR data, followed by employing the simple mean scan subtraction. A-scan far away from the target location might be used instead of mean scan subtraction. Although these methods have been used widely in GPR applications, the direct wave from the received data cannot be completely removed due to the air-earth interface and time-shift (jitter) of the radar system.

Various researchers have shown interest in subspace techniques including Singular Value Decomposition (SVD) [18], [23], Linear Discriminant Analysis (LDA) [24], Principal Component Analysis (PCA) [25], [26], and Independent Component Analysis (ICA) [6], [25], [27], [28]. These techniques are effective in reducing the higher-order statistical dependencies and established in yielding performance improvements in clutter reduction. The returns are divided into clutter and landmine signals by selecting reasonable principal components (PCs) and independent components (ICs) which are used to determine the subspace for reconstruction.

Detection and identification of buried landmines is an important, yet difficult, task [29]. The backscattered echoes in GPR data, which are usually distorted by multiple scattering are used for target identification [8], [30]. However, a serious drawback of a GPR system relies in the limitation of its
ability in discriminating between landmine and non-landmine echoes. Thus, this leads in need of advanced signal processing and target classification techniques. Several algorithms have been introduced in classification, characterisation, and recognition of GPR and other areas signals such as Prony Analysis (PA) [29], [31], [32], Matrix Pencil (MP) [33], [34], and Linear Prediction Singular Value Decomposition (LPSVD) [35].

Prony proposed fitting a sum of exponentials to equally spaced data points and extended the model to interpolate at intermediate points [31]. In [29], the backscattered GPR echoes is modelled separately from a landmine target as a linear combination of exponential functions. The complex parameters (such as poles) are expected to be characteristic for each target-type, and invariant with respect to target orientation. However, as the Prony Analysis method is based on linear least-squares, it is highly sensitive to the additive measurement noise. Thus, LPSVD method which is known as Kumaresan and Tufts [35] is introduced that offers an improved performance in the application of signals with lower signal-to-noise ratio (SNR). This technique uses backward linear prediction and low-rank data matrix approximation based on SVD.

Another approach that can be used to extract poles of complex frequencies is Matrix Pencil [33] which is described to be more robust in the presence of noise in the sampled data and computationally more efficient. It offers a lower statistical variance of the estimated parameters than a
polynomial-type method (such as PA). This technique is explained in [36], in which the number of spurious poles are reduced by using the dominant eigenvalues in estimating the model number since underestimating its value result in errors whilst overestimating it generates both the correct and spurious poles.

1.4 Thesis Structure

The thesis is organised as follows:

**Chapter 2** Ground Penetrating Radar with its principles is reviewed. The chapter looks at the important principles in conducting GPR survey in field experiments, as well as modelling GPR data using dedicated GPR simulation package.

**Chapter 3** Commercially available professional GPR software – ReflexW is discussed and used to process measured GPR data in order to reduce the amount of clutter and noise present in the data. The results of this pre-processing are used for comparison with the results achieved using signal pre-processing techniques developed in this work in order to evaluate the performance of each method in clutter reduction.

**Chapter 4** In order to tackle the problem of clutter in measured GPR images, multivariate data analysis techniques are studied in detail in this chapter. Those techniques are then applied to signals measured on-site and in laboratory as described in Chapter 2. The main algorithms modified and
applied in this section are: one-dimensional (1D) and two-dimensional (2D) Principal Component Analysis (PCA), Independent Component Analysis (ICA), and a combination of Principal and Independent Component Analysis (SVDPICA). Whilst both PCA and ICA have recently been applied and tested on GPR images, all of the work has been done in landmine detection type tasks. In addition to that, two-dimensional PCA as well as the SVDPICA algorithms have not been proposed for GPR data processing before. The problems of evaluating the performance of each individual algorithm is another issue related to the work done in this chapter. Whilst in most of image or signal processing tasks “ideal” image of “noiseless” signal is available to accurately assess the performance of the applied technique, it does not apply in GPR area. Therefore, high signal-to-noise ratio can imply the significant reduction of clutter and noise present in the raw image but can also hide the fact that some of the important features present in the measured image and related to buried target might have been removed from the image during the applied processing method.

Chapter 5 Signal processing methods used for the analysis and characterisation of GPR signals are studied and applied in this section. Extended Prony, Linear Prediction Singular Value Decomposition, and Matrix Pencil techniques are related but significantly different methods. In this chapter those methods are introduced and utilised to analyse pre-processed GPR signals. Accuracy of estimated signal parameters including
frequency, damping, amplitude, and phase is assessed and compared for various situations and underground objects where large amount of clutter is present in the images.

**Chapter 6** Conclusions and the summary of the performed work and results are given and some recommendations for further work are presented in the last chapter of the thesis.
Chapter 2

Ground Penetrating Radar

Overview

This chapter explains the principles of Ground Penetrating Radar and its electromagnetic properties. The issues of GPR measurements despite of its advantages have been highlighted. GPR simulation software package is introduced and simulated data is compared to the measured experimental data.

2.1 Introduction

Application of Ground Penetrating Radar (GPR) is a rapidly growing field that has seen significant progress in the development of theory, techniques and technology over the past 15 – 20 years [37]. It is a non-destructive method that generates EM pulses to record, locate and evaluate the depth of buried objects or subsurface features that cannot be seen visually [38].

GPR system normally consists of data collection unit, transmitting and receiving antennas. The system that contains transmit and receive functions performed by the same antenna is known as mono-static, otherwise it is
called bi-static. GPR systems typically have both antennas combined in a single housing that makes the separation between the two antennas is most often fixed [39]. Most common elements of GPR system depicted in Figure 2.1 are:

- a display unit that provides a ‘real time’ display of the recorded data such as laptop computer or a dedicated control unit (i.e. data logger);
- a control unit that generates EM pulses and records the received energy;
- a survey cart with some form of positioning device (GPS receiver or odometer) or, alternatively, a sled, frame or other device that allows the antennas to be pulled across the ground manually;
- a battery pack (usually 12V) to power the unit;
- an antenna housing containing antenna source electronics and two or more antenna units – transmitter and receiver(s).

Figure 2.1: Typical GPR system.
The operation of a GPR device is based upon the propagation and reflection characteristics of an EM wave or pulse in material. An EM wave is reflected at positions where the penetrated material changes which caused by a change of the EM properties in the subsurface; the permittivity ($\varepsilon$), the permeability ($\mu$) and electrical conductivity ($\sigma$) [13], [40]. However, not all these three parameters provide useful information to the GPR [41].

The conductivity generally affects the penetration of the GPR due to absorption of the radar signals in the medium. Any medium with high moisture content increases the electrical conductivity, thus decreasing penetration. The electrical conductivity of most engineering materials is usually low and the magnetic permeability is often negligible for materials not rich in iron-oxide [13]. Therefore, due to the lack of magnetic content in the earth’s soil, the relative permeability hardly provides any useful information because it offers little contrast in the radiated EM pulses. The relative permittivity which corresponds to the dielectric constant of the medium provides the highest degree of contrast in the reflected wave. This resulted in a good characterisation of the ground and therefore leads to the reflection in the EM pulse [41].

GPR operates in a manner similar to navigational radar systems but it sends pulses of EM waves into the ground in order to identify the shape, size, and locations of subsurface features. In general, GPR is mostly used in
reflection mode where a signal is emitted via a transmitter antenna into the structures and soil under investigations as depicted in Figure 2.2 [42].

Figure 2.2: Basic functional principle of a GPR device.

When the transmitted EM wave encounter changes in subsurface materials, the properties of the wave are altered, and part of the wave is reflected back to the surface, where data on its amplitude, wavelength, and two-way travel time are collected for analysis and interpreted accordingly [43]. The control device measures the amplitude of the reflected energy for each position at different times. The amplitude and phase of the “echoes” can give information about the scattering properties of the target whilst the
time of arrival of different pulses gives indirect distance indications to the subsurface layers of reflectors [7]. This process produces a data matrix with columns containing the reflected amplitudes at certain times for each measurement position along the trace. GPR data assembled in this way can therefore be displayed as a two-dimensional image [40].

Figure 2.3: Principles of GPR in CMP mode which the transmitting and receiving antennas are moved away from each other [44].

A common mid-point dataset (CMP) [44] is another mode that can sometimes be used for GPR measurements. It is also called a velocity sounding, since the technique is commonly used for signal velocity establishment. In CMP mode, the antenna separation is increased for each recording whilst keeping over a common mid-point as in Figure 2.3. This mode of GPR measurements was not used in this project and will therefore not be discussed in more details in this work.

The performance of GPR depends upon the surface and subsurface conditions and its specifications. These include requirements and information about reflections, depth of investigation, and resolution. The
ability of a GPR to detect objects depends on the wavelength of the input signal, so the quality of the image improves as the wavelength decreases and the frequency increases [41]. At high frequencies, the resolution of GPR system therefore increases but depth of penetration decreases and vice versa for lower frequencies.

The detection of the target however becomes difficult in the situations where [8]:

- the target is buried close to the surface of the ground;
- the target has very small geometrical dimensions;
- the object has a small permittivity difference to that of the ground.

In these entire situations, the reflected signal of the target is very weak and is usually overlapped with the clutter – represents any unwanted reflections present in the GPR signal, making it difficult to distinguish between both signals and noise – caused by interference from other radio-wave-emitting devices that are present in the surveyed area during data collection, without applying proper signal processing techniques. The data received from GPR system is essentially consist of [41]:

(a) coupling induced by transmitting and receiving antennas;

(b) reflection by ground surfaces;

(c) stochastic EM, interference either from GPR system itself or from environment;

(d) reflection by subsurface targets;
(e) dispersion or reflection by the underground inhomogeneities.

The signals (a) and (b) are usually known as “clutter”, which are much more intense than the buried-target reflection due to short distance from transmitting to receiving antennas and the attenuation of the target reflection by underground medium like soil, etc. The need for deploying proper clutter reduction methods is therefore necessary in order to enable effective and accurate target detection and identification.

2.2 Advantages and Disadvantages of GPR

The use of GPR as a geophysical method of site investigation has several beneficial advantages compared to other subsurface sensing technologies such as [23], [38], [41], [43]:

- non-destructive nature;
- its ability to maximise research efficiency and minimise the cost;
- its capability to cover large areas quickly;
- the capability in detecting both metallic and non-metallic targets.

The graphic displays of GPR data often resemble geologic cross sections. This gives benefit in protecting and preserving particularly in archaeology sites where cultural sensitive features (such as human burials) are concerned. Indeed, its ability to survey an area in front of it, unlike other sensors that can only survey an area directly beneath them, gives the potential not only to identify these culturally sensitive area, but also in
detecting dangerous objects before the system moves over and past them [43].

In comparison to other subsurface sensing technologies, GPR has the potential to increase research efficiency by conducting surveys quickly at relatively low cost as well as covering larger areas. The antennas of GPR itself may be pulled by hand or with a vehicle which can produce considerable data/unit time that makes the GPR measurements relatively easy to make [43], [45].

Despite of its advantages described earlier, GPR has some limiting factors [23], [38], [41]:

- its usage in inhomegeneous soil can increases the number of false alarm and the depth of signal penetration is limited. A trade-off between resolution of the returned signal and penetration depth is therefore required;
- GPR method is sensitive to unwanted signals (noise) caused by various geologic (e.g., boulders, tree roots) and cultural factors (i.e. nearby vehicles, buildings, etc.). EM transmission from cellular phones, two-way radios, television, and microwave transmitters may cause noise on GPR records;
- the data collected through GPR methods is highly subjective to the interpretation, especially if interferences are not identified correctly.
It is therefore important to define clearly the target parameters as there is a considerable difference between the target response from different situations such as buried pipe, buried mine, a void, or a planar interface that has a major impact on signal processing strategy. The resolution and the required depth of penetration should be clearly identified in order to set the frequency and bandwidth of operation, which will in turn influence the choice of modulation technique and hence the hardware design [13].

2.3 Basic Principles of GPR

The outcome and quality of raw data obtained from GPR survey is primarily dependent upon two factors [14], [43], [45]:

- data acquisition parameters which control the waveform transmitted into the ground;
- characteristics of the materials through which the EM wave propagates.

To set the data acquisition parameters properly, it is important to understand the properties of the propagating medium and target objects properly. Important fundamental EM properties of materials are discussed in the following section.
2.3.1 Electromagnetic Properties in Dielectric Materials

The interactions of EM waves with physical media can be quite complex and the most exact models known for EM interactions use quantum mechanics [46]. The approximation to the quantum representation is given by Maxwell’s equations describing the EM interactions in terms of electric $(E)$ and magnetic $(H)$ field vectors that propagate and reflect as waves [47]. Often, the propagation of EM waves can be represented by scalars instead of vectors, which simplifies the mathematics at the expense of a loss directional information, such as polarisation.

2.3.1.1 Permeability

Magnetic permeability $(\mu)$ [39], [45], [48] describes the ability of a material to become magnetised in the presence of an EM field. It is a measure of the magnetic polarisation of a material that measured in henries per metre (H/m). Materials that are more magnetically permeable will more readily interfere with the magnetic part of the EM wave, thereby attenuating the wave and resulting in shallow subsurface images. The relative permeability, $\mu_r$ of a material is given as,

$$\mu_r = \frac{\mu}{\mu_0} \quad (2.1)$$

where $\mu_0$ represents the permeability of free space and is given as,

$$\mu_0 = 4\pi \times 10^{-7} \approx 1.256 \times 10^{-6} \text{ H/m} \quad (2.2)$$
in which left the relative permeability with a unit-less quantity.

Ferromagnetic materials \((\mu \gg 1)\) are considered to be magnetically lossy and may have a frequency dependent permeability [39]. Asphalt and concrete are however considered to be non-ferromagnetic materials \((\mu \approx 1)\) thus it can be assumed that the permeability of these materials is the same as for the space \((\mu = 1)\).

### 2.3.1.2 Electrical Conductivity

Electrical conductivity \((\sigma)\) describes the ability of a material to conduct the electric portion of the EM wave and is measured in Siemens per metre \((\text{S/m})\). Materials which are more electrically conductive will more readily conduct the electric part of the EM wave, thereby dissipating or attenuating the wave and resulting in shallow subsurface images. Conversely, materials with a low electrical conductivity will allow greater depth of EM wave permeation [45], [48].

Since GPR is an EM energy, it is subject to attenuation (natural absorption) as it moves through a material [49]. The signal is able to penetrate a great deal of material if the energy is moving through a resistive (low conductivity) material such as very dry sand, ice, or dry concrete which makes the signal stays intact longer, thus being able to go further into the material. However, if the material is conductive such as salt water and
wet concrete, the GPR energy will get absorbed before it has had the chance to go very far into the material. As a rule of thumb based on this electrical conductivity, the greater the water content of the material, the greater the conductivity.

2.3.1.3 Permittivity

Permittivity ($\varepsilon$) [48], [49] describes the ability of a material to store and transmit an electric charge induced by an EM field that measured in Farads per metre ($\text{F/m}$). It is a descriptive number that indicates how fast radar energy travels through a material. Radar energy will always move as quickly as possible through a material, but certain materials slow down the energy more than others. The speed moving GPR energy can be deduced by knowing the value of permittivity, e.g. the higher the permittivity, the slower the radar wave moves through the medium. The wet material will for example slow down the radar signals due to the water presence causing the raise of the overall dielectric of the material.

The permittivity of free space ($\varepsilon_0$) is assumed to be the same as in a vacuum which is,

$$\varepsilon_0 = \frac{1}{c^2 \mu_0} \approx 8.854 \times 10^{-12} \text{ F/m}$$

(2.3)

The relative permittivity ($\varepsilon_r$) of a material is therefore can be considered as in equation (2.4) and is unit-less.
\[ \varepsilon_r = \varepsilon / \varepsilon_0 \]  

(2.4)

2.3.2 Data Collection Parameters

There are numerous other parameters that can be varied and controlled to achieve the targeted depth and resolution during a GPR survey. These parameters are described in the following section.

2.3.2.1 Wave velocity

The angular frequency \( \omega \) [39] of an EM wave is measured in radians per second \( \text{rads/s} \) and is proportional to its frequency, \( f \) measured in Hertz:

\[ \omega = 2\pi f \]  

(2.5)

The wave number, or propagation constant, \( \beta \) is measured in radians per metre and is given by [50]:

\[ \beta = \omega \sqrt{\mu \varepsilon / 2} \left[ \sqrt{1 + \left( \frac{\sigma}{\omega \varepsilon} \right)^2} + 1 \right] \]  

(2.6)

For the material with no conductivity, equation (2.6) can be reduced to:

\[ \beta = \omega \sqrt{\mu \varepsilon} \]  

(2.7)

The velocity at which a wave travels through a material \( v \) is measured in metres per second and is given by [50]:

\[ v = \frac{\omega}{\beta} \]  

(2.8)
The wave velocity of free space \((c)\) can be obtained from the equations (2.7) and (2.8) respectively [39]:

\[
c = \frac{\omega}{\beta_0} = \frac{\omega}{\omega \sqrt{\mu_0 \varepsilon_0}} = \frac{1}{\sqrt{\mu_0 \varepsilon_0}} \approx 3 \times 10^8 \text{ m/s} \tag{2.9}
\]

For materials with no conductivity \((\sigma = 0)\) and permeability equals to that of free space \((\mu_r = 1)\), equation (2.9) can be stated as:

\[
\nu_r = \frac{\omega}{\beta} = \frac{\frac{1}{\sqrt{\mu_r \mu_0 \varepsilon_r \varepsilon_0}}}{\sqrt{\varepsilon_r}} = c \tag{2.10}
\]

2.3.2.2 Wavelength

Wavelength \((\lambda)\) [39] can be described as a propagating wave which repeats itself at a particular distance and is measured in metres:

\[
\lambda = \frac{2\pi}{\beta} = \frac{2\pi \sqrt{\mu_r \mu_0 \varepsilon_r \varepsilon_0}}{\omega} = \frac{\nu}{f} \tag{2.11}
\]

GPR antennas can typically be distinguished when moving along the concrete features aligned perpendicular to the direction of travel if they are spaced at least one half of a wavelength apart \((\lambda/2)\). Features that are stacked on top of each other can typically be distinguished if they are at least a quarter of a wavelength apart \((\lambda/4)\) [49]. However, features that are closer to each other than these distances may appear as a single feature in GPR data.
2.3.2.3 Relationship between Frequency and Depth

Choosing a frequency for a GPR survey is quite critical. Lower frequency with long wavelengths provide the deepest penetration, whereas high frequency with short wavelengths are only able to image shallow features [45], [51], [52].

2.3.2.4 Relationship between Frequency and Resolution

Resolution of subsurface features is in part affected by antenna wavelength which is also directly related to the frequency. Higher frequency radar provides higher resolution than lower frequency radar [45], [51], [52], [53]. The shorter wavelengths of high frequency produce a narrower cone of transmission and give the ability to focus on smaller areas, thereby resolve smaller features. This is differs to lower frequency radar which has more spread out transmission cones due to longer wavelengths [45].

2.3.3 Determination of Layer Thickness and Feature Depth

An absolute measurement of depth or thickness can easily be recognised if the propagation velocity can be measured, or derived. For homogeneous and isotropic material, the relative propagation velocity (in metre per second), \( v_r \), can be obtained as in equation (2.10) [13].

The data recorded from a GPR system are voltages of the received signal versus time [39]. If the velocity of propagation through a material is known,
the depth (in metre) to an object \( d \) can be calculated from the recorded two-way travel time \( t \). Notice that since the wave travels from the transmitter to the object and back to the receiver, the two-way travel time is halved to obtain the actual time to the feature. The layer thickness is therefore can be computed as [13]:

\[
    d = \frac{\nu_r t}{2}
\]  

(2.12)

2.3.4 Electromagnetic Wave Physics

2.3.4.1 Target Reflection and Feature Identification

A typical GPR section has distance along the top and two-way travel time down the side [54]. At each position in the survey, the recorded trace is displayed as a coloured or grey-scale strip of pixels with the entire section being built up from each of these traces. The layers of GPR recorded trace show up as coloured or black-white-black lines (reflections) and isolated ‘blocky’ or smaller-scaled features (such as pipes, foundations, etc.) produce curved, hyperbolic-shaped features (called diffraction hyperbola). When the antenna crosses a target at a right angle, the resulting image looks like an inverted U or V – a hyperbola is the descriptive term for its shape as depicted in Figure 2.4 (b).
Figure 2.4: Linescan and oscilloscope displays (right) show the (a) direct coupling and (b) reflections and diffraction hyperbola of GPR data scanned over the plywood top panel of the pipe in a box.
The very first signal in GPR scan is often called “direct coupling” between transmitter and receiver and is used to identify the surface position. In Figure 2.4 (a), direct coupling looks like a straight horizontal band on top of the data window. It is a combination of the transmit pulse in the air and surface reflection from the top of the material (plywood). It arrives at the receiver before the signal penetrates into the material, thus direct coupling carries little information about the structure [49].

Hyperbola is a form of point source reflection that is due to the fact that GPR energy is emitted in a cone, which radiates outwards with depth [43]. As such, energy is reflected from objects that are not directly below the antenna; the reflections, however, is recorded as being directly below the antenna, and at a greater depth due to the oblique transmission of the wave. The hyperbola shape happens because the radiated antenna beam has the shape of a wide cone, thus the radar can see the target not only when on top of it, but also in several scans before and after that position. The apex of the hyperbola denotes the actual location of the target. The shape of a hyperbola depends on two parameters [49]:

- scan spacing: smaller scan spacing (more scans per inch/cm) produces wider hyperbolas;
- radar wave velocity: higher velocity (lower dielectric) produces wider hyperbolas.
Target of larger diameter produce brighter reflections. The shape of hyperbola does not change significantly with target size for any diameter less than 2 inches – all such targets are point-like for the radar as their size is a fraction of the wavelength. The target size can be roughly estimated from the width of the hyperbola’s flat top. Therefore, any targets below 2 inches in diameter will produce hyperbolas of the same size and shape [49].

2.3.4.2 Background Noise and Clutter

Background noise [13], [43] termed clutter in case of a GPR can be defined as those signals that are unrelated to the target scattering characteristics but occur in the same sample time window and have similar spectral characteristics to the target wavelength. It refers to point targets and small discontinuities that reflect energy and obscure the signals of other more important reflected waves. The clutter can be caused by breakthrough between the transmitter and receiver antennas as well as multiple reflections between the antenna and the ground surface. Local variations in the characteristic impedance of the ground can also cause clutter, as well as inclusions of groups of small reflection sources within the material. The background noise may also be generated by other nearby sources of EM waves, including televisions, cell phones, and radio transmission antennas [45], [55].
2.3.5 Distance Normalisation

In GPR, the position of the targets should be given in relation to surrounding objects or to a reference grid on the ground surface. The acquired data can be of any practical use if there is enough spatial positioning information exists in the data to correlate any detected features with these surface coordinates [7].

![Figure 2.5: Collecting the position of cracks in practical approach with measuring tape.](image)

The most common method for incorporating this spatial positioning information into the radar file, considering the radar operating system in continuous mode is by placing event markers in the data when the antenna covers a given lateral distance [7]. Practically, these markers are often introduced by a manual switch or button controlled by the radar operator, with the distance calibrated by means of a measuring tape, regularly-spaced survey pegs or spray markers, or in certain circumstances by human
estimation. The yellow dots depicted in Figure 2.4 (b) represent the markers position being taken during the measurements whilst Figure 2.5 illustrates the situation where measuring tape is being used in practice.

2.4 Types of Data Collection

The objective of GPR data presentation is to provide a visual display of the processed data that closely approximates an image of the subsurface, with the anomalies that are associated with the objects of interest located in their proper spatial positions [41]. Data display is a key to data interpretation and in fact, producing a good display is an integral part of interpretation. This section explains three types of data representations used in GPR for a better understanding of the terminology [9], [39], [41], [47].

(a) A-scan

A-scan (or one-dimensional data presentation) is a time-amplitude plot that is obtained by a stationary measurement, emission, and collection of a signal after placing the antenna above the position of interest.

(b) B-scan

Most of GPR data analysis is based on interpretation of a series of GPR signals presented as a B-scan. It is a two-dimensional dataset or image obtained as the horizontal collection from the ensemble of A-scans. The horizontal axis of the two-dimensional image is the
surface position, and the vertical axis is the round-trip marvel time of the EM wave. A matrix is then performed where each row represents a sample point (or time) and column represents a trace. The value of the element of the matrix is the amplitude of the voltage for the associated trace and sample point. Assigning intensity to the amplitude of the voltage allows the matrix to be displayed as an image.

(c) C-scan

C-scan (or three-dimensional data presentation) signal is obtained from the ensemble of B-scans, measured by repeated line scans along the plane. Three-dimensional displays are fundamentally block views of GPR traces that are recorded at different positions on the surface. Obtaining good three-dimensional images are very useful for interpreting specific targets. Targets of interest are generally easier to identify and isolate on three-dimensional datasets than on conventional two-dimensional profile lines.
Figure 2.6: GPR measurement taken across a pipe in a box of (a) one-dimensional and (b) two-dimensional GPR data. The one dimensional data is taken at the distance of 0.4m in two-dimensional image. (c) Three-dimensional GPR data represents an empty box.

2.5 Types of GPR System

The School of Engineering at the University of Portsmouth has purchased a number of GPRs designed and developed by Utsi Electronics Ltd. Some of these devices have been utilised to perform laboratory experiments. Two types of radars used in this work are:
(a) Groundvue 3 (GV3)

It is used with a dual bowtie antenna pair. There are two types of frequency used during the measurements which are 1 GHz radar system (transmitter and receiver) and 1.5 GHz Crack Detection Head (CDH). The CDH is specifically developed for the detection and measurement of vertical cracks in concrete and asphalt/tarmac. The operating depth (measured in travel time) and the distance between samples must be defined accurately whilst other parameters can be reset after the survey. The depth range for this radar is from 0 – 5 metres in dry conditions, and a little less in water logged grounds. Figure 2.8 shows the GPR of GV3 system.

(b) Groundvue 5 (GV5)

It is a 4 GHz horn antenna system with high resolution capability. Similar to GV3, the operating depth and distance between samples must be defined clearly. The GV5 can be used on its own as one (or more) of the four channels of the multi-channel GV3.

Figure 2.7: GPR Groundvue 5.
Figure 2.8: GPR Groundvue 3 with 400 MHz, 1 GHz, and Crack Detection Head antenna systems.
Most of GPR systems operate in the frequency range from DC to the highest frequency, which can reach to the buried objects. GPR is therefore a very wide frequency system, and its centre frequency is referred as its operational frequency. For concrete inspection, GPR is typically operates around 1 – 1.5 GHz, where the object is 1 cm large and depth is 20 – 50 cm.

2.6 GPR Data

Two types of data have been used in determining the effectiveness of the proposed algorithms, simulated and experimental (measured) data. GPRMax software has been used in simulating the GPR data whilst three types of experimental data have been obtained in three different sets of measurements.

2.6.1 Simulated Data

In this work, the presented GPR simulations have been computed in two-dimensional only via Finite-Difference Time-Domain (FDTD) [56], [57] algorithm implemented via GPRMax2D. It is an EM wave simulator for GPR modelling that is based on the FDTD numerical method [58], [59].

The general structure of an input file of the GPRMax software is a plain American Standard Code for Information Interchange (ASCII) text file which can be prepared with any editor or word-processing program. The software used the hash character (#) to denote the beginning of a command
with its parameters occupied in a single line of the input file. Other characters can be used to comment or describe the input file. In the presence of misspelled command, the program will be terminated and hence issued an error message. A white space character is used to separate a command that requires more than one parameter [58].

2.6.1.1 Basic Concepts of GPR Modelling

The GPR response from a particular target or set of targets can be simulated by solving the Maxwell’s equations subject to the geometry of the problem and the initial condition. The initial condition (i.e. the excitation of the GPR transmitting antenna) is defined to avoid the propagation through space reaching a zero value at infinity. This is due to the reason of no specific boundary limits the problem’s geometry and where the EM fields can take a pre-determined value. The nature of the GPR forward problem can therefore be classified as an initial value – open boundary problem [58].

In general, the FDTD approach to the numerical solution of Maxwell’s equations is to discretise both the space and time continua [58]. The discretisation spatial $\Delta x$, $\Delta y$, and $\Delta z$, and temporal $\Delta t$ steps are therefore play a significant role since the smaller they are, the closer the FDTD model is to a real representation of the problem. However, the values of the discretisation steps always have to be finite since computers have a limited amount of storage and finite processing speed. Hence, the FDTD model represents a discretised version of the real problem and of limited size. The
building block of the discretised FDTD grid is called Yee cell [60] name after Kane Yee who pioneered the FDTD method.

A FDTD simulation starts by defining the structure to be modelled and identifying the range of wavelengths for which to compute a response. A grid is defined based on the grid spacing which must be small enough to sufficiently resolve the shortest wavelength of interest (typically $\lambda/10$ to $\lambda/30$) and resolve the finest feature of the structure [61].

The numerical solution is obtained in an iterative fashion where the EM fields propagate in the FDTD grid. The FDTD simulated the fields for a given time window by specifying the number of iterations and each iteration corresponds to an elapsed simulated time of one $\Delta t$. Energy can never travel farther than one grid spacing in a single time step due to the manner in which the fields are updated. The time step is therefore should be small enough to accurately portray physical fields. The values of $\Delta x$, $\Delta y$, $\Delta z$, and $\Delta t$ are therefore can not be assigned independently due to its conditionally stable numerical process that is known as the CFL condition, which is after the initials of Courant, Freidrichs, and Lewy, and is given by [58], [61]:

$$\Delta t \leq \frac{1}{c \sqrt{\frac{1}{(\Delta x)^2} + \frac{1}{(\Delta y)^2} + \frac{1}{(\Delta z)^2}}}$$ (2.13)
where \( c \) is the speed of light. Hence, \( \Delta t \) is bounded by the values of \( \Delta x \), \( \Delta y \), and \( \Delta z \). The stability condition for the two-dimensional case is easily obtained by letting \( \Delta z \to \infty \).

As stated previously, the excitation of the GPR transmitting antenna should be defined first to terminate the computational region at a finite distance from sources and targets. An approximate condition known as Absorbing Boundary Condition (ABC) is introduced at a sufficient distance from the source to truncate and limit the computational space by absorbing any waves impinging on it. This program used ABC as its default, however, when modelling more than six layers in the GPR, Perfectly Matched Layer (PML) – a layer of non-physical absorbing material is advisable to be used due to its superior performance compare to ABC. Figure 2.9 illustrates the basic difference between the problem to be modelled using ABC and the actual FDTD modelled space [58].

![Figure 2.9: The two-dimensional GPR forward problem and its GPRMax2D domain bounded by ABC.](image)
2.6.1.2 Coordinate System of GPRMax2D

The GPRMax2D FDTD algorithm is implemented in the \((x-y)\) plane where the origin of the coordinate system is the lower left corner at \((0,0)\).

The smallest space that can be allocated to represent a specific medium is a 2D cell \((\Delta x \times \Delta y)\) with the reference point at its centre. The space coordinates are however range from the left edge of the first cell to the right edge of the last one as depicted in Figure 2.10. For a given set of space coordinates \((x, y)\), the actual positions of the EM field components are [58]:

- \(\left( x + \frac{\Delta x}{2}, y + \frac{\Delta y}{2} \right)\) for \(E_z\);
- \(\left( x + \frac{\Delta x}{2}, y \right)\) for \(H_x\);
- \(\left( x, y + \frac{\Delta y}{2} \right)\) for \(H_y\);

which are due to the staggered arrangement of field components in the FDTD algorithm. Therefore, the interference between two cells of different constitutive parameters is located on the positions of the magnetic \((H)\) field components and all sources are actually located at the positions of the electric \((E)\) field component.
2.6.1.3 GPR Modelling with GPRMax2D

There are two important points should be concerned when modelling the GPR with GPRMax2D. This is to cover the most fundamental aspects of using FDTD based program as well as avoiding the common errors which are [58]:

- discretisation;
- Absorbing Boundary Condition (ABC).

In general, discretisation depends on the required accuracy, the frequency content of the source’s pulse, and the size of the targets [58]. Errors associated with numerical induced dispersion are another factor that influences discretisation. This is opposing to the real world where EM waves propagate with the same velocity irrespectively of their direction and frequency (assuming no dispersive media and far-field conditions). This error [56], [59] can be kept minimum by following the rule of thumb which
the discretisation step should be at least ten times smaller than the smallest wavelength of the propagating EM fields and is described as:

\[ \Delta l = \frac{\lambda}{10} \]  \hspace{1cm} (2.14)

The ABCs employed in GPRMax2D perform well (i.e., without introducing significant artificial reflections) if all sources and targets are kept at least 15 cells away from them. Although the PML absorbing boundaries has a user adjustable thickness, it absorbs very efficient most waves that propagate in them. Placing a source and output in these layers should be avoided to obtain a correct GPR modelling. These requirements should be considered when deciding the size of the model and free space (i.e. air) should be always included above a source for at least 15 to 20 cells for the same reason [58].

2.6.1.4 General Structure of GPRMax2D Input File

Commands

There are some principles in modelling GPR to describe the GPRMax2D commands and their parameters [58]. Parameters associated with simulated space and spatial coordinates should be specified in metres. Time and frequency parameters are however identified in seconds and Hertz.

An example of input file in constructing a model by using GPRMax2D is depicted in Figure 2.11. The output is illustrated in Figure 2.12 together with an image representation of the model’s space constructed using the
information stored in the geometry file. Based on Figure 2.12 of GPR trace, the GPR receiving antenna records amplitude at a particular sampling rate, for a particular time duration. GPRMax2D program converts spatial and temporal parameters given in metres and seconds respectively to integer values corresponding to FDTD cell coordinates and iteration numbers. The time between sample points can be identified using this information which 2827 samples were generated in 10 ns (the value of time_window in the input file given in Figure 2.11) yielding a time between samples of approximately 3.5373 ps. Time scale is referred to as two-way travel time taken to transmit and reflect from the target. This information should be considered when computing the distance between features [39].
Ground Penetrating Radar

Figure 2.11: Example of GPRMax2D input file in simulating concrete slab with joints using 1.5 GHz frequency radar. Red arrows represent lines that were deactivated and ignored by the program as the first character of each command starts without the hash character (#).
(a) **Geometry image representation of the model**

The free space above the source is included in the model according to ABC rules.

(b) **Simulated GPR scan of the model**

The GPRMax2D converts spatial and temporal parameters to integer values corresponding to FDTD cell coordinates and iteration numbers.

(c) **A-scan representation**

The A-scan represents the GPR trace number 60 of the GPR scan depicted in (b).

Figure 2.12: Example of GPRMax2D model.
2.6.2 Experimental (Measured) Data

There are three types of experimental data used for this research which was prepared by:

- School of Engineering, University of Portsmouth – to evaluate the performance of GPR GV5 system in detecting and characterising three PVC pipes samples of different diameters with and without the presence of sand;

- Department of Mechanical and Design Engineering, University of Portsmouth and Utsi Electronics Ltd – to investigate the potential and applicability of GPR GV5 and GV3 CDH systems in detecting cracks and defect within their water reservoirs and infrastructure (floor slab and slab joints);

- School of Engineering, University of Greenwich – to assess the performance of GPR GV5 system in detecting cracks on a concrete metal fibre mixtures slab.

The description and physical model of each experiment is illustrated and discussed in the subsequent chapters.

2.6.2.1 PVC Pipes Detection

The experiment was performed on a prototype laboratory model that was carefully designed and constructed with the intention of detecting pipes with different diameters. Two open-top wooden container boxes were
constructed using a minimum amount of metals (nails, screws, etc.) to prevent metal interference with the radar signal. The construction of the container boxes and the preparation of these models for experimental purposes are described in Figure 2.13.

![Figure 2.13: (a) Experimental setup used in collecting GPR data for pipes detection, (b) sand was placed in the small container box, (c) top panel with grid drawn on it, and (d) position of three pipes in the container box.](image)

The container boxes were made in two different sizes from 2.5 cm plywood measuring $1.23 \text{ m} \times 1.23 \text{ m} \times 0.175 \text{ m}$ (small) and
Ground Penetrating Radar

1.27 m × 1.27 m × 0.265 m (big) sitting on four supporting beams with measurements of 0.07 m × 0.095 m × 1.3 m. PVC pond liner material was used to line both boxes that were securely stapled into position. In order to carry out the GPR measurements, plywood measuring 1.4 m × 1 m × 0.018 m was placed on top of the constructed container boxes and scanned lines along the 10 cm grid drawn on the top panel.

Three different sizes of PVC pipes with thickness of 0.2 cm, which were sealed with tape at each end, were used during the GPR measurements. The dimensions of the pipes are given in Table 2.1. In the presence of sand, the PVC pipes were buried at a depth of 15 cm from the sand surface. During data collection, markers were introduced manually into the data at the beginning and end points of the top panel as depicted in Figure 2.13 (c).

Table 2.1: Pipe dimensions used in the experiment performed for pipes detection.

<table>
<thead>
<tr>
<th>Pipe size</th>
<th>Diameter (cm)</th>
<th>Length (cm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Big</td>
<td>10.8</td>
<td>25</td>
</tr>
<tr>
<td>Medium</td>
<td>4.3</td>
<td>31.5</td>
</tr>
<tr>
<td>Small</td>
<td>2.8</td>
<td>30</td>
</tr>
</tbody>
</table>
2.6.2.2 Cracks and Defect within Water Reservoirs Detection

The physical model was built to provide a realistic laboratory representation of joint structures with and without the presence of water. Two concrete slab test-beds, A and B, were used to simulate storage reservoirs with leakage problems and also those, which remain intact. The construction of the slabs and the preparation of these models for experimental purposes are described in [42].

In order to perform experiment of concrete with and without the presence of water, the concrete slabs required container boxes to allow the slab to be submerged in the water. Two container boxes were made from 2 cm plywood measuring 1.2 m × 1.2 m × 0.15 m sitting on four under supported beams with the measurements of 0.07 m × 0.09 m × 1.3 m. Both boxes were lined with PVC pond liner material, which was securely stapled into position.

A simple PVC drain was fitted through the base in the corner of each box. The right angle that bends the drain was turned into a “U” tube by using a 2.2 cm PVC tube with the outside leg is able to turn down to allow for free drainage of the boxes. The drain was filled with plastic pan scrub material to prevent the sand from flowing down the drain whilst allowing the flow water when required.
The sand was prepared with bags of washed sand that were spread out on polythene sheets on the laboratory floor with a blowing fan heater to dry the sand. The 9 × 25 kg bags of dry sand were placed in the slab A whilst the slab B contained the same quantity of damp/wet sand. The sand was levelled and covered with a weed control fabric to prevent sand interfering with the joints when the concrete blocks were laid out (Figure 2.15).

Each slab was constructed by building four concrete blocks with the measurements of 0.915 m × 0.25 m × 0.125 m. The first block was laid on the fabric on its side and 7 mm joint of F-Bondpolyester resin FIP 300SF was applied to the side of the second block to be matted to the first block. This was repeated with the third block with the exception of the thickness of the joint which is 25 mm. The last block was jointed with an interior white sealant (used in UPVC windows) with 12 mm thickness.
During data collection, markers were introduced manually into the data when the radar crossed the concrete slab joints or a large crack. In order to carry out experiments involving holes and cracks, both slabs were drilled and hit with a club-hammer. The approximate positions of joints, holes, and cracks are depicted in Figure 2.17. Application of hammer resulted in the development of a large crack along the middle joint and goes all the way to the bottom of the slabs.
Figure 2.17: (a) Depiction of joints, holes, and cracks position and (b) the approximate positions of the radar traverses during the experiments.

2.6.2.3 Cracks on a Concrete Slab Detection

A laboratory experiment was conducted at dedicated concrete testing facility in detecting cracks on a concrete slab \((6\,\text{m} \times 6\,\text{m} \times 0.25\,\text{m})\) reinforced with metal fibre mixtures \((3\,\text{cm} \times 0.1\,\text{cm})\). During data collection, 20 cm grid was used to collect each scan that covers the whole concrete slab and
markers were introduced manually into the data when the radar crossed the crack. In addition to this, measuring the position of each crack was also made using measuring tape.

(a) Concrete slab containing metal fibre mixtures.  
(b) The 20 cm grid measurement.

(c) Cracked concrete block.  
(d) Practical approach measuring the crack position.

Figure 2.18: Data collection over a cracked concrete metal fibre mixtures slab.
### 2.7 Glossary of Terms

The following section provides description of some terms related to GPR technology, commonly used in this section [54]:

<table>
<thead>
<tr>
<th>Term</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Antenna</strong></td>
<td>It is the transmitting or receiving transducer of the GPR system. The transmitting antenna produces the radiating GPR wave whilst the receiving antenna captures the reflected/scattered energy from the targets.</td>
</tr>
<tr>
<td><strong>Attenuation</strong></td>
<td>The loss of signal energy affecting GPR waves as they propagate through the subsurface. The attenuation is caused by a combination of material losses and the spreading loss of energy as the wave travels outwards.</td>
</tr>
<tr>
<td><strong>EM energy</strong></td>
<td>Propagating energy made up of coupled, alternating electrical and magnetic fields. GPR uses radio wave EM energy at MHz.</td>
</tr>
<tr>
<td><strong>GPR section</strong></td>
<td>The cross-sectional image of the subsurface generated by the GPR system in the form of a collection of coloured or grey-scale strips that relate to the amplitude of the recorded voltage signal trace. The GPR section usually has</td>
</tr>
</tbody>
</table>
Ground Penetrating Radar

distance along the survey line as the horizontal axis and recorded time (as two-way travel time in ns) as the vertical axis.

GPS
Global Positioning System – device for spatial navigation and positioning. Only differential GPS systems at lower frequencies antenna are accurate enough for GPR.

Grey-scale
A sequence of grey shades from black through greys to white used to show the amplitude of a GPR trace. It produces a GPR section that has the same colouring as a ‘black and white’ photograph.

Radio waves
It is an EM energy at a frequency of MHz. At these frequencies, EM energy is able to travel through air or ground relatively easily – called radiating EM energy. These frequencies are referred to as ‘radio waves’ as this is the frequency band of AM radio stations.

Two-way travel time
The time it takes the GPR wave to travel down to a target and back up to the receiving antenna in ns.
Trace

This is the recorded output of the received GPR signal. It is a simple recording of the voltage at the receiving antenna over a specific period of time from the start or ‘triggering’ of the GPR system. The trace takes the form of wiggles or wavelets and is usually converted into a coloured or grey-scale strip of pixels in a GPR section.

Velocity

The ‘speed’ of a propagating GPR wave through a material in m/s or m/ns.
Chapter 3

Pre-processing using ReflexW

Overview

This chapter describes the details of dedicated GPR software package, ReflexW™, used to pre-process GPR data analysed in this project. It gives details of the processing steps available in the software to improve the raw data quality before further processing and techniques developed in this work are applied.

3.1 Introduction

GPR has been used extensively for a variety of applications in many fields. Processing GPR data using computers are commonly used nowadays, due to the inexpensive access to computer facilities [62]. The processing procedure consists of a sum of operations applied to the raw data in order to enhance the signal-to-noise ratio (SNR) and produce a more realistic image of the ground. GPR provides a quick and easy way to image the first few metres of the subsurface with a relative good resolution in determining shallow void
characteristics where individual anomalies are not distinguishable by the other non-destructive geophysical methods [63].

A GPR display typically consists of a collection of time-series returns viewed stacked side-by-side giving a depth profile of the subsurface, resulting in either the B-scan displays common in seismic surveys, or a hard-limited intensity modulated raster plot of echo-strength vs. travel time [7]. These plots of radargrams are complex and difficult to comprehend without some level of signal processing to put them in a format that is amenable to human examination [47].

The purpose of GPR data processing is to improve the raw data quality by manipulating the acquired data into an image that can be used to infer the subsurface structure. Minimal processing is required in the present of a perfect acquisition system. In general, the processing routines fall into one of the following categories [64]:

- enhancing signal at the expense of noise;
- providing velocity information;
- collapsing diffractions and placing dipping events in their true subsurface locations (migration);
- increasing resolution.

A variety of data processing techniques have been developed to aid in the interpretation of the data. Ultimately, most users process GPR data with dedicated proprietary software [37], either system-specific (e.g., RADAN™...
Pre-processing using ReflexW

[49]) or independent packages that can import a range of different data types (e.g., IXGPR\textsuperscript{TM} [65], Radar Unix [66], RadExplorer\textsuperscript{TM} [67], ReflexW\textsuperscript{TM} [68]).

Two types of GPR data processing have been classified in this work called pre-processing and one-dimensional processing. In the pre-processing step, two different approaches have been applied to the GPR raw data using ReflexW\textsuperscript{TM} [68] and some other manual simple steps such as simple mean removal, PCA, and ICA which explained further in the next chapter. The one-dimensional signal processing techniques are discussed in Chapter 5.

Figure 3.1: Overview of GPR data processing flow [37], [62].
Pre-processing using ReflexW

Typical processing flow for GPR data is depicted in Figure 3.1 [37], [62]. Data processing focuses on the purple highlighted areas: data editing, basic processing, advanced processing, and visualisation/interpretation processing. Processing is usually an iterative activity in which the data will flow through the processing loop several times before it is finalised in the visualisation step. Batch processing with limited interactive control may be applied on large datasets after initial testing on selected data samples have been performed.

3.2 ReflexW Software

In this work, the collected GPR data was imported and processed using ReflexW™ [68], which is an independent package software that can import a range of different data types. It has been developed by K. J. Sandmeier for the processing and interpretation of reflection and transmission data specifically in GPR application, and reflection and refraction of seismic and ultrasound data [69].

3.3 Basic GPR Data Processing Steps

The basic processing steps are usually applied to the raw data (often automatically) and introduce minimal operator bias into the data without the need for additional subsurface information, typically in the form of trace
Pre-processing using ReflexW editing, filtering or data correction [37]. These procedures are in general applicable to most collection modes.

Figure 3.2 describes the processing flow used for analysing GPR data using ReflexW. The basic descriptions of each step are as follows [37]:

- data editing – removal and correction of bad/poor data and sorting of data files;
- dewow filtering – correction of low frequency and DC bias in data;
- topographic corrections – correction of topography effects;
- time-zero correction – correction of start time to match with surface position;
- filtering – 1D and 2D filtering to improve SNR and visual quality;
Pre-processing using ReflexW

- gain function – improvement of data display and interpretation;
- migration – correction of the survey geometry and spatial distribution of energy effects.

During data processing in ReflexW, the filenames of the GPR data is recommended to be numerically sequential for the purpose of viewing data in three-dimensional before the processing steps taken into action. Moreover, the sequence processing stage is easier to be applied in processing batch of data. Details analysis of each step is discussed in the subsequent chapters.

3.3.1 Data Editing

In any processing sequence, data editing is often the most time-consuming since the files usually need sorting and rearranging as the first step in the post collection. Effective maintenance of the data from the start is vital for good-quality interpretation, particularly with large volumes of data [44]. The incorrect recording of survey parameters (e.g., spatial increment, section interval, and start position) results inaccuracy of data file headers, thus affecting the way of the data being imported into the processing program. It is therefore a good practice to view the header files of each section before the processing to check the consistency and accuracy of the survey parameters [37].
Incoherent, noisy or missed traces require editing out and/or filtering to improve the visual nature of the section [37]. This is commonly caused by the overenthusiastic triggering, external noise sources, equipment failure/problems, or traversing too fast if odometer is being used. In most cases, only the occasional trace is corrupted and a simple interpolation between traces is sufficient.

Figure 3.3: Example of 'desaturation' function is used in correcting the clipped form of saturated traces [37].

‘Desaturation’ or ‘declipping’ feature is a useful editing tool available in some programs [37]. Under certain circumstances, the initial ground wave signal can become ‘clipped’ as the GPR receiver system saturates with
strong ground coupling, and thus the recorded trace does not represent the true peak amplitude of the returned signal (Figure 3.3). If trace normalising is being used to account for differences in antenna-ground coupling, the late arrivals of saturated traces will be artificially enhanced in comparison to the non-saturated traces. The ‘desaturation’ function attempts to correct this effect by reconstructing the form of the ground wave pulse using interpolation methods (Figure 3.3) as long as the trace is not too noisy and the form of the saturated ground wave wavelet is ‘clean’.

3.3.2 Dewow Filtering

Due to the close position of transmitter and receiver antennas, the fields near the transmitter contain low frequency energy associated with electrostatic and inductive fields, which decay rapidly with distance. This low frequency energy caused the base level of the received signal to bow up or down, and this effect is known as baseline ‘wow’ in the GPR technology [37].

Dewow filtering is used to remove the low frequency components presented in the data by applying a running average filter to each trace and eliminate a long waved part of the signal that is caused by the EM induction [43], [44], [70]. It is a vital step as it reduces the data to a mean zero level and therefore, allows positive-negative colour filling to be used in the recorded traces (Figure 3.4) [37].
3.3.3 Topographic Corrections

Topographic correction is an essential component of GPR data editing, especially in cases where there is significant topographic variation in order to ‘position’ the data in its correct spatial context [37], [43]. Compensation for topography is often important due to the shallow exploration depth of GPR. For minor surface variations, time-shifting data traces can largely compensate for topographic variations [37]. Two types of topographic corrections were applied to GPR data which are elevation static and maximum phase corrections.

Figure 3.4: Dewow filter correction on a raw GPR trace [37].
3.3.3.1 Elevation Static Correction

Flattening a GPR profile based on the airwave does not account when near surface has significantly different radar velocities than underlying units [71]. The elevation and velocity static correction is therefore should be performed to obtain more realistic subsurface images in order to correct any changes in topography over the survey [43]. Elevation static correction is normally performed with a simple constant velocity correction that acts in a vertical sense [72]. It corrects the two-way travel time of the traces to a flat datum level some distance above the air/ground interface. Figure 3.5 describes the datum statics terminology of static correction [64].

![Datum Statics Terminology](image)

Figure 3.5: Datum statistics terminology [64].

ReflexW requires the user to specify the subsurface velocity profile independently to shift the data to positive (or negative) times from a given
datum reference, usually either lowest or the highest point in the whole survey line. Table 3.1 [73] represents the value of distance and time correspond to the selection of radar frequencies as the interactive input of the static correction values. The position of the Sender/Receiver distance is depicted in Figure 3.6.

Table 3.1: Sender/Receiver distance values.

<table>
<thead>
<tr>
<th>GPR Frequency</th>
<th>S/R Distance (m)</th>
<th>Velocity (m/s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>GV3 400MHz</td>
<td>0.3</td>
<td>0.1</td>
</tr>
<tr>
<td>GV3 1GHz</td>
<td>0.12</td>
<td>0.1</td>
</tr>
<tr>
<td>GV3 1.5GHz and CDH</td>
<td>0.07</td>
<td>0.1</td>
</tr>
<tr>
<td>GV5 4GHz</td>
<td>0.03</td>
<td>0.1</td>
</tr>
</tbody>
</table>

Figure 3.6: The position of Sender/Receiver distance.
3.3.3.2 Maximum Phase Correction

Maximum phase correction is used to correct the phase of maximum amplitude in a zone near the coupling and direct energy by flatten on one of these for a better look of the GPR data. This step is often useful in the presence of rough ground [70]. The application of maximum phase correction can be seen in Figure 3.7.

Figure 3.7: The application of maximum phase correction step to measured GPR data (a) before and (b) after.
3.3.4 Time-Zero Correction

The zero time may not have been detected precisely by the instrument in the field and should therefore be repacked to ensure correct depths in the profile. Furthermore, drift of the zero time along the profile can occur because of the temperature difference between the instrument electronics and the air temperature or damaged cables. The drift cases misalignment of the reflections and the zero time has to be re-positioned for all traces along the profile [44]. The time-zero positioning is usually achieved using some particular criteria such as the air wave first break point or first negative peak of the trace [37].

During data time-zero positioning, it is important to ensure that the mean value of the data (A-scan) is near to zero, which assumes that the amplitude probability distribution of the data is symmetric about the mean value and not skewed. The short time mean value is therefore constant over the time duration of the A-scan [13]. In general, the grey area on top of each radar signal in GPR image (usually applied to low frequency radar) is removed since it contains very little information for a better processing (Figure 3.8) [74]. This step might be unnecessary for high frequency radar which has good resolution that gives better view on the surface rather than depth.
Figure 3.8: Time-zero variations in a selected 1.5 GHz GPR data. (a) Time-zero position on the original GPR data (white dashed-line) and (b) the white dashed-line change in time-zero position.

3.3.5 Filtering

Filtering [37] is generally applied to the data to remove cultural (i.e., human-induced), system noise (e.g., high frequency ‘speckle’ from radio transmission [75]) or the striping effect from the antenna ringing [76] for the visual quality improvement of the data. It is useful in extracting particular aspects of the data and aiding interpretation (e.g., emphasising flat-lying reflectors, diffractions [62], [77]). Simple filters are often very effective at removing high/low frequency noise, whilst sophisticated methods are more appropriate for specific problems (e.g., excessive ringing or noise spikes [62], [78]).

Filters can be applied before and after gains but pre-gain filters do operate on the data in its truest form [37]. If filters are being applied post-
gain, the effect of the gain on the amplitude and spectral content of the data is therefore must be fully understood first. In general, filters can be classified into two basic types:

- temporal – applied down the individual traces in time;
- spatial – applied across a number of traces in distance.

These are often combined to produce advanced two-dimensional filters that operate on the data by removing, suppressing or enhancing signals of given frequency or across a specific number of traces in space of samples (time) [37].

3.3.5.1 Temporal Filters

Temporal filters alter the shape of single traces in the vertical (time) direction to enhance or eliminate noise at frequencies either higher or lower than the main GPR signal bandwidth and, as such, are ultimately just ‘clean-up’ filters that make the GPR section visually better. The typical one-dimensional temporal filters include [37]:

- simple mean – takes the mean of the data across a specified time window and smoothes the data to remove excessive higher frequency noise;
- simple median – takes the median of the data across a specified time window for power spike removal;
Pre-processing using ReflexW

- low- or high-pass (frequency domain filters) – the low frequency component is used to remove the noise in the data whilst the high frequency components eliminate the signal drift and low frequencies;

- band-pass (frequency domain filters) – it is a combination of both high- and low-pass filters that will let through a specific range of frequency components which are defined by a ‘pass region’.

Band-pass Butterworth filter (Figure 3.9) is used in processing the GPR data using ReflexW. The decision of pass region should be selective as excessive in narrowing down the pass region cause in removing the actual recorded signal components and the outcome of GPR section will become poorer [37]. Table 3.2 [74] shows the cut-off frequencies used with band-pass Butterworth filter.

### Table 3.2: Band-pass Butterworth filter cut-off frequencies.

<table>
<thead>
<tr>
<th>Radar type</th>
<th>Lower frequency</th>
<th>Upper cut-off frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>GV3 400 MHz</td>
<td>200 MHz</td>
<td>600 MHz</td>
</tr>
<tr>
<td>GV3 1 GHz</td>
<td>500 MHz</td>
<td>1500 MHz</td>
</tr>
<tr>
<td>GV3 1.5 GHz and CDH</td>
<td>750 MHz</td>
<td>Between 2250 – 3000 MHz</td>
</tr>
<tr>
<td>GV5 4 GHz</td>
<td>700 MHz</td>
<td>600 MHz</td>
</tr>
</tbody>
</table>
Pre-processing using ReflexW

Figure 3.9: Principle of a simple band-pass filter in the frequency domain and the form of two common filter functions: a notch and Butterworth filters [37].
3.3.5.2 Spatial Filters

Spatial filters operate in the same manner as temporal filters but, instead, perform the filtering operation across the traces in distance to enhance or eliminate certain frequencies and emphasise or suppress features. The most common of spatial filters are [37], [44]:

- simple running average – takes the mean of a large number of traces to smooth the data horizontally and emphasises flat-lying reflectors whilst suppressing dipping reflectors and diffractions;
- average subtraction – takes the mean of a small number of traces in a window to be subtracted from each individual trace in sequence to suppress flat-lying reflectors and emphasises dipping reflectors and diffractions;
- background removal – takes the mean of all traces in a section to be subtracted from each trace to remove background noise and antenna ringing, and eliminates continuous flat-lying reflectors.

One of the major problems in processing GPR data is the removal of the air ground reflection [79]. Spatial filters operate well with good-quality, low-clutter data and, in general, tend to be good at removing the strong air/ground wave response and ringing from the datasets. Background removal and average subtraction were used during data processing in ReflexW. The background removal was able to remove the striping effects from the image and enhance the curved and dipping reflectors whilst the
average subtraction is very effective to allow subtle weaker signals that are lost to become visible in a processed section [62], [70].

4 GHz GPR data: Basic Processing
GPR section collected over a concrete slab with crack structure at 0.18 m. Basic processing – Dewow filtering.

High frequency noise ("speckle")

4 GHz GPR data: Band-pass filter applied
Band-pass Butterworth filter was applied to all traces equally. Pass region is 700 – 6000 MHz.

High frequency noise ("speckle") removed
3.3.6 Gain Function

Radar signals can be rapidly attenuated during the propagation into the ground. Signals from greater depths are very small compared to signals from shallower depth [37]. The simultaneous display of these signals
Pre-processing using ReflexW requires conditioning before visual display due to the attenuation and spherical EM wave spreading of the signals [44]. The related amplitude information can be recovered by applying gain function to compensate the propagation losses at higher depths as well as increasing the visibility of deep objects [80].

The data structure is altered in some way (i.e., relative amplitude and/or phase relationships are changed) in the application of gain function [37], [43]. It is therefore very important to understand the effect of gain function before its application and the data are treated with care when interpreting [62]. Figure 3.11 shows the compensation of decreased signal amplitude at increasing travel by designing a custom gain to increase the signal strength. Gain is applied to the data which increases in time after the transmit pulse and the rise of the gain function is tailored to accommodate the drop-off and signal amplitude versus time [81].

![Figure 3.11: Concept of time-varying gain where signal amplification varies with time [81].](image-url)
Several procedures can be used in gain function such as constant gain, exponential gain and automatic gain control (AGC) [82]. All gain functions tend to operate in a similar fashion by applying some multiplying factor to successive regions of the trace in time (referred to as time window) and can be easily changed, typically the window length (in ns), the procedure type (linear, exponential, user-defined, etc.) and the maximum gain allowed. In general, both noise and coherent signals are usually amplified together in an indiscriminate way [37].

The choice of the range gain function applied during acquisition depends on the data handling. The gain can either be designed to enhance the visibility of the data for on-site interpretation which is the most common among radar operators, or set to a constant value throughout to preserve the information content of the signal that is more beneficial when further digital signal processing is to be performed on the data [7].

![Figure 3.12: The effect of inordinate acquisition gain selection](image)

(a) Excessive gain applied  (b) After removal of gain

Figure 3.12: The effect of inordinate acquisition gain selection [7].
Pre-processing using ReflexW

It is however possible for the receiver to clip some portions of the signal that exceed the range gain. This clipping enhances the appearance of the image by highlighting the edges of any anomalies and the contrast for the detection of weak signals is increased. However, this situation is known as data distortion (Figure 3.12 (a)) and is therefore important to include this effect during the range gain selection [7].

![Images showing moderate gain applied and after removal of gain](image)

(a) Moderate gain applied  
(b) After removal of gain

Figure 3.13: Acquisition gain removal during pre-processing [7].

In general, gain should be selected based on some priori physical model and the objective should be to modify the data whilst retaining its full utility without introducing artefacts. Time gain is a non-linear operation, thus the application of filtering operations before and after time gain should be accounted as the output will not be equivalent [37], [81].
3.3.7 Migration

Migration [13], [37] is an image processing technique that used to improve section resolution and develop more spatially realistic images of the subsurface. The objective is to reconstruct, on the basis of reflection data, acquired at the surface, the entire reflecting structure that is present in the subsurface, and is causing said reflection data [79]. It applies a synthetic aperture image reconstruction process to focus scattered signals, collapse hyperbolas to their apices, and re-position dipping reflections [43]. The idea of migration is illustrated in Figure 3.14 [79].

![Figure 3.14: The problem – the reflected energy of the object is spread out over the different receiver positions leading to the measurement hyperbola. The idea of migration is to recombine this energy in one position [79].](image)

Migration requires an accurate radar velocity structure and the origin knowledge of the distorted reflections and wave travel paths before it can be applied to the data [37], [62]. In general, migration often enhances the
display of the reflections significantly by moving the collapsed diffraction hyperbolas and dipping reflections to the true geometrical position [44].

3.4 ReflexW Processing of the Measured GPR Image in this Work

Figure 3.15 represents the example of the ReflexW process on the GPR data to detect the presence of joints within water reservoirs with 4 GHz frequency radar. Note that the time-zero correction step is not applied to this data due to the high frequency GPR is used.
Pre-processing using ReflexW

**Basic processing – dewow filtering applied**
The low frequency presented in the data is removed.

**Topographic correction – static correction applied**
‘Positioning’ the data in its correct spatial context since the data is collected over unflattened surface.

**Topographic correction – maximum phase correction applied**
Correct the maximum phase in a zone near the coupling and direct energy due to rough concrete surface.
Pre-processing using ReflexW

Filtering – band-pass applied
Band-pass Butterworth filter was applied to all traces equally. Pass region is 700 – 6000 MHz.

High frequency noise (“speckle”) removed

Filtering – background removal applied
The striping effect is removed.

Air-ground wave removed

Filtering – average subtraction applied
The subtle weaker signals become visible.
Pre-processing using ReflexW

3.5 Conclusion

Traditional approach of using commercially available ReflexW software for geophysical and GPR data processing has been investigated as an alternative way to pre-process data using signal processing techniques. Basic and simple processing steps such as one-dimensional filtering, topographic and

Figure 3.15: Example of ReflexW process on GPR data to detect leakages from joints.

Gain function applied
- Improve the data display and interpretation.
- Gain function selected – exponential.

Migration applied
- Correct the effects of survey geometry and spatial distribution energy.
- Joint structures are more visible
- Sand layers
Pre-processing using ReflexW

time-zero corrections, band-pass filtering, background removal and average subtraction were used to process the data without altering the main features or targets present in the data. However, highly skilled and trained human operator is usually needed to process the measured data using this dedicated software to make sense of the captured images and interpret accordingly.
Chapter 4

Multivariate Pre-processing Techniques for GPR Images

Overview

Pre-processing methods for clutter and unwanted signal reduction in GPR data presented in this chapter is accomplished using multivariate techniques before further analysis can be undertaken. PCA and ICA algorithms are discussed in details as well as the two novel approaches adopted in this work – SVDPICA and 2DPCA. A comparison of pre-processed images between these algorithms and ReflexW is given as part of the research reported in this thesis.

4.1 Introduction

GPR radargram provides information about the subsurface layers and anomalies that can be used to detect and investigate buried targets. Radargram is a two-dimensional dataset representing a collection of reflected traces corresponding to a vertical slice through the surveyed
ground. One of the most attractive features of GPR technology is the high resolution of subsurface imagery provided although shallower penetration depth is usually achieved compared to some other geophysical methods. The depth of GPR investigation may become even shallower due to the contamination of GPR data by severe ringing which can mask the reflected signals [83]. In the presence of a target, the GPR reflection results in an “inverted hyperbola” shape and image processing techniques can be used to extract and analyse this signature from the radargram [84]. It is however important that the background noise, clutter effects, antenna and ground coupling be removed or suppressed as much as possible before those techniques are applied. The obtained GPR profiles, acquired in “real-time” can be analysed in a raw format but some pre-processing is usually necessary to clean up the data and enable its reliable interpretation [85].

4.2 Review of Prior Research

4.2.1 Previous Work in Signal Processing

A number of methods have been developed in order to reduce the presence of unwanted signals which can often dominate the GPR data and obscure the “real” reflections containing target information. Time gating technique [10], [21] has been used to distinguish a target from the backscattered signal of the surface in the situations where target is buried deep below the surface. This technique has been used to remove the ground reflection from GPR
measurements of deep targets such as buried pipes, exploded ordnance and tunnels [86]. However, choosing an appropriate time gate can be challenging for shallow targets buried near the surface. This is due to the overlapping responses from the targets and clutter, which the backscattered signal from both the targets and the surface will arrive almost simultaneously [21].

The simplest clutter reduction technique is background subtraction method [8], [83], [87] whereby the background signal is estimated as the mean of the unprocessed ensemble of GPR signals without a buried object [18], [88]. This method is most effective if the ground is flat and uniform over the antenna path – an ideal situation, rarely encountered in the field [86]. This approach also tends to remove parts of the target of any of the used scan paths that include the target response. In the case of soil which is an inhomogeneous medium and its statistical properties vary with position along the surface, subtracting the average may not lead to sufficient clutter reduction [21]. Alternatively, the background subtraction can be performed by estimating the background to be subtracted from the unprocessed GPR signals by using either moving average [89], [90] or median filter [18]. However, the estimation of the background will be affected in the presence of an object and the estimation will not be accurate.

It is worth noting again that most of the above listed and discussed methods depend on the background signal estimation [10]. The background
signal is estimated first, as the mean value of the unprocessed ensemble of the collected GPR data before the simple mean scan subtraction method is employed. Alternatively, A-scan far away from the landmine location might also be used. Although this method cannot completely remove the direct wave from the received data due to the air-earth interface and time-shift of the radar system, it has been widely used in GPR applications.

4.2.2 Multivariate Analysis in GPR Application

Another group of techniques, known as multivariate data analysis techniques has also been applied to process the obtained radargram before further analysis of the recorded GPR image takes place. The application of PCA analysis to GPR data for mines detection using cross track-depth scans is presented in [91]. In [92] clutter was reduced by reconstructing the GPR image from the most significant eigenvectors, and generalised SVD has been used in [86] to separate noise and signal spaces. A different unsupervised approach is introduced in [27] where characteristics of the source signals (principal components) and associated eigenimages were used to determine the subspace for reconstruction.

ICA is another technique widely used in many applications such as brain imaging, econometrics, feature extraction [93], acoustical signal processing, edge detection, and face recognition [94]. This technique has been applied in [25] by implementing the Joint Approximate Diagonalisation of Eigen-matrices (JADE) to reduce clutter for anti-personal (AP) landmine. The
spatial and temporal ICA approaches is compared in [26] and [27] on buried iron and plastic AP landmine dummies. In [6], ICA has been used to analyse the GPR signal characteristics via Fixed-Point ICA (FastICA).

Other related multivariate methods also include Factor Analysis (FA) [95], [96] and Singular Spectrum Analysis (SSA) [85]. Factor Analysis [41], [96] is a method for modelling the correlations in multidimensional data, by low-dimensional latent variables typically called factors or sources. A set of factors will be extracted from the dataset which are orthogonal and ordered according to the proportion of the variance of the original data. In general, only a (small) subset of factors is kept for further consideration and the remaining factors are considered as either irrelevant or non-existent (i.e., they are assumed to reflect measurement error or noise) [97].

Singular Spectrum Analysis [80], [85], [98] is an approach closely related to PCA. SSA decomposes the original time series (GPR traces) into a sum of a small number of independent and interpretable components such as low-frequency trends, narrowband quasi-periodic signals and noise. It consists of four steps which are embedding (i.e. a one-dimensional time series of observations is transformed into the multidimensional series), the SVD application to obtain a set of vectors that are linearly independent to be formed into several groups before the full reconstruction of the given radar traces time series by using a diagonal average procedure.
The application of these multivariate techniques analysis represents the main approach and contribution of this work. Fundamentals of these techniques are explained in the continuation of this chapter. Two particular techniques from this group – Principal Component Analysis and Independent Component Analysis are discussed in details and their application in processing obtained GPR data explained. Modifications introduced to both techniques in order to improve their performance in processing GPR data are given followed by some results and conclusions.

4.3 Multivariate Analysis

The measured GPR data can be considered to consists of a large number of variables [41], where each corresponding sample in returned GPR trace represents one variable. It is to be expected that those variables are highly correlated with each other thus affecting the accuracy and reliability of target detection and characterisation methods. Unnecessary variables can also increase the data-storage and data-processing costs of deploying a model on a large database. One of the key reasons for applying multivariate analysis in processing GPR data is to find the ways to reduce dimensionality of GPR data without sacrificing the accuracy of measured datasets.

4.3.1 Principal Component Analysis

Principal Component Analysis (PCA) is one of the most popular techniques for pattern recognition and signal processing [99], [100]. In fact, it is
important in many engineering and scientific disciplines, e.g., in data compression, feature extraction, noise filtering, signal restoration and classification [101].

PCA [102] is a technique usually employed to reduce dimensionality of a multivariate dataset without significant loss of information by finding the principal components (PCs) with which the given data can be most compactly represented. Among the purposes of PCA are the following [99]:

- dimensionality reduction;
- determination of linear combinations of variables;
- feature selection: the choosing of the most useful variables;
- visualisation of multidimensional data;
- identification of underlying variables;
- identification of groups of objects or of outliers.

PCA is a second-order statistical method which uses covariances between the observed variables in the estimation in order to express datasets in terms of linear combinations of a small set of factors that are assumed to be uncorrelated and Gaussianly distributed [41]. It is therefore worth pointing out that for the already uncorrelated variables contained in a dataset, application of PCA is of no value.

In addition to being uncorrelated, the PCs are orthogonal and ordered in terms of the variability they represent [103]. That is, the first PC accounts for as much of the variability in the data as possible and each succeeding
Multivariate Pre-processing Techniques for GPR Images

A component accounts for as much of the remaining variability as possible. In most cases, the representation is sought as a linear transform of the observed variables [41]. Often the PCs (i.e., directions on which the input data have the largest variances) are regarded as important, whilst those components with the smallest values are regarded as unimportant since they are considered to be associated with noise and not contribute very much information to the overall dataset and, hence can be eliminated. It is sometimes possible to eliminate a large number of variables that contribute only marginally to the total information with the appropriate transformation [99].

![Graphical model for Principal Component Analysis.](image)

The motivating idea behind the use of PCA in GPR data processing is the high dimensionality of GPR signals. It is therefore very likely that those variables can be highly correlated with each other. The accuracy and reliability of various detection methods will suffer if the analysed datasets
are containing highly correlated set of variables that are unrelated to the outcome of interest. One of the key steps in processing the GPR data is finding ways to reduce dimensionality without sacrificing accuracy, thus PCA has to be one of the most suitable candidate approaches.

Geometrically, PCA treats the samples from each GPR scan at a given time moment as a point in the space of dimension $n$ equal to the number of performed GPR scans. One instance of this data forms a cloud of points in $n$-dimensional space. This cloud can spread in some directions more than in the others where variance of data in a given direction represents the measure of the data spread. The PCA finds a set of orthogonal axes, a base, such that each consecutive axis spans the directions with consecutively decreasing variance. The projections of the points onto these axes constitute the components. Each component can be visualised as a time series. The original time series can be recovered as a linear combination of these components. Components corresponding to very small variances can be neglected hence a data dimensionality can be reduced. Besides, the clutter information for GPR measurements is usually contained in one or a couple of PCs with the highest variance, therefore clutter and noise-free GPR dataset can be obtained by discarding those components [104].

Two types of PCA – one-dimensional PCA (1DPCA) and two-dimensional PCA (2DPCA) were used in this work. The subsequent chapters described each of these algorithms in details.
4.3.1.1 One-Dimensional Principal Component Analysis

(1DPCA)

The PCA algorithm can be performed via two approaches which are Eigenvalue Decomposition of Covariance Matrix and Singular Value Decomposition. The differences of these methods are described in the following sections.

4.3.1.1.1 PCA via Eigenvalue Decomposition of Covariance Matrix

For the purpose of implementing PCA algorithm on GPR dataset, a $m \times n$ GPR data matrix $X$ containing $n$ measurements – GPR traces, each consisting of $m$ data samples, variables can be considered as indicated in Figure 4.2. Dimensionality of measured data in this case is the length of each GPR trace from the set $m$. The main goal of PCA is to transform the original dataset to another domain – variable space where same data can be represented with smaller number of variables. In this way, dimensionality of the data is reduced and better understanding of the measured data enabled. The linear transformation to accomplish this can be defined as:

$$ Y = SX $$

(4.1)

The original dataset $X$, where columns denote particular measurements and rows stand for individual data samples in each measurement, is now expressed by a new transformed matrix $Y$ through the transformation
defined with matrix $S$. Calculation of matrix $S$ such that the transformed dataset contained in $Y$ is mutually uncorrelated, represents the main objective of the PCA algorithm. The structure of data matrix $X$ can be represented as:

$$X_{m \times n} = \begin{bmatrix} x_{1,1} & x_{1,2} & \cdots & x_{1,n} \\ x_{2,1} & x_{2,2} & \cdots & x_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ x_{m,1} & x_{m,2} & \cdots & x_{m,n} \end{bmatrix} = \begin{bmatrix} x_1 & x_2 & \cdots & x_i & \cdots & x_n \end{bmatrix} \quad (4.2)$$

Figure 4.2: (a) GPR image contained in data matrix $X$ and (b) single GPR trace $x_i (i = 120)$ from $X$ – A-scan.

Covariance matrix $C_X$ of the original data $X$ has to be calculated first in order to obtain $S$. The covariance matrix consists of the correlation characteristics (covariances and variances) between each pair of data from $X$. The covariance between two measured values indicates the degree of mutual similarity, where large absolute value denotes a high correlation and therefore redundancy of respective data. Zero covariance, on the other side
indicates completely uncorrelated data. To construct the covariance matrix for the dataset $X$, it is first necessary to calculate and remove the mean value from each measurement:

$$\bar{x}_j = x_j - \frac{1}{n} \sum_{j=1}^{n} x_j$$

(4.3)

where $x_j$ denotes the $i$-th data sample in the $j$-th measurement (i.e., $j$-th column of matrix $X$). In this way, each dimension is centred around zero and new, centred data matrix is obtained:

$$X_0 = \begin{bmatrix}
\bar{x}_{1,1} & \bar{x}_{1,2} & \cdots & \bar{x}_{1,n} \\
\bar{x}_{2,1} & \bar{x}_{2,2} & \cdots & \bar{x}_{2,n} \\
\vdots & \vdots & \ddots & \vdots \\
\bar{x}_{m,1} & \bar{x}_{m,2} & \cdots & \bar{x}_{m,n}
\end{bmatrix} = [\bar{x}_1 \quad \bar{x}_2 \quad \cdots \quad \bar{x}_n]$$

(4.4)

The covariance matrix is obtained as a dot product of centred data matrix $X_0$ divided by the number of measurements, traces in the analysed dataset $X$:

$$C_X = \text{cov}(X) = \frac{1}{n-1} X_0 X_0^T$$

(4.5)

In (4.5), $n-1$ is usually used instead of full number of measurements $n$ in order to obtain unbiased covariance estimate. Obtained covariance can now be decomposed via spectral or eigenvalue decomposition into product of three matrices:

$$C_X = U \cdot D \cdot U^T$$

(4.6)

where $U$ is an orthonormal matrix (matrix of orthogonal unit vectors, i.e.

$$U^T U = I \quad \text{or} \quad \sum_{k=1}^{n} u_k i u_j = u_i^T u_j = \delta_{ij}, \quad u_i \text{ is the } i\text{-th column of } U) \text{ and } D \text{ is a}$$
diagonal matrix. The columns of $U$ are the eigenvectors of matrix $X$ and the diagonal elements of $D$ are the eigenvalues of $X$. If $X$ is positive-definite, the eigenvalues will all be positive. Equation (4.6) can be multiplied with $U$ and rewritten as:

$$C_X \cdot U = U \cdot D \cdot U^T \cdot U = U \cdot D$$

(4.7)

This can then be written as a normal eigenvalue equation:

$$C_X \cdot u_i = \lambda_i \cdot u_i$$

(4.8)

where $u_i$ represents the $i$-th column of $U$, i.e. the $i$-th eigenvector and $\lambda_i = D_{ii}$ is the $i$-th eigenvalue.

Using $S = U^T$ for the linear transformation defined in (4.1), covariance matrix of transformed dataset $Y$ can be expressed as:

$$C_Y = \text{cov}(Y) = \frac{1}{n-1} YY^T = \frac{1}{n-1}(U^T X)(U^T X)^T = \frac{1}{n-1} U^T XX^T U$$

(4.9)

Equation (4.9) can be rewritten in the form similar to matrix diagonalisation equation given with (4.6), i.e.,

$$C_X = UC_Y U^T$$

(4.10)

It is clear that the linear transformation to achieve diagonal covariance matrix $C_Y$ can be accomplished using matrix $S$ whose columns are the eigenvectors of the original data matrix $X$, thus $S \equiv U^T$. Rows of transformed data matrix $Y$ are usually referred to as PCs. Slight ambiguity
in using this term should be pointed out at this stage as some of the surveyed literature and websites uses this term for the columns of transformation vector $U$, i.e. eigenvectors which is not correct.

$Y$ can be mapped back to $X$ by inversing the transformation, which is simplified since the inverse of the transformation is equal to its transpose, i.e.,

$$U^{-1} = U^T$$ \hspace{1cm} (4.11)

This can be proven by reconsidering the form given with the equation (4.10)

$$\left( U^T \right)^{-1} C_Y U^{-1} = C_X$$ \hspace{1cm} (4.12)

Since the covariance matrix is symmetric, i.e. $C_X = C_X^T$ then

$$\left( U^T \right)^{-1} C_Y U^{-1} = \left( U^{-1} \right)^T C_Y \left( \left( U^T \right)^{-1} \right)^T$$ \hspace{1cm} (4.13)

which implies that

$$\left( U^T \right)^{-1} = \left( U^{-1} \right)^T \text{ and } U^{-1} = \left( \left( U^T \right)^{-1} \right)^T$$ \hspace{1cm} (4.14)

These equations can only be true if the inverse of $U$ is equal to its transpose. Thus, to obtain $X$ from the transformed dataset $Y$, equation (4.1) can be rewritten as:

$$Y^T = X^T S^T = X^T \left( U^T \right)^T = X^T U$$ \hspace{1cm} (4.15)

which after multiplication with $U^{-1}$ results in:

$$Y^T U^{-1} = X^T U U^{-1}$$ \hspace{1cm} (4.16)
or

\[ X = (Y^T U^{-1})^T \]  \hspace{1cm} (4.17)

Taking into consideration that \( U^{-1} = U^T \) equation (4.17) can be rewritten as

\[ X = (Y^T U^{-1})^T = (Y^T U^T)^T = U Y \]  \hspace{1cm} (4.18)

This completes the round trip from the original data \( X \) to the PCs contained in \( Y \) and back to the original data. It is worth pointing out that to completely recover the original data contained in \( X \), mean for each dimension removed in (4.3) should be added back to the data recovered in (4.18). This step has been avoided for simplicity, but should be considered when implementing PCA into software.

4.3.1.1.2 PCA via Singular Value Decomposition of Data Matrix

In practice, PCs of data matrix \( X \) can also be obtained directly – via Singular Value Decomposition (SVD) of data matrix \( X \). SVD decomposes \( X \) into three new matrices as shown in Figure 4.3.

![Figure 4.3: Graphical depiction of SVD of data matrix X.](image-url)
SVD is a known algorithm for “diagonalisation” of rectangular matrices [105]. The diagonalising process produces diagonal and two orthonormal matrices. Mathematically, the SVD of data matrix \( X \) is given by:

\[
X = U \Sigma V^T
\]  

(4.19)

In the context of GPR data, the centred data matrix is usually transposed first to perform the SVD:

\[
X^T = (U \Sigma V^T)^T = V \Sigma^T U^T = V \Sigma U^T
\]  

(4.20)

since matrix \( \Sigma \) is a diagonal matrix and \( \Sigma = \Sigma^T \). The covariance matrix \( C_X \), can now be expressed as:

\[
C_X = XX^T = U \Sigma V^T V \Sigma U^T = U \Sigma \Sigma U^T = U \Sigma^2 U^T
\]  

(4.21)

Result obtained in (4.21) compared to equation (4.6) indicates that matrix \( U \) obtained via SVD again contains the eigenvectors of data matrix \( X \). Matrix \( \Sigma^2 = D \) contains the eigenvalues of \( X \), i.e.,

\[
\Sigma^2 = \text{diag} \{ \lambda_1, \lambda_2, \ldots, \lambda_n \}
\]  

(4.22)

The columns of \( U \ (m \times n) \) and \( V \ (n \times n) \) are sometimes called the left and right singular vectors, which are the eigenvectors of two covariance matrices, \( XX^T \) and \( X^T X \) respectively. Consequently,

\[
UU^T = U^T U = I
\]

\[
VV^T = V^T V = I
\]  

(4.23)

where \( I \) is the identity matrix.
Elements of matrix $\Sigma$ are non-negative values, known as singular values of $X$, whilst the columns of matrix $U$ and rows of matrix $V$ represent the eigenvectors of matrices $XX^T$ and $X^TX$ respectively. This is easily observed from spectral decomposition of $XX^T$ and $X^TX$:

$$XX^T = U \Sigma V^T \left( U \Sigma V^T \right)^T = U \Sigma V^T \Sigma U^T = U \Sigma^2 U^T$$

(4.24)

and

$$X^TX = \left( U \Sigma V^T \right)^T \Sigma V^T = \Sigma U^T \Sigma V^T = \Sigma^2 V^T$$

(4.25)

### 4.3.1.1.3 Dimensionality Reduction using PCA

This section explains the process of dimensionality reduction of original dataset $X$. Regardless of the approach used to obtain transformation matrix $U$, the process is the same as explained in the rest of this section.

The amount of information, variance, contained in each eigenvector from $U$ is indicated by the corresponding eigenvalues from $\Sigma^2 = D$. Singular values in $\Sigma$ or eigenvalues from $D$ can thus be sorted in descending order and the corresponding eigenvectors in $U$ and $V$ are reordered in the same way. Denoting the reordered matrices with $\hat{U}$ and $\hat{V}$, PCA transformation of data matrix $X$ can again be obtained as:

$$Y = \hat{U}^T X$$

(4.26)

By selecting only the first $k$ rows from the $n \times n$ matrix $Y$, the $k \times n$ matrix $Y_k$ is obtained. Thus data from $X$ is in this way projected from $n$ to $k$.
dimensions. Since the original data contained in matrix $X$ can be fully recovered using:

$$X = \hat{U}Y$$  \hspace{1cm} (4.27)

partial recovery of the original data can be achieved by:

$$X_{\text{rec}} = \hat{U}_kY_k$$  \hspace{1cm} (4.28)

where $\hat{U}_k$ represents the matrix obtained from $\hat{U}$ by discarding the last $n-k$ columns of matrix $\hat{U}$. Therefore, the matrix $X_{\text{rec}}$ now contains most of the information from the original matrix having lost the details contained in the last $n-k$ eigenvectors.

Summary of both approaches, including the dimensionality reduction and data reconstruction steps is given in Table 4.1 using Matlab code. Identical result can be achieved via modified SVD equation:

$$X_{\text{rec}} = \hat{U}_k \hat{\Sigma}_k \hat{V}_k^T$$  \hspace{1cm} (4.29)

where $\hat{U}_k$, $\hat{\Sigma}_k$ and $\hat{V}_k$ represent the reordered matrices $U$, $\Sigma$ and $V$. The reordered diagonal matrix $\hat{\Sigma}_k$ with $n-k$ diagonal elements is set to zero, the multiplication given by equation (4.29) effectively rejects the last $n-k$ eigenvectors from $\hat{U}$ (columns) and $\hat{V}$ (rows).
Multivariate Pre-processing Techniques for GPR Images

Table 4.1: Two ways of performing PCA including common dimensionality reduction and data reconstruction steps implemented in Matlab; left column – Eigenvalue Decomposition approach and right column – SVD approach.

Data set/matrix: $\mathbf{X}$ (m×n)  
- $m$ – number of dimensions/variables (samples in each recorded trace)  
- $n$ – number of measurements/trials (individual traces)

% get the size of the data set
[m,n] = size(X);
% obtain the mean trace and remove it from each measurement
% note: this effectively removes the mean for each dimension
mn = mean(X,2);
X = X - repmat(mn,1,n);

%% via: eigenvalue decomposition
% calculate the covariance matrix
Cx = 1 / (n-1) * X * X';
% find the eigenvectors and eigenvalues
[U, D] = eig(Cx);
% extract eigenvalues to vector
D = diag(D);
% sort the eigenvalues and eigenvectors in decreasing order
[junk, rindices] = sort(-1*D);
D = D(rindices);
U = U(:,rindices);

% transform the original data
Y = U'*X;
% reduce data dimensionality
% i.e. consider only the first r rows of Y
Yr = Y(:,1:k);

%% via SVD
% transpose and scale the data
% matrix needed because Matlab “svd” function assumes
% measurements in rows, not in columns as with our GPR data
X = X' / sqrt(n-1);
% SVD
[U,S,V] = svd(X);
% calculate the eigenvalues
S = diag(S);
D = S .* S;
% no need for sorting, SVD did it all

% recover the original data
% addition of mean trace is optional
% for image analysis tasks
Xrec = U'*Y + mn;
% recover the original data with losses
% due to dimensionality reduction
Xkrec = U(:,1:k)'*Yr + mn;
4.3.1.1.4 GPR Clutter Reduction using PCA

For most natural images, eigenvalues decrease so rapidly and it is usually necessary to retain only a few largest values from $\Sigma$ and the corresponding eigenvectors from $U$ to significantly reduce dimensionality of data from $X$ with minimal loss of information, particularly taking into consideration that the last eigenvectors usually represent measurement noise. Decluttering of GPR images requires somewhat reverse reasoning. For GPR images where clutter or other non-target related signals usually dominate the GPR image, most of this information is contained in the first few eigenvectors from $U$. It is therefore possible to reduce the amount of clutter for the SVD-reconstructed image by setting the largest (i.e. first) or a couple of largest singular values from diagonal matrix, $\Sigma$ to zero.

As previously explained, discarding the last $n-k$ columns and rows from $\hat{\Sigma}$ would also reduce the amount of noise present in dataset $X$ without affecting the amount of information contained in the GPR image. Therefore, in order to remove both clutter and noise present in the original dataset $X$, the following SVD reconstruction can be used:

$$X_{rec} = \hat{U}_R \hat{\Sigma}_R \hat{V}_R^T$$  \hspace{1cm} (4.30)

where:

- $\hat{U}_R$ is an $m \times r$ orthonormal matrix ($r = k - l$) obtained from matrix $U$ by reordering the columns corresponding to the descending values
of diagonal matrix $\Sigma$ and discarding the first few $l$ as well as the last $n-k$ columns;

- $\hat{V}_r$ is an $r \times n$ orthonormal matrix obtained from matrix $V$ by reordering the rows corresponding to the descending values of diagonal matrix $\Sigma$ and discarding the first few $l$ as well as the last $n-k$ rows;

- $\hat{\Sigma}_r$ is an $r \times r$ diagonal matrix obtained from reordered matrix $\hat{\Sigma}$ by discarding the first few element $l$ as well as the last $n-k$ columns and rows, i.e., $\hat{\Sigma}_r = \text{diag}\{\sigma_2, \sigma_3, \ldots, \sigma_k\}$ and $\sigma_2 > \sigma_3 > \ldots > \sigma_k$.

The above approach can also be recast in terms of superposition of individual eigenimages of the original GPR image $X$. Each eigenimage $E_i$ is obtained by multiplying a single row from matrix $U, u_i$, with the corresponding but transposed row from matrix $V, v_i^T$:

$$E_i = u_i v_i^T$$  (4.31)

The contribution of each eigenimage to the original dataset is described by the corresponding singular value from $\Sigma$. The original GPR image is then easily obtained by scaling and adding all eigenimages, where the scaling factor for each eigenimage is the corresponding singular value from $\Sigma$.

$$X = \sum_{i=1}^{a} \sigma_i E_i$$  (4.32)
Thus, by discarding the strongest $l$ eigenimages as well as the weakest $n-k$ eigenimages from the measured GPR image, a decluttered and denoised GPR image can be obtained:

$$X_{\text{rec}} = \sum_{i=l}^{n-k} \sigma_i E_i$$  \hspace{1cm} (4.33)

Figure 4.4: One-dimensional Principal Component Analysis (1DPCA).

4.3.1.2 Two-Dimensional Principal Component Analysis (2DPCA)

One-dimensional PCA (1DPCA) can also be performed on the group of images or other 2D datasets. The traditional approach is to turn each of the 2D datasets into a 1D sequence by placing each column of the individual image below the previous one and then generating a new 2D dataset by
placing each “column image” into a new data matrix. 1DPCA via SVD is then being performed on this new dataset. Recently, an extension of PCA, called two-dimensional PCA (2DPCA) [107] has been developed in order to address some shortcomings of this traditional way that 1DPCA is performed on 2D datasets. 2DPCA starts by computing a new, 2D data matrix:

$$H = \frac{1}{m} \sum_{i=1}^{m} (X_i - \mu)$$

(4.34)

Here $X_i$ represents the $i$-th image or some other 2D dataset from the group of $m$ such images and $\mu$ refers to the average of all images in the set, calculated as:

$$\mu = \frac{1}{m} \sum_{i=1}^{m} X_i$$

(4.35)

The PCs are then computed in the usual way as in 1DPCA using SVD of data matrix $H$. This approach has been used in some image recognition tasks and has been reported to provide a higher recognition rate with a smaller computational complexity compared to 1DPCA [108]. Detailed analysis of this approach has also been performed in [109].

In this work, 2DPCA approach has been used to remove clutter and other types of strong interference from the set of GPR data images obtained in the same measurement session, under the assumption that the first $l$ PCs for each collected image contains most of the unwanted information contained in all images from the group. Once the PCs are obtained using
2DPCA technique, decluttering procedure is performed in the same way as in the case of 1DPCA.

Figure 4.5: Two-dimensional Principal Component Analysis (2DPCA).

Unfortunately, whilst PCA can be, and is, used to transform one set of variables into another smaller set, the newly created variables are not usually easy to interpret. PCA has been most successful in applications such as image compression where data reduction – and not interpretation – is of primary importance. In many applications, PCA is used only to provide information on the true dimensionality of a dataset [103].
4.3.1.3 Order Selection

The obtained eigenvalues of PCA describe the importance of the variance to the associated PC and can be very helpful to determine the components that are really significant as well as how much the dataset can be reduced when ordering by size. The associated PCs contribute little to the data if several eigenvalues are zero or close to zero, thus can be eliminated. The eigenvalue threshold can be determined by [103]:

- eliminate some percentage below the sum of all eigenvectors;
- eigenvalues is plotted against the order number to look for breakpoints in the slope of the curve which the eigenvalues of noise should not change much in value and hence produce a flatter slope.

4.3.2 Independent Component Analysis (ICA)

The multivariate techniques PCA discussed previously differ from ICA in the criteria applied to the transformation. In PCA, the object is to transform the dataset so as to produce a new set of variables (termed PCs) that are uncorrelated. The goal is to reduce the dimensionality of the data, not necessarily to produce more meaningful variables. The application of both PCA approaches described earlier show that decorrelating the data is not sufficient to produce independence between the variables, at least when the variables have non-Gaussian distributions and resulting in the mutually orthogonal set of eigenvectors which makes it difficult to interpret or assign
any physical meaning to the obtained PCs. Although obtained PCs are decorrelated, they still represent a linear combination of independent variables from the original dataset. This inability to make the extracted components independent of each other provides the motivation to try and use an alternative technique known as Independent Component Analysis (ICA) [93], [110] for further improvement of PCA decluttered GPR images. The goal of ICA is however more ambitious which is finding new variables (termed independent components, ICs) that are both statistically independent and non-Gaussian [103].

The fundamental principle of ICA [93], [110] consists in estimating the set of independent signals from the mixture of the given signals by estimating the un-mixing matrix. ICA seeks to transform the original dataset into a number of independent variables. The motivation for this transformation is primarily to uncover more meaningful variables and not to reduce the dimensions of the dataset. Thus, pre-processing the dataset using PCA is needed when it involves in dataset reduction [103].

ICA is sometimes referred to as a Blind Signal Separation (BSS). The term ‘blind’ is intended to imply that BSS separates data into source signals when no or very little information about the nature of those source signals is available [110], [111] although it is worth noting that some researchers make distinction between the ICA and BSS methods.
The spirit of the suggested method for ICA clutter reduction resembles that of the PCA based technique. The most significant computational difference between ICA and PCA is the fact that PCA uses only second-order statistics whilst ICA employs higher-order statistics. In contrast to correlation-based transformations such as PCA, ICA not only decorrelates the signals (second-order statistics) but also reduces higher-order statistical dependencies, attempting to make the signals as independent as possible.

Restricting the analysis to use only second-order statistics gives satisfactory results for signals with a Gaussian distribution where statistical moments above second-order are zero. However, the majority of signals measured in nature (natural signals), do not have a Gaussian distribution and have higher-order moments. This property is exploited by ICA as it assumes the statistical independence and non-Gaussian nature of original signal components [25]. The subspace formed by ICA is therefore not orthogonal as is the case with components extracted by PCA.

The mathematical model for ICA, assuming the same structure of data matrix $X$ as in the PCA case can be written succinctly as:

$$X = SA$$  \hspace{1cm} (4.36)

where $m \times n$ matrix $S$ now represents the matrix containing $n$ original source signals as its columns and $A$ is the mixing matrix. Matrix $X$ is therefore represents a set of $n$ mixtures of original signals contained in $S$ and element $a_{ij}$ from the mixing matrix $A$ specifies the contribution of the $i$-th
source signal (column) from $S$ to $j$-th mixture (column) of $X$. The main computational issue in ICA process is the estimation of mixing matrix $A$, so that the ICs, i.e. columns of matrix $S$ can be obtained by:

$$S = XA^{-1}$$  \hspace{1cm} (4.37)

To estimate the mixing matrix, ICA makes two basic assumptions: that the source variables are truly independent and that they are non-Gaussian. Both conditions are usually satisfied when the sources are real signals. A third restriction is that the mixing matrix $A$ must be square, i.e. the number of sources should equal to the number of measured signals. This problem is usually solved by using PCA to pre-process the matrix $X$ and reduce its dimensions, i.e. new matrix $X_K$ can be formed in such a way that the number of columns in $X_K$ becomes equal to the number of columns in $S$.

The ICA technique can in analogy to PCA be reconsidered in terms of independent images, where $i$-th independent image of the original GPR image $X$, can be obtained by multiplying the $i$-th column from $S$, $s_i$ with the $i$-th row, from the mixing matrix $A$, $a_i^T$:

$$I_i = s_i a_i^T$$  \hspace{1cm} (4.38)

Original image can be reconstructed through the addition of all independent images. GPR image with much stronger target information compared to original image can be obtained by rejecting non-target related independent images before the reconstruction.
One of the problems related to ICA used in this context is that, unlike PCA, there is no direct method to estimate the importance of each extracted independent variable, as can be done in PCA by sorting the eigenvalues in $\Sigma$. Although some approaches to automatically order components extracted by ICA procedure have been reported [112], [113], [114], this question remains open in research literature. Thus, the visual inspection of obtained independent images has been used in this work to decide which of the independent images represents clutter and other non-target related information in order to reject those images. In addition to that, ICs can be randomly scaled and permutated after the ICA process. The ICA source separation and source estimation process is depicted in Figure 4.6 [111].

![Blind Source Separation (BSS) block diagram](image)

Figure 4.6: Blind Source Separation (BSS) block diagram. $s(t)$ are the sources, $x(t)$ are the recordings, $\hat{s}(t)$ are the estimated sources, $A$ is a mixing matrix and $W$ is an un-mixing matrix [111].
4.3.2.1 Independent Component Analysis Algorithms

Implementation of an ICA algorithm is not a trivial task, however, a number of excellent algorithms can be found such as Fixed-Point ICA (FastICA) [115], information maximisation (Infomax) [116], temporal decorrelation source separation (TDSEP) [117] and Joint Approximate Diagonalisation of Eigen-matrices (JADE) [25], [99], [118] which have been used widely in numerous signal processing applications. Each algorithm used a different approach to solve equation [111]:

Figure 4.7: ICA method in GPR application.
• FastICA – it employs higher-order statistics for the recovery of independent sources that uses simple negentropy estimation based on the maximum entropy principle [93], [115];

• Infomax – it is a gradient-based neural network algorithm that uses higher-order statistics with a learning rule for information maximisation which is attained by maximising the joint entropy of a transformed vector [116];

• TDSEP – it is based on the simultaneous diagonalisation of several time-delayed correlation matrices that able to separate signals whose amplitude distribution is near Gaussian [117];

• JADE – it is based on the joint diagonalisation of cumulant matrices under the assumption that the sources have non-Gaussian distributions. This algorithm uses second and fourth-order cumulants that measured the mutual information between cross cumulants. Second-order cumulant is used to decorrelate the data, i.e. to obtain a whitening matrix $\hat{Z}$. The separation matrix is estimated as $R'\hat{Z}$, where $R'$ is a rotation matrix used to make the cumulant matrices as diagonal as possible [119].

In this work, ICA is used via JADE algorithm to process the GPR data.
4.3.3 Proposed Novel Algorithm

The objective of PCA is primarily to reduce the dimensionality of the data whilst ICA is to uncover more meaningful variables of the dataset. Majority of measured signals including GPR are non-Gaussian and above the second-order moments. This work is therefore proposed a combination of these two methods, called SVDPICA – Singular Value Decomposition of PCA and ICA for data reduction and interpretation.

The motivating behind this idea is the problem of orthogonality of the extracted PCs in PCA method (obtained by reconsidering the SVD data reconstruction equation (4.30)) and the problem of ordering the obtained ICs of ICA algorithm. Realising that the columns of matrix $\hat{U}_R$ in equation (4.30) are orthogonal but not mutually independent, ICA can be performed on matrix $\hat{U}_R$, thus estimating the mixing matrix $A$ such that:

$$\hat{U}_R = \hat{U}_R A \hspace{3.5cm} (4.39)$$

New matrix $\hat{U}_R$ obtained in this ICA process now represents a matrix with un-mixed PCs from matrix $X$ as its columns. Those are now identical to ICs that can be obtained by performing the ICA on the original data matrix $X$. This can therefore be combined with the SVD reconstruction equation (4.30) such that:

$$X_{rec} = \hat{U}_R \hat{\Sigma}_R \hat{V}_R^T = \hat{U}_R A \hat{\Sigma}_R \hat{V}_R^T = \hat{U}_R \hat{T}_R \hspace{3.5cm} (4.40)$$

where
\[ T_R = A \hat{\Sigma}_R \hat{V}_R^T \]  

(4.41)

Newly obtained matrix \( T_R \) can be “decomposed” into two new matrices \( \hat{V}_R^T \) and \( \hat{\Sigma}_R \), i.e.:

\[ T_R = A \hat{\Sigma}_R \hat{V}_R^T = \hat{\Sigma}_R \hat{V}_R^T \]  

(4.42)

in such a way that one of the two matrices, \( \hat{\Sigma}_R \), becomes diagonal matrix. This can be achieved by setting each element of new diagonal matrix \( \hat{\Sigma}_R \) to be:

\[ \sigma_{i,j} = norm(\hat{t}_{Ri}) \]  

(4.43)

where \( \hat{t}_{Ri} \) represents the \( i \)-th column of matrix \( T_R \). Each column of new matrix \( \hat{V}_R \), \( \hat{v}_{Ri} \) is obtained by scaling the columns from \( T_R \) in the following way:

\[ \hat{v}_{Ri} = \frac{\hat{t}_{Ri}}{norm(\hat{t}_{Ri})} \]  

(4.44)

Thus, a “new” SVD like decomposition of data matrix \( X \) is obtained:

\[ X_{rec} = \hat{U}_R \hat{\Sigma}_R \hat{V}_R^T \]  

(4.45)

with the additional “quality” of containing the matrix of un-mixed eigenvectors \( \hat{U}_R \) as well as the “importance” matrix \( \hat{\Sigma}_R \). Discarding one or more first and last eigenvectors from \( \hat{U}_R \) can now further refine and improve decluttering and denoising effects achieved via PCA or simple mean removal in the first stage of this process. This can be easily automated
as the information about the importance of each eigenvector from $\tilde{U}_r$ is contained in the diagonal elements of matrix $\tilde{\Sigma}_r$. Summary of the proposed SVDPICA method is illustrated in Figure 4.8.

Figure 4.8: Three stages of SVDPICA method – SVD decomposition of original data matrix $X$, discarding of the first $l$ and last $n-k-l$ PCs, ICA unmixing the reduced matrix of eigenvectors $\tilde{U}_r$ to obtain matrix $\tilde{U}_r$.

4.4 Results

The main objective of this phase of the project is to reduce the amount of clutter and other unwanted signals present in the GPR data which are not
related to the target characteristics. Techniques described in the previous section have at this stage been used and applied in two related but different tasks:

- uniform (or close to uniform) clutter reduction – for this purpose, PCA and SVDICA methods have been applied and tested on the simulated and measured data to detect pipes, leakages from joints, holes and cracks, and cracked concrete;

- non-uniform clutter reduction – significant amount of GPR data was measured and collected in the experiment aimed at detecting and characterising buried pipework. This data contained strong unwanted signals originating from the sides of the boxes used to hold buried pipes. Those signals varied non-uniformly across the image and as such could not be reduced with other, conventional clutter reduction techniques. The two-dimensional PCA and ICA algorithms have been applied to this data with the aim to improve image quality by removing target uncorrelated features from the image and prepare data for classification tasks which is envisaged to be the next stage in the project.

### 4.4.1 Algorithm Evaluation Parameters

The performance of applied algorithms in clutter and unwanted signal reduction tasks has been evaluated in terms of Signal-to-Noise Ratio (SNR). SNR can be described as the ratio of average energy of the image matrix
after clutter and unwanted signals reduction to the average of matrix containing clutter and noise [120]:

\[
\text{SNR} = \frac{P_{\text{clutter and unwanted signals reduced image}}}{P_{\text{clutter + noise}}}
\]  

(4.46)

where the average energy of the clutter and noise can be retrieved by subtracting it from raw B-scan image. The idea behind this approach is that the processed GPR image should contain the information about the target only with no clutter, noise and unwanted signals present in the data. It is however difficult to obtain a clutter-and-noise-free dataset for GPR images to be used as a reference for evaluating its performance in terms of SNR and related measures. Therefore, in addition to Peak-Signal-to-Noise Ratio (PSNR), Mean Structural Similarity (M-SSIM) Index measure has also been used in this work to evaluate the performance of the proposed techniques and give a better sense of validity as well as to indicate the potential advantages of the techniques for clutter and unwanted signals reduction developed and applied in this project. Both evaluation measures PSNR and M-SSIM index are discussed in details in the following sections.

4.4.1.1 Peak-Signal-to-Noise Ratio (PSNR)

Given two 8-bit grey-level, \(m \times n\) images, a reference image, \(e\) and a test image, \(g\), the PSNR between \(e\) and \(g\) is defined by [121]:

\[
\text{PSNR}(e, g) = 10\log_{10}\left(255^2 / \text{MSE}(e, g)\right)
\]

(4.47)
where

\[
\text{MSE}(e, g) = \frac{1}{mn} \sum_{i=1}^{m} \sum_{j=1}^{n} (e_{ij} - g_{ij})^2
\] (4.48)

For identical images, the MSE approaches zero and the PSNR value tends to infinity. At the other end of the scale, a small value of PSNR implies high numerical differences between two images which in the case of GPR images could indicate high amount of noise removed from the raw image. This can however be misleading as the removed components could also contain a useful part of the GPR signal. To account for this possibility, in addition to PSNR measure, measure of structural similarity is also used in this work.

4.4.1.2 Mean Structural Similarity (M-SSIM) Index

Measure of Structural Similarity (SSIM) index is used for quality assessment that indicates the resemblance between the image processed via one of the proposed methods and the “ideal” image with the values between 0 and 1, where 1 represents the identical images. Diagram in Figure 4.9 [122] indicates how the SSIM Index between two images is calculated through comparison in three stages – luminance, contrast and structure. Assuming two signals, \( x \) and \( y \), the similarity measure is defined by considering one of the signals having perfect quality which can then be used as a quantitative measurement of the second signal quality.
Figure 4.9: Diagram of the structural similarity (SSIM) measurement system [122].
The luminance, contrast and structure comparisons of two signals \((x, y)\) are computed using the following functions [122]:

- **Luminance comparison:**

\[
l(x, y) = \frac{2 \mu_x \mu_y + C_1}{\mu_x^2 + \mu_y^2 + C_1} \tag{4.49}
\]

- **Contrast comparison:**

\[
c(x, y) = \frac{2 \sigma_x \sigma_y + C_2}{\sigma_x^2 + \sigma_y^2 + C_2} \tag{4.50}
\]

- **Structure comparison:**

\[
s(x, y) = \frac{\sigma_{xy} + C_3}{\sigma_x \sigma_y + C_3} \tag{4.51}
\]

where

- Mean intensity, \(\mu_x = \frac{1}{n} \sum_{i=1}^{n} x_i\)
- Mean intensity, \(\mu_y = \frac{1}{n} \sum_{i=1}^{n} y_i\)
- Base contrast, \(\sigma_x = \left( \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \mu_x)^2 \right)^{\frac{1}{2}}\)
- Base contrast, \(\sigma_y = \left( \frac{1}{n-1} \sum_{i=1}^{n} (y_i - \mu_y)^2 \right)^{\frac{1}{2}}\)
- Base contrast, \(\sigma_{xy} = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \mu_x)(y_i - \mu_y)\)
- Constant \(C_1 = (K_1L)^2\)
- Constant \(C_2 = (K_2L)^2\)
- Constant \(C_3 = \frac{C_2}{2}\)
constant $K_1 = 1$
constant $K_2 = 1$
$L = \text{dynamic range of pixel values}
\quad (255 \text{ for 8-bit grayscale image}).$

Combining these three relations results in similarity measure the SSIM index between signals $x$ and $y$ which is:

$$SSIM(x, y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{\left(\mu_x^2 + \mu_y^2 + C_1\right)\left(\sigma_x^2 + \sigma_y^2 + C_2\right)}$$ (4.52)

The above equations are given for two vectors, i.e. one-dimensional signals only. In the context of image processing, there are more than two vectors to be compared and a single overall quality measure of the entire image is usually required. Therefore, a mean of SSIM index of the corresponding columns of the two images is computed resulting in the mean structural similarity (M-SSIM) index [122] given as:

$$M-SSIM(X, Y) = \frac{1}{m} \sum_{j=1}^{m} SSIM(x_j, y_j)$$ (4.53)

where $X$ and $Y$ are the reference ("ideal") and the processed images respectively, $x_j$ and $y_j$ are the image contents at the $j$-th local window, and $m$ is the number of local windows of the image.

The SSIM index is a method for measuring the similarity between two images that can be viewed as a quality measure for one of the images being compared if the other image is regarded as the image of desired image quality. It is an improved version of the universal image quality index
proposed earlier [123] in order to improve the image quality assessment in a number of applications by measuring structural distortion of the image. In comparison to MSE or PSNR type measures which are mainly sensitive to the error energy, the SSIM index exhibits higher consistency with subjective quality measurement but avoids using complex human visual system (HVS) based indicators which usually result in a high computational complexity without clear advantage over simple MSE or PSNR type measures.

4.4.2 Simulated Data

As the overall aim of this research is to evaluate the potential and applicability of GPR in detecting pipes, leakages from joints, holes and cracks, and cracked concrete, simulations and laboratory measurements have been performed in order to replicate those conditions. However, due to data complexity of metal fibre presence in the cracked concrete mixture slab, the simulated data was not able to be reproduced. Analysis of this data is therefore can not be performed to any further extent.

The evaluation of the simulated data has been done in terms of PSNR and M-SSIM index to indicate the performance of the proposed clutter reduction methods. The simulated data has the benefit of controlled the amount of clutter and noise present in the GPR image, thus the “ideal” image of simulated data can be simulated directly or obtained by subtracting the clutter image from the original raw GPR image to be used for the evaluation purposes. The performance of PSNR and M-SSIM index is
calculated by comparing the degraded image to the “ideal”, undistorted image. High values of PSNR and M-SSIM index indicates high similarity between the “ideal” and distorted image thus indicating a possible advantage of the proposed pre-processing technique.

4.4.2.1 Clutter Reduction

Two multivariate techniques – PCA and SVDPICA, have been applied to the simulated GPR images in testing the effectiveness of the proposed methods in reducing the amount of clutter presented in GPR data. Initial results and the entire process in this section illustrated using the image simulated on joints with 1.5 GHz radar frequency. Similar procedure is repeated for other simulated images and results are summarised at the end of the chapter.

Figure 4.10 (a) illustrates the eigenspectra of selected GPR data via SVD. The dimensionality reduction has been done by removing the first eigenvector which represents clutter with the highest variance as well as the components with very small variances – claimed to be noise. The selection number of PCs to be used is 30 and the results of clutter reduction approaches are given Figure 4.11. Basic mean removal (MR) technique has also been included to evaluate the performance of the proposed methods. It can be observed that the joint features are well recovered using both techniques, as well as the basic mean removal method. Also, the strong signals from the edge of the box (both sides) can easily be distinguished as
much as the joint signals. Table 4.2 and Table 4.3 summarise the performance of the decluttering techniques of the remaining simulated data of the performed experiments described in Chapter 2.

Figure 4.10: Plot of eigenvalues against space sample (component number) of simulated GPR data represent joints (a) with PCA via SVD and (b) selected 30 PCs via SVDPICA method.
Figure 4.11: Results of GPR image decluttering on simulated data represent joints using different techniques (a) original raw image, (b) mean removal, (c) PCA via SVD, (d) SVDPCA and (e) “Ideal” data which is obtained by subtracting the image of concrete from the image of concrete containing joints.
Table 4.2: PSNR and M-SSIM index indicators for described clutter reduction algorithms of simulated data experiment to detect leakages from joints, holes and cracks.

<table>
<thead>
<tr>
<th>Data</th>
<th>Features</th>
<th>PSNR</th>
<th>M-SSIM Index</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>MR</td>
<td>PCA</td>
</tr>
<tr>
<td>1</td>
<td>Joints</td>
<td>70.4701</td>
<td>71.9973</td>
</tr>
<tr>
<td>2</td>
<td>Holes</td>
<td>74.3522</td>
<td>72.1234</td>
</tr>
<tr>
<td>3</td>
<td>Crack1</td>
<td>61.8449</td>
<td>62.1144</td>
</tr>
<tr>
<td>4</td>
<td>Crack2</td>
<td>61.4218</td>
<td>61.6739</td>
</tr>
<tr>
<td>5</td>
<td>Crack3</td>
<td>61.4232</td>
<td>61.6747</td>
</tr>
<tr>
<td>6</td>
<td>Crack4</td>
<td>61.2889</td>
<td>61.5340</td>
</tr>
</tbody>
</table>

Figure 4.12: Evaluation of clutter reduction methods in terms of PSNR for simulated data to detect leakages from joints, holes and cracks.
Table 4.3: PSNR and M-SSIM index indicators for described clutter reduction algorithms of simulated data experiment to detect pipes.

<table>
<thead>
<tr>
<th>Data</th>
<th>Features</th>
<th>PSNR</th>
<th>M-SSIM Index</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>MR</td>
<td>PCA</td>
</tr>
<tr>
<td></td>
<td></td>
<td>MR</td>
<td>PCA</td>
</tr>
<tr>
<td>1</td>
<td>Big</td>
<td>78.1583</td>
<td>80.0656</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.9999</td>
<td>0.9999</td>
</tr>
<tr>
<td>2</td>
<td>Medium</td>
<td>70.5832</td>
<td>70.4405</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.9992</td>
<td>0.9992</td>
</tr>
<tr>
<td>3</td>
<td>Small</td>
<td>71.4841</td>
<td>70.7910</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.9994</td>
<td>0.9993</td>
</tr>
<tr>
<td>4</td>
<td>All pipes</td>
<td>73.7019</td>
<td>73.0961</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.9996</td>
<td>0.9995</td>
</tr>
</tbody>
</table>

Figure 4.13: Evaluation of clutter reduction methods in terms of PSNR for simulated data in detecting pipes.
4.4.2.2 Unwanted Signals Reduction

The proposed unwanted signals reduction methods used 2DPCA and ICA to encounter the problem of unwanted features appears in GPR images. Due to inability of simulation software to generate equivalent set of data, this section only provides results achieved by pre-processing measured datasets.

4.4.3 Measured Data

The laboratory experiments have been performed on three different sets of experiments as described in Chapter 2. The types of radar used to perform each experiment are:

- pipes detection – GPR GV5;
- detection of leakages from joints, holes and cracks – GPR GV5 and GV3 CDH system;
- cracks detection on metal fibre mixtures concrete slab – GPR GV5.

The evaluation results for GPR images employed both PSNR and M-SSIM index indicators are summarised in the next sections. In the situations where degraded image can be compared to original, undistorted image, high PSNR value indicates high similarity between the original and distorted image, i.e. low level of noise in the distorted image. However, in reality, this undistorted image is hardly available thus the processed images can only be compared to original, distorted image. Lower PSNR value should therefore indicates “better” algorithm performance as this would imply increase in
RMS difference between the raw and the processed images. However, care should be taken when interpreting the PSNR values of measured GPR data as the large RMS difference between the original and processed images might also imply that some of the important, target related features in the raw image are not very well preserved.

For the similarity performance, the “ideal” image is used as a reference to be compared to the processed data. This “ideal” image is obtained from a trained human – GPR image analyst by using ReflexW software. Although some of the steps in ReflexW can actually distort the original image, M-SSIM index measure can still be used to assess the validity of automated clutter reduction techniques in comparison to supervised GPR image processing.

4.4.3.1 Clutter Reduction

The same procedures of decluttering methods to simulated data have been applied to evaluate the performance of the proposed methods on experimental measured data. As stated previously, clutter usually dominates the GPR data by having the strongest signal and most of the clutter information is contained in the first eigenvector with the highest variance whilst noise corresponds to the components with very small variances. A dimensionality reduction can therefore be obtained by discarding those components.
Figure 4.14: Plot of eigenvalues against space sample (component number) of measured GPR data in detecting pipes – big size (a) with PCA via SVD and (b) selected 30 PCs via SVDPICA method.
Figure 4.15: Results of GPR image decluttering on measured data – big pipe using different techniques (a) original raw image, (b) mean removal, (c) PCA via SVD, (d) SVDPCA and (e) ReflexW software.
Figure 4.15 shows the obtained results of clutter reduction for one measured GPR image – big pipe. Pipe feature is located towards the middle of the image (traces 70 – 90) and although it is well recovered using both proposed techniques, the SVDPCA method depicted in Figure 4.15 (d) removes more clutter from the original image, thus improved the visibility of the pipe related feature compared to the standard PCA method depicted in Figure 4.15 (c).

Table 4.4 illustrates the obtained PSNR and M-SSIM index results of clutter reduction methods for pipes detection whilst experiments to detect leakages from joints, holes and cracks are presented in Table 4.5 (GPR GV3) and Table 4.6 (GPR GV5). The evaluation of cracked concrete with metal fibre mixtures experiment is described in Table 4.7. Basic mean removal method has also been included for comparison purposes in assessing these methods.
Table 4.4: PSNR and M-SSIM index indicators for described clutter reduction algorithms of measured data experiment to detect pipes.

<table>
<thead>
<tr>
<th>Data</th>
<th>Features</th>
<th>PSNR</th>
<th>M-SSIM Index</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>MR</td>
<td>PCA</td>
</tr>
<tr>
<td>1</td>
<td>Big</td>
<td>63.0997</td>
<td>63.0092</td>
</tr>
<tr>
<td>2</td>
<td>Medium</td>
<td>64.0282</td>
<td>64.0084</td>
</tr>
<tr>
<td>3</td>
<td>Small</td>
<td>64.0399</td>
<td>64.0545</td>
</tr>
<tr>
<td>4</td>
<td>All pipes</td>
<td>63.8272</td>
<td>63.8189</td>
</tr>
</tbody>
</table>

Figure 4.16: Evaluation of clutter reduction methods in terms of PSNR for pipes detection of measured data.
Table 4.5: PSNR and M-SSIM index indicators for described clutter reduction algorithms of measured data experiment to detect leakages from joints, holes and cracks (GPR GV3).

<table>
<thead>
<tr>
<th>Data</th>
<th>Features</th>
<th>PSNR</th>
<th>M-SSIM Index</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>MR</td>
<td>PCA</td>
</tr>
<tr>
<td>1</td>
<td>Joints</td>
<td>66.7388</td>
<td>65.8015</td>
</tr>
<tr>
<td>2</td>
<td>Holes</td>
<td>68.8345</td>
<td>67.8838</td>
</tr>
<tr>
<td>3</td>
<td>Crack1</td>
<td>66.6830</td>
<td>63.4232</td>
</tr>
<tr>
<td>4</td>
<td>Crack2</td>
<td>66.4185</td>
<td>62.3803</td>
</tr>
<tr>
<td>5</td>
<td>Crack3</td>
<td>64.3476</td>
<td>64.6666</td>
</tr>
<tr>
<td>6</td>
<td>Crack4</td>
<td>63.0630</td>
<td>66.8946</td>
</tr>
</tbody>
</table>

Figure 4.17: Evaluation of clutter reduction methods in terms of PSNR for measured data to detect leakages from joints, holes and cracks (GPR GV3).
Table 4.6: PSNR and M-SSIM index indicators for described clutter reduction methods of measured experiment to detect leakages from joints, holes and cracks (GPR GV5).

<table>
<thead>
<tr>
<th>Data</th>
<th>Features</th>
<th>PSNR</th>
<th>M-SSIM Index</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>MR</td>
<td>PCA</td>
</tr>
<tr>
<td>1</td>
<td>Joints</td>
<td>66.3082</td>
<td>65.2599</td>
</tr>
<tr>
<td>2</td>
<td>Holes</td>
<td>65.0095</td>
<td>63.9253</td>
</tr>
<tr>
<td>3</td>
<td>Crack1</td>
<td>64.1771</td>
<td>63.5505</td>
</tr>
<tr>
<td>4</td>
<td>Crack2</td>
<td>64.3296</td>
<td>64.0813</td>
</tr>
<tr>
<td>5</td>
<td>Crack3</td>
<td>66.3176</td>
<td>65.9510</td>
</tr>
<tr>
<td>6</td>
<td>Crack4</td>
<td>61.0167</td>
<td>61.4956</td>
</tr>
</tbody>
</table>

Figure 4.18: Evaluation of clutter reduction methods in terms of PSNR for measured data to detect leakages from joints, holes and cracks (GPR GV5).
Table 4.7: PSNR and M-SSIM index indicators for described clutter reduction methods of measured experiment to detect cracked concrete.

<table>
<thead>
<tr>
<th>Data</th>
<th>PSNR</th>
<th>M-SSIM Index</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MR</td>
<td>PCA</td>
</tr>
<tr>
<td>1</td>
<td>62.6275</td>
<td>63.1238</td>
</tr>
<tr>
<td>2</td>
<td>62.1551</td>
<td>62.0268</td>
</tr>
<tr>
<td>3</td>
<td>65.8766</td>
<td>65.8387</td>
</tr>
<tr>
<td>4</td>
<td>63.7718</td>
<td>63.7623</td>
</tr>
<tr>
<td>5</td>
<td>62.7176</td>
<td>62.8797</td>
</tr>
<tr>
<td>6</td>
<td>63.8803</td>
<td>64.0282</td>
</tr>
</tbody>
</table>

Figure 4.19: Evaluation of clutter reduction methods in terms of PSNR for measured data to detect cracked concrete.
4.4.3.2 Unwanted Signals Reduction

In the pipes detection experiment, the main feature present in the GPR images involved with an empty container box is the strong signal on the left and right sides indicated with black and white arrows (Figure 4.20) which caused by the reflection of the EM waves from the left and right sides of the box. Both side reflections which are the unwanted signals of the GPR data are the strongest features in the image and as such they heavily obscure the hyperbolas corresponding to plastic pipes to be detected and characterised. The proposed techniques – 2DPCA and ICA are aiming to reduce those features in the image without significantly affecting the content of the remaining image, thus making the hyperbolas related to the pipes in the box more prominent in the final, processed image.

Figure 4.20: GPR images with removed mean trace (a) empty box and (b) box with three plastic pipes.
In Figure 4.20, mean trace was calculated and removed from both images before applying to any other processing to reduce the amount of clutter corresponds to the air-wooden panel interface present in both images. Several images were obtained by scanning along the lines before and after the interest line (i.e., the interest line is line 7, thus the scan was made from line 5 – 9) of the measurement grid to be used as input images for both algorithms. Spectrum of eigenvalues obtained after 2DPCA shown in Figure 4.21 (a) indicates the most of the image energy is concentrated in the first 5 – 10 eigenimages.

Figure 4.21: (a) Eigenvalues spectrum of PCs obtained using 2DPCA procedure and (b) reconstructed side reflection using 2DPCA eigenimages 1 – 6.

The side reflections were removed by adding the first 6 eigenimages to be subtracted from the original image (Figure 4.20 (b)). Cumulative image containing the first 6 eigenimages obtained using 2DPCA decomposition is
depicted in Figure 4.21 (b) whilst Figure 4.22 represents the first 6 independent images, results of ICA un-mixing procedure. The first independent image has been identified to contain most of the unwanted features by visual inspection. This image is therefore subtracted from the original image as in 2DPCA procedure to achieve the same aim – unwanted features reduction (i.e. side reflections) present in the original image and improvement of features corresponding to pipes placed in the box.

![Independent images obtained using ICA in reconstructing unwanted features (side reflections).](image)

Figure 4.22: Independent images obtained using ICA in reconstructing unwanted features (side reflections).

Results of the side reflections reduction using both approaches – 2DPCA and ICA are depicted in Figure 4.23. Both techniques can be seen as effective in reducing significant amount of unwanted features from the original image. Hyperbolas caused by the reflections from the largest and
middle size pipes can now be seen more clearly in both images of 2DPCA and ICA. The weakest hyperbola feature indicates the presence of the smallest pipe in box can also be observed.

![2D-PCA](image)

![2D-ICA](image)

Figure 4.23: Unwanted signals (side reflections) reduction results via (a) 2DPCA and (b) ICA algorithms.
The performance of both algorithms for images measured in pipes detection experiment is summarised in Table 4.8. The processed GPR images using both algorithms were compared to the original, raw image in terms of PSNR. The field labelled “difference” in Table 4.8 indicates the PSNR between the raw image and image obtained using direct subtraction (i.e. subtraction of the pipes image to the empty box image). The idea of using the “difference” approach is that the “empty” image which contains only interference features can be subtracted from the measured image to obtain significantly improved GPR image. Although this “technique” can not be applied in practical situations as the interference-only image is never available, the “difference” approach was used in this work to compare and evaluate the other two techniques.

![Difference Image](image.png)

Figure 4.24: The "difference" image is obtained by subtracting the empty box image from the image with three pipes.
Table 4.8: Comparison of proposed algorithms in reducing unwanted signals using PSNR measure of pipes detection experiment.

<table>
<thead>
<tr>
<th>Data</th>
<th>Pipe size</th>
<th>PSNR 2DPCA</th>
<th>PSNR ICA</th>
<th>PSNR Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Big</td>
<td>72.7788</td>
<td>69.4238</td>
<td>74.8429</td>
</tr>
<tr>
<td>2</td>
<td>Medium</td>
<td>66.4575</td>
<td>68.1856</td>
<td>68.3998</td>
</tr>
<tr>
<td>3</td>
<td>Small</td>
<td>67.3531</td>
<td>68.5401</td>
<td>68.3570</td>
</tr>
<tr>
<td>4</td>
<td>All three</td>
<td>64.8607</td>
<td>68.9848</td>
<td>66.0883</td>
</tr>
</tbody>
</table>

Figure 4.25: Evaluation results of unwanted signals reduction techniques for pipes detection of measured data.
4.5 Conclusion

Four groups of multivariate techniques to improve the quality of raw GPR images in reducing the amount of clutter and unwanted signals have been described and evaluated in this work using simulated and measured GPR images. The proposed techniques to be used at the pre-processing stage is aim to improve GPR image quality and enable easier detection and characterisation features (i.e. pipes and leakage from joints, holes and cracks) in the final stage of this work. GPR measurements obtained in those conditions suffer from the presence of noise and clutter as well as the unwanted signals. The proposed signal processing techniques represent an effective way of solving these problems. Clutter reduction algorithms (PCA and SVDPCA) and unwanted signals reduction algorithms (2DPCA and ICA) have been used, and the obtained results demonstrate the effectiveness of each approach.

The proposed clutter reduction algorithms show improvements in reducing the amount of clutter and noise presented in the simulated GPR data. Although these amounts are controlled during the simulations, the obtained results demonstrate that these techniques can enhance the GPR data more than the basic mean removal method for further processing. Despite the benefits of being able to generate “ideal” data which is clutter-and-noise-free image that can be used as a reference in the evaluation of GPR data performance, the simulated data has its drawback; it produces the
same GPR scans when it is run with the same condition settings. This differs from the measured data because GPR scans in one condition can produce different images each time a scan is made. For this reason, it is difficult to use experimentally simulated data to evaluate algorithms developed to reduce unwanted signals e.g. 2DPCA and ICA.

The measured data have been processed using the clutter reduction algorithms as well as the unwanted signals reduction techniques. The index numbers of PSNR and M-SSIM of the proposed clutter reduction algorithms indicate the potentials of these techniques to reduce the amount of clutter and noise presented in the GPR data. The obtained results suggest that the automated use of both algorithms to pre-process GPR images in order to detect and characterise features of interest in those images is possible. The “difference” method used in the unwanted signals reduction performs just as well, if not better than the proposed techniques – 2DPCA and ICA.
Chapter 5

Evaluation of Developed Pre-processing Techniques and Characterisation of Targets using Signal Modelling

Overview

The proposed clutter reduction techniques described in Chapter 4 require further analysis to evaluate its performance. This chapter introduces signal modelling techniques – Extended Prony (EP), Linear Prediction Singular Value Decomposition (LPSVD) and Matrix Pencil (MP), and subspace method – PCA as an alternative way of evaluating the performance of developed pre-processing techniques as well as for the target characterisation and classification tasks. Singular Entropy method is described and proposed to be used with MP approach to estimate the model order.
5.1 Introduction

The pre-processing techniques investigated in Chapter 4 indicate possibilities in reducing the clutter and unwanted signals present in the data in order to aid target detection. Although those techniques seem to have a potential to improve target detection, the additional problem of the identification of subsurface targets is a more difficult and as yet unsolved task.

On the other hand, there is also a need to further evaluate clutter reduction methods developed and implemented in Chapter 4. In this part, methods to extract resonant frequencies are used and combined with subspace methods described in the previous chapter to both develop an alternative way of evaluating the performance of pre-processing techniques and investigate the use of similar techniques for the target characterisation and classification tasks. In summary, the objective of the methods used in this chapter is to extract information from the signal that can help to characterise the physical/natural properties of the subsurface rather than just help the user to “see something in a radargram”.

In this chapter three candidate methods to estimate resonant poles from individual GPR traces are first presented. Those methods include Extended Prony (EP) [124], [125], Linear Prediction combined with Singular Value Decomposition (LPSVD) [29], [126], and Matrix Pencil (MP) [127], [128] algorithms. Those methods are first tested using synthetic signal and the
influence of noise on their performance examined. MP method shows the most robust performance and was used in the rest of the chapter to analyse both the simulated and measured GPR data. Important problem of correct model order estimation for simulated and measured signal has also been addressed in this chapter and the method using Differential Singular Entropy proposed as a suitable technique to most accurately estimate the model order, i.e. the number of resonant poles present in the signal.

5.2 Techniques

5.2.1 Prony Analysis

Prony Analysis [31], [129], [130], [131] is a method of fitting a linear combination of exponential terms to observed, damped oscillatory signal \( y(n) \). If such signal is represented by \( P \) complex exponentials, its estimate can be given with:

\[
\hat{y}(n) = \sum_{k=1}^{P} A_k e^{(\alpha_k+j\omega_k)(n-1)} e^{j\phi_k} \quad n = 1, 2, \ldots, N
\] (5.1)

Each of \( P \) terms, also known as the modes of the original signal \( y(n) \), is defined with four elements: the magnitude \( A_k \), the damping \( \alpha_k \), the angular frequency \( \omega_k \) and the initial phase \( \phi_k \).

It can be useful to recast the above equation in a slightly different form:

\[
\hat{y}(n) = \sum_{k=1}^{P} h_k z_k^{n-1}
\] (5.2)
Evaluation of Developed Pre-processing Techniques and Characterisation of Targets using Signal Modelling

where

\[ h_k = A_k e^{i\phi_k} \]  \hspace{1cm} (5.3)

is a complex amplitude representing a time-independent parameter and

\[ z_k = e^{(\alpha_k + i\omega_k)t} \]  \hspace{1cm} (5.4)

is a complex exponent that represents time-dependent parameter. Those complex exponents are usually referred to as signal poles.

The problem of estimating two sets of parameters \( h_k \) and \( z_k \) is based on the minimisation of the squared error over \( N \) samples of the observed signal \( y(n) \):

\[
\delta = \sum_{n=1}^{N} |e(n)|^2 = \sum_{n=1}^{N} |y(n) - \hat{y}(n)|^2 = \sum_{n=1}^{N} |y(n) - \sum_{k=1}^{p} h_k z_k^{n-1}|^2
\]  \hspace{1cm} (5.5)

This is a difficult non-linear problem and can be solved using Prony method which utilises linear equations solution.

Since the observed signal \( y(n) \) is fitted to a given exponential model, i.e.

\[ y(n) = \sum_{k=1}^{p} h_k z_k^{n-1} \]  \hspace{1cm} (5.6)

the above equation can be written for \( 1 \leq n \leq P \) in the matrix form:

\[
\begin{bmatrix}
    z_1^0 & z_2^0 & \cdots & z_p^0 \\
    z_1^1 & z_2^1 & \cdots & z_p^1 \\
    \vdots & \vdots & \ddots & \vdots \\
    z_1^{P-1} & z_2^{P-1} & \cdots & z_p^{P-1} \\
\end{bmatrix}
\begin{bmatrix}
    h_1 \\
    h_2 \\
    \vdots \\
    h_p \\
\end{bmatrix}
= 
\begin{bmatrix}
    y(1) \\
    y(2) \\
    \vdots \\
    y(P) \\
\end{bmatrix}
\]  \hspace{1cm} (5.7)
From the above equation, the complex amplitude $h_k$ can be obtained providing $z_k$’s are known. Prony method provides a way to determine $z_k$’s without having to resort to non-linear minimisation. This method proposed to define polynomial $A(z)$ which has $z_k$’s as its roots:

$$A(z) = \prod_{k=1}^{\rho} (z - z_k) = \sum_{m=0}^{\rho} a(m) z^{\rho-m}$$

(5.8)

Shifting the index in (5.6) from $n$ to $n - m$ and multiplying by the parameter $a(m)$ yields:

$$a(m) y(n-m) = a(m) \sum_{k=1}^{\rho} h_k z_k^{n-m-1}$$

(5.9)

Equation (5.9) can be further modified by summing from $m = 0$ to $m = P$ to:

$$\sum_{m=0}^{P} a(m) y(n-m) = \sum_{k=1}^{\rho} h_k z_k^{n-P} \sum_{m=0}^{P} a(m) h_k z_k^{\rho-m-1} = 0$$

(5.10)

This is a forward linear prediction equation which can be expressed in matrix form as:

$$\begin{bmatrix} y(P) & y(P-1) & \ldots & y(1) \\ y(P+1) & y(P) & \ldots & y(2) \\ \vdots & \vdots & \ddots & \vdots \\ y(2P-1) & y(2P-2) & \ldots & y(P) \end{bmatrix} \begin{bmatrix} a(1) \\ a(2) \\ \vdots \\ a(P) \end{bmatrix} = \begin{bmatrix} y(P+1) \\ y(P+2) \\ \vdots \\ y(2P) \end{bmatrix}$$

(5.11)

Thus, the $a(m)$ coefficient can be determined using Prony method which decouples the problem of determining $h_k$ and $z_k$ parameters. The entire procedure is usually split into three steps:
1. solve (5.11) to determine coefficient $a(m)$ where $m = 1, 2, \ldots, P$;

2. determine the complex exponents, roots $z_k$ of polynomial $A(z)$ defined in (5.8). At this point, damping and frequency of each mode can be determined from (5.4) as:

$$\alpha_k = \text{Re}\left(\log(z_k)\frac{1}{T}\right)$$

(5.12)

$$\omega_k = \text{Im}\left(\log(z_k)\frac{1}{T}\right)$$

(5.13)

3. determine complex amplitude $h_k$ from (5.7) and use (5.3) to find amplitude $A_k$ and initial phase $\varphi_k$ as:

$$A_k = |h_k|$$

(5.14)

$$\varphi_k = \text{Im}\left(\ln \frac{h_k}{|h_k|}\right)$$

(5.15)

5.2.2 Extended Prony Method

It is clear that the method proposed by Prony reduces the non-linear fitting problem to two numerical tasks of solving linear systems of equations (step 1) and finding the roots of polynomial (step 2). The length of data record should and usually does exceed the number of data points needed to fit a model, i.e. $N > 2P$. In practical situations $N$ should be at least three times larger than the model order $P$. In this case, matrix form of linear prediction equation (5.11) is modified to:
Evaluation of Developed Pre-processing Techniques and Characterisation of Targets using Signal Modelling

\[
\begin{bmatrix}
y(P) & y(P-1) & \ldots & y(1) \\
y(P+1) & y(P) & \ldots & y(2) \\
\vdots & \vdots & \ddots & \vdots \\
y(N-1) & y(N-2) & \ldots & y(N-P)
\end{bmatrix}
\begin{bmatrix}
a(1) \\
a(2) \\
\vdots \\
a(P)
\end{bmatrix}
= -
\begin{bmatrix}
y(P+1) \\
y(P+2) \\
\vdots \\
y(N)
\end{bmatrix}
\] (5.16)

This equation can be written in the condensed, matrix form:

\[ \mathbf{Y} \mathbf{a} = -\mathbf{y} \] (5.17)

and solved in the least-squares (LS) sense:

\[ \mathbf{a} = -\left(\mathbf{Y}^{\prime} \mathbf{Y}\right)^{-1} \mathbf{Y}^{\prime} \mathbf{y} \] (5.18)

This method is known as “least-squares Prony method” or “extended Prony method”.

5.2.3 Linear Prediction Singular Value Decomposition (Kumaresan-Tufts) Method

One of the problems Prony method faces are large variance and bias when analysing noisy signals [29], [129], [132]. Influence of noise can be reduced by setting the number of exponential components to be estimated to \( L \), where \( L > P \). Using SVD, the matrix \( \mathbf{Y} \) from equation (5.17) can be written as:

\[ \mathbf{Y} = \mathbf{U} \mathbf{\Sigma} \mathbf{V}^{\prime} \] (5.19)

where \( \mathbf{\Sigma} \) is a \( (N-L) \times L \) matrix with the singular values on the diagonal arranged in decreasing order. Noise can be reduced by considering the reduced rank approximation.
\[
\hat{Y} = U \hat{\Sigma} V^H
\]  
(5.20)

with

\[
\hat{\Sigma} = \begin{bmatrix}
\Sigma_p & 0 \\
0 & 0
\end{bmatrix}_{(N-L) \times L}
\]  
(5.21)

where \( \Sigma_p \) is the top-left \( P \times P \) minor of \( \Sigma \). An estimate for the coefficients \( a(m) \) is then:

\[
\hat{a} = -Y^\dagger y
\]  
(5.22)

where \( Y^\dagger \) is the pseudo-inverse of \( \hat{Y} \). The use of truncated SVD improves the SNR, providing a better estimate of the vector \( \hat{a} \) and consequently of the exponential parameters. This approach is known as Kumaresan-Tufts (KT) algorithm [35] but is sometimes also referred to as Linear Prediction Singular Value Decomposition (LPSVD).

### 5.2.4 Matrix Pencil Method

Another method to estimate exponential parameters from noisy signals is Matrix Pencil (MP) method [33], [127]. It is generally more robust than LPSVD by having a lower variance on the estimated parameters with a slightly larger bias. The MP starts with the application of SVD to data matrix \( Y \), where \( Y \) is in the form of:
Evaluation of Developed Pre-processing Techniques and Characterisation of Targets using Signal Modelling

\[ Y = \begin{bmatrix} y(0) & y(1) & \ldots & y(L) \\ y(1) & y(2) & \ldots & y(L+1) \\ \vdots & \vdots & \ddots & \vdots \\ y(N-L-1) & y(N-L) & \ldots & y(N-1) \end{bmatrix}_{(N-L)\times(N-L)} \] (5.23)

Here \( L \) denotes the pencil parameter that plays the role of the prediction order parameter in the LPSVD/KT method where \( P \leq L \leq N - P \), and is usually \( \frac{N}{3} \leq L \leq \frac{2N}{3} \). Noise can be reduced by considering the reduced rank approximation

\[ Y = U^\prime \Sigma^\prime V'^T \] (5.24)

by taking the first \( P \) singular values. Two matrices \( Y_0 \) and \( Y_1 \) can be obtained from equation (5.24) by eliminating the last and first row of \( V' \) and are defined as:

\[ Y_0 = \begin{bmatrix} y(0) & y(1) & \ldots & y(L-1) \\ y(1) & y(2) & \ldots & y(L) \\ \vdots & \vdots & \ddots & \vdots \\ y(N-L-1) & y(N-L) & \ldots & y(N-2) \end{bmatrix}; \] (5.25)

\[ Y_1 = \begin{bmatrix} y(1) & y(2) & \ldots & y(L) \\ y(2) & y(3) & \ldots & y(L+1) \\ \vdots & \vdots & \ddots & \vdots \\ y(N-L) & y(N-L+1) & \ldots & y(N-1) \end{bmatrix}; \] (5.26)

Matrices \( Y_0 \) and \( Y_1 \) can now be decomposed as:

\[ Y_0 = Z_0 H Z_r \] (5.26)

\[ Y_1 = Z_1 H Z_r \] (5.27)
where

\[
Z_i = \begin{bmatrix}
1 & 1 & \cdots & 1 \\
z_1 & z_2 & \cdots & z_p \\
\vdots & \vdots & \ddots & \vdots \\
\phantom{1} & \phantom{1} & \cdots & \phantom{1}
\end{bmatrix},
\]

\[
Z_r = \begin{bmatrix}
1 & z_1 & \cdots & z_1^{L-1} \\
1 & z_2 & \cdots & z_2^{L-1} \\
\vdots & \vdots & \ddots & \vdots \\
1 & z_p & \cdots & z_p^{L-1}
\end{bmatrix}
\]

(5.28)

\[
H = \text{diag}(h_1, h_2, \ldots, h_p)
\]

(5.29)

\[
Z = \text{diag}(z_1, z_2, \ldots, z_p)
\]

(5.30)

Matrix Pencil \( Y_i - zY_0 \) can be rewritten using decomposition of two matrices as:

\[
Y_i - zY_0 = Z_i H (Z - zI_p) Z_r
\]

(5.31)

When \( z \neq z_i \), the matrix \( Z - zI_p \) is of rank \( P \). However, for \( z = z_i \) it is of rank \( P - 1 \). Therefore, the poles of the signal reduce the rank of the MP for \( P \leq L \leq N - P \). This is equivalent to saying that the poles \( z_i \) are the generalised eigenvalues of \( (Y_i, Y_0) \), in the sense that \( (Y_i - zY_0)v = 0 \), with \( v \) is the eigenvector of \( Y_i - zY_0 \). To find the poles \( z_i \), one can use the fact that \( Y_0^t Y_i \) has \( P \) eigenvalues equal to the poles \( z_i \) and \( L - P \) null eigenvalues. Here \( Y_0^t \) denotes a pseudo-inverse (Moore-Penrose) of \( Y_0 \).
In practice, it is impossible to have access to noise-less signal and work directly with the noisy data requires the use of SVD to select singular values of the signal. The basic steps of the MP method can be summarised as follows:

1. build the matrices $Y_0$ and $Y_1$ as in (5.25);
2. perform SVD of $Y_1$, i.e. $Y_1 = U\Sigma V^T$;
3. estimate the signal subspace of $Y_1$ by considering the $P$ largest singular values of $\Sigma$: $\bar{Y}_1 = U_p \Sigma_p V_p^T$, where $U_p$ and $V_p$ are built from the first $P$ columns of $U$ and $V$, and $\Sigma_p$ is the top-left $P \times P$ minor of $\Sigma$;
4. the matrix $Z_L = Y_1^T Y_0 = V_p \Sigma_p^{-1} U_p^T Y_0$ has $P$ eigenvalues which provide estimates of the inverse poles $1/z_i$; the other $L - P$ eigenvalues are zero. Since $Z_L$ has only $P$ non-zero eigenvalues, it is convenient to restrict attention to a $P \times P$ matrix $Z_p = \Sigma_p^{-1} U_p^T Y_0 V_p$.

The MP technique exploits the MP structure of the underlying signal, rather than the prediction equations satisfied by it. Nevertheless, there are strong similarities between the MP and LPSVD techniques.
5.3 Results

Methods discussed in the second part of this chapter have been applied to a number of signals in the rest of this section. First, a simple signal containing low number of poles has been simulated and the performance of all three algorithms in the analysis of this signal containing various levels of noise assessed. Most accurate algorithm is selected and used further in the analysis of real and simulated GPR data. Results are presented in the following sections.

5.3.1 Synthesised Signals

Simple synthesised signal has been used to assess the performance of all three signal modelling algorithms contained three resonant modes. All three modes are stable, i.e. three related damping coefficients are positive. Complete set of parameters used to generate those signals is given in the first row of Table 5.1. Sampling frequency used in this simulation is 100 Hz and various levels of noise are added to the simulated signal in order to investigate sensitivity to noise of each of proposed algorithms.

For the purpose of this study, it is assumed that the correct model order is known. Since the complex signal poles always come in complex-conjugate pairs, $L = 6$ is used as the model order for each run of analysis algorithms in order to extract three poles present in the signal. Obtained results for all three cases are summarised in Table 5.1.
First section of this table represents results achieved with each of the algorithms applied on noise-free signal (SNR = ∞). Each of applied algorithms was able to recover the original parameters of the simulated, noise-free signal with perfect accuracy. For relatively low amount of noise present in the signal (SNR = 80 dB), both LPSVD and MP algorithms were able to estimate correct signal parameters whilst the EP method introduced small errors in the estimated set of parameters. For a very noisy signal (SNR = 10 dB), EP was unable to recover any of the signal parameters correctly. Both LPSVD and MP algorithms performed well although MP method results in slightly more accurate values. For this reason, only MP algorithm is used to analyse measured GPR signals in the next section of the thesis.

It is worth noting that the EP algorithm was unable to recover any of the original parameters correctly during the application of the analysis of noisy signals with the exact model order = 6, even for a very low levels of noise present in the signal (SNR = 80 dB). As the model order is increased to 20, the same number of different frequencies present in the signal is obtained upon the application of the EP algorithm. However, most of the amplitudes related to the estimated frequencies were insignificant and therefore have been discarded from the obtained results set. This procedure left only three relevant amplitudes and damping factors in the final set.
Table 5.1: Performance of three signal modelling algorithms using simulated signals with different SNRs.

<table>
<thead>
<tr>
<th>Frequency (Hz)</th>
<th>Amplitude</th>
<th>Damping</th>
<th>Phase (rad)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Original Signal</td>
<td>2.0 3.0 4.0</td>
<td>0.6 0.5 0.3</td>
<td>0.1 0.2 0.05</td>
</tr>
<tr>
<td><strong>SNR = ∞ (noise-free signal)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EP</td>
<td>2.0 3.0 4.0</td>
<td>0.6 0.5 0.3</td>
<td>0.1 0.2 0.05</td>
</tr>
<tr>
<td>LPSVD</td>
<td>2.0 3.0 4.0</td>
<td>0.6 0.5 0.3</td>
<td>0.1 0.2 0.05</td>
</tr>
<tr>
<td>MP</td>
<td>2.0 3.0 4.0</td>
<td>0.6 0.5 0.3</td>
<td>0.1 0.2 0.05</td>
</tr>
<tr>
<td><strong>SNR = 80 dB</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EP</td>
<td>2.0001 3.0005 4.0001</td>
<td>0.6027 0.5084 0.3008</td>
<td>0.1018 0.2086 0.0512</td>
</tr>
<tr>
<td>LPSVD</td>
<td>2.0 3.0 4.0</td>
<td>0.6 0.5 0.3</td>
<td>0.1 0.2 0.05</td>
</tr>
<tr>
<td>MP</td>
<td>2.0 3.0 4.0</td>
<td>0.6 0.5 0.3</td>
<td>0.1 0.2 0.05</td>
</tr>
<tr>
<td><strong>SNR = 10 dB</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EP</td>
<td>2.8 26.18 43.05</td>
<td>2.36 0.51 0.87</td>
<td>7.15 30.014 42.594</td>
</tr>
<tr>
<td>LPSVD</td>
<td>1.998 2.996 4.077</td>
<td>0.6383 0.5632 0.3518</td>
<td>0.1288 0.253 0.0943</td>
</tr>
<tr>
<td>MP</td>
<td>1.999 2.997 4.006</td>
<td>0.5996 0.5222 0.3326</td>
<td>0.1026 0.2 0.05</td>
</tr>
</tbody>
</table>
The values given in Table 5.1 of EP algorithm are well related to the initial parameters of the simulated signal. The signal has been successfully analysed using this “thresholding” approach and noise-free version of the signal reconstructed. Same procedure is applied in the analysis of signal with high noise level (SNR = 10 dB) but, unlike in the SNR = 80 dB case, the obtained results were not correct as indicated in Table 5.1.

Figure 5.1: Performance of studied algorithms for SNR = 10 dB case, (a) original signal, (b), (c) and (d) reconstructed signals and residual using (b) EP, (c) LPSVD and (d) MP.
Simulated signal containing high level of noise (SNR = 10 dB) as well as the signals reconstructed using estimated parameter sets and corresponding residual errors are shown in Figure 5.1. Powers of residual signals shown on plots (b), (c) and (d) are 0.1639 (EP), 0.0013 (LPSVD) and 0.000537 (MP) indicate poor performance of EP algorithm and its sensitivity to noise and somewhat superior performance of MP algorithm.

5.3.2 GPR Signals and Model Order Estimation

5.3.2.1 Model Order Estimation Problem

In this section, a number of simulated and measured GPR traces are analysed using MP method. Unlike the analysis process described in the previous section where the model order of simulated signal is known a priori, for most real signals, measured on the grid, model order is unknown although experienced user may have some idea of the expected model order. If the selected order is too low, some modes will not be estimated, whilst too high order specified for the analysis introduces extra components that are not present in the original signal. Therefore, model order selection is a trade-off between increased resolution and decreased variance in the estimated spectrum.

There are several methods and criteria to estimate model order described in literature although guidelines on the usage of some of those criteria in practical situations are rather limited. Some of better known criteria
investigated in recent decades include Akaike Information Criterion [133] and Minimum Descriptive Length [134] but both of those criteria are known to perform poorly for closely spaced modes [133]. In this work, two simple approaches for model order estimation recommended in the literature have been used and investigated. Power of the residue signal – error between the produced model and the measured signal can be estimated for various model orders and the order resulting in minimum residue power adopted. Singular Entropy (SE) approach is another method used to estimate the optimal model order in this study. As an additional check, Fast Fourier Transform (FFT) based analysis is performed and the obtained power spectrum of the measured signal is compared to the signal estimated using adopted model order.

Singular Entropy (SE) approach [135] starts by performing SVD of Hankel matrix in the form of:

$$Y = \begin{bmatrix} y(1) & y(2) & \cdots & y(L+1) \\ y(2) & y(3) & \cdots & y(L+2) \\ \vdots & \vdots & \ddots & \vdots \\ y(N-L) & y(N-L+1) & \cdots & y(N) \end{bmatrix}$$

(5.32)

Here, $L$ represents the MP parameter usually chosen to be $L = N/4 \div N/3$ and $y(n)$ is the measured signal where $n = 1, 2, \ldots, N$. SVD of matrix $Y$ results in three new matrices such that

$$Y = U\Sigma V^T$$

(5.33)
Matrix $\Sigma$ is a diagonal matrix containing singular values $\sigma_i (i = 1, 2, \ldots, m)$ where $m = \min \{ N - L, L + 1 \}$, whilst $U$ and $V$ contain left and right eigenvectors of $Y$ respectively.

A Singular Entropy Increment (SEI) is now defined as:

$$
\Delta E_i = - \left( \frac{\sigma_i}{\sum_{j=1}^{m} \sigma_j} \right) \log \left( \frac{\sigma_i}{\sum_{j=1}^{m} \sigma_j} \right) \quad i = 1, 2, \ldots, m
$$

(5.34)

and $k$-th order Differential of Singular Entropy Increment (DSEI) is:

$$
\Delta E_k' = \frac{\Delta E_k - \Delta E_{k-1}}{k - (k-1)} = \Delta E_k - \Delta E_{k-1} \quad k = 2, 3, \ldots, m
$$

(5.35)

When the effective signal saturates, the SEI converges to a bounded value rapidly and there is an inflection point which corresponds to the signal modal order. After that point, the SEI is small, which can be considered as the effect of noise. Due to this property of SEI, no matter how much the noise is, the increment of Singular Entropy will converge to a bounded value with a noticeable jump when the information of effective signal tends to saturation.

With the help of DSEI, it is easy to extract the order of the corresponded inflection point, namely, the model order $L$. To illustrate this approach, Figure 5.2 shows the SEI and DSEI for simulated signal analysed in the previous section. After the turning point at $L = 6$ which the SEI is approximately zero, accurately indicates the model order of the simulated
noisy signal. Since the extracted complex poles always appear in complex conjugate pairs, turning at $L = 6$ indicates the presence of three modes and therefore three frequency, amplitude, damping and phase sets extracted and shown in Table 5.1.

![Figure 5.2: (a) Singular Entropy Increment and (b) Differential of Singular Entropy Increment characteristics for simulated signal (SNR = 10 dB) depicting correct signal model order at $L = 6$.](image)

5.3.2.2 Analysis of Measured GPR Signals

In this section, analysis of GPR images obtained in the experiment involving measurement pipes, described in Chapter 2 is presented. To evaluate the effects of processing techniques described in Chapter 4, same trace is extracted from the image processed in 4 different ways and analysed using MP algorithm. Initial results and the entire process in this section illustrated using the image of big pipe only. Similar procedure is repeated for other measured images and results summarised at the end of the chapter.
Figure 5.3: Big pipe image after pre-processing via (a) mean removal, (b) 2DPCA, (c) ICA and (d) difference techniques. The white line in (a) represents the position of analysed trace – column 83.

Figure 5.3 shows the pre-processed measured image of big pipe with the application of unwanted signals reduction techniques. Image with only mean removal applied to it (i.e., closest to “raw” image) as well as images after 2DPCA and ICA techniques are depicted in Figure 5.3 (a), (b) and (c) respectively whilst Figure 5.3 (d) contains the “difference” image obtained by subtracting the big pipe image to the empty box image. Careful aligning
of two images is attempted using correlation method to align these two images before subtraction and improve the obtained “difference” result. In this way, removal of most of the side reflections is attempted. Although this approach could not be attempted on the images obtained in practical GPR surveys, as the “empty” image is usually not available, this result is used in this study as an “ideal” image against which the results of other processing techniques are compared as explained in Chapter 4. Position of analysed trace (column 83 in image matrix) is also indicated in Figure 5.3 (a).

Figure 5.4: Extracted GPR traces of pre-processed data via (a) mean removal, (b) 2DPCA, (c) ICA and (d) difference techniques.
Evaluation of Developed Pre-processing Techniques and Characterisation of Targets using Signal Modelling

Figure 5.5: (a) SEI and (b) DSEI for each of extracted traces.
Figure 5.4 illustrates the traces extracted from each image shown in Figure 5.3. The model order estimation using Singular Entropy approach is depicted in Figure 5.5 and columns (a) and (b) show the SEI and DSEI characteristics for each trace respectively.

Figure 5.6 and Figure 5.7 illustrate the results of MP analysis on the first trace, extracted from the image shown in Figure 5.3 (a) for the models of different orders. The original and estimated signal is depicted in Figure 5.6 whilst columns (a) and (b) of Figure 5.7 show the frequency and amplitude of the extracted poles and damping of each of those poles. FFT has also been performed on this signal to verify the results of MP analysis and result plotted over the frequency-amplitude characteristic shown in column (a) of Figure 5.7 indicates level of agreement between these two methods.

Model orders $L = 12, 16, 20, 22$ and $26$ correspond to the last five prominent peaks on DSEI characteristics from column (b) of Figure 5.5 have been used to perform this analysis. Column (b) of Figure 5.7 indicates the amount of damping corresponds to each extracted complex pole, i.e. frequency present in the analysed trace. It is interesting to note the small amount of negative damping in the lower end of the frequency spectrum. This is obviously incorrect but is due to either noise present in the signal or the length of the analysed signal which indicates the penetration depth of the radar. This parameter, as explained in Chapter 2 can be set before the GPR
survey and it would be good to increase this setting in future experiments in order to remove this artefact.

Figure 5.6: The signal reconstructions using selected model order of a single trace.
Evaluation of Developed Pre-processing Techniques and Characterisation of Targets using Signal Modelling

$L = 12$

$L = 16$

$L = 20$

$L = 22$

$L = 26$

(a) (b)

Figure 5.7: Results of Matrix Pencil on a single trace.
To further evaluate the quality of the obtained results, residual signal, i.e. error between the measured and estimated signal is also considered. Figure 5.8 shows the power of the modelling error and the maximal amplitude of the residual signal for different model orders. It can be observed that increasing the model orders to values significantly higher than 20 contribute very little to further reduction of residual power and maximal amplitude. This finding is therefore in a good agreement with the Singular Entropy approach. Further in the analysis of the results and application of MP algorithm, combination of those two methods was used to determine the correct model order. Residual is used to estimate the approximate model order, with the ability of Singular Entropy model to depict the exact model order to be more accurate by detecting the peaks in the DSEI characteristics.

![Figure 5.8](image_url)

Figure 5.8: (a) Power and (b) maximal amplitude of the residual signal for different model orders $L$. 

178
Results of MP analysis for images measured with big pipe are illustrated in the first row of Figure 5.9. Second and third rows of the same figure represent similar results obtained from the MP analysis of images captured with medium and small pipes. Columns (a) and (b) represent the frequency-amplitude and frequency-damping characteristics. It can be observed that both methods developed and tested in Chapter 4 (2DPCA and ICA techniques) aimed at reducing the side reflections related clutter can achieve appreciable reduction of this clutter compared to raw image with only a simple mean removal applied to it.

Similarity between two processed traces and the corresponding trace from the difference image is higher for the big pipe case (first row of Figure 5.9) compared to both medium and small pipes traces. Effects of the application of those techniques on frequency-damping characteristic are not so pronounced but are visible. This is expected as the damping is more related to the property of the materials of both target and scatterers which did not change in this experiment.
Evaluation of Developed Pre-processing Techniques and Characterisation of Targets using Signal Modelling

Big pipe

Medium pipe

Small pipe

(a) (b)

Figure 5.9: Comparative results of MP analysis obtained for a single pipe in the box (a) frequency-amplitude and (b) frequency-damping characteristics.
Similar analysis is applied to images shown in Figure 5.10 obtained with all three pipes in the box. Frequency-amplitude and frequency-damping characteristics for traces correspond to the middle of big, medium and small pipes are depicted in Figure 5.11.

Figure 5.10: Image of all three pipes in the box after pre-processing via (a) mean removal, (b) 2DPCA, (c) ICA and (d) difference techniques. The white lines in (a) represent the positions of analysed traces – column 25, 83 and 138.
Figure 5.11: Comparative results of MP analysis obtained for all three pipes in the box (a) frequency-amplitude and (b) frequency-damping characteristics (for traces 25, 83 and 138 from Figure 5.10).
Here, due to the excessively large amount of side reflections present in the traces close to the edges of the box (i.e. big and small pipes), the effects of pre-processing are not as pronounced as in the single image pipe case illustrated before. The trace further away from the edges of the box and less affected by the side reflections clutter (i.e. middle pipe) however exhibits similar behaviour to the one noticed earlier in the case of pre-processing on the single pipe trace, i.e. the frequency-amplitude characteristic of the applied pre-processing techniques is similar to the one obtained on the analysed trace extracted from the difference image.

5.3.3 Towards Target Characterisation – Analysis of GPRMax Simulated Signals

In the previous section, MP method has been applied to measured raw and pre-processed images in order to assess the performance of pre-processing algorithms proposed and developed in this work. Although results and clutter reduction achieved with those algorithms have been illustrated, another important task in development of signal processing algorithms for GPR technology is feature extraction for subsequent target characterisation and recognition. This section indicates the potential of using signal modelling and super-resolution techniques discussed in this chapter for those tasks.
For the initial assessment of suitability of those techniques, simulated rather than measured data has been employed and analysed in this section. GPR images correspond to the experiment in leakages detection from joints, holes and cracks discussed in Chapter 2 were used and analysed using MP algorithm, following the same procedure outlined in the previous section. Availability of clutter-free, simulated data makes it easy to assess the potential of the proposed analysis algorithm before it is applied to measured data, which, as it has been indicated in the previous chapter contained excessively of noise and clutter, thus making it impossible to properly assess the suitability of those algorithms for target characterisation tasks.

DSEI characteristics have been generated for each extracted trace before the application of MP algorithm to estimate the model order to be used in the MP analysis. It is interesting to note that the indicated model orders are very similar in value (8 – 12 in most of the cases) which denotes how the presence of clutter increases the complexity of the model to be used for data analysis. Due to the absence of any noise or clutter, analysis of multiple traces corresponds to the same target resulted in almost identical results. For clarity purposes, only the most representative results have been selected and presented in this section.

Images with indicated extracted traces are shown in Figure 5.12. The extracted traces correspond to the three main features to be detected in the concrete slab survey via GPR – joint amongst the concrete slab, hole at the
joint position and cracks in the slab. In addition to that, an extra trace corresponds to a scan along the healthy part of the concrete slab has also been used in the analysis. From the results shown in Figure 5.13, it can be seen that the model orders used in the MP analysis are relatively low, $L = 10$ is used for modelling cracks whilst $L = 12$ model order is used to model healthy concrete and joints including the hole at the joint position.

Figure 5.12: Simulated GPRMax images with indicated traces (white lines) used in further analysis (a) joint, (b) crack, (c) crack and healthy concrete and (d) hole at the joint position.
Figure 5.13: Comparative results of MP analysis obtained from simulated images containing various targets in concrete slab (a) frequency-amplitude and (b) frequency-damping characteristics.

The reason for the investigation of signal modelling techniques and the subsequent adoption of MP algorithm is now apparent. Although the frequency-amplitude characteristic reveals little about the features analysed in this study, frequency-damping characteristic for each feature is more significant. Each group of analysed traces – joints (including hole at the joint position), cracks, and healthy concrete can be easily distinguished from the corresponding frequency-damping characteristic. Difference between healthy joint and joint containing hole is also encouraging indicate the potential of both GPR technology on its own as well as the proposed signal processing methods in the construction industry.
5.3.4 Target Characterisation using Principal Component Analysis on Simulated Data

Analysis of GPR traces using MP technique with the aim of extracting resonant poles and dominant frequencies, amplitudes and damping coefficients has been performed in the previous part of this chapter. Results indicate possibility in characterising different targets using the extracted MP parameters. Another approach of using PCA in evaluating the performance and results of pre-processing algorithms and characterising GPR traces of interest is tested in this section. The procedure of PCA in analysing GPR traces for target characterisation is described as follow:

1. single or a group of successive GPR traces are first extracted from GPR images and stored into data matrix $X$;
2. PCs of the stored traces are obtained via SVD of data matrix $X$ and most of the PCs discarded whilst the most significant ones retained;
3. the extracted traces are then projected back on those PCs. In this way, the dimensionality of the extracted traces is reduced and each trace is represented with a small number of coordinates where number of coordinates correspond to the number of retained PCs. Distance between those points is calculated to establish possible group membership of each trace. Points correspond to each trace are plotted in 2D and 3D cases to visualise their positions in the new (feature) space.
The extracted traces correspond to joints, holes and cracks in leakages detection experiment are depicted in Table 5.2. Mean removal method was applied to the corresponding images before the extraction of those traces. From the obtained results of PCA based on dimensionality reduction illustrated in Figure 5.14, the cracks can be clustered easily. However, this is differ to joints which the positions are slightly more scattered in the diagram, that is mostly due to the significant differences in size of the simulated joints (i.e. Joint 2 = 25 mm and Joint 3 = 12 mm). As expected, holes performed similar to joints since they were placed at the same position as the joints and scattered away from their cluster. Table 5.3 described the calculated distance between points correspond to each extracted trace for the case of two retained PCs.

Table 5.2: Extracted traces of GPR simulated data on different target types.

<table>
<thead>
<tr>
<th>Features</th>
<th>GPR trace</th>
<th>Extracted signals</th>
</tr>
</thead>
<tbody>
<tr>
<td>Joint 1</td>
<td>58 – 62</td>
<td></td>
</tr>
<tr>
<td>Joint/ Hole</td>
<td>Range</td>
<td></td>
</tr>
<tr>
<td>-------------</td>
<td>---------</td>
<td></td>
</tr>
<tr>
<td>Joint 2</td>
<td>117 – 121</td>
<td></td>
</tr>
<tr>
<td>Joint 3</td>
<td>177 – 181</td>
<td></td>
</tr>
<tr>
<td>Joint 4</td>
<td>58 – 62</td>
<td></td>
</tr>
<tr>
<td>Hole 1</td>
<td>117 – 121</td>
<td></td>
</tr>
<tr>
<td>Hole 2</td>
<td>177 – 181</td>
<td></td>
</tr>
</tbody>
</table>
Evaluation of Developed Pre-processing Techniques and Characterisation of Targets using Signal Modelling

<table>
<thead>
<tr>
<th>Crack</th>
<th>Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>Crack 1</td>
<td>62 – 66</td>
</tr>
<tr>
<td>Crack 2</td>
<td>75 – 79</td>
</tr>
<tr>
<td>Crack 3</td>
<td>74 – 78</td>
</tr>
<tr>
<td>Crack 4</td>
<td>104 – 108</td>
</tr>
</tbody>
</table>
Figure 5.14: PCs represent different target types on simulated data in (a) two-dimensional and (b) three-dimensional.
Table 5.3: Calculated distance of each target depicted in Figure 5.14 (a).

<table>
<thead>
<tr>
<th></th>
<th>Joint 1</th>
<th>Joint 2</th>
<th>Joint 3</th>
<th>Joint 4</th>
<th>Hole 1</th>
<th>Hole 2</th>
<th>Crack 1</th>
<th>Crack 2</th>
<th>Crack 3</th>
<th>Crack 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Joint 1</td>
<td>0</td>
<td>113.6</td>
<td>30.2</td>
<td>4.2</td>
<td>134.4</td>
<td>51.1</td>
<td>95.7</td>
<td>106.5</td>
<td>106.3</td>
<td>103.7</td>
</tr>
<tr>
<td>Joint 2</td>
<td>113.6</td>
<td>0</td>
<td>84.2</td>
<td>117.0</td>
<td>23.8</td>
<td>81.1</td>
<td>128.5</td>
<td>127.5</td>
<td>127.1</td>
<td>126.6</td>
</tr>
<tr>
<td>Joint 3</td>
<td>30.2</td>
<td>84.2</td>
<td>0</td>
<td>34.1</td>
<td>104.5</td>
<td>28.2</td>
<td>98.9</td>
<td>106.8</td>
<td>106.5</td>
<td>104.3</td>
</tr>
<tr>
<td>Joint 4</td>
<td>4.2</td>
<td>117.0</td>
<td>34.1</td>
<td>0</td>
<td>138.1</td>
<td>55.2</td>
<td>94.4</td>
<td>105.5</td>
<td>105.3</td>
<td>102.7</td>
</tr>
<tr>
<td>Hole 1</td>
<td>134.4</td>
<td>23.8</td>
<td>104.5</td>
<td>138.1</td>
<td>0</td>
<td>96.6</td>
<td>151.7</td>
<td>150.1</td>
<td>149.7</td>
<td>149.3</td>
</tr>
<tr>
<td>Hole 2</td>
<td>51.1</td>
<td>81.1</td>
<td>28.2</td>
<td>55.2</td>
<td>96.6</td>
<td>0</td>
<td>126.0</td>
<td>133.2</td>
<td>132.9</td>
<td>130.8</td>
</tr>
<tr>
<td>Crack 1</td>
<td>95.7</td>
<td>128.5</td>
<td>98.9</td>
<td>94.4</td>
<td>151.7</td>
<td>126.0</td>
<td>0</td>
<td>13.3</td>
<td>13.4</td>
<td>10.7</td>
</tr>
<tr>
<td>Crack 2</td>
<td>106.5</td>
<td>127.7</td>
<td>106.8</td>
<td>105.5</td>
<td>150.1</td>
<td>133.2</td>
<td>13.3</td>
<td>0</td>
<td>0.4</td>
<td>2.8</td>
</tr>
<tr>
<td>Crack 3</td>
<td>106.3</td>
<td>127.1</td>
<td>106.5</td>
<td>105.3</td>
<td>149.7</td>
<td>132.9</td>
<td>13.4</td>
<td>0.4</td>
<td>0</td>
<td>2.8</td>
</tr>
<tr>
<td>Crack 4</td>
<td>103.7</td>
<td>126.6</td>
<td>104.3</td>
<td>102.7</td>
<td>149.3</td>
<td>130.8</td>
<td>10.7</td>
<td>2.8</td>
<td>2.8</td>
<td>0</td>
</tr>
</tbody>
</table>
5.4 Conclusion

Extended Prony (EP), Linear Prediction Singular Value Decomposition (LPSVD) and Matrix Pencil (MP) have been investigated and used:

- as an alternative to PSNR measure used in Chapter 4;
- for target characterisation and classification.

Each of these methods has been applied to the simulated data to evaluate its performance in the presence of additive noise in the GPR data. As stated previously, GPR data suffer from large amount of noise and clutter, thus using simulated data enabled different levels of noise to be added to the data to assess the performance of the algorithms.

MP method shows its robustness in analysing noisy data compared to EP and LPSVD, thus was used for the rest of the analysis. The problem of selecting the correct model order in analysing signal has been tackled using Singular Entropy (SE) approach. With the help of Differential Singular Entropy Increment (DSEI), it is easier to extract the exact model order of the observed signal. This was verified by obtaining a minimum of residue power (i.e. error between the produced model and the measured signal).

The estimated signal produced by the MP method using the SE approach has been compared to the power spectrum of the measured signal obtained from the Fast Fourier Transform (FFT) based analysis. The results of frequency-amplitude produced by FFT shows an agreement with the MP
method although the peak of the amplitude using FFT is not exactly the same as that from MP. However, since the MP method has super resolution, it allows the frequencies to be placed exactly at the locations of the dominant frequencies.

For target characterisation, the extracted information using the MP method was able to help in characterising the natural properties of the subsurface. The frequency-damping characteristic obtained from the MP analysis was able to characterise features according to their group since the damping is more related to the property of the materials of both target and scatterers. In addition to MP method, subspace method developed in Chapter 4 – Principal Component Analysis (PCA) was used to characterise features based on the extracted Principal Components (PCs) obtained via Singular Value Decomposition (SVD).
Chapter 6

Findings and Conclusions

Overview

This chapter summarises the research carried out and the major findings of this thesis. The author’s contributions and recommendations for further work, based on this research are outlined.

6.1 Extended Summary of the Main Ideas and Work Done on the Project

The general aim of the work described in this thesis was to investigate the possibility of application of various signal processing algorithms to process, improve and understand data measured by GPR in various situations that can arise in construction industry. The GPR data measured in those conditions usually suffer from excessively large amount of noise and clutter which can completely obscure the target of the GPR survey and make the images and data extremely difficult to understand and interpret. Highly skilled and trained human operator is usually needed in those situations to process measured data and make sense of the captured images. The process
of analysing those images can therefore become both time demanding and very costly. For this reason, the research into signal and artificial intelligence techniques aimed at helping the process of improving and understanding the measured data has been gathered pace in the recent decades.

One of the main tasks tackled in this project dealt with the problem of removing clutter and unwanted signals present in GPR data. However, judging how well the applied technique has performed, i.e. what is the overall improvement in the measured data is another important issue for GPR aimed algorithms and techniques. In the most image processing and image improvement problems (image compression and decompression for examples), an “ideal”, i.e. clean or undistorted dataset or image is available to assess the performance of the applied algorithms. Standard image quality metrics (Mean Square Error (MSE), Signal-to-Noise Ratio (SNR) or Peak-Signal-to-Noise Ratio (PSNR)) can then be applied to quantify and subsequently decide on the suitability of the algorithm for the image improvement tasks. Unfortunately, this approach is not usually applicable for the GPR data and techniques used to improve this data. Ideal or undistorted image is not readily available and the use of simulated data (obtained using GPR simulation software) can produce only a completely clean and undistorted set of data. Any distortion additionally introduced to
simulated data would be too artificial, not close to real experimental situations.

For the purpose of evaluation of decluttering techniques, PSNR metrics has been initially used in this project. The use of PSNR in evaluating the quality of decluttered GPR images is however not a straightforward task. In traditional image processing applications, high PSNR values of well denoised images indicate high similarity between the original (ideal) and denoised image. At the other end of the scale, a small PSNR value implies poor denoising results caused by large numerical differences between the processed and original (ideal) image. In case of GPR, this approach has to be inverted – high differences between the original (raw and noisy) and decluttered image can indicate good performance of decluttering algorithm and well decluttered resulting image. This result however needs to be treated with caution as the removed components could also contain useful, target-related parts of the GPR image.

In addition to PSNR measure, this project also adopted and evaluated an alternative approach of using signal modelling algorithms to evaluate pre-processing techniques. In this way, twofold aim can be achieved – assessment of the quality of pre-processing algorithms as well as extracting useful signal features for further automatic or semi-automatic characterisation and qualification of the target under survey.
Decluttering algorithms and techniques investigated in this work relied on two related techniques for multivariate data analysis methods – Principal Component Analysis (PCA) and Independent Component Analysis (ICA). Whilst the details of both approaches have been explained in this thesis, it is worth emphasising the general reasons behind the investigation of those two techniques. In contrast to most of traditional techniques used for GPR data processing, multivariate methods such are PCA and ICA, approach the analysis of multiple variables (or measurements) by treating those multiple variables as a single entity. Thus, both PCA and ICA seek to produce results that take into account the relationship between the multiple variables as well as within the variables, and use tools that operate on all of the data.

A major concern of both approaches is to find transformations of the multivariate data that make the dataset smaller or easier to understand, i.e. they seek to find the answer to the following questions: Is it possible that the relevant information contained in a multidimensional variable could be expressed using fewer dimensions (i.e. variable) and might the reduced set of variables be more meaningful than the original dataset? If the latter were true, we call the more meaningful variables hidden, or latent, in the original data and can assume that the new variables better represent the underlying processes that produced the original dataset in the first place. This approach has been used in this work in order to try and discard the set of data that might only be related to noise or other non-target objects present in the
surveyed area. Whilst the results demonstrated in this work certainly encourage this philosophy, the problem of identifying the useful variables or set of signals and distinguishing it from the other artefacts present in the data are not an easy one to solve. Each approach has its own advantages and disadvantages.

PCA which is a more established and traditional approach, uses the spectrum of eigenvalues of the analysed data matrix to decide and distinguish between the clutter, useful set of signals and noise present in the signal. Still, the useful set of signals might further contain features related to non-target objects present in the area.

ICA on the other hand improves on the PCA analysis which merely decorrelates the data by seeking to transform the original dataset into number of truly independent variables. The motivation for this transformation is primarily to uncover more meaningful variables, not just to reduce the dimensions of the dataset. Although both techniques seemed to produce interesting and useful sets of components representing the analysed data, it is difficult to decide, in a completely unsupervised manner, which of the independent components produced by ICA relate to target and which are originating from other, non-target objects in the area.

The described problem was approached in this work by suggesting a different way of combining PCA and ICA. This resulted in ranking of extracted components and an opportunity to automatically discard some of
those components from the set. Results given in this work suggest that this approach can, in some situations, be used as a good replacement for supervised clutter reduction approach in use today. For removing strong but non-uniform clutter present in some GPR images, new experiment has been designed and the used of 2DPCA algorithm resulting in significant clutter reduction. Application of 2DPCA algorithm has not been attempted or reported in GPR area elsewhere in literature.

In the continuation of this project, signal modelling has been applied to pre-processed data in order to extract useful features that can be used to characterise targets in the survey – possible damages in the structure (cracks, holes and joints) or underground pipework. The main aim of this modelling was to detect and extract resonant poles and use the parameters of those poles – resonant frequency, amplitude and damping of each detected resonant pole to characterise target. Three signal processing algorithms have been proposed for this investigation – Extended Prony (EP), Linear Prediction Singular value Decomposition (LPSVD) and Matrix Pencil (MP) methods. Initial analysis, using synthesised, resonant type signals confirmed high sensitivity of EP method to noise present in the data. This was in agreement with the majority of literature available on this subject. LPSVD and MP methods were much more robust and resilient to presence of noise in the signal but the final set of tests on measured and simulated data was
Findings and Conclusions

carried out using mainly MP analysis approach as it was found that MP outperforms LPSVD in the analysis of GPR type data.

One of the main problems of using MP method is determining the optimal model order to be used for the analysis. A number of methods suggested elsewhere in the literature to determine the model order have been tested in this project but none of those was found to perform very well for the signals obtained in GPR measurements. Instead, a novel method using Singular Entropy (SE) has been adopted in this project. Although the extraction of resonant poles has been tried in GPR signal processing, MP approach combined with SE technique has only been used and recently reported in the analysis of oscillations in power grids and the adoption and testing of this approach in GPR field represents another novel aspects of this project.

To assess the performance of suggested algorithms, detailed investigation into GPR technology and theory behind the operation of GPR system has been undertaken in the first part of this project. Extensive and varied measurements have been performed in three dedicated laboratory facilities, designed to closely resemble conditions in the construction industry. This has been followed by simulations with dedicated GPRMax software, used by most of the researchers in the field. These algorithms have been tested on both simulated and experimental data. The GPR simulation has been implemented via GPRMax2D to replicate the actual experiments
performed in two different controlled laboratories only i.e., pipes detection and leakages detection from joints, holes and cracks. The third experiment to detect cracked concrete slab with metal fibre mixtures is however unable to be simulated due to data complexity. In order to evaluate the performance of the selected processing methods, the measured data have been processed with dedicated proprietary software, ReflexW designed specifically for the purpose of processing GPR and other geophysical type of data. Traditional processing steps have been applied during ReflexW pre-processing such as one-dimensional filtering, topographic and time-zero corrections, band-pass filtering, background removal and average subtraction.

In addition to PSNR criteria used in evaluating the performance of different pre-processing techniques in reducing and eliminating the clutter and unwanted signals in the GPR data, structural similarity (SSIM) between the obtained GPR image and the one produced by highly trained GPR image processing expert using ReflexW has been proposed in this work. Mean Structural Similarity (M-SSIM) Index has been used to give a better sense of validity and indicate potential advantages of the proposed pre-processing techniques. This type measure is recently proposed approach for image quality assessment which indicates the similarity between two images with values between 0 and 1 where 1 indicates the test and reference images are identical. The SSIM is based on human visual system and is designed to
improve on traditional metrics like PSNR and MSE, which have been proved to be inconsistent with human eye perception.

The measurements performed in detecting cracked concrete slab with metal fibre mixtures, although useful in providing good insight into some problems related to GPR survey, resulted in a complex set of data which did not allow the final conclusions on the performance of the algorithms developed in this work to be made. The pipes detection experiment provided more suitable data for the application of decluttering algorithms, analysis and evaluation of achieved results at this stage of work. This lead to some recommendations and suggestions for further work and continuation of this project outlined at the end of this chapter.

6.2 Author’s Contributions

The author’s achievements have been the successful implementation of clutter and unwanted signal reduction techniques as well as identifying and characterising features present in GPR data. The specific contributions can be summarised as follows:

- design and implementation of simulated data using dedicated GPRMax software to understand the problem of GPR data clutter and target characterisation;
- investigation and interpretation of pre-processing GPR data using commercially available ReflexW software to be used in the
evaluation of subspace and data decomposition techniques applied to GPR data;

- analysis of subspace and data decomposition techniques – Principal Component Analysis (PCA) and Independent Component Analysis (ICA) on GPR data;

- implementation and modification of traditional PCA and ICA methods to improve their performance in image processing and GPR clutter removal tasks – this resulted in two novel algorithms which are the combination of Singular Value Decomposition of PCA and ICA (SVDPICA) and 2DPCA, not previously suggested in GPR field;

- traditional image quality measures (PSNR and M-SSIM in particular) implemented and tested in order to assess the quality of developed pre-processing algorithms;

- analysis and implementation of signal modelling techniques – Extended Prony (EP), Linear Prediction Singular Value Decomposition (LPSVD) and Matrix Pencil (MP) was twofold – evaluation of pre-processing algorithms as well as characterisation of surveyed targets;

- implementation of Singular Entropy based model order estimation method to be used with MP analysis on GPR signals;

- analysis of subspace method – PCA to characterise features.
6.3 Recommendations for Further Work

6.3.1 GPR Simulations

Although GPR simulations via GPRMax software proved very useful in this work, there has been significant disagreement between the measured and simulated data despite taking great care to replicate all measurement conditions in those simulations. Those differences and disagreements should be resolved, i.e. further insight into behaviour of simulation algorithm and electromagnetic theory used in those simulations gained. This is certainly a research project in itself and as such was beyond the boundaries of this project.

6.3.2 GPR Measurements

Great care needs to be taken when designing and conducting the GPR surveys intended to be further used for the evaluation and development of new DSP algorithms aimed at helping GPR operators. Some of the “real industrial” conditions should be avoided at first in order to collect a dataset that can easily be interpreted and reproduced later in the lab and simulation. If at all possible, “ideal” or “close-to-ideal” dataset should be obtained in order to assess and evaluate the performance of designed or implemented signal processing algorithms.
6.3.3 Pre-processing Algorithms

Multivariate techniques such as PCA and ICA are showing significant promise in processing GPR data. Modifications proposed in this work did enhance the performance of those techniques in GPR applications although some problems remain. Ordering extracted components and deciding what to discard and what to retain in the final set seems to be the main issue to be tackled and solved if those algorithms are to be applied for unsupervised pre-processing of GPR data.

Problem of evaluating the performance of those algorithms is an open issue, mainly due to a lack of “ideal” data to compare against and evaluate achieved results. In this work, both traditional and novel approaches have been implemented and tested.

6.3.4 Analysis Algorithms

Although the use of signal modelling has shown some promise in the analysis of measured GPR data, it should be further investigated; perhaps utilising simulated GPR data to a greater extent in the continuations of this work. Simple experimental setups should be initially analysed in more details and the influence of each element established before tackling more complex structures and problems. The use of Singular Entropy approach for model order estimation was appropriate although other methods can still be
investigated, especially if those can result in completely automatic or unsupervised optimal order calculation and detection.
Bibliography


Bibliography


Appendix

Appendix A: Chapter 2

A.1 Input Files of GPRMax Simulation on Joints, Holes and Cracks Detection

A.1.1 GPR 4 GHz Frequency for Joints and Holes

```plaintext
-- GPR joints, holes and cracks simulation based on Amir's -------
-- experiment at UoP ---------------------------------------------
-- NB, 29/11/2012 -----------------------------------------------
--- GprMax model for holes + joints in a concrete block for GV5 ---
--- The size of the holes are 10 mm (width) by 40 mm (depth) -----
--- Frequency: 4 GHz ---------------------------------------------
--- hole1 = at the position of joint2 -----------------------------
--- hole2 = at the position of joint3 -----------------------------
--- joint1 = 7 mm width -------------------------------------------
--- joint2 = 25 mm width ------------------------------------------
--- joint3 = 12 mm width ------------------------------------------
#media_file: d:\gpr\media\media_nurul.dat
--------- #domain defines size of domain ---------
#domain: 1.27 0.6
--------- #dx_dy defines spatial steps = radius of smallest pipe/10
--------- or smallest wavelength to be resolved/10
#dx_dy: 0.0015 0.0015
--------- #time_window defines length of each trace
--------- it is related to the depth to be reached
#time_window: 5.0e-9
--------- Geometry of the problem defined here
----- Container box
#box: 0.0 0.0 1.27 0.15 concrete
#box: 0.0 0.24 1.27 0.26 wood -- bottom panel
#box: 0.0 0.26 0.02 0.50 wood -- left panel
#box: 1.25 0.26 1.27 0.50 wood -- right panel
#box: 0.02 0.26 1.25 0.36 wet_sand
----- Concrete slabs
#box: 0.113 0.36 0.363 0.485 concrete -- slab 1
#box: 0.370 0.36 0.620 0.485 concrete -- slab 2
#box: 0.645 0.36 0.895 0.485 concrete -- slab 3
#box: 0.907 0.36 1.157 0.485 concrete -- slab 4
----- Joints
#box: 0.363 0.36 0.370 0.485 fbond_polyester -- joint 1
```

227
A.1.2 GPR 4 GHz Frequency for Crack

-- GPR joints, holes and cracks simulation based on Amir’s -----
-- experiment at UoP ---------------------------------------------
-- NB, 29/11/2012 ------------------------------------------------
------------------------------------------------------------------
------ GprMax model of crack in a concrete slab 1 for GV5 --------
------ Crack dimension is 0.4 cm (width) and 12 cm (depth) ------
------ Crack slab1 = at the position 27 cm from the edge --------
------ of the concrete -------------------------------------------
------ Frequency: 4 GHz ------------------------------------------
------------------------------------------------------------------
#media_file: d:\gpr\media\media_nurul.dat
------------------------------------------------------------------
#title: 2 holes + 3 joints in a concrete block (GV5) model
#messages: y
#geometry_file: amirmodelv1_5.geo

APPENDIX

#box: 0.620 0.36 0.645 0.485 fbond_polyester -- joint 2
#box: 0.895 0.36 0.907 0.485 fbond_polyester -- joint 3
----- Holes
#box: 0.6275 0.445 0.6375 0.485 free_space -- hole 1
#box: 0.8960 0.445 0.9070 0.485 free_space -- hole 2
-------------------------------------------------------------------
Appendix

A.2 Input Files of GPRMax Simulation on Pipes Detection

A.2.1 GPR 4 GHz Frequency for Three Pipes in a Box

-- GPR pipes simulation based on July 2012 experiment at UoP ------
-- NB, 06/12/2012 ---------------------------------------------

--------- GprMax model for 3 pipes in the box experiment --------
--------- with the ground added below the box --------------------
--------- For accurate model air gap is added ---------------------
--------- between the bottom of the box and soil surface --------
--------- Frequency: 4 GHz --------------------------------------
--------- Dimension of the pipes: -------------------------------
---------- Big: 10.8 cm ----------------------------------------
---------- Medium: 4.3 cm --------------------------------------
---------- Small: 2.8 cm ---------------------------------------

#media_file: d:\gpr\media\media_nurul.dat

--------- #domain defines size of domain ----------
#domain: 1.27 0.6
--------- #dx_dy defines spatial steps = radius of smallest pipe/10
--------- or smallest wavelength to be resolved/10
#dx_dy: 0.0015 0.0015
--------- #time_window defines length of each trace
--------- it is related to the depth to be reached
#time_window: 10.0e-9

--------- Geometry of the problem defined here
#box: 0.0 0.217 1.27 0.242 wood     -- bottom panel
#box: 0.0 0.482 1.27 0.50 plywood -- top panel
#box: 0.0 0.242 0.025 0.482 wood -- left panel
#box: 1.245 0.242 1.27 0.482 wood -- right panel
#box: 0.0 0.0 1.27 0.122 dry_claysoil
--------- three pipes = plastic cylinders with air cylinders inside
#cylinder: 0.37 0.296 0.054 plastic          -- big
#cylinder: 0.37 0.296 0.052 free_space
#cylinder: 0.675 0.2635 0.0215 plastic       -- medium
#cylinder: 0.675 0.2635 0.0195 free_space
A.3 Media File of the GPRMax Data

The media file has the role of a simple database of frequently used materials which is very simple and the (#) character is reserved, similarly to an input file. Each definition of a medium parameters should occupy a single line in the media file. A general entry in a media file is:

```
# f1  f2  f3  f4  f5  f6  strl
```

Each parameter represents:

- **f1** the DC (static) relative permittivity of the medium $\varepsilon_{rs}$
- **f2** the relative permittivity at the theoretically infinite frequency $\varepsilon_{\infty}$
- **f3** the relaxation time of the medium $\tau$ (seconds)
- **f4** the DC (static) conductivity of the medium $\sigma$ (Siemens/metre)
- **f5** the relative permeability of the medium $\mu_r$
- **f6** the magnetic conductivity of the medium $\sigma^*$
- **strl** a string characterising the medium
A.4 Matlab Code for GPRMax

The following code is used to run the input files of GPRMax.

```matlab
%% Script to start GprMax simulation and display the results
% BV, NB, June 2009
clear all;
close all;
cl;
filename = 'amirmodelv1_1';
geo = '.geo';
sca = '.sca';
geofile = [filename geo];
scafile = [filename sca];

%% Invoking GPR simulation
% *** Select output/results directory
directory_out = 'C:\GPR\Amir_Simulations\Output';
% *** Select exe and input directories
dos 'C:\GPR\GprMax2D
C:\GPR\Amir_Simulations\Input\amirmodelv1_1.in'

%% Results display
% Geometry data
[mesh, header, media] = gprmax2g(geofile);
model title = header.title;
dx = header.dx;  % cell size in x direction - horizontal for GprMax
dy = header.dy;  % cell size in y direction - vertical for GprMax
nx = header.nx;
ny = header.ny;
x = nx*dx;
y = ny*dy;

% Scan data
[Header, Fields] = gprmax(scafile);
model title = Header.title;
Nsteps = Header.NSteps;  % no. of scans in the file
```
iterations = Header.iterations; % no. of iterations in the given
% time window

dt = Header.dt;                  % time for each iteration

timew = Header.removed;         % time window (s)
ez = reshape(Fields.ez, iterations, Nsteps); % reshape scan data
% into matrix

%% Display GPR image - geometry and scan

% Geometry
figure(1);
imagesc(0:dx:x, 0:dy:y, mesh);
set(gca,'YDir','normal')
colorbar(gray);
xlabel('[m]');
ylabel('[m]');
title(modeltitle);

% Scan
figure(2);
imagesc(0:Nsteps, (0:dt:timew)*10^9, ez); % *10^9 to get time in ns
colorbar(gray);
xlabel('[Trace no]');
ylabel('[ns]');
title(modeltitle);

%% Display GPR scan image after mean removal
ezmr = mrem(ez);
figure(3);
imagesc(ezmr);
colorbar(gray);
xlabel('[m]');
ylabel('[ns]');

%% Save data
save(filename);
Appendix B: Chapter 4

B.1 Clutter Reduction Algorithms Code

% Analysis of measured or simulated GPR scans on reducing clutter.
% Basically, GPR image is applied using these methods:
% 1. Decluttering via clutter subtraction (Difference)
% 2. Mean subtraction
% 3. PCA
% 4. SVDPICA
%
% The decluttering via clutter subtraction is only possible in the
% availability of clutter-only image.
% PSNR and M-SSIM are used to evaluate the performance of each
% method.
%
% For the SVDPICA method, the following steps are used:
% 1. Perform PCA on the mean removed image.
% 2. Get rid of clutter (first PC) and noise (last n-k PCs)
%    or just retain all significant PCs (dimensionality reduction
%    only).
% 3. Find and retain the most significant ICs in the reconstructed
%    image
%    by performing ICA on retained PCs set.
%
% The GPR data matrix needs to be transposed and scaled in the
% application of PCA. This is due to the Matlab 'svd' function that
% assumes the measurements in row instead of columns as with our
% GPR data.
%
% BV, NB December 2011.

clear all;
close all;
clc;

%% Load GPR data
current_directory = pwd;

select_data = input('Option Data: '); % Option Data
if (select_data == 1)
    % Simulated data
    directory_data = 'C:\GPR\Amir_Simulations\Output';
    cd(directory_data);
    data = input('Data: ','s');
    load([data '.mat']);
    X = ez;
elseif (select_data == 2)
    % Measured data
    directory_data = 'C:\REFLEX\Portsmouth\Portsmouth Data July';
    cd(directory_data);
    data = input('Data: ','s');
    X = load([data '.ASC']);
end

[m, n] = size(X);
% X - data matrix
% m - number of time samples, i.e. dimensions/ variables
Appendix

```matlab
% n - number of space samples, i.e. trials/ measurements
cd(current_directory);
figure(1);
imagesc(X);
colormap(gray);
title('Original Image');

%% 1. Decluttering using clutter subtraction (Difference)
% Only possible when clutter-only image is available
% (i.e. in simulation)
if (select_data == 1)
    % Simulated data
directory_data = 'C:\GPR\Amir_Simulations\Output';
cd(directory_data);
data = input('Clutter-Only Data: ','s');
load([data '.mat']);
Xn = ez;
elseif (select_data == 2)
    % Measured data
error('No Clutter-Only Data Available!')
end

cd(current_directory);
Xideal = X - Xn;            % remove clutter from the data
figure(2);
imagesc(Xn);
colormap(gray);
title('Clutter-Only Data');
figure(3);
imagesc(Xideal);
colormap(gray);
title('Decluttered Data - Full Clutter Subtraction');

%% 2. Simple mean subtraction
Xmr = mrem(X);
figure(4);
imagesc(Xmr);
colormap(gray);
title('Decluttered Data - Mean Removal');

%% 3. PCA method 1
% Uses SVD of data matrix
dataX = X'/sqrt(n-1);       % transpose and scale the data matrix
[U, S, V] = svd(dataX,0);   % PCA via "economy" SVD
D = diag(S).^2;             % calculate the eigenvalues
figure(5);
plot(20*log10(D),'k-*');    % scree plot
grid on;
xlabel('Eigenvalue no');
ylabel('Eigenvalue (dB)');
title('Eigenspectra');
k = input('Number of PCs: ');
Y = U(:,1:k);               % transform the original data
% Reduce data dimensionality
```
% i.e. consider only the first r rows of Y
Yr = Y(1:k,:);

% Recover the original data
% just for checking
Xrec = U*Y;
Xrec = Xrec';               % this should now be the same with X

% Recover the original data with losses due to dimensionality
% reduction
% just for checking
Xkrec = U(:,1:k)*Yr;
Xkrec = Xkrec';

% Set the largest singular values to zero for clutter reduction
% since the clutter usually dominates the largest eigenvalues.
Xsvd = U(:,2:k)*Y(2:k,:);
Xsvd = Xsvd';

% PCA via SVD Results - Plots
% Plot the first k eigenimages
E = zeros(k,m,n);
for i = 1:k
    E(i,:,:) = (U(:,i) * S(i,i) * (V(:,i))');
end
figure(6);
for i = 1:k
    subplot(ceil(sqrt(k)),ceil(sqrt(k)),i)
    colormap(gray);
    imagesc(reshape(E(i,:,:),m,n));
    title(["Eigenimage # ", num2str(i)]);
end

% Plot the reconstructed image
figure(7)
imagesc(Xsvd);
colormap(gray);
title('Decluttered Data - SVD')

%% 4. PCA method 2
% Uses eigenvalue decomposition of data covariance matrix

dataX2 = X';                 % transpose the data matrix
cov = 1/(n-1)*dataX2*dataX2';     % calculate the covariance matrix
[U, D] = eig(cov);           % find the eigenvectors and
D = diag(D);                 % extract diagonal eigenvalues
U = U(:,rindices);

% Sort the eigenvalues and eigenvectors in decreasing order
[junk, rindices] = sort(-1*D);
D = D(rindices);
U = U(:,rindices);

% Plot eigenvalue spectra
figure(8);
plot(20*log10(D),'k-*');     % scree plot
grid on;
xlabel('Eigenvalue no');
ylabel('Eigenvalue (dB)');
title('Eigenspectra');

k = input('Number of PCs: ');   % select number of PCs to retain
Y = U'*dataX2;                 % project the original data set

% Reduce data dimensionality
% i.e. consider only the first r rows of Y
Appendix

\begin{verbatim}
Yr = Y(1:k,:);
% Recover the original data
Xrec = U*Y;
Xrec = Xrec';               % this should now be the same with X

% Recover the original data with losses due to dimensionality reduction
Xkrec = U(:,1:k)*Yr;
Xkrec = Xkrec';

% Set the largest singular values to zero for clutter reduction since the clutter usually dominates the largest eigenvalues.
Xeig = U(:,2:k)*Y(2:k,:);
Xeig = Xeig';

% PCA via SVD Results - Plots
% Plot the first k eigenimages
figure(9);
Xreceig = zeros(size(dataX2));
for i = 1:k
    Y1 = zeros(size(Y));
    Y1(i,:) = Y(i,:);
    eigeigen = U * Y1;
    subplot(ceil(sqrt(k)),ceil(sqrt(k)),i);
    colormap(gray);
    imagesc(eigeigen');
    title(['Eigenimage # ', num2str(i)]);
    Xreceig = Xreceig + eigeigen;
end

% Plot the reconstructed image
figure(10);
colormap(gray);
imagesc(Xeig);
title('Decluttered Data - EIG');

% Note - Xeig is identical to Xsvd!

%% 5. SVDPICA method
k = input('Number of PCs: ');
[U, S, V] = svd(Xmr,0);     % do the SVD on mean removed data
Uk = U(:,1:k);              % consider first k PCs only
Vk = V(:,1:k);
Sk = S(1:k,1:k);
W = jadeR(Uk',k);           % do the ICA
Ukti = W * Uk';            % un-mix the mixed PCs
B = pinv(W);                % estimate the mixing matrix
Tktil = Vk * Sk * B;

% Decompose Tktil into two new matrices such that: Tktil = Vkti * Sktil
% The data has been transposed, thus, we have rows of measurements instead of columns
[m1 n1] = size(Tktil);
Vkti = [ ];
Sktil = zeros(n1,n1);
for i = 1:n1
    beta = norm(Tktil(:,i));
    Vkti = [Vkti Tktil(:,i)/beta];
    Sktil(i,i) = beta;
end

% Sort new ICS in descending order of associated singular values from Sktil
\end{verbatim}
Appendix

\[ \text{Stemp} = \text{diag}(\text{Sktil}); \] % extract diagonal of matrix as vector
\[ [\text{junk}, \text{rindices}] = \text{sort}(-1*\text{Stemp}); \] % sort in decreasing order
\[ \text{Uktil} = \text{Uktil}(\text{rindices},:); \] % reorder
\[ \text{Sktil} = \text{Sktil}(\text{rindices},\text{rindices}); \]
\[ \text{Vktil} = \text{Vktil}(:,\text{rindices}); \]
\[ \text{eigentil} = \text{diag}(\text{Sktil})^2; \] % get the eigenvalues

\% Plot the new eigenvalue spectra
\text{figure}(11);
\text{plot}(20*\text{log10}(\text{eigentil}),'k-*'); % scree plot
\text{grid on; xlabel('Eigenvalue no'); ylable('Eigenvalue (dB)'); title('New Eigenspectra');}

\% Obtain independent images and reconstruct the data
\[ \text{l} = \text{zeros}(k,m,n); \]
\for i = 1:k
\[ l(i,:,:,:) = (\text{Vktil}(:,i) * \text{Sktil}(i,i) * (\text{Uktil}(i,:)))'; \]
\end
\[ \text{k1} = \text{input}('Number of ICs [nfirst nlast]: '); \]
\[ \text{Xrec1} = \text{sum}([\text{l}(\text{k1}(1):\text{k1}(2),:,:)]); \]
\[ \text{Xrec1} = \text{reshape}(\text{Xrec1},m,n); \]

\% SVDPICA Results - Plots
\% Plot individual independent images
\text{figure}(12);
\for i = 1:(\text{k1}(2)-\text{k1}(1)+1)
\[ \text{subplot}(\text{ceil(sqrt(\text{k1}(2)-\text{k1}(1)+1))},\text{ceil(sqrt(\text{k1}(2)-\text{k1}(1)+1))},i) \]
\[ \text{colormap(gray); imagesc(\text{reshape}([\text{l}(i,:,:),m,n])); title(['Independent image # ', num2str(i)])); \]
\end

\% Plot reconstructed image
\text{figure}(13);
\text{imagesc}(\text{Xrec1});
\text{colormap(gray); title('Decluttered Data - PCA via ICA');}

\% Evaluation via PSNR
\% i.e. Compare to "ideal" or reference image
\% For simulated data "ideal" image can be used as a reference
\% For measured data original raw image is used as a reference
\if (\text{select\_data} == 1)
\[ \text{Xref} = \text{X\_ideal}; \]
\else
\if (\text{select\_data} == 2)
\[ \text{Xref} = \text{X}; \]
\end
\end

\[ \text{XR} = \text{imscale}(\text{Xref}); \]
\[ \text{XMR} = \text{imscale}(\text{Xmr}); \]
\[ \text{XSVD} = \text{imscale}(\text{Xsvd}); \]
\[ \text{XEIG} = \text{imscale}(\text{Xeig}); \]
\[ \text{XSVDPICA} = \text{imscale}(\text{Xrec1}); \]

\% Calculate PSNRs for all obtained images, using XR as reference
\% iimage
\[ \text{psnr\_MR} = \text{PSNR}(\text{XMR},\text{XR}); \]
\[ \text{psnr\_SVD} = \text{PSNR}(\text{XSVD},\text{XR}); \]
\[ \text{psnr\_EIG} = \text{PSNR}(\text{XEIG},\text{XR}); \]
\[ \text{psnr\_SVDPICA} = \text{PSNR}(\text{XSVDPICA},\text{XR}); \]

\% Evaluation via M-SSIM
% For simulated data = "Ideal" vs. processed images
% For measured data = Reflexw vs. processed images

if (select_data == 1)
    XI = XR;
elseif (select_data == 2)
    current_directory = pwd;
    directory_data = 'C:\REFLEX\Portsmouth\Portsmouth Data July';
    cd(directory_data);
    ideal = input('Ideal data for M-SSIM: ','s');
    XI = load([ideal '.ASC']);
    figure(14);
    imagesc(XI);
    colormap(gray);
    title('Ideal Data');
    cd(current_directory);
    XI = imscale(XI);
end

[mssim_MR, ssim_map_MR] = ssim(XI, XMR);
[mssim_SVD, ssim_map_SVD] = ssim(XI, XSVDSVD);
[mssim_EIG, ssim_map_EIG] = ssim(XI, XEIG);
[mssim_SVDPICA, ssim_map_SVDPICA] = ssim(XI, XSVDPICA);

% Display results
% PSNR results
format short
disp('PSNR values:');
disp([psnr_MR psnr_SVD psnr_EIG psnr_SVDPICA]);

% Similarity result
format short
disp('SSIM values:');
disp([mssim_MR mssim_SVD mssim_EIG mssim_SVDPICA]);

% The PSNR result:
% SIMULATED DATA:
% Ideal image can be obtained to be used as the "reference" image
% in the evaluation using PSNR. Thus, higher PSNR indicates better
% algorithm.
% MEASURED DATA:
% Original raw data is used as the "reference" image. Lower values
% of PSNR indicates better algorithm.

B.2 Unwanted Signals Reduction Algorithms Code

% Program for removing the unwanted signals/features in the image.
% Technique 1: 2D-PCA
% Technique 2 : ICA
% Improved images obtained from these techniques are evaluated
% using various image quality measures:
% Measure 1: PSNR
% Measure 2: M-SSIM
% This program able to do the shift/align among the images to have
% similar size of each image. This aligning part is most beneficial
% in obtaining an "ideal" image from the subtraction of the image
% contains clutter and target to the image with the clutter only.
% The subtraction method is known as "Difference" technique which
% will be used in the evaluation of the proposed techniques of the
% unwanted signals removal. The original image will be used in
% calculating both the PSNR and M-SSIM values.
% BV, NB, September 2012

close all;
clear all;
icl;

[4]all: ');
% Setup to analyse: 1 = Big pipe; 2 = Medium pipe;
% 3 = Small pipe; 4 = All 3 pipes

% Load GPR data

current_directory = pwd;
directory_data = '"D:\Users\My Documents\My Dropbox\Nuruls simulation\Aligning"';
c(directory_data);

% Empty box scans needed for 2D-PCA
load FUN1L51raw.ASC;
load FUN1L61raw.ASC;
load FUN1L71raw.ASC;
load FUN1L81raw.ASC;
load FUN1L91raw.ASC;
load FUN1L72raw.ASC;

if stup == 1
  load FUN2B71raw.ASC; % Big pipe
elseif stup == 2
  load FUN2M71raw.ASC; % Medium pipe
elseif stup == 3
  load FUN2S71raw.ASC; % Small pipe
elseif stup == 4
  load FUN2L71raw.ASC; % All 3 pipes
end
c(current_directory);

% Remove means and select images to process

X1 = mrem(FUN1L51raw); X2 = mrem(FUN1L61raw);
X3 = mrem(FUN1L71raw); X4 = mrem(FUN1L81raw);
X5 = mrem(FUN1L91raw); X6 = mrem(FUN1L72raw);

if stup == 1
  X = mrem(FUN2B71raw);
elseif stup == 2
  X = mrem(FUN2M71raw);
elseif stup == 3
  X = mrem(FUN2S71raw);
elseif stup == 4
  X = mrem(FUN2L71raw);
% Empty image
Xe = mrem(FUN1L1Raw);

%% Aligning along the x-axis
[Xd, Xed, displ] = align(X, Xe);
fprintf('Displacement along x-axis was: %.2f
', displ);

%% Aligning along y-axis
[med ned] = size(Xed);
[md nd] = size(Xd);
if ned~=md
    fprintf('Matrices have different number of columns!!!
');
    return
end

% Upsample each column of both matrices
uf = 10;
for j = 1:ned
    Xedi(:,j) = interp(Xed(:,j),uf);
    Xdi(:,j) = interp(Xd(:,j),uf);
end

% Transpose both matrices before aligning along the y-axes
Xedi = Xedi';
Xdi = Xdi';

[Xdid, Xedid, displ] = align(Xdi, Xedi);
fprintf('Displacement along y-axis was: %.2f
', displ/uf);

% Transpose back
Xedid = Xedid';
Xdid = Xdid';

%% Display results and calculate PSNR and M-SSIM of 'Difference'
%% method
figure(1)
imagesc(Xedid);
colorbar(gray);
title('Empty box');
xlabel('GPR traces');
ylabel('Time samples');
figure(2)
imagesc(Xdid);
colorbar(gray);
title('Box with pipe(s)');
xlabel('GPR traces');
ylabel('Time samples');
figure(3)
Xdiff = Xdid-Xedid;
imagesc(Xdiff);
colorbar(gray);
title('Difference');
xlabel('GPR traces');
ylabel('Time samples');

% PSNR calculation
Xdids = imscale(Xdid);          % scale to [0 1] for PSNR
Xdiffs = imscale(Xdiff);
psnr_diff = PSNR(Xdiffs,Xdids);
% Similarity measure
[mssim_diff, ssim_map_diff] = ssim(X_diffs, X_diffs);

% 2D-PCA
% The 2D data sets of the individual images will be added in
% finding the average total image to be applied to the SVD

% Check the size of all images and take the minimum size to be the
% reference for all images
[m1, n1] = size(X1);
m2, n2 = size(X2);
m3, n3 = size(X3);
m4, n4 = size(X4);
m5, n5 = size(X5);
m6, n6 = size(X6);

n = min([n1 n2 n3 n4 n5 n6]);
X1 = X1(:,1:n);
X2 = X2(:,1:n);
X3 = X3(:,1:n);
X4 = X4(:,1:n);
X5 = X5(:,1:n);
X6 = X6(:,1:n);

% Add all images and calculate the average image
X_total = (X1 + X2 + X3 + X4 + X5 + X6)/6;

% SVD of the total image
[V, S, U] = svd(X_total);

% Plot the eigenvalues to decide on the exact number of eigenimages
% to be reconsidered
eigen = diag(S).^2;
sev = sum(eigen);
figure(4);
plot(eigen,'k','LineWidth',2);
hold on;
plot(eigen,'ko','LineWidth',2);
title('2D-PCA Eigenvalue Spectrum');
xlabel('Eigenvalue Number');
ylabel('Eigenvalue');

% Construct eigenimage group best representing the unwanted feature
% Obtain first eno eigenimages
eno = input('Select number of eigenimages to be removed from the image: ');
St = zeros(size(S));
for i = 1:eno
    St(i,i) = S(i,i); % keep the first eno eigenvalues for reconstruction
end
Xeig = V * St * U';

% Plot the individual eigenimages
Xetot = zeros(m2,n);
figure(5);
for i = 1:eno
    St = zeros(size(S));
    St(i,i) = S(i,i);
    Xe = V * St * U';
    subplot(ceil(sqrt(eno)),ceil(sqrt(eno)),i)
    imagesc(Xe);
    colormap(gray);
    title(['2DPCA Eigenimage #', num2str(i)]);
    Xetot = Xetot + Xe;
end
end

% "Total" eigenimage - first 'eno eigenimages added together
figure(6);
imagesc(Xetot);           % plot the reconstructed total eigenimage
colormap(gray);
title('Total Eigenimage of the Unwanted Features');

%% Plot 2D-PCA results
%
% Align original and total eigenimage of 2D-PCA
[Xd, Xeigd, displ] = align(X, Xeig);

% Remove the reconstructed total eigenimage from the selected
% individual image
Xsvd2 = Xd - Xeigd;
figure(7);
imagesc(Xsvd2);           % plot the residual
colormap(gray);
title('2D-PCA');

%% Calculate PSNR for 2D-PCA
%
% The dimension of image obtained from 2D-PCA (Xsvd2) must be the
% same with the image before the application of this method (i.e.
% Xdids) for PSNR comparison. Thus:
% 1. Do the upsample to each column of Xsvd2
% 2. Sort the number of columns
[nsvd2 msvd2] = size(Xsvd2);
[ndids mids] = size(Xdids);
m = min(msvd2,mdids);
Xdids = Xdids(:,1:m);

for j = 1:m
  Xsvd2(:,j) = interp(Xsvd2(:,j),uf);
end

% Sort the number of columns
[nsvd2 msvd2] = size(Xsvd2);
[ndids mids] = size(Xd2ids);
n = min(nsvd2,ndids);
Xd2ids = Xd2ids(1:n,:);
Xsvd2i = Xsvd2(1:n,:);

% Calculate the PSNR
psnr_2DPCA = PSNR(imscale(Xsvd2i),Xd2ids);
% Similarity measure
[msim_2DPCA, ssim_map_2DPCA] = ssim(imscale(Xsvd2i),Xd2ids);

%% 2D-ICA
% Turn each image into 1D mixture and construct matrix of mixtures
[m n] = size(X1);
X11D = reshape(X1, m*n, 1);
X21D = reshape(X2, m*n, 1);
X31D = reshape(X3, m*n, 1);
X41D = reshape(X4, m*n, 1);
X51D = reshape(X5, m*n, 1);
X61D = reshape(X6, m*n, 1);
Xm = [X11D X21D X31D X41D X51D X61D]'; % mixtures matrix

% Do PCA and plot the eigenvalues to get some idea of independent
% components to be extracted from the mixture
[U,S,pcl] = svd(Xm', 0);
\texttt{eigen = diag(S).^2; figure(8); plot(eigen,'k'); hold on; plot(eigen,'rx'); title('Scree Plot'); xlabel('Eigenvalue Number'); ylabel('Eigenvalue')}

\texttt{%% Un-mix}
\texttt{nu_ICA = 6; \% nu_ICA equals to number of mixed images}
\texttt{W = jadeR(Xm,nu_ICA); ic = (W * Xm)';}
\texttt{\% Reshape and plot individual images figure(9); for k = 1:nu_ICA subplot(ceil(sqrt(nu_ICA)),ceil(sqrt(nu_ICA)),k)
  imagesc(reshape(ic(:,k),m,n));
  colormap(gray);
  title(sprintf('Independent image # %d',k)); end}

\texttt{%% Obtain residual image and plot}
\texttt{XO = imscale(X); \% original image scaled to [0 1]
XO = imscale(X); \% original image scaled to [0 1]
\% Select the independent image containing most of the unwanted features
ICselect = input('Select the independent image that contains most of the unwanted features: ');
\% Reshape and scale to [0 1]
XI1 = imscale(reshape(ic(:,ICselect),m,n));
\% Align along x-axis before obtaining residual image (resizing included)
[XOd, XI1d, displ] = align(XO, XI1);
\% Obtain residual image
Xica2 = imscale(XOd - XI1d);
\% Calculate PSNR for ICA method
psnr_ICA = PSNR(Xica2,XOd);
\% Similarity measure
[mssim_ICA, ssim_map_ICA] = ssim(Xica2,XOd);
\% Display results
format short
disp(' '); disp(' '); disp(' '); if stup == 1
disp(' PSNR values - Big pipe:');
elseif stup == 2
disp(' PSNR values - Medium pipe:');
elseif stup == 3
disp(' PSNR values - Small pipe:');
elseif stup == 4
disp(' PSNR values - All 3 pipes in the box:');
end
disp(' Difference 2D-PCA ICA '); disp('-----------------------------');
disp([psnr_diff psnr_2DPCA psnr_ICA]);
%% Similarity measure display
format short
disp(' ');
disp(' ');
if stup == 1
disp(' M-SSIM values - Big pipe: ');
e elseif stup == 2
disp(' M-SSIM values - Medium pipe: ');
e elseif stup == 3
disp(' M-SSIM values - Small pipe: ');
e elseif stup == 4
disp(' M-SSIM values - All 3 pipes in the box: ');
e end
disp(' Difference   2D-PCA     ICA     ');
disp(' ------------------------------- ');
disp([mssim_diff mssim_2DPCA mssimICA]);

B.3 Built-in Functions Code

B.3.1 mrem.m

function [Xmr] = mrem(Xraw)
% Simple function for mean removal from GPR image
% Input: Xraw - raw image
% Output: Xmr - mean removed image
% June 2012

[m,n] = size(Xraw);
mean_trace = mean(Xraw,2);
Xmr = Xraw - repmat(mean_trace,1,n);

B.3.2 imscale.m

function [Aout] = imscale(Ain)
% Simple function to scale down GPR image
% [Aout] = imscale(Ain)
% Scales data in Ain from any range to 0-1 range.
% Scaled data is stored in Aout.
% June 2012

Amin = min(min(Ain));
Amax = max(max(Ain));
if (Amin < 0)
    Ain = Ain - Amin;
end
Amax = max(max(Ain));
Aout = Ain/Amax;
a) `align.m`

```matlab
function [X1d, X2d, displ] = align(X1, X2)

% Function to align GPR image X1 with GPR image X2 along the
% x-axis.
% Uses 1D correlation of individual image rows to find misalignment
% between the images (i.e. image rows).
% Also, it resizes re-aligned images to the same size.
% July 2012
%
% Output:   X1d - re-aligned and resized image
%           displ - displacement between the images

[m1 n1] = size(X1);
[m2 n2] = size(X2);

% Autocorrelate to find displacement between the images
corsum = 0;
for i = 1:min(m1,m2)
corsum = corsum + xcorr(X1(i,:),X2(i,:));
end

[maxc ind] = max(corsum);
d = ind - max(n1,n2);

% Align, if needed
if d < 0
    X2d = X2(:,1:d);
    X1d = X1;
    displ = -d;
elseif d > 0
    X1d = X1(:,d+1:end);
    X2d = X2;
    displ = d;
else % for d = 0 retain original images
    X1d = X1;
    X2d = X2;
    displ = 0;
end

% Make both images same size, if needed
[m1d n1d] = size(X1d);
[m2d n2d] = size(X2d);

if n2d > n1d
    X2d = X2d(:,1:n1d);
elseif n1d > n2d
    X1d = X1d(:,1:n2d);
end
```

```
Appendix C: Chapter 5

C.1 One-dimensional Analysis Code

% Filename: onedimensional_analysis
% Version: 1
% Analysis of measured or simulated GPR scans using 3 different
% methods:
%         1. PCA - Principal Component Analysis
%         2. LPSVD - Linear Prediction Singular Value Decomposition
%         3. MP - Matrix Pencil Method
% 1. GPR traces are extracted from GPR images and stored into matrix.
% Single trace from image can be selected or a group of successive
% traces.
% 2. Principal components of the stored traces are then obtained and
% extracted traces (or any new set of traces, from new GPR image)
% projected on those principal components.
% In this way dimensionality of selected traces is significantly
% reduced.
% Each trace or set of successive traces is now represented with a
% small number of points.
% Scatter plots are then produced to show different class
% memberships.
% 3. Alternatively, or in addition to this analysis, each trace is
% also analysed by extracting dominant frequencies, corresponding
% amplitudes and damping using MP method. Those points are also
% plotted in 3D space to detect possible clusters or groupings in
% frequency domain or to just observe frequency properties of
% extracted traces.
% BV, NB, December 2012

clear all;
close all;
clc;

%% Set parameters
notraces = 0;    % no. of successive traces to be used in simulation
                % i.e. before and after the selected trace
nsignals = 30;   % maximum number of signals/images to be analysed
unwrappit = 1;   % if 1, concatenate traces from the same image
                 % in a single trace
seltraces = [];  % matrix to hold all selected traces
i = 0;           % current number of analysed signals

ColorSet = varycolor(nsignals);
Cc = hsv(nsignals); % define colormap

%% Select data directory and load images
currentFolder = pwd;  % get current directory
select_data = input('Option Data: ');
if (select_data == 1)
    % ***** Simulated data
    GPRdir = 'C:\GPR\Pipes_GPRmax_Simulations\Output';
    GPRdir = 'C:\GPR\Amir_Simulations\Output';
    cd(GPRdir);
    base_name = '1pipemodelv5_';
    nos = [7 9 11];
    base_name = 'amirmodelv1_';
    colsel = [64 77 60 119];
    elseif (select_data == 2)
    % ***** Measured data
    GPRdir = 'C:\REFLEX\Portsmouth\Portsmouth Data\SRR methods 1';
    % GPRdir = 'C:\REFLEX\Amir Data';
    cd(GPRdir);
    base_name = 'FUN2_';
    base_name = 'WB_';
    nos = [1 5 9];
    colsel = [109 108 107];
end

else if (select_data == 2)
    % ***** Measured data
    GPRdir = 'C:\REFLEX\Portsmouth\Portsmouth Data\SRR methods 1';
    % GPRdir = 'C:\REFLEX\Amir Data';
    cd(GPRdir);
    base_name = 'FUN2_';
    base_name = 'WB_';
    nos = [1 5 9];
    colsel = [109 108 107];
end

% Extract traces
for i = 1:length(nos)
    no = num2str(nos(i));
    load([base_name no '.mat']);
    if (select_data == 1)
        X = ezmr;
    elseif (select_data == 2)
        X = X;
    end
    figure(1);
    imagesc(X);
    colormap(gray);
    title(['Image #' no]);
    pause(2);
    tr = X(:,colsel(i)-notraces:colsel(i)+notraces);
    [m1,n1] = size(tr);
    seltraces = [seltraces tr];
    elseif (select_data == 1)
        M = ezmr;
        elseif (select_data == 2)
        M = M;
end

% Extract group of traces from the image
tr = X(:,colsel(i)-notraces:colsel(i)+notraces);
[m1,n1] = size(tr);

% Resample traces if necessary
if (i == 1)  % if no size at start, take the size of first trace
    [m2,n2] = size(tr);
elseif (i > 1)
    [m2,n2] = size(seltraces);
end

% Check if those sizes are identical
if (m1 == m2)  % use the first m1 samples in each trace
    tr = tr(1:m1,:);
else if (unwrappit == 0)
    % Save wrapped traces and their labels
    seltraces = [seltraces tr];
elseif (unwrappit == 1)
% Or unwrap and save unwrapped traces and their labels
trunwrapped = reshape(tr,m1*n1,1); % all traces in 1 column
seltraces = [seltraces trunwrapped];
end

% Plot unwrapped traces
% Option - comment out to speed up the process
figure(2)
plot(trunwrapped);
grid on;
title('Selected Trace(s)');
pause(2);
end

cd(currentFolder);

%% ------------- ANALYSIS 1 - PCA -------------
% Use PCA to reduce data dimensionality and analyse
figure(3);
imagesc(seltraces);
colormap(gray);
dsmplf = 10;        % downsample factor
% use it when the data is too large for PCA in Matlab
[m,n] = size(seltraces);
% Downsampling might be needed when unwrapped traces are used
input_data = (resample(seltraces,1,dsmplf))';  % downsample and transpose
% After transposition
% rows = observations = individual traces
% columns = variables = time instants/pixels
% Normalise traces
input_data = zscore(input_data);
[coeff, scores, latent] = myprincomp(input_data); % does it all
% Matrix scores contains the projection of input_data onto the PCs
c_var = (cumsum(latent) ./ sum(latent)); % cumulative variance
% for all dimensions
figure(3)
plot(c_var);
hold on;
plot(c_var,'r*');
grid on;
xlabel('PC #');
ylabel('% of data explained');

% Plot first 4 PCs from the set
figure(4)
for i = 1:4
    subplot(2,2,i)
    plot(coeff(:,i));
    grid on;
    title(['PC # ', num2str(i)]);
end

%% Scatter plot for 2D or 3D PCs set
% Different setup will be used different colours and numbers
Y = scores(:,1:3);
[nosig ndim] = size(Y);
% ndim = number of dimensions, plot in 2D or 3D
% nosig = number of signals/points to be plotted
n d i m = 3;
% change here to select 2D or 3D plot
Y = scores(:,1:ndim);
figure(6);
[m, n] = size(Y);
traceslab = 1:m;
T = traceslab;
msize = 15; % marker size
spread(Y(1:nosig,:),'.',T(1:nosig)',msize, 0);
tittle(['First', num2str(ndim) ' PCs indicating different setups']);
xlabel('PC #1');
ylabel('PC #2');
if ndim == 3
    zlabel('PC #3');
end
% Print numbers for easier understanding of the plot
for i = 1:nosig
    if ndim == 2
        text(Y(i,1),Y(i,2)+0,num2str(i),'HorizontalAlignment','center')
    elseif ndim == 3
        text(Y(i,1),Y(i,2)+0,Y(i,3)+0,num2str(i),'HorizontalAlignment','center')
    end
end
% Calculate the distance between each point
D = squareform(pdist(Y));

% % % ------------- ANALYSIS 2 - LPSVD/MP -------------

th = 0.1; % percentage of max signal value used to detect start % and end of the actual signal in the considered trace % obviously - lower "th" = longer part extracted
M = 20; % model order

% Choose the sampling frequency, Fs
if (select_data == 1)
    Fs = 1/dt;
elseif (select_data == 2)
    Fs = 1/3.90625e-11;
end
% Sampling frequency calculation:
% tw = 10e-9; % time window
% dt = tw/m; % time step = time window/number of samples
% Fs = 1/dt = 1/(time window/number of iterations)
% time step or time increment for measured data: 3.90625e-11
[Mtr, Ntr] = size(seltraces);
counter = 0;
paras = [ ]; % 3D matrix to save all extracted parameters
for i = 1:Ntr
    tr = seltraces(:,i); % extract each "multiple" trace
    trm = reshape(tr,m1,n1); % reshape back to single traces
        counter = counter+1;
tr = trm(:,j); \% analyse each trace individually

% Extract portion of the signal to be analysed
TH = (th/100)*max(abs(tr)); \% threshold value
sign_ind = find(abs(tr)>=TH); \% part of the trace contain
% the actual signal
sign = tr(min(sign_ind):max(sign_ind));

% Plot the original trace and extracted,
% “signal present” part of the trace
figure(9)
cif;
plot(tr); hold on; grid on;
plot((min(sign_ind):max(sign_ind)),sign,'r');

% LPSVD/MP analysis
figure(10)
cif
if (analysis == 1)
[hpoles, spoles, para] = lpsvd(sign,Fs,M); \% LPSVD
elseif (analysis == 2)
[hpoles, spoles, para, M] = mp(sign,Fs,M); \% MP
end

% Check the number of estimated parameters
[mpara, npara] = size(para);
% Fix for insufficient number of poles estimated
if(mpara < M/2)
parae = zeros(M/2,4); \% create empty matrix of
para = parae;
end

% Store obtained results - parameter sets
paras(:,:,counter) = para;
pause;
end

\% Plot extracted parameters in 3D parameter space

[mp np kp] = size(paras);

% Extract and reshape into 1D
damps = reshape(paras(:,1,:),mp*kp,1);
freqs = reshape(paras(:,2,:),mp*kp,1);
amps = reshape(paras(:,3,:),mp*kp,1);

% Make matrix of features
Z = [damps freqs amps];

% Make vector of parameter labels for same trace
% (i.e. give parameters from same trace with the same label/number)
trparaslab = [];
for i = 1:kp
trparaslab = [trparaslab; ones(mp,1)*i];
end

% Make vector of parameter labels same for each image
% (i.e. give parameters from same signal with the same
% label/number)
imparaslab = [];
j = 1;
for i = 1:counter:kp*mp
imparaslab = [imparaslab; ones(mp*n1,1)*j];
j = j+1;
end
C.2 Selecting Model Order Code

% Analysis of LPSVD & MP in selecting model order
% Includes:
% 1. The singular entropy in selecting model order.
% 2. The threshold part has been removed from the
% onedimensionalanalysis version 1 in order to select the data
% manually.
% 3. Add the FFT and compare with the amplitude of the extracted
% poles of MP.
% Version: 2 of onedimensionalanalysis
% NB, BV 12/02/2013

clear all;
close all;
clc;

%% Set parameters and select data directory
notraces = 0; % no. of successive traces to use in simulation
% i.e. before and after the selected trace
nsignals = 30; % maximum number of signals/images to be analysed
unwrappit = 1; % if 1, concatenate traces from the same image
% in a single trace
seltraces = []; % matrix to hold all selected traces
i = 0; % current number of analysed signals
ColorSet = varycolor(nsignals);% define colormap
Cc = hsv(nsignals); %

%% Select data directory and load images
currentFolder = pwd; % get current directory
select_data = input('Option Data: ');  
if (select_data == 1)  
    % ***** Simulated data  
    GPRdir = 'C:\GPR\Pipes_GPRmax_Simulations\Output';  
    GPRdir = 'C:\GPR\Amir_Simulations\Output';  
    cd(GPRdir);  
    base_name = '1pipemodelv5_';  
    nos = [7 9 11];  
    colsel = [64 77 60 119];  
elseif (select_data == 2)  
    % ***** Measured data  
    GPRdir = 'C:\REFLEX\Portsmouth\Portsmouth Data July';  
    GPRdir = 'C:\REFLEX\Amir Data';  
    base_name = 'FUN2_';  
    nos = [5];  
    colsel = [119];  
end

% Extract traces  
for i = 1:length(nos)  
    no = num2str(nos(i));  
    load([base_name no '.mat']);  
    if (select_data == 1)  
        X = ezmr;  
    elseif (select_data == 2)  
        X = X;  
    end
    % Display image  
    figure(1);  
    imagesc(X);  
    colormap(gray);  
    title(['Image #' no]);  
    pause(2);  
    % Plot selected trace  
    figure(2)  
    plot(X(:,colsel(i)));  
    grid on;  
    title('Selected Trace(s)');  
    % Extract group of traces from the image  
    Xn = input('Select data samples [xn1, xn2]: ');  
    tr = X(Xn(1):Xn(2),colsel(i):notraces:colsel(i)+notraces);  
    [m1,n1] = size(tr);  
    % Resample traces if necessary  
    % Get the size of matrix to hold selected traces  
    if(i == 1)  
        [m2,n2] = size(tr);  
    elseif(i > 1)  
        [m2,n2] = size(seltraces);  
    end
    % Check if those sizes are identical  
    % Sort if they are unequal  
    if(~(m1 == m2) && (i > 1))  
        tr = tr(1:m1,:);  
    end

252
Appendix

% Save traces
if (unwrappit == 0)
    % Save wrapped traces and their labels
    seltraces = [seltraces tr];
else if (unwrappit == 1)
    % Or unwrap and save unwrapped traces and their labels
    trunwrapped = reshape(tr,m1*n1,1);  % all traces in one column
    seltraces = [seltraces trunwrapped];
end

% Plot unwrapped traces
% Option - comment out to speed up the process
figure(3)
plot(trunwrapped);
grid on;
title('Selected Trace(s)');
pause(2);
end
cd(currentFolder);

%% ------------- ANALYSIS 2 - LPSVD/MP -------------
% Choose the sampling frequency, Fs
if (select_data == 1)
    Fs = 1/dt;
elseif (select_data == 2)
    Fs = 1/3.90625e-11;
end

% Sampling frequency calculation:
% tw = 10e-9;  % time window
% dt = tw/m;  % time step = time window/number of samples
% Fs = 1/dt = 1/(time window/number of iterations)
% time step or time increment for measured data: 3.90625e-11
[Mtr,Ntr] = size(seltraces);

% 3D matrix to save all extracted parameters


for i = 1:Ntr
    tr = seltraces(:,i);        % extract each "multiple" trace
    trm = reshape(tr,m1,n1);    % reshape back to single traces

    % Singular Entropy Increment and Differential of Singular
    % Entropy Increment for simulated signal
    [dE dEp] = sen(tr);

    n = 0:80;
    figure(7)
    plot(n,dE(1:81));
    grid on;
    xlabel('n');
    ylabel('Singular Entropy Increment');

    figure(8)
    plot(n,dEp(1:81));
    grid on;
    xlabel('n');
    ylabel('Differential of Singular Entropy Increment');

    for j = 1:n1
        M = input('Select model order: ');

end
counter = counter+1;
tr = trm(:,j);            % analyse each trace individually

% MP analysis
figure(10)
clf
if (analysis == 1)
    [hpoles, spoles, para] = lpsvd(sign,Fs,M);  % LPSVD
elseif (analysis == 2)
    [hpoles, spoles, para, M] = mp(tr,Fs,M);    % MP
end

% Check the number of estimated parameters
[mpara, npara] = size(para);
% Fix for insufficient number of poles estimated
if(mpara < M/2)
    parae = zeros(M/2,4);  % create empty matrix of parameters of proper size
    parae(1:mpara,:) = para;
    para = parae;
end

[p1 p2] = size(para);
[p3 p4] = size(paras);
if(p1 < p3)
    parase = zeros(p3,4);
    parase(1:p1,:) = para;
    para = parase;
end

% Store obtained results - parameter sets
paras(:,:,counter) = para;
pause;
end

%% Plot extracted parameters in 2D parameter space

[mp np kp] = size(paras);

% Extract and reshape into 1D
damps = reshape(paras(:,1,:),mp*kp,1);
freqs = reshape(paras(:,2,:),mp*kp,1);
amps = reshape(paras(:,3,:),mp*kp,1);

% Make matrix of features
Z = [damps freqs amps];

% Make vector of parameter labels for same trace
% (i.e. give parameters from same trace with same label/number)
trparaslab = [];
for i = 1:kp
    trparaslab = [trparaslab; ones(mp,1)*i];
end

% Make vector of parameter labels same for each image
% (i.e. give parameters from same signal with same label/number)
imparaslab = [];
for i = 1:counter:kp*mp
    imparaslab = [imparaslab; ones(mp*n1,1)*i];
    j = j+1;
end

[z1 z2] = size(Z);
T = imparaslab(1:z1,1);
msize = 10;                     % marker size
Appendix

\[ Z_{\text{dim}1} = \{\text{freqs, damps}\}; \]
\[ \text{figure}(12); \]
\[ \text{spread}(Z_{\text{dim}1}',T',\text{msize},1); \]
\[ \text{title}('Extracted 2D parameters indicating different setups'); \]
\[ \text{xlabel}('Frequency'); \]
\[ \text{ylabel}('Damping'); \]

\[ Z_{\text{dim}2} = \{\text{freqs, amps}\}; \]
\[ \text{figure}(13); \]
\[ \text{spread}(Z_{\text{dim}2}',T',\text{msize},1); \]
\[ \text{title}('Extracted 2D parameters indicating different setups'); \]
\[ \text{xlabel}('Frequency'); \]
\[ \text{ylabel}('Amplitude'); \]

% Power Spectrum of the signal using FFT
\[ nfft = 2^{15}; \]
\[ X = \text{fft}(x,nfft); \]
\[ P_{yy} = X.*\text{conj}(X); \]
\[ P_{yy\_crop} = P_{yy}[1:nfft/2]; \]
\[ f = 0 : Fs/(nfft-1) : Fs/2; \]
\[ \% Plot the FFT and amplitude of LPSVD/MP on the same figure \]
\[ \text{figure}(14); \]
\[ \text{plot}(f,P_{yy\_crop}/\text{max}(P_{yy\_crop}),'b','\text{LineWidth}',2); \]
\[ \text{hold on}; \]
\[ \text{grid on}; \]
\[ \text{plot}(\text{freqs},\text{amps}/\text{max}(\text{amps}),'r','\text{LineWidth}',2); \]
\[ \text{title}('Power Spectrum of the Signal'); \]
\[ \text{xlabel}('Frequency (Hz)'); \]
\[ \text{ylabel}('Power Spectrum (db)'); \]

C.3 Built-in Functions Code

C.3.1 myprincomp.m

\begin{verbatim}
function [coeff, score, latent] = myprincomp(x)
  \% MYPRINCOMP Principal Components Analysis
  \% Input:
  \%   x - data matrix (rows are observations, columns are variables)
  \%   rows = observations = individual traces
  \%   columns = variables = time instants/pixels
  \% Output:
  \%   coeff - principal component axis
  \%   score - projection of data matrix onto principal component axis (rows are observations, columns are components)
  \%   latent - principal component variances, i.e. eigenvalues of the covariance matrix of X
  \% Center x by removing column means
  [n,p] = size(x);
  x0 = x - repmat(mean(x,1),n,1);
  \% Calculate principal component coefficients
  \% i.e. eigenvectors of S = x0'*x0/(n-1), but use SVD.
  [U,sigma,coeff] = svd(x0,0); \% put in 1/sqrt(n-1) later
  \% Project x0 onto the principal component axes to get the scores.
  sigma = diag(sigma);
  score = x0*coeff;
  sigma = sigma ./ sqrt(n-1);
\end{verbatim}
Appendix

% Calculate the variances of the principal components
% i.e. calculate the eigenvalues of S = x0'x0./(n-1)

% latent = sigma.^2;

C.3.2 spread.m

function spread(X, label, ms, fill)
% Plot samples of different labels with different colors.
% Written by Michael Chen (sth4nth@gmail.com).
% Modified by Branislav Vuksanovic (boban1962@gmail.com)
% X - data
% label - data labels
% ms - marker size
% fill - fill in the markers (1) or not (0)
[d, n] = size(X);
if nargin == 1
    label = ones(n,1);
end
assert(n == length(label));
c = max(label);
ColorSet = varycolor(c);
set(gca, 'ColorOrder', ColorSet);
hold all;
figure(gcf);
cf;
hold on;
switch d
    case 2
        view(2);
        for i = 1:c
            idc = label == i;
            if fill == 0
                plot(X(1,idc),X(2,idc), 'o','Color', ... 
                    ColorSet(i,:), 'MarkerSize', ms);
            elseif fill == 1
                plot(X(1,idc),X(2,idc),'o', 'MarkerEdgeColor','k', ... 
                    'MarkerFaceColor', ColorSet(i,:), 'MarkerSize', ms);
            end
        end
    case 3
        view(3);
        for i = 1:c
            idc = label == i;
            if fill == 0
                plot3(X(1,idc),X(2,idc),X(3,idc), 'o','Color', ... 
                    ColorSet(i,:), 'MarkerSize', ms);
            elseif fill == 1
                plot3(X(1,idc),X(2,idc),X(3,idc),'o', 'MarkerEdgeColor','k', ... 
                    'MarkerFaceColor', ColorSet(i,:), 'MarkerSize', ms);
            end
        end
    otherwise
        error('ERROR: only support data of 2D or 3D.');
end
axis equal
grid on
hold off
C.3.3 sortpoles.m

function poles = sortpoles(para, criteria, nf)
% Analyse obtained parameter set according to one
% of defined criteria and produce a set of poles.
% Inputs:     para - matrix of obtained parameters
% criteria - selection criteria
% nf - number of poles to retain [for criteria 1]
%       - percentage of highest amplitude to retain [for criteria 2]
% criteria 1 - select nf poles corresponding
% to nf highest amplitudes in the set
% criteria 2 - select all poles with amplitudes higher than
% nf% of highest amplitude in the set
% Output:     poles - set of selected poles

if para(:,1) < 0
    para(:,1) = -para(:,1);
else
    para(:,1) = para(:,1);
end
[para1 ind] = sort(para(:,3));   % sort according to amplitudes
paras = para(flipud(ind),:); % parameters sorted in descending order

% Selecting poles according to criteria 1
% i.e. nf poles with highest amplitudes
if criteria == 1
    parass = paras(1:nf,:);     % selected parameters
    poles = parass(:,1:2);      % select poles
end

% Selecting poles according to criteria 2
% i.e. all poles with amplitudes higher than
% nf% of the highest amplitude
if criteria == 2
    lim = max(paras(:,3))*nf/100; % find max amplitude to retain
    ind = find(paras(:,3)>lim);   % select poles
    poles = paras(ind,1:2);        % select poles
end